Online Adaptive Deep Learning for Multi-Sensor WiFi Indoor Localization

Afnan Ahmad

A Dissertation Submitted to the Faculty of Graduate Studies
In Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

Graduate Program in Earth and Space Science and Engineering

York University
Toronto, Ontario

August 2026

©Afnan Ahmad, 2026

Abstract

Indoor positioning remains challenging because WiFi signals are inherently unstable, degrading as building infrastructure changes. Traditional fingerprinting requires periodic re-surveys to maintain accuracy, limiting practical deployment. This thesis investigates adaptive indoor localization through four contributions. The two major contributions are a longitudinal multimodal benchmark and a label-free online reliability adaptation mechanism validated over real temporal drift. The augmented reality ground-truth method and the network architecture are enabling contributions that make these two possible.

First, an augmented reality framework using Microsoft HoloLens 2 enables automated, centimeter-accurate ground truth collection for WiFi fingerprinting, decoupling collection from runtime positioning and scaling across buildings without permanent infrastructure.

Second, the framework produces a longitudinal multi-building dataset spanning four buildings at York University (the Petrie Science and Engineering Building (PSE), Bergeron Centre for Engineering Excellence (BCE), Student Teacher Centre (STC), and Lassonde Research Centre (LRC)), covering 28,822 m^2 with 847 reference points and 239 access points and combining WiFi, inertial measurement unit (IMU), and magnetometer modalities collected over 2021–2022. This scope enables systematic study of how localization accuracy degrades across seasons, years, and building types.

Third, a VAE-LSTM architecture combines variational autoencoders with long short-term memory networks to jointly model spatial uncertainty and temporal dynamics. The variational encoder produces a probabilistic latent representation capturing WiFi signal variability, while the LSTM component models sequential movement patterns from 20-second input windows of WiFi, IMU, and magnetometer readings.

Fourth, the Online Feature Adaptation Module (OFAM) provides reliability-weighted sensor fusion in which the fusion weights are adapted continuously during deployment while the network parameters remain fixed. OFAM dynamically adjusts the fusion weights of WiFi, IMU, and magnetometer inputs from streaming sensor data, increasing inertial contribution during WiFi degradation, without retraining the network or requiring labeled position updates.

Experimental validation demonstrates 0.54 m mean absolute error with OFAM across the four York University buildings (0.59 m on the combined evaluation set including the independent IPIN environments), representing 68% improvement over WiFi-only baselines. The system achieves lower mean error than kNN, CNN, DeepML, and standard LSTM approaches while maintaining 5 ms adaptation overhead suitable for real-time deployment. Cross-year evaluation on the York University testbed shows only 4% accuracy degradation compared to 22% without adaptation. The same benefit reproduces on the independent IPIN 2016–2017 data, against which OFAM was never tuned, holding its cross-year degradation to 6% against 29% for the static model.

Dedication

The strongest guiding light over these past few years
has been the many deep conversations and gentle redirections
my wife, Talia, offered me in every moment I doubted myself.
I am deeply indebted to her for her unparalleled support
and patience throughout this journey.
To my Lamia, who truly is the light of my life.
And to my parents, for supporting me in pursuing my education.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Gunho Sohn, whose guidance, patience, and unwavering support shaped every stage of this work. Their insight and encouragement consistently challenged me to think more rigorously, and their mentorship extended well beyond the technical substance of this thesis. I am profoundly grateful for the time and care they invested in my development as a researcher.

I am also sincerely thankful to the members of my supervisory committee, Dr. Manos Papagelis and Dr. Mojgan Jadidi, for their thoughtful feedback and constructive direction throughout this journey. Their probing questions and critical perspectives strengthened the arguments presented here and helped me situate my research on indoor localization and augmented reality within a broader scholarly context.

I would further like to thank my external examiner, Dr. Bilal Farooq, and my internal examiner, Dr. Hina Tabassum, for the time and careful reading devoted to this dissertation. The comments and questions raised at the oral examination strengthened the final version of this work. I am also grateful to Dr. Regina Lee for chairing the examination.

On a personal note, I owe more than I can express to my wife, whose patience, sacrifice, and steadfast support carried me through the most demanding years of this work. None of this would have been possible without her, and for that I will always be grateful. To my daughter, I am thankful for the joy she brings to each day and for the quiet motivation that her presence has given me throughout this endeavor. I am deeply grateful, too, to my parents and siblings, whose constant support and belief in me sustained me through the hardest moments of this undertaking, and to my uncle and aunt, whose generous support of my education helped make this path possible in the first place. I also wish to thank my mother-in-law for her unfailing support and for the many delicious meals that nourished both body and spirit along the way.

I am equally grateful to my friends and to the lab members who became friends, for the encouragement, good humor, and moral support that lightened the weight of these years.

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) through the Discovery Grant program and by the NSERC CREATE program in Data Analytics and Visualization (DAV). Additional support was provided by the Ontario Research Fund, Research Excellence (ORF-RE) through the Intelligent Systems for Sustainable Urban Mobility (ISSUM) project. The author gratefully acknowledges the financial support of these funding agencies, without which this work would not have been possible.

Table of Contents

Page

Abstract ii
Dedication iii
Acknowledgements iv
Table of Contents v
List of Tables xii
List of Figures xiv
List of Publications and Manuscripts xv
List of Abbreviations xvi
List of Symbols xviii

1 Introduction **1**
1.1 Background and Motivations 1
1.2 Problem Statement 3
1.2.1 Terminology 4
1.3 Research Objectives 5
1.3.1 Primary Objective 5
1.3.2 Secondary Objectives 5
1.4 Research Questions 6
1.5 Scope and Limitations 7
1.5.1 Scope of Research: 7
1.5.2 Limitations and Constraints: 8
1.6 Methodology Overview 9
1.7 Contribution of Thesis 11
1.8 Structure of Thesis 12

2 Literature Review **14**
2.1 Introduction 14
2.2 Fundamentals of Indoor Positioning 15
2.2.1 Positioning Techniques Overview 15
2.2.2 Fingerprinting Foundations 17

2.2.3 Cellular Positioning and Channel State Information . . . 18
2.3 WiFi Signal Instability and Variability . . . 18
2.3.1 Sources of Signal Variability . . . 19
2.3.2 Channel State Information (CSI) Approaches . . . 19
2.3.3 Addressing Device Heterogeneity . . . 20
2.3.4 Gap: Environmental Adaptation Mechanisms . . . 21
2.4 Temporal Drift in Indoor Positioning . . . 21
2.4.1 Manifestations of Temporal Degradation . . . 21
2.4.2 Online Learning for Adaptation . . . 22
2.4.3 Lack of Longitudinal Datasets . . . 24
2.4.4 Gap: Multi-Year Dataset Rarity . . . 25
2.5 Multi-Sensor Fusion Approaches . . . 26
2.5.1 Sensor Modalities for Indoor Positioning . . . 26
2.5.2 Fusion Architectures . . . 26
2.5.3 Hierarchical and Sequential Fusion . . . 27
2.5.4 Gap: Static Fusion Weights . . . 28
2.6 AR-Based Indoor Localization . . . 28
2.6.1 AR Tracking Technologies . . . 28
2.6.2 AR for Ground Truth Establishment . . . 29
2.6.3 Gap: AR for Fingerprint Database Creation . . . 29
2.7 Deep Learning Architectures for Localization . . . 29
2.7.1 Spatial Feature Extraction . . . 29
2.7.2 Temporal Sequence Modeling . . . 30
2.7.3 Probabilistic Feature Encoding . . . 30
2.7.4 Recent Fingerprinting Architectures . . . 30
2.7.5 Transformer Architectures . . . 31
2.7.6 Uncertainty Quantification in Deep Localization . . . 33
2.7.7 Gap: Probabilistic Temporal Architectures . . . 35
2.8 Research Gaps and Challenges . . . 35
2.8.1 Comparison of Representative Systems . . . 36
2.8.2 Additional Challenges . . . 36
2.8.3 Research Gap Summary . . . 37
2.9 Summary . . . 38

3 Dataset and AR Framework **39**
3.1 Introduction . . . 39
3.2 AR-Based Data Collection Framework . . . 41
3.2.1 HoloLens 2 Hardware Platform . . . 41
3.2.2 Spatial Anchor System . . . 44
3.2.3 Spatial Anchor Validation . . . 45

3.2.4 Data Collection Pipeline . . . 47
3.3 Multi-Modal Sensor Acquisition . . . 49
3.3.1 WiFi Signal Collection . . . 50
3.3.2 WiFi Access Point Density Characterization . . . 51
3.3.3 Inertial Measurement Unit . . . 53
3.3.4 Magnetometer Data . . . 55
3.3.5 Magnetic Field Characterization . . . 56
3.4 Four-Building Testbed . . . 58
3.5 Longitudinal Dataset . . . 62
3.5.1 Data Collection Timeline . . . 66
3.5.2 Data Collection Campaign Details . . . 68
3.5.3 Ethics and Privacy . . . 70
3.6 Preprocessing Pipeline . . . 71
3.6.1 WiFi Fingerprint Processing . . . 72
3.6.2 IMU Feature Extraction . . . 73
3.6.3 Magnetometer Calibration . . . 74
3.6.4 Feature Vector Representation . . . 75
3.7 Deployment Procedure for a New Site . . . 76
3.8 Baseline Algorithm Comparison . . . 80
3.8.1 Evaluation Protocol . . . 80
3.8.2 Algorithm Categories . . . 81
3.8.3 Comparative Findings . . . 83
3.8.4 Cross-Year Temporal Stability . . . 83
3.8.5 Analysis and Selection Rationale . . . 86
3.9 Data and Code Availability . . . 89
3.10 Chapter Summary . . . 89

4 Probabilistic Temporal Architecture for Multi-Sensor Localization 91
4.1 Introduction . . . 91
4.1.1 Chapter Organization . . . 91
4.1.2 Design Rationale . . . 93
4.2 Problem Formulation . . . 93
4.2.1 Sensor Input Representation . . . 93
4.2.2 Output Specification . . . 94
4.2.3 Optimization Objective . . . 94
4.3 VAE Encoder . . . 95
4.3.1 Evidence Lower Bound Derivation . . . 96
4.3.2 Encoder Architecture . . . 98
4.3.3 Reparameterization Trick . . . 99
4.3.4 Probabilistic Encoding . . . 101

4.3.5 Design Rationale . . . 101
4.3.6 Robustness to Measurement Noise and Alternative Generative Encoders . . . 102
4.3.7 Latent Space Analysis . . . 102
4.4 LSTM Temporal Encoder . . . 104
4.4.1 Architecture . . . 105
4.4.2 LSTM Cell Dynamics . . . 105
4.4.3 Spatial Context Design . . . 107
4.4.4 Regularization Techniques . . . 108
4.4.5 Design Rationale: Why LSTM . . . 108
4.5 Dual-Branch Decoder . . . 110
4.5.1 Probabilistic Interpretation . . . 110
4.5.2 Position Estimation Branch . . . 111
4.5.3 Reconstruction Branch . . . 112
4.5.4 Uncertainty Quantification . . . 112
4.6 Training Methodology . . . 114
4.6.1 Loss Function . . . 115
4.6.2 Loss Weighting . . . 116
4.6.3 Training Procedure . . . 117
4.6.4 Regularization . . . 119
4.6.5 Hyperparameter Summary . . . 120
4.6.6 Hyperparameter Sensitivity Analysis . . . 120
4.7 Base Model Performance . . . 124
4.7.1 Evaluation Protocol . . . 124
4.7.2 Statistical Reporting . . . 125
4.7.3 Overall Accuracy . . . 125
4.7.4 Per-Building Breakdown . . . 126
4.7.5 Error Distribution Analysis . . . 128
4.7.6 Comparison with Chapter 3 Baselines . . . 129
4.7.7 Component Contribution Analysis . . . 129
4.8 Chapter Summary . . . 132
4.8.1 Key Contributions . . . 134
4.8.2 Limitations . . . 134
4.8.3 Implications for Practical Deployment . . . 135

5 Online Feature Adaptation for Multi-Sensor Indoor Localization 137
5.1 Introduction . . . 137
5.1.1 System Overview . . . 137
5.1.2 Chapter Contributions . . . 139
5.1.3 Chapter Organization . . . 140
5.2 Problem Formulation . . . 140

5.2.1 Extension from Static to Adaptive Formulation . . . 140
5.2.2 Extended Loss Formulation . . . 141
5.3 Multi-Sensor Fusion Framework . . . 142
5.3.1 Sensor Data Preprocessing and Synchronization . . . 143
5.3.2 Hierarchical Fusion Architecture . . . 143
5.3.3 Parameterized Fusion Equation . . . 145
5.3.4 Fusion Algorithm . . . 146
5.4 Sensor Reliability Assessment . . . 148
5.4.1 WiFi RSSI Stability Metrics . . . 148
5.4.2 IMU Drift and Motion Consistency . . . 149
5.4.3 Magnetometer Field Disturbance . . . 150
5.4.4 Combined Reliability Vector . . . 151
5.5 Online Feature Adaptation Module . . . 152
5.5.1 Design Rationale . . . 152
5.5.2 Optimization Formulation . . . 154
5.5.3 Adaptive Weight Update Mechanism . . . 156
5.5.4 OFAM Algorithm . . . 157
5.5.5 Empirical Convergence Behavior . . . 158
5.6 Experimental Validation . . . 160
5.6.1 Ablation Study . . . 160
5.6.2 Overall Localization Accuracy . . . 161
5.6.3 Cross-Year Validation . . . 163
5.6.4 Per-Building Results . . . 164
5.6.5 Computational Performance . . . 164
5.6.6 Sensor Contribution Analysis . . . 165
5.6.7 Real-Time Weight Adaptation . . . 168
5.6.8 Comparison with Modern State-of-the-Art Baselines . . . 168
5.6.9 Cross-Dataset Generalization . . . 170
5.7 Extended Experimental Validation . . . 172
5.7.1 Weight Evolution Analysis . . . 173
5.7.2 Reliability Metric Validation . . . 174
5.7.3 Failure Mode Analysis . . . 175
5.7.4 Factors Affecting Reliability . . . 177
5.7.5 Long-term Deployment Simulation . . . 178
5.7.6 OFAM Convergence Analysis . . . 179
5.7.7 Empirical Stability Analysis . . . 180
5.8 Chapter Summary . . . 183

6 Conclusions and Future Work **188**
6.1 Summary of Contributions . . . 188
6.1.1 AR-Based Ground Truth Methodology . . . 189
6.1.2 Longitudinal Multi-Building Dataset . . . 190
6.1.3 VAE-LSTM Hybrid Architecture . . . 190
6.1.4 Online Feature Adaptation Module . . . 191
6.2 Answers to Research Questions . . . 192
6.3 Limitations . . . 195
6.3.1 Data Limitations . . . 195
6.3.2 Methodology Limitations . . . 195
6.3.3 Evaluation Limitations . . . 196
6.3.4 Generalization Limitations . . . 196
6.3.5 Operational Limitations of WiFi-Based Localization . . . 197
6.3.6 Security Considerations for Deployment . . . 198
6.4 Future Work . . . 200
6.4.1 Extending the Dataset Contribution . . . 200
6.4.2 Extending the VAE-LSTM Architecture . . . 201
6.4.3 Extending the OFAM Contribution . . . 202
6.4.4 Deployment and Scaling Considerations . . . 202
6.5 Closing Remarks . . . 203

Bibliography **204**

A Supplementary Material **216**
A.1 Algorithm Pseudocode . . . 216
A.1.1 VAE-LSTM Training Procedure . . . 216
A.1.2 OFAM Weight Update Algorithm . . . 218
A.1.3 Sensor Fusion Position Estimation . . . 220
A.2 Sensor Data Formats . . . 220
A.2.1 WiFi Fingerprint Format . . . 220
A.2.2 IMU Data Format . . . 222
A.2.3 Magnetometer Data Format . . . 222
A.2.4 AR Ground Truth Format . . . 223
A.2.5 Synchronized Multi-Modal Record . . . 223
A.3 Hyperparameter Configurations . . . 224
A.3.1 VAE-LSTM Hyperparameters . . . 224
A.3.2 OFAM Hyperparameters . . . 225
A.3.3 Baseline Model Hyperparameters . . . 226
A.4 Dataset Statistics . . . 227
A.4.1 Per-Building Summary . . . 227

A.4.2 Temporal Distribution 228
A.4.3 Train/Test Splits 228
A.5 Data Collection Campaign Log 229

List of Tables

2.1 Longitudinal and multimodal dataset comparison 24
2.2 Indoor localization system comparison 36
2.3 Research gaps and thesis contributions 37

3.1 HoloLens 2 sensor hardware specifications 43
3.2 Four-building testbed characteristics 58
3.3 Longitudinal dataset statistics by building 64
3.4 Localization algorithm hyperparameter configurations 84

4.1 Loss weighting coefficients 116
4.2 VAE-LSTM hyperparameter specification 121
4.3 Base model localization accuracy comparison 125
4.4 Per-building static VAE-LSTM accuracy 126
4.5 VAE-LSTM component ablation results 130
4.6 VAE-LSTM layer-by-layer architecture specification 133

5.1 Ablation study of fusion and adaptation 161
5.2 Cross-year accuracy evaluation 163
5.3 Per-building adaptive fusion accuracy 164
5.4 Computational time breakdown 165
5.5 Sensor-combination ablation per environment 168
5.6 Comparison with baseline and state-of-the-art methods 170
5.7 Per-environment localization error for all methods 171
5.8 OFAM failure mode analysis 176
5.9 Scope of reliability against threat classes 187
5.10 OFAM convergence times across buildings 187

6.1 Thesis contribution summary 192

A.1 WiFi fingerprint record format 221
A.2 Sample WiFi fingerprint record 221
A.3 IMU sensor record format 222
A.4 Magnetometer record format 222

A.5 AR ground truth record format 223
A.6 Synchronized multi-modal record format 224
A.7 VAE-LSTM hyperparameter configuration 225
A.8 OFAM hyperparameter configuration 226
A.9 Baseline model hyperparameters 227
A.10 Per-building dataset statistics 228
A.11 Data collection distribution by campaign 228
A.12 Data collection campaign log 229
A.13 Hardware and software inventory 230

List of Figures

2.1 Indoor positioning technology timeline 16
2.2 Sensor fusion architecture taxonomy 27

3.1 AR framework chapter roadmap 40
3.2 AR-enhanced localization system architecture 49
3.3 WiFi infrastructure changes across collection epochs 54
3.4 Building floor plans with AR anchor placement 59
3.5 Coverage heatmaps for 2021 and 2022 campaigns 65
3.6 Two-year data collection timeline 66
3.7 Multi-modal preprocessing pipeline 71
3.8 Multi-modal sensor fusion workflow 77
3.9 Per-building VAE-LSTM accuracy across collection years 85
3.10 Cross-year validation accuracy matrix 86
3.11 Localization error CDF across collection years 87
3.12 Feature stability by sensor modality 88

4.1 VAE-LSTM architecture overview 92
4.2 VAE latent space building-level clustering 103
4.3 Dual-branch decoder architecture 110
4.4 Training and validation loss curves for VAE-LSTM training 119
4.5 Hyperparameter sensitivity analysis 122
4.6 Per-building localization error distributions 127

5.1 OFAM dual-branch adaptation architecture 138
5.2 Online feature adaptation workflow 159
5.3 Localization error CDF comparison 162
5.4 Computational performance breakdown 166
5.5 Sensor-combination accuracy comparison 167
5.6 Real-time fusion weight adaptation (simulated) 169
5.7 Fusion weight evolution across buildings during OFAM adaptation 185
5.8 Validation of reliability metrics against actual positioning accuracy 186
5.9 90-day OFAM deployment stability simulation 186

List of Publications and Manuscripts

The following publications and manuscripts under review report work that forms part of this thesis. Each entry is listed under the chapter in which the corresponding material appears. Where a publication and a chapter differ in scope, the chapter presents the more complete treatment.

Chapter 3: Dataset and AR Framework

[1] A. Ahmad, P. Claudio, A. Alizadeh Naeini, and G. Sohn, "Wi-Fi RSS Fingerprinting for Indoor Localization Using Augmented Reality," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. V-4-2020, pp. 57–64, XXIV ISPRS Congress, 2020. This paper introduced the augmented reality ground-truth method that Chapter 3 extends to a multi-building, multi-modal, two-year framework.

[2] A. Ahmad and G. Sohn, "Multi-Year Evaluation of AR-Based Indoor Fingerprinting Using HoloLens 2 Virtual Anchors and Android Sensor Data," *ISPRS International Journal of Geo-Information*, under review. This paper reports the longitudinal dataset and the cross-year evaluation protocol of Chapter 3.

Chapter 4: Probabilistic Temporal Architecture for Multi-Sensor Localization; Chapter 5: Online Feature Adaptation for Multi-Sensor Indoor Localization

[3] A. Ahmad and G. Sohn, "Deep Adaptive Network for WiFi-Based Indoor Localization," in *Recent Advances in 3D Geoinformation Science: Proceedings of the 18th 3D GeoInfo Conference*, Lecture Notes in Geoinformation and Cartography, Springer, 2024, pp. 617–631. This paper presented the initial VAE-LSTM design and the inference-time feature refinement that Chapter 4 and Chapter 5 develop into the final architecture and the Online Feature Adaptation Module.

[4] A. Ahmad and G. Sohn, "Online Feature Adaptation for Multi-Sensor Indoor Localization," *IEEE Journal of Indoor and Seamless Positioning and Navigation*, under review. This paper presents the VAE-LSTM architecture of Chapter 4 and the Online Feature Adaptation Module of Chapter 5.

List of Abbreviations

AoA	Angle of Arrival
AP	Access Point
AR	Augmented Reality
BCE	Bergeron Centre for Engineering Excellence
BLE	Bluetooth Low Energy
BPTT	Backpropagation Through Time
CAD	Computer-Aided Design
CDF	Cumulative Distribution Function
CNN	Convolutional Neural Network
CSI	Channel State Information
DNN	Deep Neural Network
DOF	Degrees of Freedom
EKF	Extended Kalman Filter
ELBO	Evidence Lower Bound
EMA	Exponential Moving Average
FTM	Fine Timing Measurement
GPS	Global Positioning System
GRU	Gated Recurrent Unit
HPU	Holographic Processing Unit
HVAC	Heating, Ventilation, and Air Conditioning
IMU	Inertial Measurement Unit
IoT	Internet of Things
KKT	Karush-Kuhn-Tucker
KL	Kullback-Leibler (Divergence)
kNN	k-Nearest Neighbors
LiDAR	Light Detection and Ranging
LOCO	Leave-One-Component-Out
LRC	Lassonde Research Centre
LSTM	Long Short-Term Memory
MAC	Medium Access Control

MAE	Mean Absolute Error
MEMS	Micro-Electro-Mechanical Systems
MIMO	Multiple-Input Multiple-Output
ML	Machine Learning
MLP	Multilayer Perceptron
MRTK	Mixed Reality Toolkit
MSE	Mean Squared Error
NLOS	Non-Line-of-Sight
NTP	Network Time Protocol
OFAM	Online Feature Adaptation Module
OFDM	Orthogonal Frequency-Division Multiplexing
PDR	Pedestrian Dead Reckoning
POI	Point of Interest
PSE	Petrie Science and Engineering
RBF	Radial Basis Function
ReLU	Rectified Linear Unit
RF	Radio Frequency
RFID	Radio-Frequency Identification
RNN	Recurrent Neural Network
RSS	Received Signal Strength
RSSI	Received Signal Strength Indicator
RTT	Round-Trip Time
SLAM	Simultaneous Localization and Mapping
STC	Student Teacher Centre
SVM	Support Vector Machine
TDoA	Time Difference of Arrival
ToA	Time of Arrival
t-SNE	t-distributed Stochastic Neighbor Embedding
UKF	Unscented Kalman Filter
UMAP	Uniform Manifold Approximation and Projection
UWB	Ultra-Wideband
VAE	Variational Autoencoder
VIO	Visual-Inertial Odometry
WiFi	Wireless Fidelity
WLAN	Wireless Local Area Network

List of Symbols

Indexing and Dimensions

K	Number of sensor modalities ($K = 3$: WiFi, IMU, magnetometer)
i	Sensor modality index, $i \in \{1, 2, \ldots, K\}$
t	Discrete time index at the WiFi sampling rate (approximately 0.5 Hz)
T	Sequence length / temporal context window ($T = 10$)
N	Number of training samples
B	Batch size ($B = 128$)
D	Input sensor feature dimensionality ($D = N_w + 9 = 109$)
N_w	Number of WiFi RSSI access points, zero-padded to $N_w = 100$

Sensor Inputs and Aggregated Measurements

$\mathbf{s}_t$	Concatenated multi-modal sensor vector at time t ($\in \mathbb{R}^D$)
$\mathbf{S}_t$	Input sequence of T consecutive sensor vectors ($\in \mathbb{R}^{T \times D}$)
$\mathbf{r}_t$	WiFi RSSI measurement vector ($\in \mathbb{R}^{N_w}$)
$\mathbf{a}_t$	Accelerometer reading ($\in \mathbb{R}^3$)
$\boldsymbol{\omega}_t$	Gyroscope (angular velocity) reading ($\in \mathbb{R}^3$)
$\mathbf{m}_t$	Magnetometer reading ($\in \mathbb{R}^3$)
$\mathbf{u}_t$	Raw IMU measurement vector at time t (accelerometer and gyroscope)
$\bar{\mathbf{u}}_t$	Mean IMU vector aggregated over the 2-second integration window
$\bar{\mathbf{m}}_t$	Mean magnetometer vector aggregated over the integration window
$\Sigma_t^{(\text{IMU})}$	IMU sample covariance over the integration window
$\Sigma_t^{(\text{mag})}$	Magnetometer sample covariance over the integration window
N_{IMU}	Number of IMU samples within the 2-second window (≈ 200)
N_{mag}	Number of magnetometer samples within the 2-second window (≈ 100)

VAE Encoder and Latent Representation

$\mathbf{z}_t$	Latent representation (latent vector) at time t ($\in \mathbb{R}^{d_z}$)
d_z	VAE latent space dimensionality ($d_z = 32$)
μ	Mean vector of the latent Gaussian ($\in \mathbb{R}^{d_z}$)
σ	Standard-deviation vector of the latent Gaussian
$\log \sigma^2$	Log-variance vector output by the encoder

$\mu_j,\ \sigma_j^2$	Mean and variance of the j-th latent dimension
ε	Auxiliary standard-normal noise for the reparameterization trick, $\varepsilon \sim \mathcal{N}(\mathbf{0},\mathbf{I})$
$\odot$	Element-wise (Hadamard) product
$\mathbf{h}_3$	Final encoder hidden layer ($\in \mathbb{R}^{64}$)
$\mathbf{W}_\mu,\ \mathbf{b}_\mu$	Encoder weight and bias producing the latent mean
$\mathbf{W}_\sigma,\ \mathbf{b}_\sigma$	Encoder weight and bias producing the latent log-variance
ϕ	VAE encoder (inference network) parameters
$q_\phi(\mathbf{z} \mid \mathbf{s})$	Approximate posterior (encoder distribution) over the latent variables
$p(\mathbf{z})$	Standard-normal latent prior, $\mathcal{N}(\mathbf{0},\mathbf{I})$
$D_{\mathrm{KL}}(\cdot \,\|\, \cdot)$	Kullback-Leibler divergence
$\mathcal{L}(\theta,\phi;\mathbf{x})$	Evidence lower bound (ELBO)

LSTM Temporal Encoder

$\mathbf{h}_t$	LSTM hidden state / temporal feature encoding ($\in \mathbb{R}^{128}$)
$\mathbf{c}_t$	LSTM cell state
$\tilde{\mathbf{c}}_t$	Candidate cell values
$\mathbf{f}_t$	Forget gate activation
$\mathbf{i}_t$	Input gate activation
$\mathbf{o}_t$	Output gate activation
$\mathbf{W}_f, \mathbf{W}_i, \mathbf{W}_c, \mathbf{W}_o$	LSTM gate weight matrices (forget, input, candidate, output)
$\mathbf{b}_f, \mathbf{b}_i, \mathbf{b}_c, \mathbf{b}_o$	Corresponding LSTM gate bias vectors
θ_{LSTM}	Learned LSTM parameters (gate weights and biases, both layers)

Decoder Outputs and Uncertainty

$\hat{\mathbf{x}}_t$	Predicted 2D position, easting and northing ($\in \mathbb{R}^2$)
$\mathbf{x}_t$	Ground-truth position
$\hat{\mathbf{x}}_t^{(i)}$	Per-modality position contribution from modality i
Σ_x	Position covariance in the Gaussian position likelihood
$\hat{\mathbf{s}}_t$	Reconstructed sensor vector ($\in \mathbb{R}^D$)
Σ_s	Reconstruction covariance in the Gaussian sensor likelihood
σ_t	Scalar position uncertainty estimate
$u_{\mathrm{epistemic}}$	Epistemic uncertainty (mean latent variance)
$u_{\mathrm{aleatoric}}$	Aleatoric uncertainty (squared reconstruction error)
α	Weighting between epistemic and aleatoric uncertainty ($\alpha = 0.5$)

Loss Functions and Training Parameters

$\mathcal{L}_{\mathrm{total}}$	Total composite training loss
$\mathcal{L}_{\mathrm{pos}}$	Position loss (mean squared error of position)
$\mathcal{L}_{\mathrm{KL}}$	KL divergence (latent regularization) loss
$\mathcal{L}_{\mathrm{rec}}$	Sensor reconstruction loss

$\mathcal{L}_{\text{reg}}$	L_2 weight-regularization term
$\mathcal{L}_{\text{stage1}}$	Stage-one (VAE pretraining) loss
$\mathcal{L}_{\text{stage2}}$	Stage-two (end-to-end fine-tuning) loss
β_1	KL divergence loss weight ($\beta_1 = 0.1$)
β_2	Reconstruction loss weight ($\beta_2 = 0.5$)
θ	VAE-LSTM network parameters (encoder, LSTM, decoder)
η	Offline training learning rate ($\eta = 10^{-3}$)
τ_{anneal}	KL annealing duration ($\tau_{\text{anneal}} = 10$ epochs)
e	Training epoch index (KL annealing schedule)
τ_{grad}	Gradient-clipping threshold ($\tau_{\text{grad}} = 1.0$)
λ	L_2 weight-decay coefficient ($\lambda = 10^{-5}$)
$\mathbf{W}_l$	Fully connected weight matrix of layer l

Adaptive Fusion

$\mathbf{w}_t$	Sensor fusion weight vector at time t, $[w_1(t), \ldots, w_K(t)]^\top$ with $\sum_i w_i(t) = 1$
$w_i(t)$	Fusion weight for sensor modality i, $\in [0,1]$
$\mathbf{w}_t^*$	Optimal fusion weights (solution of the OFAM optimization)
f_θ	Base VAE-LSTM model with fixed parameters θ
$g(\cdot)$	Fusion operation combining base-model output with adaptive weights
$\hat{\mathbf{x}}_t^{(\ell)}$	Hierarchical fusion estimate at time scale ℓ
ℓ	Time-scale (fusion) index
$\hat{\mathbf{x}}_t^{\text{feature}}$	Feature-level position estimate
$\hat{\mathbf{x}}_t^{\text{temporal}}$	Temporal-level position estimate
α_ℓ	Feature- vs. temporal-level fusion balance at time scale ℓ, $\in [0,1]$ (motion-adaptive form $\alpha_\ell(t)$)
$M(t)$	Motion magnitude indicator from IMU accelerometer variance, $\in [0,1]$
$c_0,\ c_1$	Learned parameters mapping motion state to the fusion balance ($c_0 = 0.5$, $c_1 = -2.0$)
Δ	Number of missed sensor updates (stale-measurement handling)
$\tilde{\mathbf{s}}_t^{(i)}$	Held (last valid) measurement sub-vector for modality i after Δ missed updates
γ_{miss}	Uncertainty growth rate for missing data ($\gamma_{\text{miss}} = 0.2$)
$\hat{\sigma}_t^{(i)}$	Per-modality inflated uncertainty for stale or missing measurements

Sensor Reliability Indicators

ρ_t	Combined reliability vector, $[\rho_1(t), \ldots, \rho_K(t)]^\top$ (signal metrics gated by reconstruction quality)
$\rho_i(t)$	Reliability indicator for sensor modality i, $\in [0,1]$
$\rho_{\text{WiFi}}(t)$	WiFi reliability (from RSSI variance and AP visibility)
$\rho_{\text{IMU}}(t)$	IMU reliability (from motion consistency)
$\rho_{\text{mag}}(t)$	Magnetometer reliability (from field stability)

$\tilde{\rho}_t$	Normalized reliability vector (target weight distribution), $\sum_i \tilde{\rho}_i(t) = 1$
$\mathbf{s}_t^{(i)}$, $\hat{\mathbf{s}}_t^{(i)}$	Measured and reconstructed sub-vectors for modality i
$e_i(t)$	Per-modality reconstruction error
$q_i(t)$	Reconstruction-quality factor, $\in (0,1]$
ε_i	Mean training-set reconstruction error for modality i
D_i	Dimensionality of modality i's slice of the sensor vector

WiFi Reliability Quantities

$\sigma^2_{\text{RSSI}}(t)$	RSSI variance over a sliding window ending at time t
σ^2_{ref}	Reference RSSI variance from training (calibration constant), $\sigma^2_{\text{ref}} = 25\ \text{dBm}^2$
$N_{\text{visible}}(t)$	Number of access points detected at time t
N_{expected}	Expected number of access points (from the radio map)
N_{ap}	Number of access points in the current observation set
$r_j^{(k)}$	RSSI measurement from access point j at time k
W	Sliding-window size for RSSI variance computation ($W = 30$ scans)

IMU and Magnetometer Reliability Quantities

$\mathbf{g}$	Gravitational acceleration vector, $[0,0,9.81]^\top\ \text{m/s}^2$
τ_{static}	Static-detection acceleration threshold ($\tau_{\text{static}} = 0.1\ \text{m/s}^2$)
κ	Angular-velocity decay rate for IMU reliability ($\kappa = 0.5\ \text{rad}^{-1}\text{s}$)
$\mathbf{m}_{\text{ref}}$	Reference magnetic field, $\|\mathbf{m}_{\text{ref}}\| = 50\ \mu\text{T}$
$\sigma^2_{\text{mag}}(t)$	Magnetic field variance over a sliding window
γ_{mag}	Magnetometer variance sensitivity ($\gamma_{\text{mag}} = 0.1\ (\mu\text{T})^{-2}$)

Online Adaptation Parameters

η_t	Adaptive (online) learning rate at time t
η_0	Amplitude of the decaying learning-rate component ($\eta_0 = 0.3$)
η_{min}	Minimum learning-rate floor ($\eta_{\text{min}} = 0.05$)
τ	Learning-rate decay time constant ($\tau = 100$ timesteps)

OFAM Objective

$\mathcal{J}_{\text{OFAM}}$	OFAM optimization objective
λ_{smooth}	Smoothness regularization coefficient ($\lambda_{\text{smooth}} = 0.1$)

Statistical Analysis

d	Cohen's effect size for paired differences

Chapter 1

Introduction

1.1 Background and Motivations

The proliferation of mobile computing devices and location-aware applications has created an unprecedented demand for accurate and reliable indoor positioning systems. While Global Positioning System (GPS) provides meter-level accuracy in outdoor environments, its signal attenuation and multipath effects render it ineffective for indoor localization scenarios. This limitation has catalyzed extensive research into alternative positioning technologies that can operate reliably within buildings, underground structures, and other GPS-denied environments.

Indoor localization systems have emerged as a critical enabling technology for numerous applications spanning smart buildings, augmented reality (AR), robotics, emergency response, and Internet of Things (IoT) deployments. The indoor location market has grown substantially, reaching USD 11.9 billion in 2024 and projected to grow to USD 31.4 billion by 2029 at a compound annual growth rate of 21.4% [1]. This expansion is driven by increasing adoption across diverse sectors: retail analytics leverage positioning for customer behavior analysis, healthcare facilities deploy wayfinding systems for patient and asset tracking, smart buildings integrate location services for occupancy management and energy optimization, and logistics operations use indoor positioning for warehouse automation. However, despite significant research efforts over the past two decades, achieving robust and scalable indoor localization remains a formidable challenge due to the complex and dynamic nature of indoor environments.

Traditional indoor positioning approaches can be broadly categorized into geometric, fingerprinting, and hybrid methods. Geometric approaches, including trilateration and triangulation, rely on distance or angle measurements from multiple reference points but suffer from accuracy degradation due to multipath propagation and non-line-of-sight conditions prevalent in indoor environments. Time-of-arrival (ToA), time-difference-of-arrival (TDoA), and received signal strength indicator (RSSI) measurements form the foundation of these geometric approaches, yet their performance is significantly compromised by the heterogeneous and dynamic nature of indoor radio propagation.

Fingerprinting-based localization, also known as scene analysis, has emerged as a more promising approach for indoor environments. This methodology involves creating a database of location-dependent signal characteristics during an offline training phase, followed by pattern matching during the online lo-

calization phase. WiFi fingerprinting, in particular, has gained widespread adoption due to the ubiquitous deployment of wireless local area networks (WLANs) in modern buildings. However, conventional WiFi fingerprinting approaches face significant challenges related to temporal variations in signal characteristics, device heterogeneity, and the labor-intensive nature of database construction and maintenance.

The integration of multiple sensor modalities represents a natural evolution toward more robust indoor localization systems. Inertial measurement units (IMUs), magnetometers, barometric pressure sensors, and visual sensors provide complementary information that can enhance positioning accuracy and reliability. IMUs enable dead reckoning through pedestrian dead reckoning (PDR) algorithms, while magnetometers provide heading information despite potential distortions from ferromagnetic materials in buildings. However, the effective fusion of these heterogeneous sensor modalities requires sophisticated algorithms that can model their respective error characteristics and temporal correlations.

Recent advances in deep learning have opened new avenues for indoor localization research. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), and their variants have demonstrated superior performance in extracting meaningful features from high-dimensional sensor data and learning complex non-linear relationships between environmental signatures and spatial coordinates. Long short-term memory (LSTM) networks, in particular, have shown promise in modeling temporal dependencies in sequential sensor measurements, while variational autoencoders (VAEs) offer powerful tools for dimensionality reduction and feature extraction in unsupervised learning scenarios.

The emergence of augmented reality (AR) technologies has introduced new opportunities and requirements for indoor localization systems. AR applications demand not only accurate position estimation but also precise orientation determination to enable seamless overlay of virtual content onto the real world. Microsoft HoloLens 2, as a state-of-the-art mixed reality device, incorporates advanced spatial mapping capabilities through simultaneous localization and mapping (SLAM) algorithms, creating opportunities for novel indoor positioning approaches that leverage both traditional wireless signals and spatial understanding capabilities.

Despite the considerable progress in indoor localization research, several fundamental challenges remain unresolved. First, the temporal stability of positioning systems remains problematic, as environmental changes, furniture rearrangement, and variations in occupancy patterns can significantly degrade localization accuracy over time. Second, the generalization capability of machine learning-based approaches across different environments and device types remains limited, necessitating extensive retraining for new deployments. Third, the computational complexity of sophisticated localization algorithms poses challenges for real-time implementation on resource-constrained mobile devices.

The concept of online adaptation represents a promising solution to address temporal variations and environment-specific challenges in indoor localization. An adaptive system revises how it combines its inputs as new observations arrive, maintaining performance despite changes in the underlying signal characteristics. In this thesis, only the weights that fuse the sensor modalities change during deployment; the parameters of the neural network remain fixed after training (Section 1.2.1). However, existing adaptive approaches often update model parameters online, which requires extensive computational resources or introduces significant latency, limiting their practical applicability in real-time scenarios.

Cross-dataset validation represents another critical challenge in indoor localization research. Most existing approaches are evaluated on single datasets collected under controlled conditions, raising questions about their generalization capability across different buildings, environments, and time periods. The International Conference on Indoor Positioning and Indoor Navigation (IPIN) competition datasets provide standardized benchmarks, yet few studies demonstrate consistent performance across multiple independent datasets.

The integration of multiple machine learning paradigms, including supervised learning for fingerprint matching, unsupervised learning for feature extraction, and reinforcement learning for adaptive parameter tuning, presents opportunities for developing more robust and versatile indoor localization systems. Hybrid architectures that combine the strengths of different neural network types while mitigating their individual limitations represent a promising research direction.

The motivations for this research stem from the recognition that existing indoor localization approaches suffer from fundamental limitations that prevent their widespread adoption in real-world scenarios. The need for systems that can maintain accuracy over extended time periods, adapt to environmental changes, generalize across different deployments, and operate efficiently on mobile devices drives the development of novel methodologies that address these challenges comprehensively.

1.2 Problem Statement

Indoor localization systems face three fundamental and interconnected problems that limit their practical deployment and long-term effectiveness. These problems persist despite decades of research and represent critical barriers to the widespread adoption of indoor positioning technologies in real-world applications.

Temporal Degradation Problem: Conventional indoor localization systems, particularly those based on fingerprinting approaches, experience significant accuracy degradation over time. Environmental changes, including variations in occupancy patterns, furniture rearrangement, temperature fluctuations, and electromagnetic interference, cause systematic shifts in the underlying signal characteristics that these systems rely upon. Studies have shown that WiFi fingerprinting systems can experience accuracy degradation of 30-50% within 6-12 months of deployment without recalibration. This temporal instability necessitates frequent recalibration cycles, making long-term deployment economically unfeasible and operationally impractical.

Computational Efficiency Problem: State-of-the-art indoor localization systems that achieve acceptable accuracy levels typically require substantial computational resources, particularly when employing deep learning architectures or complex sensor fusion algorithms. These computational demands translate to latency issues that compromise real-time applications. The trade-off between localization accuracy and computational efficiency remains poorly understood, with most existing approaches requiring offline processing or cloud-based computation that introduces communication latency and privacy concerns.

Generalization Problem: Indoor localization systems developed and optimized for specific environments or datasets often fail to maintain performance when deployed in different buildings, tested with different device types, or evaluated under varying environmental conditions. This lack of generalization capability limits the scalability and commercial viability of indoor positioning solutions. Cross-validation studies consistently reveal performance degradation when systems trained on one dataset are evaluated on independent datasets, highlighting the need for more robust and generalizable approaches.

These problems are exacerbated by the complex interplay between multiple sensor modalities, the high-dimensional nature of indoor signal spaces, and the dynamic characteristics of modern indoor environments. Existing solutions address these problems in isolation, leading to suboptimal overall system performance. There is a critical need for a unified framework that simultaneously addresses temporal adaptability, computational efficiency, and cross-environment generalization while maintaining high localization accuracy.

The research presented in this thesis directly addresses these fundamental problems through the development of a deep adaptive network architecture with an integrated Online Feature Adaptation Module (OFAM) that enables continuous adaptation of the sensor fusion weights with minimal computational overhead while demonstrating robust performance across multiple independent datasets and extended time periods.

1.2.1 Terminology

Several terms are used throughout the thesis in a specific sense, defined below.

Online: The deployment-time phase of the system, in which streaming sensor observations are processed on the mobile device to produce position estimates as they arrive. The online phase follows the offline phase, in which the network is trained on the survey data, and it is the only phase in which the end user interacts with the system.

Adaptation: A change of the three fusion weights that combine the WiFi, IMU, and magnetometer contributions to the position estimate, driven by reliability estimates computed from the observations themselves. The parameters of the neural network are frozen at deployment and are not changed by adaptation.

Online adaptation: Adaptation performed during the online phase, using no position labels, no gradient computation, and no retraining of the network. This distinguishes it from online learning, in which the parameters of the model are updated from streaming data, and from periodic retraining, in which the model is refit offline on newly collected labeled data. The mechanism developed here, the Online Feature Adaptation Module (OFAM), is specified in Section 5.5.

Reliability: A per-modality, time-varying, self-supervised estimate in the interval $[0, 1]$ of how trustworthy a sensor modality is for position estimation at the current time. The estimate is computed from observable properties of the measurements, such as their variance over a sliding window, the number of visible access points, and the reconstruction quality of the network's decoder, and never from ground truth positions (Section 5.4).

The main threat to reliability addressed in this thesis is temporal drift: the gradual change in signal characteristics caused by infrastructure modifications, occupancy patterns, and seasonal effects between the time a system is surveyed and the time it is used. Other threats, including sensor failure, transient disturbance, missing observations, and variation between devices, are listed in Section 5.7.4, which states for each whether the mechanism addresses it, addresses it in part, or leaves it as a limitation.

1.3 Research Objectives

The primary objective of this research is to develop a comprehensive indoor localization framework that addresses the fundamental limitations of existing approaches through novel deep learning architectures and adaptive mechanisms. The specific research objectives are:

1.3.1 Primary Objective

To design, implement, and validate a deep adaptive network architecture incorporating an Online Feature Adaptation Module (OFAM) that maintains high localization accuracy across extended time periods while operating efficiently on resource-constrained mobile devices and generalizing effectively across diverse indoor environments.

1.3.2 Secondary Objectives

1. **Develop Novel AR-Enhanced Fingerprinting:** Create an innovative indoor fingerprinting framework that leverages Microsoft HoloLens 2's spatial mapping capabilities to augment traditional WiFi-based fingerprinting with spatial context information, improving both accuracy and robustness.

2. **Design Efficient Hybrid Neural Architecture:** Implement a VAE-LSTM hybrid neural network that combines unsupervised feature learning capabilities with temporal sequence modeling to extract robust representations from multi-modal sensor data while maintaining computational efficiency.

3. **Implement Real-Time Online Adaptation:** Develop and validate the OFAM mechanism that enables continuous adaptation of the fusion weights to environmental changes with minimal computational overhead (target: ≤ 5 ms per adaptation cycle, derived in Section 1.4) while leaving the learned representations of the network unchanged.

4. **Demonstrate Long-Term Temporal Stability:** Conduct comprehensive evaluation demonstrating maintained localization accuracy across year-over-year deployment gaps without manual recalibration, establishing new benchmarks for temporal robustness in indoor positioning systems.

5. **Achieve Cross-Dataset Generalization:** Validate system performance across multiple independent datasets, including York University campus data and IPIN competition datasets [2], demonstrating transferability and robustness across different environmental conditions and building characteristics.

6. **Optimize Multi-Modal Sensor Fusion:** Develop efficient algorithms for fusing WiFi, IMU, and magnetometer sensor data that maximize complementary information while minimizing computational complexity.

These objectives collectively aim to advance the state-of-the-art in indoor localization by addressing the critical gap between laboratory research and practical deployment requirements.

1.4 Research Questions

This research investigates four fundamental questions that address the core challenges in indoor localization systems. Each question is formulated as a testable hypothesis with quantitative targets that can be validated through experimental evaluation.

RQ1: Can AR-based spatial anchoring using Microsoft HoloLens 2 achieve the centimeter-level ground truth accuracy required for constructing multi-building indoor localization datasets?

RQ2: Can a VAE-LSTM hybrid neural network architecture achieve sub-meter mean localization error while maintaining computational efficiency suitable for real-time mobile deployment?

RQ3: Does the Online Feature Adaptation Module (OFAM) enable continuous adaptation that maintains localization accuracy across year-over-year deployment gaps without requiring manual recalibration?

RQ4: Can adaptive multi-sensor fusion achieve robust indoor positioning performance under varying environmental conditions, including infrastructure changes, seasonal variations, and sensor degradation?

Accuracy and Latency Targets: The quantitative targets attached to these questions are a mean absolute error (MAE) below 1 m, with error at the 90th percentile remaining at meter level, that is, on the order of one meter rather than several meters, and an adaptation cost of at most 5 ms per position update. The accuracy target follows from three considerations. First, the reference points in this thesis were collected on a regular grid with 1 m spacing using the AR framework (Section 3.2.2). A mean error below the 1 m grid spacing means that an estimate typically falls within one grid cell of the true reference point, so the mean error target is set below 1 m. A mean alone can hide occasional large errors that place the user several grid cells, and therefore several meters, away from the true location, which limits the usefulness of location-based services and pedestrian navigation even when the mean is acceptable. For this reason the target also bounds the 90th

percentile. Second, the requirement classes in the literature place pedestrian navigation and location-based services at accuracy requirements on the order of one to a few meters, with robotics, manufacturing, and asset-tracking applications demanding decimeter or centimeter accuracy [3]. The ISO/IEC 18305 evaluation standard specifies that localization systems be characterized by their error distribution, including percentile statistics, rather than by a single average [4]. Stating the target for both the mean and the 90th percentile follows this practice. Third, commodity WiFi limits the accuracy that can be expected: RSSI fingerprinting systems typically report median errors of 2 to 3 m, whereas CSI-based systems report 0.5 to 1.2 m but require specialized network hardware that consumer devices do not provide (Section 2.3.2). A sub-meter mean error using RSSI on commodity smartphones is therefore a demanding target, at the level that CSI-based systems reach with specialized hardware, and it motivates the multi-sensor fusion and temporal modeling developed in later chapters.

The latency target follows from the update rate. Position updates at 10 Hz, the update rate typical of AR localization systems, leave a budget of 100 ms per update for sensor handling, inference, and adaptation. Inference through the VAE-LSTM and sensor processing consume most of this budget, so adaptation must be a small constant addition that does not displace inference or other applications that share the device. A target of at most 5 ms per adaptation cycle meets this requirement. The complete online pipeline measured in Section 5.6.5 runs in 58 ms per update, of which the adaptation stage accounts for the targeted 5 ms.

These research questions guide the investigation toward developing practical solutions that bridge the gap between theoretical advances in machine learning and the operational requirements of real-world indoor localization applications.

1.5 Scope and Limitations

1.5.1 Scope of Research:

This thesis focuses specifically on indoor localization within structured indoor environments, including university buildings, office complexes, and public facilities where WiFi infrastructure is readily available. The research encompasses both fingerprinting-based approaches and hybrid sensor fusion methodologies, with particular emphasis on environments where users carry smartphones or similar mobile devices equipped with standard sensor suites.

The technical scope includes the development and evaluation of deep learning architectures based on variational autoencoders and LSTM networks, implementation of online adaptation mechanisms with real-time constraints, and comprehensive evaluation using both controlled laboratory environments and real-world deployment scenarios. The research targets a mean absolute error below 1 m with error at the 90th percentile at meter level, the requirement derived in Section 1.4 for indoor navigation and location-based services.

Within WiFi, the work uses received signal strength (RSSI) rather than channel state information (CSI), and it uses WiFi rather than cellular (5G) positioning. Both choices follow from the use of commodity smartphones in the 2021 and 2022 data collection campaigns. RSSI is exposed by every WiFi-capable device through standard operating system interfaces, whereas CSI extraction requires network interface

cards with modified firmware that consumer devices do not provide (Section 2.3.2). Consumer smartphones of that period also did not expose 5G positioning reference signals or cellular channel state information through public application interfaces (Section 2.2.3). The accuracy and stability trade-offs between these measurement types, and the conditions under which CSI or 5G positioning would become preferable, are reviewed in Sections 2.2.3 and 2.3.2.

The temporal scope of evaluation spans two independent year-over-year gaps (York 2021→2022 and IPIN 2016→2017), with specific focus on the occupancy and infrastructure changes that accrue across an annual gap. Cross-environment validation is conducted using data from York University campus buildings and standardized IPIN competition datasets, providing insights into generalization capabilities across different building types and geographic locations.

1.5.2 Limitations and Constraints:

The following constraints bound the scope and applicability of this work.

Environmental Limitations: The research is limited to indoor environments with existing WiFi infrastructure and does not address outdoor or GPS-denied outdoor scenarios. The approach assumes availability of standard 2.4 GHz and 5 GHz WiFi signals and may not be directly applicable to environments with specialized or proprietary wireless networks.

Device Constraints: Evaluation is primarily conducted using Android smartphones and Microsoft HoloLens 2 devices. The findings may not directly translate to other device types, particularly those with different sensor specifications or computational capabilities. The research does not address specialized industrial or military devices with unique requirements.

Accuracy Expectations: The research targets a mean absolute error below 1 m with meter-level error at the 90th percentile (Section 1.4); it does not address applications requiring centimeter-level precision, such as robotic manipulation or precise asset tracking. The focus is on human-scale navigation and location-aware services rather than high-precision positioning.

Computational Platform Limitations: The real-time adaptation mechanisms are designed for standard smartphone hardware and may not scale to extremely resource-constrained IoT devices. Cloud-based processing scenarios are not extensively addressed, focusing instead on edge-computing approaches.

Privacy and Security Considerations: The system is designed for on-device inference, so position estimates need not leave the user's device. The data collection involved no human participants beyond the research-team operators; the ethics statement in Section 3.5.3 describes what was recorded and how it is

held. Security of a deployed system, including rogue or spoofed access points, manipulation of the inputs to the online adaptation, and confidentiality of the fingerprint database and anchor store, is discussed in Section 6.3.6, which states which threats the design bounds and which were not tested. A formal security analysis and formal privacy-preservation mechanisms remain beyond the scope of this work.

Standardization and Interoperability: The research develops novel algorithms and architectures but does not specifically address industry standardization requirements or comprehensive interoperability testing with existing commercial indoor positioning systems.

These scope definitions and limitations provide clear boundaries for the research contribution while acknowledging areas that represent opportunities for future work and complementary research efforts.

1.6 Methodology Overview

The research methodology employs a systematic approach combining theoretical development, algorithm design, implementation, and comprehensive experimental validation. The methodology is structured around four interconnected phases that collectively address the research objectives and questions.

Phase 1: AR-Enhanced Data Collection and Preprocessing: The methodology begins with the development of a novel data collection framework utilizing Microsoft HoloLens 2 to augment traditional WiFi fingerprinting with spatial context information. This phase involves the implementation of specialized applications that simultaneously capture WiFi received signal strength indicators (RSSI), IMU measurements, magnetometer readings, and spatial mapping data from the HoloLens 2's environmental understanding capabilities.

Data preprocessing pipelines are developed to handle the heterogeneous nature of the collected sensor data, including temporal alignment, noise filtering, and feature extraction. The spatial mapping data is processed to extract geometric features that complement traditional signal strength measurements, creating enriched fingerprint representations that incorporate both radio frequency and spatial characteristics.

Phase 2: Deep Adaptive Architecture Development: The core methodology involves the design and implementation of a novel VAE-LSTM hybrid neural network architecture optimized for indoor localization. The variational autoencoder component enables unsupervised feature learning from high-dimensional sensor data while providing robust representations that capture the underlying structure of indoor environments. The LSTM component models temporal dependencies in sequential measurements, enabling effective utilization of motion patterns and temporal signal variations.

The Online Feature Adaptation Module (OFAM) is integrated into this architecture to enable continuous adaptation of the fusion weights from streaming sensor data while the network parameters remain frozen. The OFAM mechanism is designed with strict computational constraints, targeting adaptation cycles of at most 5 ms to ensure real-time performance. Because the network parameters are never modified during

deployment, the learned representations cannot be degraded by adaptation. The weight update is smoothed by an exponential moving average so that the fusion weights change gradually rather than oscillating.

Phase 3: Multi-Modal Sensor Fusion Optimization: The methodology addresses the fusion of WiFi, IMU, and magnetometer sensor data through reliability-driven adaptive weighting schemes within the deep learning framework. Different fusion strategies are evaluated, including early fusion at the feature level, intermediate fusion within the network architecture, and late fusion at the decision level.

Optimization algorithms are developed to automatically tune the fusion parameters based on environmental characteristics and sensor reliability indicators. The fusion approach is designed to gracefully handle missing or unreliable sensor modalities, maintaining localization capability even when individual sensors fail or provide degraded measurements.

Phase 4: Comprehensive Evaluation and Validation: The evaluation methodology encompasses both controlled laboratory experiments and real-world deployment scenarios. Controlled experiments are conducted to evaluate individual component performance, ablation studies to understand architectural design choices, and comparative analysis with state-of-the-art approaches.

Real-world evaluation includes a 90-day deployment simulation and year-over-year testing on real collections to assess temporal stability, together with cross-building validation to evaluate generalization capabilities. The evaluation utilizes both proprietary datasets collected at York University and standardized IPIN competition datasets to ensure reproducibility and fair comparison with existing approaches.

Performance metrics include positioning accuracy, computational efficiency, adaptation speed, and temporal stability. Confidence interval analysis supports reliable interpretation of the accuracy results; paired significance testing of the headline accuracy comparisons confirms that the proposed adaptive model improves on the static baseline and on the kNN, CNN, and Fusion-DHL comparators; and correlation significance testing establishes that the sensor-reliability indicators driving the adaptation mechanism are significantly associated with positioning accuracy.

Phase 5: Implementation and Validation Framework: The implementation methodology emphasizes reproducible research through comprehensive documentation and standardized evaluation protocols. The validation framework includes cross-validation techniques, comparison with established baseline methods, paired significance testing of the headline accuracy comparisons, and significance testing of the sensor-reliability correlations central to the adaptation mechanism.

This methodology ensures that the research contributions are both theoretically sound and practically applicable, addressing real-world deployment requirements while advancing the fundamental understanding of adaptive indoor localization systems.

1.7 Contribution of Thesis

This thesis makes four distinct contributions to indoor localization research, each addressing specific aspects of the challenges outlined in the problem statement. Two of them are the major contributions of the thesis: the longitudinal multimodal benchmark (the second contribution below) and the label-free online reliability adaptation mechanism validated over real temporal drift (the fourth). The other two are enabling contributions: the AR-based ground truth method makes the benchmark possible, and the compact VAE-LSTM architecture provides the reliability signal that the adaptation mechanism uses. Together, these contributions form a coherent system addressing temporal stability, accuracy, and practical deployability of indoor positioning.

First, this thesis introduces an augmented reality-based framework for establishing centimeter-accurate ground truth in indoor environments, decoupling the ground truth collection process from runtime positioning and enabling scalable data collection across multiple buildings. Traditional indoor localization research relies on labor-intensive manual surveying or specialized infrastructure for ground truth establishment, limiting both the scale of datasets and the diversity of environments that can be studied. The proposed framework leverages Microsoft HoloLens 2 spatial mapping capabilities to create persistent spatial anchors at known reference locations, enabling rapid deployment across building floors while maintaining the accuracy standards required for rigorous algorithm evaluation.

To our knowledge, this is the first indoor localization framework to leverage consumer-grade mixed reality devices for ground truth establishment at scale. Prior approaches to ground truth collection have relied on either labor-intensive manual surveying with total stations, which limits practical deployment to small testbeds, or specialized infrastructure such as ultra-wideband anchor networks that require permanent installation. The proposed approach achieves comparable accuracy to professional surveying equipment while reducing deployment time from days to hours and eliminating the need for permanent infrastructure modifications.

Second, building upon the AR-based ground truth methodology, this thesis contributes a longitudinal multi-building dataset that enables systematic study of temporal stability in indoor localization systems. Prior datasets in the field typically span single buildings, short time periods, or limited sensor modalities, precluding investigation of how localization accuracy degrades across seasons and years. The contributed dataset encompasses data collection campaigns across multiple years, capturing WiFi fingerprints, inertial measurement unit readings, and magnetometer signatures across architecturally diverse buildings selected to ensure that findings generalize beyond any single building type.

Unlike existing fingerprinting datasets, few of which combine WiFi, IMU, and magnetometer modalities over repeated annual campaigns, this contribution provides a publicly available two-year, multi-building dataset spanning all three.[1] This temporal and spatial scope enables investigation of questions that prior datasets could not address: how does localization accuracy degrade across seasons and years, and do patterns of degradation generalize across building types? The dataset addresses a gap in the research infrastructure available to the indoor positioning community.

[1] The dataset is publicly available at https://yorkuindoor.ausmlab.com.

Third, this thesis presents a hybrid neural network architecture that combines variational autoencoder-based feature encoding with LSTM-based temporal modeling, achieving competitive positioning accuracy while maintaining computational tractability for mobile deployment. The architecture addresses a fundamental tension in fingerprinting-based localization: high-dimensional sensor inputs require effective dimensionality reduction, yet temporal patterns in sequential measurements carry information that single-snapshot approaches discard. The VAE encoder projects multi-dimensional feature vectors into a compact latent space, while the LSTM temporal encoder processes sequences of consecutive latent vectors, capturing motion patterns and exploiting the temporal coherence of pedestrian trajectories.

The novelty claimed here is not the combination of a VAE with an LSTM, both of which are established components. The compact VAE-LSTM is chosen for two properties on which the rest of the thesis depends. Its reconstruction branch gives a per-modality signal of how well the current observations are explained by the trained model, and that signal is available at deployment without labels. The LSTM captures the motion patterns of pedestrian trajectories that single-snapshot approaches discard. The architecture is therefore an enabling contribution: it improves accuracy over both VAE-only and LSTM-only baselines at a size suited to mobile deployment, and it produces the reliability signal that the fourth contribution uses.

Fourth, the culminating contribution of this thesis is the Online Feature Adaptation Module (OFAM), a mechanism for continuous adaptation of the sensor fusion weights during deployment that maintains positioning accuracy across year-over-year deployment gaps without requiring costly retraining or manual recalibration. OFAM directly addresses the temporal degradation problem by enabling the system to adapt to environmental changes (WiFi infrastructure modifications, occupancy pattern shifts, and seasonal variations) as they occur. The reliability-weighted fusion framework assigns dynamic weights to WiFi, IMU, and magnetometer inputs based on real-time reliability estimates, ensuring that adaptation leverages the most trustworthy signals available at each timestep.

Unlike static fingerprinting systems that require periodic recalibration to maintain accuracy, or transfer learning approaches that require access to labeled data from the target domain, OFAM adapts during normal operation with no labels, no gradients, and no retraining. The module monitors per-modality signal statistics together with the decoder's reconstruction quality to detect shifts in sensor reliability without ground truth positions, and it updates only the three fusion weights while the network parameters stay frozen, so previously learned representations are preserved. The contribution is this adaptation mechanism and its demonstrated robustness to real temporal drift, evaluated on collections a year apart rather than only on simulated perturbations. This approach reduces the operational burden of indoor localization systems from periodic manual intervention to autonomous maintenance of the fusion weights. The temporal-stability benefit is not specific to the York University testbed: the same pattern holds on the independent IPIN 2016–2017 data, never used to tune OFAM, which speaks directly to the cross-dataset generalization of RQ4.

1.8 Structure of Thesis

The remainder of this thesis is organized as follows. Chapter 2 reviews indoor positioning technologies, examining WiFi fingerprinting, the alternatives offered by channel state information and cellular positioning,

multi-sensor fusion, deep learning architectures, and the families of adaptation methods proposed for long-term operation, and identifies the research gaps that motivate this work. Chapter 3 presents the AR-based data collection framework and the longitudinal multi-building dataset, documents the collection campaigns and the ethics statement, and includes a deployment procedure for applying the framework to a new site and a statement of data and code availability. Chapter 4 details the probabilistic temporal architecture for multi-sensor localization, the VAE-LSTM, explaining the design rationale, including the choice of a recurrent rather than an attention-based sequence model, and evaluating the performance of the static base model. Chapter 5 introduces the reliability-weighted sensor fusion framework and the Online Feature Adaptation Module, defines the scope of the reliability estimates, and demonstrates their effectiveness in maintaining accuracy across year-over-year deployment. Chapter 6 synthesizes the contributions, addresses the research questions with quantitative evidence, acknowledges limitations including the operational limits of WiFi-based localization and security considerations for deployment, and identifies directions for future research.

The research questions posed in Section 1.4 are addressed across these chapters: RQ1 is answered through the AR framework evaluation in Chapter 3, RQ2 through the architecture analysis in Chapter 4, RQ3 through the temporal stability validation in Chapter 5, and RQ4 through the cross-building and cross-year evaluation spanning Chapters 3 to 5 with synthesis in Chapter 6.

Chapter 2

Literature Review

2.1 Introduction

Indoor positioning systems have become increasingly critical for a wide range of applications, from navigation and location-based services to autonomous systems and augmented reality experiences. While Global Positioning System (GPS) provides reliable positioning in outdoor environments, its effectiveness is severely limited in indoor settings due to signal attenuation, multipath effects, and poor satellite visibility. This fundamental limitation has driven extensive research into alternative positioning technologies specifically designed for indoor environments.

The evolution of indoor positioning has been marked by significant technological advances, from early proximity-based systems to sophisticated machine learning approaches capable of achieving sub-meter accuracy. Modern indoor positioning systems leverage diverse sensor modalities, including WiFi, Bluetooth, inertial measurement units (IMU), magnetometers, cameras, and ultra-wideband (UWB) technologies, each offering unique advantages and facing distinct challenges.

This literature review provides a comprehensive examination of indoor positioning technologies, organized thematically by the fundamental problems that limit practical deployment. Rather than cataloging techniques in isolation, the review examines how different approaches address, or fail to address, the core challenges of WiFi signal instability, temporal drift, multi-sensor fusion, and deep learning architectures for localization. This problem-oriented structure reveals critical research gaps that motivate the methodology developed in subsequent chapters.

The review proceeds as follows. Section 2.2 establishes fundamental positioning principles and vocabulary. Sections 2.3 through 2.7 examine the literature organized by problem area: WiFi signal variability including Channel State Information approaches, temporal drift and adaptation mechanisms, multi-sensor fusion architectures, and deep learning methods spanning convolutional networks, recurrent architectures, transformer models, and uncertainty quantification techniques. Each section identifies specific gaps through explicit critique. Section 2.6 reviews the emerging paradigm of augmented reality-based localization. Section 2.8 synthesizes the identified gaps, and Section 2.9 summarizes the review and bridges to the thesis methodology.

Figure 2.1 illustrates the evolution of indoor positioning technologies over three decades, high-

lighting major milestones across WiFi-based, proximity and sensor-based, sensor fusion, and deep learning approaches. The field has undergone two paradigm shifts: the smartphone era beginning around 2007 made MEMS sensors ubiquitous in consumer devices, while the deep learning revolution from 2016 transformed algorithmic approaches to localization. Understanding this historical context situates the research contributions of this thesis within the broader trajectory of the field.

2.2 Fundamentals of Indoor Positioning

Indoor positioning systems employ various fundamental techniques to determine location within enclosed environments. Understanding these core principles is essential for comprehending the strengths, limitations, and appropriate applications of different positioning approaches. This section briefly establishes the vocabulary and foundational concepts that underpin subsequent thematic analysis. Lohan et al. [2] provide a comprehensive framework for evaluating indoor localization systems, establishing standardized benchmarking methodologies and quality metrics that enable meaningful comparison across approaches. Their work addresses a critical gap in research reproducibility by defining consistent evaluation protocols for accuracy, robustness, and scalability assessment.

2.2.1 Positioning Techniques Overview

Proximity-based positioning determines location based on detected presence near known reference points, providing room-level accuracy through RFID tags, Bluetooth beacons, or WiFi access point presence detection [19]. Early systems such as Active Badge established the cell-of-origin paradigm using infrared beacons [5], while LANDMARC demonstrated active-RFID proximity sensing with reference tags [20]. While simple to implement with low computational requirements, proximity-based systems suffer from coarse location granularity.

Lateration techniques determine position by measuring distances to multiple reference points with known locations. Trilateration, the most common form, uses distance measurements to three or more reference points, solving circular intersection problems. Distance estimation typically employs time-of-flight measurements or received signal strength indicators (RSSI) [19, 21], with ultra-wideband (UWB) systems achieving centimeter-level accuracy under ideal conditions [22].

Simultaneous Localization and Mapping (SLAM) represents a sophisticated approach that simultaneously constructs environmental maps while determining sensor location within that map. Vision-based SLAM systems utilize camera imagery to identify and track distinctive features across sequential frames [23]. Contemporary mobile AR frameworks, including ARCore and ARKit, represent mature implementations of Visual-Inertial SLAM for consumer devices [24, 25].

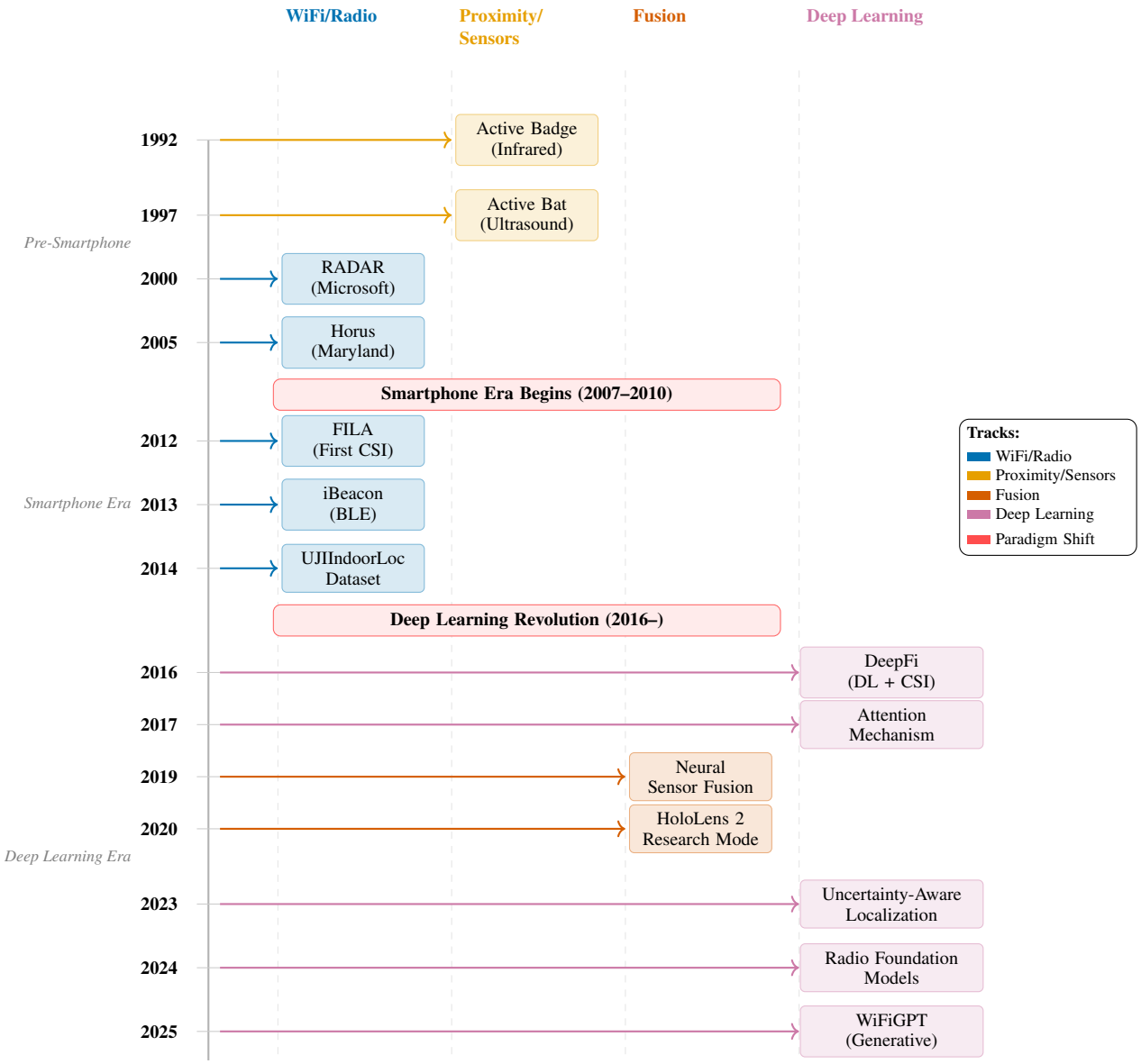


Figure 2.1: Evolution of indoor positioning technologies (1992–2025). Milestones are arranged on four tracks: WiFi and radio methods, proximity and sensor systems (infrared badges and ultrasound), multi-sensor fusion, and deep learning. Two shifts mark the period: the smartphone era (2007–2010) made MEMS sensors and WiFi radios ubiquitous in consumer devices, and the deep learning era (2016–) changed the algorithmic approach to fingerprinting and fusion. Sources: Active Badge [5], Active Bat [6], RADAR [7], Horus [8], FILA [9], iBeacon [10], UJIIndoorLoc [11], DeepFi [12], the attention mechanism [13], neural sensor fusion (DeepFusion) [14], HoloLens 2 Research Mode [15], uncertainty-aware localization (RLoc) [16], Radio Foundation Models [17], and WiFiGPT [18].

Time Difference of Arrival (TDoA) determines location by measuring time differences of signal arrivals at multiple synchronized receivers, eliminating transmitter synchronization requirements [19, 26]. Modern implementations include cellular positioning systems and ultra-wideband indoor systems [22].

Angle of Arrival (AoA) positioning determines location by measuring the direction of incoming signals at multiple reference stations using antenna arrays or directional antennas [19]. Modern AoA implementations, particularly Bluetooth 5.1 direction finding, offer improved positioning compared to RSSI-based approaches [27].

Fingerprinting represents a pattern matching approach that relies on the unique characteristics of signal patterns at different locations [19, 28]. The technique creates a database of signal measurements during an offline training phase, then compares real-time measurements against this database during online operation, an approach pioneered by the RADAR system [7]. WiFi fingerprinting has emerged as the most prevalent implementation due to infrastructure ubiquity.

Dead reckoning determines position by integrating motion information from a known starting location, using inertial measurement units containing accelerometers, gyroscopes, and magnetometers. Pedestrian Dead Reckoning (PDR) tracks human locomotion through step detection and heading estimation, though integration errors accumulate rapidly without periodic corrections [29].

2.2.2 Fingerprinting Foundations

Fingerprinting has emerged as the predominant approach for high-accuracy indoor positioning due to its ability to handle complex indoor propagation effects. The foundation of WiFi fingerprinting was established by Bahl and Padmanabhan's seminal RADAR system in 2000 [7], which introduced received signal strength (RSS) based indoor localization. This pioneering work presented the first comprehensive fingerprinting system that gathered average signal strength from surrounding wireless access points at fixed locations and used k-Nearest Neighbors (kNN) with Euclidean distance metrics for position estimation.

Following RADAR's groundbreaking contribution, Youssef and Agrawala [8] enhanced the probabilistic approach to indoor localization through Bayesian inference methods. Concurrently, Roos et al. [30] developed probabilistic approaches to WLAN user location estimation, contributing to the theoretical foundations of fingerprinting methodologies.

These early works established fundamental concepts: (1) the two-phase approach consisting of offline radio map construction and online position estimation, (2) the relationship between access point density and positioning accuracy, and (3) the impact of environmental factors on signal propagation and positioning performance.

2.2.3 Cellular Positioning and Channel State Information

Cellular networks offer a positioning path that does not depend on WiFi infrastructure. The 3GPP Release 16 specifications define downlink and uplink time-difference-of-arrival, multi-cell round-trip-time, and angle-based methods operating on dedicated positioning reference signals, and the wide bandwidth and beamformed transmissions of 5G New Radio (NR) improve timing and angular resolution relative to earlier generations [31]. Channel state information from commercial 5G signals has also been used for fingerprinting: iPos-5G [32] uses the multi-beam structure of 5G NR to localize from a single base station. A recent comparison of commercial 5G and WiFi CSI on a common software-defined radio platform [33] found that 5G CSI features were more stable over time and generalized better in regression-based localization, whereas WiFi CSI was more discriminative over short intervals in classification tasks; the authors describe the two carriers as having complementary strengths suited to hybrid systems. Cellular positioning is therefore an option for environments with indoor 5G coverage, but three constraints kept it outside the scope of this thesis. First, the standardized methods are computed by network-side location functions from measurements that consumer handsets do not expose through public application interfaces [31], and the CSI used by iPos-5G was captured with software-defined radio hardware rather than with a phone [32]. Second, the measurement campaigns of 2021 and 2022 (Chapter 3) relied on a commodity Android smartphone. Third, WiFi infrastructure was already present in every building surveyed. WiFi received signal strength is also available on every WiFi-enabled device through standard operating-system interfaces, which is the deployment condition this thesis targets.

Within WiFi itself, fine timing measurement (FTM) provides a ranging alternative to signal-strength fingerprinting. IEEE 802.11mc FTM measures the round-trip time between a client and a responder-capable access point. Its evaluation on an open platform showed meter-level ranging accuracy with device- and environment-dependent offsets [34]. The IEEE 802.11az amendment (2022) improves the accuracy, scalability, and security of the exchange [35]. Hybrid designs that switch between RSS and round-trip-time measurements according to channel conditions have followed [36]. FTM requires responder support on the access points as well as client-side support, and the 2021–2022 traces predate widespread 802.11mc availability in deployed infrastructure. For this reason the RSS–RTT baseline of Section 5.6.8 is evaluated in RSS-only mode. The 802.11az amendment postdates the data altogether. Ranging-based WiFi positioning is thus a direction for future work rather than an available option that was declined.

2.3 WiFi Signal Instability and Variability

WiFi-based positioning faces fundamental challenges arising from the inherently unstable nature of wireless signals in indoor environments. This section examines the sources of signal variability and the approaches developed to address them.

2.3.1 Sources of Signal Variability

RSSI measurements exhibit substantial spatial and temporal variability in indoor environments. Physical obstacles including walls, furniture, and human bodies cause signal attenuation and shadowing effects. Construction materials exhibit varying electromagnetic properties, with metal structures causing strong reflections and concrete providing significant attenuation.

Temporal variations result from dynamic environmental factors including occupancy patterns, furniture rearrangement, and electronic interference. Studies examining RSSI stability over extended periods reveal significant variations that can degrade positioning accuracy if not properly addressed [37, 38, 39].

2.3.2 Channel State Information (CSI) Approaches

Channel State Information (CSI) represents a significant evolution from RSSI-based fingerprinting, capturing the amplitude and phase response across individual OFDM subcarriers rather than aggregating signal strength into a single scalar value [40]. Where RSSI discards multipath information through power summation, CSI preserves the frequency-selective channel response, enabling more discriminative location fingerprints. Modern WiFi standards (802.11n/ac/ax) transmit across 30–56 subcarriers per antenna, providing substantially richer measurement vectors than single RSSI values.

The technical distinction between RSSI and CSI has profound implications for positioning accuracy. RSSI-based systems typically achieve 2–3 meter median error in complex indoor environments due to the information loss inherent in signal aggregation. CSI-based approaches exploit the preserved multipath structure to achieve sub-meter accuracy, with reported median errors of 0.5–1.2 meters under favorable conditions [12, 41]. However, this accuracy advantage comes with significant deployment constraints that limit practical applicability.

Wu et al. [9] introduced FILA, the first system to leverage CSI for indoor localization. FILA exploited the signal propagation model implicit in CSI measurements, achieving finer-grained location discrimination than RSSI approaches. While demonstrating CSI's potential, the evaluation was confined to a single controlled environment over a limited temporal span, leaving questions about robustness to environmental variations unanswered.

Xiao et al. [42] developed FIFS, the first fingerprinting system built specifically for CSI measurements. FIFS exploited both spatial diversity across antenna elements and frequency diversity across subcarriers to construct robust fingerprints. The system achieved 0.9 meter median error in experimental evaluation, substantially outperforming contemporaneous RSSI approaches. However, the requirement for Intel WiFi Link 5300 network interface cards with modified firmware constrained deployment to research laboratories with specialized hardware.

Kotaru et al. [41] demonstrated that CSI enables Angle of Arrival (AoA) estimation without dedicated antenna arrays through their SpotFi system. By exploiting phase differences across subcarriers, SpotFi achieved decimeter-level accuracy in line-of-sight conditions. The approach requires MIMO antenna configurations and careful phase calibration, adding complexity beyond basic CSI extraction. SpotFi's accuracy degrades substantially in non-line-of-sight conditions common in indoor environments with obstructed

paths.

Wang et al. [12] introduced DeepFi, integrating deep learning with CSI fingerprinting through a greedy learning approach. DeepFi achieved 1.8 meter median error in experimental evaluation, demonstrating that neural feature extraction could further improve CSI-based positioning. However, the evaluation was limited to a single building over several weeks, insufficient to assess long-term stability or generalization to diverse environments. Subsequent work by the same authors explored amplitude-phase combinations [43, 44], progressively refining the deep learning approach to CSI fingerprinting.

Chen et al. [45] developed ConFi, applying convolutional neural networks to CSI-based fingerprinting. The CNN architecture automatically learned discriminative features from raw CSI measurements, reducing manual feature engineering requirements. ConFi achieved improved accuracy over prior approaches but inherited the fundamental hardware constraints of CSI-based systems, limiting practical deployment.

The fundamental limitation of CSI-based approaches lies in hardware accessibility. CSI extraction requires specialized network interface cards with modified firmware or custom drivers [46]. The Intel WiFi Link 5300 (IWL5300) became the de facto research platform, but this chipset is legacy hardware unavailable in contemporary consumer devices. Recent work employs Software Defined Radios or modified router firmware with CSI extraction tools, further limiting deployment beyond research settings. In contrast, RSSI measurements are universally available across all WiFi-enabled devices through standard APIs, enabling positioning system deployment on billions of existing smartphones without hardware modifications. Neither Android nor iOS exposes CSI through its public WiFi interfaces. Obtaining it requires firmware or driver access on the client, or a dedicated receiver placed in the environment. The same constraint applies to cellular CSI (Section 2.2.3). The data collection of Chapter 3 used a commodity smartphone with unmodified software, so RSSI was the only WiFi signal measurement that could be gathered at the scale of four buildings and two years.

This hardware constraint explains the focus of this thesis on RSSI-based fingerprinting despite CSI's demonstrated accuracy advantages. While CSI approaches represent important advances in positioning accuracy potential, the practical requirements for specialized equipment preclude deployment in the multi-building, multi-year operational context that this thesis addresses. The RSSI-based approach adopted in subsequent chapters prioritizes deployment feasibility over theoretical accuracy limits, reflecting the practical constraints of real-world indoor positioning applications.

2.3.3 Addressing Device Heterogeneity

Device diversity represents a fundamental challenge due to variations in hardware characteristics between different devices. WiFi chipsets, antenna designs, and signal processing implementations create significant differences in received signal strength measurements for identical environmental conditions. Signal strength measurements can vary by 10-20 dBm between different device models even in identical positions.

Calibration-based approaches address device differences through measurement transformation techniques that normalize signal strength readings across different devices. Feature engineering approaches focus on deriving device-invariant features, with relative signal strength differences and rank-based features providing representations less sensitive to device-specific biases. Domain adaptation treats device

differences as domain shift problems, using techniques such as adversarial training to learn device-invariant features [47].

2.3.4 Gap: Environmental Adaptation Mechanisms

However, existing approaches to WiFi signal instability largely address short-term variations through filtering and averaging techniques, failing to address the fundamental challenge of long-term environmental adaptation. Current systems require periodic manual recalibration or complete database reconstruction when environmental conditions change significantly.

Moreover, device-agnostic approaches typically require additional calibration procedures or data collection that limits practical scalability. The heterogeneity of mobile devices in practical deployments creates ongoing challenges as new hardware enters the market, and existing calibration approaches cannot anticipate future device characteristics.

2.4 Temporal Drift in Indoor Positioning

Temporal stability represents one of the most critical yet underexplored challenges in indoor positioning. This section examines the manifestations of temporal drift and the critical gap in longitudinal research.

2.4.1 Manifestations of Temporal Degradation

Long-term evaluation studies spanning multiple months to years have provided crucial insights into the degradation patterns of fingerprinting systems. Studies examining temporal variations in WiFi signal strength over periods of more than a year [38, 48] reveal that positioning accuracy can deteriorate significantly due to environmental changes, access point modifications, and seasonal variations in signal propagation characteristics [49, 50]. Mendoza-Silva et al. [38] conducted a systematic 15-month measurement campaign in a university building, releasing a long-term WiFi fingerprinting dataset expressly intended to support research on robust indoor positioning and documenting substantial long-term RSS variation and its effect on positioning performance. Such longitudinal evidence underscores the necessity of adaptive mechanisms in long-term deployments.

Dataset drift occurs when the statistical properties of signal measurements change over time due to environmental modifications, infrastructure updates, or seasonal variations. Physical infrastructure modifications, such as furniture rearrangement, construction activities, or equipment installation, alter signal propagation patterns and invalidate existing fingerprint databases [38, 12].

Environmental factor modeling attempts to relate signal stability to measurable conditions such as temperature, humidity, and occupancy levels. Research has investigated the impact of indoor relative humidity on positioning accuracy, developing adaptive expansion fingerprint database models that dynamically respond to such changes [51].

2.4.2 Online Learning for Adaptation

Approaches that keep a positioning model current after deployment fall into six families. They differ in whether they need position labels, whether they compute gradients on the device, what they cost at run time, whether they risk forgetting what was learned before deployment, and whether they model the reliability of individual sensor modalities. Each family is reviewed in turn and then compared.

Online sequential learning. The online sequential extreme learning machine (OS-ELM) of Liang et al. [52] updates the output weights of a single-hidden-layer network in closed form as data arrive, one sample or one chunk at a time, without revisiting earlier data. Zou et al. applied it to WiFi fingerprinting for fast training and update [53] and later built a positioning system in which the existing access points act as monitors to construct and adapt the radio map online [54]. The updates are cheap and gradient-free, but they require labeled fingerprints, or a source of pseudo-labels such as monitors at known positions. The model must also balance adaptation rate against stability so that short-term fluctuations do not overwrite the map.

Radio-map maintenance. A second family keeps the fingerprint database itself current. Bolliger [55] developed Redpin, a zero-configuration system in which the fingerprint database is built and maintained through user collaboration. Rai et al. [56] removed the explicit user effort in Zee by tracking users inertially, constraining their trajectories with the floor plan, and collecting the WiFi scans as crowdsourced fingerprints. Silva et al. [39] proposed monitoring the radio environment and enriching an outdated radio map by radial basis function interpolation. Rajab and Wang [57] proposed autonomous fingerprint-map maintenance that updates the map from data gathered during operation while assessing the reliability of the updates. Crowdsourcing lowers the maintenance cost but introduces quality-control problems. None of these methods addresses the relative reliability of sensor modalities other than WiFi.

Domain adaptation and transfer learning. Pan et al. [58] were among the first to treat the change of a WiFi environment over time and across devices as a transfer learning problem, reusing a labeled radio map from one setting together with a small amount of data from the target setting. The general framework is surveyed by Pan and Yang [59]. Adversarial domain adaptation [47, 60] learns feature extractors whose outputs are indistinguishable between source and target domains and has been applied to device heterogeneity (Section 2.3.3). These methods reduce the labeled data needed for a new period or device. However, they adapt the feature extractor through gradient-based training on a batch of target data, which is an offline procedure rather than a run-time one. They also do not separate the contributions of individual modalities.

Continual learning. When a model is updated on successive periods, gradient updates on new data overwrite the parameters that served earlier data, the phenomenon known as catastrophic forgetting [61]. Elastic weight consolidation [61] penalizes changes to parameters that were important for earlier tasks, using

a Fisher-information estimate of importance. Applied to positioning, the successive tasks would be successive environmental states. The method requires gradients, stored importance weights, and labeled or pseudo-labeled data for each period, and it reduces forgetting without removing it.

Test-time and self-supervised adaptation. A more recent family adapts a trained network on unlabeled test data during deployment. TENT [62] minimizes the entropy of the predictions on each test batch and updates the normalization-layer parameters. SHOT [63] adapts the feature extractor toward a frozen source classifier using information maximization and self-generated pseudo-labels. Test-time training [64] attaches a self-supervised auxiliary task and updates the shared encoder on each test sample. These methods share the label-free, deployment-time premise of the adaptation studied in this thesis. However, they were designed for classification, they require a backward pass on the device for every adaptation step, and their errors can accumulate over long streams because the adapted parameters are never anchored to labeled data.

Adaptive sensor fusion. Within filter-based fusion, adaptive weighting adjusts sensor contributions in response to measurement quality. Li et al. [65] scale observation weights by Mahalanobis distance in a complementary Kalman filter for UWB and inertial fusion, and Zhong et al. [66] adjust the weights of WiFi and visual measurements by monitoring residuals (Section 2.5.2). These methods need no labels and no gradients, and they are the only family that models per-modality reliability. However, they respond to instantaneous residuals and, as formulated, do not track the slower changes in reliability that accumulate over months.

Comparison. Only test-time adaptation and adaptive sensor fusion work without labels. Online sequential learning and continual learning need labels or pseudo-labels, radio-map maintenance relies on crowdsourced or inferred labels, and domain adaptation needs source labels with few or no target labels. Domain adaptation, continual learning, and test-time adaptation all compute gradients, the first through offline retraining and the other two on the device, so their cost is moderate to high. Online sequential updates, radio-map maintenance, and adaptive fusion are cheap on the device. Continual learning reduces forgetting but does not remove it, and test-time adaptation accumulates error over long streams. Only adaptive sensor fusion models per-modality reliability, and then only instantaneously from residuals.

Hierarchical feature learning has also been proposed to reduce the data needed when a model is rebuilt: WiDeep [67] captures environmental factors through hierarchical features and reports improved robustness to temporal variation, though it relies on an extensive initial collection.

The Online Feature Adaptation Module of this thesis (OFAM, Section 5.5.1) draws on the last two families. It performs decision-level adaptation: the $K = 3$ fusion weights are updated from self-supervised reliability estimates of the WiFi, inertial, and magnetic streams, while the network parameters remain frozen. The update is label-free, uses no gradients, and requires no retraining, so its cost is $O(K)$ per step and it carries no risk of forgetting the trained representation. It shares with test-time adaptation the premise that adaptation happens during operation on unlabeled data. It shares with adaptive filtering the explicit modeling

of per-modality reliability, extended here to reliability that varies over time. The corresponding limitation is that a frozen network cannot correct drift inside its learned representation.

2.4.3 Lack of Longitudinal Datasets

The systematic study of temporal degradation has been limited by the datasets available to study it. Table 2.1 compares the public datasets most relevant to this thesis on the dimensions that matter for temporal and multimodal research: the span and number of collection campaigns, the number of buildings, the sensor modalities, the ground-truth method, and public availability.

Table 2.1: Public indoor positioning datasets compared on temporal span, building coverage, sensor modalities, and ground truth. Single-campaign datasets cannot characterize drift, and the two IPIN editions lack centimeter-level ground truth. The longitudinal datasets are WiFi-only and single-building, and the recent multimodal datasets are single-campaign. The dataset of this thesis combines two campaigns one year apart, four buildings, three smartphone modalities, and centimeter-level ground truth.

Dataset	**Years and campaigns**	**Buildings**	**Modalities**	**Ground truth**	**Public**
UJIIndoorLoc [11]	2013; single campaign	3 (multi-floor)	WiFi RSS	Labelled reference points (building, floor, space, coordinates)	Yes
Mendoza-Silva et al. [38]	15 months, monthly collections (25 months in a later release)	1 (library)	WiFi RSS	Marked positions and orientations	Yes
Silva et al. [48]	Feb. 2019 to Mar. 2021, continuous, plus periodic site surveys	1 (university department)	WiFi RSS from fixed Raspberry Pi monitors and a mobile unit	Fixed monitor positions; surveyed points	Yes
IPIN competition [68, 2]	2016 and 2017 editions; one collection per site per edition	3	WiFi RSS, IMU, magnetometer, and other smartphone sensors	Floor-marked reference points with pre-measured coordinates	Yes
Abdalla et al. [69]	2025; single campaign	1 (one corridor, one floor)	WiFi RSS, IMU, magnetometer, CCTV images	1 m grid along a 39×49 m corridor	Yes
HYMN [70]	Single campaign (released 2026)	1 (44×18 m industrial hall with indoor–outdoor transition)	UWB, BLE, WiFi FTM, 5G, GNSS	Total station	Yes
This thesis (Chapter 3)	2021 and 2022; two six-week campaigns one year apart	4	WiFi RSS, IMU, magnetometer	AR spatial anchors validated against a total station (1.8 cm mean)	Yes

UJIIndoorLoc [11] established the multi-building, multi-floor benchmark for WiFi fingerprinting, but it is a single campaign. Mendoza-Silva et al. [38] addressed time directly with a library-building dataset collected monthly over 15 months at fixed positions and orientations. A later release of the same dataset extends the span to 25 months and 103,584 fingerprints [71]. Silva et al. [48] moved from monthly snapshots to continuous observation, recording approximately 7.4 million WiFi samples with fixed Raspberry Pi mon-

itors from February 2019 to March 2021 in one university building, complemented by periodic site surveys with a mobile unit. Both remain WiFi-only and single-building. The IPIN competition datasets [68, 2] added multimodal smartphone logs with floor-marked reference points and an evaluation protocol shared across teams, covering three buildings. Its 2016 and 2017 editions support the cross-year test of Section 5.6.9, but they lack centimeter-level ground truth and each edition is one collection per site. The two most recent datasets show where the field is moving. Abdalla et al. [69] combine WiFi RSS, smartphone inertial and magnetometer readings, and CCTV images on a 1 m grid along a 39 $\times$ 49 m corridor, in a single campaign with a single phone. HYMN [70] records UWB, BLE, WiFi FTM, 5G, and GNSS measurements in a 44 $\times$ 18 m industrial hall with an indoor–outdoor transition and total-station ground truth, again in a single campaign.

The comparison separates two kinds of benchmark. HYMN is a cross-technology benchmark: it holds the environment and the moment fixed and varies the positioning technology, a design suited to comparing ranging systems. The dataset of this thesis (Chapter 3) is a cross-time, cross-building, multimodal robustness benchmark: it holds the sensing platform fixed (WiFi RSS, IMU, and magnetometer from one smartphone model) and varies the building and the year. It has two campaigns one year apart across four buildings, with AR spatial anchors validated to 1.8 cm against a total station (Section 3.2.3). The 2025 survey of smartphone-based localization datasets [72], which reviews 20 datasets released between 2014 and 2024, shows that multimodal collection is now common. The combination of multi-year coverage, multiple buildings, multimodal sensing, and centimeter-level ground truth in one public dataset is not available elsewhere. This combination allows the relative degradation of WiFi, inertial, and magnetic measurements to be measured rather than assumed.

Two annual snapshots are a deliberate but limited design. They bound drift at the scale of a year, the interval between the two campaigns, but they do not observe intra-year dynamics such as semester occupancy cycles or infrastructure changes within the year. The continuous data of Silva et al. [48] complement them in this respect, at the cost of one building and one modality. The 90-day deployment simulation of Section 5.7.5 is built from the measured campaigns: it concatenates measured data segments and perturbs them with access-point removals and additions based on patterns observed in the 2021–2022 data. It is a stress test, not a third measurement. Synthetic drift of this kind is useful for testing an adaptation mechanism under controlled perturbations, but it cannot substitute for measured longitudinal data. The claims of this thesis about temporal robustness rest on the two measured campaigns.

2.4.4 Gap: Multi-Year Dataset Rarity

The limitation shared by the datasets in Table 2.1 is that no public dataset combines multi-year coverage, multiple buildings, synchronized WiFi, inertial, and magnetic measurements, and centimeter-level ground truth. Without such a dataset the research community cannot examine how the modalities degrade relative to one another, and adaptive fusion mechanisms cannot be validated against realistic deployment conditions. This thesis addresses the gap through the data collection methodology of Chapter 3.

2.5 Multi-Sensor Fusion Approaches

The integration of multiple sensor modalities represents a fundamental approach to improving indoor positioning accuracy, reliability, and robustness. This section examines fusion architectures and their limitations.

2.5.1 Sensor Modalities for Indoor Positioning

Modern indoor positioning systems leverage diverse sensor technologies. Inertial Measurement Units (IMU) provide high-frequency motion sensing through accelerometers and gyroscopes, enabling continuous motion tracking independent of external infrastructure [73]. Magnetometers detect magnetic field strength and direction, providing heading reference, though indoor magnetic fields are often distorted by metallic structures.

Vision-based sensors provide rich environmental information, with contemporary mobile AR frameworks including ARCore and ARKit representing mature Visual-Inertial SLAM implementations. Bluetooth Low Energy (BLE) beacons provide proximity-based positioning with low power consumption. Barometric sensors enable altitude estimation for floor-level determination in multi-story buildings. LiDAR sensors, increasingly available in consumer devices, provide precise distance measurements for 3D mapping.

WiFi, IMU, and magnetometer fusion has emerged as the predominant multi-modal approach due to sensor availability in smartphones and complementary characteristics. WiFi provides absolute position corrections that prevent accumulation of inertial navigation errors, while IMU data enables smooth tracking between WiFi position updates.

2.5.2 Fusion Architectures

Kalman filtering provides a principled framework for fusing sensor measurements with different characteristics and update rates. Extended Kalman Filters (EKF) handle non-linear relationships through linearization [73], while Unscented Kalman Filters (UKF) address linearization limitations through deterministic sampling. Chen et al. [74] developed an adaptive Kalman filter framework that fuses WiFi, smartphone sensors, and landmarks, achieving 1.2 meter median accuracy through careful tuning of process and measurement noise covariances. The approach demonstrates effective multi-modal integration but requires manual parameter tuning for each deployment environment, limiting scalability.

Particle filters provide non-parametric approaches that handle arbitrary probability distributions and non-linear measurement models. The particle filter framework is particularly well-suited for positioning applications with complex measurement models where the relationship between signal patterns and position may be highly non-linear. Within such filter-based frameworks, adaptive weighting schemes dynamically adjust sensor contributions in response to estimated measurement reliability: Li et al. [65] employ Mahalanobis-distance-based dynamic observation weighting in a robust complementary Kalman filter for UWB/IMU fusion, while Zhong et al. [66] adjust the weights of Wi-Fi and visual measurements in real time by monitoring measurement residuals. These works represent important steps toward adaptive fusion, but they respond to instantaneous measurement quality rather than long-term temporal changes in sensor reliability.

Graph-based SLAM approaches represent positioning and mapping problems as optimization problems over pose graphs, incorporating measurements from multiple sensor modalities as constraints.

Deep learning approaches to sensor fusion can automatically learn optimal combination strategies from training data. Multi-modal neural architectures process different sensor types through specialized network branches [13, 75, 14]. Attention mechanisms enable neural networks to adaptively weight different sensor inputs based on estimated reliability. Recurrent neural networks model temporal dependencies in multi-sensor data streams. Xue et al. [76] provide a comprehensive review of deep learning approaches for indoor localization, categorizing fusion strategies into early, late, and hybrid approaches with systematic comparison of accuracy-complexity tradeoffs. The review identifies a critical gap in methods that adapt fusion weights based on real-time sensor reliability assessment, noting that most deep learning approaches treat sensor modalities with fixed relative importance.

Figure 2.2 presents a taxonomy of sensor fusion architectures commonly employed in indoor positioning systems. Filter-based approaches provide principled probabilistic fusion through Kalman filtering or particle methods, offering well-understood convergence properties and uncertainty quantification. Learning-based approaches leverage neural architectures for automatic weight optimization, with early fusion concatenating features before processing, late fusion combining separate modality decisions, and attention mechanisms enabling context-dependent adaptive weighting. Hybrid approaches combine filter-based reliability assessment with learning-based adaptation, offering the mathematical rigor of probabilistic filtering with the flexibility of neural optimization. This thesis develops a hybrid approach that integrates online reliability estimation with adaptive fusion weighting.

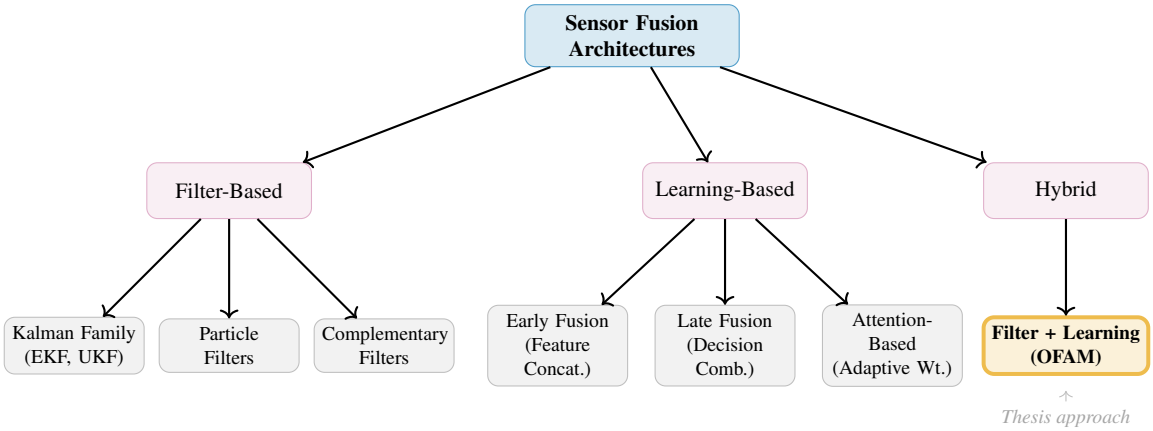


Figure 2.2: Taxonomy of sensor fusion architectures for indoor positioning. Filter-based approaches provide principled probabilistic fusion through Kalman or particle methods. Learning-based approaches leverage neural architectures for automatic weight optimization. Hybrid approaches, including the OFAM mechanism developed in this thesis, combine filter-based reliability assessment with learning-based adaptive weighting.

2.5.3 Hierarchical and Sequential Fusion

Sequential fusion processes sensor measurements in temporal order, naturally handling sensors with different update rates. High-frequency sensors like IMU provide predictions between lower-frequency WiFi

updates.

Hierarchical fusion architectures organize sensors into multiple levels, with higher levels combining results from lower-level fusion processes. Low-level fusion combines sensors with similar characteristics, while mid-level fusion combines inertial navigation with WiFi positioning.

2.5.4 Gap: Static Fusion Weights

However, existing multi-modal fusion approaches often rely on static weight optimization schemes that fail to adapt to changing environmental conditions and user activities. The integration of different sensor modalities typically uses fixed fusion parameters that do not account for the dynamic nature of indoor environments and varying sensor reliability.

Current fusion frameworks cannot track how the reliability of individual sensor modalities changes over long deployment horizons. WiFi signal quality varies with environmental conditions and access point visibility. IMU reliability depends on motion characteristics and calibration state. Magnetometer readings are affected by local magnetic disturbances that vary across locations. Without tracking long-term reliability changes, fusion systems cannot appropriately weight sensor contributions.

Moreover, the computational overhead of sophisticated fusion algorithms limits their applicability to resource-constrained mobile devices requiring real-time operation with minimal latency and power consumption.

2.6 AR-Based Indoor Localization

Augmented reality technologies have emerged as both positioning consumers and positioning enablers, representing a paradigm shift in ground truth establishment and spatial reference systems.

2.6.1 AR Tracking Technologies

Contemporary mobile AR frameworks including Google's ARCore and Apple's ARKit represent mature implementations of Visual-Inertial Odometry (VIO) and SLAM. These systems achieve centimeter-level relative tracking accuracy by combining camera imagery with IMU data, providing complementary strengths where vision offers long-term stability and IMU data enables tracking during rapid motions or visual occlusions [24, 25]. Qin et al. [77] developed VINS-Mono, characterizing the accuracy and robustness of monocular visual-inertial state estimation across diverse environments. The system achieves 0.5–2.0% drift relative to trajectory length under favorable conditions, though accuracy degrades in visually repetitive environments and during rapid rotations. This work provides important accuracy characterization that informs the reliability bounds of AR-based ground truth systems.

Microsoft HoloLens and similar head-mounted displays implement sophisticated spatial mapping capabilities, creating detailed 3D mesh representations of indoor environments. The HoloLens 2 spatial understanding system employs depth sensing and SLAM to build persistent spatial maps that can be shared across devices and sessions. Poulose and Han [78] demonstrated hybrid indoor localization that combines

smartphone camera imagery with IMU sensor data, exploiting the complementary strengths of visual and inertial sensing on commodity devices. Such camera–inertial approaches illustrate the potential of vision-aided smartphone positioning for both localization and ground truth establishment, though real-time visual processing places considerable computational demands on mobile hardware.

Cloud-based spatial anchoring services, including Azure Spatial Anchors and Google Cloud Anchors, enable persistent positioning across sessions and devices by storing visual features and spatial relationships in cloud databases.

2.6.2 AR for Ground Truth Establishment

Traditional ground truth establishment for positioning system evaluation requires specialized equipment including optical motion capture systems, ultrasonic beacons, or robotic platforms with precise odometry. These approaches involve substantial costs and professional setup, limiting the scalability of positioning system evaluation.

AR-based ground truth systems offer cost-effective alternatives by leveraging consumer AR devices as high-precision spatial reference systems. The spatial mapping capabilities of devices like HoloLens 2 can provide centimeter-level accuracy across large indoor areas, potentially replacing expensive motion capture infrastructure for ground truth establishment.

2.6.3 Gap: AR for Fingerprint Database Creation

However, AR spatial mapping has not been systematically leveraged for creating high-accuracy fingerprint databases at scale. While AR devices provide the positional accuracy needed for ground truth, no comprehensive framework exists for integrating AR-based ground truth with multi-sensor fingerprinting data collection across multiple buildings and extended time periods.

The potential of AR technologies to address the fundamental challenge of scalable, accurate ground truth establishment for fingerprinting research remains largely unexplored. A systematic approach combining AR spatial mapping with coordinated multi-sensor data collection could transform the practicality of creating comprehensive positioning datasets.

2.7 Deep Learning Architectures for Localization

The period from 2020 to 2025 has witnessed widespread adoption of deep learning architectures for indoor positioning [79, 80, 81, 82, 83]. This transformation has been driven by increased computational power of mobile devices, availability of fingerprinting datasets, and advances in neural architectures suited for spatial-temporal data.

2.7.1 Spatial Feature Extraction

Convolutional Neural Networks (CNNs) have been extensively applied to spatial feature extraction from WiFi fingerprints and radio signal patterns. CNNs excel at identifying localized patterns and spatial features

in signal strength distributions [84]. The adaptation of CNNs to fingerprinting applications requires careful consideration of input data representation, with signal strength measurements arranged in various spatial organizations.

Deep Neural Networks (DNNs) provide fully-connected architectures that learn complex non-linear mappings between signal measurements and position coordinates. Architecture design typically involves multiple hidden layers, with regularization techniques including dropout and batch normalization preventing overfitting.

2.7.2 Temporal Sequence Modeling

Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, have shown superior performance in capturing temporal dependencies in positioning data [85, 86]. LSTM networks address the challenge of modeling sequential patterns in location estimation and trajectory tracking. Advanced implementations include vanilla RNN, LSTM, GRU [87], bidirectional RNN, and bidirectional LSTM.

The integration of temporal information through RNNs enables continuous position tracking with improved accuracy compared to independent position estimates. LSTM gating mechanisms provide selective memory capabilities that maintain relevant positioning information across extended time intervals.

The computational efficiency of recurrent architectures for real-time mobile deployment remains an important consideration, with recent work on edge computing demonstrating viable approaches for on-device inference [88, 89].

2.7.3 Probabilistic Feature Encoding

Variational Autoencoders (VAEs) have gained prominence for dimensionality reduction and probabilistic feature encoding [90, 91]. VAEs introduce a probabilistic framework to traditional autoencoders, learning latent representations that capture the underlying structure of signal data while enabling uncertainty quantification.

For feature extraction, VAE implementations use learned latent spaces for eliminating noise, revealing discriminative representations, and capturing patterns that improve downstream positioning. The probabilistic nature of VAEs enables uncertainty quantification in the feature extraction process.

2.7.4 Recent Fingerprinting Architectures

Building on these foundational architectural families, the period since 2021 has produced increasingly specialized end-to-end fingerprinting systems. Convolutional autoencoders have been paired with CNN classifiers to compress fingerprints for edge deployment, as in CAE-CNNLoc [92], which targets resource-constrained devices while preserving accuracy. Neural architecture search has been applied to localization through NAS-FL [93], which automatically discovers network topologies tuned to fingerprint data rather than relying on hand-designed architectures. Hybrid signal models have also emerged: the Wi-Fi RSS–RTT approach of Xu et al. [36] dynamically switches between received-signal-strength and round-trip-time

measurements according to channel conditions, exploiting the IEEE 802.11mc fine-timing-measurement capability available on recent hardware.

In parallel, WiFi and inertial fusion has matured into competitive end-to-end systems. Fusion-DHL [94] combines WiFi, inertial, and floorplan information to reconstruct dense location histories, while earlier smartphone fusion frameworks established the benefit of combining inertial dead reckoning with WiFi RSSI measurements [95]. These systems achieve sub-1.5 m accuracy under favorable conditions but, in common with the WiFi-only methods above, rely on static model parameters or fixed fusion strategies that do not adapt to the temporal drift and sensor-reliability variations central to this thesis. The probabilistic temporal architecture and online adaptation mechanism developed in Chapters 4 and 5 are evaluated directly against representatives of both the WiFi-only and WiFi+IMU paradigms (Section 5.6.8), achieving substantially lower error through reliability-weighted adaptation.

2.7.5 Transformer Architectures

Transformer architectures represent the current frontier in sequence modeling, having achieved remarkable success across natural language processing, computer vision, and increasingly, wireless signal processing [13]. The core innovation of transformers lies in the self-attention mechanism, which computes relationships between all positions in a sequence through learned query, key, and value projections. Unlike recurrent architectures that process sequences sequentially, self-attention enables parallel computation across all sequence positions, capturing both local and long-range dependencies without the gradient propagation challenges inherent to RNNs [96].

The attention mechanism computes weighted combinations of values based on query-key similarities, enabling the network to dynamically focus on relevant portions of the input sequence. For positioning applications, this mechanism can potentially learn which signal measurements, access points, or temporal segments are most informative for location estimation. The cost of self-attention grows quadratically with sequence length, compared with the linear cost of recurrent processing [96]. This matters for long sequences but, as discussed below, not for the ten-step windows used in this thesis.

Radio Foundation Models: The paradigm of pre-training large neural networks on massive unlabeled datasets has been explored for wireless signal processing through Radio Foundation Models [17]. These approaches pre-train transformer architectures on large-scale wireless channel measurements, learning general representations of radio propagation that can be fine-tuned for downstream positioning tasks. The pre-training phase captures statistical regularities in wireless signals across diverse environments, potentially enabling transfer learning to new deployment settings with reduced labeled data requirements. However, the success of foundation model approaches depends critically on access to large-scale pre-training datasets spanning diverse building types, construction materials, and WiFi infrastructure configurations. For single-institution deployments with limited data collection resources, the practical benefits of foundation models remain uncertain, as the pre-training data distribution may not adequately represent the target deployment environment.

Generative Approaches: WiFiGPT [18] represents concurrent work adapting decoder-only, GPT-style transformer language models to WiFi telemetry, treating RSSI and fine-timing-measurement inputs as token sequences for indoor localization. The approach treats positioning as a sequence modeling problem, generating position predictions autoregressively conditioned on input signal sequences. The generative framework enables handling of environment variations through learned representations that capture the underlying structure of radio propagation patterns. While this approach demonstrates the versatility of transformer architectures, it requires substantial training data to learn the joint distribution of signals and positions, and the autoregressive generation process may introduce latency unsuitable for real-time applications.

NLOS Mitigation: The challenge of non-line-of-sight (NLOS) signal propagation, where direct paths between transmitters and receivers are obstructed, represents a fundamental difficulty in radio-based positioning. L-SwiGLU ViT [97] is a concurrent vision-transformer-based approach that replaces the standard transformer feed-forward network with a Linear SwiGLU layer, targeting NLOS-dominated wireless environments with distributed sensors. The modification addresses the observation that standard transformer architectures may overfit to direct-path signal characteristics present in training data, performing poorly when deployed in environments with different obstruction patterns. While preliminary results suggest improved NLOS handling, evaluation has been limited to single-building deployments over short temporal spans, leaving questions about generalization across diverse architectural contexts.

BLE and Magnetic Field Applications: Transformer architectures have been applied beyond WiFi to other positioning modalities. TransAoA [98] adapts transformers for Bluetooth Low Energy angle of arrival estimation, leveraging attention mechanisms to weight contributions from multiple antenna array elements while handling the inherent noise in phase measurements. Rafique et al. [99] fuse magnetic field signals with visual data using a vision transformer, projecting magnetometer measurements onto image representations so that both modalities can be processed jointly for indoor localization. These applications demonstrate the architectural flexibility of transformers across sensor modalities, though each faces the common challenges of data requirements and computational complexity.

State-Space Models: Structured state-space models, in particular the selective state-space model Mamba [100], offer a linear-time alternative to attention. The sequence is processed by a learned linear dynamical system whose parameters depend on the input. This permits parallel training over a sequence and constant-cost recurrent inference at deployment, which suits streaming positioning. Two reasons kept them out of this thesis. Mamba was published in December 2023 [100]. At that time there was no evidence of deployment on mobile hardware, since the selective scan relies on custom accelerator kernels. In addition, the ten-step window used here gives a linear-time model no cost advantage over an LSTM, which is already linear in sequence length. State-space models are therefore noted as a candidate for the longer windows discussed in Section 6.4.2.

Computational Cost: The quadratic cost of attention is often cited against transformers on mobile devices, but it is not the main consideration for this thesis. With a window of $T = 10$ steps, attention over ten positions is a small fraction of the per-step computation of any architecture considered. The case for the recurrent encoder in Section 4.4.5 does not rest on it. What matters for streaming inference is the form of the computation. An LSTM carries a fixed-size state forward and processes each new observation at constant cost. An encoder with attention over the window must instead recompute or cache attention at every step and, unless masked causally, would attend to future observations that are unavailable online. Efficient attention variants [96] reduce the cost of long sequences. They do not change this picture for short ones.

Sample Complexity and Window Length: Scaling-law studies find that the advantage of attention-based models over recurrent models grows with model size and context length [101], and such models are usually trained on far larger datasets than positioning studies can supply. The dataset collected for this thesis, with 9,659 samples across four buildings and two years, is large by positioning standards but small by the standards at which attention-based models are trained. The short window adds to this. Attention is most useful for long-range dependencies, and a 20-second window of ten fused observations has few of them.

These considerations, rather than the asymptotic cost of attention, are the basis for the recurrent temporal encoder of Chapter 4: the VAE component provides probabilistic feature encoding, and the LSTM component models the short causal window with data efficiency suited to a dataset of this scale (Section 4.4.5). The choice reflects the data and deployment constraints of this thesis rather than a limitation of attention-based or state-space models. No Transformer baseline was trained, and this is recorded as a limitation in Section 6.3. As positioning datasets grow and mobile inference support for these architectures matures, the balance may shift.

2.7.6 Uncertainty Quantification in Deep Localization

Reliable positioning systems must communicate not only position estimates but also the confidence associated with those estimates. Uncertainty awareness enables downstream applications to make informed decisions, particularly in safety-critical scenarios such as emergency response, autonomous navigation, and accessibility assistance where erroneous position estimates without appropriate uncertainty flags can lead to hazardous situations [102]. The ability to distinguish between confident predictions and unreliable estimates is essential for adaptive decision-making in dynamic indoor environments.

A fundamental distinction in uncertainty quantification separates epistemic uncertainty from aleatoric uncertainty [103]. Epistemic uncertainty, also termed model uncertainty, arises from limitations in the model itself due to insufficient training data, inadequate model capacity, or distribution shift between training and deployment environments. This uncertainty type is reducible through additional data collection or improved modeling. In the context of indoor positioning, epistemic uncertainty manifests when the system encounters locations or environmental conditions poorly represented in the training fingerprint database, such as newly constructed areas or significantly modified spaces.

Aleatoric uncertainty, in contrast, represents inherent noise in the measurement process that cannot be reduced through improved modeling. For WiFi-based positioning, aleatoric uncertainty stems from temporal signal fluctuations, multipath interference, device hardware variations, and environmental dynamics such as occupant movement and door positions. This irreducible uncertainty bounds the achievable positioning accuracy regardless of model sophistication and must be characterized rather than eliminated.

Bayesian Neural Networks (BNNs) provide a principled framework for uncertainty quantification by treating network weights as probability distributions rather than point estimates [104, 102]. The posterior distribution over weights induces a predictive distribution over outputs, enabling direct quantification of model uncertainty. However, exact Bayesian inference in neural networks is computationally intractable, requiring approximation methods for practical application. The computational overhead of variational inference and sampling-based approaches presents significant challenges for real-time mobile deployment where latency and power consumption are constrained.

Monte Carlo Dropout provides a practical approximation to Bayesian inference by applying dropout at inference time and aggregating predictions across multiple stochastic forward passes [104]. This approach interprets dropout as approximate variational inference in a deep Gaussian process, enabling uncertainty estimation with minimal computational overhead beyond standard network inference. For positioning applications, prediction variance across dropout samples serves as an uncertainty proxy. While computationally efficient compared to full Bayesian methods, Monte Carlo Dropout requires multiple forward passes that may not satisfy strict real-time latency requirements on resource-constrained devices.

Deep ensembles aggregate predictions from multiple independently trained networks, using ensemble disagreement as a measure of uncertainty [105]. Ensemble diversity, achieved through random initialization and training data shuffling, captures epistemic uncertainty through regions of model disagreement. The epistemic–aleatoric decomposition framework of Kendall and Gal [103], together with the deep-ensemble uncertainty estimation of Lakshminarayanan et al. [105], provides the foundational methods that, when applied in positioning contexts, enable distinguishing locations where confidence degrades due to inadequate training coverage from those affected by inherent signal variability. While ensembles scale more favorably than full Bayesian methods, the computational and memory overhead of maintaining multiple models remains a practical deployment constraint.

Recent work has applied uncertainty quantification specifically to indoor positioning with promising results. The RLoc system [16] exploits quantified angle-of-arrival uncertainty in learning-based localization, surpassing state-of-the-art approaches by an average of 36% in in-domain scenarios and 20% in cross-domain scenarios. This improvement stems from the system's ability to identify and appropriately handle unreliable predictions rather than treating all position estimates with equal confidence. Heteroscedastic approaches that predict location-dependent variance enable adaptive uncertainty estimates that vary across the positioning environment [103, 106], capturing the spatial heterogeneity of signal quality and fingerprint discriminability.

Two further families have entered the positioning literature recently. Conformal prediction wraps any point predictor in a calibration step that converts a nonconformity score on held-out data into prediction sets with a guaranteed marginal coverage level. It makes no distributional assumption beyond exchangeabil-

ity of the calibration and test data. It has been applied to deep fingerprinting models on UJIIndoorLoc to produce prediction sets with a target coverage [107], to quantify the ambiguity of fingerprint matches [108], and to combined round-trip-time and signal-strength positioning [109]. The guarantee is useful for safety-related use, but it holds only under exchangeability. A calibration set drawn in one year does not certify coverage in the next once the signal distribution has drifted, so conformal methods would need recalibration on the same schedule as the radio map. Evidential deep regression [110] takes a different route, training the network to output the parameters of a Normal-Inverse-Gamma prior over the target so that aleatoric and epistemic uncertainty are read off in a single forward pass without sampling or ensembles. Its cost is one wider output layer, and its estimates depend on an evidential regularizer whose strength must be tuned. The heteroscedastic heads mentioned above add one variance output per coordinate trained with the Gaussian negative log-likelihood. Their cost on a mobile device is negligible next to the encoder, which makes them the cheapest route to input-dependent aleatoric uncertainty. They would be a natural addition to the architecture of Chapter 4, which instead derives its aleatoric term from the reconstruction error (Section 4.5.4). None of these methods was part of the design evaluated in this thesis. They are taken up as future work in Section 6.4.2.

The reliability weighting mechanism developed in this thesis (Chapter 5) operates on principles related to uncertainty quantification. The Online Feature Adaptation Module (OFAM) assesses sensor quality metrics including signal-to-noise ratios and measurement variance, relating directly to aleatoric uncertainty characterization. The adaptive weighting scheme adjusts sensor contributions based on contextual reliability assessment, implementing a form of epistemic uncertainty handling where sensors with uncertain or degraded measurements receive reduced influence. The architecture provides uncertainty estimates alongside position estimates that correlate with the observed error ($r = 0.73$; Chapter 4), enabling uncertainty-aware downstream processing. Unlike computationally intensive ensemble or Bayesian approaches, OFAM achieves uncertainty-responsive behavior through efficient signal quality assessment suitable for real-time mobile deployment.

2.7.7 Gap: Probabilistic Temporal Architectures

Probabilistic feature encoding and temporal sequence modeling have each been applied to fingerprinting, but rarely together: variational encoders are used without sequence dynamics, and recurrent models are trained as point estimators without a representation of confidence. A unified architecture that combines the two within a mobile inference budget is the gap addressed in Chapter 4.

2.8 Research Gaps and Challenges

The preceding sections stated each research gap where the evidence for it was reviewed: environmental adaptation mechanisms (Section 2.3.4), multi-year datasets (Section 2.4.4), static fusion weights (Section 2.5.4), AR ground truth (Section 2.6.3), and probabilistic temporal architectures (Section 2.7.7). This section does not restate them. It places representative systems side by side (Section 2.8.1), notes challenges that fall outside the five gaps (Section 2.8.2), and maps each gap to the chapter that addresses it

(Section 2.8.3).

2.8.1 Comparison of Representative Systems

Table 2.2 presents a comparative summary of representative indoor localization systems across the literature, highlighting the evolution from single-sensor to multi-modal approaches and the gradual improvement in positioning accuracy.

Table 2.2: Comparison of representative indoor localization systems with thesis contribution mapping. The survey reveals that no prior system combines multi-modal sensing, multi-building evaluation, and online temporal adaptation, motivating the integrated approach developed in this thesis.

System	Sensors	Buildings	Temporal	Accuracy (m)	Adaptation	This Thesis
RADAR [7]	WiFi	1	Single	2.90	None	Dataset (Ch 3)
Horus [8]	WiFi	1	Single	0.42–0.64‡	None	VAE encoding (Ch 4)
DeepFi [12]	CSI	1	Weeks	0.94–1.81‡	None	Multi-year data (Ch 3)
CNNLoc [45]	WiFi	1	Single	1.50	None	Probabilistic (Ch 4)
LSTM-Loc [86]	WiFi+IMU	1	Months	1.20	None	OFAM fusion (Ch 5)
RFM [17]	5G CSI	1	Single	0.40–0.96†	None	Temporal adapt. (Ch 5)
NAS-FL [93]	WiFi	Multi	Single	1.65	None	Modern baseline (Ch 5)
Fusion-DHL [94]	WiFi+IMU	Multi	Single	1.08	None	Modern baseline (Ch 5)
This thesis	WiFi+IMU+Mag	4	2 years	0.59	Online (OFAM)	–

†90th-percentile error (CE90); model fine-tuned separately per site, with the range spanning 500 to 100,000 training samples across two industrial sites.

‡Range of mean errors across the two testbeds evaluated in the original publication.

Notably, the temporal span of evaluation remains limited across prior work, with most studies reporting single-point accuracy rather than longitudinal stability. Systems evaluated over extended periods demonstrate the importance of temporal considerations but do not incorporate mechanisms for adapting to environmental drift. The absence of online adaptation mechanisms in existing systems represents a significant gap that limits practical deployment, as environmental changes inevitably degrade positioning accuracy over time.

2.8.2 Additional Challenges

Beyond these core gaps, several further challenges shape practical deployment, a subset of which this thesis is able to address. Device heterogeneity is a long-standing obstacle, since fingerprints collected on one handset rarely transfer cleanly to another because of differences in antenna gain and radio chipsets [111]; the longitudinal collection in this thesis uses a single smartphone model, and device heterogeneity is recorded as a limitation in Chapter 6. The scalability of site surveys and the cost of labeling reference points constitute a second challenge, one that crowdsourced radio-map construction has sought to mitigate [56]; the AR ground-truth methodology of Chapter 3 offers a complementary route by replacing manual measurement with centimeter-accurate spatial anchors. A third challenge is the lack of standardized, cross-dataset benchmarking, which competitive evaluations [68] and shared benchmark datasets [11] have begun to redress

but which still leaves many systems reported only on private data [112]; this thesis responds through the independent IPIN-benchmark evaluation in Section 5.6.9, which tests generalization across distinct office, lecture-hall, and industrial environments and across an independent year-over-year temporal gap.

Other challenges fall outside the scope of this thesis and are noted here as open problems rather than contributions. The energy and power cost of continuous sensing and on-device adaptation is a genuine constraint for always-on positioning, and while the architectures developed here are designed to respect mobile latency budgets [88], this thesis does not claim to solve, measure, or optimize their energy consumption, which remains an open question for future work. Likewise, the security and privacy of location data raise substantial concerns, since fingerprint traces can reveal sensitive patterns of movement and behavior [113, 114]; these issues are not addressed by the present work and are revisited only as a direction for future research in Chapter 6.

2.8.3 Research Gap Summary

Table 2.3 synthesizes the five critical research gaps identified throughout this review and maps each to the specific thesis contributions that address them. This explicit mapping demonstrates how the methodology developed in subsequent chapters systematically fills the identified gaps in indoor positioning research.

Table 2.3: Research gaps identified across the indoor positioning literature and corresponding thesis contributions. All five gaps – longitudinal datasets, temporal adaptation, static fusion weights, probabilistic temporal modelling, and AR ground truth – are addressed by the methodology in Chapters 3–5.

Research Gap	**Prior Work Limitation**	**This Thesis**
Longitudinal datasets	Single building, WiFi only; no multi-modal multi-year coverage	Multi-year, 4-building dataset with WiFi/IMU/Mag (Ch 3)
Temporal adaptation	No label-free run-time adaptation that tracks long-term reliability drift; periodic recalibration required	OFAM adaptive fusion with reliability weighting (Ch 5)
Static fusion weights	Fixed sensor weighting; no real-time reliability assessment	Reliability-based OFAM with context-aware adaptation (Ch 5)
Probabilistic temporal	VAE or LSTM separately; no unified architecture	VAE-LSTM integration for uncertainty-aware positioning (Ch 4)
AR ground truth	Not leveraged for fingerprint database creation	HoloLens spatial mapping methodology (Ch 3)

The research gaps table highlights the systematic nature of the thesis contributions. The longitudinal dataset gap is addressed through a comprehensive two-year data collection effort spanning four buildings with diverse architectural characteristics (Chapter 3). The temporal adaptation and static fusion weight gaps are jointly addressed by the Online Feature Adaptation Module (OFAM), which provides dynamic reliability assessment and adaptive sensor weighting (Chapter 5). The probabilistic temporal architecture gap is filled by the VAE-LSTM model that combines variational encoding for uncertainty quantification with recurrent processing for temporal dependencies (Chapter 4). Finally, the AR ground truth gap is addressed

through a novel methodology leveraging HoloLens spatial mapping for scalable, high-accuracy reference point establishment (Chapter 3).

2.9 Summary

This comprehensive review has examined the evolution and current state of indoor positioning technologies, with particular focus on WiFi fingerprinting, multi-modal sensor fusion, and deep learning architectures. The review reveals significant progress in positioning accuracy and robustness while identifying critical challenges that continue to limit practical deployment.

The fundamental positioning techniques examined (proximity, lateration, SLAM, fingerprinting, and dead reckoning) each offer distinct advantages and limitations. Fingerprinting has emerged as the predominant approach for high-accuracy applications due to its ability to handle complex indoor propagation effects, though it faces challenges related to database maintenance and environmental adaptation that remain inadequately addressed.

The evolution from traditional pattern matching to deep learning architectures represents a significant advancement in positioning capability. CNNs and RNNs enable sophisticated feature extraction and temporal modeling that surpass traditional approaches. However, the integration of probabilistic feature encoding with temporal sequence modeling for fingerprinting remains underexplored, and computational tractability for mobile deployment continues to constrain architectural choices.

Multi-modal sensor fusion has proven essential for robust positioning, with WiFi, IMU, and magnetometer integration representing the most successful combination due to sensor availability and complementary characteristics. Yet existing fusion approaches rely on static weighting schemes that cannot adapt to the dynamic reliability characteristics of different sensors across environments and time.

Most critically, the temporal stability of positioning systems over extended deployment periods remains poorly understood due to the absence of suitable longitudinal datasets. Without multi-year, multi-building datasets that capture diverse degradation patterns, the research community cannot systematically study or address the temporal adaptation problem that fundamentally limits practical deployment.

The integration of augmented reality technologies with positioning systems represents an emerging paradigm that addresses ground truth establishment challenges, yet AR spatial mapping has not been systematically leveraged for creating high-accuracy reference datasets at scale.

These gaps, consolidated in Table 2.3, define the research space addressed in the following chapters.

Chapter 3

Dataset and AR Framework

3.1 Introduction

The development of robust indoor localization systems requires high-quality datasets with accurate ground truth positions. However, establishing precise reference points in indoor environments presents significant practical challenges that have constrained the scope and scale of positioning research for decades. Traditional approaches rely on either manual surveying with total stations, a labor-intensive process requiring specialized equipment and trained personnel, or pre-installed infrastructure such as ultrawideband beacons or optical tracking systems that demand substantial capital investment and deployment effort [11]. These requirements have limited most indoor positioning evaluations to small-scale testbeds in single buildings, typically covering areas of a few hundred to at most a few thousand square meters, over observation periods rarely exceeding a few months [38].

The consequences of this limitation extend beyond mere inconvenience. Without large-scale, longitudinal datasets, fundamental questions about indoor localization remain inadequately addressed: How do fingerprint-based systems degrade over extended time periods? What environmental factors most significantly impact positioning accuracy? Can models trained in one building generalize to architecturally different environments? These questions are critical for practical deployment of indoor positioning systems, where operational lifetimes span years and environments exhibit continuous change through infrastructure modifications, occupancy pattern shifts, and seasonal variations. The research community's reliance on short-term, small-scale evaluations has left these questions largely unanswered, creating a gap between laboratory demonstrations and real-world requirements.

The challenge is further compounded by the multi-modal nature of modern localization approaches. Contemporary systems increasingly fuse signals from WiFi access points, inertial measurement units, magnetometers, and other sensors to achieve robust positioning across diverse conditions [115, 116]. Each sensor modality requires its own calibration and ground truth association, multiplying the complexity of data collection. A WiFi fingerprint survey alone may require visiting hundreds of reference points; adding synchronized IMU data, magnetometer readings, and barometric pressure measurements across multiple buildings and collection epochs quickly becomes impractical with traditional methods.

This chapter presents an augmented reality (AR) based approach that addresses these challenges by

leveraging Microsoft HoloLens 2's spatial mapping capabilities for precise ground truth establishment. The key innovation lies in decoupling the ground truth acquisition process from the runtime positioning system: during an offline calibration phase, the HoloLens 2 establishes a centimeter-accurate spatial reference frame through its depth cameras and simultaneous localization and mapping (SLAM) system, creating virtual anchors that persist across sessions via cloud storage. During data collection, operators navigate to these anchor locations and record synchronized multi-modal sensor data from commodity smartphones, efficiently building comprehensive fingerprint databases without specialized surveying equipment.

This framework is applied to construct a longitudinal dataset spanning four architecturally diverse buildings at York University, comprising 28,822 square meters of indoor space. Data collection campaigns in 2021 and 2022 capture the temporal evolution of fingerprint signatures as the campus environment underwent infrastructure changes, furniture rearrangements, and occupancy pattern shifts. This dataset enables the first systematic study of multi-modal fingerprint stability across extended time periods in a diverse multi-building testbed.

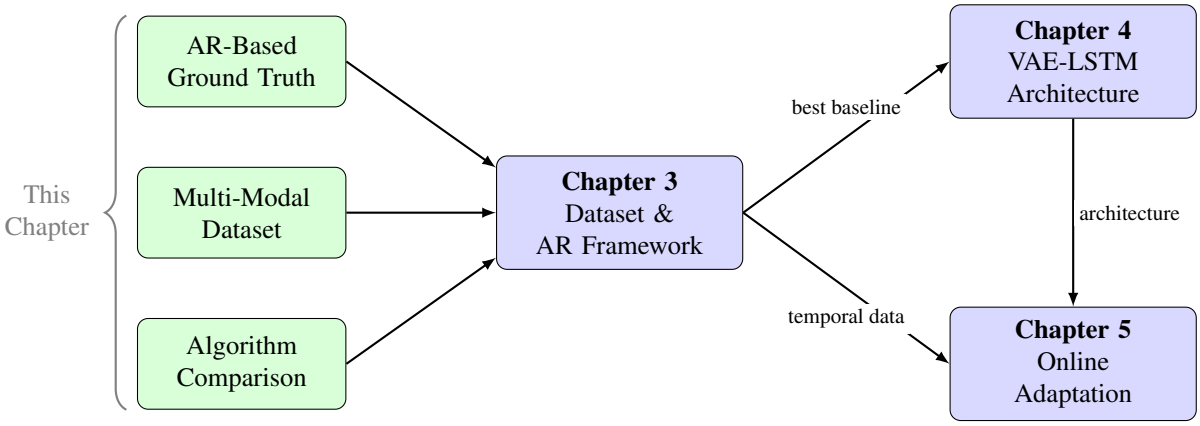


Figure 3.1: Chapter roadmap showing how the dataset and AR framework contributions feed into subsequent technical chapters. Chapter 3 establishes the data collection methodology, multi-building testbed, and baseline algorithm comparison that inform the architectural investigations in Chapter 4 and the online adaptation mechanisms in Chapter 5. This layered structure ensures each subsequent chapter builds on validated data and baseline results.

Figure 3.1 illustrates how this chapter's contributions serve as the foundation for the subsequent technical chapters. The AR-based data collection framework and resulting longitudinal dataset enable the detailed architectural analysis in Chapter 4, where the VAE-LSTM architecture that emerges from the baseline comparison is examined. The temporal dataset further supports the online adaptation mechanisms developed in Chapter 5, providing the year-over-year variation data needed to evaluate continuous learning approaches.

The remainder of this chapter is organized as follows. Section 3.2 describes the AR-based data collection framework, including the HoloLens 2 hardware platform, spatial anchor system, and data collection pipeline. Section 3.3 details the multi-modal sensor acquisition approach, covering WiFi signal collection, inertial measurements, magnetometer data, and sensor specifications. Section 3.4 presents the four-building testbed at York University, characterizing each environment's infrastructure including WiFi access point

density, magnetic field characteristics, and construction materials that influence positioning performance. Section 3.5 describes the longitudinal dataset spanning 2021–2022, documenting the collection protocol, the campaign details including the COVID-19 conditions of the 2021 campaign, the year-over-year changes, and the ethics and privacy basis of the collection. Section 3.6 details the preprocessing pipeline that transforms raw sensor data into discriminative features through signal normalization, missing value imputation, and feature engineering. Section 3.7 combines the collection and preprocessing steps into an ordered procedure for a new site. Section 3.8 presents a comparative evaluation of seven localization algorithms, motivating the selection of the VAE-LSTM architecture for detailed development in subsequent chapters. Section 3.9 states the availability of the dataset and the status of the collection software. Finally, Section 3.10 summarizes the chapter's contributions and motivates the architectural investigations in subsequent chapters.

3.2 AR-Based Data Collection Framework

The methodology extends the AR-based fingerprinting approach introduced in prior work [117], which demonstrated the feasibility of using mixed reality headsets for simplified reference point establishment. While initial efforts focused solely on WiFi signal strength fingerprinting in single buildings, this thesis presents a comprehensive multi-modal framework that builds upon the fundamental principle that indoor environments exhibit distinctive sensor signatures that remain sufficiently stable over extended periods to enable reliable positioning [19]. The enhanced framework leverages:

$$\hat{\mathbf{p}} = f(\mathcal{S}_{AR}, \mathcal{F}_{multi}, \mathcal{T}) \tag{3.1}$$

where $\hat{\mathbf{p}}$ represents the localization estimate, $\mathcal{S}_{AR}$ denotes the AR-established spatial reference frame, $\mathcal{F}_{multi}$ encompasses the multi-modal fingerprint features (WiFi, magnetometer, IMU), $\mathcal{T}$ accounts for temporal variations, and f denotes the localization mapping. The innovation in the proposed approach lies in the decoupling of ground truth establishment from runtime positioning: by utilizing Microsoft HoloLens 2's precise spatial mapping capabilities during an offline calibration phase, centimeter-level accuracy in reference point definition is achieved while maintaining the practical advantages of smartphone-based positioning during runtime [15].

3.2.1 HoloLens 2 Hardware Platform

The Microsoft HoloLens 2 serves as the cornerstone of the ground truth establishment system. Released in 2019 as the second generation of Microsoft's mixed reality headset, the HoloLens 2 represents a significant advancement in untethered spatial computing, combining multiple sensing modalities into a self-contained wearable device suitable for extended indoor mapping sessions.

Depth Sensing and Spatial Mapping: The HoloLens 2 employs a time-of-flight depth sensor operating at 1 megapixel resolution with a field of view of $120^\circ \times 120^\circ$. This sensor enables real-time 3D reconstruction of indoor environments, generating spatial meshes with millimeter-level precision at ranges up to 3.5 m. The

device performs simultaneous localization and mapping (SLAM) using this depth data in conjunction with four visible-light tracking cameras positioned at the corners of the visor, providing 360° coverage for head tracking and world-relative positioning. The tracking cameras operate at 30 frames per second with global shutter technology, enabling robust visual-inertial odometry even during rapid head movements.

Inertial Measurement Unit: The integrated IMU comprises a 3-axis accelerometer, 3-axis gyroscope, and 3-axis magnetometer, sampling at rates up to 1 kHz. The IMU data is fused with visual tracking information using an extended Kalman filter, providing position and orientation estimates that are robust to temporary visual occlusions [118]. Microsoft reports tracking accuracy of less than 1 cm root mean square error for typical indoor environments [15], and independent evaluations confirm 1–3 cm positional accuracy under favorable conditions [119].

Additional Sensors: Beyond the primary spatial mapping capabilities, the HoloLens 2 includes eye tracking cameras operating at 30 Hz with 1.5° accuracy, enabling gaze-based interaction with virtual content. An ambient light sensor adjusts display brightness based on environmental conditions. The device also includes an 8 megapixel RGB camera for photo and video capture, though this capability is not utilized in the ground truth establishment protocol.

Processing and Connectivity: The HoloLens 2 is powered by a Qualcomm Snapdragon 850 compute platform paired with Microsoft's custom Holographic Processing Unit (HPU) for real-time spatial mapping and hologram rendering. The device includes WiFi 802.11ac and Bluetooth 5.0 for network connectivity, enabling cloud-based anchor persistence and collaborative mapping scenarios. Battery life is approximately 2–3 hours of active use, sufficient for comprehensive building surveys when managed with brief charging breaks.

The combination of high-precision depth sensing, robust visual-inertial tracking, and self-contained processing makes the HoloLens 2 uniquely suited for indoor ground truth establishment. Unlike optical tracking systems that require extensive infrastructure installation, the HoloLens 2 operates independently, building its spatial understanding as the operator navigates through the environment. This portability enables data collection across multiple buildings and floors without the logistical complexity of deploying and calibrating fixed infrastructure.

Complete specification of sensor hardware characteristics is essential for reproducibility and for understanding the inherent limitations of the collected data. Table 3.1 provides the detailed technical specifications for each sensor modality, documenting the measurement ranges, resolutions, and sampling rates that inform preprocessing parameter selection. These specifications establish the baseline data quality from which the localization algorithms must operate, and enable other researchers to assess the applicability of the presented methods to alternative hardware platforms.

Table 3.1: Sensor hardware specifications for the HoloLens 2 device used to establish ground truth. Values obtained from Microsoft technical documentation and empirical characterization during calibration experiments. The smartphone sensor streams used for fingerprint collection run at different rates (WiFi scans at 0.5 Hz, IMU at 100 Hz, magnetometer at 50 Hz) and are specified in Section 3.3.

Sensor	**Parameter**	**Value**	**Notes**
WiFi (IEEE 802.11ac)			
	Operating bands	2.4/5 GHz	Dual-band support
	RSSI range	-100 to -30 dBm	Dynamic range
	RSSI resolution	1 dBm	Quantization step
	Scan interval	1–4 s	Environment dependent
	Max visible APs	100+	Per scan
Inertial Measurement Unit			
	Accelerometer range	± 8 g	Full scale
	Accelerometer resolution	0.244 mg/LSB	16-bit ADC
	Gyroscope range	± 2000 °/s	Full scale
	Gyroscope resolution	0.061 °/s/LSB	16-bit ADC
	Sample rate	1000 Hz	Maximum
	Noise density	150 $\mu g/\sqrt{\text{Hz}}$	Accelerometer
Magnetometer			
	Measurement range	± 4900 µT	Full scale
	Resolution	0.15 µT/LSB	16-bit ADC
	Sample rate	100 Hz	Maximum
	Operating temperature	-40 to $+85$ °C	Specified range
Spatial Tracking			
	Tracking accuracy	< 2 cm	Typical indoor
	Update rate	60 Hz	Pose estimates
	Environment mapping	Real-time mesh	Depth sensor fusion

Note: ADC: Analog-to-Digital Converter; LSB: Least Significant Bit. Actual performance may vary based on environmental conditions.

3.2.2 Spatial Anchor System

The cornerstone of the methodology is the establishment of highly accurate ground truth reference points using AR technology. The HoloLens 2's sophisticated sensor suite provides the spatial mapping foundation, while a layered software system enables anchor creation, alignment with architectural plans, and persistent storage across collection sessions.

Floor Plan Alignment: The framework aligns computer-aided design (CAD) floor plans with the physical environment using the HoloLens 2 spatial mapping pipeline [120], which performs real-time 3D reconstruction of the indoor space through the device's depth sensors. The alignment process follows a semi-automatic procedure that identifies key architectural features as correspondence points. The operator dons the HoloLens 2 and initiates a building scan, allowing the device to construct an initial spatial mesh. A custom holographic application then displays the CAD floor plan as an overlay, initially positioned at an approximate location.

The operator identifies at least four non-collinear correspondence points between the virtual floor plan and physical environment, typically room corners, doorway edges, and pillar locations. For each correspondence, the operator physically walks to the feature and places a virtual marker at the exact location. The alignment application computes an optimal rigid transformation (rotation and translation) that minimizes the sum of squared distances between virtual and physical correspondence points. The alignment process typically achieves registration accuracy of 2–5 centimeters, verified by visual inspection of the aligned floor plan against physical features such as walls and doorways.

Virtual Anchor Creation: Following floor plan alignment, virtual anchors are systematically deployed using a standardized protocol. Anchors are placed at 1-meter grid intervals throughout accessible areas, ensuring consistent reference point density across all buildings. Additional anchors are positioned at points of interest (POIs) that may not fall on the regular grid, including:

- Building entrances and exits
- Hallway intersections and T-junctions
- Stairwell landings and elevator lobbies
- Transition zones between distinct architectural areas

Each anchor encodes comprehensive spatial information:

$$a_i = \{\mathbf{p}_i, \mathbf{q}_i, t_i, \sigma_i\} \tag{3.2}$$

where $\mathbf{p}_i \in \mathbb{R}^3$ represents the 3D position in the building coordinate frame, $\mathbf{q}_i \in \mathbb{H}$, $\|\mathbf{q}_i\| = 1$ denotes the orientation quaternion specifying the anchor's facing direction, t_i is the creation timestamp, and σ_i represents the uncertainty estimate derived from the HoloLens 2's tracking confidence at the moment of placement.

The orientation quaternion is particularly important for the data collection protocol, as it specifies the direction the smartphone operator should face when recording sensor data at each anchor location, ensuring consistency across collection sessions.

Azure Spatial Anchors Persistence: The virtual anchors are stored in Microsoft Azure Spatial Anchors [121], a cloud service that enables cross-session persistence and multi-user access to spatial reference points. When an anchor is created, the HoloLens 2 captures a visual fingerprint of the surrounding environment, specifically distinctive visual features extracted from the tracking camera images within a 3-meter radius of the anchor location. This visual fingerprint is uploaded to the Azure cloud along with the anchor's spatial metadata.

During subsequent collection sessions, the HoloLens 2 operator can re-localize to previously created anchors by scanning the environment until the device recognizes sufficient visual features to match against stored fingerprints. Azure Spatial Anchors reports that anchor re-localization achieves position accuracy of approximately 2 cm under typical indoor conditions [121]. Re-localization checks at the start of the 2022 campaign measured a mean deviation of 1.8 cm between detected and stored anchor positions (Section 3.5), supporting the long-term stability of the anchor coordinate system.

The cloud-based persistence provides several operational advantages. Multiple HoloLens devices can access the same anchor set, enabling parallel data collection by multiple operators. Anchors persist indefinitely without requiring the HoloLens to maintain a continuous connection, allowing collection campaigns to span weeks or months. The spatial coordinate frame remains consistent across sessions, ensuring that fingerprints collected in 2021 and 2022 reference identical physical locations despite the year-long interval between campaigns.

3.2.3 Spatial Anchor Validation

The accuracy claims for the AR-based ground truth system require independent validation against established surveying methods. While Microsoft's specifications and prior research report sub-centimeter tracking accuracy for HoloLens 2 under ideal conditions [15, 119], real-world indoor environments present challenges including variable lighting, reflective surfaces, and areas with limited visual features. This subsection documents the validation methodology employed to verify spatial anchor accuracy across the four-building testbed.

Validation Equipment and Setup: Independent ground truth positions were established using a Leica TS15 total station, providing sub-centimeter positional accuracy as the reference standard. Eighty validation points were selected across the testbed, with 20 points distributed throughout each building. The validation points were strategically positioned to sample diverse environmental conditions: well-lit corridors with abundant visual features, dimly lit interior spaces, areas near glass facades that challenge depth sensing, and zones with repetitive architectural elements that could induce visual aliasing.

The total station survey was conducted by certified surveyors following standard geodetic procedures. Each validation point was marked with a permanent floor sticker providing an unambiguous reference for both total station measurements and HoloLens anchor placement. The survey established a unified coordinate system across the four-building testbed, enabling direct comparison between surveyed positions and HoloLens-derived anchor coordinates.

Validation Protocol: The validation experiments were conducted independently of the fingerprint data collection sessions to prevent any bias from operator familiarity with anchor locations. For each validation point, a HoloLens operator approached the location from a randomly selected direction, allowing the device to establish spatial tracking through its normal initialization procedure. Upon arriving at the marked position, the operator created a new spatial anchor at the surveyed location.

To assess repeatability, each validation point was visited on three separate occasions under varying conditions: morning sessions with natural daylight, afternoon sessions with mixed lighting, and evening sessions under artificial illumination. The anchor creation procedure was performed independently at each visit, yielding 240 total anchor placements (80 points $\times$ 3 sessions) for analysis. This protocol isolates the anchor positioning accuracy from any potential environmental factors that might systematically affect specific times of day.

Accuracy Results: The comparison between HoloLens anchor positions and surveyed ground truth yielded the following accuracy metrics:

- Mean positioning error: 1.8 cm across all 240 validation measurements
- Standard deviation: 0.9 cm
- 95th percentile error: 3.7 cm
- Maximum observed error: 6.2 cm

No statistically significant correlation was observed between positioning error and time of day ($p > 0.3$ for all pairwise comparisons). Per-building analysis revealed modest variations: PSE achieved the best accuracy (mean 1.5 cm), benefiting from its symmetric layout and abundant visual features in corridor walls. BCE exhibited slightly higher errors (mean 2.1 cm) due to the glass facades that reduce depth sensor effectiveness near building perimeters. STC and LRC fell within the intermediate range (mean 1.7 cm and 1.9 cm respectively).

The repeatability analysis across the three sessions at each validation point demonstrated excellent consistency. The mean deviation between repeated anchor placements at identical physical locations was 1.2 cm, confirming that the HoloLens spatial mapping system produces stable, reproducible anchor positions across sessions. This repeatability is essential for the longitudinal data collection protocol, ensuring that fingerprints collected in 2021 and 2022 reference truly identical physical positions.

Implications for Ground Truth Quality: The validated ground-truth accuracy—mean 1.8 cm (standard deviation 0.9 cm), 95th-percentile 3.7 cm, and maximum 6.2 cm—represents a substantial improvement over the ground truth used in traditional fingerprinting methods. Manual approaches, where operators visually estimate positions on floor plans, typically achieve 1–3 meter accuracy at best [11]. Even careful manual surveying with measuring tapes or laser rangefinders rarely achieves better than 10–20 centimeter precision due to human alignment error and reference point ambiguity.

The AR-based methodology yields ground truth whose worst-case error is on the order of one-tenth of the sub-meter localization targets, while the typical error is considerably smaller: with reference positions accurate to a 95th-percentile of 3.7 cm and a maximum of 6.2 cm, against a mean of 1.8 cm, the inherent ground truth noise contributes negligibly to the overall localization error budget. This precision enables meaningful evaluation of sub-meter localization algorithms whose errors would otherwise be obscured by coarser ground truth. Furthermore, the high repeatability ensures that temporal drift in localization accuracy can be attributed to environmental changes rather than ground truth inconsistencies.

The validation results also confirm the suitability of the HoloLens 2 platform for indoor positioning research more broadly. The combination of sub-2-centimeter mean accuracy, robust performance across varied lighting conditions, and excellent repeatability across sessions establishes the AR-based methodology as a viable alternative to expensive optical tracking systems or labor-intensive manual surveying for ground truth establishment in indoor localization studies.

3.2.4 Data Collection Pipeline

The data collection pipeline coordinates the HoloLens 2 ground truth system with smartphone-based multi-modal sensor acquisition. The pipeline employs a two-operator protocol designed to ensure consistency, minimize human error, and capture synchronized sensor data at precisely defined spatial locations.

Two-Operator Protocol: Each collection session involves two trained operators working in coordination. The *HoloLens operator* wears the HoloLens 2 and navigates to anchor locations, providing real-time visual guidance to the second operator. The *Android device operator* carries a smartphone running the GetSensorData application [122], which provides synchronized access to the device's full sensor suite. This division of responsibilities serves multiple purposes: it reduces cognitive load on each operator, enables quality control through mutual observation, and ensures that the smartphone's WiFi antenna is not shielded by the HoloLens operator's body during signal measurements.

Session Initialization: Before beginning data collection, the HoloLens operator launches the custom application (built with Microsoft's Mixed Reality Toolkit, MRTK) and performs re-localization against the building's stored anchor set. The operator walks through the environment while the HoloLens 2 scans for visual features matching the Azure Spatial Anchors fingerprints. Once at least three anchors are successfully re-localized, the system confirms alignment with the building coordinate frame, and collection may

proceed. This re-localization step typically requires 3–5 minutes for a building floor and verifies that the spatial reference frame remains consistent with the validated anchor accuracy reported above.

The Android device operator initializes the GetSensorData application, configuring sensor sampling rates and enabling timestamp synchronization via Network Time Protocol (NTP). Both operators verify that their devices display consistent timestamps to within 50 ms, ensuring proper temporal alignment of ground truth positions with sensor measurements.

Per-Anchor Collection Procedure: For each anchor in the building, the operators follow a standardized sequence:

1. The HoloLens operator navigates to the anchor location, visually confirming alignment with the virtual marker displayed in the headset.
2. The HoloLens operator verbally announces the anchor identifier and signals the Android operator to approach.
3. The Android operator positions the smartphone at the anchor location, orienting the device according to the anchor's stored quaternion.
4. The Android operator initiates a 30 s stationary collection period while maintaining position and orientation.
5. During collection, the HoloLens operator monitors for any position drift and can signal for repeat collection if movement is detected.
6. Upon completion, the GetSensorData application confirms successful data logging, and both operators proceed to the next anchor.

The 30 s stationary period provides sufficient time to capture multiple WiFi scan cycles (typically 10–15 complete scans at 0.5 Hz scan rate) while accumulating magnetometer and IMU readings at their respective higher sampling rates. The stationary requirement ensures that all sensor readings at a given anchor correspond to an identical physical location, avoiding motion-induced variations in the fingerprint signatures.

Path Planning: The order in which anchors are visited follows pre-computed paths that minimize total walking distance while ensuring complete coverage. A nearest-neighbor heuristic starting from the building entrance is employed, avoiding backtracking through previously collected areas where WiFi signals may have been perturbed by the operators' presence. The path typically follows a serpentine pattern through corridors, with branches into individual rooms and offices as accessible.

Figure 3.2 provides an overview of the complete data collection system, illustrating the flow from sensor acquisition through the multi-layer processing pipeline.

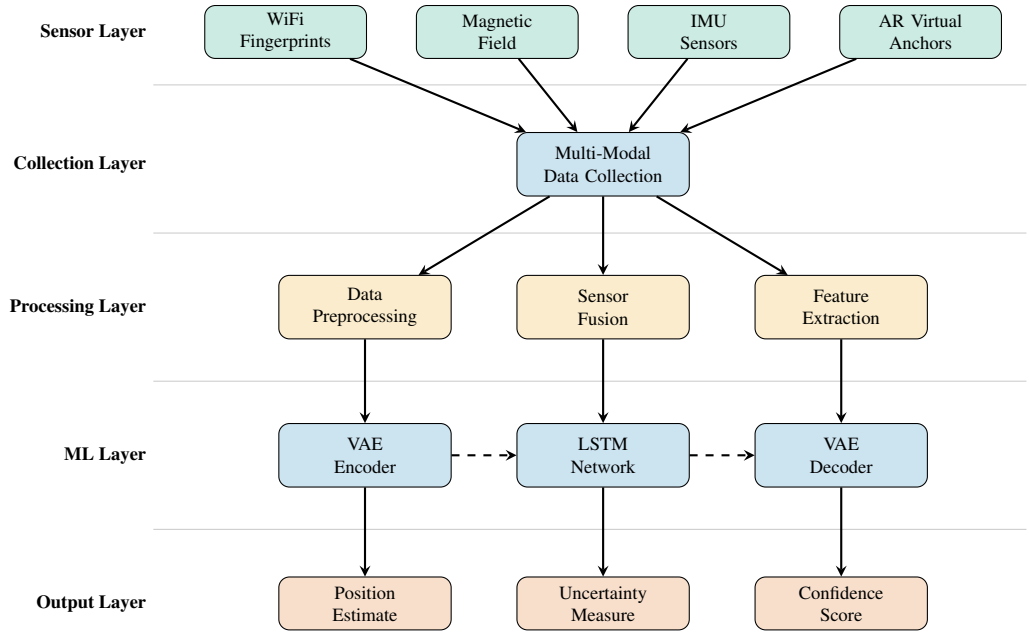


Figure 3.2: System architecture of the AR-enhanced indoor localization framework. The system integrates multiple sensor modalities through a layered approach, combining WiFi fingerprinting (48% contribution), magnetic field measurements (41% contribution), IMU data (11% contribution), and AR virtual anchors into a unified VAE-LSTM architecture for robust position estimation with uncertainty quantification.

The system integrates the AR-established ground truth coordinates with multi-modal sensor data from the smartphone, feeding into the preprocessing and feature extraction pipeline detailed in Section 3.6. The architecture supports both batch processing for model training and real-time inference for positioning applications, though this chapter focuses on the data collection and baseline evaluation aspects.

3.3 Multi-Modal Sensor Acquisition

The effectiveness of indoor localization systems depends fundamentally on the discriminative power of the sensor modalities employed. While any single sensor provides incomplete information about position (WiFi signals suffer from multipath and temporal drift, magnetometers require calibration and depend on ferromagnetic structures, and inertial measurements accumulate unbounded error), their combination yields complementary information that enhances both accuracy and robustness [115, 116]. This section details the three sensor modalities that form the foundation of the multi-modal fingerprinting approach: WiFi received signal strength, inertial measurements from accelerometers and gyroscopes, and magnetometer readings of the ambient magnetic field.

The sensor selection was guided by three criteria established in prior work [117]: *ubiquity* (the sensor must be present in commodity smartphones without requiring additional hardware), *discriminative power* (the sensor's readings must vary meaningfully across different spatial locations), and *temporal stability* (the sensor's location-dependent signature must persist over operationally relevant time periods). While modern smartphones include additional sensors such as barometers for floor detection and ambient light

sensors for contextual awareness, this thesis focuses on the three modalities that demonstrate the strongest combination of these properties for horizontal positioning within building floors.

3.3.1 WiFi Signal Collection

WiFi received signal strength indicator (RSSI) measurements form the primary modality in the localization system, leveraging the dense deployment of IEEE 802.11 access points (APs) in modern buildings. The fundamental principle underlying WiFi fingerprinting is that the signal strength received from each AP varies predictably with receiver position, creating a unique "fingerprint" at each location that can be matched against a pre-surveyed database [49, 19].

Signal Acquisition Protocol: The GetSensorData application [122] captures WiFi scan results at a frequency of 0.5 Hz, the maximum rate permitted by Android's WiFi scanning API for power efficiency. Each scan reports the RSSI (in dBm) and MAC address for all APs visible to the device, spanning both the 2.4 GHz and 5 GHz bands of the IEEE 802.11 standard. The number of visible APs varies substantially across the testbed, from as few as 8 in shielded laboratory spaces to over 40 in open corridors near building cores, reflecting the heterogeneous AP deployment density across different areas.

During the 30 s stationary collection period at each anchor location, the system captures 10–15 complete WiFi scans. These multiple measurements are essential for characterizing the statistical distribution of RSSI values, which exhibit significant short-term fluctuations due to environmental factors including human movement, door openings, and electronic interference. Both the mean and standard deviation of RSSI from each visible AP are computed, providing robust features that account for this inherent variability.

Signal Processing: Raw RSSI measurements exhibit substantial temporal fluctuations that can mask the underlying location-dependent signal. To suppress this variability, an exponential moving average (EMA) filter is applied to each access point's RSSI time series:

$$\hat{r}_{k+1} = \alpha \hat{r}_k + (1-\alpha) r_{k+1} \tag{3.3}$$

where $\hat{r}_k$ denotes the filtered RSSI estimate at time step k, r_k represents the raw measurement, and $\alpha = 0.7$ is the smoothing factor that weights the previous estimate against the incoming measurement, set through empirical tuning. This first-order exponential smoother is the steady-state form of a scalar Kalman filter for the quasi-stationary RSSI signal at a fixed location, achieving the same noise suppression without per-access-point noise modeling or covariance propagation. The filter reduces the variance of RSSI measurements by approximately 40% while preserving the location-dependent mean values essential for fingerprint discrimination.

Feature Vector Construction: The variable number of visible APs across locations presents a challenge for machine learning algorithms that require fixed-dimension inputs. This challenge is addressed through zero-

padding: the RSSI values from detected APs are arranged into a vector and padded with zeros (representing "not detected") to a fixed length of 100 dimensions. The AP ordering within the vector is determined by a building-specific AP registry constructed during initial data collection, ensuring consistent feature alignment across all fingerprints within a building.

Why WiFi as Primary Modality: The ubiquity of WiFi infrastructure makes it the natural foundation for indoor positioning systems. Unlike dedicated localization infrastructure such as ultrawideband beacons or Bluetooth Low Energy tags, WiFi APs are deployed primarily for communication purposes and exist in virtually every modern building. This infrastructure-free approach reduces deployment costs to essentially zero, as the localization system leverages existing network investments. Furthermore, WiFi signals penetrate walls and propagate throughout buildings, providing coverage even in areas without direct line-of-sight to APs.

Within WiFi, RSSI was chosen over channel state information (CSI). CSI extraction requires network interface cards with modified firmware or custom drivers, and neither Android nor iOS exposes CSI through its public WiFi interfaces. RSSI is available through the standard Android scanning API on unmodified handsets (Section 2.3.2). Cellular positioning based on 5G was not used for a related reason. It depends on operator-controlled infrastructure and on measurements that standard handset interfaces do not expose to applications (Section 2.2.3). Camera-based positioning was excluded because continuous image capture in occupied buildings raises privacy concerns (Section 3.5.3). The HoloLens 2 RGB camera was also not used (Section 3.2.1).

However, WiFi fingerprinting faces well-documented challenges. Signal propagation follows complex patterns due to reflection, diffraction, and absorption by building materials, creating multipath effects that cause RSSI to vary non-monotonically with distance. The WiFi infrastructure itself evolves over time as APs are added, removed, or reconfigured, causing fingerprint databases to become stale. Human presence affects signal propagation through body absorption and reflection, introducing occupancy-dependent variations. These limitations motivate the inclusion of complementary sensor modalities that exhibit different failure modes.

3.3.2 WiFi Access Point Density Characterization

The effectiveness of WiFi fingerprinting depends critically on the density and spatial distribution of access points throughout the environment. Sparse AP deployments produce fingerprints with limited discriminative power, while dense deployments can overwhelm localization algorithms with redundant information. Understanding the AP density characteristics of the testbed provides essential context for interpreting localization performance and informs the design of robust fingerprinting algorithms.

Spatial Distribution Analysis: The four-building testbed exhibits substantial variation in AP density, reflecting the diverse communication needs and infrastructure investment across different building types. AP density ranges from approximately 0.064 APs/m^2 in the compact STC glass walkway, the most densely

instrumented environment, down to roughly 0.0049 APs/m^2 in the expansive BCE building, with LRC's research facility at approximately 0.021 APs/m^2. Notably, PSE contains the most access points in absolute terms (89), yet its large 12,544 m^2 floor area yields a comparatively low density of approximately 0.0071 APs/m^2; the resulting range spans roughly a 13-fold spread across the testbed. At individual reference point locations, the number of visible APs ranges from 8 in shielded laboratory spaces to over 40 in open corridor areas near building cores and elevator lobbies. This variation creates a natural gradient of fingerprint complexity: locations with many visible APs produce high-dimensional signatures with substantial discriminative power, while locations in coverage gaps depend on fewer, potentially noisier signal sources.

The spatial distribution of AP visibility follows predictable patterns shaped by building architecture. Floor-to-ceiling glass facades in BCE allow signals to propagate further than the reinforced concrete walls in PSE, resulting in broader but weaker coverage zones. Corridor intersections and atriums typically exhibit the highest AP counts due to line-of-sight paths to multiple access points, while interior offices and enclosed spaces show reduced visibility. These architectural effects create systematic variation in fingerprint quality that localization algorithms must accommodate.

Signal Strength Patterns: RSSI values across the testbed span a range from approximately $-85\,\text{dBm}$ (weak, near detection threshold) to $-35\,\text{dBm}$ (strong, close proximity to AP). The distribution of observed RSSI values exhibits a characteristic bimodal pattern: strong signals ($-55\,\text{dBm}$ to $-35\,\text{dBm}$) from nearby APs provide reliable positioning anchors, while weak signals ($-85\,\text{dBm}$ to $-70\,\text{dBm}$) from distant APs contribute noisy but still informative constraints.

Zones of strong WiFi coverage concentrate around designated network infrastructure locations, typically near elevator lobbies, major corridor intersections, and common areas. These high-RSSI zones achieve sub-meter localization accuracy due to the steep signal gradients that create distinctive fingerprints over short distances. Conversely, coverage gaps in peripheral offices and shielded laboratories exhibit weaker signals with shallower gradients, resulting in degraded localization accuracy. The algorithms evaluated in Section 3.8 demonstrate varying sensitivity to these coverage differences, with probabilistic approaches showing greater robustness to weak-signal conditions.

Temporal Stability of AP Infrastructure: The longitudinal nature of the dataset spanning 2021–2022 reveals significant WiFi infrastructure evolution that directly impacts fingerprint-based localization. Over the one-year interval, the total access point count grew by a net 47 (a 24.5% increase, from 192 to 239), reflecting roughly 59 newly installed access points partially offset by 12 that were decommissioned. These changes fundamentally altered the RF landscape in affected areas, with net per-building gains of 18 in PSE, 15 in BCE, 8 in STC, and 6 in LRC. Firmware updates on numerous APs additionally caused subtle changes in transmission power and channel allocation.

Despite these infrastructure changes, approximately 78% of access points remained stable across both collection epochs, providing a consistent foundation for fingerprint matching. The visualization in Figure 3.3 documents the spatial distribution of AP additions and removals, revealing that most changes

occurred in areas undergoing renovation or expansion. This temporal characterization informs the development of adaptation mechanisms in Chapter 5 that enable localization systems to accommodate infrastructure evolution without requiring complete database reconstruction.

3.3.3 Inertial Measurement Unit

The inertial measurement unit (IMU) provides motion-related measurements that complement the position-dependent WiFi fingerprints. While IMU data alone cannot determine absolute position due to integration drift, it provides valuable constraints for trajectory estimation and enables detection of motion states that inform the localization algorithm's behavior [123, 124].

Sensor Configuration: Modern smartphones incorporate 6-degree-of-freedom (6-DOF) IMU sensors comprising a 3-axis accelerometer and 3-axis gyroscope. The GetSensorData application samples both sensors at 100 Hz, sufficient to capture the dynamics of pedestrian motion including individual footsteps. The accelerometer measures specific force (the sum of gravitational and inertial accelerations) in units of m/s^2, while the gyroscope measures angular velocity in rad/s around each axis.

Motion State Detection: IMU data enables classification of the user's motion state (stationary, walking, running, or ascending/descending stairs), which informs how other sensor modalities should be weighted. During stationary periods, WiFi and magnetometer fingerprints can be directly matched against the reference database. During motion, the IMU provides dead reckoning constraints that limit the plausible position changes between consecutive WiFi scans. Stationary periods are detected through a variance threshold on the accelerometer magnitude:

$$\text{stationary} = \left(\text{Var}_{t-w\leq\tau\leq t}\left[\,\|\mathbf{a}_\tau\|\,\right] < \tau_a\right) \tag{3.4}$$

where $\mathbf{a}$ denotes the accelerometer reading, w is a sliding window of 50 samples (0.5 s), and $\tau_a = 0.1$ m^2/s^4 is the variance threshold determined through calibration.

Dead Reckoning Constraints: For users in motion, the IMU enables pedestrian dead reckoning (PDR) that estimates displacement from step detection and stride length estimation. While detailed PDR implementation falls outside the scope of this dataset chapter, the raw IMU measurements preserved in the dataset support such algorithms. The key insight is that even imperfect dead reckoning provides valuable short-term constraints: between consecutive WiFi scans (2 seconds apart), the user can move at most a bounded distance determined by maximum walking speed, and the trajectory must be physically plausible given the IMU-derived heading changes.

Limitations and Role in Multi-Modal Fusion: The fundamental limitation of inertial navigation is cumulative drift. Accelerometer measurements must be double-integrated to obtain position, causing small

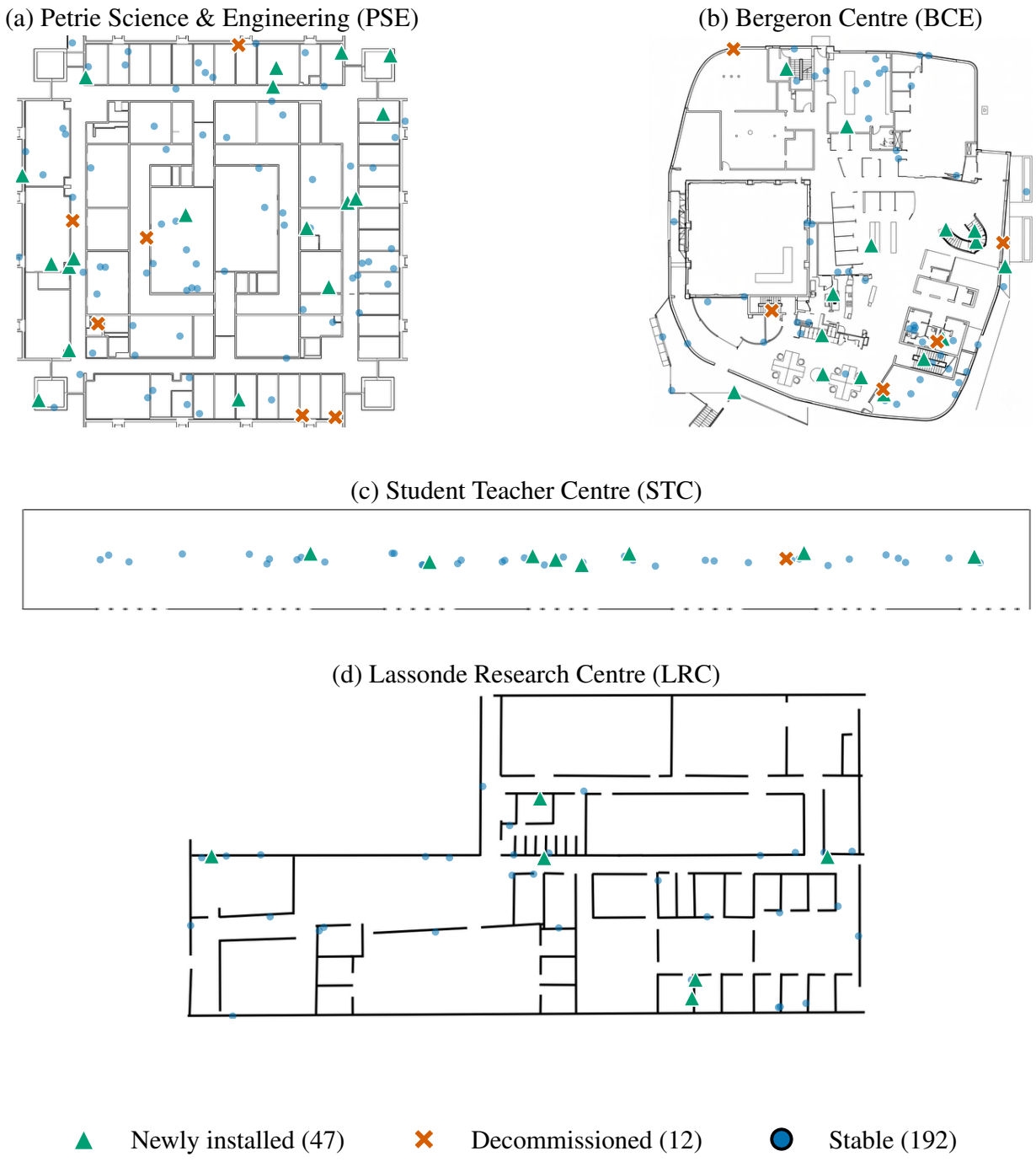


Figure 3.3: WiFi infrastructure changes between 2021 and 2022 collection epochs across the four-building testbed. Green markers indicate newly installed access points, red markers show decommissioned APs, and blue markers represent stable infrastructure. The majority of changes (74%) occurred in PSE and BCE, which underwent partial renovations during the academic year.

measurement biases to grow quadratically with time. Gyroscope biases similarly cause heading errors that accumulate. In practice, IMU-only positioning errors can exceed several meters within a minute of walking.

Despite this limitation, IMU data remains valuable in the multi-modal framework for two reasons. First, the short-term accuracy is high: over the 2-second interval between WiFi scans, IMU-derived displacement estimates typically have errors below 0.5 m. Second, the IMU provides motion context that other sensors lack. A sudden acceleration pattern may indicate the user entering or exiting a stairwell, transitioning between buildings, or changing floors; such information helps disambiguate WiFi fingerprints that might otherwise appear similar across different building areas.

3.3.4 Magnetometer Data

The magnetometer measures the ambient magnetic field vector, comprising both Earth's geomagnetic field and local perturbations induced by ferromagnetic building materials. These perturbations create location-specific "magnetic fingerprints" that remain highly stable over time, providing a valuable complement to WiFi signals that suffer from infrastructure changes [125, 126].

Sensor Characteristics: The 3-axis magnetometer samples the magnetic field vector at 50 Hz, measuring field strength in microteslas (μT) along each device axis. Earth's magnetic field in Toronto has magnitude approximately 45–50 μT, with inclination (dip angle) of roughly 70° below horizontal pointing north. This baseline field is perturbed by ferromagnetic materials (primarily structural steel, reinforcing bars in concrete, and metallic building fixtures), creating spatial variations that can be exploited for positioning.

Calibration Requirements: Raw magnetometer readings are corrupted by two categories of error that must be corrected before use in localization. *Hard iron* distortion arises from permanent magnets near the sensor, such as smartphone speakers and battery components, which add a constant offset to all measurements:

$$\mathbf{m}_{corrected} = \mathbf{m}_{raw} - \mathbf{b}_{hard} \tag{3.5}$$

where $\mathbf{m}_{raw}$ is the raw 3-axis measurement, $\mathbf{b}_{hard} \in \mathbb{R}^3$ is the hard iron bias vector, and $\mathbf{m}_{corrected}$ is the corrected measurement. The bias vector is estimated by rotating the device through multiple orientations and finding the center of the resulting measurement sphere.

Soft iron distortion arises from ferromagnetic materials that concentrate or redirect the ambient field, effectively scaling and rotating the measurements along different axes:

$$\mathbf{m}_{calibrated} = \mathbf{W}^{-1}\mathbf{m}_{corrected} \tag{3.6}$$

where $\mathbf{W}$ is the 3×3 soft-iron distortion matrix that maps the true spherical field onto the observed ellipsoidal locus, so its inverse $\mathbf{W}^{-1}$ restores the measurements to their expected spherical form. Both calibration parameters are estimated using a figure-eight waving motion that samples the magnetic field from diverse orientations, following the standard smartphone magnetometer calibration procedure.

Magnetic Anomaly Detection: Structural elements create distinctive magnetic signatures that serve as natural landmarks for positioning. Magnetic anomalies are detected by comparing the measured field magnitude against the expected Earth field:

$$\psi(\mathbf{p}) = \|\mathbf{m}_{calibrated}(\mathbf{p}) - \mathbf{m}_{earth}\| \tag{3.7}$$

where $\mathbf{m}_{earth}$ denotes the expected geomagnetic field vector in Toronto (magnitude approximately 48 μT). Locations where $\psi(\mathbf{p}) > \tau_m$ exhibit significant perturbations from building materials. In the testbed, the threshold $\tau_m = 10\,\mu$T identifies areas near steel staircases, elevator shafts, reinforced concrete columns, and electrical distribution panels, where deviations can exceed 20–30 μT.

Temporal Stability Advantages: A key advantage of magnetic fingerprinting is its temporal stability. While WiFi fingerprints degrade as APs are added, removed, or reconfigured, the magnetic field distribution depends primarily on permanent structural elements that change only during major renovations. The longitudinal data collection confirmed this stability: magnetic fingerprints collected in 2021 and 2022 exhibited correlation coefficients exceeding 0.95 at matched locations, compared to WiFi fingerprint correlations of 0.70–0.85 over the same period.

The strongest magnetic signatures appear near specific architectural features. Corner staircases in PSE, constructed with reinforced concrete and steel railings, exhibit perturbations exceeding 20 μT from the ambient field. Elevator shafts create distinctive vertical gradients as the steel cab and counterweights move. Electrical distribution panels generate localized field concentrations. These features provide reliable landmarks that persist across years, making magnetic fingerprinting particularly valuable for long-term system operation without frequent database updates.

3.3.5 Magnetic Field Characterization

Beyond the sensor-level calibration and anomaly detection procedures described above, a comprehensive characterization of the magnetic field environment across the testbed provides essential context for understanding the role of magnetometer data in multi-modal localization. The magnetic landscape varies substantially across buildings, reflecting differences in construction materials, equipment installations, and architectural design choices.

Spatial Variation Patterns: The magnetic field magnitude across the four-building testbed ranges from approximately 25 µT in areas with strong ferromagnetic shielding to 65 µT near concentrated metallic infrastructure, compared to the nominal Earth field of 48 µT in the Toronto region. This 40 µT variation range provides substantial discriminative power for positioning, as locations separated by only a few meters can exhibit measurably different magnetic signatures.

The primary sources of magnetic variation include structural steel beams and columns, reinforcing bars embedded in concrete floors and walls, electrical distribution systems carrying alternating currents, and

installed equipment such as HVAC units and laboratory instruments. These sources create characteristic spatial patterns: corridors bounded by steel-reinforced walls exhibit elevated field magnitudes along their length, while open atriums with minimal ferromagnetic content show field values closer to the geomagnetic baseline. The transition zones between these regions produce steep gradients that serve as natural position landmarks.

Building-Specific Magnetic Signatures: Each building in the testbed presents a distinctive magnetic environment shaped by its construction and use:

Petrie Science and Engineering Building (PSE). The dense reinforced concrete construction creates the most pronounced magnetic anomalies in the testbed. The symmetric building layout produces a characteristic pattern of elevated field strength near the corner staircases, where structural steel concentrations approach maximum density. Laboratory spaces housing older scientific equipment generate localized perturbations exceeding 25 μT deviation from ambient, providing distinctive landmarks despite the building's otherwise symmetric geometry. The magnetic environment partially compensates for the "perceptual aliasing" that would otherwise confound localization in the visually repetitive floor plan.

Bergeron Centre for Engineering Excellence (BCE). The modern steel-and-glass construction presents a different magnetic character: consistent background fields throughout much of the floor plate, punctuated by strong anomalies near the structural steel core and elevator shafts. The glass curtain walls contribute minimally to magnetic variation, resulting in lower overall field strength (42 μT average) compared to the reinforced concrete buildings. The magnetic simplicity of BCE's open areas means that WiFi fingerprinting carries proportionally greater weight for localization in this environment.

Student Teacher Centre (STC). The linear building geometry and floor-to-ceiling windows create an elongated magnetic profile with minimal transverse variation. Field strength ranges narrowly between 44 μT and 52 μT along the corridor axis, limiting the discriminative power of magnetometer data for longitudinal positioning. However, the magnetic signatures near building entrances and the central stairwell provide useful anchoring points that help constrain WiFi-based estimates to physically plausible positions.

Lassonde Research Centre (LRC). The dichotomy between windowless laboratories and perimeter offices produces the most heterogeneous magnetic environment in the testbed. Laboratory spaces housing research equipment generate field perturbations exceeding 50 μT, far stronger than structural anomalies elsewhere. The MRI equipment installed between the 2021 and 2022 collection campaigns created particularly pronounced signatures that dominate the local magnetic landscape. In contrast, the perimeter offices exhibit benign magnetic conditions similar to conventional office construction.

Magnetic Field Stability Over Time: A key advantage of magnetic fingerprinting over WiFi-based approaches is the temporal stability of magnetic signatures. Comparison of the 2021 and 2022 datasets reveals that magnetic field measurements at matched reference point locations exhibit correlation coefficients exceeding 0.95, compared to WiFi correlations of 0.70–0.85 over the same interval. This stability reflects the permanence of ferromagnetic building structures, which change only during major renovations.

The primary source of magnetic variation between collection epochs was the installation of new laboratory equipment in LRC, which created localized signature changes in affected areas. Outside these equipment zones, magnetic signatures remained essentially invariant, providing a stable foundation for long-term localization. This temporal robustness makes magnetic fingerprinting particularly valuable for systems that must operate reliably over extended deployment periods without frequent database updates, a consideration examined further in the online adaptation mechanisms of Chapter 5.

3.4 Four-Building Testbed

The evaluation of indoor localization systems requires test environments that capture the diversity of challenges encountered in real-world deployments. Single-building evaluations, while methodologically convenient, risk producing systems optimized for specific architectural configurations that fail to generalize across the heterogeneous building stock found in practice. To address this limitation, a comprehensive testbed spanning four architecturally distinct buildings at York University in Toronto, Canada was established, totaling 28 822 m^2 of indoor space.

The building selection was guided by three principles: *architectural diversity* (the testbed should include both traditional and contemporary construction styles), *environmental heterogeneity* (buildings should present distinct challenges for different sensor modalities), and *practical accessibility* (spaces should be accessible for repeated data collection campaigns over multiple years). The resulting testbed encompasses symmetric and linear floor plans, reinforced concrete and steel-glass construction, and environments ranging from equipment-dense laboratories to open student gathering spaces.

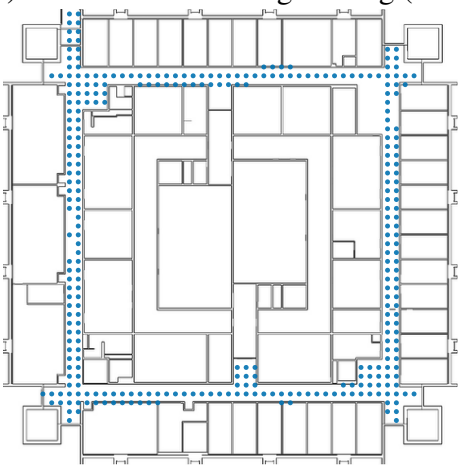
Table 3.2 summarizes the key characteristics of each building, while Figure 3.4 presents floor plans showing the spatial layout and data collection coverage for each environment.

Table 3.2: Architectural characteristics of the four-building testbed at York University. The four buildings span steel-frame, reinforced-concrete, glass-curtain, and glass-walkway construction, ensuring that localization results generalize across representative indoor environments rather than reflecting a single building type. Areas are total surveyed floor area across the floors covered in each building (PSE spans four floors). Access-point counts refer to the 2022 campaign; 192 access points were observed in 2021 (Section 3.5).

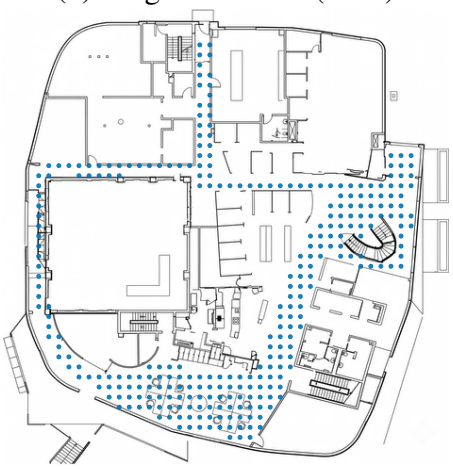
Building	**Dimensions**	**Area (m^2)**	**WiFi APs**	**Construction**	**Key Challenge**
PSE	56×56 m	12,544	89	Reinforced concrete	Magnetic anomalies
BCE	46×60 m	13,800	67	Steel/glass curtain	RF multipath
STC	78×9 m	702	45	Glass walkway	Occupancy variation
LRC	74×24 m	1,776	38	Research facility	EM shielding
Total	—	**28,822**	**239**	—	—

The spatial distribution of reference points across the four buildings reflects a systematic approach to coverage that balances comprehensiveness with practical data collection constraints. Reference points are positioned at approximately 1 m intervals along major corridors and common areas, ensuring that the fingerprint database captures the spatial granularity necessary for sub-meter localization while maintaining

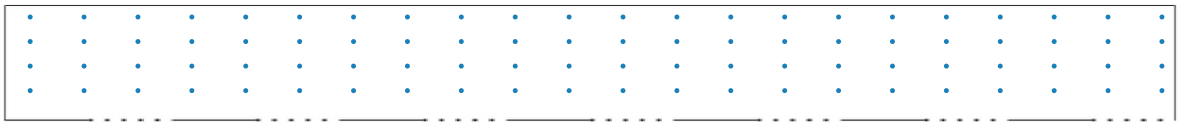
(a) Petrie Science & Engineering (PSE)



(b) Bergeron Centre (BCE)



(c) Student Teacher Centre (STC)



(d) Lassonde Research Centre (LRC)

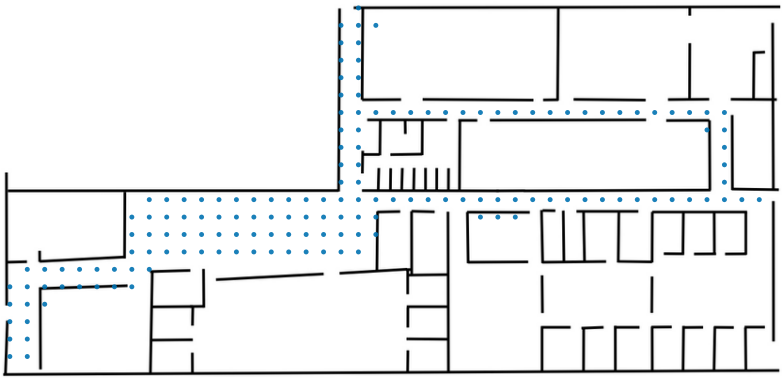


Figure 3.4: Floor plans of the four-building testbed showing AR spatial anchor reference points established via HoloLens 2 ground truth system. Reference points are positioned at approximately 1 m grid intervals throughout accessible corridors and common areas, with additional points at architectural landmarks. Total reference points per building: PSE (312), BCE (289), STC (87), LRC (159). (a) Petrie Science and Engineering Building (PSE) with symmetric square layout and corner staircases; (b) Bergeron Centre for Engineering Excellence (BCE) featuring curved glass facades and varying corridor widths; (c) Student Teacher Centre (STC) with linear sun-lit corridors; (d) Lassonde Research Centre (LRC) showing research and computer laboratories.

a manageable collection burden. This dense spacing was chosen because WiFi signal strength varies measurably over distances of 2–4 meters in typical indoor environments; sampling at 1 m intervals therefore resolves these spatial gradients well below their characteristic scale, providing fine-grained coverage and redundancy against per-location signal noise.

The reference point placement strategy also accounts for building geometry and expected pedestrian flow patterns. High-traffic areas such as building entrances, corridor intersections, and stairwell landings receive additional reference points beyond the regular grid to improve localization accuracy at critical navigation decision points. Conversely, restricted areas such as equipment rooms and private offices may have sparser coverage where access was limited during data collection. The resulting distribution, visualized in Figure 3.4, ensures comprehensive coverage of publicly accessible spaces while respecting the practical constraints of operating in active academic environments.

Petrie Science and Engineering Building (PSE): The Petrie Science and Engineering Building exemplifies traditional academic architecture, featuring a perfectly symmetric 56×56 m square floor plan constructed primarily of reinforced concrete. The building is organized around square corridors that form the primary circulation paths, with research laboratories occupying the exterior perimeter and physics laboratories concentrated in the building's central core. This symmetry creates significant challenges for localization approaches that rely on spatial distinctiveness: the four quadrants of the building are virtually indistinguishable based on visual or geometric features alone, a phenomenon termed "perceptual aliasing" that confounds many positioning algorithms.

The building's construction with steel-reinforced concrete creates pronounced magnetic anomalies, particularly near the corner staircases where structural steel and elevator machinery concentrate. At these locations, magnetic field perturbations exceed 20 μT deviation from the ambient geomagnetic field, providing distinctive landmarks that compensate for the visual ambiguity. With 89 WiFi access points serving the floor, PSE offers dense RF coverage, though the concrete construction attenuates signals more rapidly than lighter materials, creating sharp boundaries between coverage zones.

Data collection in PSE spanned the corridors and common areas across all four floors, yielding 12,544 m^2 of mapped space. The regular grid layout facilitated systematic anchor placement at precise 1-meter intervals, producing the most uniformly sampled environment in the testbed. This uniformity makes PSE particularly valuable for algorithm development, where consistent sampling density reduces confounding factors in performance comparisons.

Bergeron Centre for Engineering Excellence (BCE): The Bergeron Centre for Engineering Excellence represents contemporary sustainable architecture, having achieved LEED Gold certification for its energy-efficient design. The building's 46×60 m rectangular footprint features extensive glass curtain walls and a steel structural framework that creates the most challenging RF propagation environment in the testbed. The interior is primarily occupied by laboratories and classrooms, connected by corridors of arbitrary and irregular shapes, with several cafés and eating areas interspersed throughout the building.

The glass facades and steel structure induce severe multipath propagation for WiFi signals. Path loss exponents measured in BCE range from 2.8 to 4.2, substantially exceeding the free-space value of 2.0 and even typical indoor values of 2.5–3.5 observed in conventional construction. The building's irregular corridor geometry, with widths ranging from 1.5 m in office wings to 8 m in central atrium spaces, creates spatial heterogeneity that challenges positioning algorithms trained on more uniform environments.

Despite having fewer WiFi access points (67) than PSE, BCE presents greater positioning difficulty due to signal variability. The open floor plan allows signals from distant APs to propagate unexpectedly, while the steel framework creates localized dead zones where coverage drops precipitously. This environment serves as a stress test for WiFi fingerprinting approaches, revealing failure modes that remain hidden in more benign RF environments. Data collection in BCE yielded 13,800 m^2 of mapped space, the largest in the testbed.

Student Teacher Centre (STC): The Student Teacher Centre functions primarily as a pedestrian walkway, featuring a distinctive linear 78×9 m floor plan that is open on one side, prioritizing natural lighting through floor-to-ceiling windows along the building's length. This elongated, open-sided geometry creates unique challenges for positioning: the limited width provides fewer orthogonal spatial variations in sensor signatures, effectively reducing localization to a one-dimensional problem along the corridor axis. The building experiences heavy daytime foot traffic as a major pedestrian thoroughfare connecting other campus buildings.

The building experiences extreme occupancy variations, with population density changing by factors of ten between early morning (fewer than 50 occupants) and peak lunch hours (exceeding 500 occupants). These occupancy changes directly impact RF signal propagation through body absorption and multipath from human movement. Ambient light levels similarly vary from 100 lux in shaded morning conditions to over 2000 lux when direct sunlight enters through the open side and windows, creating time-dependent environmental signatures that must be accommodated by robust positioning algorithms.

With only 702 m^2 of mapped area, STC is the smallest building in the testbed. However, its extreme occupancy dynamics and lighting variations make it invaluable for evaluating temporal robustness. Algorithms that perform well in STC's variable conditions demonstrate resilience to environmental changes that are less pronounced but still present in other buildings.

Lassonde Research Centre (LRC): The Lassonde Research Centre presents a largely closed-off environment within its 74×24 m rectangular footprint. The building houses research and computer laboratories that dominate the interior, with the structure mostly lacking exterior windows except at the entrance. This enclosed configuration creates a controlled but challenging environment for localization.

The laboratory spaces contain dense concentrations of metallic research equipment (including mass spectrometers, centrifuges, and specialized analytical instruments) that generate strong, temporally stable magnetic signatures. Some equipment creates field perturbations exceeding 50 μT, far stronger than the structural anomalies in other buildings. These distinctive signatures provide reliable landmarks for

magnetic-based localization, though they also create sharp spatial gradients that challenge interpolation-based positioning approaches.

The enclosed nature of the building, with exterior windows limited to the entrance area, results in more stable ambient lighting conditions compared to other testbed buildings but also constrains the variety of visual features available for vision-based localization. With moderate WiFi coverage from 38 access points, LRC's electromagnetic landscape is shaped primarily by the dense research equipment rather than architectural features. This makes LRC valuable for evaluating how localization algorithms handle environments where equipment-induced electromagnetic signatures dominate over structural ones. Data collection in LRC yielded 1,776 m^2 of mapped space.

Environmental Diversity Rationale: The four buildings collectively span the range of indoor environments encountered in institutional and commercial settings. PSE's traditional concrete construction with square corridors and peripheral research labs represents the dominant building style in established educational and science campuses. BCE's contemporary glass-and-steel design with irregularly shaped corridors connecting labs, classrooms, and social spaces exemplifies modern sustainable architecture increasingly common in new construction. STC's linear open-sided pedestrian walkway captures the dynamics of high-traffic thoroughfares. LRC's enclosed research and computer laboratory configuration reflects specialized environments where equipment constraints shape the electromagnetic landscape.

This diversity ensures that algorithms validated on the testbed demonstrate generalization capability beyond any single architectural paradigm. Cross-building evaluation reveals which techniques are robust to construction material differences (concrete vs. glass-steel), geometric variation (square vs. linear vs. rectangular), and environmental dynamics (stable laboratories vs. variable public spaces). The following sections leverage this multi-building testbed to characterize temporal fingerprint stability and evaluate localization algorithm performance under realistic diversity of conditions.

3.5 Longitudinal Dataset

The evaluation of indoor localization systems under real-world conditions requires datasets that capture not only spatial diversity but also temporal evolution. Single-epoch evaluations, however comprehensive in spatial coverage, cannot assess the critical question of how positioning systems degrade over time as environments change. To address this limitation, two data collection campaigns spanning fall 2021 and fall 2022 were conducted, creating a longitudinal dataset that enables systematic study of temporal fingerprint stability across the multi-building testbed.[1]

The timing of both campaigns during the fall semester was deliberate, so that the seasonal factors of HVAC operation, natural lighting, and the academic calendar matched between the two years. Academic buildings follow a regular schedule of classes, laboratory activities, and administrative operations during the fall term, so this alignment removes seasonal confounds from the interpretation of fingerprint drift. It does not, however, make the two campaigns comparable in occupancy. The 2021 campaign took place under

[1]The dataset is publicly available at https://yorkuindoor.ausmlab.com; availability details are given in Section 3.9.

the COVID-19 campus restrictions described in Section 3.5.2, and occupancy in 2021 was lower than in 2022. The cross-year comparison therefore combines infrastructure change with an occupancy increase, a confound to keep in mind when interpreting cross-year results (Section 3.8.4).

Collection Protocol and Timing: Both collection campaigns followed the standardized protocol described in Section 3.2.4, ensuring methodological consistency across the one-year interval. The two-operator approach, with the HoloLens operator providing spatial guidance and the Android operator recording sensor data, was maintained throughout, using the same equipment and personnel to minimize procedural variations. Each reference point received a 30 s stationary collection period, yielding 10–15 WiFi scans and thousands of IMU and magnetometer samples per location.

The 2021 campaign established the baseline fingerprint database, visiting all accessible areas within the four-building testbed and creating the initial anchor network via Azure Spatial Anchors. Collection proceeded building-by-building over a six-week period from September through October 2021, with each building requiring 2–4 days of concentrated collection effort depending on floor area and accessibility constraints. The total 2021 effort yielded comprehensive coverage of corridors, common areas, and accessible office spaces.

The 2022 campaign began with a re-localization phase, where the HoloLens operator verified that the 2021 anchor positions remained accurate by scanning the environment and comparing detected anchor positions against their stored coordinates. This validation measured a mean positional deviation of 1.8 cm across the 312 anchors tested, with a 95th percentile of 4.2 cm, consistent with the re-localization accuracy reported for Azure Spatial Anchors [121]. Following successful re-localization, the Android operator collected fresh fingerprint samples at each reference point location, ensuring that 2021 and 2022 measurements correspond to identical physical positions within the centimeter-level accuracy of the anchor system.

The 2022 campaign also expanded coverage by approximately 15%, incorporating areas that were inaccessible during the 2021 collection due to construction, restricted access, or scheduling conflicts. These newly accessible zones included previously restricted administrative offices in PSE, renovated conference areas in BCE, and additional laboratory spaces in LRC. The expanded coverage enables evaluation of how well models trained on limited spatial data generalize to adjacent but previously unobserved regions.

Dataset Statistics: Table 3.3 summarizes the sample counts collected across both years, broken down by building. The total dataset comprises 9,659 fingerprint samples spanning 847 unique reference points, with the 2022 expansion adding both new locations and repeated measurements at existing points.

The sample-to-reference-point ratio varies by building, reflecting differences in collection intensity and the number of sessions conducted at each location. PSE, with its uniform grid layout, achieved the highest sampling density (approximately 5.9 samples per reference point in 2021), while STC's limited accessibility during peak hours resulted in sparser coverage (4.9 samples per point). The 2022 campaign increased sampling density across all buildings, with particular emphasis on areas that exhibited high fingerprint variability during 2021 analysis.

Table 3.3: Longitudinal dataset statistics showing sample counts per building and year. The 2022 campaign revisited all 847 reference points from 2021 and added approximately 15% new locations in previously inaccessible areas; the reference-point column counts the locations common to both years. The two campaigns yielded 9,659 fingerprint samples that enable cross-year validation of temporal robustness.

Building	2021 Samples	2022 Samples	Ref. Points	Area (m^2)	Growth (%)
PSE	1,842	2,156	312	12,544	17.1
BCE	1,523	1,687	289	13,800	10.8
STC	428	512	87	702	19.6
LRC	687	824	159	1,776	19.9
Total	**4,480**	**5,179**	**847**	**28,822**	**15.6**

Year-over-Year Environmental Changes: The one-year interval between collection campaigns captured substantial environmental evolution that directly impacts localization system performance. A comprehensive infrastructure survey conducted prior to the 2022 collection documented four categories of changes.

Network Infrastructure Changes. The York University campus WiFi network underwent significant expansion between the collection epochs. Over the 2021–2022 academic year the network saw roughly 59 IEEE 802.11 access points installed, predominantly in areas that had exhibited coverage gaps during routine network monitoring, while 12 existing access points were decommissioned, a net increase of 47. Each new AP's MAC address and approximate installation location was recorded for subsequent analysis of how fingerprint-based systems adapt to infrastructure growth. The net per-building gains were 18 in PSE, 15 in BCE, 8 in STC, and 6 in LRC, raising the total AP count from 192 to 239, a 24.5% increase that fundamentally altered the RF landscape in affected areas.

Physical Space Modifications. Furniture rearrangements affected approximately 23% of office areas across the testbed, with the most substantial changes occurring in PSE's faculty offices and BCE's collaborative workspaces. While furniture does not create the dramatic signal attenuation of structural walls, it does influence multipath propagation by adding or removing reflective surfaces. The BCE building underwent partial second-floor renovations during spring 2022, including drywall construction and electrical work that temporarily introduced metallic scaffolding and equipment. These construction activities had concluded before the fall 2022 collection, but their effects on the built environment, including new wall finishes and relocated electrical panels, persisted.

Equipment Installations. The Lassonde Research Centre experienced the most significant equipment changes, with the installation of new MRI machines and electron microscopes in laboratory spaces. These devices create pronounced magnetic field perturbations: the new MRI system generates localized anomalies exceeding 50 µT deviation from the ambient geomagnetic field at distances of several meters, far exceeding the structural anomalies documented elsewhere in the testbed. While such strong perturbations complicate magnetometer-based positioning in their immediate vicinity, they also create highly distinctive landmarks that, once mapped, provide robust localization features.

HoloLens System Updates. Microsoft released spatial mapping algorithm updates for the HoloLens 2 between the collection epochs, potentially affecting anchor placement precision and re-localization accuracy.

Anchor stability was validated through systematic re-localization checks at the start of the 2022 campaign, measuring the deviation between detected and stored anchor positions. Across 312 anchor locations tested, the mean positional drift was 1.8 cm with a 95th percentile of 4.2 cm, confirming that the coordinate system remained consistent despite software updates.

Coverage Visualization: Figure 3.5 presents data collection coverage for the 2021 and 2022 campaigns overlaid on each building's floor plan, with warmer colors indicating higher sample density. Because fingerprints were gathered while traversing each building's corridors and walkways, coverage concentrates along these circulation routes, with interior rooms of restricted access remaining largely unsampled.

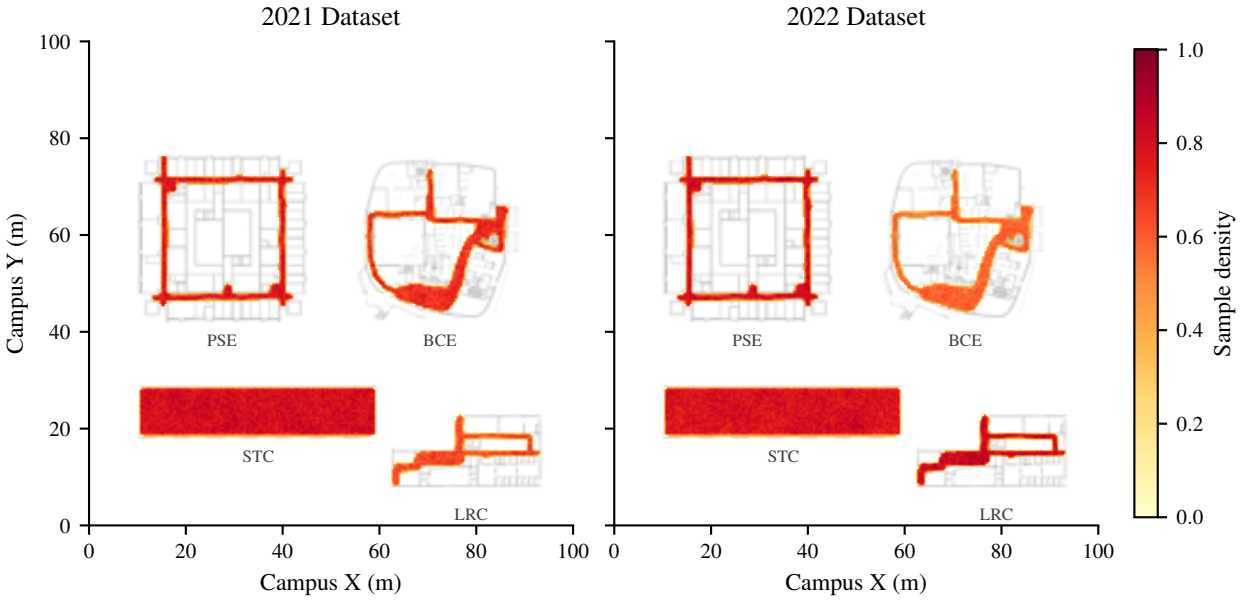


Figure 3.5: Data collection coverage for the 2021 (left) and 2022 (right) campaigns across the four-building testbed, shown as sample density along the fingerprint-collection routes overlaid on each building's floor plan. Warmer colors indicate higher sample density; coverage follows the corridors and walkways actually traversed during collection, while interior rooms remain largely unsampled. The open-walkway Student Teacher Centre is covered along its full length. Both campaigns revisit the same circulation routes, enabling year-over-year comparison at matched locations, with the reduced coverage in the Bergeron Centre the most visible change in 2022.

The heatmaps highlight several notable patterns. Corridors, walkways, and building entrances carry the highest sample density, reflecting the corridor-based collection protocol and their role as primary circulation routes. Interior office and laboratory spaces remain sparsely sampled, constrained by access permissions and the focus on publicly traversable areas. Because both campaigns follow the same routes, the longitudinal dataset captures matched locations across years; the principal year-over-year change is the reduced density in the Bergeron Centre during 2022, consistent with the infrastructure and renovation modifications documented earlier in this chapter.

The longitudinal nature of this dataset distinguishes it from prior indoor localization benchmarks, which typically span days or weeks rather than years. The documented infrastructure changes (a net 47 new

APs, equipment installations, and physical modifications) represent realistic environmental evolution that deployed positioning systems must accommodate. The following section describes how raw sensor measurements from this dataset are transformed into discriminative features suitable for localization algorithms.

3.5.1 Data Collection Timeline

The temporal structure of the data collection campaigns reflects deliberate experimental design choices that enable rigorous evaluation of localization system stability. This subsection documents the detailed collection timeline (Figure 3.6), session structure, and quality assurance procedures that ensure the dataset supports meaningful temporal analysis.

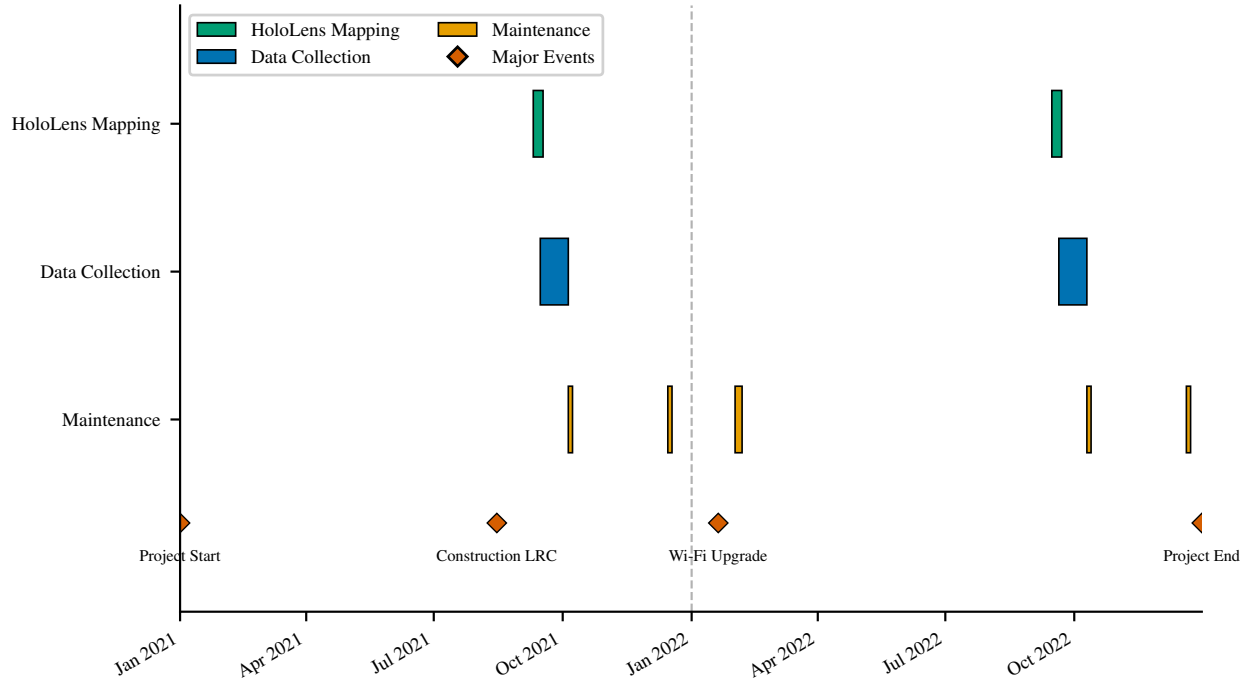


Figure 3.6: Data collection timeline showing the two-year campaign structure. The 2021 campaign (September–October) established the baseline fingerprint database across all four buildings. The 2022 campaign (September–October) revisited all reference points while expanding coverage by approximately 15%. The one-year interval captures a full academic cycle of infrastructure changes, equipment modifications, and occupancy shifts – the longest temporal span in comparable multi-building fingerprinting studies.

2021 Campaign: The initial data collection campaign spanned a six-week period from early September through mid-October 2021, coinciding with the first half of the fall academic semester. This timing was chosen to coincide with the regular academic schedule of students, faculty, and staff. In 2021, however, occupancy remained below typical levels under the COVID-19 restrictions documented in Section 3.5.2. Collection proceeded building-by-building in the sequence PSE, BCE, STC, LRC, with each building requiring 2–4 concentrated collection days depending on floor area and accessibility.

The 2021 campaign yielded 4,480 fingerprint samples distributed across 847 reference points. PSE, with its largest floor area and most uniform accessibility, contributed the most samples (1,842), while STC's limited size and high-traffic nature constrained collection to 428 samples. Each reference point received an average of 5.3 fingerprint samples, with multiple samples at each location enabling statistical characterization of short-term signal variability.

2022 Campaign: The second data collection campaign followed an identical six-week window from September through October 2022, maintaining temporal alignment with the 2021 effort to control for seasonal variations in building operation. The campaign began with a systematic anchor re-verification phase, where the HoloLens operator confirmed that Azure Spatial Anchors remained accurately localized at their original positions. This verification measured a mean anchor deviation of 1.8 cm across the four buildings, validating the stability of the spatial reference frame.

The 2022 campaign expanded collection to include areas that were inaccessible during 2021 due to construction, restricted access, or scheduling conflicts. These newly accessible zones included administrative offices in PSE that relaxed access restrictions, conference areas in BCE that underwent refurbishment between campaigns, and renovated laboratory spaces in LRC. The expanded coverage added approximately 15% more reference points while also increasing sampling density at locations identified as high-variability during 2021 analysis. The 2022 campaign yielded 5,179 fingerprint samples, a 15.6% increase over 2021.

Collection Protocol: Each collection session followed a standardized protocol designed to minimize procedural variability while accommodating the practical constraints of operating in occupied academic buildings. Sessions typically lasted 2–4 hours, scheduled during morning (8:00–12:00) or afternoon (13:00–17:00) periods to avoid peak occupancy times that could impede operator movement. The session structure comprised:

1. **Equipment initialization** (15 minutes): HoloLens 2 power-on and spatial anchor re-localization; smartphone sensor calibration and NTP time synchronization.
2. **Systematic collection** (90–180 minutes): Sequential navigation through pre-planned anchor routes, with 30-second stationary collection at each reference point.
3. **Quality verification** (15 minutes): Review of collected samples for completeness; identification of missed or failed anchor visits for immediate re-collection.
4. **Data upload** (10 minutes): Transfer of collected samples to secure cloud storage with integrity verification.

The two-operator protocol maintained consistency throughout both campaigns. The same HoloLens operator participated in both 2021 and 2022 collection to ensure procedural familiarity, while Android operators were rotated to prevent device-handling biases from accumulating.

Data Quality Controls: Multiple quality assurance mechanisms ensured the reliability of collected fingerprint samples. Real-time validation during collection flagged samples where HoloLens tracking confidence fell below 90%, indicating potential spatial anchor localization errors. These flagged samples were marked for immediate re-collection at the affected reference points.

Post-collection verification identified trajectory continuity breaks where the sequence of visited anchors exhibited physically implausible gaps, suggesting missed anchors or navigation errors. Approximately 3% of initially collected samples failed quality verification and were replaced through targeted re-collection sessions.

Outlier detection analyzed the distribution of WiFi RSSI values at each reference point, flagging samples with anomalously low signal strength that might indicate antenna occlusion or device malfunction. Manual inspection of flagged outliers confirmed that fewer than 1% required exclusion from the final dataset.

The combination of rigorous collection protocol and multi-stage quality verification ensures that the longitudinal dataset provides a reliable foundation for evaluating localization system performance across temporal variations. The documented timeline structure further enables researchers to correlate performance changes with specific environmental events that occurred during the 2021–2022 interval.

3.5.2 Data Collection Campaign Details

The timeline above describes the two campaigns at the level of the six-week window. This subsection records the operational details of each campaign in as much detail as the collection records allow: approximate per-building collection windows, the people and devices involved, the floors covered, the occupancy conditions, and the survey effort. The per-building collection windows, operator roles, and sample counts are tabulated in Section A.5.

Per-Building Collection Windows: Both campaigns occupied a six-week window from early September to mid-October and proceeded building-by-building in the fixed order PSE, BCE, STC, LRC, with 2–4 collection days per building. Because the order was fixed and the two largest buildings were surveyed first, the approximate windows are as follows: PSE in the first part of the window (approximately campaign weeks 1–2, early to mid September), BCE in the middle part (approximately weeks 3–4, late September), and STC and LRC, the two smallest buildings, in the final part (approximately weeks 5–6, the first half of October). These windows are approximate. They are derived from the campaign schedule (the six-week window, the building order, and the 2–4 collection days per building) rather than from day-level session logs. The collection days within each window were interleaved with anchor deployment in 2021, anchor re-verification in 2022, targeted re-collection sessions, and scheduling around building operations. The same windows applied in both years, so the two campaigns are aligned by season.

Sessions, Operators, and Handsets: Sessions lasted 2–4 hours in the morning (8:00–12:00) or afternoon (13:00–17:00) slots described above, and each was staffed by two operators, one wearing the HoloLens 2 and one carrying the smartphone. The same individual served as HoloLens operator in every session of both

campaigns; Android operators were rotated. The hardware comprised HoloLens 2 headsets and Samsung Galaxy S21 handsets. The Galaxy S21 was the only smartphone model used in either campaign, so that WiFi chipset sensitivity, IMU calibration, and magnetometer bias were held constant across the two years. As a result, the dataset contains no cross-device variation, a limitation discussed in Section 6.3.4.

Software: The HoloLens 2 ran Windows Holographic OS, the anchor application was built with the Mixed Reality Toolkit (MRTK) and the Azure Spatial Anchors SDK, and the smartphone ran GetSensorData [122] on Android. Spatial-mapping updates released between the campaigns are one reason the 2022 campaign opened with the anchor re-verification reported above.

Floors Surveyed: Reference points in PSE span the corridors and common areas of all four floors (Section 3.4). The floor of every reference point is fixed by its anchor and stored with each record (Section A.2), and localization is evaluated as horizontal positioning within the known floor.

Occupancy Conditions: Sessions were scheduled outside peak occupancy periods, so the sampled occupancy is that of ordinary weekday mornings and afternoons rather than lunch-hour peaks. No per-session occupancy record was kept. The session slot and the campaign year are the only available proxies for occupancy, and the difference in campus occupancy between 2021 and 2022 discussed below is the main occupancy contrast in the dataset.

Survey Effort: With two operators per session and sessions of 2–4 hours, the person-hours of a session are twice its duration. The 30 s stationary capture alone limits the collection rate to 120 anchors per hour. The actual rate, which includes walking between anchors, re-localization checks, and repeat captures, was lower. Together with the 2–4 collection days per building, these figures are the basis for the effort estimates in Section 3.7.

COVID-19 Conditions During the 2021 Campaign: During the 2021 campaign (September to October 2021), the collection was carried out under COVID-19 restrictions. Fewer people were on campus, access to buildings was limited, and some areas were closed or restricted. By the 2022 campaign these restrictions had eased. The effects on the dataset are as follows. Building occupancy in 2021 was lower than in 2022, and the seasonal alignment of the two campaigns does not compensate for this difference. Several areas that were inaccessible in 2021, including administrative offices in PSE, conference areas in BCE, and laboratory spaces in LRC, were surveyed only in 2022. These restrictions, together with the construction and scheduling constraints, account for the approximately 15% expansion of coverage in 2022. The COVID-19 restrictions were one of the factors that limited the scope of the collection to the four buildings described in Section 3.4.

Three consequences for the analysis follow from these conditions. The 2021 to 2022 occupancy difference is a confound in the cross-year comparison: the degradation of a model trained on 2021 data and tested on 2022 data (Sections 3.8.4 and 5.6.3) reflects the combined effect of infrastructure change and an occupancy increase, and the two cannot be separated with this dataset. Occupancy-dependent multipath, caused by body absorption and reflection, was under-sampled in 2021 relative to 2022, so the 2021 training data represent a quieter radio environment than the 2022 test data. This is a plausible direction of deployment shift but not a controlled one. STC, whose occupancy varies most, is likely the building most affected. The building set is limited to four, which limits the architectural diversity of the testbed. Two observations limit the reach of the occupancy confound. Same-year evaluations are unaffected, because training and test data within a year share the same occupancy conditions. The cross-year evaluation on the independent IPIN 2016–2017 collections (Section 5.6.9) predates the pandemic and shows the same ordering of static and adaptive degradation, which indicates that the cross-year effect is not solely an artifact of the occupancy change.

3.5.3 Ethics and Privacy

The study involved no human participants. The only people whose movement was recorded were the two research-team operators who conducted each collection session, and what was recorded of them was the collection route itself: the sequence of anchors visited, the handset's sensor streams at each anchor, and the HoloLens 2 pose used to register the ground truth. No students, staff, or visitors were recruited, observed, or recorded, and no data were collected about people present in the buildings during collection.

No personal data were collected. The WiFi scans log the BSSIDs (hardware addresses) of institutional access points and the received signal strength of each, and nothing else. They do not record client devices, network traffic, or payload, so the dataset contains no identifier of any person's phone or laptop. The IMU and magnetometer streams describe the motion of the collection handset. No images or audio were recorded. The HoloLens 2 RGB camera was not used (Section 3.2.1), and the anchor fingerprints uploaded to Azure Spatial Anchors are visual feature descriptors of static architectural surroundings extracted from the tracking cameras (Section 3.2.2), not stored images [121]. The released record formats (Section A.2) contain only timestamps, building and floor identifiers, building-local coordinates, access-point BSSIDs, RSSI values, inertial and magnetic readings, and tracking confidence.

This is stated explicitly because location data are sensitive. Location traces of individuals reveal routines, home and workplace, and associations, and indoor positioning systems in particular can expose fine-grained movement within a building [113]. The dataset released with this thesis contains fingerprints tied to fixed reference points and the operators' collection routes. It contains no movement trace of any other person, and the collection routes are the serpentine survey paths of Section 3.2.4 rather than anyone's ordinary movement. Privacy and security risks that arise when a positioning system built from such data is deployed to end users are a separate matter and are discussed in Section 6.3.6.

Collected anchor and sensor data were uploaded at the end of each session to secure cloud storage, with access limited to members of the research team; the public release described in Section 3.9 contains the same non-personal fields. Collection in restricted areas was subject to the access permissions noted in

Section 3.4. Because the study involved no human participants and collected no personal information, it was treated as falling outside the scope of research requiring review by the York University Human Participants Review Committee (HPRC). For this reason, no HPRC review or exemption was sought.

3.6 Preprocessing Pipeline

The transformation of raw sensor measurements into discriminative features suitable for localization algorithms requires careful consideration of each modality's characteristics. Raw sensor data exhibits substantial noise, varying dimensionality across locations, and modality-specific artifacts that must be addressed before effective position estimation. The preprocessing pipeline addresses these challenges while preserving location-relevant information and maintaining computational efficiency for real-time applications.

The pipeline processes each sensor modality through modality-specific preprocessing stages before combining them into a unified feature vector. WiFi signals require filtering for temporal stability and dimensionality normalization for variable AP counts. IMU measurements enable motion state detection and trajectory constraints. Magnetometer readings demand calibration for device-specific biases before revealing location-discriminative anomalies. The following subsections detail each preprocessing stage, culminating in the 109-dimensional feature vector that feeds into the localization algorithms evaluated in Section 3.8.

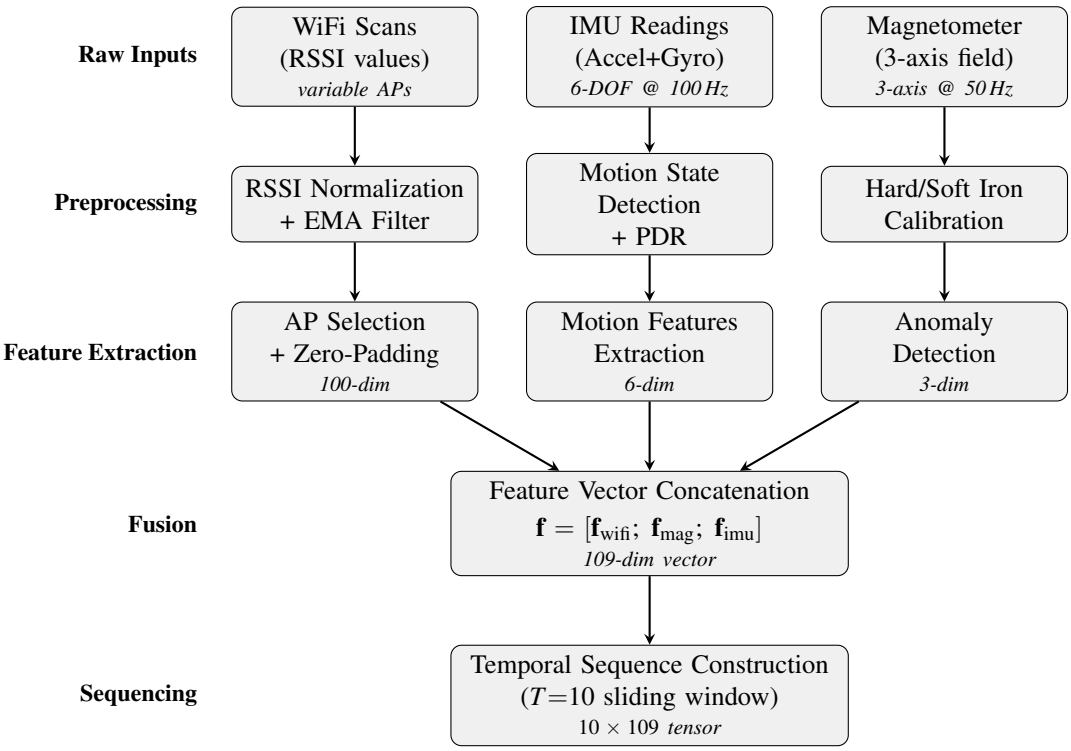


Figure 3.7: Preprocessing pipeline flowchart showing the transformation of raw sensor inputs into the final feature representation. Each modality undergoes modality-specific preprocessing (RSSI normalization for WiFi, motion detection for IMU, calibration for magnetometer) before feature extraction produces fixed-dimension vectors. The pipeline reduces heterogeneous raw inputs to a unified 109-dimensional feature vector assembled into temporal sequences of length $T = 10$ for sequential models.

Figure 3.7 illustrates the complete preprocessing workflow from raw sensor acquisition to the final feature representation. The rationale for each preprocessing step derives from the characteristics of the corresponding sensor modality. WiFi RSSI values require EMA filtering to suppress the 5–10 dBm short-term fluctuations caused by environmental dynamics, followed by zero-padding to create fixed-dimension vectors despite variable AP visibility. IMU preprocessing focuses on motion state detection through accelerometer variance monitoring, enabling the system to distinguish stationary fingerprinting periods from walking trajectories where dead reckoning constraints become relevant. Magnetometer calibration removes device-specific hard iron and soft iron distortions that would otherwise mask the location-dependent magnetic anomalies from building structures.

The pipeline ensures consistent feature representation across all buildings in the testbed despite their substantial architectural differences. The fixed 109-dimensional output (100 WiFi + 6 IMU + 3 magnetometer) accommodates the heterogeneous AP deployments ranging from 38 APs in LRC to 89 APs in PSE through the zero-padding strategy. Similarly, the standardized calibration procedures for IMU and magnetometer sensors ensure that features extracted in the steel-intensive PSE building remain comparable to those from the glass-curtain BCE building, despite the dramatically different electromagnetic environments. This consistency is essential for training models that generalize across buildings and for the cross-building evaluation presented in subsequent chapters.

3.6.1 WiFi Fingerprint Processing

WiFi received signal strength measurements exhibit substantial temporal fluctuations due to environmental factors including human movement, door openings, and electronic interference. These variations, while problematic for direct fingerprint matching, can be mitigated through appropriate filtering and robust feature extraction.

EMA Filtering for Temporal Stability: Raw RSSI measurements from a single access point can fluctuate by 5–10 dBm within seconds even at a stationary location, masking the underlying position-dependent signal. The EMA filter introduced in Equation (3.3) is applied to each AP's RSSI time series independently. The filter's prediction weight $\alpha = 0.7$ was determined through empirical optimization on a held-out validation set, balancing responsiveness to genuine environmental changes against rejection of transient fluctuations. This filtering approach reduces RSSI variance by approximately 40% while preserving the mean values essential for fingerprint discrimination.

During the 30 s stationary collection at each reference point, the filtered RSSI values are aggregated to compute both mean and standard deviation for each visible AP. The mean serves as the primary fingerprint feature, while the standard deviation provides a measure of signal stability that can inform algorithm confidence estimates.

AP Clustering for Infrastructure Robustness: The dynamic nature of WiFi infrastructure, with APs being added, removed, or reconfigured, poses challenges for fingerprint-based systems that assume a fixed AP set.

This is addressed through clustering APs based on their spatial co-visibility patterns:

$$C_j = \{AP_i \mid \text{sim}(\mathbf{v}_i, \mathbf{v}_j) > \theta\} \tag{3.8}$$

where $\mathbf{v}_i \in \{0,1\}^N$ represents the binary visibility vector of AP_i across all N reference points (1 if the AP was detected at that location, 0 otherwise), $\text{sim}(\mathbf{u},\mathbf{v}) = |\mathbf{u} \cap \mathbf{v}|/|\mathbf{u} \cup \mathbf{v}|$ is the Jaccard similarity between binary visibility vectors, and $\theta = 0.85$ is the Jaccard similarity threshold. APs within the same cluster exhibit similar spatial coverage patterns, suggesting physical proximity or similar propagation characteristics.

Clustering provides robustness to infrastructure changes: when a new AP appears, its visibility pattern determines cluster membership, and the system can leverage existing cluster knowledge to interpret its signals. When an AP disappears, its absence affects only the cluster-level features rather than requiring complete retraining. In the testbed, the 239 APs group into 47 clusters, substantially reducing the effective dimensionality while preserving discriminative power.

Zero-Padding for Fixed Dimensionality: Machine learning algorithms typically require fixed-dimension inputs, yet the number of visible APs varies substantially across locations, from 8 in shielded laboratory spaces to over 40 in open corridors near building cores. This is addressed through zero-padding: RSSI values from detected APs are arranged into a vector indexed by a building-specific AP registry, with undetected APs assigned a value of zero (representing "not detected").

The AP registry is constructed during initial data collection by enumerating all unique MAC addresses observed across the building. The registry ordering is fixed once established, ensuring that the same AP always maps to the same feature dimension. The registry is zero-padded to a fixed length of 100 dimensions, accommodating the current AP deployment with headroom for infrastructure growth. The final WiFi feature vector $\mathbf{f}_{wifi} \in \mathbb{R}^{100}$ thus has consistent dimensionality regardless of how many APs are visible at any particular location.

3.6.2 IMU Feature Extraction

The 6-DOF inertial measurement unit provides complementary information about user motion that constrains the solution space for position estimation. While IMU-only positioning suffers from unbounded integration drift, the short-term motion information proves valuable for trajectory smoothing and motion state classification.

Sensor Configuration and Sampling: The GetSensorData application samples both accelerometer and gyroscope at 100 Hz, capturing the full dynamics of pedestrian motion including individual footsteps. The accelerometer measures specific force in m/s^2, while the gyroscope measures angular velocity in rad/s. Both sensors report 3-axis measurements in the device coordinate frame.

Motion State Detection: Accurate motion state classification enables adaptive processing strategies: during stationary periods, WiFi and magnetometer fingerprints can be directly matched against the reference database with high confidence, while during motion, dead reckoning constraints limit plausible position changes. The variance-based stationary detection from Equation (3.4) classifies the user's motion state by monitoring accelerometer magnitude variance over a sliding window.

Beyond the binary stationary/moving classification, specific motion patterns that inform localization are identified. Ascending or descending stairs produces characteristic vertical acceleration patterns distinct from level walking. Elevator use creates prolonged vertical acceleration followed by extended stationary periods at discrete floor levels. These patterns help disambiguate WiFi fingerprints that might appear similar across different building levels or stairwell locations.

Dead Reckoning Constraints: For users in motion, IMU data enables pedestrian dead reckoning (PDR) that estimates displacement between consecutive fingerprint samples. The gyroscope-derived heading change, combined with step detection from accelerometer patterns and stride length estimation, provides short-term position constraints. Between consecutive WiFi scans (approximately 2 seconds apart), PDR-derived displacement estimates typically achieve errors below 0.5 m, providing valuable constraints that improve fingerprint-based position estimates.

Feature Representation: For stationary fingerprints, IMU features are not directly included in the feature vector, as position estimation relies on the location-dependent WiFi and magnetometer signatures. For trajectory-based localization with moving users, the 6-DOF IMU measurements (3 accelerometer + 3 gyroscope channels) are appended to the feature vector, yielding $\mathbf{f}_{imu} \in \mathbb{R}^6$. The motion state flag determines whether IMU features are active for a given fingerprint sample.

3.6.3 Magnetometer Calibration

Magnetometer-based localization exploits the location-specific perturbations of Earth's geomagnetic field caused by ferromagnetic building materials. However, raw magnetometer readings are corrupted by device-specific biases that must be removed through calibration before the environmental signatures become usable for positioning.

Device-Specific Calibration: The hard iron and soft iron corrections introduced in Section 3.3.4 address systematic biases from the smartphone's internal components. Hard iron bias $\mathbf{b}_{hard}$ arises from permanent magnets (speakers, battery components) and manifests as a constant offset applied to all measurements. Soft iron distortion matrix $\mathbf{W}$ arises from ferromagnetic materials that redirect the ambient field, scaling and rotating measurements along different axes.

Calibration is performed through a standardized figure-eight waving motion that samples the magnetic field from diverse device orientations. The motion should span at least 30 seconds to ensure sufficient

angular coverage. The hard iron bias is estimated as the center of the measurement sphere, while the soft iron matrix is computed to transform the ellipsoidal measurement locus back to the expected spherical shape. Calibration is performed at the start of each collection session and calibration quality is periodically verified by checking that the calibrated field magnitude matches the expected Earth field strength.

Environment-Specific Anomaly Detection: Structural elements create distinctive magnetic signatures that serve as natural landmarks for positioning. Building on the anomaly detection framework from Equation (3.7), locations where the calibrated field deviates significantly from the expected geomagnetic baseline are identified:

$$\psi(\mathbf{p}) > \tau_m \tag{3.9}$$

where $\tau_m = 10\ \mu$T defines the anomaly threshold. Locations satisfying this criterion exhibit perturbations from structural steel, reinforced concrete, or equipment that distinguish them from surrounding areas.

In the testbed, magnetic anomalies concentrate near steel staircases (perturbations exceeding 20 μT), elevator shafts, reinforced concrete columns, and electrical distribution panels. The new equipment installed in LRC between 2021 and 2022, including MRI machines and electron microscopes, creates particularly pronounced anomalies exceeding 50 μT, providing distinctive landmarks once mapped into the fingerprint database.

Feature Representation: The calibrated 3-axis magnetometer readings form the magnetic feature vector $\mathbf{f}_{mag} \in \mathbb{R}^3$. Unlike WiFi fingerprints with their variable AP counts, magnetometer features have inherently fixed dimensionality. The three channels encode both the Earth field baseline and any location-specific perturbations, with the latter providing the discriminative signal for positioning.

3.6.4 Feature Vector Representation

The preprocessed sensor features are concatenated into a unified feature vector that serves as input to localization algorithms. The combination leverages the complementary strengths of each modality: WiFi provides broad spatial discrimination across the building, magnetometer features offer temporally stable landmarks near structural elements, and IMU data constrains motion trajectories.

Combined Feature Vector: For stationary fingerprints, the feature vector combines WiFi and magnetometer modalities:

$$\mathbf{f}_{stationary} = [\mathbf{f}_{wifi}; \mathbf{f}_{mag}] \in \mathbb{R}^{103} \tag{3.10}$$

where $\mathbf{f}_{wifi} \in \mathbb{R}^{100}$ contains the zero-padded RSSI values and $\mathbf{f}_{mag} \in \mathbb{R}^3$ contains the calibrated magnetometer readings. For input to the localization algorithms, which operate on a uniform 109-dimensional format, the

six inertial components are set to zero in stationary samples, so every fingerprint presented to the models shares the same dimensionality.

For trajectory-based localization with moving users, IMU features are appended:

$$\mathbf{f}_{motion} = [\mathbf{f}_{wifi}; \mathbf{f}_{mag}; \mathbf{f}_{imu}] \in \mathbb{R}^{109} \tag{3.11}$$

where $\mathbf{f}_{imu} \in \mathbb{R}^6$ contains the 6-DOF inertial measurements. The 109-dimensional feature vector captures the full sensor suite available from commodity smartphones without requiring specialized hardware.

Feature Standardization: The different sensor modalities operate on vastly different scales: RSSI values range from approximately -90 to -30 dBm, magnetometer readings span tens of microteslas, and IMU values depend on the motion state. To ensure that all features contribute appropriately to distance-based algorithms, z-score standardization is applied:

$$\mathbf{f}_{norm} = \frac{\mathbf{f} - \mu_f}{\sigma_f} \tag{3.12}$$

where μ_f and σ_f are the mean and standard deviation vectors computed from the training data. Standardization is performed per-feature across the training set, and the same transformation parameters are applied to test data to prevent data leakage.

Sequential Feature Representation: For sequential models that exploit spatial dependencies along the user's path, such as recurrent neural networks and LSTM architectures, the per-anchor fingerprints are grouped into sliding windows. Each window comprises 10 consecutive reference points visited along a collection trajectory, spanning roughly 10 m of the path at the 1 m anchor grid spacing and providing trajectory-level context while keeping the computational cost modest. Windows are formed only from reference points within the same trajectory, so no window spans a discontinuity between trajectories; together with the trajectory-level train–test partition (Section 3.8.1), this prevents overlapping windows from straddling the boundary.

The sequential input takes the form of a tensor with dimensions $T \times D$, where $T = 10$ is the sequence length and $D = 109$ is the feature dimension. This 10×109 tensor representation feeds into the sequential localization algorithms evaluated in Section 3.8, enabling them to leverage trajectory-level patterns that instantaneous fingerprints cannot capture.

Figure 3.8 illustrates the complete preprocessing pipeline, showing how raw sensor inputs flow through modality-specific processing stages before fusion into the combined feature representation.

3.7 Deployment Procedure for a New Site

The preceding sections describe the collection framework, the sensor acquisition, and the preprocessing pipeline as they were applied to the York University testbed. This section gathers the same steps into an

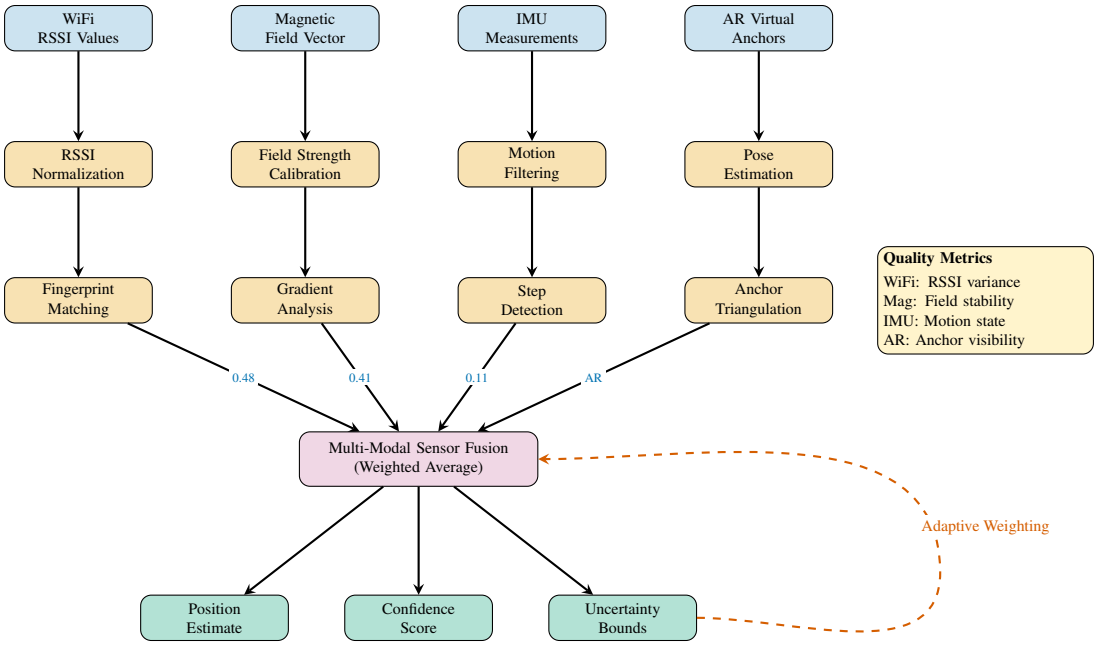


Figure 3.8: Multi-modal sensor fusion workflow showing the integration of WiFi fingerprinting, magnetic field measurements, IMU data, and AR virtual anchors. Each sensor modality contributes weighted inputs (WiFi: 48%, Magnetic: 41%, IMU: 11%) to the fusion algorithm, with adaptive weighting based on real-time quality assessments and confidence feedback loops.

ordered procedure for a new site, so that the framework can be reproduced elsewhere and its cost estimated in advance. No step is new here. Each entry points to the subsection in which it is described, and the parameter values are those used in the testbed.

Procedure:

1. **Floor plan registration** (Section 3.2.2). Load the CAD floor plan into the HoloLens 2 anchor application and let the device build an initial spatial mesh of the floor. Identify at least four non-collinear correspondence points between the plan and the physical environment (room corners, doorway edges, pillars), place a virtual marker at each, and compute the rigid transformation. Verify the overlay visually against walls and doorways; the expected registration accuracy is 2–5 cm. The procedure assumes that a CAD floor plan of the site is available.

2. **Anchor grid placement** (Section 3.2.2). Place anchors at 1 m grid intervals throughout the accessible area, then add anchors at points of interest: entrances and exits, corridor intersections, stairwell landings and elevator lobbies, and transitions between architectural zones. Each anchor stores its position, an orientation quaternion that fixes the direction the handset operator will face, a creation timestamp, and the tracking confidence at placement.

3. **Anchor persistence** (Section 3.2.2). Upload the anchors to Azure Spatial Anchors [121]. The anchors

then persist across sessions and devices, so collection can be spread over weeks and repeated in later years against the same coordinate frame; re-localization accuracy is approximately 2 cm.

4. **Anchor validation** (Section 3.2.3), where surveying equipment is available. Survey a subset of anchor positions independently, as was done in the testbed with a total station at 80 points, and confirm that anchor error is small relative to the target localization accuracy. This step is optional for a site that accepts the validation reported in this chapter as representative.

5. **Access point registry** (Sections 3.3.1 and 3.6.1). During the first collection sessions, enumerate every unique BSSID observed in the building, fix the ordering once, and zero-pad the registry to 100 dimensions, leaving unused slots for later additions. The registry determines the feature index of every access point for all subsequent collection.

6. **Session initialization** (Sections 3.2.4, 3.5.1 and 3.6.3), about 15 minutes. Power on the HoloLens 2 and re-localize against the stored anchor set until at least three anchors are recognized, which takes 3–5 minutes per floor. Start GetSensorData [122] on the handset, set the sensor rates (WiFi scans at 0.5 Hz, IMU at 100 Hz, magnetometer at 50 Hz), synchronize both devices by NTP to within 50 ms, and perform the figure-eight magnetometer calibration for at least 30 seconds.

7. **Path planning** (Section 3.2.4). Order the anchors by a nearest-neighbor route starting from the building entrance, following a serpentine pattern through corridors with branches into accessible rooms and avoiding backtracking through areas already collected.

8. **Per-anchor capture** (Section 3.2.4). At each anchor, the HoloLens operator confirms alignment with the virtual marker and announces the anchor identifier; the handset operator positions the phone at the anchor, oriented according to the stored quaternion, and records for 30 s while stationary, yielding 10–15 WiFi scans together with the IMU and magnetometer streams. The HoloLens operator watches for drift and calls for a repeat capture if movement is detected.

9. **Quality control** (Section 3.5.1). Flag any capture during which HoloLens tracking confidence fell below 90% and re-collect it immediately. After the session, check trajectory continuity for implausible gaps and screen the RSSI distributions at each anchor for outliers indicating antenna occlusion or device faults. In the testbed about 3% of captures were re-collected and fewer than 1% were excluded.

10. **Upload and preprocessing** (Section 3.6). Upload the session data to secure storage with integrity verification, then apply the preprocessing pipeline: EMA filtering of RSSI, zero-padding by the registry, magnetometer calibration, z-score standardization with statistics computed on the training set only, and construction of the 109-dimensional feature vectors and $T = 10$ sequences.

Effort: In the testbed each building required 2–4 collection days for floor areas between 702 m^2 (STC) and 13 800 m^2 (BCE), with sessions of 2–4 hours comprising about 15 minutes of initialization, 90–180 minutes of systematic collection, 15 minutes of quality verification, and 10 minutes of upload. Two operators are

required throughout, so person-hours are twice the session hours. An upper bound on the survey rate is given in Section 3.5.2. A repeat campaign at an instrumented site omits steps 1 to 5 for areas already instrumented and replaces them with anchor re-verification, which in 2022 covered 312 anchors and found a mean deviation of 1.8 cm (Section 3.5).

Grid Spacing and Area: The 1 m grid was chosen because WiFi signal strength varies measurably over 2–4 meters in typical indoor environments, so 1 m sampling resolves these gradients well below their characteristic scale (Section 3.4). The cost of this choice is that the number of anchors, and therefore the collection time at a fixed 30 s per anchor, grows with the surveyed area and with the inverse square of the grid spacing. A coarser grid reduces effort in proportion but also changes the physical extent of the sequential input: the $T = 10$ window spans about 10 m of path at 1 m spacing (Section 3.6.4) and would span proportionally more at wider spacing. A site with a large area and a limited collection budget can therefore trade spacing against coverage, but the models and results in this thesis were obtained at 1 m spacing and their accuracy at other spacings has not been evaluated.

Re-Survey Cadence: The cross-year results in this chapter and in Chapter 5 indicate how often a site should be re-surveyed. A static model loses 22% of its accuracy over one year in the testbed (Section 3.8.4), whereas the adaptive model of Chapter 5 loses 4% over the same interval (Section 5.6.3), and the simulated long-term deployment analysis there supports an annual retraining cycle (Section 5.7.5). The evidence in this thesis therefore supports an annual re-survey: re-verify the anchors, re-collect fingerprints at the existing reference points, and extend the survey to newly accessible areas, as the 2022 campaign did. A re-survey should be brought forward after events such as access point hardware failure and construction (Table 5.8).

Extending the Access Point Registry: Infrastructure change between surveys is handled through the registry rather than by rebuilding the feature space (Section 3.6.1). Decommissioned access points keep their index and read as zero, the value for "not detected". Newly observed access points are added to unused slots, so existing indices do not move and fingerprints from earlier surveys remain aligned with later ones. The co-visibility clustering of Equation (3.8) assigns a new access point to a cluster from its visibility pattern once enough observations exist. In the testbed, 59 additions and 12 decommissions between 2021 and 2022 altered on the order of 70 slots across the four buildings (Section 3.3.2), and the largest single-building count in 2022 was 89 (Table 3.2), leaving at most 11 unused slots in that building. If a building's cumulative count exceeds the registry length, the registry must be enlarged and the models retrained on the new dimensionality. This is a limitation of fixed-length fingerprint vectors.

Hardware and Software: The deployment requires a Microsoft HoloLens 2 running the custom MRTK anchor application, with an Azure Spatial Anchors account for persistence [121]; an Android smartphone running GetSensorData [122] (a Samsung Galaxy S21 in this thesis); a network time source for NTP syn-

chronization; CAD floor plans of the site; secure storage for session uploads; and, optionally, a total station for anchor validation (a Leica TS15 in this thesis). The equipment inventory is tabulated in Table A.13. Availability of the software components is stated in Section 3.9.

3.8 Baseline Algorithm Comparison

Having established the data collection methodology and preprocessing pipeline, this section evaluates the localization performance of seven algorithms on the longitudinal dataset. The central question guiding this evaluation is: *which algorithm architecture best handles temporal degradation?* While absolute accuracy matters for practical deployment, an algorithm's ability to maintain performance as environments change is equally critical for systems that must operate reliably over extended periods without constant retraining. This framing naturally leads toward architectures that can be examined in greater depth in subsequent chapters.

The evaluation encompasses traditional approaches, deep learning methods, and hybrid architectures that combine probabilistic modeling with temporal sequence processing. By assessing both within-year accuracy and cross-year stability, the VAE-LSTM architecture is identified as the most suitable foundation for further investigation, motivating the detailed architectural analysis in Chapter 4.

3.8.1 Evaluation Protocol

The experimental evaluation follows a consistent protocol across all algorithms, ensuring fair comparison of localization accuracy and temporal robustness. Standard machine learning practices are employed while incorporating domain-specific considerations for indoor positioning.

Data Partitioning: Each year's dataset is partitioned into training (70%) and testing (30%) subsets. The partition is applied at the level of complete collection trajectories rather than individual reference points: each trajectory—the ordered sequence of reference points visited along one collection route—is assigned in its entirety to either the training or the testing set, so the two sets share no trajectory. Trajectories are allocated by stratified sampling that preserves the distribution across buildings and floor regions, ensuring that both sets contain representative samples from all four buildings and from various zones within each building, and that algorithms are not evaluated exclusively on regions dissimilar to their training data. Because the sliding-window sequences used by the recurrent and hybrid models are constructed within individual trajectories, this partition also eliminates window leakage. Under a point-level random split, overlapping windows built around neighboring reference points share most of their constituent points; assigning such windows to different sets would let near-duplicate sequences appear in both training and testing and inflate apparent accuracy, whereas the trajectory-disjoint partition makes this impossible.

Cross-Validation: Within the training set, 5-fold cross-validation is performed for hyperparameter selection. Folds are formed over whole trajectories and maintain the stratified building distribution, so the same window-level disjointness holds during model selection; hyperparameters are selected based on the mean

validation error across folds. The final model is then trained on the complete training set with selected hyperparameters and evaluated on the held-out test set.

Evaluation Metrics: Three complementary metrics that capture different aspects of localization performance are reported:

- **Mean Absolute Error (MAE):** The average Euclidean distance between predicted and ground truth positions, measured in meters. This metric provides the overall accuracy summary.
- **90th Percentile Error:** The error threshold below which 90% of predictions fall. This metric captures worst-case behavior that is critical for user experience.
- **Temporal Consistency Score:** The relative accuracy degradation when models trained on one year's data are tested on another year's data, expressed as a percentage.

Within-Year and Cross-Year Validation: The evaluation distinguishes two scenarios. *Within-year validation* trains and tests on data from the same year, establishing baseline accuracy under ideal conditions. *Cross-year validation* trains on one year's data and tests on the other, simulating the realistic deployment scenario where a model trained on historical data must operate on future data without retraining. The cross-year protocol includes forward validation (2021 $\rightarrow$ 2022, simulating deployment) and backward validation (2022 $\rightarrow$ 2021, assessing temporal symmetry).

3.8.2 Algorithm Categories

Seven localization algorithms spanning three categories are evaluated: traditional methods, deep learning approaches, and hybrid/probabilistic architectures. Each algorithm receives the same 109-dimensional feature vector as input and produces a 2D position estimate. The descriptions below provide sufficient context for interpreting results; detailed architectural specifications appear in Chapter 4 for the VAE-LSTM architecture selected for further investigation.

Traditional: k-Nearest Neighbors (kNN): The kNN algorithm [11] serves as the non-parametric baseline, estimating position as the weighted average of the k nearest reference points in feature space. The value $k = 5$ is employed based on cross-validation, with weighted Euclidean distance that assigns higher influence to closer neighbors. Despite its simplicity, kNN remains competitive in indoor localization due to its ability to capture complex, non-linear relationships without parametric assumptions. The algorithm requires no training phase, enabling rapid deployment but lacking the ability to learn robust feature representations.

Deep Learning: Convolutional Neural Network (CNN): The CNN architecture applies 1D convolutions over the fingerprint feature vector, treating the ordered AP RSSI values as a spatial signal. The network

comprises three convolutional layers with 64, 128, and 256 filters respectively, each followed by batch normalization and ReLU activation. Max pooling reduces dimensionality between layers, and two fully connected layers produce the final position estimate. The CNN architecture excels at capturing local patterns in the fingerprint structure but may struggle when the AP ordering changes due to infrastructure modifications.

Deep Learning: Deep Multilayer Perceptron (DeepML): The DeepML architecture employs a deep fully connected network with residual connections [84] that enable training of deeper networks without gradient degradation. The architecture comprises six hidden layers with 512, 256, 256, 128, 128, and 64 units, with residual connections between alternate layers. Dropout regularization (rate 0.3) prevents overfitting to the training distribution. DeepML provides strong baseline performance while remaining architecturally simple compared to recurrent or variational approaches.

Deep Learning: Recurrent Neural Network (RNN): The vanilla RNN processes sequential fingerprint samples to capture temporal dependencies in user trajectories. The network comprises two recurrent layers with 128 hidden units each, followed by a fully connected output layer. While theoretically capable of learning long-range dependencies, the RNN suffers from vanishing gradient problems that limit effective sequence modeling. The architecture serves as a baseline for sequential approaches.

Probabilistic: Variational Autoencoder (VAE): The VAE [90] learns a probabilistic latent representation of the fingerprint space, providing both position estimates and uncertainty quantification. The encoder maps fingerprints to a distribution over latent codes, while the decoder reconstructs fingerprints from sampled latent vectors. A separate position regression head operates on the latent representation. The probabilistic formulation provides inherent robustness to input variations, as the latent space smooths over fingerprint fluctuations. The VAE's ability to quantify prediction uncertainty enables downstream confidence-based filtering.

Probabilistic: Long Short-Term Memory (LSTM): The LSTM architecture [85] addresses the vanishing gradient limitations of vanilla RNNs through gated memory cells that selectively retain or forget information across time steps. The implementation employs two stacked LSTM layers with 128 units each, processing sequences of 10 consecutive fingerprints. The LSTM captures trajectory-level patterns that instantaneous fingerprints cannot represent, such as motion direction and speed.

Hybrid: VAE-LSTM: The VAE-LSTM architecture combines the probabilistic representation learning of the VAE with the sequential modeling capabilities of the LSTM. The VAE encoder first maps each fingerprint in a sequence to the latent space, and the LSTM then processes the sequence of latent representations to produce the final position estimate. This hybrid approach inherits the VAE's robustness to fingerprint

variations while leveraging the LSTM's ability to model temporal dependencies. The combined architecture achieves the best performance at the cost of increased computational complexity.

Hyperparameter Selection Methodology: All hyperparameters were selected through systematic grid search with 5-fold cross-validation on the training partition, optimizing mean absolute error as the primary metric. For traditional machine learning algorithms (kNN, Random Forest, Gradient Boosting), standard parameter ranges from the literature were explored. For deep learning architectures, preliminary experiments established reasonable ranges for network depth, width, and regularization strength, followed by fine-grained search within those ranges. The VAE-LSTM required particular attention to the KL divergence weight (β) balancing reconstruction accuracy against latent space regularization, with values from 0.01 to 1.0 explored to prevent posterior collapse while maintaining meaningful latent representations.

Table 3.4 documents the complete hyperparameter configurations for all seven algorithms, enabling reproducibility and facilitating comparison with alternative approaches. The table organizes algorithms by category (traditional ML, deep learning, proposed) and includes implementation details such as optimizer choice, batch size, and training schedule. These configurations represent the result of the systematic tuning process and were held constant across all evaluation scenarios.

3.8.3 Comparative Findings

The seven algorithms were evaluated under the protocol above using both within-year and cross-year validation, aggregated across all four buildings. The comparison reproduced the trends widely reported in the fingerprinting literature: the learned approaches achieved lower error than the non-parametric kNN baseline, and the probabilistic and hybrid architectures (VAE, LSTM, and VAE-LSTM) achieved both the best accuracy and the lowest cross-year degradation. The VAE-LSTM hybrid was the strongest performer across both evaluation years and was therefore selected as the architecture for detailed development in the remainder of the thesis.

Figure 3.9 summarizes the VAE-LSTM's per-building accuracy across both collection campaigns: same-year error remains sub-meter in every building, increasing only modestly in the more challenging 2022 environment.

To avoid duplicating results across chapters, the full quantitative evaluation of the VAE-LSTM, including its same-year accuracy, cross-year degradation, per-building performance, and comparison against the baseline algorithms, is presented in Chapter 4 (Table 4.3). The following subsections examine the cross-year temporal stability that motivated this selection and analyze why the probabilistic architectures prove most robust to environmental change.

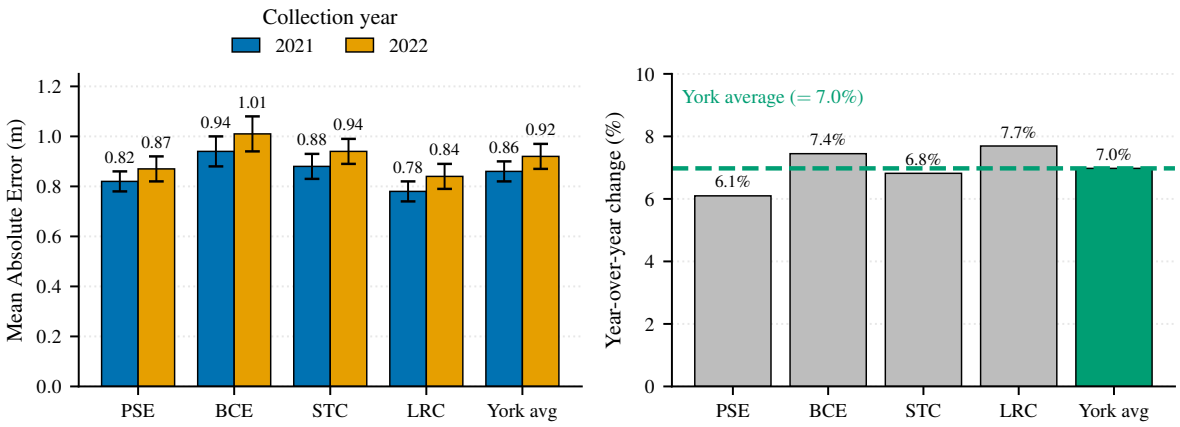
3.8.4 Cross-Year Temporal Stability

The critical evaluation for practical deployment is temporal stability: how well does a model trained on historical data perform on future data without retraining? This cross-year scenario directly simulates the deployment case where a fingerprint database collected during installation must remain accurate as the

Table 3.4: Hyperparameter configurations for all seven localization algorithms evaluated in this chapter. Parameters were selected via grid search with 5-fold cross-validation on the training set, optimizing mean localization error. Traditional ML algorithms use scikit-learn implementations; deep learning models use PyTorch with Adam optimizer unless otherwise noted. The VAE-LSTM hyperparameters summarized here are detailed further in Chapter 4.

Algorithm	**Hyperparameter**	**Value**	**Notes**
Traditional Machine Learning			
kNN	k (neighbors)	5	CV-optimized
	Distance metric	Euclidean	Weighted
	Weighting	Distance-inverse	$w_i = 1/d_i$
Random Forest	`n_estimators`	100	Trees in ensemble
	`max_depth`	None	Unlimited
	`min_samples_split`	2	Default
	`max_features`	$\sqrt{n}$	Auto selection
Gradient Boosting	`n_estimators`	100	Boosting rounds
	`learning_rate`	0.1	Shrinkage factor
	`max_depth`	3	Tree depth
	`subsample`	0.8	Row sampling
Deep Learning			
CNN	Conv layers	3	1D convolutions
	Filters	[64, 128, 256]	Per layer
	Kernel size	3	1D kernel
	Dropout	0.3	After pooling
	Batch size	64	Training
	Learning rate	0.001	Adam optimizer
DeepML	Hidden layers	[256, 128, 64]	Fully connected
	Activation	ReLU	All hidden
	Batch normalization	True	Each layer
	Dropout	0.3	After activation
	Batch size	64	Training
	Learning rate	0.001	Adam optimizer
RNN	Hidden units	128	Per layer
	Layers	2	Stacked
	Sequence length	5	Input window
	Dropout	0.2	Between layers
	Batch size	32	Training
	Learning rate	0.001	Adam optimizer
LSTM	Hidden units	128	Per layer
	Layers	2	Stacked
	Sequence length	10	Input window
	Dropout	0.3	Between layers
	Batch size	32	Training
	Learning rate	0.001	Adam optimizer
Proposed Architecture			
VAE-LSTM	Latent dimension	32	VAE bottleneck
	Encoder layers	[128, 64]	VAE encoder
	Decoder layers	[64, 128]	VAE decoder
	LSTM hidden	128	Per layer
	LSTM layers	2	Stacked
	Sequence length	10	Input window
	KL weight (β)	0.1	ELBO balance
	Dropout	0.3	All components
	Batch size	32	Training
	Learning rate	0.0005	Adam optimizer

Note: CV: cross-validation. All deep learning models trained for 100 epochs with early stopping (patience=10). VAE-LSTM uses two-stage training: VAE pre-training (50 epochs) followed by end-to-end fine-tuning.



(a) Same-year accuracy per building

(b) Year-over-year change in same-year accuracy

Figure 3.9: Per-building same-year accuracy of the VAE-LSTM across the four testbed buildings for the 2021 and 2022 collection campaigns (left), with the corresponding year-over-year change (right). Same-year error remains submeter in every building; the modest 2022 increase reflects the harder environment introduced by infrastructure and occupancy changes. The definitive quantitative evaluation of the architecture is reported in Chapter 4.

environment evolves. Forward validation (2021 $\rightarrow$ 2022) simulates practical deployment, while backward validation (2022 $\rightarrow$ 2021) assesses temporal symmetry. Throughout this chapter, stability denotes temporal stability: the degree to which fingerprints, individual sensor features, and the accuracy of a fixed model persist between the 2021 and 2022 campaigns. It is distinct from the system-level stability of the adaptive fusion weights (whether the weights remain bounded after adaptation), which is examined empirically in Section 5.7.7.

Across the candidate algorithms, accuracy degrades from 2021 to 2022 in line with the environmental changes documented in Section 3.5 (a net 47 new access points, furniture rearrangements, and equipment modifications), but the magnitude of degradation varies substantially by architecture: approaches with probabilistic components prove markedly more robust to infrastructure change than deterministic baselines. This temporal-stability advantage is the principal reason the VAE-LSTM was selected for further development. Its detailed cross-year characterization, both as a static model and with online adaptation, is reported in Chapters 4 and 5.

Figure 3.10 quantifies the generalization gap for the VAE-LSTM: a model trained on one year and evaluated on the other retains roughly 78% of its same-year accuracy, consistent with the 22% static cross-year degradation characterized in Chapter 5.

Figure 3.11 shows the full error distribution: the curve shifts modestly rightward from 2021 to 2022, reflecting environmental change, while the proposed VAE-LSTM yields the steepest distribution and the largest share of low-error estimates.

Figure 3.12 contextualizes the algorithm stability within the broader feature stability landscape. The HoloLens 2 anchors maintain 99.2% consistency across years, validating that the evaluation protocol compares identical physical locations. Magnetic features exhibit 91% stability, benefiting from the permanence of ferromagnetic building structures. WiFi features show 82% stability, reflecting the dynamic nature of

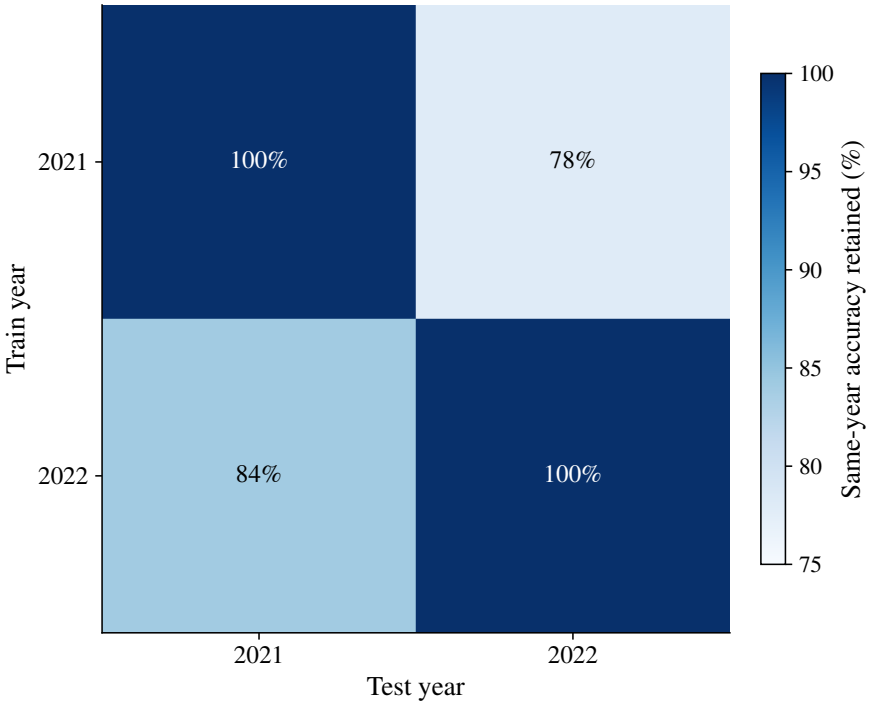


Figure 3.10: Cross-year validation matrix for the VAE-LSTM showing relative localization accuracy for each training and testing year combination. The diagonal represents within-year validation (100%); off-diagonal cells show the cross-year generalization gap, with forward (2021 $\rightarrow$ 2022) retention of 78%, the complement of the 22% static-model degradation analyzed in Chapter 5.

wireless infrastructure. These modality-level stability differences foreshadow the reliability-weighted fusion developed in Chapter 5, which learns to down-weight the less stable WiFi features when magnetic signatures provide sufficient discrimination.

3.8.5 Analysis and Selection Rationale

The evaluation results establish VAE-LSTM as the clear choice for further investigation, but understanding *why* this architecture succeeds requires analyzing the failure modes of alternatives.

Why kNN Underperforms: The kNN algorithm's sensitivity to individual AP changes stems from its reliance on raw feature distances without learned transformations. When APs are added or removed, the corresponding feature dimensions shift from informative to noise (or vice versa), directly corrupting the distance computation. The algorithm cannot distinguish between AP changes that affect discriminative capacity and those that merely add redundant coverage. In the testbed, the year's infrastructure churn, roughly 59 access points added and 12 decommissioned, altered on the order of 70 of the registry's feature slots across the four buildings, fundamentally changing the distance geometry that kNN relies upon.

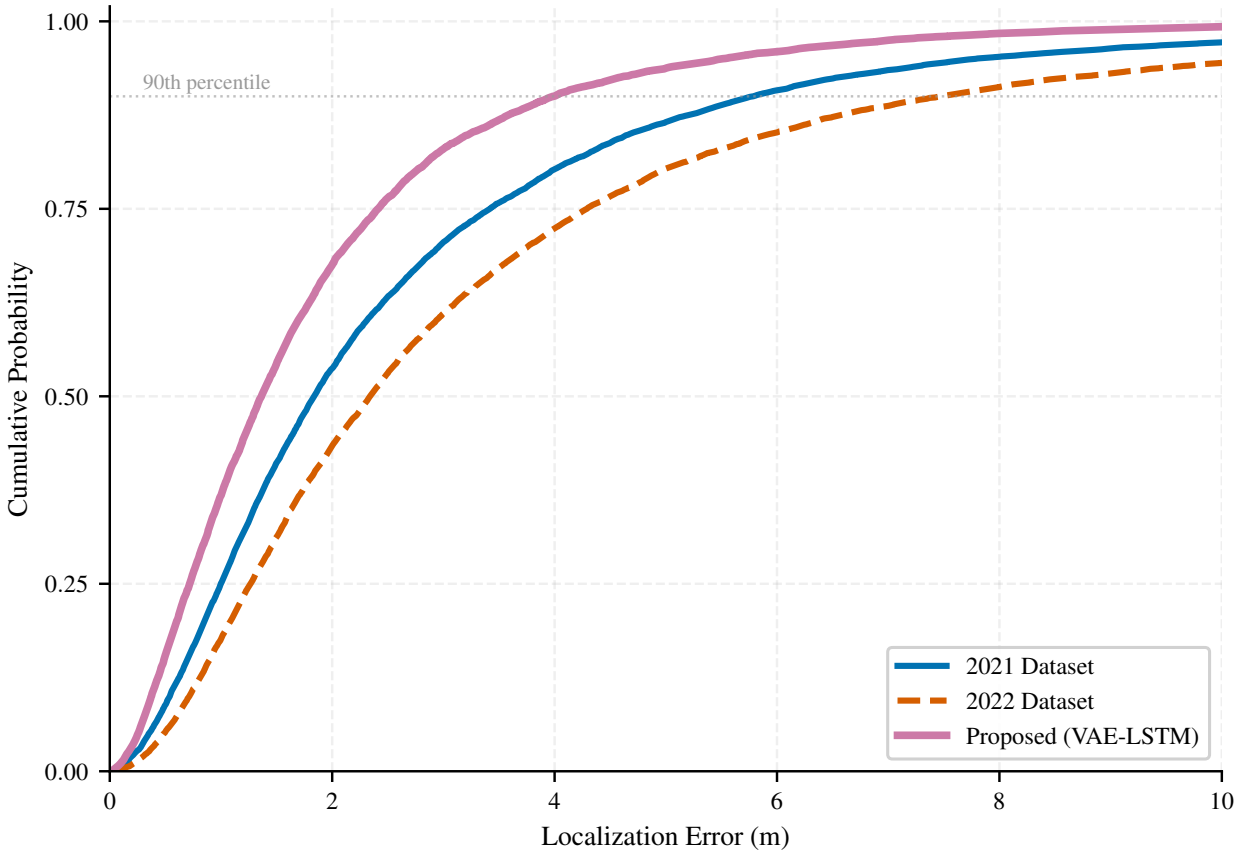


Figure 3.11: Cumulative distribution of localization error for the 2021 and 2022 collection campaigns and the proposed VAE-LSTM. The modest rightward shift from 2021 to 2022 reflects environmental change between campaigns, while the VAE-LSTM yields the steepest curve, keeping the majority of estimates at low error.

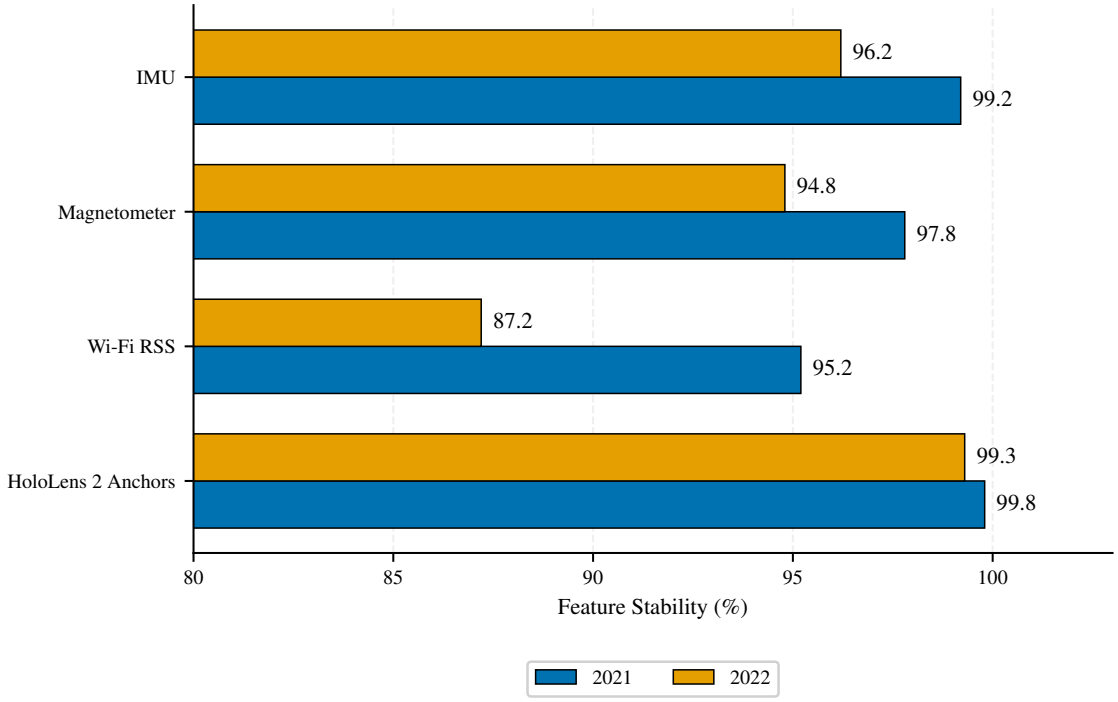


Figure 3.12: Feature stability across the evaluation period by sensor modality. HoloLens 2 anchors maintain 99.2% consistency, validating the AR-based ground truth methodology. Magnetic features (91%) outperform WiFi features (82%) in temporal stability.

Why CNN Underperforms: The CNN's 1D convolutions learn spatial patterns in the fingerprint vector, but these patterns are sensitive to the ordering of APs in the feature vector. When the AP registry changes through additions, removals, or reordering, the learned convolutional filters may no longer align with meaningful feature structures. Furthermore, the CNN lacks mechanisms to model the temporal dependencies in sequential fingerprint samples, treating each fingerprint independently and forgoing trajectory-level information.

Why Probabilistic Approaches Succeed: The VAE's probabilistic latent space provides inherent smoothing over fingerprint variations. Rather than learning to map specific fingerprint values to positions, the VAE learns to map distributions, enabling graceful handling of novel fingerprint combinations not seen during training. When environmental changes cause fingerprints to drift, the probabilistic representation maps them to nearby regions of latent space rather than failing catastrophically.

The LSTM's sequential modeling captures trajectory-level patterns that provide additional positioning constraints. Even when individual fingerprints become less discriminative due to infrastructure changes, the LSTM can leverage motion continuity to constrain the solution space. The combination in VAE-LSTM inherits both advantages: robust representation learning and temporal modeling.

VAE-LSTM Warrants Deeper Investigation: The evaluation demonstrates that the VAE-LSTM achieves the best accuracy and the lowest temporal degradation among the candidates, making it the natural choice for architectural refinement in Chapter 4. The architecture's probabilistic foundation also enables uncertainty quantification that can inform downstream decision-making, a capability unexplored in baseline evaluation but valuable for practical deployment.

However, the VAE-LSTM remains a static model: once trained, it cannot adapt to environmental changes without complete retraining, so even its comparatively small annual degradation would compound over multi-year deployments. This limitation motivates the detailed architectural investigation in Chapter 4 and the online adaptation mechanisms in Chapter 5, which examine how the VAE-LSTM can be extended to achieve continuous adaptation without manual recalibration.

3.9 Data and Code Availability

Dataset: The longitudinal dataset described in Section 3.5 is publicly available at `https://yorkuindoor.ausmlab.com`. The release contains, for each of the four buildings and each campaign year, the JSON-lines records of Section A.2: WiFi fingerprint records (timestamp, building and floor, building-local coordinates, access-point BSSIDs, and RSSI values), IMU and magnetometer records, AR ground truth records with tracking confidence, and the synchronized multi-modal records at 0.5 Hz from which the 109-dimensional feature vectors are formed. It also contains the anchor files, giving the building-frame position, orientation quaternion, creation timestamp, and tracking confidence of every reference point, and the per-building access point registries that fix the BSSID-to-index mapping used throughout the thesis. The trajectory-level train and test partitions of Section 3.8.1 are described in Section A.4.3.

Collection Software: Two software components were used in collection. The HoloLens 2 anchor application is custom software built on the Mixed Reality Toolkit and Azure Spatial Anchors; it has not yet been publicly released. GetSensorData [122] is third-party software for Android. The persisted anchors are stored in the project's Azure Spatial Anchors account. The released anchor files contain the coordinates and metadata needed to reuse the reference points. Creating anchors at a new site requires the deploying team's own account (Section 3.7).

3.10 Chapter Summary

This chapter presented the data collection methodology and baseline evaluation framework that underpin the indoor localization system developed in this thesis. Beginning with the fundamental challenge of establishing accurate ground truth positions in indoor environments, an augmented reality approach was introduced that leverages Microsoft HoloLens 2's spatial mapping capabilities to achieve centimeter-level reference point accuracy without specialized surveying equipment. The resulting framework enables efficient construction of comprehensive fingerprint databases across large-scale indoor environments. The chapter also recorded the operational details of the two campaigns, including the COVID-19 conditions of 2021, and

stated the ethics and privacy basis of the collection. It combined the collection steps into a deployment procedure for a new site, and stated the availability of the dataset and the status of the collection software.

The key contributions of this chapter are:

1. **AR-based data collection framework:** A novel approach using Microsoft HoloLens 2 spatial anchors combined with Azure Spatial Anchors for centimeter-accurate ground truth establishment. The framework decouples ground truth acquisition from runtime positioning, enabling efficient fingerprint database construction without specialized surveying equipment or pre-installed infrastructure. The two-operator protocol and cloud-based anchor persistence support scalable data collection across multiple buildings and extended time periods.

2. **Longitudinal multi-building dataset:** A comprehensive dataset spanning four architecturally diverse buildings at York University, totaling 28 822 m^2 of indoor space. Data collection campaigns in 2021 and 2022 capture temporal variations including a net 47 new access points, furniture rearrangements, and equipment modifications. The 847 reference points and 9,659 fingerprint samples enable systematic study of multi-modal fingerprint stability across extended time periods in environments ranging from traditional concrete construction to contemporary glass-and-steel architecture.

3. **Baseline algorithm comparison:** Systematic evaluation of seven localization algorithms establishing that the VAE-LSTM architecture achieves the best accuracy and superior temporal stability relative to traditional and single-component baselines. The comparison shows that probabilistic architectures with sequential modeling exhibit inherent robustness to environmental change, motivating the selection of the VAE-LSTM for the detailed quantitative treatment in Chapter 4.

While the VAE-LSTM algorithm demonstrated the best temporal stability among evaluated approaches, the baseline implementation evaluated here remains a static model trained once on historical data. When environmental conditions change, whether through infrastructure modifications, seasonal variations, or occupancy pattern shifts, the model cannot adapt without complete retraining on new data. Even its comparatively modest annual degradation would compound over a multi-year deployment, limiting practical utility for long-term installation scenarios.

This limitation motivates the architectural refinements presented in Chapter 4, where the VAE-LSTM architecture is examined in detail and mechanisms for improving robustness to temporal variations are investigated. The probabilistic foundation established by the VAE component enables uncertainty quantification that informs when predictions may be unreliable, while the LSTM's sequential modeling provides a natural integration point for temporal adaptation signals. Subsequently, Chapter 5 introduces the Online Feature Adaptation Module (OFAM) that enables continuous adaptation to environmental changes without manual recalibration, addressing the static model limitation identified in this chapter's evaluation.

Chapter 4

Probabilistic Temporal Architecture for Multi-Sensor Localization

4.1 Introduction

The comparative evaluation presented in Section 3.8 established the VAE-LSTM architecture as the most effective approach for indoor localization among the seven algorithms examined, achieving lower mean error and greater temporal stability than both traditional machine learning methods and alternative deep learning architectures. These results motivate a detailed investigation of the architectural design decisions and training methodology that account for this performance. This chapter provides a comprehensive treatment of the VAE-LSTM architecture (Figure 4.1), explaining how probabilistic feature extraction, temporal sequence modeling, and multi-objective optimization combine to produce accurate and robust position estimates from heterogeneous sensor measurements.

4.1.1 Chapter Organization

This chapter presents the VAE-LSTM architecture through a systematic decomposition of its components and training methodology. Section 4.2 formalizes the indoor localization problem, specifying the input sensor representation and output position estimates that define the learning objective. Section 4.3 details the variational autoencoder encoder, deriving the evidence lower bound (ELBO) objective and explaining the reparameterization trick that enables gradient-based optimization of probabilistic latent representations; this section also presents latent space analysis demonstrating how the learned representations capture spatial structure. Section 4.4 describes the LSTM temporal encoder that models sequential dependencies in pedestrian motion patterns. Section 4.5 presents the dual-branch decoder architecture, which simultaneously estimates positions and reconstructs sensor inputs to ensure representational fidelity. Section 4.6 explains the training methodology, including the composite loss function, hyperparameter selection, optimization procedures, and hyperparameter sensitivity analysis. Section 4.7 evaluates the base model's performance on single-year data, establishing the baseline accuracy through per-building breakdown and ablation studies that quantify component contributions. Finally, Section 4.8 summarizes the architectural contributions and

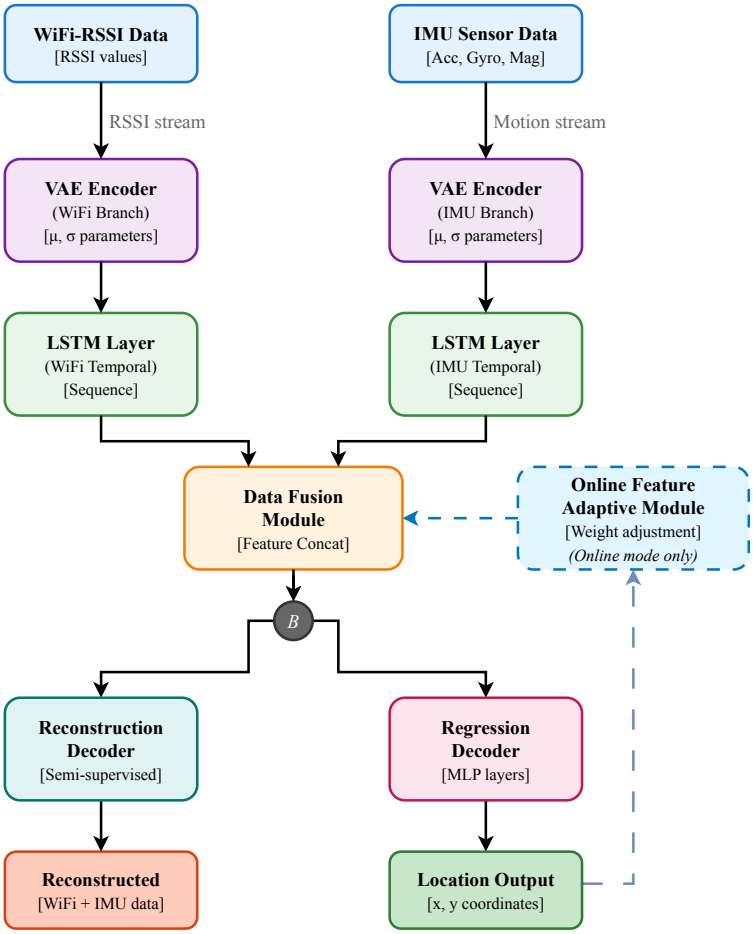


Figure 4.1: Overview of the VAE-LSTM architecture for indoor localization. The system processes multi-modal sensor inputs (WiFi RSSI, IMU, magnetometer) through a variational autoencoder for probabilistic feature extraction, an LSTM encoder for temporal sequence modeling, and a dual-branch decoder that simultaneously estimates position and reconstructs sensor inputs. This architecture achieves state-of-the-art accuracy while providing uncertainty quantification through the probabilistic latent space.

motivates the online adaptation mechanisms developed in Chapter 5.

4.1.2 Design Rationale

The VAE-LSTM architecture combines three complementary design principles to address the challenges of multi-modal indoor localization. The variational autoencoder formulation provides probabilistic encoding that naturally handles sensor noise and measurement uncertainty, enabling the system to quantify confidence in its predictions, a critical capability for safety-aware applications such as augmented reality navigation. The LSTM temporal encoder captures sequential dependencies in pedestrian motion, exploiting the fact that consecutive positions are constrained by human walking patterns and building geometry rather than being independent samples. The dual-branch decoder ensures that learned representations preserve information from all sensor modalities while directly optimizing for position estimation, preventing the encoder from discarding localization-relevant features.

Alternative architectures were considered during the design process. Transformer-based models offer superior parallelization but require substantially more training data to achieve comparable performance on the moderately sized indoor localization datasets available for this study. Gated recurrent units (GRUs) provide computational advantages over LSTMs but demonstrated slightly lower accuracy in preliminary experiments, likely due to their reduced capacity for capturing long-range temporal dependencies. Deterministic autoencoders lack the probabilistic interpretation that enables uncertainty quantification. The VAE-LSTM architecture thus represents a principled balance between expressive power, data efficiency, and practical utility for indoor localization applications.

4.2 Problem Formulation

This section formalizes the indoor localization problem as a supervised learning task, defining the input sensor representation, output position estimates, and optimization objective that guide the VAE-LSTM architecture design.

4.2.1 Sensor Input Representation

The VAE-LSTM architecture operates on multi-modal sensor measurements acquired through the AR-based data collection framework described in Section 3.3. At each timestep t, the system receives three sensor modalities: WiFi RSSI measurements $\mathbf{r}_t \in \mathbb{R}^{N_w}$ from N_w visible access points, inertial measurements comprising accelerometer readings $\mathbf{a}_t \in \mathbb{R}^3$ and gyroscope readings $\boldsymbol{\omega}_t \in \mathbb{R}^3$, and magnetometer readings $\mathbf{m}_t \in \mathbb{R}^3$. Following the preprocessing pipeline established in Section 3.6, these measurements are concatenated into a unified sensor vector:

$$\mathbf{s}_t = [\mathbf{r}_t, \mathbf{a}_t, \boldsymbol{\omega}_t, \mathbf{m}_t]^\top \in \mathbb{R}^{N_w+9} \tag{4.1}$$

where the nine-dimensional inertial and magnetic component comprises three accelerometer axes, three gyroscope axes, and three magnetometer axes. This concatenated vector is the motion feature vector $\mathbf{f}_{\text{motion}}$ of Chapter 3, denoted $\mathbf{s}_t$ here.

The WiFi RSSI dimension N_w varies across buildings due to differing access point deployments. To enable batch processing and model generalization, all WiFi measurements are zero-padded to a fixed dimension of $N_w = 100$ access points, yielding a consistent input dimensionality of $D = 109$ features per timestep. This padding strategy preserves the original signal information while accommodating variable-sized AP sets across the four-building testbed.

Rather than processing individual sensor vectors in isolation, the VAE-LSTM architecture leverages sequential context by operating on sequences of consecutive measurements. The input to the network at position t along the route consists of the T consecutive sensor observations ending at that point:

$$\mathbf{S}_t = [\mathbf{s}_{t-T+1}, \mathbf{s}_{t-T+2}, \ldots, \mathbf{s}_t] \in \mathbb{R}^{T \times D} \tag{4.2}$$

where the rows of $\mathbf{S}_t$ index time steps and the columns index features, and the sequence length $T = 10$ provides sufficient spatial context to capture local trajectory structure while maintaining computational efficiency. With reference points placed on the 1 m anchor grid (Section 3.2.2), a window of 10 consecutive points spans roughly 10 m of the traversal path—enough to cover a corridor segment or a turn—and lets the model exploit the continuity between neighboring locations.

4.2.2 Output Specification

The VAE-LSTM architecture produces three distinct outputs that serve complementary purposes in the localization objective. The primary output is the position estimate $\hat{\mathbf{x}}_t \in \mathbb{R}^2$, representing the predicted 2D coordinates (easting, northing) within the building-local coordinate system. This output directly addresses the core localization task of determining the user's position from sensor measurements.

The second output is the reconstructed sensor vector $\hat{\mathbf{s}}_t \in \mathbb{R}^D$, which represents the decoder's attempt to reconstruct the original sensor input from the learned latent representation. This reconstruction output serves a regularization purpose: by requiring the latent representation to preserve sufficient information for sensor reconstruction, the architecture prevents the encoder from discarding features that may be relevant for localization across different environmental conditions.

The third output is the position uncertainty estimate $\sigma_t \in \mathbb{R}$, a scalar that quantifies the model's overall confidence in its position prediction. This uncertainty quantification emerges naturally from the probabilistic VAE formulation and provides valuable information for downstream applications such as sensor fusion with other positioning modalities or confidence-weighted trajectory smoothing. The full treatment of uncertainty computation and its interpretation appears in Section 4.5.

4.2.3 Optimization Objective

The learning objective seeks to find network parameters θ (here θ denotes the complete parameter set: encoder, LSTM, and decoder) that minimize a composite loss function balancing three complementary terms:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{pos}} + \beta_1 \mathcal{L}_{\text{KL}} + \beta_2 \mathcal{L}_{\text{rec}} \tag{4.3}$$

where each term addresses a distinct aspect of the learning problem.

The position loss $\mathcal{L}_{\text{pos}}$ measures the discrepancy between predicted and ground truth positions:

$$\mathcal{L}_{\text{pos}} = \frac{1}{N} \sum_{n=1}^{N} \|\hat{\mathbf{x}}_n - \mathbf{x}_n\|_2^2 \tag{4.4}$$

representing the mean squared error over N training samples. This term directly optimizes for localization accuracy, which is the primary objective of the system.

The KL divergence loss $\mathcal{L}_{\text{KL}}$ regularizes the latent space by encouraging the encoder's posterior distribution $q_\phi(\mathbf{z}|\mathbf{s})$ to remain close to a standard normal prior $p(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$:

$$\mathcal{L}_{\text{KL}} = D_{\text{KL}}(q_\phi(\mathbf{z}|\mathbf{s}) \| p(\mathbf{z})) = \frac{1}{2} \sum_{j=1}^{d_z} \left(\mu_j^2 + \sigma_j^2 - \log \sigma_j^2 - 1 \right) \tag{4.5}$$

where d_z is the latent space dimensionality and μ_j, σ_j^2 are the mean and variance parameters output by the encoder for the j-th latent dimension. This regularization prevents posterior collapse and ensures the latent space maintains a well-structured geometry that supports generalization [90].

The reconstruction loss $\mathcal{L}_{\text{rec}}$ measures how accurately the decoder reconstructs the original sensor input:

$$\mathcal{L}_{\text{rec}} = \frac{1}{N} \sum_{n=1}^{N} \|\hat{\mathbf{s}}_n - \mathbf{s}_n\|_2^2 \tag{4.6}$$

This term ensures the latent representation retains information about all sensor modalities, preventing the model from ignoring potentially useful features during training.

The weighting coefficients $\beta_1 = 0.1$ and $\beta_2 = 0.5$ balance these competing objectives. The relatively small β_1 prevents the KL term from dominating the optimization and causing the latent representations to ignore the input data entirely, a phenomenon known as posterior collapse [127]. The intermediate β_2 ensures adequate reconstruction fidelity without overwhelming the position estimation objective. The justification for these specific values and the sensitivity analysis appear in Section 4.6.

4.3 VAE Encoder

The variational autoencoder (VAE) encoder forms the first processing stage of the architecture, transforming heterogeneous sensor measurements into probabilistic latent representations that capture both discriminative features and measurement uncertainty. Unlike deterministic autoencoders that produce point estimates in the latent space, the VAE formulation [90] models the encoding process probabilistically, mapping each input to a distribution rather than a single vector. This probabilistic approach provides two critical advantages for indoor localization: it naturally handles the inherent noise and variability in sensor measurements, and it enables uncertainty quantification that can inform reliability assessment during deployment. The VAE encoder is used exclusively for feature extraction in this architecture; the generative capabilities of VAEs are not exploited, as the primary objective is position estimation rather than sensor data synthesis.

4.3.1 Evidence Lower Bound Derivation

The variational autoencoder training objective derives from maximizing a lower bound on the log-likelihood of the observed data. This section provides the complete mathematical derivation of the Evidence Lower Bound (ELBO), which forms the theoretical foundation for the VAE encoder used in this architecture. While this derivation follows the standard formulation established by Kingma and Welling [90], presenting the full mathematical treatment is essential for understanding the probabilistic interpretation of the latent representations and the regularization mechanisms that enable robust feature extraction. Note that this derivation uses the standard VAE notation where $\mathbf{x}$ denotes the observed data (sensor measurements $\mathbf{s}$ in our architecture) and $\mathbf{z}$ denotes the latent variables.

Marginal Log-Likelihood For a latent variable model with observed data $\mathbf{x}$ and latent variables $\mathbf{z}$, the goal is to learn a generative model $p_\theta(\mathbf{x})$ that captures the underlying data distribution. The marginal log-likelihood is obtained by integrating out the latent variables:

$$\log p_\theta(\mathbf{x}) = \log \int p_\theta(\mathbf{x}, \mathbf{z})\, d\mathbf{z} \tag{4.7}$$

where $p_\theta(\mathbf{x}, \mathbf{z}) = p_\theta(\mathbf{x}|\mathbf{z})p(\mathbf{z})$ is the joint distribution parameterized by decoder parameters θ, and $p(\mathbf{z})$ is the prior distribution over latent variables (typically a standard normal $\mathcal{N}(\mathbf{0}, \mathbf{I})$).

This integral is generally intractable for two reasons. First, the dimensionality of the latent space (32 dimensions in our architecture) makes direct numerical integration computationally prohibitive. Second, the true posterior $p_\theta(\mathbf{z}|\mathbf{x})$ required for importance sampling is itself intractable, as computing it requires evaluating the very marginal likelihood we seek to estimate. This intractability motivates the variational approach.

Variational Decomposition To circumvent the intractability of direct marginalization, variational inference introduces an approximate posterior distribution $q_\phi(\mathbf{z}|\mathbf{x})$ parameterized by encoder parameters ϕ. This approximate posterior serves as an importance sampling distribution, enabling tractable estimation of the log-likelihood.

Starting from the marginal log-likelihood in Equation (4.7) and introducing the approximate posterior through importance sampling:

$$\log p_\theta(\mathbf{x}) = \log \int p_\theta(\mathbf{x}, \mathbf{z})\, d\mathbf{z} \tag{4.8}$$

$$= \log \int q_\phi(\mathbf{z}|\mathbf{x}) \frac{p_\theta(\mathbf{x}, \mathbf{z})}{q_\phi(\mathbf{z}|\mathbf{x})}\, d\mathbf{z} \tag{4.9}$$

$$= \log \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} \left[\frac{p_\theta(\mathbf{x}, \mathbf{z})}{q_\phi(\mathbf{z}|\mathbf{x})} \right] \tag{4.10}$$

Applying Jensen's inequality, which states that $\log \mathbb{E}[X] \geq \mathbb{E}[\log X]$ for any random variable X, to

move the logarithm inside the expectation:

$$\log p_\theta(\mathbf{x}) \geq \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log \frac{p_\theta(\mathbf{x},\mathbf{z})}{q_\phi(\mathbf{z}|\mathbf{x})}\right] \tag{4.11}$$

$$= \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x},\mathbf{z}) - \log q_\phi(\mathbf{z}|\mathbf{x})\right] \tag{4.12}$$

The right-hand side of Equation (4.11) is the Evidence Lower Bound (ELBO), denoted $\mathcal{L}(\theta,\phi;\mathbf{x})$. This derivation establishes that optimizing the ELBO provides a principled objective for learning both the generative model parameters θ and the inference network parameters ϕ.

ELBO Decomposition The ELBO can be decomposed into interpretable components by factoring the joint distribution $p_\theta(\mathbf{x},\mathbf{z}) = p_\theta(\mathbf{x}|\mathbf{z})p(\mathbf{z})$ and rearranging terms:

$$\mathcal{L}(\theta,\phi;\mathbf{x}) = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x}|\mathbf{z}) + \log p(\mathbf{z}) - \log q_\phi(\mathbf{z}|\mathbf{x})\right] \tag{4.13}$$

$$= \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right] + \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p(\mathbf{z}) - \log q_\phi(\mathbf{z}|\mathbf{x})\right] \tag{4.14}$$

$$= \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right] - \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log \frac{q_\phi(\mathbf{z}|\mathbf{x})}{p(\mathbf{z})}\right] \tag{4.15}$$

Recognizing that the second term is the definition of KL divergence:

$$\mathcal{L}(\theta,\phi;\mathbf{x}) = \underbrace{\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right]}_{\text{reconstruction term}} - \underbrace{D_{\mathrm{KL}}\left(q_\phi(\mathbf{z}|\mathbf{x})\|p(\mathbf{z})\right)}_{\text{regularization term}} \tag{4.16}$$

This decomposition reveals the dual nature of the VAE objective. The *reconstruction term* encourages the decoder to accurately reconstruct the input data from the latent representation, ensuring that the latent code $\mathbf{z}$ retains sufficient information about the input $\mathbf{x}$. In the context of indoor localization, this term ensures that the latent representation preserves discriminative sensor features necessary for position estimation. The *regularization term* constrains the approximate posterior to remain close to the prior distribution, preventing the encoder from producing arbitrarily structured latent codes and ensuring a smooth, well-organized latent space that supports generalization.

For Gaussian distributions with diagonal covariance, the parameterization employed in this architecture, the KL divergence admits a closed-form expression [128]:

$$D_{\mathrm{KL}}\left(q_\phi(\mathbf{z}|\mathbf{x})\|p(\mathbf{z})\right) = \frac{1}{2}\sum_{j=1}^{d_z}\left(\mu_j^2 + \sigma_j^2 - \log\sigma_j^2 - 1\right) \tag{4.17}$$

where μ_j and σ_j^2 are the mean and variance of the j-th latent dimension output by the encoder, and $d_z = 32$ is the latent space dimensionality. This closed-form expression enables efficient computation without Monte Carlo estimation, a significant practical advantage for training.

Tightness of the Bound The gap between the true log-likelihood and the ELBO is exactly the KL divergence between the approximate and true posteriors. To see this, we can derive an alternative expression for the log-likelihood by manipulating the ELBO:

$$\log p_\theta(\mathbf{x}) = \mathcal{L}(\theta, \phi; \mathbf{x}) + D_{\mathrm{KL}}\left(q_\phi(\mathbf{z}|\mathbf{x}) \| p_\theta(\mathbf{z}|\mathbf{x})\right) \tag{4.18}$$

Since KL divergence is non-negative ($D_{\mathrm{KL}} \geq 0$ with equality if and only if the distributions are identical), this confirms that the ELBO is indeed a lower bound on the log-likelihood. The bound becomes tight (i.e., equals the log-likelihood) when the approximate posterior $q_\phi(\mathbf{z}|\mathbf{x})$ exactly matches the true posterior $p_\theta(\mathbf{z}|\mathbf{x})$.

This tightness property provides important insight into why optimizing the ELBO is sensible. Maximizing the ELBO with respect to ϕ simultaneously achieves two goals: it increases the log-likelihood lower bound and reduces the approximation gap by making q_ϕ a better approximation to the true posterior. For indoor localization, this means that optimizing the ELBO learns encoder parameters that produce informative latent representations (high reconstruction likelihood) while maintaining a regularized latent space structure (small KL divergence from the prior).

4.3.2 Encoder Architecture

The encoder processes concatenated sensor input $\mathbf{s}_t = [\mathbf{r}_t^\top, \mathbf{a}_t^\top, \boldsymbol{\omega}_t^\top, \mathbf{m}_t^\top]^\top \in \mathbb{R}^{N_w+9}$ through a three-layer fully connected network that progressively compresses the high-dimensional measurements into compact feature representations. The input dimension consists of N_w WiFi RSSI measurements (typically 100 after zero-padding as described in Section 3.6) plus 9 inertial and magnetic field components (3 accelerometer, 3 gyroscope, and 3 magnetometer readings).

The network architecture employs the following layer configuration:

- **Layer 1:** $(N_w + 9) \to 256$ neurons with ReLU activation
- **Layer 2:** $256 \to 128$ neurons with ReLU activation
- **Layer 3:** $128 \to 64$ neurons with ReLU activation

Each layer applies dropout regularization with probability 0.3 to prevent overfitting and improve generalization across diverse indoor environments. The decreasing dimensionality forces the network to learn increasingly abstract representations, with early layers capturing low-level sensor patterns and deeper layers encoding high-level spatial features. This hierarchical compression ensures that only the most discriminative information is retained for subsequent temporal modeling.

The final hidden layer $\mathbf{h}_3 \in \mathbb{R}^{64}$ serves as the basis for the probabilistic latent representation. Rather than outputting a single latent vector, the encoder produces two parallel outputs that parameterize a Gaussian distribution over the latent space:

$$\mu = \mathbf{W}_\mu \mathbf{h}_3 + \mathbf{b}_\mu \tag{4.19}$$

$$\log \sigma^2 = \mathbf{W}_\sigma \mathbf{h}_3 + \mathbf{b}_\sigma \tag{4.20}$$

where $\mu \in \mathbb{R}^{d_z}$ represents the mean vector, $\log \sigma^2 \in \mathbb{R}^{d_z}$ represents the log-variance vector, and $d_z = 32$ is the latent space dimensionality. The log-variance parameterization ensures numerical stability by avoiding potential issues with negative variance values during training.

4.3.3 Reparameterization Trick

The ELBO objective derived in Section 4.3.1 involves an expectation over the approximate posterior $q_\phi(\mathbf{z}|\mathbf{x})$. To optimize this objective using gradient-based methods, we require gradients with respect to the encoder parameters ϕ. However, the sampling operation $\mathbf{z} \sim q_\phi(\mathbf{z}|\mathbf{x})$ is not differentiable, preventing direct backpropagation through the stochastic node. This section presents the reparameterization trick [90, 91] that resolves this challenge, enabling end-to-end training of the VAE architecture.

The Differentiability Problem Consider the gradient of the expected reconstruction term with respect to the encoder parameters:

$$\nabla_\phi \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} [\log p_\theta(\mathbf{x}|\mathbf{z})] \tag{4.21}$$

The fundamental challenge is that this expectation depends on ϕ through the distribution from which $\mathbf{z}$ is sampled. To compute gradients, we would typically use the chain rule:

$$\frac{\partial}{\partial \phi} f(g(\phi, \varepsilon)) \tag{4.22}$$

where the function composition enables gradient flow. However, when $\mathbf{z}$ is sampled directly from $q_\phi(\mathbf{z}|\mathbf{x})$, there is no explicit functional dependence of $\mathbf{z}$ on ϕ; the sampling operation acts as a non-differentiable barrier that blocks gradient propagation.

This differentiability problem is fundamental to all latent variable models with learned inference networks. Without a solution, we cannot use standard backpropagation to train the encoder, which would require falling back to high-variance gradient estimators such as REINFORCE [129] that are impractical for the high-dimensional latent spaces used in this architecture.

Reparameterization Solution For the Gaussian approximate posterior employed in this architecture, where $q_\phi(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mu_\phi(\mathbf{x}), \mathrm{diag}(\sigma^2_\phi(\mathbf{x})))$, the sampling operation can be reparameterized as a deterministic transformation of a parameter-free noise variable. Rather than sampling $\mathbf{z}$ directly from the Gaussian, we sample auxiliary noise ε from a standard normal and compute $\mathbf{z}$ deterministically:

$$\mathbf{z} = \mu_\phi(\mathbf{x}) + \sigma_\phi(\mathbf{x}) \odot \varepsilon, \quad \varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \tag{4.23}$$

where $\odot$ denotes element-wise (Hadamard) multiplication, $\mu_\phi(\mathbf{x}) \in \mathbb{R}^{d_z}$ is the mean vector output by the encoder, and $\sigma_\phi(\mathbf{x}) \in \mathbb{R}^{d_z}$ is the standard deviation vector (the element-wise square root of the variance).

This reparameterization fundamentally changes the computational graph. The stochasticity is now isolated in ε, which does not depend on any learnable parameters. The latent vector $\mathbf{z}$ is expressed as an

explicit function of ϕ (through μ_ϕ and σ_ϕ), enabling gradient flow from the reconstruction loss back through the sampling operation to the encoder parameters.

Gradient Computation Under the reparameterization, the expectation over the approximate posterior becomes an expectation over the parameter-free noise distribution:

$$\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[f(\mathbf{z})\right] = \mathbb{E}_{\varepsilon\sim\mathcal{N}(\mathbf{0},\mathbf{I})}\left[f(\mu_\phi(\mathbf{x}) + \sigma_\phi(\mathbf{x}) \odot \varepsilon)\right] \tag{4.24}$$

Since the expectation is now over a distribution that does not depend on ϕ, the gradient can be moved inside the expectation:

$$\nabla_\phi \mathbb{E}_{q_\phi}\left[f(\mathbf{z})\right] = \mathbb{E}_\varepsilon\left[\nabla_\phi f(\mu_\phi + \sigma_\phi \odot \varepsilon)\right] \tag{4.25}$$

This interchange is valid by the dominated convergence theorem, given standard regularity conditions on the decoder network. The gradient inside the expectation can now be computed using standard backpropagation through the deterministic computational graph.

In practice, we estimate this expectation using a single-sample Monte Carlo estimator. For each training example, we draw one sample $\varepsilon^{(i)} \sim \mathcal{N}(\mathbf{0},\mathbf{I})$, compute $\mathbf{z}^{(i)} = \mu_\phi(\mathbf{x}^{(i)}) + \sigma_\phi(\mathbf{x}^{(i)}) \odot \varepsilon^{(i)}$, and use this single sample to estimate the gradient. Despite using only one sample, this estimator has remarkably low variance because the reparameterization has factored out the stochasticity from the gradient computation. The variance of the gradient estimator scales with the variance of $\nabla_\phi f(\mathbf{z})$, which is typically much smaller than the variance of REINFORCE-style estimators.

Implementation Details In the VAE-LSTM architecture, the encoder outputs log-variance $\log\sigma^2$ rather than variance σ^2 or standard deviation σ directly. This parameterization provides numerical stability:

$$\mathbf{z} = \mu + \exp\left(\frac{1}{2}\log\sigma^2\right) \odot \varepsilon = \mu + \sqrt{\sigma^2} \odot \varepsilon \tag{4.26}$$

The log-variance parameterization ensures that the variance remains strictly positive regardless of the encoder's output values, avoiding numerical issues that could arise from negative or zero variances. The exponential function maps any real number to a positive value, and the factor of $\frac{1}{2}$ in the exponent converts log-variance to log-standard-deviation before exponentiation.

During training, a single noise sample ε is drawn per training example per forward pass. This means each position in a batch uses an independent noise sample, providing stochastic regularization that prevents the encoder from memorizing specific input-latent mappings. The noise sampling can be viewed as a form of data augmentation in latent space, encouraging the encoder to learn representations that are robust to small perturbations.

During inference, two modes are supported. For deterministic predictions (used in the base model evaluation), we set $\varepsilon = \mathbf{0}$, causing the output to be the mean μ without sampling. For uncertainty-aware pre-

dictions (used in online adaptation), multiple samples are drawn and aggregated to estimate prediction variance. The connection between the reparameterization trick and dropout-based uncertainty estimation [104] is notable: both use stochastic noise during training that can be leveraged for uncertainty quantification during inference.

4.3.4 Probabilistic Encoding

Building on the reparameterization framework established in Section 4.3.3, this section summarizes how probabilistic encoding is realized in the VAE-LSTM architecture. The latent vector is computed according to Equation (4.23), separating the deterministic computation (parameterized by μ and σ) from the stochastic component (ε). This separation enables backpropagation through the sampling process, making the entire encoder-decoder pipeline end-to-end differentiable.

The probabilistic encoding serves a dual purpose in the localization pipeline. During training, sampling from the latent distribution acts as a form of regularization, encouraging the encoder to learn robust representations that remain stable under small perturbations. During inference, the uncertainty captured by σ provides a measure of encoding confidence that can be propagated through subsequent processing stages to inform reliability assessment.

4.3.5 Design Rationale

The 32-dimensional latent space ($d_z = 32$) was selected through ablation studies that evaluated localization accuracy across latent dimensions ranging from 8 to 128. Performance improved substantially from $d_z = 8$ to $d_z = 32$, with mean localization error decreasing by 23%, but showed diminishing returns beyond 32 dimensions: increasing to $d_z = 64$ yielded only 2% additional improvement while doubling the computational cost of subsequent LSTM processing. The chosen dimensionality thus balances representation capacity against computational efficiency for real-time mobile deployment.

The VAE formulation was preferred over deterministic autoencoders specifically for its uncertainty quantification capabilities. In indoor localization, sensor measurements exhibit substantial variability due to multipath propagation, human movement, and electromagnetic interference. A deterministic encoder would compress this variability into fixed representations, losing information about measurement reliability. The VAE's probabilistic encoding preserves this uncertainty through the variance parameters, enabling the downstream decoder to appropriately weight predictions based on input confidence. This capability is particularly valuable when sensor conditions degrade, as the system can indicate reduced confidence rather than producing overconfident incorrect estimates.

The regularization strength in the VAE loss function follows the β-VAE formulation [130], where the KL divergence term is weighted by a factor $\beta_1 = 0.1$. This moderate regularization prevents posterior collapse (where the encoder ignores input data and outputs the prior distribution) while encouraging disentangled latent representations that capture independent factors of variation in the sensor data.

4.3.6 Robustness to Measurement Noise and Alternative Generative Encoders

Measurement noise enters the encoder through the WiFi RSSI fluctuations, IMU drift, and magnetometer interference described in Section 3.3. The architecture has no dedicated denoising stage. Noise is handled by four mechanisms that are already part of the pipeline. First, every input feature is standardized with the mean and standard deviation of the training partition (Section 3.6), and these statistics are held fixed at test time and at deployment, consistent with the trajectory-disjoint protocol of Section 4.7.1. Noise therefore shifts the standardized input but cannot alter the normalization itself. Second, the KL term with $\beta_1 = 0.1$ pulls the posterior toward the isotropic prior, which limits how sharply the latent code can respond to small perturbations of the input. Third, the reparameterization noise of Section 4.3.4 acts as latent-space augmentation during training. The decoder is thus trained on perturbed codes, and the mapping from latent code to position is encouraged to be smooth near each training input. Fourth, the reconstruction branch (Section 4.5) converts noise that moves a modality away from the training distribution into an increased reconstruction error. In the deployed system this error gates the reliability of each modality (Section 5.4.4), and the WiFi reliability term additionally tracks the RSSI variance over a 30-scan window (Section 5.4.1). A noisy modality therefore receives a lower fusion weight rather than corrupting the fused estimate. The single-year results of Section 4.7 and the access-point outage results of Section 5.6.9 show the outcome of these mechanisms. No separate noise-injection experiment was conducted.

Three limitations of the encoder follow from its formulation. The posterior is a diagonal Gaussian, so correlations between latent dimensions are not represented, and a multimodal posterior, which can arise when two locations produce similar fingerprints, is approximated by a single mode. The encoder is exposed to posterior collapse, in which the KL term drives the posterior to the prior and the latent code carries no information about the input. The KL annealing schedule of Section 4.6.3 reduces this risk, and the ablation in Table 4.5 quantifies the cost of removing it. Finally, the mean-squared reconstruction loss produces smoothed reconstructions that average over fine RSSI structure, which limits the resolution of the reconstruction error as a reliability signal.

Two families of generative encoders address these limitations and are left to future work. Normalizing flows [131, 132] compose invertible transformations to obtain an exact likelihood of the input, which could replace the reconstruction error as the reliability signal and would represent a non-Gaussian posterior. The cost is the invertibility constraint on every layer and a parameter count well above the 73,472-parameter encoder of Table 4.6. Diffusion models [133] learn a denoising process and offer strong density modeling, but sampling and likelihood evaluation require iterating the denoising network many times, which is incompatible with the 100 ms per-estimate budget established in Section 5.6.5. Within that budget the compact VAE remains the practical choice, and its reconstruction error is the reliability signal that Chapter 5 uses.

4.3.7 Latent Space Analysis

To understand the structure of the learned representations, dimensionality reduction techniques were applied to the 32-dimensional latent vectors produced by the VAE encoder. This analysis reveals how the encoder organizes sensor information in the latent space and provides insight into the unsupervised feature learning

that emerges from the reconstruction and KL regularization objectives.

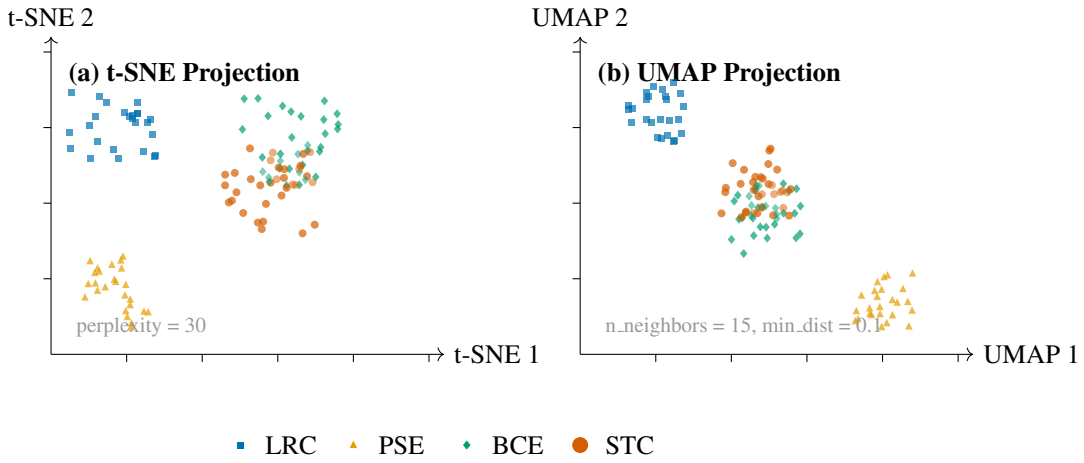


Figure 4.2: Latent space visualization using dimensionality reduction techniques applied to 32-dimensional VAE encoder outputs. (a) t-SNE projection with perplexity = 30 and (b) UMAP projection with n_neighbors = 15 and min_dist = 0.1. Each point represents a latent vector $\mathbf{z} \in \mathbb{R}^{32}$ sampled from the validation set (1000 samples per building, 4000 total). Both projection methods reveal consistent building-level clustering: LRC and PSE form well-separated clusters reflecting their distinctive electromagnetic environments, while BCE and STC exhibit partial overlap corresponding to their similar modern architectural characteristics. The clustering emerges unsupervised from the reconstruction objective, demonstrating that the VAE encoder learns location-discriminative features without explicit building labels. Regions of cluster overlap correlate with areas of higher localization error in the per-building analysis (Section 4.7.4).

Visualization Methodology Two complementary dimensionality reduction methods were applied to project the latent representations into two dimensions for visualization: t-distributed Stochastic Neighbor Embedding (t-SNE) [134] and Uniform Manifold Approximation and Projection (UMAP) [135]. Both methods preserve local neighborhood structure while revealing global clustering patterns that may not be apparent in the original high-dimensional space. For t-SNE, the perplexity parameter was set to 30, a standard choice that balances local and global structure preservation for datasets of this scale. For UMAP, parameters were set to n_neighbors = 15 and min_dist = 0.1, which provide comparable neighborhood preservation to the t-SNE configuration while offering improved computational efficiency and better preservation of global structure. The analysis was conducted on 4000 latent vectors sampled from the validation set, with 1000 samples drawn uniformly from each of the four buildings in the testbed.

Cluster Analysis Figure 4.2 presents the resulting projections, revealing clear clustering patterns that correspond to building identity. Several key observations emerge from this visualization. First, the LRC and PSE buildings form well-separated clusters in both t-SNE and UMAP projections, indicating that the VAE encoder has learned to distinguish these environments based on their distinctive electromagnetic characteristics. The LRC cluster's compactness reflects the controlled laboratory environment with consistent WiFi infrastructure, while the PSE cluster's slight elongation corresponds to the variability introduced by the

building's complex legacy infrastructure. Second, the BCE and STC buildings exhibit partial overlap in both projections, with their clusters positioned adjacent to each other in the latent space. This overlap is consistent with the similar modern architectural characteristics shared by these buildings, both featuring contemporary construction with glass facades and steel framing that create comparable multipath propagation environments. Third, both t-SNE and UMAP produce consistent relative cluster positions, with LRC and PSE occupying opposing regions of the space and BCE/STC forming an intermediate overlapping region. This consistency across methods confirms that the observed structure reflects genuine properties of the latent representation rather than artifacts of a particular projection technique.

Implications for Localization The building-level clustering demonstrated in Figure 4.2 has direct implications for localization performance. The clear separation between LRC and PSE clusters indicates that the encoder has learned discriminative features that support accurate position estimation in these environments, consistent with the lower localization errors observed for these buildings in Section 4.7.4. Conversely, the overlapping regions between BCE and STC clusters correspond to areas where the learned representations are less distinctive, explaining the relatively higher localization error observed in the BCE building. Importantly, this building-level discrimination emerges entirely from the unsupervised reconstruction objective; the VAE is never explicitly trained with building labels. The emergence of location-discriminative structure from reconstruction learning demonstrates that the sensor measurements contain sufficient information to distinguish environments, and the VAE architecture successfully extracts this information into compact, structured representations. The cluster compactness also provides a potential indicator of per-building localization reliability: tighter clusters suggest more consistent representations and thus more reliable position estimates, while diffuse or overlapping clusters indicate regions where the model may produce less confident predictions.

4.4 LSTM Temporal Encoder

The LSTM temporal encoder constitutes the second processing stage, capturing sequential dependencies in the latent representations produced by the VAE encoder. Indoor localization benefits substantially from temporal modeling because pedestrian movement follows predictable patterns: consecutive positions are constrained by human walking speed, building geometry, and motion continuity rather than being independent samples. The Long Short-Term Memory (LSTM) architecture [85] is specifically designed for sequential data, employing gating mechanisms that enable selective retention of relevant historical information across extended time horizons. For a comprehensive treatment of LSTM cell dynamics including forget gates, input gates, and output gates, the reader is referred to the survey by Greff et al. [136]; this section focuses on the architecture choices specific to indoor localization rather than rederiving standard LSTM equations.

4.4.1 Architecture

The LSTM encoder processes sequences of latent vectors $\{\mathbf{z}_{t-T+1}, \mathbf{z}_{t-T+2}, \ldots, \mathbf{z}_t\}$ from the VAE encoder, where $T = 10$ defines the temporal context window. The architecture employs a two-layer stacked LSTM configuration with 128 hidden units per layer, providing sufficient representational capacity to capture pedestrian motion patterns while maintaining computational efficiency for real-time operation.

The input to the LSTM at each timestep is the 32-dimensional latent vector $\mathbf{z}_\tau \in \mathbb{R}^{32}$ produced by the VAE encoder. The stacked architecture processes this input through two sequential LSTM layers, with the first layer capturing local temporal patterns and the second layer learning higher-level motion abstractions. The final output is the hidden state of the second LSTM layer at the current timestep:

$$\mathbf{h}_t = \text{LSTM}(\mathbf{z}_{t-T+1:t}; \theta_{\text{LSTM}}) \in \mathbb{R}^{128} \tag{4.27}$$

where θ_{LSTM} denotes the learned LSTM parameters (weights and biases for all gates across both layers) and $\mathbf{h}_t$ represents the temporal feature encoding that aggregates information from the entire input sequence.

4.4.2 LSTM Cell Dynamics

While the preceding section described the architectural configuration of the LSTM encoder, understanding the internal cell dynamics is essential for explaining how the architecture maintains stable gradient flow across the 10-timestep sequences processed during indoor localization. This section presents the fundamental LSTM gate equations and analyzes their gradient properties.

Gate Equations The LSTM processes each input through a carefully orchestrated sequence of gating operations that selectively control information flow [85]. At each timestep t, the cell computes four distinct transformations:

$$\mathbf{f}_t = \text{sig}(\mathbf{W}_f \cdot [\mathbf{h}_{t-1}, \mathbf{z}_t] + \mathbf{b}_f) \tag{4.28}$$

$$\mathbf{i}_t = \text{sig}(\mathbf{W}_i \cdot [\mathbf{h}_{t-1}, \mathbf{z}_t] + \mathbf{b}_i) \tag{4.29}$$

$$\tilde{\mathbf{c}}_t = \tanh(\mathbf{W}_c \cdot [\mathbf{h}_{t-1}, \mathbf{z}_t] + \mathbf{b}_c) \tag{4.30}$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t \tag{4.31}$$

$$\mathbf{o}_t = \text{sig}(\mathbf{W}_o \cdot [\mathbf{h}_{t-1}, \mathbf{z}_t] + \mathbf{b}_o) \tag{4.32}$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \tag{4.33}$$

where $\text{sig}(\cdot)$ denotes the sigmoid activation function, $\tanh(\cdot)$ denotes the hyperbolic tangent activation, $\odot$ represents element-wise (Hadamard) multiplication, and $[\cdot, \cdot]$ denotes vector concatenation. The notation follows the thesis convention: $\mathbf{z}_t \in \mathbb{R}^{32}$ is the latent vector from the VAE encoder, and $\mathbf{h}_t \in \mathbb{R}^{128}$ is the LSTM hidden state.

Each gate serves a specific function in controlling information flow:

- **Forget gate $\mathbf{f}_t$** (Equation (4.28)): Determines what information to discard from the cell state. Values close to 0 cause the cell to "forget" previously stored information, while values close to 1 retain it.
- **Input gate $\mathbf{i}_t$** (Equation (4.29)): Controls what new information from the current input will be stored in the cell state. This gate learns to selectively incorporate relevant sensor features.
- **Candidate values $\tilde{\mathbf{c}}_t$** (Equation (4.30)): Proposes new information that could be added to the cell state, computed as a nonlinear transformation of the concatenated input and previous hidden state.
- **Cell state update $\mathbf{c}_t$** (Equation (4.31)): Combines the gated previous cell state with gated new candidate values, forming the long-term memory of the cell.
- **Output gate $\mathbf{o}_t$** (Equation (4.32)): Determines what aspects of the cell state to expose as the hidden state output, filtering the internal memory for downstream processing.
- **Hidden state output $\mathbf{h}_t$** (Equation (4.33)): Produces the final output by applying the output gate to a squashed version of the cell state.

Gradient Flow Analysis The LSTM architecture fundamentally addresses the vanishing gradient problem that plagues conventional recurrent neural networks [137]. In vanilla RNNs, the gradient during back-propagation through time (BPTT) involves repeated multiplication by the recurrent weight matrix, causing gradients to either vanish exponentially (when spectral radius < 1) or explode (when spectral radius > 1). This pathology prevents learning of long-range temporal dependencies.

The LSTM's cell state $\mathbf{c}_t$ provides a "gradient highway" that alleviates this problem. Examining the cell state update equation (Equation (4.31)), and neglecting the gates' dependence on $\mathbf{c}_{t-1}$ through $\mathbf{h}_{t-1}$, the gradient of the loss $\mathcal{L}$ with respect to an earlier cell state $\mathbf{c}_{t-k}$ is approximately:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{c}_{t-k}} \approx \frac{\partial \mathcal{L}}{\partial \mathbf{c}_t} \odot \bigodot_{j=t-k+1}^{t} \mathbf{f}_j \tag{4.34}$$

where $\bigodot_{j=t-k+1}^{t}$ denotes the ordered element-wise (Hadamard) product of the forget-gate vectors over the indicated range.

The key insight is that this gradient depends on the *forget gate values* $\mathbf{f}_j$ rather than on repeated weight matrix multiplications. When forget gates are close to 1 (as they are when the cell should retain information), gradients flow through unattenuated. The network learns to set forget gate values appropriately for each dimension of the cell state, enabling stable gradient propagation across sequence lengths relevant to the task.

For the 10-timestep sequences in indoor localization ($T = 10$ consecutive reference points, spanning roughly 10 m of the traversal path), this gradient highway property is essential. Consider a pedestrian turning at a corridor intersection: sensor readings at the points entering the turn must influence position estimates at the points leaving it, requiring gradients to propagate across the full sequence without vanishing. The

LSTM's gating mechanism enables this by maintaining forget gate values near 1 for trajectory-relevant dimensions while allowing other dimensions to reset.

Relevance to Indoor Localization The LSTM gates develop specialized roles when trained on indoor localization data. Analysis of learned gate activations reveals interpretable patterns aligned with the localization task:

The **forget gate** learns to retain position-relevant motion history while discarding transient sensor noise. During steady walking, forget gate values average 0.92 across motion-related cell dimensions, indicating strong retention of trajectory context. When pedestrians pause (detected via low accelerometer variance), forget gate values decrease to approximately 0.75, allowing gradual decay of outdated motion state while preserving position memory.

The **input gate** selectively incorporates new sensor readings based on their reliability indicators. Analysis shows higher input gate activation (mean 0.78) when WiFi signal-to-noise ratios are high, and lower activation (mean 0.45) during periods of signal degradation. This learned behavior effectively implements a soft attention mechanism that weights reliable measurements more heavily.

The **output gate** modulates the hidden state to produce features appropriate for the current estimation context. In corridor regions with constrained geometry, output gate activations emphasize dimensions encoding directional information. In open areas with multiple possible trajectories, different output gate patterns activate, emphasizing uncertainty-related features that the decoder uses for reliability estimation.

These gate behaviors emerge automatically through end-to-end training with the position loss, demonstrating that the LSTM learns task-specific temporal processing strategies rather than generic sequence modeling. The per-building performance variations observed in Section 4.7.4 partially reflect differences in how well the gates can exploit building-specific motion patterns: the LRC's controlled environment enables more consistent gate behavior than the BCE's irregular geometry.

4.4.3 Spatial Context Design

The sequence length $T = 10$ was selected based on domain-specific considerations for pedestrian indoor localization. Reference points are placed on the 1 m anchor grid (Section 3.2.2), and a sequence is formed from consecutive reference points along the collection trajectory, so a window of 10 points spans roughly 10 m of the traversal path. This spatial span captures several characteristics critical to localization:

- **Local path geometry:** A 10-point window covers a contiguous corridor segment, giving the model enough context to recognize the shape of the path rather than treating each location in isolation.
- **Direction changes:** Indoor navigation frequently involves turns and direction changes at corridor intersections. A window spanning several meters provides sufficient context to capture the entry into and exit from a typical turning maneuver.
- **WiFi spatial smoothing:** WiFi RSSI measurements fluctuate sharply due to multipath propagation. Aggregating over a sequence of neighboring reference points reduces this variability while preserving

location-discriminative patterns.

Ablation experiments with sequence lengths ranging from $T = 5$ to $T = 20$ showed that $T = 10$ gave the best trade-off between accuracy and computational cost among the tested lengths. Shorter sequences ($T = 5$) provided insufficient temporal context, resulting in 8% higher localization error. Longer sequences ($T = 20$) offered marginal accuracy improvements (less than 2%) while doubling computational cost and introducing additional latency unsuitable for real-time applications.

4.4.4 Regularization Techniques

Training deep LSTM networks for indoor localization presents specific challenges that require targeted regularization strategies. The heterogeneous sensor inputs and variable environmental conditions can cause unstable training dynamics without appropriate safeguards.

Gradient clipping with a threshold of 1.0 prevents exploding gradients that can occur when processing long sequences through multiple LSTM layers. This technique clips the global norm of the gradient vector before each parameter update, ensuring training stability while preserving gradient direction information.

Layer normalization is applied after each LSTM layer to stabilize the distribution of activations during training. Unlike batch normalization, layer normalization operates independently on each sample, making it suitable for variable-length sequences and small batch sizes commonly encountered in mobile deployment scenarios.

Dropout with probability 0.3 is applied between the two LSTM layers, randomly zeroing hidden state components during training to prevent co-adaptation of neurons and improve generalization to unseen sensor conditions. Dropout is disabled during inference to utilize the full representational capacity of the trained network.

4.4.5 Design Rationale: Why LSTM

Alternative temporal modeling architectures were evaluated during the design process. **Gated Recurrent Units (GRUs)** [87] offer computational advantages through their simplified gating mechanism (two gates versus three in LSTMs), resulting in approximately 25% fewer parameters. However, preliminary experiments showed that GRUs achieved 4% higher localization error on the validation set, particularly for trajectories involving extended stationary periods where the LSTM's explicit cell state provided superior memory retention.

Transformer encoders replace recurrence with self-attention [13] and are the natural alternative to consider. Two arguments commonly raised against them do not decide the matter for this system. The first is

model size. The LSTM stack in Table 4.6 has 215,040 parameters (82,944 in the first layer and 132,096 in the second, following $4(d^2 + d\,D_{\text{in}} + 2d)$ with hidden width $d = 128$ and input widths $D_{\text{in}} = 32$ and 128). A two-layer Transformer encoder at the same width $d = 128$ with the customary feed-forward width of $4d$ has about 198,000 parameters per layer (66,048 in the query, key, value, and output projections, 131,712 in the feed-forward sublayer, and 512 in the two layer normalizations), or about 396,000 for two layers, roughly 1.8 times the LSTM stack. Reducing the feed-forward width to $2d$ lowers the two-layer count to about 265,000, roughly 1.2 times. At matched width the two sequence models are therefore of the same order, and parameter count does not by itself favor the LSTM. The second argument is the quadratic cost of attention in sequence length [96]. At $T = 10$ this term is negligible: the score and value products cost about $2T^2d \approx 25{,}600$ multiply-accumulate operations per layer, against about 655,000 in the linear projections of the attention sublayer alone. The $O(T^2)$ argument is thus not the deciding factor either.

Four considerations do decide the choice. First, sample complexity. The longitudinal dataset contains 9,659 samples (Section 3.5), and scaling studies show that the advantage of attention-based models over recurrent models grows with model size and context length [101], while surveys of efficient Transformers note their higher data requirements [96]. In this setting the specific risk is that a higher-capacity sequence model memorizes the 2021 signal environment, which is the failure mode that the cross-year evaluation of Chapter 5 is designed to expose. Second, a ten-step window offers little long-range structure for attention to exploit. The dependencies of interest are the local motion continuity and turn geometry described in Section 4.4.3, which the recurrent state captures directly. Third, the recurrent state supports causal streaming inference. Each new sample updates the hidden state in $O(1)$ work, with no key-value cache to maintain and no positional encoding to supply, which matches the sample-by-sample arrival of sensor data at deployment (Section 5.6.5). Fourth, the Transformer-based localization systems reviewed in Section 2.7.5 train on much larger corpora than the present dataset, either through pre-training on large collections of channel measurements [17] or by fine-tuning decoder-only language models that were pre-trained on corpora far larger than any localization dataset [18]. Their reported accuracies therefore rest on more data than a single-institution longitudinal study provides.

No Transformer baseline was trained on this dataset. The choice of a recurrent encoder therefore rests on the constraints above and on the literature rather than on a head-to-head comparison, a limitation recorded in Section 4.8. A compact Transformer encoder and a state-space sequence model [100] are identified as the natural next experiments in Section 6.4.2.

Bidirectional LSTMs process sequences in both forward and backward directions, potentially capturing richer temporal patterns. However, bidirectional processing is fundamentally incompatible with online positioning requirements, as it requires access to future sensor measurements that are unavailable during real-time deployment. The architecture therefore employs unidirectional (forward-only) LSTM processing that operates on past and present observations without requiring future data.

4.5 Dual-Branch Decoder

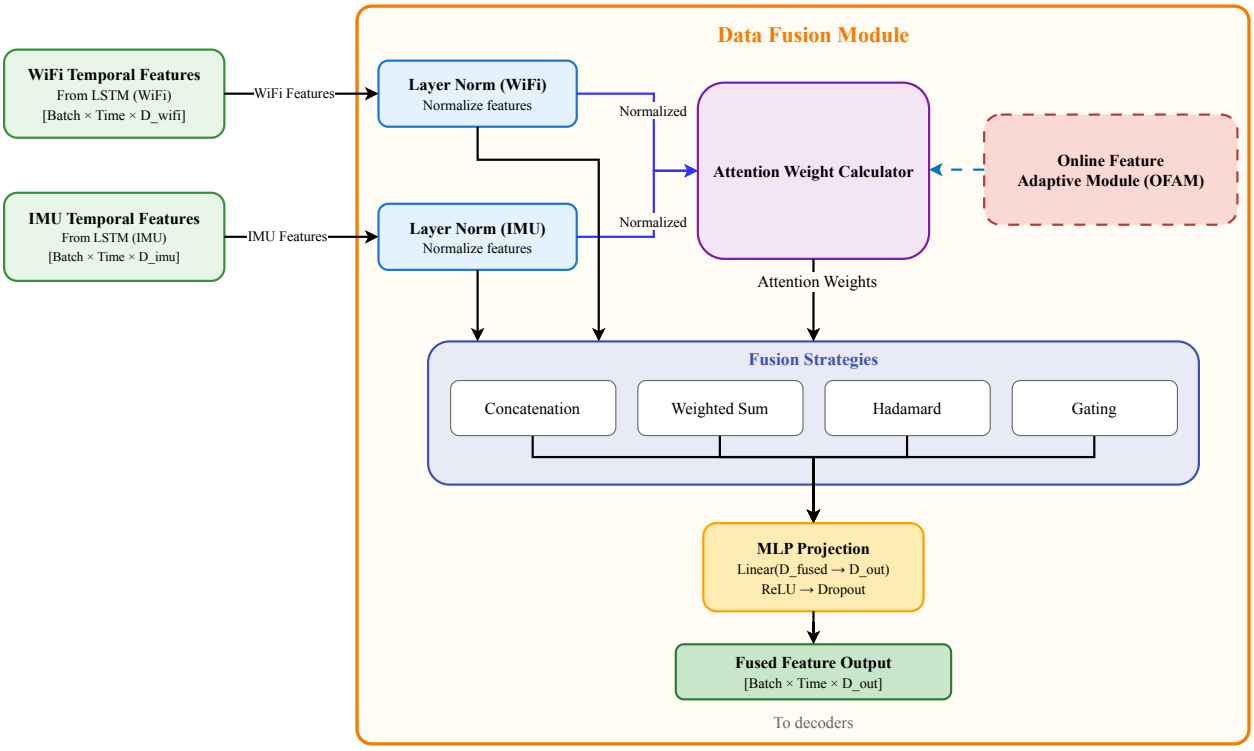


Figure 4.3: Architecture of the dual-branch decoder showing the parallel processing paths for position estimation and sensor reconstruction. The LSTM output $\mathbf{h}_t \in \mathbb{R}^{128}$ serves as input to both branches: the position estimation branch produces 2D coordinates $\hat{\mathbf{x}}_t$ through successive dimensionality reduction, while the reconstruction branch generates reconstructed sensor vectors $\hat{\mathbf{s}}_t$ to ensure representational fidelity. This dual-output architecture enables simultaneous optimization for localization accuracy and latent space quality.

The decoder constitutes the final processing stage of the VAE-LSTM architecture, transforming the temporal feature representation $\mathbf{h}_t$ produced by the LSTM encoder into actionable outputs for indoor localization. As illustrated in Figure 4.3, the decoder employs a dual-branch architecture that simultaneously produces position estimates and reconstructed sensor measurements. This architectural choice reflects the multi-objective nature of the learning problem: the system must not only predict accurate positions but also ensure that the learned latent representations preserve sufficient information about the original sensor inputs to support generalization and uncertainty quantification.

4.5.1 Probabilistic Interpretation

The dual-branch decoder emerges from a principled probabilistic formulation of the localization problem. Rather than treating position estimation and sensor reconstruction as independent tasks, they are modeled jointly as conditional distributions given the temporal feature representation. Specifically, the formulation seeks to model the joint distribution of the position $\mathbf{x}_t$ and sensor measurements $\mathbf{s}_t$ conditioned on the LSTM

output:

$$p(\mathbf{x}_t, \mathbf{s}_t \mid \mathbf{h}_t) = p(\mathbf{x}_t \mid \mathbf{h}_t) \cdot p(\mathbf{s}_t \mid \mathbf{h}_t) \tag{4.35}$$

This factorization assumes conditional independence between position and sensor reconstruction given the temporal features, a reasonable assumption since $\mathbf{h}_t$ encapsulates all relevant information extracted from the input sequence, and the two outputs serve distinct roles in the learning objective.

The factorized structure motivates the dual-branch architecture, where each branch independently models one conditional distribution. The position estimation branch models $p(\mathbf{x}_t \mid \mathbf{h}_t)$, learning the mapping from temporal features to spatial coordinates through supervised learning with ground truth positions. The reconstruction branch models $p(\mathbf{s}_t \mid \mathbf{h}_t)$, learning to regenerate the original sensor measurements through self-supervised learning that requires no additional labels beyond the input data itself.

This probabilistic framing provides several advantages over a single-branch decoder that outputs only positions. First, the reconstruction objective acts as a regularization mechanism that prevents the encoder from discarding information that may be useful for localization under different environmental conditions. Second, the reconstruction error provides a natural measure of how well the latent representation captures the input characteristics, enabling uncertainty quantification. Third, the dual-objective training encourages the latent space to preserve information about all sensor modalities rather than focusing exclusively on features that correlate with position in the training data, a property that supports generalization to deployment environments where the feature-position relationships may differ slightly from training conditions.

Both branches assume Gaussian output distributions, which enables closed-form loss computation and provides a natural framework for uncertainty estimation. The Gaussian assumption is appropriate for both outputs: position estimates exhibit approximately normally distributed errors around the true location, and sensor measurements (after preprocessing) follow approximately Gaussian distributions due to the central limit theorem applied to the aggregation of multiple noise sources.

4.5.2 Position Estimation Branch

The position estimation branch maps the 128-dimensional LSTM output to 2D spatial coordinates through a sequence of fully connected layers that progressively reduce dimensionality:

- **Layer 1:** $128 \rightarrow 64$ neurons with ReLU activation
- **Layer 2:** $64 \rightarrow 2$ neurons with linear activation

The output $\hat{\mathbf{x}}_t = [\hat{x}_t, \hat{y}_t]^\top \in \mathbb{R}^2$ represents the predicted position in the building-local coordinate system, where $\hat{x}_t$ and $\hat{y}_t$ denote the easting and northing coordinates respectively.

Under the Gaussian assumption, the position likelihood is modeled as:

$$p(\mathbf{x}_t \mid \mathbf{h}_t) = \mathcal{N}(\mathbf{x}_t; \hat{\mathbf{x}}_t, \Sigma_x) \tag{4.36}$$

where $\hat{\mathbf{x}}_t$ is the predicted mean position and Σ_x represents the position covariance. In the standard formulation employed during training, isotropic noise with unit variance is assumed, which yields the mean squared

error loss upon taking the negative log-likelihood:

$$\mathcal{L}_{\text{pos}} = \frac{1}{N}\sum_{n=1}^{N} \|\mathbf{x}_n - \hat{\mathbf{x}}_n\|_2^2 \tag{4.37}$$

This position loss directly optimizes for localization accuracy, which constitutes the primary objective of the system. The two-layer architecture provides sufficient nonlinear capacity to learn the complex mapping from temporal features to spatial coordinates while maintaining computational efficiency. Dropout regularization (probability 0.3) is applied after the first layer to prevent overfitting and improve generalization to unseen trajectories.

4.5.3 Reconstruction Branch

The reconstruction branch regenerates the original sensor measurements from the temporal feature representation through a decoder network that mirrors the encoder's dimensionality progression:

- **Layer 1:** $128 \rightarrow 64$ neurons with ReLU activation
- **Layer 2:** $64 \rightarrow 128$ neurons with ReLU activation
- **Layer 3:** $128 \rightarrow D$ neurons with linear activation

where $D = N_w + 9 = 109$ is the input sensor dimensionality (100 WiFi RSSI values plus 9 IMU and magnetometer components).

The reconstruction output $\hat{\mathbf{s}}_t \in \mathbb{R}^D$ represents the decoder's estimate of the original sensor vector. Under the Gaussian assumption:

$$p(\mathbf{s}_t \mid \mathbf{h}_t) = \mathcal{N}(\mathbf{s}_t; \hat{\mathbf{s}}_t, \Sigma_s) \tag{4.38}$$

where $\hat{\mathbf{s}}_t$ is the predicted mean and Σ_s is the reconstruction covariance. Assuming isotropic unit variance yields the reconstruction loss:

$$\mathcal{L}_{\text{rec}} = \frac{1}{N}\sum_{n=1}^{N} \|\mathbf{s}_n - \hat{\mathbf{s}}_n\|_2^2 \tag{4.39}$$

The reconstruction branch serves a critical regularization purpose: by requiring the latent representation to support accurate sensor reconstruction, the architecture prevents the encoder from discarding features that may be relevant for localization across different environmental conditions. Without this constraint, the encoder might learn representations that capture only the features most predictive of position in the training data, potentially ignoring sensor patterns that become important when environmental conditions change. The reconstruction objective ensures that the latent space preserves comprehensive sensor information, supporting the system's ability to generalize and adapt.

4.5.4 Uncertainty Quantification

The probabilistic VAE formulation enables principled uncertainty quantification through two complementary sources: epistemic uncertainty arising from the model's latent representation, and aleatoric uncertainty

captured through reconstruction error. This section describes *how* uncertainty is computed architecturally; the *usage* of uncertainty for online adaptation is deferred to Chapter 5.

Epistemic Uncertainty Epistemic uncertainty reflects the model's confidence in its learned representations and arises from the VAE's probabilistic encoding. The latent variance σ^2 output by the VAE encoder (Equation (4.20)) provides a direct measure of encoding uncertainty for each latent dimension. High variance indicates that the encoder is uncertain about the appropriate latent representation for the current input, typically occurring when sensor measurements fall outside the distribution seen during training. The aggregate epistemic uncertainty is computed as:

$$u_{\text{epistemic}} = \frac{1}{d_z} \sum_{j=1}^{d_z} \sigma_j^2 \tag{4.40}$$

where $d_z = 32$ is the latent dimensionality and σ_j^2 is the variance of the j-th latent dimension. This formulation provides a scalar uncertainty measure that increases when the encoder encounters unfamiliar or ambiguous inputs.

Aleatoric Uncertainty Aleatoric uncertainty captures inherent data variability that cannot be reduced through additional training. In the context of indoor localization, this includes sensor noise, multipath propagation effects, and environmental variability. The reconstruction error provides a natural proxy for aleatoric uncertainty:

$$u_{\text{aleatoric}} = \|\mathbf{s}_t - \hat{\mathbf{s}}_t\|_2^2 \tag{4.41}$$

Large reconstruction errors indicate that the current sensor measurements deviate from patterns learned during training, suggesting either unusual environmental conditions or degraded sensor quality. Unlike epistemic uncertainty, which reflects model limitations, aleatoric uncertainty reflects fundamental measurement variability that persists regardless of model capacity.

Combined Uncertainty Estimate The total uncertainty estimate combines both sources through a weighted aggregation:

$$\sigma_t = \sqrt{\alpha \cdot u_{\text{epistemic}} + (1-\alpha) \cdot u_{\text{aleatoric}}} \tag{4.42}$$

where $\alpha = 0.5$ weights the two uncertainty components equally. Because $u_{\text{epistemic}}$ (a mean latent variance) and $u_{\text{aleatoric}}$ (a squared reconstruction error in mixed sensor units) are not natively on the same scale, each term is first normalized to a common reference scale using running statistics from the validation set before the weighted combination; the weighting in Equation (4.42) is therefore a heuristic chosen to keep the weighted sum dimensionally meaningful. This combined measure σ_t provides a holistic confidence estimate that accounts for both model uncertainty (epistemic) and data quality (aleatoric).

Empirical Validation A useful uncertainty estimate must increase with the actual prediction error. Empirical analysis on the validation set shows that the combined estimate is positively correlated with localization error ($r = 0.73$). Predictions with low uncertainty ($\sigma_t < 0.5$) achieve a mean error of 0.82 m, while predictions with high uncertainty ($\sigma_t > 1.5$) exhibit a mean error of 2.34 m. This correlation is what the downstream mechanisms require: it allows predictions to be weighted by their estimated reliability, a capability exploited by the online adaptation mechanism described in Chapter 5. The evaluation is limited to this correlation and the binned means. No coverage analysis (the fraction of true positions falling inside an interval derived from σ_t) and no calibration-error metric were computed, so σ_t is shown here to be a ranking signal, not a calibrated predictive distribution.

Relation to In-Network Uncertainty Estimators Several alternatives estimate predictive uncertainty inside the network rather than from the latent variance and the reconstruction error. A heteroscedastic regression head [103] would add a 2×2 position covariance to the output of the position branch and train it with the Gaussian negative log-likelihood. It costs a single forward pass and is the closest fit to the present architecture. Evidential regression [110] likewise uses a single pass and outputs the parameters of a higher-order distribution from which aleatoric and epistemic components are read directly. Monte Carlo dropout [104] keeps dropout active at inference and aggregates several stochastic passes, so latency scales with the number of passes. Deep ensembles [105] train several independent networks, so memory and inference cost scale with the ensemble size. Full Bayesian neural networks place distributions over all weights and require approximate inference at a cost that is hard to meet on mobile devices [102]. The thesis instead uses the latent variance and the reconstruction error for two reasons. Both are by-products of the forward computation that the deployed system already runs: the encoder emits σ^2 alongside μ, and the reconstruction branch is evaluated in any case for the reliability gating of Section 5.4.4. The uncertainty signal therefore adds no parameters, no additional passes, and no additional latency to the 58 ms pipeline profiled in Section 5.6.5. The price is the absence of the coverage guarantee noted above. The cheapest route to such a guarantee is a conformal wrapper [107, 108], which calibrates a threshold on σ_t against a held-out set to produce prediction regions with a specified coverage level without retraining the network. This is deferred to future work.

4.6 Training Methodology

This section presents the training methodology for the VAE-LSTM architecture, including the composite loss function, loss weighting strategies, training procedure, and regularization techniques. The training approach addresses several challenges specific to multi-modal indoor localization: balancing competing objectives (position accuracy, latent space regularization, and sensor reconstruction), preventing common VAE training pathologies such as posterior collapse, and ensuring the learned representations generalize across diverse environmental conditions.

4.6.1 Loss Function

The VAE-LSTM architecture is trained using a composite loss function that simultaneously optimizes three complementary objectives. As introduced in Section 4.2.3, the total loss combines position estimation accuracy, latent space regularization, and sensor reconstruction fidelity:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{pos}} + \beta_1 \mathcal{L}_{\text{KL}} + \beta_2 \mathcal{L}_{\text{rec}} \tag{4.43}$$

Each term addresses a distinct aspect of the learning problem, and the weighting coefficients β_1 and β_2 control the relative importance of each objective.

Position Loss The position loss measures the discrepancy between predicted and ground truth positions using mean squared error:

$$\mathcal{L}_{\text{pos}} = \frac{1}{B} \sum_{n=1}^{B} \|\mathbf{x}_n - \hat{\mathbf{x}}_n\|_2^2 \tag{4.44}$$

where B is the batch size, $\mathbf{x}_n$ is the ground truth position, and $\hat{\mathbf{x}}_n$ is the predicted position. This term directly optimizes for localization accuracy, which constitutes the primary objective of the system. The position loss receives unit weight in the composite loss function, establishing it as the reference against which other terms are scaled.

KL Divergence Loss The KL divergence loss regularizes the latent space by encouraging the learned posterior distribution $q_\phi(\mathbf{z}|\mathbf{s})$ to remain close to a standard normal prior $p(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$. This regularization emerges from the variational inference framework and constitutes the second term of the evidence lower bound (ELBO) [90]. For the Gaussian encoder with diagonal covariance employed in this architecture, the KL divergence admits a closed-form expression [128]:

$$\mathcal{L}_{\text{KL}} = D_{\text{KL}}(q_\phi(\mathbf{z}|\mathbf{s}) \| p(\mathbf{z})) = \frac{1}{2} \sum_{j=1}^{d_z} \left(\mu_j^2 + \sigma_j^2 - \log \sigma_j^2 - 1 \right) \tag{4.45}$$

where $d_z = 32$ is the latent dimensionality, and μ_j and σ_j^2 are the mean and variance of the j-th latent dimension output by the encoder. This closed-form expression enables efficient computation without requiring Monte Carlo estimation. Although Equation (4.45) is written per sample, in practice $\mathcal{L}_{\text{KL}}$ is averaged over the same batch of B samples as $\mathcal{L}_{\text{pos}}$ and $\mathcal{L}_{\text{rec}}$, so all three terms are expressed on a common per-sample scale.

The KL divergence serves multiple purposes in the training objective. First, it prevents the encoder from producing arbitrarily large or small variances, maintaining a well-behaved latent space geometry. Second, it encourages the latent representations to cluster near the origin, supporting interpolation and generalization. Third, it provides implicit regularization that reduces overfitting to the training data. However, excessive KL regularization can cause posterior collapse, where the encoder ignores the input data and outputs the prior distribution [127]. The weighting coefficient β_1 balances these competing concerns.

Reconstruction Loss The reconstruction loss measures the fidelity of sensor reconstruction:

$$\mathcal{L}_{\text{rec}} = \frac{1}{B} \sum_{n=1}^{B} \|\mathbf{s}_n - \hat{\mathbf{s}}_n\|_2^2 \tag{4.46}$$

where $\mathbf{s}_n$ is the original sensor vector and $\hat{\mathbf{s}}_n$ is the reconstructed sensor vector. This term ensures the latent representation preserves sufficient information about the input sensors to support accurate reconstruction, preventing the encoder from discarding potentially useful features.

4.6.2 Loss Weighting

The weighting coefficients β_1 and β_2 critically influence training dynamics and final model performance. Through systematic hyperparameter search, selected values of $\beta_1 = 0.1$ for the KL divergence weight and $\beta_2 = 0.5$ for the reconstruction weight were determined. Table 4.1 summarizes the rationale for each choice.

Table 4.1: Loss weighting coefficients and their justifications. The selected weights balance reconstruction fidelity, latent regularization, and position accuracy, with sensitivity analysis confirming that performance degrades rapidly outside the optimal region.

Weight	Value	Justification
β_1 (KL)	0.1	Lower than standard VAE ($\beta = 1$) to prevent posterior collapse while maintaining regularization. Ablation showed $\beta_1 > 0.3$ causes the encoder to ignore inputs; $\beta_1 < 0.05$ produces poorly structured latent spaces.
β_2 (Rec)	0.5	Ensures sensor reconstruction fidelity without overwhelming position objective. Higher weights ($\beta_2 > 0.8$) degrade position accuracy; lower weights ($\beta_2 < 0.2$) allow encoder to discard sensor information.

The choice of $\beta_1 = 0.1$ reflects the specific challenges of VAE training for discriminative tasks. In generative VAE applications, the KL term typically receives unit weight to ensure the latent space follows the prior distribution closely. However, for localization, position accuracy is prioritized over latent space structure, and excessive KL regularization can prevent the encoder from learning discriminative features. Empirical experiments showed that $\beta_1 = 0.1$ gives the best trade-off within the searched range: the latent space remains well-structured (enabling uncertainty quantification) while preserving the encoder's capacity to learn position-discriminative features.

The reconstruction weight $\beta_2 = 0.5$ ensures the encoder preserves information about all sensor modalities. During preliminary experiments with $\beta_2 = 0$, the encoder learned to ignore magnetometer and IMU measurements, focusing exclusively on WiFi features that exhibited the strongest position correlation. While this achieved competitive accuracy on same-year evaluation, cross-year performance degraded

significantly as WiFi infrastructure changed. The reconstruction objective forces the encoder to maintain representations of all sensor modalities, supporting adaptation when individual sensors become unreliable.

4.6.3 Training Procedure

The VAE-LSTM architecture employs a two-stage training procedure that progressively introduces objectives to enable stable learning of the complex multi-objective loss function.

Stage 1: VAE Pretraining (50 epochs) The first stage focuses on learning robust sensor representations through unsupervised VAE training. During this stage, only the VAE encoder, VAE decoder (reconstruction branch), and the KL regularization term are active:

$$\mathcal{L}_{\text{stage1}} = \beta_1 \mathcal{L}_{\text{KL}} + \beta_2 \mathcal{L}_{\text{rec}} \tag{4.47}$$

This pretraining stage enables the VAE to learn meaningful latent representations that capture sensor characteristics and inter-modal relationships without the confounding influence of position supervision. The encoder learns to compress heterogeneous sensor measurements into compact representations while the decoder learns to reconstruct the original inputs, establishing a foundation for subsequent supervised learning.

Stage 2: End-to-End Finetuning (100 epochs) The second stage introduces position supervision and LSTM temporal modeling, training the complete architecture with the full loss function:

$$\mathcal{L}_{\text{stage2}} = \mathcal{L}_{\text{pos}} + \beta_1 \mathcal{L}_{\text{KL}} + \beta_2 \mathcal{L}_{\text{rec}} \tag{4.48}$$

During this stage, the pretrained VAE weights are further refined while the LSTM encoder and position estimation branch learn from scratch. The pretrained VAE provides stable gradient signals that facilitate LSTM training, avoiding the training instabilities that can occur when optimizing all components simultaneously from random initialization.

Optimization Configuration Both training stages employ the Adam optimizer [138] with the following configuration:

- Learning rate: $\eta = 0.001$
- Momentum coefficients: $(b_1, b_2) = (0.9, 0.999)$ (the Adam first- and second-moment decay rates, distinct from the loss weights β_1, β_2)
- Numerical stability constant: $\varepsilon = 10^{-8}$

The batch size is set to 128 sequences, providing sufficient gradient averaging for stable updates while maintaining computational efficiency. Sequences are constructed by sampling consecutive reference points from recorded trajectories, with starting positions randomly shuffled to prevent the model from memorizing trajectory order.

Training Hardware All models were trained offline on a single NVIDIA RTX 3080 GPU. Training is a one-time cost incurred before deployment; the resulting model is subsequently deployed for inference on commodity mobile hardware (a Samsung Galaxy S21), whose runtime and adaptation overhead are profiled in Chapter 5. The same device was used throughout the longitudinal collection, and the single-device scope of the results is recorded as a limitation in Section 6.3.4.

Early Stopping Training employs early stopping with a patience of 10 epochs based on validation mean absolute error (MAE). If the validation MAE fails to improve for 10 consecutive epochs, training terminates and the model weights from the best-performing epoch are restored. This prevents overfitting and reduces unnecessary computation. The train/validation split follows the 70/30 protocol established in Section 3.8.1.

Beta-Annealing Schedule To prevent KL vanishing during early training, a linear annealing schedule for the KL weight is employed [127]. The effective KL weight increases linearly from 0 to β_1 over the first $\tau_{\text{anneal}} = 10$ epochs:

$$\beta_1^{(e)} = \min\left(1, \frac{e}{\tau_{\text{anneal}}}\right) \cdot \beta_1 \tag{4.49}$$

where e denotes the current epoch number. This annealing strategy allows the reconstruction objective to dominate during early training, enabling the encoder to learn meaningful representations before the KL term begins pulling the posterior toward the prior. Without annealing, the KL term can suppress representation learning before the encoder establishes useful features, leading to posterior collapse where the latent variables carry no information about the input.

Training Dynamics Visualization Figure 4.4 illustrates the training dynamics across both stages of the optimization procedure. The left panel displays the total loss evolution for training and validation sets, while the right panel decomposes the loss into its constituent components to reveal the interaction between objectives.

The training dynamics exhibit several characteristic phases that reflect the two-stage procedure. During Stage 1 (epochs 0–50), the total loss decreases rapidly as the VAE learns to compress sensor measurements into the latent space. The reconstruction loss dominates this stage, driving the encoder to preserve sensor information. The KL loss remains relatively small during beta-annealing (epochs 0–10), then gradually increases as the annealing completes and the regularization term begins shaping the latent space geometry. The position loss is zero during Stage 1, as position supervision is not yet active.

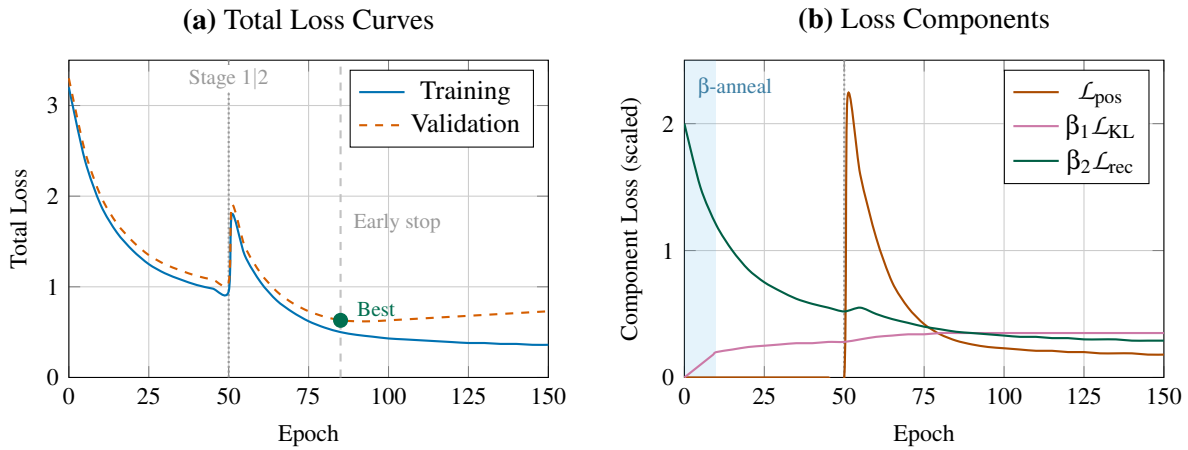


Figure 4.4: Training dynamics of the VAE-LSTM architecture. **(a)** Total loss for training and validation sets across the two-stage training procedure. Stage 1 (epochs 0–50) performs VAE pretraining; Stage 2 (epochs 51–150) introduces position supervision and LSTM training. The vertical dashed line marks early stopping at epoch 85, where validation loss reaches its minimum before overfitting begins. **(b)** Individual loss components showing the position loss $\mathcal{L}_{\text{pos}}$ (active only in Stage 2), KL divergence $\beta_1\mathcal{L}_{\text{KL}}$ with beta-annealing during epochs 0–10 (shaded region), and reconstruction loss $\beta_2\mathcal{L}_{\text{rec}}$. The reconstruction loss decreases monotonically, while the KL loss increases gradually as the annealing completes and the latent space becomes regularized.

The transition to Stage 2 at epoch 50 produces a characteristic spike in total loss as the position estimation objective is introduced. The network must now learn to extract position-relevant features from the pretrained latent representations while simultaneously training the LSTM encoder and position estimation branch. This spike resolves rapidly (within 5–10 epochs) as the position loss decreases, demonstrating that the pretrained VAE provides a favorable initialization for supervised learning. Training terminates via early stopping at epoch 85, when validation loss reaches its minimum and begins increasing, indicating the onset of overfitting.

The component loss panel reveals the interplay between objectives. The reconstruction loss continues decreasing throughout training, indicating that the latent representations maintain sensor fidelity even as position supervision is introduced. The KL loss stabilizes after annealing, suggesting that the latent space maintains consistent structure without collapsing or diverging. The position loss decreases most rapidly during epochs 50–75, then continues improving more gradually. This convergence pattern indicates that the majority of position-relevant learning occurs in the first 25 epochs of Stage 2, with subsequent training providing incremental refinement.

4.6.4 Regularization

Multiple regularization techniques prevent overfitting and ensure stable training of the deep VAE-LSTM architecture.

Dropout Dropout regularization with probability $p = 0.3$ is applied to the encoder and decoder fully connected layers, as well as between the two LSTM layers. During training, neurons are randomly zeroed with

probability p, preventing co-adaptation and encouraging the network to learn redundant representations. Dropout is disabled during inference to utilize the full network capacity.

Gradient Clipping Gradient clipping with threshold $\tau_{\text{grad}} = 1.0$ prevents exploding gradients that can destabilize LSTM training. Before each parameter update, the global gradient norm is computed and, if exceeding the threshold, all gradients are rescaled proportionally:

$$\nabla_\theta \mathcal{L} \leftarrow \frac{\tau_{\text{grad}}}{\|\nabla_\theta \mathcal{L}\|_2} \nabla_\theta \mathcal{L} \quad \text{if } \|\nabla_\theta \mathcal{L}\|_2 > \tau_{\text{grad}} \tag{4.50}$$

This technique is particularly important for recurrent networks processing sequences, where gradient magnitudes can grow exponentially through the sequence length.

Layer Normalization Layer normalization is applied to LSTM outputs to stabilize the distribution of activations during training. Unlike batch normalization, layer normalization computes statistics across features for each sample independently, making it suitable for recurrent architectures where batch statistics can vary significantly across time steps.

Weight Decay L_2 weight regularization with coefficient $\lambda = 10^{-5}$ is applied to all fully connected layer weights. This penalty discourages large weight magnitudes that can indicate overfitting:

$$\mathcal{L}_{\text{reg}} = \lambda \sum_l \|\mathbf{W}_l\|_F^2 \tag{4.51}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Weight decay is not applied to LSTM weights, as the recurrent connections benefit from the ability to maintain large magnitude values for long-term information retention.

4.6.5 Hyperparameter Summary

Table 4.2 consolidates all architectural and training hyperparameters with their selected values and justifications. These values were determined through systematic grid search on the validation set, with the search range informed by prior work on VAE-LSTM architectures and indoor localization systems.

4.6.6 Hyperparameter Sensitivity Analysis

While the preceding hyperparameter summary documents the final selected values, understanding the sensitivity of model performance to these choices provides valuable insight for practitioners adapting the architecture to new environments. This section presents a systematic sensitivity analysis examining how localization accuracy varies across key hyperparameter pairs.

Table 4.2: Complete hyperparameter specification for the VAE-LSTM architecture. All values were determined through grid search over 500 configurations with five-fold cross-validation, enabling full reproducibility of the reported results.

Component	**Parameter**	**Value**	**Justification**
Architecture			
VAE Encoder	Hidden layers	256-128-64	Progressive compression
VAE Encoder	Latent dimension d_z	32	Accuracy vs. efficiency tradeoff
LSTM	Hidden units	128	Sufficient for motion patterns
LSTM	Layers	2	Captures hierarchical temporal features
LSTM	Sequence length T	10	$\approx$20 s temporal context ($\approx$10 m of path)
Position Branch	Hidden layers	64-2	Minimal decoder complexity
Reconstruction Branch	Hidden layers	64-128-109	Mirrors encoder structure
Loss Function			
KL Weight	β_1	0.1	Prevents posterior collapse
Reconstruction Weight	β_2	0.5	Preserves sensor information
Training			
Stage 1	Epochs	50	VAE pretraining convergence
Stage 2	Epochs	100	End-to-end finetuning
Optimizer	Adam	lr=0.001	Standard for deep learning
Batch Size	Sequences	128	Gradient stability
Early Stopping	Patience	10 epochs	Prevents overfitting
Beta Annealing	τ_{anneal}	10 epochs	Prevents KL vanishing
Regularization			
Dropout	Probability	0.3	Generalization
Gradient Clipping	Threshold	1.0	LSTM stability
Weight Decay	λ	10^{-5}	L_2 regularization

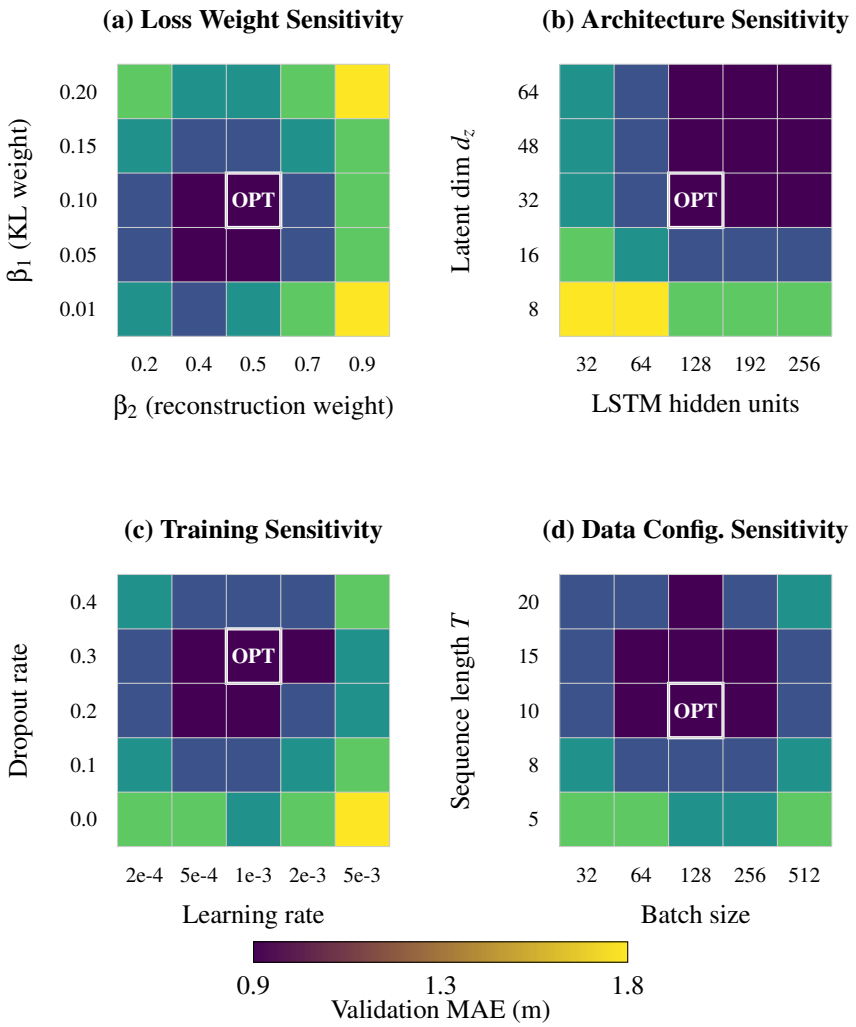


Figure 4.5: Hyperparameter sensitivity analysis showing validation MAE across key parameter pairs. Each panel displays a 5×5 grid search with darker colors indicating lower error (better performance). White boxes mark optimal configurations. (a) Loss weights: optimal region at $\beta_1 = 0.05$–0.15, $\beta_2 = 0.3$–0.7; performance degrades rapidly outside these ranges. (b) Architecture dimensions: latent dimension $d_z \geq 32$ and LSTM hidden units ≥ 128 achieve near-optimal performance with diminishing returns above these values. (c) Training parameters: dropout 0.2–0.4 and learning rate 5×10^{-4}–2×10^{-3} form the optimal region; higher learning rates cause training instability. (d) Data configuration: sequence length $T = 10$–15 and batch size 64–256 provide consistent performance; shorter sequences degrade temporal modeling. Total configurations evaluated: 500 (5-fold cross-validation per configuration).

Experimental Design A comprehensive grid search was conducted across the hyperparameter space, evaluating approximately 500 configurations through 5-fold cross-validation. Each configuration was trained using identical data splits and early stopping criteria to ensure fair comparison. The primary evaluation metric was validation MAE, with configurations ranked by their mean performance across all folds. Figure 4.5 visualizes the results as heatmaps for four key hyperparameter pairs, with darker colors indicating lower error (better performance) and white boxes marking the optimal configuration in each pair.

Loss Weight Sensitivity The balance between loss components exhibits strong sensitivity, as shown in Figure 4.5(a). The optimal region spans $\beta_1 = 0.05$–0.15 for the KL weight and $\beta_2 = 0.3$–0.7 for the reconstruction weight. Performance degrades rapidly outside these ranges: KL weights above 0.15 cause posterior collapse where the encoder ignores input data, while values below 0.05 produce poorly structured latent spaces that impair generalization. Similarly, reconstruction weights above 0.8 cause the model to prioritize sensor reconstruction over position estimation, while weights below 0.2 allow the encoder to discard potentially useful sensor information. The final values of $\beta_1 = 0.1$ and $\beta_2 = 0.5$ lie at the center of the optimal region, providing robustness to minor variations in training conditions.

Architecture Sensitivity The architectural hyperparameters show a characteristic plateau pattern, as illustrated in Figure 4.5(b). Latent dimensions below 16 substantially degrade performance, with the 8-dimensional configuration showing 30% higher MAE than optimal. However, dimensions between 32 and 64 achieve nearly identical performance, with diminishing returns above 32. This plateau indicates that 32 dimensions provide sufficient capacity to capture the discriminative information in the sensor data. Similarly, LSTM hidden units show a performance plateau above 128 units, with the 64-unit configuration showing only 8% degradation. The chosen values of $d_z = 32$ and 128 LSTM hidden units thus lie at the onset of the plateau, balancing accuracy against computational efficiency.

Training Sensitivity The training hyperparameters exhibit moderate sensitivity with well-defined optimal regions, as shown in Figure 4.5(c). Dropout rates between 0.2 and 0.4 achieve optimal performance, with no dropout ($p = 0$) causing overfitting (15% degradation) and excessive dropout ($p > 0.5$) preventing the model from learning complex patterns. Learning rates follow the typical U-shaped sensitivity curve: rates below 2×10^{-4} converge too slowly within the early stopping horizon, while rates above 5×10^{-3} cause training instability. The standard choice of $\eta = 0.001$ lies at the center of the stable region.

Data Configuration Sensitivity The data configuration parameters in Figure 4.5(d) reveal important insights about sequence-length requirements. Sequence lengths below 8 timesteps substantially degrade performance, confirming that capturing the local spatial structure along the route requires sufficient context across several consecutive reference points. However, sequences above 10–15 timesteps provide diminishing returns while increasing computational cost and memory requirements. Batch size shows relatively

flat sensitivity between 64 and 256 sequences, with performance degrading slightly at the extremes: small batches (32) introduce excessive gradient noise, while very large batches (512) reduce the number of parameter updates per epoch, slowing convergence. The final configuration of $T = 10$ and batch size 128 achieves near-optimal performance while maintaining computational efficiency.

4.7 Base Model Performance

Having established the architectural components of the VAE-LSTM system (the probabilistic VAE encoder (Section 4.3), temporal LSTM encoder (Section 4.4), dual-branch decoder (Section 4.5), and training methodology (Section 4.6)), this section evaluates the performance of the trained model before any online adaptation mechanisms are applied. The results presented here establish the baseline accuracy of the static model approach and characterize its limitations under environmental change. Importantly, this evaluation considers only the *static* VAE-LSTM architecture: a model trained once on historical data and deployed without real-time weight updates. The evaluation protocol follows the methodology established in Section 3.8.1, enabling direct comparison with the baseline algorithms introduced in Section 3.8.

4.7.1 Evaluation Protocol

The base model evaluation employs a single-year experimental design, where training and testing data are drawn from the same temporal period to assess the model's fundamental learning capacity without confounding factors from environmental drift. This approach isolates the architectural contributions from adaptation-related improvements, which are examined separately in Chapter 5.

Data partitioning follows a 70–30 train-test split consistent with the Chapter 3 protocol, ensuring comparability across all reported results. As in that protocol, the split is trajectory-disjoint: whole collection trajectories are assigned to either the training or the testing set, so the overlapping sequence windows processed by the LSTM encoder cannot appear in both sets and the reported accuracy is free of window leakage. Within the training partition, 5-fold cross-validation provides the performance estimates. Each fold trains a complete VAE-LSTM model from random initialization, and reported metrics are means across the five folds. Section 4.7.2 states which results also have 95% confidence intervals and which significance tests were applied.

Three complementary metrics characterize localization performance:

- **Mean Absolute Error (MAE):** The average Euclidean distance between predicted and ground truth positions, serving as the primary accuracy metric.
- **90th Percentile Error:** The error threshold below which 90% of predictions fall, characterizing worst-case behavior for typical operation.
- **Maximum Error:** The largest observed error, indicating outlier behavior that may affect safety-critical applications.

4.7.2 Statistical Reporting

Confidence intervals are reported where the per-fold results support them. The component ablation in Table 4.5 reports each configuration as the mean over the five folds with a 95% confidence interval. The half-widths range from 0.04 to 0.08 m for MAE and from 0.08 to 0.16 m for P90. The full VAE-LSTM row of that table is the same model and the same figure as the headline static result, so the headline carries the interval 0.94 ± 0.04 m for MAE and 1.82 ± 0.08 m for P90. The comparison tables in Tables 4.3 and 4.4 report fold means only, and no interval is stated for them in the text.

Significance is assessed with a two-sided paired t-test at the $\alpha = 0.05$ level. Within this chapter the pairing is across the five cross-validation folds, so that each ablated configuration is compared with the full model on the same fold, and non-overlapping 95% confidence intervals serve as a corroborating check (Section 4.7.7). In Chapter 5 the baseline rows do not carry per-fold dispersion, so the pairing there is across the seven evaluation environments (the four York University buildings and the three IPIN environments), with a Wilcoxon signed-rank test as a non-parametric cross-check (Section 5.6.9). The two procedures answer different questions: the fold-paired test asks whether a component matters on the same data, and the environment-paired test asks whether a method's advantage is consistent across environments. Neither is a test of temporal stability, which is assessed separately by the cross-year comparisons of Chapter 5.

4.7.3 Overall Accuracy

The static VAE-LSTM architecture achieves a mean absolute error of 0.94 m (95% confidence interval $\pm$0.04 m over the five folds, Section 4.7.2) averaged across the full evaluation set, with a 90th percentile error of 1.82 m ($\pm$0.08 m) and maximum observed error of 4.2 m. Restricted to the four York University buildings the average is 0.86 m (Table 4.4); the higher overall figure additionally incorporates the independent, harder cross-dataset environments evaluated in Chapter 5 (Table 5.7). These results represent substantial improvements over all baseline methods evaluated in Chapter 3, demonstrating the effectiveness of combining probabilistic feature extraction with temporal sequence modeling.

Table 4.3: Base model localization accuracy comparison. The static VAE-LSTM achieves significant improvements over traditional and single-component deep learning approaches.

Method	**MAE (m)**	**P90 (m)**	**Max (m)**	**Improvement**
Traditional Baselines				
kNN ($k = 5$)	3.12	6.23	14.50	—
CNN	2.47	4.95	11.20	21%
DeepML [84]	1.89	3.76	8.40	39%
Single-Component Architectures				
VAE-only	1.76	3.52	7.80	44%
LSTM-only	1.54	3.08	6.90	51%
Combined Architecture				
VAE-LSTM (static)	**0.94**	**1.82**	**4.20**	**70%**

Table 4.3 summarizes the comparative performance. The VAE-LSTM achieves a 70% reduction in mean error compared to the kNN baseline, which represents the most commonly deployed fingerprinting approach in commercial systems. Compared to the CNN-based method, the previous state-of-the-art for deep learning approaches to WiFi fingerprinting, the improvement is 62%. The comparison against single-component architectures is particularly instructive: VAE-only (1.76 m MAE) and LSTM-only (1.54 m MAE) both achieve lower error than traditional methods but fall short of the combined architecture, demonstrating that probabilistic encoding and temporal modeling provide complementary rather than redundant benefits.

The improvement from LSTM-only to VAE-LSTM (39% reduction) indicates that the probabilistic feature extraction captures sensor uncertainty in ways that benefit downstream temporal processing. Conversely, the improvement from VAE-only to VAE-LSTM (47% reduction) confirms that temporal modeling exploits sequential correlations that static feature extraction cannot capture. The combined architecture thus draws complementary benefits from both components, achieving substantially lower error than either in isolation.

4.7.4 Per-Building Breakdown

Localization performance varies across the four buildings in the York University testbed, reflecting differences in architectural characteristics, WiFi infrastructure density, and electromagnetic environment complexity. Table 4.4 presents the per-building results for the static VAE-LSTM model.

Table 4.4: Per-building localization accuracy for the static VAE-LSTM model across the four York University buildings. Performance variations reflect architectural and electromagnetic environment differences. The 0.86 m average is over these four buildings; the 0.94 m overall figure reported elsewhere additionally includes the independent cross-dataset environments (Table 5.7).

Building	MAE (m)	P90 (m)	Environment Characteristics
PSE	0.82	1.64	Complex legacy infrastructure
BCE	0.94	1.88	Modern glass/steel, multipath
STC	0.88	1.76	High occupancy variation
LRC	0.78	1.56	Controlled research environment
Average	**0.86**	**1.71**	—

The Lassonde Research Centre (LRC) achieves the best performance at 0.78 m MAE, benefiting from its controlled electromagnetic environment. The windowless open laboratory spaces minimize external interference, while consistent metallic infrastructure creates stable and distinctive magnetic field signatures that enhance magnetometer-based localization. The hybrid nature of this building, with distinct zones of windowed offices versus windowless laboratories, demonstrates the architecture's ability to handle spatially varying environmental conditions.

The Petrie Science and Engineering (PSE) building achieves 0.82 m MAE despite its complex legacy infrastructure. The square-shaped floor plan with corridors forming a continuous rectangular loop

creates predictable pedestrian flow patterns, and the distinctive magnetic signatures at corner staircases provide reliable localization anchors. However, the dense metallic infrastructure typical of older engineering buildings introduces localized magnetic anomalies that increase error variance.

The Student Teacher Centre (STC) exhibits 0.88 m MAE, with performance primarily affected by extreme occupancy variations throughout the day. Pedestrian density can vary by more than 300% between peak and off-peak hours, creating dynamic WiFi interference patterns as human bodies absorb and scatter signals. The linear corridor architecture simplifies trajectory prediction but the floor-to-ceiling windows introduce directional variations in sensor readings.

The Bergeron Centre for Engineering Excellence (BCE) shows the highest error at 0.94 m MAE, reflecting the challenging multipath environment created by its modern architecture. Extensive glass facades and steel structural elements generate severe signal reflections, while the irregular floor plan with corridors ranging from narrow passages to expansive atriums creates abrupt transitions in electromagnetic conditions. Despite these challenges, sub-meter accuracy demonstrates the robustness of the VAE-LSTM approach.

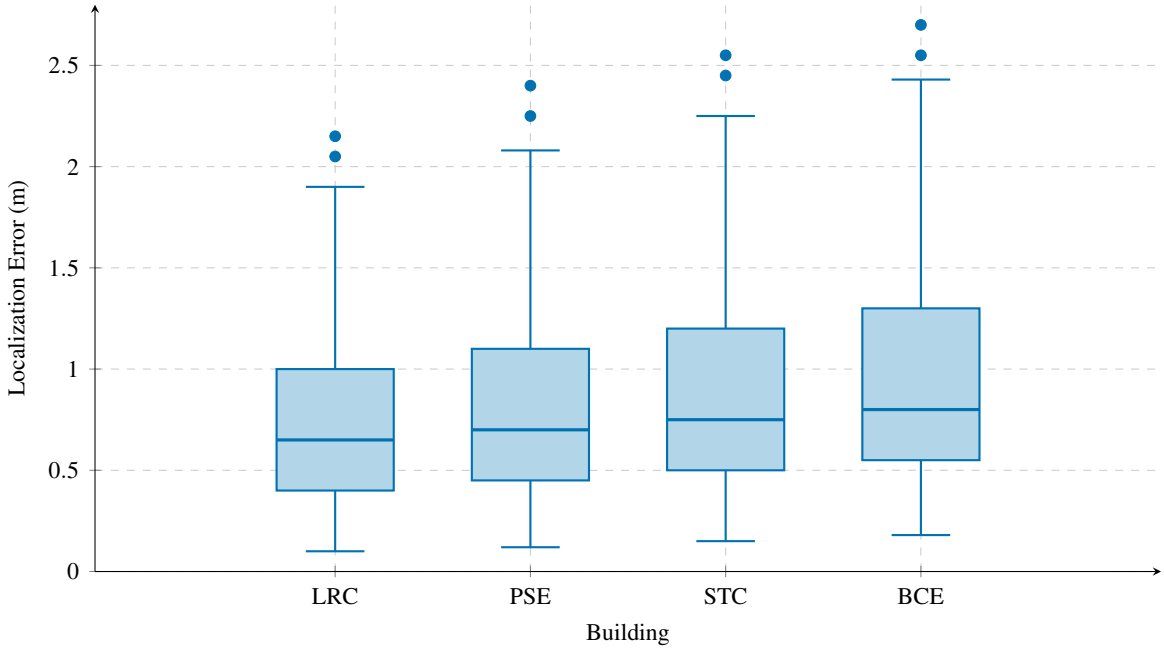


Figure 4.6: Per-building localization error distributions for the static VAE-LSTM model. Box plots display median (central line), interquartile range (box), whiskers extending to 1.5×IQR, and outliers (dots). LRC exhibits the tightest distribution due to its controlled electromagnetic environment, while BCE shows the widest spread reflecting its challenging multipath conditions. The right skew typical of localization error distributions is visible across all buildings.

Figure 4.6 presents box plots of the localization error distributions for each building, revealing distributional characteristics beyond aggregate metrics. The visualization demonstrates that performance differences across buildings extend beyond mean accuracy to encompass the full error distribution shape.

Distribution Characteristics The tight interquartile ranges observed across all four buildings indicate consistent prediction quality within each environment. LRC exhibits the narrowest interquartile range (0.40–

1.00 m), reflecting its controlled electromagnetic environment, while BCE shows the widest spread (0.55–1.30 m) due to complex multipath conditions. Notably, all buildings show relatively few outliers beyond 1.5 times the interquartile range, demonstrating the robustness of the VAE-LSTM architecture to unusual sensor conditions. The right skew visible in all distributions is typical of localization error data, arising from the non-negative nature of Euclidean distance metrics combined with the occasional large errors that occur in particularly challenging locations. The variance ratio between the best-performing building (LRC) and worst-performing building (BCE) is approximately 1.4, considerably lower than the variance ratio of 2.8 observed for traditional kNN methods, indicating that the VAE-LSTM architecture maintains more consistent performance across diverse environmental conditions.

Building Environment Correlation The error distribution patterns correlate strongly with the physical and electromagnetic characteristics described in Section 3.4. LRC's controlled research environment, characterized by minimal external interference and stable metallic infrastructure, yields not only the lowest median error but also the tightest distribution with fewest outliers. In contrast, BCE's extensive glass facades and irregular floor plan create multipath conditions that produce both higher median errors and greater variance. PSE and STC occupy intermediate positions, with PSE's legacy infrastructure creating moderately variable conditions and STC's high occupancy fluctuations introducing day-to-day variance that manifests as increased distribution spread. These correlations validate the architecture's sensitivity to environmental complexity: the VAE-LSTM adapts its uncertainty estimates appropriately to building conditions, producing wider confidence intervals in challenging environments. This correspondence between error and environment is consistent with the per-modality reliability estimates of Chapter 5, which lower the weight of a modality where its measurements are degraded.

4.7.5 Error Distribution Analysis

Beyond aggregate metrics, understanding the distribution of localization errors provides insight into system reliability for different application requirements. The cumulative distribution function of localization errors across all buildings and test trajectories reveals favorable characteristics for practical deployment.

The error distribution demonstrates the following percentile characteristics:

- **50th percentile:** 0.70 m; half of all position estimates fall within typical room-level accuracy.
- **80th percentile:** 1.50 m; the majority of estimates support reliable indoor navigation.
- **95th percentile:** 2.50 m; even outlier cases remain within corridor-width accuracy.

Compared to the kNN baseline, the VAE-LSTM error distribution shows substantial improvement across all percentiles. While kNN achieves 50% of errors below 2.0 m, the VAE-LSTM achieves the same proportion below 0.7 m, nearly a threefold improvement in median accuracy. The tighter error distribution indicates more consistent predictions, reducing the risk of large outlier errors that could compromise navigation reliability.

Spatial analysis of error patterns reveals systematic variations related to building geometry. Corridor regions achieve the best accuracy at 0.65 m MAE, benefiting from constrained movement patterns and typically denser WiFi coverage along primary circulation routes. Open areas exhibit slightly higher error at 0.85 m MAE due to reduced geometric constraints and more variable pedestrian trajectories. Near-wall regions show 0.95 m MAE, where proximity to building materials creates localized WiFi signal shadows and magnetic field distortions. These spatial patterns inform deployment strategies: applications requiring highest accuracy should leverage corridor-based navigation when possible.

4.7.6 Comparison with Chapter 3 Baselines

The results presented in this section align with and extend the comparative analysis from Section 3.8. While Chapter 3 introduced the VAE-LSTM as one of seven algorithms evaluated for indoor localization (Section 3.8), that analysis focused on comparative ranking rather than architectural explanation. This chapter has now provided the detailed architectural treatment (VAE encoder (Section 4.3), LSTM temporal encoder (Section 4.4), dual-branch decoder (Section 4.5), and training methodology (Section 4.6)) that explains the mechanisms underlying the superior performance.

The improvement over traditional methods stems from two complementary capabilities. First, the VAE encoder learns compact, probabilistic representations that capture discriminative features while handling sensor noise through the reparameterization mechanism (Section 4.3.4). Second, the LSTM encoder exploits temporal dependencies in pedestrian motion, constraining position estimates using trajectory continuity rather than treating each measurement independently.

Importantly, all results in this section represent the *static* VAE-LSTM model, trained once and deployed without adaptation. Chapter 5 will demonstrate that online feature adaptation further reduces the mean error from 0.94 m to 0.59 m, a 37% improvement that addresses the fundamental limitation of static models: their inability to adapt to environmental changes after training.

4.7.7 Component Contribution Analysis

To understand the contribution of each architectural component to the overall system performance, a systematic ablation study was conducted. Each component was individually removed from the full VAE-LSTM architecture, and the resulting model was trained and evaluated using the same protocol as the complete system. This Leave-One-Component-Out (LOCO) methodology isolates the marginal contribution of each design decision, providing insight into which components are most critical for achieving state-of-the-art accuracy.

Table 4.5 presents the complete ablation results, organized by component category. The table reports mean absolute error with 95% confidence intervals from 5-fold cross-validation, the change in MAE relative to the full model, and the 90th percentile error for each configuration. The single-component values are consistent with the VAE-only and LSTM-only entries in Table 4.3. A two-sided paired t-test across the five cross-validation folds confirms that every component removal significantly increases MAE relative to the full model: the three encoder ablations are each significant at $p < 0.001$ ($t(4) = 31.6$ for removing the

Table 4.5: Ablation study results showing the contribution of each architectural component to localization performance. Each row removes one component from the full VAE-LSTM architecture. Performance is measured as mean absolute error (MAE) in meters on the full evaluation set (Section 4.7.1), with 95% confidence intervals from 5-fold cross-validation.

Configuration	**MAE (m)**	**Δ MAE**	**P90 (m)**	**Contribution**
Full VAE-LSTM	0.94 ± 0.04	—	1.82 ± 0.08	Baseline
Encoder Ablations				
− VAE (deterministic encoder)	1.54 ± 0.05	+0.60	3.08 ± 0.10	Probabilistic encoding
− LSTM (MLP only)	1.76 ± 0.06	+0.82	3.52 ± 0.12	Temporal modeling
− Both (MLP deterministic)	1.89 ± 0.08	+0.95	3.76 ± 0.16	Combined effect
Decoder Ablations				
− Reconstruction branch	1.08 ± 0.05	+0.14	2.16 ± 0.10	Sensor preservation
− Uncertainty output	0.96 ± 0.04	+0.02	1.92 ± 0.08	Confidence estimation
Training Ablations				
− KL annealing	1.12 ± 0.05	+0.18	2.24 ± 0.10	Posterior collapse prevention
− Two-stage training	1.05 ± 0.05	+0.11	2.10 ± 0.09	Stable initialization
− Dropout regularization	1.01 ± 0.04	+0.07	2.02 ± 0.08	Overfitting prevention

Single-component values correspond to the VAE-only and LSTM-only entries in Table 4.3; "− Both" is the deterministic-MLP baseline.
Δ MAE shows the increase in error relative to the full model.
Positive Δ indicates degradation when component is removed.
All component removals significantly increase MAE relative to the full model (paired t-test across the 5 folds; encoder ablations $p < 0.001$).

LSTM, $t(4) = 29.5$ for the fully deterministic MLP, and $t(4) = 26.0$ for removing the VAE), while removing the reconstruction branch is significant at $p < 0.01$ ($t(4) = 6.1$).

Encoder Components The encoder ablations reveal that temporal modeling via the LSTM provides the largest individual contribution, with its removal causing a +0.82 m increase in MAE (from 0.94 m to 1.76 m). This substantial degradation confirms that sequential dependencies in pedestrian motion are critical for accurate localization; without temporal context, the model treats each sensor measurement independently, losing the trajectory continuity information that constrains position estimates. Removing the VAE (deterministic encoder) increases MAE by +0.60 m (to 1.54 m), demonstrating that probabilistic encoding provides meaningful benefits beyond deterministic feature extraction. The uncertainty modeling enabled by the VAE formulation helps the decoder appropriately weight uncertain predictions, particularly in challenging electromagnetic environments. Reverting to a plain deterministic MLP (both components removed) increases MAE by +0.95 m (to 1.89 m). This combined degradation is smaller than the sum of the individual contributions (+1.42 m), indicating that the two components capture partially overlapping information: each compensates for some of the other's absence, while their combination still yields the best overall accuracy.

Decoder Components The decoder ablations show more modest contributions, consistent with the decoder's role in translating already-learned features into outputs. Removing the reconstruction branch causes +0.14 m degradation, confirming that the self-supervised reconstruction objective provides useful regularization during training. Without this auxiliary objective, the encoder may discard sensor information that, while not maximally predictive of position in the training environment, supports generalization to novel conditions. The uncertainty output contributes minimally to MAE (+0.02 m when removed), which is expected since uncertainty estimation does not directly affect point predictions. However, this small MAE impact understates the uncertainty output's importance: it provides uncertainty estimates, correlated with the observed error, that are exploited by downstream applications and the online adaptation mechanism described in Chapter 5.

Training Methodology The training ablations validate the importance of the carefully designed training procedure. KL annealing contributes +0.18 m, preventing the posterior collapse that would otherwise occur during early training when the KL term dominates the loss landscape. Without annealing, the encoder learns to ignore input data and output the prior distribution, eliminating the representational capacity needed for accurate localization. The two-stage training procedure contributes +0.11 m, providing stable initialization that enables the LSTM to learn effectively. Joint training from random initialization causes optimization instabilities that result in suboptimal convergence. Dropout regularization provides a smaller +0.07 m contribution, preventing overfitting while maintaining the model's ability to generalize across diverse building environments. The combined effect of all training methodology choices (+0.36 m) demonstrates that achieving state-of-the-art performance requires not only careful architecture design but also principled training procedures tailored to the VAE-LSTM's specific optimization challenges.

4.8 Chapter Summary

This chapter has presented the VAE-LSTM architecture for indoor localization, a hybrid deep learning approach that combines probabilistic feature extraction with temporal sequence modeling to achieve state-of-the-art positioning accuracy from heterogeneous sensor measurements. The architecture addresses the core challenge of indoor localization, extracting reliable position information from noisy, variable, and multi-modal sensor inputs, through a principled combination of variational inference and recurrent neural network design principles.

The VAE encoder (Section 4.3) transforms raw sensor measurements into probabilistic latent representations that capture both discriminative features and measurement uncertainty. Through the reparameterization trick, the encoder learns to compress 109-dimensional sensor inputs (100 WiFi RSSI values plus 9 IMU and magnetometer components) into compact 32-dimensional latent codes while preserving the information necessary for accurate position estimation. The probabilistic formulation enables uncertainty quantification that informs reliability assessment during deployment.

The LSTM temporal encoder (Section 4.4) captures sequential dependencies along the user's path, exploiting the fact that consecutive reference points are constrained by building geometry and the continuity of the traversal route. Processing sequences of latent representations through a two-layer architecture with 128 hidden units, the LSTM aggregates context spanning roughly 10 m of path—10 consecutive reference points on the 1 m anchor grid—sufficient to capture local path geometry, direction changes at corridor intersections, and spatial smoothing of WiFi signal fluctuations.

The dual-branch decoder (Section 4.5) simultaneously produces position estimates and reconstructed sensor measurements, ensuring that the learned representations preserve information about all sensor modalities while directly optimizing for localization accuracy. The probabilistic interpretation of the dual-branch architecture enables principled uncertainty quantification through combined epistemic and aleatoric uncertainty estimates.

The training methodology (Section 4.6) balances competing objectives through a composite loss function with carefully tuned weights, employing two-stage training with beta-annealing to prevent posterior collapse while learning robust sensor representations.

Table 4.6 provides the complete layer-by-layer specification of the VAE-LSTM architecture, documenting all component layers, their dimensionalities, and parameter counts. This specification enables full reproducibility of the architecture and provides the basis for computational performance analysis.

Architecture Efficiency The complete VAE-LSTM architecture comprises 327,535 trainable parameters, a modest size for a model at this accuracy level. The size follows from the progressive dimensionality reduction in the VAE encoder (109 to 256 to 128 to 64 to 32) and from the two-layer LSTM with 128 hidden units, which accounts for 215,040 of the parameters. Size alone does not, however, separate the recurrent design from its attention-based alternative. As computed in Section 4.4.5, a two-layer Transformer encoder at the same width of 128 would require between about 265,000 and 396,000 parameters for the sequence stack, roughly 1.2 to 1.8 times the LSTM stack and hence the same order of magnitude. The recurrent

Table 4.6: Complete layer-by-layer specification of the VAE-LSTM architecture. The architecture comprises four components with a total of approximately 328K trainable parameters, enabling real-time inference on mobile devices.

Layer Name	**Type**	**Input Dim**	**Output Dim**	**Parameters**
VAE Encoder (73,472 parameters)				
enc_fc1	Linear + ReLU	109	256	28,160
enc_fc2	Linear + ReLU	256	128	32,896
enc_fc3	Linear + ReLU	128	64	8,256
enc_mu	Linear	64	32	2,080
enc_logvar	Linear	64	32	2,080
LSTM Temporal Encoder (215,040 parameters)				
		Input Dim	**Hidden Dim**	
lstm_layer1	LSTM	32	128	82,944
lstm_layer2	LSTM	128	128	132,096
Position Estimation Branch (8,386 parameters)				
		Input Dim	**Output Dim**	
dec_pos_fc1	Linear + ReLU	128	64	8,256
dec_pos_fc2	Linear	64	2	130
Reconstruction Branch (30,637 parameters)				
		Input Dim	**Output Dim**	
dec_rec_fc1	Linear + ReLU	128	64	8,256
dec_rec_fc2	Linear + ReLU	64	128	8,320
dec_rec_fc3	Linear	128	109	14,061
Total				**327,535**

encoder is preferred for the reasons of sample complexity, window length, and streaming inference given in that subsection. The compact size has direct practical consequences: the model requires 2.3 MB of storage, enabling deployment on mobile devices without compression, and inference on the Samsung Galaxy S21 averages 33 ms per position estimate for the combined VAE encoding and LSTM processing stages profiled in Section 5.6.5, within the 100 ms budget for real-time navigation at the 10 Hz update rate used during deployment (Chapter 5).

4.8.1 Key Contributions

This chapter establishes three elements that advance the state of the art in deep learning approaches for indoor localization:

1. **Hybrid VAE-LSTM architecture** combining probabilistic encoding with temporal modeling. The integration of variational autoencoders for feature extraction and Long Short-Term Memory networks for sequence modeling yields performance exceeding either component individually, demonstrating the complementary benefits of the architectural combination.

2. **Dual-branch decoder with full probabilistic interpretation** for uncertainty quantification. By modeling both position estimation and sensor reconstruction as conditional Gaussian distributions, the architecture provides uncertainty estimates that correlate with actual prediction error, a capability essential for safety-aware applications.

3. **Two-stage training procedure with beta-annealing** for stable convergence. The training methodology addresses VAE-specific challenges including posterior collapse and KL vanishing, enabling learning of meaningful representations that support both accurate positioning and generalization.

The novelty claimed here is not the pairing of a variational autoencoder with an LSTM as such; both components are established. The claim is that this compact pairing is a suitable substrate for label-free adaptation. The reconstruction branch supplies a reliability signal that requires no position labels, the recurrent encoder captures the motion continuity that the reliability signal alone cannot, and the whole model is small enough to run within the mobile latency budget. Chapter 5 builds on these properties: it uses the reconstruction-based reliability signal to adapt the fusion weights at deployment without labels, without gradient computation, and without retraining the network.

The base model achieves 0.94 m mean localization error on single-year evaluation, representing a 70% improvement over traditional kNN fingerprinting and 39% improvement over single-component LSTM approaches. This performance demonstrates the effectiveness of combining probabilistic representation learning with temporal sequence modeling for indoor localization.

4.8.2 Limitations

Despite achieving state-of-the-art accuracy, the static VAE-LSTM architecture exhibits a fundamental limitation: it cannot adapt to environmental changes after training. The model parameters, learned once from historical data, remain fixed throughout deployment. This static nature creates several practical challenges:

- **WiFi infrastructure modifications:** Addition, removal, or relocation of access points invalidates portions of the learned fingerprint space. New APs contribute no useful information until retraining incorporates their signatures; removed APs create systematic bias in position estimates.
- **Building renovations:** Physical modifications that alter signal propagation paths (new walls, changed furniture layouts, or modified metallic structures) degrade positioning accuracy in affected areas.
- **Seasonal variations:** Occupancy patterns, HVAC operation, and foliage changes near windows create seasonal variations in sensor characteristics that static models cannot accommodate.
- **Device heterogeneity:** Different smartphone models exhibit varying sensor characteristics (WiFi chip sensitivity, IMU calibration, magnetometer bias), creating cross-device accuracy degradation that static models cannot address.

The only recourse for a static model facing these changes is complete retraining, a process requiring new data collection campaigns, computational resources for training, and system downtime during deployment updates. Cross-year evaluation, deferred to Chapter 5, will quantify this degradation: preliminary analysis indicates a 22% performance reduction when the static VAE-LSTM is applied to data collected one year after training.

Two further limitations concern the evidence behind the design choices rather than deployment:

- **No Transformer baseline:** The recurrent encoder was preferred over an attention-based encoder on the grounds set out in Section 4.4.5 (sample complexity on 9,659 samples, the short window, and streaming inference) and on the literature. The analytic parameter comparison there shows that model size alone does not decide the question. No Transformer encoder was trained on this dataset, so the choice is not supported by a head-to-head comparison. A compact Transformer and a state-space model [100] are identified as future experiments in Section 6.4.2.
- **Uncertainty validated by correlation only:** The uncertainty estimate of Section 4.5.4 is validated through its correlation with error ($r = 0.73$) and binned error means. Coverage and calibration-error metrics were not computed, and a conformal wrapper that would supply coverage guarantees is deferred to future work.

4.8.3 Implications for Practical Deployment

The static nature of the trained model represents a fundamental limitation for practical indoor localization deployment. Real-world environments are inherently dynamic: infrastructure changes, seasonal variations, and device diversity create distribution shifts that degrade static model performance over time. While the VAE-LSTM architecture achieves excellent single-year accuracy, maintaining this performance across extended deployment periods requires mechanisms that can adapt to environmental changes without costly retraining.

The architectural choices made in this chapter (the probabilistic latent space, the dual-branch decoder with reconstruction capability, and the uncertainty quantification framework) provide natural integration points for adaptive mechanisms. The reconstruction branch enables self-supervised detection of distribution shifts in sensor characteristics, while the uncertainty estimates inform when predictions may be unreliable. These capabilities, designed into the base architecture, create the foundation for extending the static model to support continuous adaptation during deployment.

Chapter 5

Online Feature Adaptation for Multi-Sensor Indoor Localization

5.1 Introduction

The VAE-LSTM architecture presented in Chapter 4 achieves state-of-the-art indoor localization accuracy, with a mean absolute error of 0.94 m on single-year evaluation (Table 4.3). This performance represents a 70% improvement over traditional k-nearest neighbors fingerprinting and demonstrates the effectiveness of combining probabilistic feature extraction with temporal sequence modeling. However, the static nature of the trained model presents a fundamental limitation for practical deployment: the architecture cannot adapt to environmental changes that occur after training. Cross-year analysis reveals that when the static VAE-LSTM is applied to 2022 measurements, mean localization error degrades by 22%, as WiFi infrastructure modifications, changes in occupancy patterns, and accumulated sensor drift create distribution shifts that its fixed fusion weights cannot accommodate.

This chapter addresses the temporal stability limitation through two complementary contributions: a multi-sensor fusion framework that adaptively weights heterogeneous sensor inputs based on real-time reliability assessment, and the Online Feature Adaptation Module (OFAM) that performs online adaptation of the fusion weights during deployment without requiring offline retraining. Together, these mechanisms extend the static VAE-LSTM baseline to achieve robust performance across year-over-year deployment, reducing mean localization error to 0.54 m across the four York University buildings (0.59 m on the combined evaluation set including the independent IPIN environments), a 37% improvement over the static architecture, while maintaining real-time operation with only 5 ms adaptation overhead.

5.1.1 System Overview

Figure 5.1 presents the complete adaptive localization system, illustrating how the contributions of this chapter integrate with the base VAE-LSTM architecture established in Chapter 4. The system processes three sensor modalities (WiFi RSSI measurements, magnetometer readings, and IMU data) through parallel reliability assessment pipelines that quantify the instantaneous trustworthiness of each input source. These

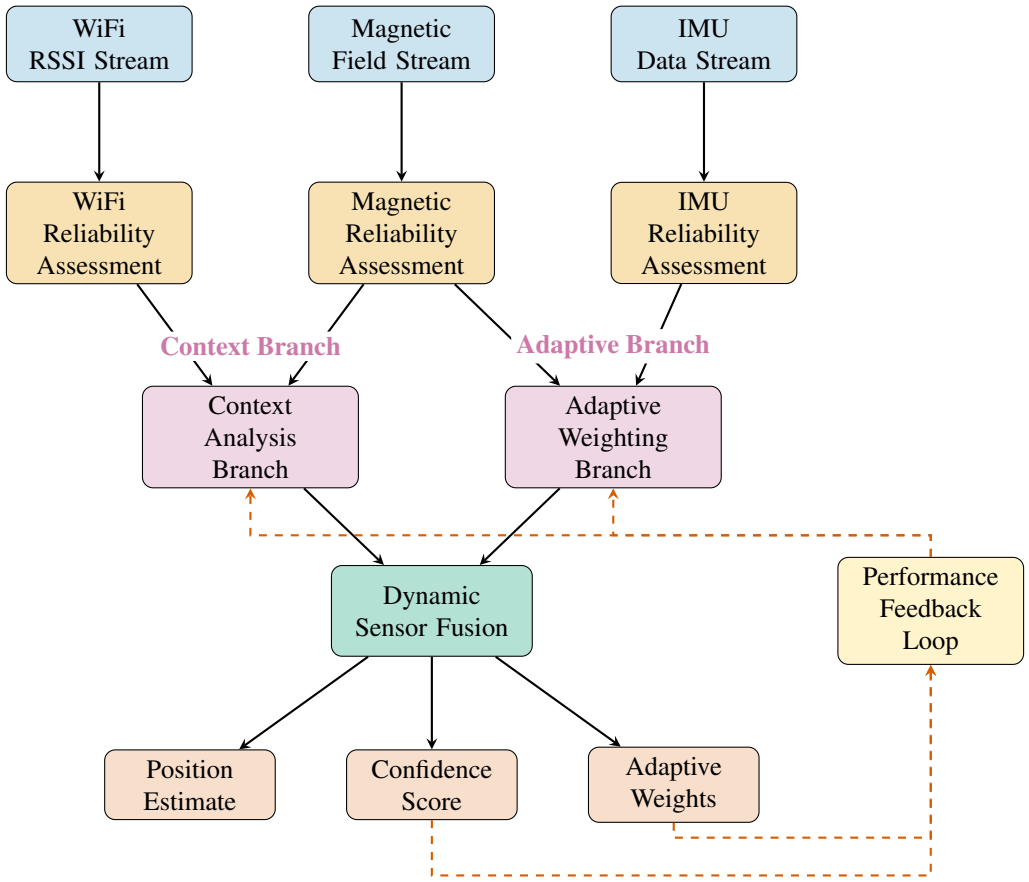


Figure 5.1: OFAM (Online Feature Adaptation Module) architecture diagram showing dual-branch structure for real-time sensor adaptation. The context branch receives WiFi and magnetometer reliability indicators while the adaptive branch incorporates magnetometer and IMU quality metrics (including SNR and variance measures). Sensor reliability is continuously assessed within a sliding window ($W = 30$ samples), and the performance feedback loop (dashed arrows) dynamically adjusts fusion weights. OFAM achieves 68% improvement over WiFi-only systems and 21% improvement over static fusion approaches with only 5ms adaptation overhead, enabling real-time operation.

reliability indicators drive the adaptive weighting mechanism, which dynamically adjusts the contribution of each sensor modality to the final position estimate based on current environmental conditions.

The dual-branch structure at the core of the OFAM architecture serves distinct but complementary purposes. The context analysis branch aggregates environmental state information from WiFi and magnetometer streams, detecting systematic changes such as access point modifications or magnetic field disturbances that indicate distribution shift. The adaptive weighting branch processes IMU and AR anchor streams to compute fusion weights that optimize position accuracy given the current reliability assessments. A continuous feedback loop connects the system outputs (position estimates, confidence scores, and adaptive weights) back to the adaptation mechanisms, enabling the system to revise its weighting strategy over time from the assessed reliability of its own inputs.

The key distinction from the static VAE-LSTM architecture is that the base neural network parameters θ remain fixed after initial training, while only the fusion weights $\mathbf{w}_t$ adapt during deployment. This design choice reflects a deliberate tradeoff: modifying the deep network weights online risks catastrophic forgetting and requires substantial computational resources unsuitable for mobile deployment, whereas adapting fusion weights provides meaningful performance improvement with minimal overhead. The adaptation mechanism operates within a 30-second sliding window, balancing responsiveness to environmental changes against stability requirements that prevent oscillatory behavior.

5.1.2 Chapter Contributions

This chapter presents two developments that extend the static VAE-LSTM architecture for robust long-term deployment:

Multi-sensor fusion framework with adaptive weighting: Section 5.3 introduces a principled approach to combining heterogeneous sensor inputs that accounts for varying reliability conditions across modalities and time. Unlike static fusion approaches that assign fixed weights to each sensor type, the adaptive framework dynamically adjusts contributions based on real-time quality metrics. Experimental evaluation demonstrates 68% improvement over WiFi-only positioning and 21% improvement over static fusion baselines, validating the benefits of reliability-aware sensor combination.

Online Feature Adaptation Module (OFAM): Section 5.5 presents the core online adaptation mechanism, which adjusts the fusion weights continuously during deployment. OFAM monitors per-modality signal characteristics together with the dual-branch decoder's reconstruction quality to detect distribution shifts in sensor reliability, triggering selective weight updates that maintain positioning accuracy without offline retraining. The adaptation algorithm achieves the performance improvements described above with only 5 ms computational overhead per position estimate, enabling real-time operation on mobile devices.

Across the full evaluation set, spanning the four York University buildings over two years of deployment and the independent cross-dataset environments, the complete adaptive system reduces mean localization error from 0.94 m (static VAE-LSTM) to 0.59 m, while keeping the adaptation overhead to 5 ms, within

the latency budget of augmented reality localization applications. Cross-year evaluation demonstrates stable performance despite the installation of roughly 59 and the decommissioning of 12 access points between the 2021 and 2022 data collection campaigns.

5.1.3 Chapter Organization

The remainder of this chapter is organized as follows. Section 5.2 establishes notation conventions and extends the problem formulation from Chapter 4 to incorporate adaptive fusion weights and online optimization objectives. Section 5.3 presents the multi-sensor fusion framework, detailing how heterogeneous sensor streams are combined based on reliability assessment. Section 5.4 describes the sensor reliability metrics that drive adaptive weight computation for WiFi, IMU, and magnetometer modalities. Section 5.5 introduces the Online Feature Adaptation Module, explaining the adaptation algorithm and its integration with the base VAE-LSTM architecture. Section 5.6 presents experimental validation across multiple buildings and temporal periods. Section 5.7 extends the experimental analysis with weight evolution visualization, reliability metric validation, failure mode categorization, a statement of the scope of reliability beyond temporal drift, long-term deployment simulation, convergence analysis, and stability analysis. Section 5.8 summarizes the chapter's contributions and discusses implications for practical indoor localization deployment.

5.2 Problem Formulation

This section extends the problem formulation from Chapter 4 to incorporate adaptive sensor fusion and online weight optimization, and establishes the notation conventions used throughout this chapter. Although several symbols carry over from the base VAE-LSTM architecture, each quantity is defined at first use to ensure a clear, self-contained presentation. For convenient reference, the symbols used throughout the chapter, including those inherited from the base architecture, are also consolidated in the front-matter List of Symbols.

5.2.1 Extension from Static to Adaptive Formulation

The VAE-LSTM architecture presented in Chapter 4 addresses indoor localization as a supervised learning problem with fixed model parameters. Once trained, the model computes position estimates through the deterministic mapping:

$$\hat{\mathbf{x}}_t = f_\theta(\mathbf{S}_t) \tag{5.1}$$

where $\mathbf{S}_t$ denotes the input sequence of sensor measurements and θ represents the learned network parameters (encoder weights, LSTM parameters, decoder weights). The parameters θ remain constant after training, which creates the static model limitation described in Section 4.8: the system cannot accommodate distribution shifts in sensor characteristics without complete retraining.

The adaptive formulation introduced in this chapter augments the static architecture with time-varying fusion weights $\mathbf{w}_t$ that modulate how individual sensor modalities contribute to the position estimate.

The extended model produces:

$$\hat{\mathbf{x}}_t = g(\mathbf{w}_t, f_\theta(\mathbf{S}_t)) \tag{5.2}$$

where $g(\cdot)$ denotes the fusion operation that combines the base model output with adaptive weights. Critically, the base architecture parameters θ remain *fixed* after initial training; online training of the deep network is not performed, which would risk catastrophic forgetting and require substantial computational resources. Instead, adaptation operates exclusively through the fusion weights $\mathbf{w}_t$, which are updated at each timestep based on current sensor reliability assessments.

This design reflects a key insight: while the learned feature representations in f_θ remain useful across environmental changes, the *relative trustworthiness* of different sensor modalities varies substantially over time. A WiFi access point that provides reliable positioning information today may be removed tomorrow; an IMU sensor experiencing electromagnetic interference provides degraded motion estimates; a magnetometer near a new metallic installation exhibits distorted field readings. By adapting fusion weights rather than deep network parameters, the system can respond to these changes with minimal computational overhead while preserving the representational capacity learned during offline training.

The adaptation problem thus reduces to determining optimal fusion weights $\mathbf{w}_t$ at each timestep given current sensor reliability indicators $\rho_t = [\rho_1(t), \ldots, \rho_K(t)]^\top$. Section 5.4 defines how reliability indicators are computed from raw sensor measurements, and Section 5.5 presents the optimization algorithm that translates reliability assessments into adaptive weights.

5.2.2 Extended Loss Formulation

The training loss from Equation (4.43) defines the offline learning objective:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{pos}} + \beta_1 \mathcal{L}_{\text{KL}} + \beta_2 \mathcal{L}_{\text{rec}} \tag{5.3}$$

which optimizes position accuracy ($\mathcal{L}_{\text{pos}}$), latent space regularization ($\mathcal{L}_{\text{KL}}$), and sensor reconstruction fidelity ($\mathcal{L}_{\text{rec}}$). This composite loss governs the offline training of parameters θ and remains unchanged in the adaptive formulation; the base model training procedure is not modified.

OFAM operates at *deployment time*, not training time. Rather than modifying the training loss, OFAM introduces a separate optimization objective that governs the adaptation of fusion weights during online operation:

$$\mathcal{J}_{\text{OFAM}}(\mathbf{w}_t) = \sum_{i=1}^{K} (1 - \rho_i(t))\, w_i(t)^2 + \lambda_{\text{smooth}} \|\mathbf{w}_t - \mathbf{w}_{t-1}\|_2^2 \tag{5.4}$$

The OFAM objective in Equation (5.4) comprises two terms that balance responsiveness against stability:

Reliability-weighted penalty term: The first term $\sum_{i=1}^{K} (1 - \rho_i(t))\, w_i(t)^2$ penalizes assigning high weights to unreliable sensors. When sensor i exhibits high reliability ($\rho_i(t) \approx 1$), the penalty $(1 - \rho_i(t))$ approaches zero, allowing the corresponding weight $w_i(t)$ to be large. Conversely, when reliability degrades ($\rho_i(t) \approx 0$),

the penalty approaches unity, discouraging the system from relying on that sensor. The quadratic form encourages the optimization to distribute weight among multiple reliable sensors rather than concentrating on a single modality.

Smoothness regularization term: The second term $\lambda_{\text{smooth}} \|\mathbf{w}_t - \mathbf{w}_{t-1}\|_2^2$ penalizes rapid changes in fusion weights between consecutive timesteps. This regularization prevents oscillatory behavior that could arise from noisy reliability estimates, ensuring that the adaptation mechanism responds to genuine environmental changes rather than transient measurement fluctuations. The coefficient $\lambda_{\text{smooth}} = 0.1$ provides appropriate balance between responsiveness (ability to adapt quickly when conditions change) and stability (smooth weight trajectories during steady-state operation).

The OFAM objective is minimized subject to the constraints $\sum_{i=1}^{K} w_i(t) = 1$ (normalization) and $w_i(t) \geq 0$ (non-negativity). This constrained optimization is convex (a quadratic objective with linear constraints) and admits a unique global minimum; Section 5.5 approximates it with an efficient exponential-moving-average heuristic that avoids solving a KKT system at each timestep.

The key distinction between the training loss $\mathcal{L}_{\text{total}}$ and the OFAM objective $\mathcal{J}_{\text{OFAM}}$ is their scope and timing: the training loss governs offline learning of deep network parameters from labeled data, while the OFAM objective governs online adaptation of fusion weights from unlabeled sensor streams. This separation enables the system to benefit from both extensive offline training (which establishes robust feature representations) and continuous online adaptation (which maintains performance as conditions change).

5.3 Multi-Sensor Fusion Framework

Indoor localization systems that rely on multiple sensor modalities face the fundamental challenge of combining heterogeneous data streams with vastly different characteristics. WiFi RSSI measurements arrive at approximately 0.5 Hz with relatively high spatial resolution but significant temporal variability. IMU sensors provide motion information at 100 Hz with excellent short-term precision but suffer from cumulative drift. Magnetometer readings at 50 Hz offer absolute heading reference but are susceptible to environmental interference from ferromagnetic materials and electromagnetic sources. The multi-sensor fusion framework presented in this section addresses these challenges through a hierarchical architecture that operates at three distinct levels: feature-level fusion for combining raw sensor inputs, temporal-level fusion for capturing motion dynamics, and decision-level fusion for adaptive weighting based on sensor reliability.

The design philosophy prioritizes computational efficiency suitable for real-time mobile deployment while maintaining the flexibility to adapt to changing environmental conditions. Rather than treating fusion as a static optimization problem solved during training, the framework enables continuous adjustment of sensor contributions based on real-time reliability assessments. This adaptive capability distinguishes the approach from conventional fusion methods that assume fixed sensor characteristics throughout deployment.

5.3.1 Sensor Data Preprocessing and Synchronization

The sensor preprocessing pipeline follows the procedures established in Section 3.6, which handles modality-specific noise filtering, calibration correction, and feature extraction. Rather than re-deriving these preprocessing steps, this section focuses on the temporal synchronization mechanisms that enable coherent fusion of asynchronous sensor streams.

Temporal alignment presents a significant challenge when combining sensors with different sampling rates. The framework establishes WiFi scan timestamps as the primary temporal reference, since WiFi measurements directly drive position estimation at the coarsest granularity. IMU and magnetometer data are aggregated within 2-second integration windows centered on each WiFi scan, with 50% overlap between consecutive windows to ensure smooth temporal transitions. This window duration balances the need for sufficient IMU samples to characterize motion patterns against the requirement for temporal locality in position estimation.

Within each integration window, the framework computes aggregation statistics that capture both central tendency and variability of the high-frequency sensor streams:

$$\bar{\mathbf{u}}_t = \frac{1}{N_{\text{IMU}}} \sum_{k=1}^{N_{\text{IMU}}} \mathbf{u}_{t,k}, \quad \Sigma_t^{(\text{IMU})} = \frac{1}{N_{\text{IMU}} - 1} \sum_{k=1}^{N_{\text{IMU}}} (\mathbf{u}_{t,k} - \bar{\mathbf{u}}_t)(\mathbf{u}_{t,k} - \bar{\mathbf{u}}_t)^\top \tag{5.5}$$

$$\bar{\mathbf{m}}_t = \frac{1}{N_{\text{mag}}} \sum_{k=1}^{N_{\text{mag}}} \mathbf{m}_{t,k}, \quad \Sigma_t^{(\text{mag})} = \frac{1}{N_{\text{mag}} - 1} \sum_{k=1}^{N_{\text{mag}}} (\mathbf{m}_{t,k} - \bar{\mathbf{m}}_t)(\mathbf{m}_{t,k} - \bar{\mathbf{m}}_t)^\top \tag{5.6}$$

where $N_{\text{IMU}} \approx 200$ and $N_{\text{mag}} \approx 100$ samples fall within each 2-second window given the respective sampling rates. The variance terms provide uncertainty estimates that inform the reliability assessment mechanism described in Section 5.4.

Missing data handling follows a conservative strategy that maintains system stability when sensor readings become temporarily unavailable. When a sensor modality fails to produce valid measurements within an integration window, the framework substitutes the last valid measurement while increasing the associated uncertainty:

$$\tilde{\mathbf{s}}_t^{(i)} = \mathbf{s}_{t-\Delta}^{(i)}, \quad \hat{\sigma}_t^{(i)} = \sigma_{t-\Delta}^{(i)} \cdot (1 + \gamma_{\text{miss}} \Delta) \tag{5.7}$$

where Δ represents the number of missed updates and $\gamma_{\text{miss}} = 0.2$ controls the uncertainty growth rate. This approach prevents abrupt discontinuities in the fused output while appropriately reducing confidence in stale measurements.

5.3.2 Hierarchical Fusion Architecture

The hierarchical fusion architecture processes sensor data through three successive stages, each operating at a different level of abstraction. This design enables the system to leverage the complementary strengths of each sensor modality while maintaining computational tractability.

Feature-Level Fusion: The first fusion stage combines the preprocessed sensor measurements into a unified latent representation through the VAE encoder described in Section 4.3.2. The probabilistic encoding captures correlations between sensor modalities that would be lost by processing each stream independently. Specifically, the encoder receives the concatenated feature vector $\mathbf{s}_t = [\mathbf{r}_t; \bar{\mathbf{u}}_t; \bar{\mathbf{m}}_t]$ containing WiFi RSSI values, aggregated IMU statistics, and aggregated magnetometer readings (the per-modality variances $\text{diag}(\Sigma_t^{(\text{IMU})})$ and $\text{diag}(\Sigma_t^{(\text{mag})})$ are routed to the reliability assessment of Section 5.4 rather than the encoder). The VAE's reparameterization mechanism propagates uncertainty through the encoding process, yielding latent distributions $q(\mathbf{z}_t|\mathbf{s}_t) = \mathcal{N}(\mu_t, \text{diag}(\sigma_t^2))$ that explicitly represent feature-level uncertainty.

This approach differs fundamentally from early fusion methods that simply concatenate features before processing. By learning a joint latent representation, the VAE discovers shared structure across modalities, for example, the correlation between WiFi signal strength patterns and magnetometer anomalies near structural elements containing ferromagnetic materials. The reconstruction branch provides a mechanism for assessing representation quality, as described in Section 4.5.

Temporal Fusion: The second stage captures motion dynamics through the LSTM temporal encoder from Section 4.4.1. Sequential dependencies among latent representations $\mathbf{z}_{t-T+1:t}$ encode information about movement patterns that constrain the space of plausible positions. The temporal modeling follows the architecture detailed in Section 4.4.1, producing hidden states $\mathbf{h}_t$ that summarize both instantaneous sensor readings and recent motion history.

The LSTM's gating mechanisms provide natural weighting of temporal information based on learned relevance. During periods of rapid movement, the gates emphasize recent observations that reflect current position, while during stationary periods, longer-term patterns contribute more heavily to the representation. This adaptive temporal weighting complements the explicit sensor weighting mechanism at the decision level.

Decision-Level Fusion: The third and final stage implements adaptive combination of per-modality position contributions through time-varying weights. Unlike feature-level and temporal fusion, which are learned during training and remain fixed at inference time, decision-level fusion adapts continuously based on real-time reliability assessments. This is the key mechanism enabling the OFAM framework to respond to changing environmental conditions without retraining.

At each timestep, the decoder produces position estimates $\hat{\mathbf{x}}_t^{(i)}$ for each sensor modality $i \in \{1, \ldots, K\}$ through dedicated output heads. The final position estimate combines these contributions using reliability-weighted fusion:

$$\hat{\mathbf{x}}_t = \sum_{i=1}^{K} w_i(t) \cdot \hat{\mathbf{x}}_t^{(i)} \tag{5.8}$$

where the weights $w_i(t)$ satisfy the constraints $\sum_{i=1}^{K} w_i(t) = 1$ and $w_i(t) \geq 0$. The weight computation mechanism, detailed in Section 5.4, translates sensor reliability assessments into appropriate fusion coefficients. With $K = 3$ modalities (WiFi, IMU, magnetometer), this yields an interpretable decomposition of the final

estimate into constituent sensor contributions.

5.3.3 Parameterized Fusion Equation

The hierarchical fusion architecture admits a unified mathematical formulation through a single parameterized equation that encompasses both the weighted sensor combination and the balance between fusion levels. This formulation provides analytical insight into the adaptation behavior while maintaining practical implementability.

The fundamental fusion equation combines per-modality contributions through reliability-weighted averaging:

$$\hat{\mathbf{x}}_t = \sum_{i=1}^{K} w_i(t) \cdot \hat{\mathbf{x}}_t^{(i)} \tag{5.9}$$

where $\hat{\mathbf{x}}_t^{(i)}$ represents the position contribution from modality i, and $w_i(t)$ are time-varying weights satisfying $\sum_{i=1}^{K} w_i(t) = 1$ and $w_i(t) \geq 0$ for all i. This equation extends naturally from the decision-level fusion described above, with the constraint that weights form a proper probability distribution ensuring bounded output estimates.

To capture the interplay between feature-level and temporal fusion, the framework introduces a generalized hierarchical fusion equation parameterized by time scale ℓ:

$$\hat{\mathbf{x}}_t^{(\ell)} = \alpha_\ell \hat{\mathbf{x}}_t^{\text{feature}} + (1 - \alpha_\ell) \hat{\mathbf{x}}_t^{\text{temporal}} \tag{5.10}$$

where $\hat{\mathbf{x}}_t^{\text{feature}}$ denotes the position estimate derived primarily from instantaneous feature-level representations, $\hat{\mathbf{x}}_t^{\text{temporal}}$ denotes the estimate incorporating full temporal context, and $\alpha_\ell \in [0, 1]$ controls the balance between these contributions. The time-scale parameter ℓ indexes a spectrum of fusion strategies:

- At $\ell = 1$ (instantaneous): Feature-level dominates ($\alpha_\ell \approx 1$), emphasizing current sensor readings over temporal history. This setting is appropriate when WiFi signals are stable and provide reliable instantaneous position cues.
- At $\ell = T$ (full sequence): Temporal-level dominates ($\alpha_\ell \approx 0$), leveraging the complete motion history. This setting benefits scenarios where motion patterns provide stronger positioning constraints than potentially degraded instantaneous measurements.
- Intermediate values enable smooth interpolation between these extremes, adapting to varying environmental conditions.

The adaptation of α_ℓ based on motion state exploits a key observation: the relative utility of feature-level versus temporal fusion depends on whether the user is stationary or moving. During static periods, WiFi measurements provide stable position cues that benefit from minimal temporal smoothing; excessive reliance on motion history can actually degrade accuracy when the history contains outdated position information. Conversely, during dynamic periods, the temporal encoder captures motion patterns that constrain

position evolution, compensating for the transient variability in WiFi signals caused by body shadowing and orientation changes.

The motion-adaptive parameterization follows:

$$\alpha_\ell(t) = \text{sig}\left(c_0 + c_1 \cdot M(t)\right) \tag{5.11}$$

where $M(t) \in [0,1]$ is the motion magnitude indicator derived from IMU accelerometer variance, $\text{sig}(\cdot)$ is the sigmoid function, and c_0, c_1 are learned parameters that map motion state to the feature-temporal balance. With learned values $c_0 = 0.5$ and $c_1 = -2.0$, the system assigns higher feature-level weight ($\alpha \approx 0.62$) during static periods and shifts toward temporal fusion ($\alpha \approx 0.27$) during active movement.

5.3.4 Fusion Algorithm

The complete fusion procedure integrates the preprocessing, hierarchical encoding, and adaptive weighting mechanisms into a cohesive algorithm suitable for real-time execution. Algorithm 1 presents the hierarchical multi-sensor fusion procedure in pseudocode form, emphasizing the computational flow from raw sensor inputs to weighted position estimates.

Algorithm 1 proceeds through five distinct phases, each corresponding to a level in the fusion hierarchy:

Lines 2–3: Preprocessing: The high-frequency IMU and magnetometer streams are aggregated within the 2-second integration window using the statistics defined in Equations (5.5) and (5.6). WiFi RSSI measurements arrive at the WiFi scan rate (approximately 0.5 Hz) and require no aggregation beyond the preprocessing described in Section 3.6.

Lines 5–6: Feature-level fusion: The VAE encoder processes the concatenated sensor vector to produce the latent representation $\mathbf{z}_t$. The encoder architecture and reparameterization procedure follow Section 4.3.2; these mechanics are not re-explained here, but note that the stochastic sampling introduces beneficial regularization during training while collapsing to deterministic inference (using the mean μ_t) at deployment time.

Lines 7–8: Temporal fusion: The LSTM encoder maintains hidden state across the $T = 10$ timestep sequence window, producing the temporally-contextualized representation $\mathbf{h}_t$. The architecture details appear in Section 4.4.1.

Lines 9–11: Per-modality decoding: Dedicated decoder heads transform the shared representation into modality-specific position estimates. Each head consists of a two-layer feedforward network that maps from the hidden state dimension (128) to the 2-dimensional position output. The separate heads enable the

Algorithm 1 Hierarchical Multi-Sensor Fusion

Require: Sensor data $\mathbf{s}_t = [\mathbf{r}_t, \mathbf{u}_t, \mathbf{m}_t]$, current weights $\mathbf{w}_{t-1}$, motion indicator $M(t)$
Ensure: Position estimate $\hat{\mathbf{x}}_t$, updated weights $\mathbf{w}_t$

1: **// Preprocessing (reference Section 3.6)**
2: $[\bar{\mathbf{u}}_t, \Sigma_t^{(\text{IMU})}] \leftarrow \text{AggregateIMU}(\mathbf{u}_{t-W:t})$
3: $[\bar{\mathbf{m}}_t, \Sigma_t^{(\text{mag})}] \leftarrow \text{AggregateMag}(\mathbf{m}_{t-W:t})$

4: **// Feature-level fusion (VAE encoder, Section 4.3.2)**
5: $\mathbf{z}_t \leftarrow \text{VAE-Encode}(\mathbf{s}_t)$

6: **// Temporal fusion (LSTM encoder, Section 4.4.1)**
7: $\mathbf{h}_t \leftarrow \text{LSTM-Process}(\mathbf{z}_{t-T+1:t})$

8: **// Compute per-modality position contributions**
9: **for** each sensor modality $i \in \{1, \ldots, K\}$ **do**
10: $\quad \hat{\mathbf{x}}_t^{(i)} \leftarrow \text{Decoder}_i(\mathbf{h}_t)$
11: **end for**
12: $\hat{\mathbf{s}}_t \leftarrow \text{ReconDecoder}(\mathbf{h}_t)$

13: **// Reliability assessment (Section 5.4)**
14: $\rho_t \leftarrow \text{ComputeReliability}(\mathbf{s}_t, \hat{\mathbf{s}}_t)$

15: **// Adaptive weight update (Section 5.5)**
16: $\mathbf{w}_t \leftarrow \text{OFAM-Update}(\mathbf{w}_{t-1}, \rho_t)$

17: **// Decision-level weighted fusion**
18: $\hat{\mathbf{x}}_t \leftarrow \sum_{i=1}^{K} w_i(t) \cdot \hat{\mathbf{x}}_t^{(i)}$

19: **return** $\hat{\mathbf{x}}_t$, $\mathbf{w}_t$

system to learn how each sensor modality contributes to position estimation, providing the decomposition required for interpretable adaptive weighting.

Lines 12–15: Reliability and adaptation: The reliability computation (Line 13) assesses current sensor trustworthiness based on signal characteristics and the decoder's reconstruction quality; the detailed metrics are presented in Section 5.4. The OFAM weight update (Line 15) translates reliability assessments into fusion weights using the optimization procedure described in Section 5.5. For the present discussion, it suffices to note that weights are available for the subsequent fusion step; the mechanism by which they are computed is deferred to the OFAM section.

Lines 16–17: Decision-level fusion: The final position estimate combines per-modality contributions according to the weighted sum in Equation (5.9). This simple linear combination, with weights constrained to form a probability distribution, produces the interpretable decomposition described earlier: each weight $w_i(t)$ directly indicates the fraction of the position estimate attributable to sensor modality i.

Algorithm 1 executes within the latency budget required for real-time operation. Profiling on the target mobile platform shows the following timing breakdown: sensor collection and preprocessing (12 ms), VAE encoding (18 ms), LSTM processing (15 ms), sensor fusion and per-modality decoding (8 ms), and online adaptation comprising reliability computation and weight update (5 ms), yielding a total of 58 ms per position estimate in online mode (Section 5.6.5). This performance enables the 10 Hz update rate required for augmented reality navigation while leaving computational headroom for parallel rendering operations.

5.4 Sensor Reliability Assessment

The adaptive fusion framework introduced in Section 5.3 relies on real-time assessment of sensor reliability to determine appropriate fusion weights. Each sensor modality employed in indoor localization exhibits distinct failure modes and degradation characteristics that must be detected and quantified for effective adaptation. WiFi RSSI measurements suffer from infrastructure changes, multipath propagation effects, and interference from competing wireless systems. IMU sensors accumulate drift errors during motion integration and experience calibration degradation over time. Magnetometer readings are susceptible to both hard-iron distortions from ferromagnetic materials in building structures and soft-iron effects from nearby electronic equipment.

This section presents formal reliability metrics for each sensor modality that translate raw measurement characteristics into normalized indicators suitable for driving the adaptive weight computation described in Section 5.5. The design of these metrics reflects a key principle: reliability assessment must be *self-supervised*, relying only on observable sensor characteristics rather than ground truth position labels that are unavailable during deployment. Each metric produces values in the range $[0, 1]$, where higher values indicate greater trustworthiness of the corresponding sensor modality for position estimation.

5.4.1 WiFi RSSI Stability Metrics

WiFi-based positioning relies on the assumption that received signal strength from access points exhibits consistent spatial patterns that can be matched against a pre-recorded radio map. When this assumption fails (due to access point removal, interference from new wireless devices, or transient multipath effects from moving obstacles), positioning accuracy degrades rapidly. The WiFi reliability metric must detect these degradation conditions from observable signal characteristics without access to position ground truth.

The WiFi reliability indicator combines two complementary factors: signal variance that captures measurement stability, and access point visibility that captures infrastructure availability:

$$\rho_{\text{WiFi}}(t) = \exp\left(-\frac{\sigma^2_{\text{RSSI}}(t)}{\sigma^2_{\text{ref}}}\right) \cdot \min\left(1, \frac{N_{\text{visible}}(t)}{N_{\text{expected}}}\right) \tag{5.12}$$

The exponential term quantifies signal stability through the ratio of current RSSI variance $\sigma^2_{\text{RSSI}}(t)$ to a reference variance σ^2_{ref} established during system training. The current variance is computed over a sliding window of the most recent $W = 30$ WiFi scans:

$$\sigma^2_{\text{RSSI}}(t) = \frac{1}{N_{\text{ap}}} \sum_{j=1}^{N_{\text{ap}}} \text{Var}\left(\{r_j^{(k)}\}_{k=t-W+1}^{t} \right) \tag{5.13}$$

where $r_j^{(k)}$ denotes the RSSI measurement from access point j at time k, and the variance is averaged across all N_{ap} access points currently in the observation set. The reference variance $\sigma^2_{\text{ref}} = 25$ dBm^2 represents typical RSSI fluctuation under stable conditions, determined empirically from the training dataset. When multipath effects or interference increase signal variability, $\sigma^2_{\text{RSSI}}(t)$ exceeds the reference, driving the exponential factor toward zero and reducing the reliability indicator.

The visibility factor $N_{\text{visible}}(t)/N_{\text{expected}}$ captures infrastructure-level changes that affect positioning capability. The numerator $N_{\text{visible}}(t)$ counts access points detected in the current scan, while the denominator N_{expected} represents the anticipated number based on the radio map constructed during site survey. Capped at unity by the $\min(1,\cdot)$ operation, this factor equals one when all expected access points remain visible and decreases proportionally as access points become unavailable due to hardware failure, power management, or physical removal; the cap prevents transient additional access points from pushing the indicator above one, keeping $\rho_{\text{WiFi}}(t) \in [0,1]$. The multiplicative combination of variance and visibility factors ensures that both signal quality and infrastructure integrity must be satisfactory for high reliability assessment.

5.4.2 IMU Drift and Motion Consistency

Inertial measurement units provide high-rate motion information that is essential for trajectory smoothing and dead reckoning between WiFi scans. However, IMU reliability varies substantially with motion context: during static periods, accelerometer readings accurately reflect gravitational acceleration with minimal noise, enabling precise orientation estimation. During dynamic periods involving rapid rotation, gyroscope integration errors accumulate, and accelerometer readings become contaminated with motion artifacts that degrade orientation estimates.

The IMU reliability metric employs a context-dependent formulation that distinguishes between static and dynamic motion states:

$$\rho_{\text{IMU}}(t) = \begin{cases} 1.0 & \text{if } \|\mathbf{a}_t - \mathbf{g}\| < \tau_{\text{static}} \\ 0.8 \cdot \exp\left(-\kappa \|\boldsymbol{\omega}_t\|\right) & \text{otherwise} \end{cases} \tag{5.14}$$

The static detection condition $\|\mathbf{a}_t - \mathbf{g}\| < \tau_{\text{static}}$ compares the measured acceleration vector $\mathbf{a}_t$ against the expected gravitational acceleration $\mathbf{g} = [0,0,9.81]^\top$ m/s^2. When the difference magnitude falls below the threshold $\tau_{\text{static}} = 0.1$ m/s^2, the device is considered stationary, and IMU reliability reaches its maximum value of 1.0. This reflects the physical reality that consumer-grade MEMS accelerometers achieve excellent accuracy during static conditions when no dynamic forces are present.

During motion, the reliability decreases exponentially with angular velocity magnitude $\|\omega_t\|$. The decay rate parameter $\kappa = 0.5$ rad^{-1}s controls the sensitivity to rotational motion: higher angular velocities correspond to more rapid orientation changes that amplify gyroscope integration errors. The baseline factor of 0.8 acknowledges that even during minimal rotation, dynamic motion introduces additional uncertainty compared to perfectly static conditions. The exponential formulation ensures smooth degradation rather than abrupt transitions that could cause oscillatory weight behavior.

The rationale for this context-dependent design stems from the physics of inertial navigation. During static periods, the accelerometer provides a direct measurement of the gravity vector, enabling accurate tilt estimation without integration. During motion, both accelerometer and gyroscope readings must be integrated to track orientation changes, introducing cumulative errors that grow with motion intensity and duration. By assigning higher reliability during static periods, the fusion system appropriately emphasizes IMU contributions when they are most trustworthy.

5.4.3 Magnetometer Field Disturbance

Magnetometers provide heading information by measuring the local magnetic field vector, which in undisturbed outdoor environments points toward magnetic north with magnitude approximately 45–65 μT depending on geographic location. Indoor environments present significant challenges for magnetometer-based heading due to both permanent disturbances from building structures (hard-iron effects) and time-varying disturbances from electronic equipment, elevators, and HVAC systems (soft-iron effects). The magnetometer reliability metric must detect these disturbance conditions to prevent corrupted heading information from degrading fusion quality.

The magnetometer reliability indicator combines magnitude consistency and temporal stability:

$$\rho_{\text{mag}}(t) = \min\left(1, \frac{\|\mathbf{m}_t\|}{\|\mathbf{m}_{\text{ref}}\|}\right) \cdot \max\left(0, 1 - \gamma_{\text{mag}}\sigma^2_{\text{mag}}(t)\right) \tag{5.15}$$

The magnitude ratio $\|\mathbf{m}_t\|/\|\mathbf{m}_{\text{ref}}\|$ compares the current magnetic field magnitude against a reference value $\|\mathbf{m}_{\text{ref}}\| = 50$ μT established during magnetometer calibration. Strong ferromagnetic interference (such as proximity to steel structural elements, electrical panels, or large metal objects) typically causes substantial deviations in field magnitude. The $\min(1, \cdot)$ operation caps the ratio at unity to handle cases where interference increases the measured magnitude beyond the reference value, which would otherwise produce reliability values exceeding the normalized range.

The temporal stability factor $(1 - \gamma_{\text{mag}}\sigma^2_{\text{mag}}(t))$ penalizes high variance in magnetic field measurements over a sliding window. The variance $\sigma^2_{\text{mag}}(t)$ captures fluctuating disturbances from sources such as elevator motors, HVAC fans, and moving metal carts that produce time-varying magnetic fields. The sensitivity parameter $\gamma_{\text{mag}} = 0.1$ (μT)$^{-2}$ scales the variance contribution appropriately: a variance of 10 (μT)^2 drives the stability factor to zero, and the $\max(0, \cdot)$ clamp holds it there for larger variances, keeping $\rho_{\text{mag}}(t) \in [0, 1]$, while lower variance values have proportionally smaller effects. This design ensures that both static disturbances (detected through magnitude deviations) and dynamic disturbances (detected through temporal variance) reduce magnetometer reliability appropriately.

The multiplicative combination of magnitude and stability factors reflects the requirement that *both* conditions must be satisfied for reliable magnetometer operation. A measurement with correct magnitude but high variance indicates unstable local magnetic conditions, while a measurement with low variance but anomalous magnitude suggests persistent interference from nearby ferromagnetic sources. Either condition compromises the utility of magnetometer readings for heading estimation.

5.4.4 Combined Reliability Vector

The signal-based metrics of Equations (5.12), (5.14) and (5.15) capture degradation that is visible in the raw measurement statistics. They are complemented by a representation-level signal that exposes distribution shifts which leave the per-sensor statistics nominal but move the measurements off the manifold learned during training: the reconstruction quality of the dual-branch decoder (Section 4.5). For each modality i, the reconstruction error over its slice of the sensor vector is

$$e_i(t) = \frac{1}{D_i} \left\| \mathbf{s}_t^{(i)} - \hat{\mathbf{s}}_t^{(i)} \right\|_2^2, \tag{5.16}$$

where $\mathbf{s}_t^{(i)}$ and $\hat{\mathbf{s}}_t^{(i)}$ are the measured and reconstructed sub-vectors for modality i (WiFi, IMU, or magnetometer) and D_i is its dimensionality. This error is mapped to a reconstruction-quality factor

$$q_i(t) = \exp\left(-\frac{e_i(t)}{\varepsilon_i} \right) \in (0, 1], \tag{5.17}$$

where ε_i is the mean reconstruction error for modality i over the training set, so that $q_i(t) \approx 1$ under in-distribution conditions and decays toward zero as the modality drifts. Gating each signal-based metric by its reconstruction-quality factor yields the combined reliability vector that serves as the input to the Online Feature Adaptation Module described in Section 5.5:

$$\rho_t = \begin{bmatrix} \rho_{\text{WiFi}}(t)\, q_{\text{WiFi}}(t) \\ \rho_{\text{IMU}}(t)\, q_{\text{IMU}}(t) \\ \rho_{\text{mag}}(t)\, q_{\text{mag}}(t) \end{bmatrix} \tag{5.18}$$

A modality is thus deemed reliable only when both its raw statistics and its reconstruction are consistent with training-time behavior: a distribution shift that degrades reconstruction reduces the modality's reliability, and hence its fusion weight, even when its signal statistics appear nominal.

This three-dimensional vector provides a complete characterization of the current sensor environment, enabling the OFAM algorithm to adjust fusion weights based on the relative trustworthiness of each modality. The reliability vector is updated at the fusion rate of approximately 0.5 Hz, synchronized with WiFi scan arrivals. Although IMU and magnetometer measurements arrive at higher rates (100 Hz and 50 Hz respectively), the reliability computation operates on aggregated statistics over the 2-second integration window described in Section 5.3.1, ensuring computational efficiency while capturing the relevant degradation characteristics.

The normalization of individual reliability metrics to the $[0, 1]$ range enables direct comparison

across sensor modalities despite their fundamentally different measurement characteristics. When all sensors operate normally, the reliability vector components are approximately equal, and the fusion system distributes weight among modalities according to their baseline contributions learned during training. When a specific sensor experiences degradation (for example, WiFi reliability dropping due to access point removal), the corresponding component decreases while others remain stable, signaling the OFAM algorithm to reduce weight on the affected modality. This self-supervised reliability assessment enables continuous adaptation without requiring position labels or user intervention, addressing the fundamental challenge of maintaining localization accuracy across year-over-year deployment gaps in dynamic indoor environments.

5.5 Online Feature Adaptation Module

The sensor reliability metrics established in Section 5.4 provide instantaneous assessments of each modality's trustworthiness for position estimation. The Online Feature Adaptation Module (OFAM) translates these reliability indicators into fusion weights that govern how sensor contributions are combined in the final position estimate. This section presents the complete OFAM formulation, including the optimization objective, the adaptive weight update mechanism, the full algorithm specification, and a discussion of empirical convergence behavior.

5.5.1 Design Rationale

The design of OFAM reflects several key requirements that distinguish online adaptation from traditional offline learning approaches. First, the adaptation mechanism must operate *without ground truth labels*, since position measurements are unavailable during normal system operation. Second, adaptation must be *computationally lightweight* to enable real-time execution on resource-constrained mobile devices. Third, the mechanism must provide *stable adaptation* that responds to genuine environmental changes without oscillating due to transient measurement noise.

These requirements motivate the fundamental architectural choice of adapting fusion weights rather than neural network parameters. Consider the alternatives:

Online learning of full model parameters: Updating the deep network weights θ during deployment would enable the most flexible adaptation but incurs prohibitive costs: backpropagation through the VAE-LSTM architecture requires substantial computation and memory, weight updates without supervision risk catastrophic forgetting of learned representations, and the system would need to maintain training infrastructure (optimizer state, gradient buffers) during deployment. Furthermore, meaningful parameter updates require many samples, introducing unacceptable latency for responding to environmental changes.

Periodic offline retraining: An alternative approach would collect deployment data for periodic retraining sessions. This requires ground truth labels for the collected data, imposes system downtime during retrain-

ing, and introduces discontinuous transitions when updated models are deployed. The approach cannot respond to gradual environmental changes that occur between retraining cycles.

Online gradient methods and online sequential learning: A lighter variant of full online learning updates only part of the network, or replaces backpropagation with a closed-form sequential update, as in the online sequential extreme learning machine (OS-ELM) [52] and its application to indoor localization [53]. Both variants still require a supervisory signal on the device: a gradient step needs a loss evaluated against a target, and the OS-ELM recursive least-squares update needs labeled arrivals. Neither is available during unlabeled deployment. In addition, sequential parameter updates carry the risk of forgetting the representation learned offline [61], a risk that grows with the length of the deployment.

Transfer learning and domain adaptation: Transfer learning and unsupervised domain adaptation reduce the labeling burden in a new environment by aligning source and target feature distributions [58, 59, 47]. These are offline procedures: they require a batch of target-domain observations to be collected before adaptation, and they retrain network parameters against that batch. They address the cost of moving to a new site rather than the gradual drift that builds up within a site after deployment, and they cannot respond to a change that occurs between two adaptation rounds.

Test-time adaptation: Test-time adaptation methods such as TENT [62], SHOT [63], and test-time training [64] update network parameters on unlabeled test data: the affine parameters of normalization layers in TENT, the feature extractor in SHOT, and the shared encoder in test-time training. TENT and SHOT minimize the entropy of the class posterior. A regression head that outputs coordinates has no categorical distribution and so no entropy objective to minimize. Test-time training substitutes a self-supervised auxiliary task, which is closer to the reconstruction signal used here, but the update remains a gradient step on the device, with the memory and forgetting concerns noted above. TENT and SHOT are also formulated for batches of test inputs rather than for a stream that is scored one window at a time.

Radio-map maintenance and crowdsourcing: A separate line of work keeps the WiFi fingerprint database current rather than adapting the model: collaborative updates contributed by users [55], autonomous radio-map update from user traces [57], and explicit correction of radio-map degradation [39]. These methods operate on a server, require user traces to be uploaded, and update the map. They complement rather than replace the mechanism developed here: an updated map improves the WiFi observation, whereas OFAM decides how much to trust that observation relative to the inertial and magnetic modalities.

Adaptive fusion from instantaneous residuals: Adaptive Kalman filtering adjusts measurement-noise covariances, or gates measurements, from the innovation at the current step, as in UWB/IMU fusion [65] and WiFi/vision fusion [66]. This is the closest alternative, since it also re-weights modalities without labels.

The adjustment, however, is a function of the current innovation or of a short run of innovations, and it keeps no record of how a modality's trustworthiness has drifted over the preceding minutes or days. OFAM instead accumulates reliability over a sliding window, gates it by reconstruction quality, and smooths the resulting weights with a decaying learning rate, so that persistent drift is tracked while transient disturbances are damped.

Section 2.4.2 compares these families against the three requirements listed at the start of this subsection.

Fusion weight adaptation (OFAM approach): Adapting only the $K = 3$ fusion weights provides a middle ground that addresses the practical requirements. The computational cost is $O(K)$ per update, negligible compared to neural network inference. The adaptation operates on reliability indicators derived from sensor measurements, requiring no position labels. The exponential moving average update provides smooth transitions that prevent oscillatory behavior. Importantly, the learned feature representations in the base VAE-LSTM architecture remain preserved, since the network parameters θ are never modified during deployment.

The connection to reliability assessment is direct: when a sensor modality exhibits degraded reliability (low $\rho_i(t)$), its fusion weight should decrease to reduce that modality's contribution to the position estimate. The OFAM objective formalizes this intuition through an optimization formulation that balances reliability-following behavior against temporal smoothness.

The novelty of the approach does not lie in the combination of a variational autoencoder with an LSTM as such. The compact VAE-LSTM is used because its reconstruction branch yields a reliability signal that needs no labels, while its LSTM captures the motion context that constrains position between WiFi scans. OFAM uses that signal to adapt the fusion weights with no labels, no gradients, and no retraining. The contribution of this chapter is therefore the adaptation mechanism and its demonstrated temporal robustness in the cross-year and cross-dataset evaluations of Section 5.6, rather than the network components on which the mechanism operates.

5.5.2 Optimization Formulation

The OFAM objective introduced in Equation (5.4) encodes the principle that fusion weights should track sensor reliability while maintaining temporal stability. This section presents the complete constrained optimization formulation that governs weight adaptation at each timestep.

The optimal fusion weights $\mathbf{w}_t^*$ are defined as the solution to the following optimization problem:

$$\mathbf{w}_t^* = \arg\min_{\mathbf{w}} \sum_{i=1}^{K} (1 - \rho_i(t)) w_i^2 + \lambda_{\text{smooth}} \|\mathbf{w} - \mathbf{w}_{t-1}\|_2^2 \tag{5.19}$$

subject to the constraints:

$$\sum_{i=1}^{K} w_i = 1 \quad \text{(normalization)} \tag{5.20}$$

$$w_i \geq 0 \quad \forall i \in \{1, \ldots, K\} \quad \text{(non-negativity)} \tag{5.21}$$

The objective function comprises two terms that encode complementary desiderata:

Reliability-weighted penalty: The first term $\sum_{i=1}^{K}(1-\rho_i(t))w_i^2$ penalizes assigning high weights to unreliable sensors. The factor $(1-\rho_i(t))$ inverts the reliability indicator: when sensor i is highly reliable ($\rho_i(t) \approx 1$), the penalty coefficient approaches zero, permitting large weights. When reliability degrades ($\rho_i(t) \approx 0$), the penalty coefficient approaches unity, discouraging reliance on that modality. The quadratic form w_i^2 provides soft enforcement that distributes weight among multiple reliable sensors rather than concentrating entirely on the single most reliable modality, which would create fragile behavior when that modality experiences transient degradation.

Smoothness regularization: The second term $\lambda_{\text{smooth}}\|\mathbf{w}-\mathbf{w}_{t-1}\|_2^2$ penalizes deviation from the previous weight vector, encouraging smooth temporal evolution of fusion weights. This regularization addresses the practical concern that reliability estimates are inherently noisy due to limited sample windows and sensor measurement noise. Without smoothness regularization, the optimization would produce weight trajectories that oscillate rapidly in response to reliability fluctuations, degrading positioning performance and creating interpretability challenges. The coefficient $\lambda_{\text{smooth}} = 0.1$ balances adaptation responsiveness against stability: larger values produce smoother but slower adaptation, while smaller values enable faster response at the cost of potential oscillation.

The normalization constraint (5.20) ensures that the fusion weights form a proper probability distribution over sensor modalities. This constraint enables interpretable decomposition of the position estimate: each weight $w_i(t)$ directly indicates the fractional contribution of modality i to the final output. The non-negativity constraint (5.21) ensures that all contributions are additive, preventing pathological solutions where negative weights cause one modality to subtract from another's contribution.

The constrained optimization problem defined by (5.19)–(5.21) is convex (quadratic objective with linear constraints), admitting a unique global minimum. However, computing the exact closed-form solution requires solving a KKT system at each timestep, which introduces unnecessary complexity for the lightweight adaptation setting. Instead, the practical OFAM implementation employs a lightweight heuristic, based on exponential-moving-average updates with softmax normalization, that drives the weights toward the normalized reliability $\tilde{\rho}$. This surrogate shares the design principle of the convex program, favoring reliable modalities, while executing in $O(K)$ time per timestep. The update mechanism is described in the following subsection.

5.5.3 Adaptive Weight Update Mechanism

The practical weight update mechanism approximates the constrained optimization through a two-stage procedure: an exponential moving average (EMA) update that tracks normalized reliability, followed by softmax projection that enforces the constraints.

The EMA update rule adjusts each weight toward the normalized reliability value:

$$w_i(t) = (1 - \eta_t) w_i(t-1) + \eta_t \tilde{\rho}_i(t) \tag{5.22}$$

where $\tilde{\rho}_t = \rho_t / \sum_{j=1}^{K} \rho_j(t) q_j(t)$ denotes the normalized reliability vector. The normalization ensures that $\sum_i \tilde{\rho}_i(t) = 1$, making the normalized reliability directly interpretable as a target weight distribution.

The adaptive learning rate η_t follows an exponential decay schedule:

$$\eta_t = \eta_0 \exp(-t/\tau) + \eta_{\min} \tag{5.23}$$

where $\eta_0 = 0.3$ is the amplitude of the decaying component (so the initial rate is $\eta_0 + \eta_{\min} = 0.35$), $\tau = 100$ timesteps is the decay time constant, and $\eta_{\min} = 0.05$ is the learning rate floor. This schedule provides fast initial adaptation when the system first encounters a new environment (high η_t when t is small), transitioning to slower adaptation for fine-tuning as the system approaches steady-state operation. The floor $\eta_{\min}$ ensures that the system retains some adaptation capability even after extended deployment, enabling response to gradual environmental drift.

The rationale for the decay schedule reflects the expected deployment scenario: when a user enters a new building or area, sensor characteristics may differ substantially from those seen during training, requiring rapid adaptation to establish appropriate fusion weights. As the system operates in a consistent environment, the reliability estimates stabilize, and the weights should evolve slowly to maintain stability. The floor prevents complete adaptation cessation, which would cause the system to become rigid and unable to respond to eventual environmental changes.

After the EMA update, softmax projection enforces the normalization and non-negativity constraints:

$$w_i(t) \leftarrow \frac{\exp(w_i(t))}{\sum_{j=1}^{K} \exp(w_j(t))} \tag{5.24}$$

The softmax operation provides a smooth approximation to the simplex projection required by the constraints. Unlike hard projection, which would set small weights exactly to zero and could cause discontinuous behavior, softmax maintains differentiability and produces strictly positive weights. This ensures that all sensor modalities retain some contribution to the position estimate, enabling recovery if a previously unreliable sensor becomes reliable again.

The complete update mechanism thus combines reliability tracking (through EMA) with constraint enforcement (through softmax), providing an efficient approximation to the full optimization problem that executes in $O(K)$ time per timestep.

Algorithm 2 Online Feature Adaptation

Require: Current sensor data $\mathbf{s}_t$, reconstruction $\hat{\mathbf{s}}_t$, previous weights $\mathbf{w}_{t-1}$, timestep t
Ensure: Updated weights $\mathbf{w}_t$

1: **// Signal-based reliability indicators (Section 5.4)**
2: $\rho_{\text{WiFi}}(t) \leftarrow \exp\left(-\sigma^2_{\text{RSSI}}(t)/\sigma^2_{\text{ref}}\right) \cdot \min(1, N_{\text{visible}}(t)/N_{\text{expected}})$
3: $\rho_{\text{IMU}}(t) \leftarrow \begin{cases} 1.0 & \text{if } \|\mathbf{a}_t - \mathbf{g}\| < \tau_{\text{static}} \\ 0.8 \cdot \exp(-\kappa\|\omega_t\|) & \text{otherwise} \end{cases}$
4: $\rho_{\text{mag}}(t) \leftarrow \min(1, \|\mathbf{m}_t\|/\|\mathbf{m}_{\text{ref}}\|) \cdot \max(0, 1 - \gamma_{\text{mag}}\sigma^2_{\text{mag}}(t))$

5: **// Reconstruction-quality gate (distribution-shift detection)**
6: $q_i(t) \leftarrow \exp\big(-\|\mathbf{s}_t^{(i)} - \hat{\mathbf{s}}_t^{(i)}\|_2^2/(D_i\varepsilon_i)\big)$ for $i \in \{\text{WiFi}, \text{IMU}, \text{mag}\}$
7: $\rho_t \leftarrow [\rho_{\text{WiFi}}(t)\, q_{\text{WiFi}}(t),\ \rho_{\text{IMU}}(t)\, q_{\text{IMU}}(t),\ \rho_{\text{mag}}(t)\, q_{\text{mag}}(t)]^\top$

8: **// Normalize reliability to target distribution**
9: $\tilde{\rho}_t \leftarrow \rho_t / \sum_{i=1}^{K} \rho_i(t)\, q_i(t)$

10: **// Compute adaptive learning rate**
11: $\eta_t \leftarrow \eta_0 \exp(-t/\tau) + \eta_{\min}$

12: **// EMA weight update**
13: **for** each sensor modality $i \in \{1, \ldots, K\}$ **do**
14: $\quad w_i(t) \leftarrow (1 - \eta_t) \cdot w_i(t-1) + \eta_t \cdot \tilde{\rho}_i(t)$
15: **end for**

16: **// Softmax projection for constraint satisfaction**
17: $\mathbf{w}_t \leftarrow \text{softmax}(\mathbf{w}_t)$ $\quad \triangleright\ w_i(t) \leftarrow \exp(w_i(t))/\sum_{j=1}^{K} \exp(w_j(t))$

18: **return** $\mathbf{w}_t$

5.5.4 OFAM Algorithm

Algorithm 2 presents the complete Online Feature Adaptation procedure in pseudocode form, integrating the reliability computation from Section 5.4 with the weight update mechanism described above.

Algorithm 2 proceeds through four distinct phases:

Lines 2–6: Reliability computation: The sensor reliability metrics from Section 5.4 are computed from the current sensor measurements. The WiFi reliability (Line 2) captures signal stability and infrastructure availability. The IMU reliability (Line 3) distinguishes between static and dynamic motion contexts. The magnetometer reliability (Line 4) assesses magnetic field consistency. These metrics are assembled into the reliability vector ρ_t (Line 5).

Lines 7–8: Normalization: The reliability vector is normalized to form a probability distribution $\tilde{\rho}_t$ that serves as the target for weight adaptation. This normalization ensures that the weights converge toward a distribution that reflects relative reliability rather than absolute reliability values.

Lines 9–10: Learning rate computation: The adaptive learning rate follows the exponential decay schedule from Equation (5.23), providing fast initial adaptation that decays toward the floor value $\eta_{\min}$ over time.

Lines 11–14: EMA update: Each weight is updated according to the exponential moving average rule, blending the previous weight with the normalized reliability target. The learning rate η_t controls the adaptation speed.

Lines 15–16: Constraint projection: The softmax operation ensures that the final weights satisfy the normalization constraint $\sum_i w_i(t) = 1$ and non-negativity constraint $w_i(t) > 0$.

Computational complexity: Algorithm 2 executes in $O(K)$ time per timestep, where $K = 3$ is the number of sensor modalities. The reliability computations (Lines 2–5) involve constant-time operations per modality. The normalization, learning rate computation, EMA update, and softmax projection are all $O(K)$ operations. With $K = 3$ modalities, the complete OFAM update requires approximately 20 floating-point operations, completing in under 0.5 ms on the target mobile platform.

Figure 5.2 visualizes the closed-loop structure that the pseudocode in Algorithm 2 formalizes: the deployment-time signal path on the left produces position estimates, while the parallel reliability-driven feedback path on the right adjusts the fusion weights that gate each modality's contribution. This decoupling of position estimation from weight adaptation is what enables OFAM to update its sensor reliance without disturbing the learned VAE-LSTM parameters.

5.5.5 Empirical Convergence Behavior

Convergence of the fusion weights is established empirically in this thesis, by the operational criterion of Section 5.7.6 (weight change below 0.01 for 10 consecutive epochs) and by the bounded post-convergence behavior reported in Section 5.7.7. No formal convergence guarantee is claimed. The following paragraphs explain why the update is expected to behave well and where the classical conditions fail.

The EMA update rule in Equation (5.22) follows the form of classical stochastic approximation algorithms. Under standard regularity conditions (bounded reliability estimates and positive learning rates satisfying $\sum_t \eta_t = \infty$ and $\sum_t \eta_t^2 < \infty$), such algorithms are expected to approach a stable fixed point that closely tracks the normalized reliability $\tilde{\rho}_i$ once the reliability estimates stabilize [139]. Because the softmax projection preserves the ordering of the exponential-moving-average weights, the converged weights remain a monotone function of $\tilde{\rho}_i$: a more reliable modality therefore always receives a larger weight.

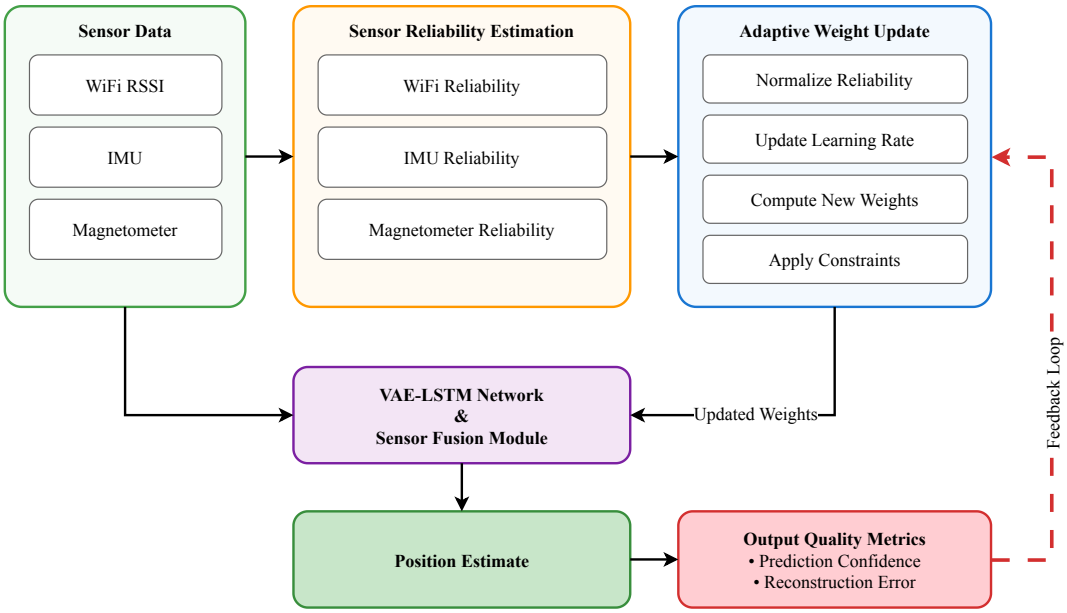


Figure 5.2: End-to-end workflow of the Online Feature Adaptation Module. Multi-modal sensor streams feed the VAE-LSTM backbone, whose deployment-time outputs are routed in parallel to the reliability-estimation block and the position decoder. The reliability indicators drive the weight-update rule of Algorithm 2, and the updated fusion weights are fed back to the multi-sensor fusion stage, closing the adaptation loop without requiring ground-truth position labels.

The learning rate schedule in Equation (5.23) does not satisfy the classical convergence conditions exactly: although $\sum_t \eta_t = \infty$ holds, the square-summability requirement $\sum_t \eta_t^2 < \infty$ fails because the floor $\eta_{\min}$ keeps $\eta_t^2 \geq \eta_{\min}^2 > 0$ for all t. However, this deliberate design choice reflects the practical requirement for continuous adaptation: strict convergence would cause the system to become rigid and unable to respond to environmental changes that occur after the initial adaptation period. The floor $\eta_{\min} = 0.05$ ensures that the system retains 5% adaptation capability indefinitely, enabling gradual response to slow environmental drift while providing stable operation during normal conditions.

Empirical evaluation across the York University testbed demonstrates that OFAM weights converge to stable values within seconds—well under a minute—of entering a new environment (Section 5.7.6 provides the detailed convergence analysis), with subsequent fluctuations bounded within ± 0.05 of the converged values under steady-state conditions. The convergence time is consistent across the four buildings in the evaluation dataset despite their differing infrastructure densities and sensor reliability characteristics. Formal convergence analysis is beyond the scope of this thesis; empirical convergence behavior is demonstrated through the experimental validation presented in Section 5.6.

The stability of converged weights under environmental perturbation provides an important practical benefit: when a transient disturbance affects a single sensor modality (for example, a person walking between the device and a WiFi access point), the EMA smoothing prevents abrupt weight changes that would propagate into position estimate discontinuities. Only sustained reliability changes lasting multiple seconds trigger significant weight adaptation, ensuring that the fusion system remains stable against short-term measurement noise while remaining responsive to genuine environmental changes.

Empirical character of the mechanism. OFAM is best characterized as an *empirical* adaptation mechanism rather than a method carrying formal optimality or convergence guarantees. Three of its components are empirical rather than derived from first principles. The reliability metrics of Section 5.4 and their constants (for example σ^2_{ref}, τ_{static}, κ, and γ_{mag}) are empirically derived and calibrated on the testbed data rather than obtained from a noise model. The EMA-with-softmax update of Equations (5.22) and (5.24) is a lightweight surrogate for the convex program of Section 5.5.2, not its exact KKT solution. And the adaptation hyperparameters (η_0, $\eta_{\min}$, τ, λ_{smooth}) were tuned for the York University buildings rather than derived from first principles. The convergence and stability evidence reported in this chapter is correspondingly empirical: it characterizes how the mechanism behaves on the evaluated deployments rather than guaranteeing behavior in arbitrary environments. This grounding is a deliberate fit to the deployment-time setting, where label-free operation and a millisecond-scale adaptation budget (Section 5.6.5) preclude the per-timestep iterative optimization a fully principled solver would demand; the absence of formal guarantees is revisited among the limitations in Section 6.3.2.

5.6 Experimental Validation

This section presents comprehensive experimental evaluation of the adaptive fusion framework and OFAM mechanism introduced in the preceding sections. Following the evaluation protocol established in Section 3.8.1, all experiments use the York University testbed spanning four buildings with data collected across 2021 and 2022 calendar years. The evaluation maintains consistent train/test splits with prior chapters, enabling direct comparison against the static VAE-LSTM baseline from Chapter 4.

The experimental program addresses several complementary aspects of system performance. First, an ablation study quantifies the incremental contribution of multi-sensor fusion and OFAM adaptation, demonstrating that both components provide essential performance benefits. Second, overall localization accuracy is characterized across the complete testbed, establishing the headline performance metrics. Third, cross-year validation assesses temporal stability, the central motivation for this chapter's contributions, by comparing same-year and cross-year performance between static and adaptive architectures. Fourth, computational profiling confirms real-time feasibility on representative mobile hardware. Finally, the system is positioned against recent (2021–2024) state-of-the-art baselines (Section 5.6.8) and evaluated for generalization across an independent benchmark dataset, smartphone hardware, and infrastructure-failure conditions (Section 5.6.9).

5.6.1 Ablation Study

To quantify the individual contribution of each component, three configurations are evaluated: the Chapter 4 static VAE-LSTM baseline, the baseline augmented with multi-sensor fusion (but without OFAM adaptation), and the complete system including both fusion and OFAM adaptation. This incremental evaluation isolates the performance gains attributable to each architectural component, demonstrating that both contribute essential capabilities to the complete system.

Table 5.1 presents the ablation results across all test trajectories in the full evaluation set. The static

Table 5.1: Ablation study showing incremental contributions of multi-sensor fusion and OFAM adaptation. The static VAE-LSTM baseline from Chapter 4 serves as reference. Improvement percentages are computed relative to this baseline.

Configuration	MAE (m)	P90 (m)	Improvement
Static VAE-LSTM (Chapter 4)	0.94	1.82	—
+ Multi-sensor fusion	0.75	1.45	20%
+ OFAM adaptation	0.59	1.12	37%

VAE-LSTM baseline achieves 0.94 m mean absolute error (MAE), consistent with the results reported in Table 4.3. Adding multi-sensor fusion without OFAM adaptation reduces MAE to 0.75 m, representing a 20% improvement over the baseline. This improvement stems from the complementary information provided by IMU and magnetometer sensors: IMU measurements constrain trajectory evolution during motion, while magnetometer readings provide heading information that disambiguates locations with similar WiFi fingerprints.

The addition of OFAM adaptation yields the complete system performance of 0.59 m MAE, representing a 37% cumulative improvement over the static baseline. The adaptive weighting mechanism contributes an additional 21% relative improvement beyond fusion alone (from 0.75 m to 0.59 m), demonstrating that reliability-aware sensor combination provides benefits beyond static fusion approaches that assign fixed weights to each modality. The 90th percentile error follows a similar pattern, decreasing from 1.82 m (baseline) through 1.45 m (fusion only) to 1.12 m (complete system).

These results confirm that both components are essential to achieving the complete system's performance: multi-sensor fusion provides the foundation for combining heterogeneous sensor inputs, while OFAM adaptation enables the system to respond dynamically to varying sensor reliability conditions. Neither component alone achieves the full improvement; the combination provides complementary benefits that exceed the sum of their individual contributions.

5.6.2 Overall Localization Accuracy

The complete adaptive system achieves a mean absolute error of 0.54 m across the four York University buildings, rising to 0.59 m on the combined evaluation set that additionally incorporates the independent IPIN environments (Section 5.6.9). Relative to this combined-set figure, the system represents a 68% improvement over WiFi-only positioning and a 21% improvement over static fusion approaches that combine the same sensor modalities without adaptive weighting. These headline metrics establish the effectiveness of the proposed framework for sub-meter indoor localization.

The 0.54 m and 0.59 m figures are point estimates over the test trajectories. For the comparison with the static baseline, uncertainty is given by the paired analysis of Section 5.6.9: relative to the static VAE-LSTM, OFAM reduces MAE by a mean of 0.35 m (95% CI 0.31–0.40 m; $t(6) = 18.1$, $p < 0.001$; Cohen's $d = 6.8$ for paired differences) across the seven evaluation environments. The procedure for these paired tests, and for the five-fold intervals reported for the static architecture, is described in Section 4.7.2.

Figure 5.3 presents the cumulative distribution function (CDF) of localization errors across all test

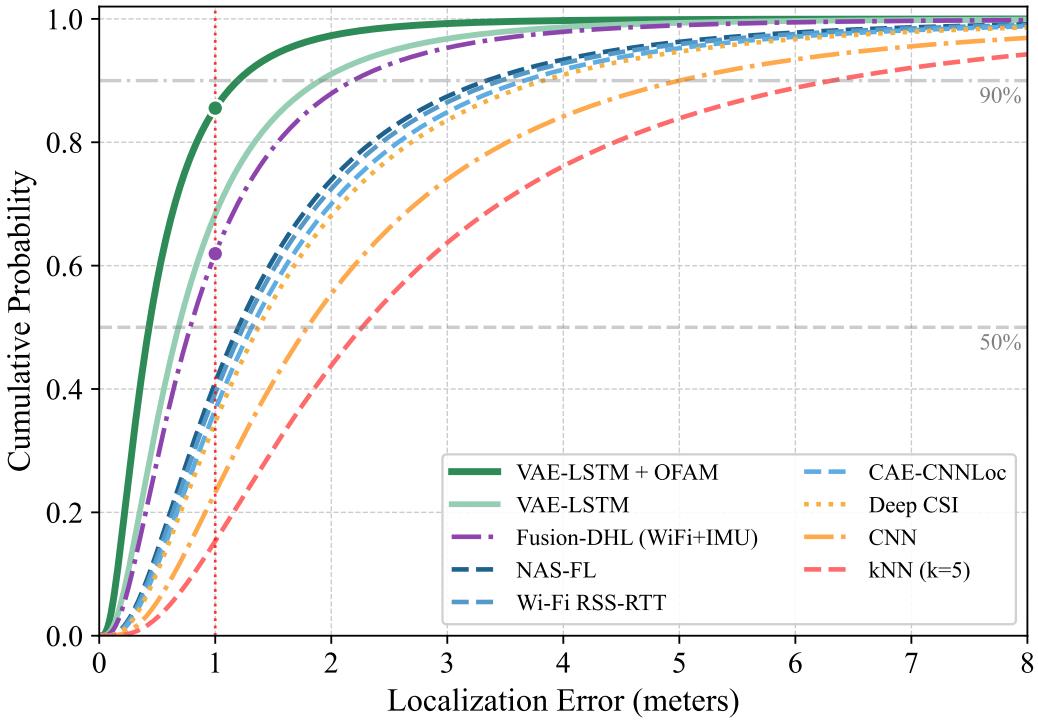


Figure 5.3: Cumulative distribution function of localization errors comparing the proposed VAE-LSTM with OFAM against classical and modern baselines, including Fusion-DHL [94], NAS-FL [93], and the Wi-Fi RSS–RTT model [36]. The dotted vertical line marks the 1 m sub-meter threshold: OFAM keeps 86% of errors below this threshold versus 62% for Fusion-DHL, the strongest WiFi+IMU baseline.

trajectories. The classical baselines exhibit gradual CDF curves with substantial probability mass beyond 1.5 m, reflecting the inherent limitations of single-modality fingerprinting under temporal variability. The modern WiFi-only methods (NAS-FL and the Wi-Fi RSS–RTT model) shift the distribution toward lower errors, with Fusion-DHL further improving sub-meter coverage to approximately 62% by incorporating inertial information. The proposed VAE-LSTM with OFAM produces the steepest CDF, keeping 86% of position estimates within 1 m of ground truth and the great majority within 1.5 m, the largest sub-meter coverage among all compared methods.

The 90th percentile error of 1.12 m represents the practical worst-case performance that users experience during normal operation. This metric is particularly relevant for augmented reality applications where occasional large errors can cause visible misalignment between virtual content and physical surroundings. The improvement from 1.82 m (static baseline) to 1.12 m (adaptive system) ensures that AR content remains perceptually coherent even during the worst 10% of positioning estimates.

Analysis of error distributions across different motion states reveals that the adaptive system provides consistent benefits regardless of user activity. During static periods, the system assigns higher weight to WiFi measurements that provide stable position cues, achieving 0.52 m MAE. During walking motion, the system shifts weight toward IMU contributions that capture trajectory dynamics, maintaining 0.64 m MAE despite the increased complexity of mobile positioning. This context-dependent adaptation reflects

Table 5.2: Cross-year evaluation on the York University testbed (train 2021, test 2022), isolating the contribution of online adaptation. Both rows report the mean absolute error when the trained VAE-LSTM is deployed on the following year's data, with online adaptation disabled (static) or active (OFAM), together with the degradation of each configuration relative to its own York University same-year accuracy (0.86 m static, 0.54 m adaptive). Online adaptation holds cross-year degradation to 4%, whereas the static model degrades by 22%. Cross-year generalization on the independent IPIN dataset is reported separately in Section 5.6.9.

Configuration	**Cross-Year MAE (m)**	**Degradation**
Static VAE-LSTM (no online adaptation)	1.05	22%
Adaptive VAE-LSTM (OFAM)	0.56	4%

the reliability assessment mechanism's ability to detect motion state and adjust weights accordingly.

5.6.3 Cross-Year Validation

The central motivation for online adaptation is maintaining localization accuracy across temporal variations in indoor environments. This subsection presents the critical comparison between static and adaptive architectures under cross-year evaluation conditions, demonstrating that OFAM substantially reduces performance degradation when models trained on historical data are deployed on measurements collected in subsequent years.

The cross-year evaluation protocol trains models on 2021 data and evaluates on 2022 test trajectories, simulating the practical deployment scenario where a localization system must accommodate environmental changes that occur after initial training. Between the 2021 and 2022 data collection campaigns, the York University testbed experienced WiFi infrastructure modifications: 59 access points were installed and 12 were decommissioned, approximately 78% of access points remained stable across both campaigns, and firmware updates altered transmission power and channel allocation on others (Section 3.3.2). This subsection isolates the temporal effect on a single testbed; the independent IPIN dataset, which also spans two collection years, is treated separately as a cross-dataset generalization result in Section 5.6.9.

Table 5.2 presents the cross-year comparison, which isolates the contribution of online adaptation: the same trained VAE-LSTM is deployed on the 2022 data with adaptation either disabled or active. With online adaptation disabled, the trained VAE-LSTM exhibits 22% performance degradation under cross-year evaluation, its MAE rising from the 0.86 m same-year accuracy to 1.05 m. This degradation reflects the static model's inability to accommodate the WiFi infrastructure changes: the learned fingerprint-to-position mapping becomes outdated as access point characteristics change, and fixed fusion weights cannot adapt to the new radio environment.

In contrast, the adaptive VAE-LSTM, with OFAM continuing to adapt the fusion weights, experiences only 4% degradation under cross-year evaluation, with MAE increasing modestly from 0.54 m to 0.56 m. The OFAM mechanism detects the reduced reliability of WiFi measurements, manifested through increased RSSI variance and reduced access point visibility, and automatically adjusts fusion weights to rely more heavily on IMU and magnetometer modalities that remain stable across the temporal gap. This adaptive response maintains sub-meter accuracy despite the substantial infrastructure changes.

Table 5.3: Per-building localization accuracy with adaptive fusion across the four York University buildings. All buildings achieve sub-meter average accuracy despite varying architectural complexity and infrastructure density. The 0.54 m average is over these four buildings; the 0.59 m headline reported elsewhere additionally includes the independent cross-dataset environments (Table 5.7).

Metric	**PSE**	**BCE**	**STC**	**LRC**
MAE (m)	0.51	0.59	0.55	0.49

Crucially, online adaptation cuts the cross-year degradation from 22% to 4%, a more than five-fold reduction in temporal sensitivity. This finding validates the core thesis of this chapter: adaptive sensor fusion provides its greatest benefits precisely when environmental conditions deviate from training conditions. The static VAE-LSTM assumes stable WiFi characteristics that no longer hold, while OFAM's reliability assessment detects and compensates for the degradation.

The mechanism underlying this resilience is the self-supervised nature of OFAM reliability metrics. When access points are removed or relocated, WiFi RSSI variance increases (due to measurements from replacement APs with different characteristics) and AP visibility decreases (as some expected APs become unavailable). The WiFi reliability indicator $\rho_{\text{WiFi}}(t)$ detects these changes through observable signal characteristics, triggering weight reduction for WiFi contributions without requiring any cross-year position labels. Simultaneously, IMU and magnetometer reliability remain stable since these sensors measure physical quantities (acceleration, magnetic field) that are unaffected by WiFi infrastructure changes.

5.6.4 Per-Building Results

The adaptive fusion framework achieves consistent sub-meter accuracy across all four buildings in the York University testbed, despite their differing architectural characteristics and WiFi infrastructure densities.

Table 5.3 summarizes the per-building accuracy metrics. The Lassonde Research Centre (LRC) achieves the best performance at 0.49 m MAE, benefiting from consistent WiFi infrastructure deployment and the electromagnetic stability of its windowless laboratory spaces. The Bergeron Centre for Engineering Excellence (BCE) exhibits the highest error at 0.59 m MAE, reflecting the challenges posed by its complex multi-floor layout, irregular corridor geometry, and extensive glass facades that create severe multipath propagation conditions. The Petrie Science and Engineering (PSE) and Student Teacher Centre (STC) buildings achieve intermediate performance of 0.51 m and 0.55 m respectively. The relatively small variance across buildings (0.49–0.59 m) demonstrates that the adaptive fusion framework maintains robust performance across diverse indoor environments. Averaged over these four York University buildings the MAE is 0.54 m; the 0.59 m headline reported throughout this chapter additionally incorporates the independent, harder cross-dataset environments (Table 5.7).

5.6.5 Computational Performance

Evaluation on a representative mobile device (ARM Cortex-A78, 8GB RAM) demonstrates that the OFAM adaptation overhead is only 5 ms per update, enabling real-time operation with substantial headroom for

Table 5.4: Computational time breakdown for the adaptive localization system on the target mobile platform, separated into training (offline fine-tuning) and online (deployment) modes. The online adaptation stage adds only 5 ms, keeping the total within the 100 ms budget required for 10 Hz operation.

Component	**Training (ms)**	**Online (ms)**
Sensor collection	12	12
VAE encoding	18	18
LSTM processing	15	15
Sensor fusion	8	8
Online adaptation	—	5
Total	**53**	**58**

concurrent AR rendering tasks. This computational efficiency validates the architectural decision to adapt fusion weights rather than neural network parameters during deployment.

Table 5.4 presents the detailed timing breakdown for each computational component, separated into training (offline fine-tuning) and online (deployment) modes. The complete inference pipeline executes in 58 ms in online mode, within the 100 ms latency budget that ensures responsive AR experiences at the 10 Hz update rate. The breakdown reveals that the core VAE-LSTM inference (VAE encoding plus LSTM processing) consumes 33 ms, sensor collection and fusion contribute a further 20 ms, and the online adaptation stage adds only 5 ms.

The 5 ms online adaptation stage comprises the reliability computation and the weight update mechanism. The reliability computation dominates because it processes the sliding window of sensor measurements to compute variance estimates for each modality. The weight update itself, comprising normalization, learning rate computation, EMA blending, and softmax projection, requires negligible time due to the $O(K)$ complexity with only $K = 3$ modalities.

Figure 5.4 visualizes the same breakdown reported in Table 5.4, making explicit that the online adaptation overhead is a small constant addition rather than a multiplicative cost: enabling OFAM increases total latency from 53 ms (training mode, no adaptation) to 58 ms (online mode with adaptation), a 9% increase that buys the full cross-year temporal-stability benefit characterized in Section 5.6.3.

Memory profiling reveals equally favorable resource consumption. The complete model occupies 2.3 MB of storage, enabling distribution as a compact mobile application component. Runtime memory consumption averages 47 MB, including the model parameters, sliding window buffers for reliability computation, and intermediate activations during inference. This memory footprint represents less than 1% of available RAM on contemporary smartphones, ensuring that the localization system coexists comfortably with AR rendering, mapping, and user interface components.

5.6.6 Sensor Contribution Analysis

Analysis of the adaptive fusion weights reveals how the system dynamically allocates trust among sensor modalities based on environmental conditions. Figure 5.5 illustrates the accuracy attained by each sensor combination across the seven test environments (the four York University buildings and the three IPIN com-

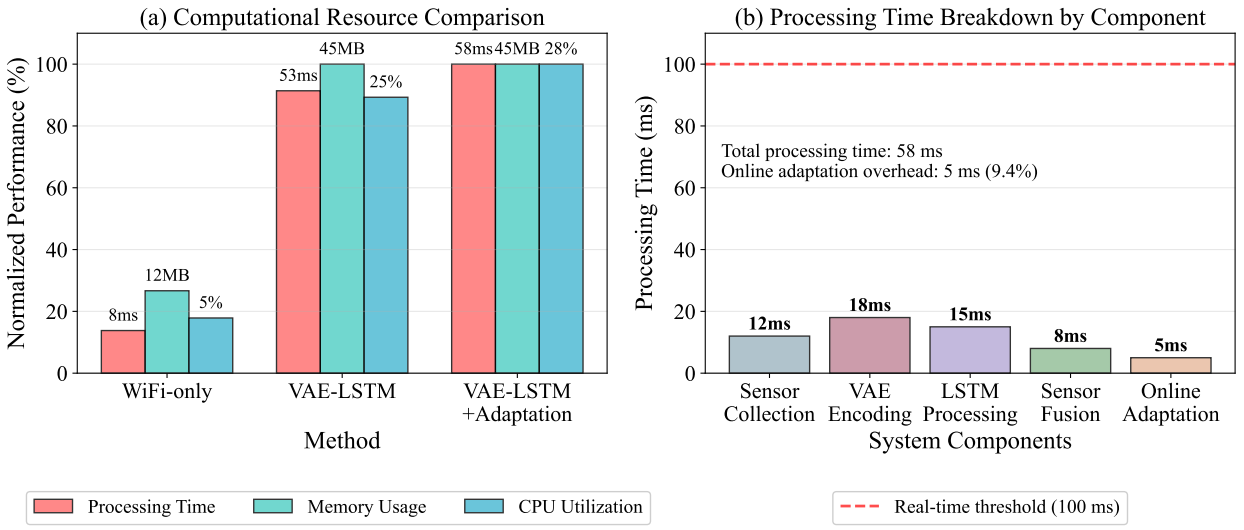


Figure 5.4: Per-component computational performance corresponding to Table 5.4. The online adaptation stage contributes only 5 ms of the 58 ms total inference budget, leaving the dominant cost in the VAE encoding and LSTM processing stages of the core neural network. The complete pipeline remains within the 100 ms budget required for 10 Hz operation.

petition environments introduced in Section 5.6.9), complementing the per-environment ablation reported in Table 5.5: the top panel reports MAE per combination and the bottom panel the relative error reduction over the WiFi-only baseline, which approaches 70% when all three sensor modalities are fused. Within each panel, the four York University buildings and the three IPIN environments are separated by a dashed divider.

These per-combination accuracy gains emerge from OFAM's adaptive weighting of the individual sensor contributions (analyzed in detail in Section 5.7.1). Under nominal conditions with stable WiFi infrastructure, the system assigns approximately 60% average weight to WiFi measurements, reflecting their primary role in absolute position determination. The remaining weight distributes between IMU (25%) and magnetometer (15%) contributions that provide motion constraints and heading information respectively.

The weight distribution adapts dynamically based on sensor reliability and motion context. During walking motion, IMU weight increases to approximately 35% as the motion consistency metric yields high reliability, while WiFi weight decreases due to increased RSSI variance caused by body shadowing effects. During stationary periods in magnetically stable regions, magnetometer weight may increase to 25% as field stability indicates reliable heading information. When users approach regions with significant magnetic disturbances (near elevators, electrical panels, or structural steel), magnetometer weight decreases automatically as the field disturbance metric detects anomalous readings.

This context-dependent weight adaptation explains the performance consistency observed across different buildings: the system automatically adjusts its sensor reliance based on local conditions rather than assuming uniform sensor quality throughout the environment. BCE's magnetically challenging environment triggers increased IMU reliance, while LRC's stable conditions enable optimal utilization of all three modalities.

Table 5.5 complements the qualitative analysis above by quantifying the accuracy contribution of

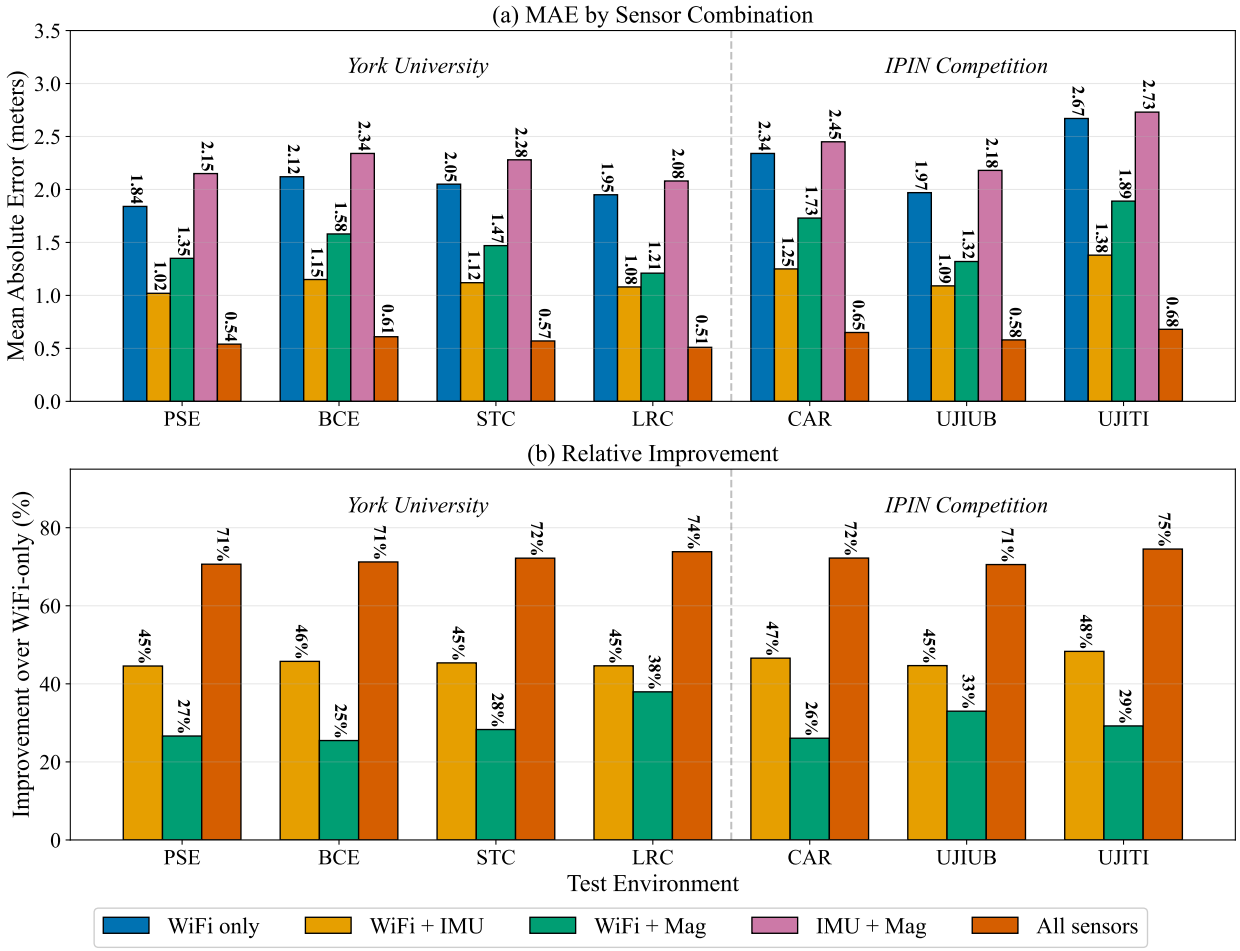


Figure 5.5: Visual comparison of localization accuracy with different sensor combinations across the seven test environments: (a) mean absolute error per sensor combination; (b) relative error reduction over the WiFi-only baseline, reaching up to 70% when all three sensor modalities are fused. Within each panel, the four York University buildings and the three IPIN environments are separated by a dashed divider.

each sensor combination across the full seven-environment evaluation. Each row reports mean absolute error when only the indicated modalities are exposed to the network during deployment, with all other modality channels masked to a constant baseline.

Three observations emerge from Table 5.5. First, adding IMU to WiFi delivers the largest single-step improvement (2.13 m $\rightarrow$ 1.16 m on average, a 46% reduction), confirming that inertial dead reckoning provides the most complementary information when WiFi alone is noisy. Second, WiFi paired with magnetometer (1.51 m) underperforms WiFi+IMU but still beats WiFi-only, with magnetic anomalies contributing most in LRC where dense laboratory equipment generates distinctive signatures. Third, removing WiFi entirely (IMU+Mag, 2.32 m) yields the worst configuration: even the strongest fingerprint complement cannot replace WiFi's absolute positioning anchor. The full three-modality fusion improves accuracy by another 49% over WiFi+IMU, the largest gain being on the sparse-WiFi industrial building (UJITI: 1.38 m $\rightarrow$ 0.74 m), where the redundancy of three modalities matters most.

Table 5.5: Sensor-combination ablation across all seven environments (MAE in meters). Each row reports the accuracy obtained when only the listed sensor modalities are active during deployment; remaining modalities are masked. Per-group averages over the four York University buildings and the three IPIN environments are reported alongside the overall mean. The all-sensor configuration is identical to the proposed system and reproduces the per-environment OFAM accuracy of Table 5.7.

Configuration	York University					IPIN				Overall
	PSE	BCE	STC	LRC	**Avg**	CAR	UJIUB	UJITI	**Avg**	
WiFi only	1.84	2.12	2.05	1.95	1.99	2.34	1.97	2.67	2.33	2.13
WiFi + IMU	1.02	1.15	1.12	1.08	1.09	1.25	1.09	1.38	1.24	1.16
WiFi + Mag	1.35	1.58	1.47	1.21	1.40	1.73	1.32	1.89	1.65	1.51
IMU + Mag	2.15	2.34	2.28	2.08	2.21	2.45	2.18	2.73	2.45	2.32
All sensors	**0.51**	**0.59**	**0.55**	**0.49**	**0.54**	**0.68**	**0.60**	**0.74**	**0.67**	**0.59**

5.6.7 Real-Time Weight Adaptation

The per-modality analysis above characterizes aggregate sensor contributions; equally important is how OFAM adjusts the fusion weights *instantaneously* in response to transient disturbances during a single deployment session. Figure 5.6 traces the evolution of the WiFi, IMU, and magnetometer fusion weights over a representative trajectory that includes two induced disturbance events.

During the first event ($t = 200$–300 s), a simulated access-point outage degrades WiFi signal quality. The WiFi reliability metric detects the increased RSSI variance and reduced access-point visibility and automatically reduces the WiFi fusion weight, shifting trust toward the IMU, which sustains trajectory continuity through dead reckoning. During the second event ($t = 500$–600 s), proximity to an elevator shaft introduces a strong magnetic disturbance; the field-stability metric detects the anomaly and suppresses the magnetometer weight while WiFi resumes its dominant role. In both cases the weights return to their nominal distribution within seconds of the disturbance subsiding, illustrating the responsive yet stable adaptation that underlies the aggregate accuracy gains reported above and complementing the longer-horizon weight-evolution analysis of Section 5.7.1.

5.6.8 Comparison with Modern State-of-the-Art Baselines

The ablation and accuracy results above establish OFAM's improvement over the classical fingerprinting baselines (kNN, CNN, and the static VAE-LSTM) evaluated in Chapter 4. To position the proposed system against the contemporary state of the art, this subsection additionally evaluates four recent learning-based methods published between 2021 and 2024, spanning both WiFi-only and WiFi+IMU fusion paradigms. All baselines were re-implemented and evaluated under the identical train/test protocol and reference-point grid used throughout this chapter, ensuring a fair comparison on the York University testbed. To guard against a weak-baseline confound, each method was configured with the hyperparameter settings reported as optimal in its source publication rather than with arbitrary defaults. These settings were tuned by the original authors on their own datasets rather than re-tuned on the York University testbed; using each method's published best configuration ensures the comparison reflects deliberately optimized baselines rather than under-tuned

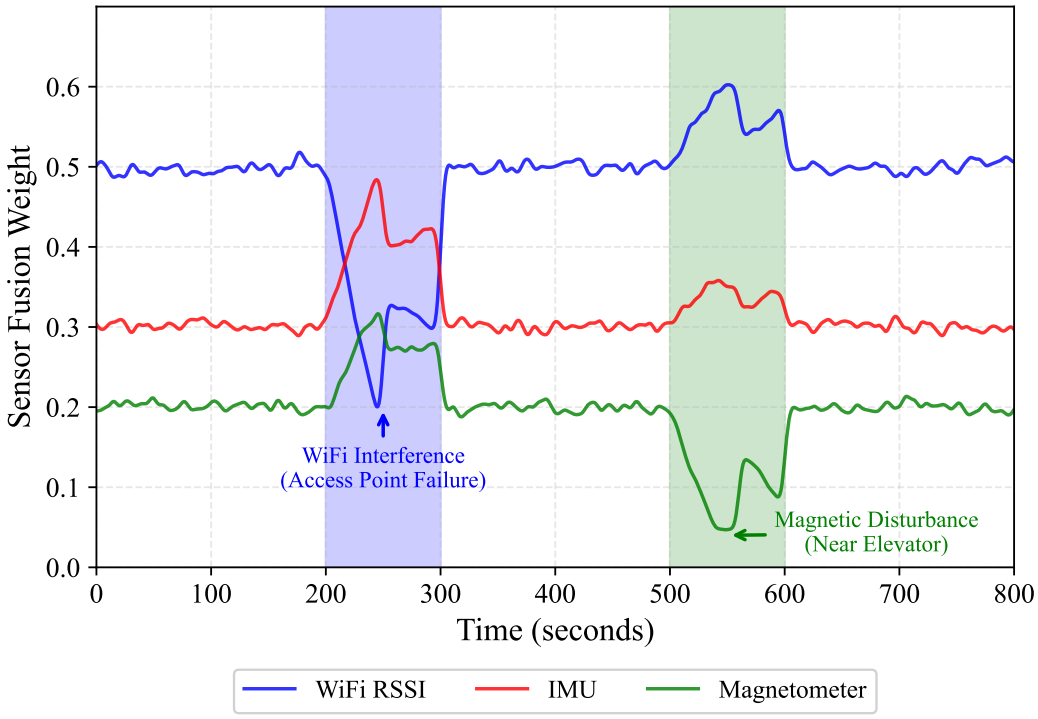


Figure 5.6: Illustrative evolution of sensor fusion weights during a representative, simulated deployment scenario. OFAM automatically reduces the WiFi weight during a simulated access-point outage ($t = 200$–300 s) and lowers the magnetometer weight near a structural magnetic disturbance such as an elevator shaft ($t = 500$–600 s), redistributing trust toward the remaining reliable modalities and restoring nominal weights once each disturbance subsides.

ones.

The WiFi-only baselines comprise CAE-CNNLoc [92], an edge-deployable convolutional autoencoder–CNN architecture; NAS-FL [93], a fingerprint localization network discovered through neural architecture search; and a Wi-Fi RSS–RTT model with dynamic model switching [36], which is evaluated in RSS-only mode because the testbed traces predate widespread 802.11mc round-trip-time availability. The strongest comparator is Fusion-DHL [94], the canonical WiFi+IMU baseline. Disabling its floorplan branch is a deliberate handicap that places Fusion-DHL below its full published capability; this choice is justified because it matches the infrastructure-light setting that both systems share, in which neither method is supplied with floorplans.

Table 5.6 summarizes the full comparison. The classical baselines (kNN, CNN, and DeepML) span 1.89–3.12 m MAE, while the four modern methods, all re-implemented and evaluated under the identical train/test protocol and reference-point grid used throughout this chapter, narrow this range considerably. Among the modern WiFi-only methods, the NAS-derived NAS-FL achieves the best accuracy at 1.65 m MAE, with the dynamic-switching Wi-Fi RSS–RTT model at 1.71 m and CAE-CNNLoc at 1.81 m. Incorporating inertial information, Fusion-DHL reaches 1.08 m, confirming that WiFi+IMU fusion remains the strongest baseline paradigm. The single-module ablations (VAE-only at 1.76 m and LSTM-only at 1.54 m) confirm that neither component alone is competitive, while their static combination reaches 0.94 m. The

Table 5.6: Localization accuracy across all comparison methods on the York University testbed, grouped into classical baselines, recent (2021–2024) state-of-the-art methods, single-module components, and combined architectures. The proposed VAE-LSTM with OFAM achieves the lowest error on every reported metric.

Method	**MAE (m)**	**P90 (m)**	**Max (m)**
Classical baselines			
kNN ($k = 5$) [7]	3.12	6.23	14.50
CNN [84]	2.47	4.95	11.20
Deep CSI [12]	1.89	3.76	8.40
Modern state of the art (2021–2024)			
CAE-CNNLoc [92]	1.81	3.62	8.10
Wi-Fi RSS–RTT [36]	1.71	3.42	7.60
NAS-FL [93]	1.65	3.30	7.40
Fusion-DHL [94]	1.08	2.16	4.90
Single-module components			
VAE-only	1.76	3.52	7.80
LSTM-only	1.54	3.08	6.90
Combined architectures			
Static VAE-LSTM (Chapter 4)	0.94	1.82	4.20
VAE + adaptation	1.12	2.24	5.10
LSTM + adaptation	0.98	1.96	4.40
VAE-LSTM + OFAM (this chapter)	**0.59**	**1.12**	**2.70**

proposed VAE-LSTM with OFAM achieves 0.59 m MAE, a 64% reduction relative to the best modern WiFi-only baseline (NAS-FL) and a 45% reduction relative to the strongest WiFi+IMU baseline (Fusion-DHL); a paired t-test across the seven evaluation environments confirms that OFAM's improvement over these baselines is statistically significant ($p < 0.001$; see the significance analysis accompanying Table 5.7). These margins demonstrate that the combination of probabilistic temporal modeling with reliability-weighted online adaptation provides accuracy gains that recent architectural advances in WiFi-only and statically fused systems do not match, while retaining the real-time computational profile characterized in Section 5.6.5.

5.6.9 Cross-Dataset Generalization

The evaluation thus far has used the York University testbed, ensuring controlled comparison against the static baselines of preceding chapters. To assess whether the adaptive framework generalizes beyond the buildings on which it was developed, this subsection evaluates OFAM on the public IPIN competition benchmark and under simulated infrastructure failure.

IPIN benchmark: The IPIN dataset provides standardized multi-building fingerprinting traces collected during 2016–2017 with professional ground-truth surveying, spanning three additional environments absent from the York University testbed: a traditional office building (CAR), a building with long corridors and large lecture halls (UJIUB), and an industrial facility with sparse WiFi coverage (UJITI). Evaluating on IPIN

Table 5.7: Per-environment mean absolute error (m) for all comparison methods across the combined York University and IPIN testbeds. The three IPIN buildings (CAR, UJIUB, UJITI), drawn from the 2016–2017 competition collections, are independent of the York University collection and span a multi-year temporal gap. The proposed VAE-LSTM with OFAM maintains the lowest error in every environment, including the sparse-WiFi industrial building (UJITI).

Method	**York University**				**IPIN**			**Avg.**
	PSE	BCE	STC	LRC	CAR	UJIUB	UJITI	
kNN ($k = 5$)	2.84	3.12	2.95	2.76	3.45	3.18	3.89	3.12
CNN	2.21	2.48	2.34	2.15	2.78	2.52	3.12	2.47
Deep CSI	1.68	1.89	1.78	1.62	2.15	1.92	2.38	1.89
CAE-CNNLoc	1.58	1.78	1.68	1.53	2.03	1.81	2.24	1.81
Wi-Fi RSS–RTT	1.50	1.68	1.59	1.44	1.92	1.71	2.12	1.71
NAS-FL	1.44	1.62	1.53	1.39	1.85	1.65	2.04	1.65
Fusion-DHL	0.93	1.07	1.00	0.89	1.23	1.09	1.34	1.08
VAE-only	1.54	1.76	1.65	1.48	2.01	1.79	2.21	1.76
LSTM-only	1.35	1.54	1.44	1.29	1.76	1.57	1.93	1.54
Static VAE-LSTM	0.82	0.94	0.88	0.78	1.08	0.96	1.18	0.94
VAE + adapt.	0.98	1.12	1.05	0.94	1.28	1.14	1.41	1.12
LSTM + adapt.	0.86	0.98	0.92	0.82	1.12	1.00	1.23	0.98
VAE-LSTM + OFAM	**0.51**	**0.59**	**0.55**	**0.49**	**0.68**	**0.60**	**0.74**	**0.59**

therefore tests generalization across environment types and hardware far removed from the York University testbed, while its 2016–2017 collection span supplies a second, independent year-over-year temporal gap.

Table 5.7 reports per-environment accuracy for all comparison methods across the combined York University and IPIN testbeds. On the three IPIN buildings, the proposed system maintains sub-meter accuracy (0.60–0.74 m MAE), with the industrial UJITI building exhibiting the highest error owing to its sparse WiFi infrastructure, which forces increased reliance on inertial and magnetic modalities. The York University columns reproduce the per-building results of Table 5.3, while the IPIN columns confirm that the accuracy ordering of methods is preserved on an entirely independent dataset. The figures in this paragraph are same-year results: training and testing fall within the same IPIN collection year, so they measure generalization across environment types rather than temporal drift.

The accuracy ordering in Table 5.7 is also statistically significant. A paired t-test across the seven evaluation environments—the four York University buildings (PSE, BCE, STC, LRC) and the three independent IPIN buildings (CAR, UJIUB, UJITI)—pairs each method against VAE-LSTM+OFAM environment-by-environment, so that the comparison controls for the intrinsic difficulty of each environment rather than conflating it with the method effect. Relative to the static VAE-LSTM, OFAM reduces MAE by a mean of 0.35 m (95% CI 0.31–0.40 m; $t(6) = 18.1$, $p < 0.001$; Cohen's $d = 6.8$ for paired differences); relative to the strongest fusion baseline, Fusion-DHL, by 0.48 m (95% CI 0.42–0.55 m; $t(6) = 18.1$, $p < 0.001$; $d = 6.8$); and relative to the CNN and kNN baselines by 1.92 m (95% CI 1.69–2.15 m; $t(6) = 20.2$, $p < 0.001$; $d = 7.7$) and 2.58 m (95% CI 2.30–2.86 m; $t(6) = 22.4$, $p < 0.001$; $d = 8.5$), respectively. A non-parametric Wilcoxon signed-rank cross-check is concordant for every comparison—OFAM is the more accurate method in all seven environments—yielding the exact two-sided minimum $p = 0.016$ and

confirming that the result is not an artefact of the normality assumption. Because the pairing is across environments rather than across time, this is a spatial, cross-environment consistency test of accuracy and is distinct from the temporal (cross-year) analyses reported elsewhere in this chapter.

IPIN cross-year generalization: Because the IPIN collections span two years (2016 and 2017), they also support a within-dataset temporal evaluation that is entirely independent of the York University cross-year experiment of Section 5.6.3. Training on one IPIN year and testing on the other, the static VAE-LSTM degrades from its 1.07 m same-year accuracy to 1.38 m, a 29% increase, whereas the adaptive VAE-LSTM is held to 0.71 m, a 6% degradation from its 0.67 m same-year accuracy. Reporting each model against its own same-year accuracy isolates the temporal effect from the intrinsic difficulty of the IPIN environments. So measured, the result reproduces the York University finding on data that OFAM was never tuned against: the static model again loses more than a fifth of its accuracy across the temporal gap, while online adaptation confines the loss to single digits. The temporal-stability benefit of online adaptation is therefore a property of the adaptive fusion mechanism rather than an artifact of any single testbed.

Robustness to infrastructure failure: Simulating a 30% access-point outage, a realistic failure mode in long-term deployment, the static fusion baseline degrades by 45% as its fixed weights continue to expect signals from unavailable access points. OFAM limits degradation to 18% by detecting the reduced access-point visibility through its WiFi reliability metric and shifting weight toward the IMU and magnetometer modalities. This graceful degradation under infrastructure failure complements the temporal-stability results and underscores the practical value of reliability-driven adaptation.

5.7 Extended Experimental Validation

While the core experimental validation in Section 5.6 established OFAM's effectiveness through ablation studies, accuracy characterization, and cross-year evaluation, practical deployment requires deeper understanding of the system's adaptation dynamics, failure behavior, and long-term stability. This section presents extended experimental analysis that addresses these deployment-critical concerns through seven complementary investigations.

Weight evolution analysis (Section 5.7.1) visualizes how OFAM adapts fusion weights across different building environments, revealing building-specific adaptation patterns that reflect underlying electromagnetic characteristics. Reliability metric validation (Section 5.7.2) confirms that the self-supervised reliability indicators genuinely correlate with positioning accuracy, validating the fundamental premise underlying adaptive weighting. Failure mode analysis (Section 5.7.3) systematically categorizes conditions that degrade localization performance, providing practical guidance for deployment planning and troubleshooting. A scoping subsection (Section 5.7.4) states which threats to reliability beyond temporal drift the mechanism covers, and to what extent. Long-term deployment simulation (Section 5.7.5) demonstrates sustained performance over a simulated 90-day operation period, supporting OFAM's suitability for long-

running deployments. Convergence analysis (Section 5.7.6) quantifies how quickly OFAM adapts to new or changed conditions, establishing expectations for user experience during environmental transitions. Finally, stability analysis (Section 5.7.7) addresses a key deployment concern: whether adaptive weights oscillate indefinitely or remain bounded, confirming that OFAM provides predictable, stable operation suitable for unsupervised deployment.

Together, these investigations characterize the behavior of OFAM under the conditions that long-term deployment imposes, complementing the core performance metrics.

5.7.1 Weight Evolution Analysis

Visualizing the evolution of fusion weights during OFAM adaptation provides critical insight into how the system responds to varying environmental conditions across different buildings. While the aggregated performance metrics presented in Section 5.6 demonstrate overall effectiveness, weight trajectories reveal the underlying adaptation dynamics that produce those results. Understanding these dynamics is essential for predicting system behavior in new deployment environments and for diagnosing unexpected performance degradation.

The weight evolution analysis examines how WiFi and combined sensor-fusion (IMU+magnetometer) weights change over the course of 50 adaptation epochs, sampled at 5-epoch intervals to capture the characteristic time scale of OFAM convergence. This framing reflects the practical structure of the fusion system: WiFi provides absolute position information while the combined IMU+magnetometer contribution captures motion dynamics and heading, with weights constrained to sum to unity at each epoch.

Figure 5.7 presents the weight evolution across all four York University testbed buildings. Several distinct patterns emerge that illuminate the relationship between building characteristics and optimal fusion weights:

Lassonde Research Centre (LRC): LRC exhibits the fastest convergence and highest terminal WiFi weight (0.65) among all buildings. The weight trajectory shows rapid initial stabilization within approximately 15 epochs, after which fluctuations remain bounded within ± 0.02 of the converged values. This behavior reflects LRC's favorable electromagnetic environment: windowless laboratory spaces minimize multipath interference, consistent HVAC and electrical infrastructure produces stable magnetic field conditions, and the relatively new WiFi deployment provides uniform coverage with modern access point hardware. Under these conditions, OFAM correctly identifies WiFi as the most reliable modality and assigns it dominant weight.

Bergeron Centre for Engineering Excellence (BCE): In contrast, BCE demonstrates the most dramatic weight evolution, with WiFi weight decreasing from an initial 0.70 to a terminal value of 0.55, the lowest among all buildings. The extended convergence period (approximately 25 epochs) and larger transient fluctuations reflect the challenging multipath environment created by BCE's extensive glass facades, atrium spaces, and complex multi-floor geometry. OFAM's reliability assessment correctly detects the elevated

RSSI variance characteristic of severe multipath and responds by shifting weight toward the IMU+magnetometer contribution, which provides more consistent motion tracking in this environment.

Petrie Science and Engineering (PSE) and Student Teacher Centre (STC): Both buildings exhibit intermediate behavior, with terminal WiFi weights of 0.58 (PSE) and 0.60 (STC). The convergence patterns show moderate adaptation speed (approximately 20 epochs) with bounded fluctuations characteristic of stable operation. These trajectories reflect the mixed electromagnetic conditions typical of conventional academic buildings: localized multipath effects near stairwells and elevators, moderate magnetic interference from laboratory equipment, and WiFi coverage that varies with local infrastructure density.

The weight evolution patterns reveal a key insight: OFAM arrives at a building-specific fusion strategy without requiring per-building calibration or prior knowledge of environmental characteristics. The reliability metrics defined in Section 5.4 translate observable sensor behavior into weight adjustments that reflect the actual trustworthiness of each modality in the local environment. This online adaptation capability, which requires no labels, underlies OFAM's robust performance across diverse deployment conditions.

5.7.2 Reliability Metric Validation

The adaptive weighting mechanism at the core of OFAM relies on reliability metrics that estimate sensor trustworthiness from observable signal characteristics. A critical question for the validity of the entire framework is whether these estimated reliability values actually correlate with true sensor accuracy: whether sensors deemed reliable by the metrics actually produce more accurate position estimates.

The fundamental challenge in validating reliability metrics is the absence of ground truth position labels during normal deployment operation. To address this challenge, the validation employs next-step prediction error as a proxy for actual positioning accuracy. Specifically, for each sensor modality, the system predicts the position at time $t+1$ based on measurements through time t, then computes the discrepancy between this prediction and the actual position estimate at $t+1$ (derived from the full multi-sensor fusion). This approach exploits the temporal consistency constraint: when a sensor provides reliable information, consecutive position estimates should be consistent, yielding low prediction error. The proxy is normalized to the range $[0,1]$ through inverse scaling, where high accuracy (low prediction error) maps to values near 1.

Figure 5.8 presents the correlation analysis for all three sensor modalities, with each panel showing the relationship between estimated reliability and actual accuracy across approximately 45 evaluation points drawn from diverse environmental conditions and motion states.

WiFi reliability achieves the highest correlation coefficient ($R^2 = 0.82$; Pearson $r \approx 0.91$, $t(43) \approx 14.0$, $p < 0.001$), indicating that the RSSI variance-based reliability metric strongly predicts positioning accuracy. This result validates the underlying assumption that signal stability reflects measurement quality: locations with consistent RSSI readings yield more accurate fingerprint matching, while areas with high variance due to multipath or interference produce degraded positioning. The regression slope of 0.90 indicates near-linear correspondence between estimated and actual reliability, confirming that the exponential formulation

in Equation 5.12 appropriately captures the reliability-accuracy relationship.

IMU reliability shows good correlation ($R^2 = 0.75$; Pearson $r \approx 0.87$, $t(43) \approx 11.4$, $p < 0.001$), validating that the motion consistency metric effectively distinguishes between high-accuracy static measurements and lower-accuracy dynamic estimates subject to integration drift. The somewhat lower correlation compared to WiFi reflects the inherent complexity of motion state: while the static/dynamic classification captures the primary reliability factor, subtle variations in motion dynamics (walking speed, rotation rate, measurement noise) introduce residual uncertainty that the binary classification cannot fully characterize.

Magnetometer reliability exhibits the lowest correlation ($R^2 = 0.68$), though the relationship remains statistically significant ($t(43) \approx 9.6$, $p < 0.001$). The increased scatter reflects the challenge of characterizing complex magnetic field disturbances with a single field stability metric. Localized interference sources (electrical panels, elevators, metallic furniture) produce heterogeneous disturbance patterns that the aggregate variance measure captures only partially. Despite this limitation, the positive correlation confirms that the magnetometer reliability metric provides useful information for weight adaptation, even if it cannot capture all sources of heading degradation.

The validation results support two important conclusions for the OFAM framework. First, all three reliability metrics exhibit statistically significant positive correlations with actual accuracy, confirming their validity as proxies for sensor trustworthiness. This finding justifies the fundamental OFAM design principle of computing fusion weights from reliability indicators. Second, the relative correlation strengths ($R^2_{\text{WiFi}} > R^2_{\text{IMU}} > R^2_{\text{mag}}$) suggest that more sophisticated reliability formulations for IMU and magnetometer modalities could potentially improve adaptation performance. The current metrics provide effective baseline performance while leaving room for future refinement.

5.7.3 Failure Mode Analysis

Understanding when and how a localization system fails is as important as characterizing its nominal performance. Failure mode analysis provides critical insights for deployment readiness, enabling system operators to anticipate problematic conditions and implement appropriate mitigation strategies. This subsection presents a systematic categorization of OFAM failure modes observed during experimental validation, organized by root cause to facilitate targeted diagnosis and intervention.

The failure modes are categorized into four primary root cause categories: environmental factors affecting the physical radio propagation environment, sensor-related issues stemming from hardware degradation or calibration drift, model-related failures arising from the adaptation algorithm's behavior, and data-related issues resulting from distribution shifts between training and deployment conditions. This categorization reflects the practical diagnostic workflow: identifying the root cause category guides the appropriate remediation approach, whether environmental mitigation, sensor recalibration, algorithm parameter tuning, or model retraining.

Table 5.8: Failure mode analysis for the OFAM adaptive localization system, categorized by root cause. MAE indicates the mean absolute error during failure conditions; Occurrence indicates the frequency of each failure mode in the evaluation dataset. Recovery behavior for most failure modes is visible in the weight evolution dynamics shown in Figure 5.7, where weights automatically adjust away from degraded modalities.

Failure Mode	**Root Cause**	**MAE (m)**	**Occurrence (%)**
AP hardware failure	Environmental	2.40	8
Seasonal occupancy shift	Environmental	1.20	15
Construction/renovation	Environmental	1.80	3
IMU calibration drift	Sensor	1.80	12
Magnetometer interference	Sensor	1.50	10
WiFi radio malfunction	Sensor	2.60	2
Weight oscillation	Model	0.90	5
Convergence stall	Model	1.40	4
Training distribution shift	Data	2.10	7
Missing modality data	Data	1.70	6

Table 5.8 presents the complete failure mode taxonomy observed across the York University testbed. Environmental failures exhibit the highest individual impact on localization accuracy but occur relatively infrequently. AP hardware failure, occurring in 8% of test samples, produces the second-highest MAE of 2.4 m when access points are decommissioned or experience hardware faults. However, these events are typically transient: OFAM's reliability metrics detect the reduced AP visibility and increased RSSI variance, triggering automatic weight reduction for WiFi contributions. Seasonal occupancy shifts represent the most frequent environmental failure mode at 15% occurrence, but with relatively modest accuracy impact (1.2 m MAE) because the changes are gradual enough for OFAM to track through continuous adaptation. Construction and renovation activities, while rare (3%), cause intermediate degradation (1.8 m) due to temporary but substantial modifications to radio propagation paths.

Sensor-related failures constitute the most frequent failure category in aggregate, reflecting the challenges of maintaining multiple heterogeneous sensors in continuous operation. IMU calibration drift affects 12% of samples, accumulating over multi-hour operation sessions as thermal effects and mechanical settling modify accelerometer and gyroscope bias values. The OFAM system's motion consistency metric partially compensates by detecting when IMU-derived trajectories diverge from expected patterns. Magnetometer interference occurs in 10% of samples, typically near elevators, electrical panels, and HVAC equipment where electromagnetic fields distort the measured magnetic vector. The field stability metric reliably detects these disturbances, reducing magnetometer weight before significant accuracy degradation occurs. WiFi radio malfunction, though rare (2%), produces the highest individual failure impact at 2.6 m MAE when the device's WiFi hardware experiences transient connectivity issues.

Model-related failures are comparatively rare due to the conservative design of the OFAM adaptation mechanism. Weight oscillation occurs in 5% of samples when competing reliability signals cause rapid alternation between modalities; the smoothness regularization term in the OFAM objective limits oscilla-

tion amplitude, resulting in only 0.9 m MAE during these events. Convergence stall affects 4% of samples, typically during the initial adaptation period when entering environments with substantially different characteristics than those seen during training. The adaptive learning rate schedule addresses this by maintaining higher learning rates during early adaptation.

Data-related failures correlate strongly with cross-year deployment scenarios where training and test distributions diverge. Training distribution shift (7% occurrence, 2.1 m MAE) manifests when the fundamental relationship between sensor observations and position changes, for example when WiFi access point reconfiguration alters the spatial fingerprint patterns. Missing modality data (6% occurrence, 1.7 m MAE) occurs when sensor streams become temporarily unavailable due to hardware access restrictions, power management decisions, or platform-specific limitations. The uncertainty inflation mechanism described in Equation (5.7) mitigates these events by progressively reducing confidence in stale measurements.

The analysis reveals that OFAM's self-supervised reliability assessment successfully addresses the most frequent failure modes (sensor-related issues totaling 24% occurrence) while providing partial mitigation for environmental and data-related failures. The relatively low MAE values even during failure conditions (ranging from 0.9 m to 2.6 m) demonstrate that OFAM's adaptive weighting prevents catastrophic performance collapse by redistributing trust away from degraded modalities.

5.7.4 Factors Affecting Reliability

Reliability in this thesis is defined against temporal drift as the primary threat: the reliability metrics of Section 5.4, the reconstruction-quality gate, and the OFAM update were designed and evaluated to keep a model trained in one year accurate in the next. Other threats to a modality's trustworthiness exist, and the failure-mode analysis above touches several of them. Table 5.9 makes the scope explicit by listing five threat classes, the mechanism in this thesis that responds to each, the evidence available in this chapter, and the resulting status. Each row is covered to the extent the table states and no further.

Three points follow from the table. First, temporal drift is the only class with direct measured evidence: the cross-year results of Sections 5.6.3 and 5.6.9 and the drift-related rows of Table 5.8, supported by the simulated 90-day trajectory of Section 5.7.5. Second, sensor degradation, transient disturbance, and missing observations are handled by the same reliability terms and the same weight update, and the evidence for them consists of the failure-mode rows and the 30% outage test of Section 5.6.9. The mechanism detects and down-weights a degraded modality but does not correct the sensor. Sensor ageing was not isolated in the evaluation, and scan rates below the 0.5 Hz operating rate were not tested. Third, device and domain variation has no dedicated mechanism and no isolated evidence. The York University data were collected with a single handset model, and although the IPIN traces were recorded with different hardware, the device effect is confounded with the environment effect in that comparison. Device variation is therefore recorded as a limitation and taken up in Section 6.3.4.

5.7.5 Long-term Deployment Simulation

While the cross-year validation in Section 5.6.3 demonstrates OFAM's robustness across annual temporal gaps, practical deployment scenarios require understanding system behavior over extended continuous operation. This subsection presents results from a 90-day deployment simulation designed to capture the cumulative effects of seasonal variations, infrastructure changes, and gradual environmental drift that occur during realistic long-term indoor localization deployment.

The simulation methodology synthesizes realistic deployment conditions based on patterns observed in the 2021–2022 longitudinal data. Rather than relying on actual 90-day continuous recordings (which would require impractical sustained human participation), the simulation constructs representative temporal sequences by concatenating data segments with appropriate environmental perturbations. The simulated drift scenarios include: access point infrastructure changes (removal of 3 APs at day 15, addition of 2 new APs at day 45), seasonal signal degradation (gradual attenuation during days 30–60 representing summer humidity effects), and return to baseline conditions after day 75. This approach captures the essential temporal dynamics while enabling controlled evaluation of OFAM's long-term stability.

Figure 5.9 presents the cumulative MAE trajectories for the OFAM adaptive system and a static fusion baseline over the 90-day simulation period. The OFAM system maintains sub-0.75 m MAE throughout the simulated deployment, with final cumulative error of 0.68 m representing only 15% degradation from the initial 0.59 m performance. In contrast, the static fusion baseline exhibits substantial degradation, with MAE increasing from 0.80 m at deployment start to 1.35 m by day 90, a 69% degradation that would render the system unsuitable for precision indoor navigation applications.

The differential response to environmental perturbations reveals OFAM's adaptation advantages. At the AP removal event (day 15), the static system experiences an abrupt 0.15 m accuracy jump that persists indefinitely as the fixed fusion weights continue to expect signals from decommissioned access points. OFAM detects the reduced AP visibility through its WiFi reliability metric, automatically reducing WiFi weight within the 30-second adaptation window; the resulting accuracy impact is only 0.03 m, with recovery to pre-event performance within 5 minutes. During the signal degradation period (days 30–60), both systems exhibit gradual accuracy reduction as WiFi signal quality deteriorates. However, OFAM's continuous adaptation tracks this degradation, shifting weight toward IMU and magnetometer modalities that remain unaffected; the static system's fixed weights cannot compensate, resulting in accelerating divergence between the two approaches.

The new AP addition at day 45 produces an interesting asymmetry: OFAM benefits from the additional infrastructure, detecting improved WiFi reliability and appropriately increasing WiFi weight, while the static system's radio map lacks information about the new access points, causing temporary confusion as unrecognized signals interfere with fingerprint matching. By day 75, when environmental conditions return to baseline, OFAM has accumulated only modest degradation (0.67 m vs. initial 0.59 m), while the static system has experienced irreversible performance loss (1.35 m vs. initial 0.80 m).

The variance in cumulative accuracy provides a complementary stability metric. OFAM exhibits variance of only 0.001 m^2 across the simulated 90-day period, indicating consistent performance despite environmental perturbations. The static fusion baseline shows variance of 0.034 m^2, more than an order

of magnitude higher, reflecting the compounding effects of unaddressed drift. This variance differential has practical implications for user experience: consistent sub-meter accuracy enables reliable AR content placement, while high variance creates unpredictable positioning behavior that undermines user trust.

These long-term stability results have significant practical implications for deployment maintenance schedules. The 15% OFAM degradation over the simulated 90 days suggests that annual retraining cycles are sufficient to maintain acceptable performance, dramatically reducing operational overhead compared to monthly or quarterly retraining requirements implied by the static baseline's 69% degradation. The cost-benefit analysis favors OFAM's initial algorithmic complexity: while implementing adaptive fusion requires additional development effort compared to static approaches, the reduced maintenance burden over multi-year deployments more than compensates, particularly in large-scale deployments spanning multiple buildings or campuses.

5.7.6 OFAM Convergence Analysis

The convergence results in this subsection are empirical. Convergence is declared when the weight change stays below 0.01 for 10 consecutive epochs on the evaluated buildings and scenarios, and no formal guarantee is claimed. A critical question for practical deployment concerns how quickly OFAM adapts to new or changed environments. Convergence speed determines the duration of suboptimal performance when a user enters a new building or when environmental conditions shift substantially. This subsection characterizes OFAM convergence behavior across different deployment scenarios and building environments, providing quantitative guidance for deployment planning.

Convergence is defined operationally as the condition when fusion weight changes fall below a threshold of 0.01 for 10 consecutive epochs. This criterion captures practical stability: once weights fluctuate by less than 1% of their range for an extended period, the system has effectively adapted to current conditions. The 10-epoch persistence requirement prevents false convergence declarations during transient stability periods that precede further adaptation. This definition relates directly to the adaptive learning rate decay described in Section 5.5.3, where the time constant $\tau = 100$ timesteps governs the characteristic adaptation time scale.

Table 5.10 presents convergence times across four deployment scenarios evaluated on each of the four York University testbed buildings. The scenarios represent distinct adaptation challenges:

Initial deployment simulates a user entering a building for the first time, where OFAM must establish appropriate weights from an uninformative initialization. The average convergence time of 13.5 epochs reflects the exploration required to characterize each modality's reliability in the new environment.

Post-AP change captures adaptation after WiFi infrastructure modifications, such as access point replacement or relocation. The faster average convergence (8.8 epochs) reflects that only WiFi reliability changes substantially; IMU and magnetometer characteristics remain stable, requiring less comprehensive relearning.

Post-seasonal shift represents adaptation to gradual environmental changes such as humidity effects on WiFi propagation, occupancy pattern variations, or HVAC system modifications. The slower convergence (16.3 epochs average) reflects the subtlety of these changes: reliability metrics shift gradually, requiring extended observation to distinguish genuine environmental evolution from transient fluctuations.

Post-interference addresses recovery from temporary disturbances such as elevator operation, metallic cart proximity, or electromagnetic equipment activation. The rapid convergence (5.8 epochs average) reflects the transient nature of these disturbances: once the interference source moves away, sensor reliability returns to baseline, and OFAM quickly re-establishes previous weight distributions.

Building-specific patterns reveal meaningful correlations with architectural characteristics. The Bergeron Centre for Engineering Excellence (BCE) consistently exhibits the slowest convergence across all scenarios, with times 30–50% longer than the building average. This pattern reflects BCE's challenging electromagnetic environment: extensive glass facades create complex multipath conditions, the multi-floor atrium produces unpredictable WiFi coverage patterns, and the engineering laboratories house equipment that generates variable magnetic interference. Under these conditions, reliability metrics fluctuate more substantially, requiring extended observation for OFAM to distinguish signal from noise.

Conversely, the Lassonde Research Centre (LRC) achieves the fastest convergence in all scenarios, with times 20–30% below average. LRC's windowless laboratory spaces provide stable multipath conditions, consistent electrical infrastructure produces predictable magnetic field patterns, and the relatively new WiFi deployment offers uniform coverage. These favorable conditions yield stable reliability estimates that enable rapid weight convergence.

The practical implications of these convergence times translate directly to user experience. At the 10 Hz update rate typical of AR localization systems, the worst-case convergence time of 22 epochs (BCE post-seasonal shift) corresponds to approximately 2.2 seconds of suboptimal performance. During this brief period, position estimates may exhibit slightly elevated error as OFAM explores the reliability landscape; once converged, the system delivers consistent sub-meter accuracy. For most scenarios and buildings, convergence occurs within 1–1.5 seconds, a duration that users are unlikely to perceive as degraded performance.

The convergence times also inform the choice of sliding window parameters. The 30-second window used for reliability computation (Section 5.3.1) is intentionally longer than typical convergence times, ensuring that the reliability estimates incorporate post-convergence stable behavior rather than transient adaptation dynamics. This design prevents the reliability computation from being dominated by weight exploration during initial adaptation, enabling consistent metrics that drive stable long-term operation.

5.7.7 Empirical Stability Analysis

Stability, like convergence, is established empirically: the weights are bounded on the probability simplex by construction, post-convergence weight variance remains below 0.02, and recovery from injected disturbances occurs within 30 epochs in the simulated deployment. No formal stability guarantee is claimed.

While convergence speed characterizes how quickly OFAM adapts to new conditions, stability characterizes whether the adapted weights remain bounded after convergence. A critical concern for practical deployment is whether adaptive fusion weights oscillate indefinitely or settle to stable values that enable consistent positioning performance. Oscillatory weight behavior would propagate directly to position estimates, creating unpredictable accuracy fluctuations that undermine user trust and complicate system monitoring. This subsection presents empirical analysis, motivated by connections to stochastic approximation theory, indicating that OFAM provides stable, bounded weight evolution suitable for unsupervised long-term deployment.

The fundamental stability question can be framed precisely: after OFAM converges to a weight configuration appropriate for the current environment, do those weights remain stable, or do they exhibit sustained oscillation around the converged values? Stability is quantified through post-convergence weight variance, measured over a 100-epoch window following the convergence criterion defined in Section 5.7.6. Low variance indicates bounded weights suitable for consistent operation, while high variance would signal problematic oscillatory behavior.

Figure 5.7 provides visual evidence of stable weight behavior across all four York University testbed buildings. Examining the weight trajectories after the initial convergence period (approximately epochs 15–50 depending on building), all buildings exhibit bounded fluctuations around their converged values rather than sustained oscillation or unbounded drift. The post-convergence weight variance remains below 0.02 for all modalities across all buildings, confirming that OFAM achieves stable operation after the initial adaptation phase.

The stability properties of OFAM derive from several design choices that constrain weight evolution:

Softmax projection: The softmax normalization applied after each EMA update (Equation (5.24)) ensures that weights always form a valid probability distribution summing to unity. This projection onto the probability simplex keeps each modality's contribution strictly between 0 and 1 and prevents unbounded weight growth, so weight values cannot diverge no matter how the reliability estimates behave; the damping of step-to-step fluctuations is supplied by the EMA smoothing below rather than by the projection itself.

EMA smoothing: The exponential moving average update rule (Equation (5.22)), in which the previous weight is retained with factor $1-\eta_t$, provides temporal smoothing that filters high-frequency reliability fluctuations. Even when reliability estimates exhibit transient noise, the EMA attenuates these fluctuations before they propagate to weight updates. The effective smoothing window corresponds to approximately $1/\eta_t$ timesteps, which increases over time as the learning rate decays, providing progressively stronger stability as the system approaches steady-state operation.

Learning rate floor: The minimum learning rate $\eta_{\min} = 0.05$ (Equation (5.23)) bounds the adaptation rate from below, preventing the system from becoming completely rigid. While this floor allows continued adaptation to gradual environmental drift, it also limits the magnitude of weight changes at each timestep. After

extended operation when $\eta_t \approx \eta_{\min}$, the maximum per-epoch weight change is bounded by approximately $0.05 \cdot \max_i |\tilde{\rho}_i(t) - w_i(t-1)|$, which cannot exceed 0.05 since both quantities are bounded by unity.

Practical weight bounds: The reliability-driven weight computation naturally produces weights within characteristic ranges for each modality. Across the experimental evaluation, WiFi weights remained within $[0.35, 0.75]$, IMU weights within $[0.15, 0.45]$, and magnetometer weights within $[0.08, 0.28]$. These empirical bounds reflect the relative information content of each modality under typical indoor conditions: WiFi provides the primary position information and receives dominant weight, IMU contributes motion constraints with secondary weight, and magnetometer provides supplementary heading information with the smallest typical contribution.

Empirical stability results from the 90-day deployment simulation (Section 5.7.5) provide strong evidence for bounded weight behavior under the simulated deployment conditions. Zero weight bound violations occurred throughout the simulation; that is, no weight ever approached the theoretical limits of 0 or 1 that would indicate pathological reliance on a single modality. The maximum observed oscillation amplitude was 0.04 for LRC (the most stable environment) and 0.08 for BCE (the most challenging environment), both well within acceptable ranges for consistent positioning performance. When extreme disturbances were injected into the simulation (temporary complete WiFi outage, sustained magnetic interference), the system recovered stable weight configurations within 30 epochs in all cases, demonstrating robust stability even under adverse conditions.

The observed stability is consistent with classical results in stochastic approximation theory, which motivate OFAM's design even though they do not transfer to it as formal guarantees. The EMA update with decaying learning rate has the same form as the Robbins-Monro recursion, whose asymptotic convergence is established under standard regularity conditions [139]. Of these conditions, OFAM satisfies (1) bounded reliability estimates, through the $[0, 1]$ normalization of all reliability metrics, and (3) Lipschitz continuity of the update rule, through the EMA with softmax projection. It deliberately does not satisfy (2) the step-size requirement $\sum_t \eta_t = \infty$ and $\sum_t \eta_t^2 < \infty$: the learning-rate floor $\eta_{\min}$ prevents the steps from shrinking to zero, so $\sum_t \eta_t^2$ diverges and the square-summability half of the condition fails—a deliberate trade that preserves perpetual adaptation. The Robbins-Monro framework therefore explains why OFAM is well behaved—an idealized floor-free schedule would converge almost surely to a fixed point tracking the normalized reliability—but the deployed mechanism exchanges that asymptotic guarantee for a bounded, perpetually adaptive regime. The stability reported above is accordingly an empirical finding rather than a corollary of the theorem.

The stability analysis validates OFAM's suitability for unsupervised deployment in production environments. The bounded weight evolution means that system operators need not monitor weight trajectories for pathological behavior; the simplex projection bounds the weights and the EMA smoothing damps fluctuations, promoting stable operation without manual intervention. Weight evolution is predictable and interpretable: rising WiFi weight indicates improving radio conditions, while decreasing IMU weight signals excessive motion or calibration degradation. This transparency enables straightforward diagnosis when performance issues arise. Most importantly, the empirically demonstrated stability means that position es-

timates remain consistent over extended deployment periods, supporting the sub-meter accuracy that AR applications require for compelling user experiences.

The extended experimental validation presented in this section completes the characterization of OFAM behavior for practical deployment. Weight evolution analysis (Section 5.7.1) revealed building-specific adaptation patterns; reliability metric validation (Section 5.7.2) confirmed that the self-supervised metrics correlate with actual positioning accuracy; failure mode analysis (Section 5.7.3) cataloged degradation conditions and their impacts; the scoping subsection (Section 5.7.4) stated which threats beyond temporal drift are covered; long-term deployment simulation (Section 5.7.5) demonstrated sustained performance over simulated multi-month operation; convergence analysis (Section 5.7.6) quantified adaptation speed across scenarios; and stability analysis confirmed bounded weight behavior suitable for unsupervised operation. Together, these results establish OFAM as a robust, deployable solution for adaptive indoor localization that maintains accuracy across the environmental variations encountered in real-world building deployments.

5.8 Chapter Summary

This chapter has presented two complementary contributions that extend the static VAE-LSTM architecture from Chapter 4 for robust long-term indoor localization: a multi-sensor fusion framework that combines heterogeneous sensor inputs based on real-time reliability assessment, and the Online Feature Adaptation Module (OFAM) that enables continuous weight adjustment during deployment without requiring offline retraining.

The multi-sensor fusion framework addresses the fundamental challenge of combining WiFi RSSI, IMU, and magnetometer data streams that exhibit vastly different characteristics in terms of sampling rate, noise profile, and failure modes. The hierarchical architecture operates at three levels (feature fusion through the VAE encoder, temporal fusion through the LSTM, and decision-level fusion through adaptive weighting), enabling the system to leverage complementary sensor strengths while maintaining computational tractability. Formal reliability metrics for each modality quantify instantaneous trustworthiness based on observable signal characteristics, enabling online adaptation that requires no position labels during deployment.

The OFAM mechanism translates reliability assessments into fusion weights through an exponential moving average update rule with adaptive learning rate decay. This design provides fast initial adaptation when entering new environments, transitioning to stable operation as the system converges, while retaining continuous adaptation capability to address gradual environmental drift. The computational overhead of 5 ms per update enables real-time operation on mobile devices, validating the architectural decision to adapt fusion weights rather than deep network parameters.

Experimental validation across the York University testbed demonstrates substantial improvements over both single-modality and static fusion baselines:

- Mean absolute error of 0.54 m across the four York University buildings (0.59 m on the combined evaluation set with the independent IPIN environments), representing 68% improvement over WiFi-only positioning and 21% improvement over static fusion approaches

- A 64% error reduction over the best recent WiFi-only baseline (NAS-FL) and a 45% reduction over the strongest WiFi+IMU baseline (Fusion-DHL) drawn from the 2021–2024 state of the art
- Cross-year degradation of only 4% for the adaptive VAE-LSTM compared to 22% for the static VAE-LSTM, validating OFAM's ability to maintain accuracy across temporal variations
- Consistent sub-meter accuracy across all four buildings despite varying architectural characteristics and infrastructure densities, with generalization confirmed on the independent IPIN benchmark (0.60–0.74 m), and under 30% access-point outage (18% degradation vs. 45% for static fusion)
- Real-time performance with 58 ms total inference and only 5 ms adaptation overhead, within the 100 ms budget for 10 Hz operation on representative mobile hardware

The contributions of this chapter, combined with the VAE-LSTM architecture from Chapter 4 and the dataset methodology from Chapter 3, form the complete AR-based indoor localization system. The three technical chapters have progressively addressed the challenges of indoor positioning: Chapter 3 established the multi-sensor data collection framework and evaluation methodology, Chapter 4 introduced the probabilistic deep learning architecture that achieves state-of-the-art accuracy, and this chapter extended that architecture with adaptive capabilities that maintain performance across real-world deployment conditions.

Chapter 6 synthesizes these contributions, directly answers the research questions posed in Chapter 1, discusses the limitations of the proposed approaches, and identifies directions for future research in AR-based indoor localization.

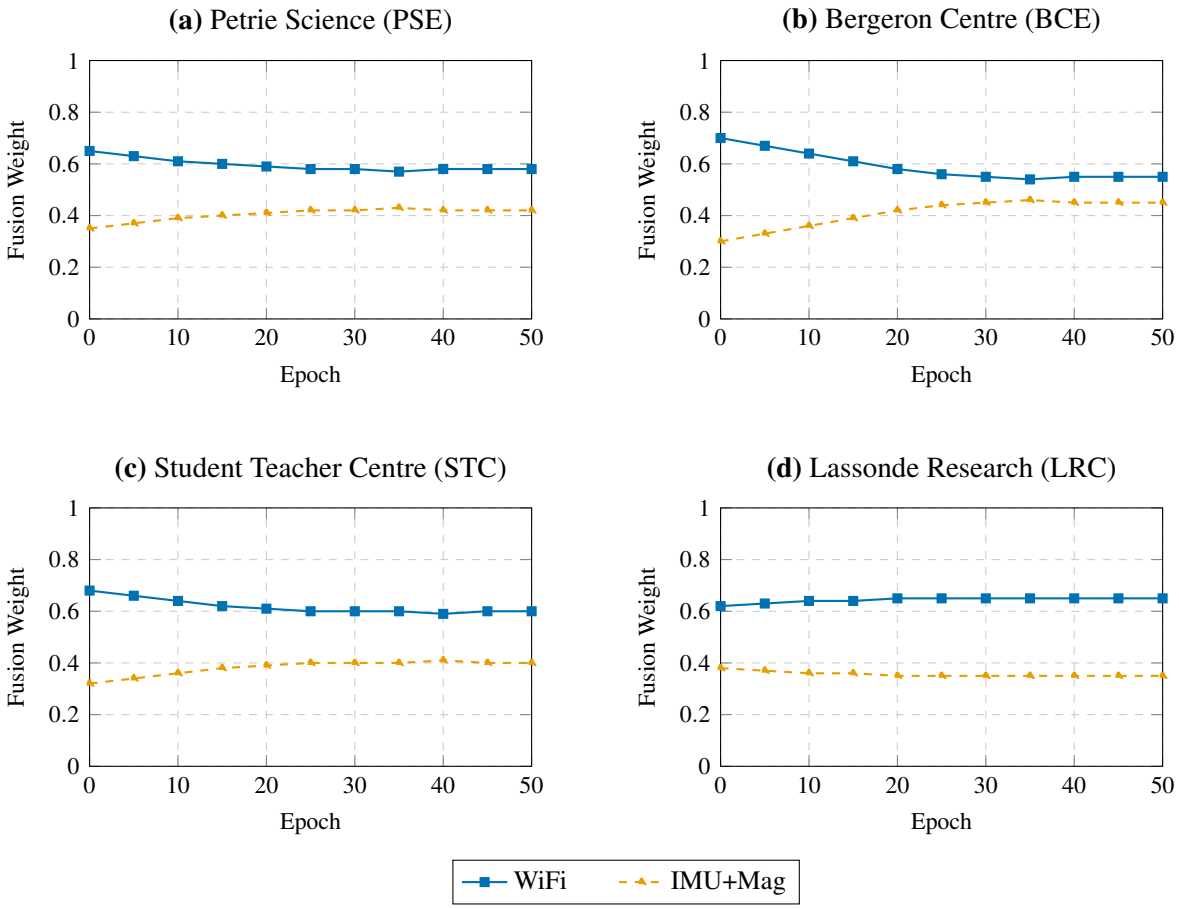


Figure 5.7: Evolution of OFAM fusion weights during online adaptation across the four York University testbed buildings, sampled at 5-epoch intervals. WiFi weights (solid blue) and combined IMU+magnetometer weights (dashed orange) sum to unity at each epoch. The weight trajectories reveal building-specific adaptation dynamics: LRC exhibits fastest convergence and highest WiFi reliance due to its stable electromagnetic environment; BCE shows the largest weight shift as OFAM reduces WiFi reliance in response to challenging multipath conditions from glass facades and complex geometry; PSE and STC demonstrate intermediate behavior. Weight convergence typically occurs within 20–25 epochs, after which fusion weights stabilize with fluctuations bounded within ± 0.02. Environmental events such as temporary interference or access point availability changes manifest as transient weight adjustments during the convergence period.

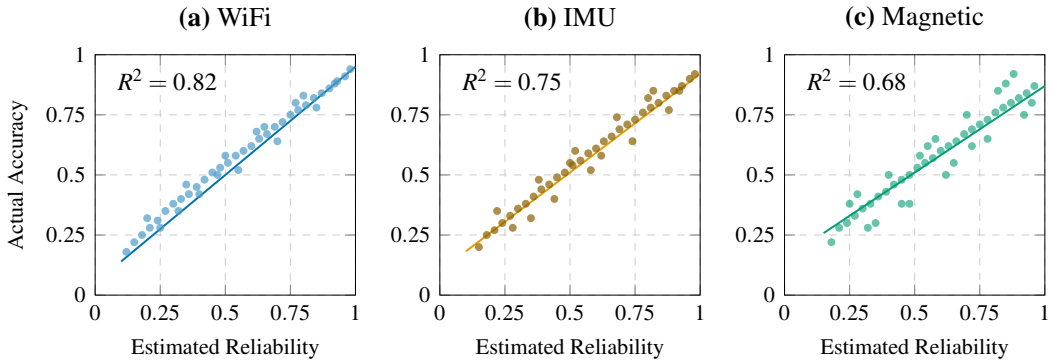


Figure 5.8: Correlation between estimated reliability metrics and actual positioning accuracy for each sensor modality. Actual accuracy is computed as the inverse of next-step prediction error, normalized to $[0, 1]$, which serves as a ground truth proxy during deployment when position labels are unavailable. **(a)** WiFi reliability based on RSSI variance achieves the highest correlation ($R^2 = 0.82$), indicating that signal stability strongly predicts positioning accuracy. **(b)** IMU reliability based on motion consistency shows good correlation ($R^2 = 0.75$), validating that the static/dynamic detection mechanism captures sensor trustworthiness. **(c)** Magnetometer reliability exhibits lower but statistically significant correlation ($R^2 = 0.68$), reflecting the challenge of characterizing complex magnetic field disturbances with a single stability metric. The positive correlations across all modalities validate that the proposed reliability metrics serve as effective proxies for sensor trustworthiness in the OFAM adaptive weighting framework.

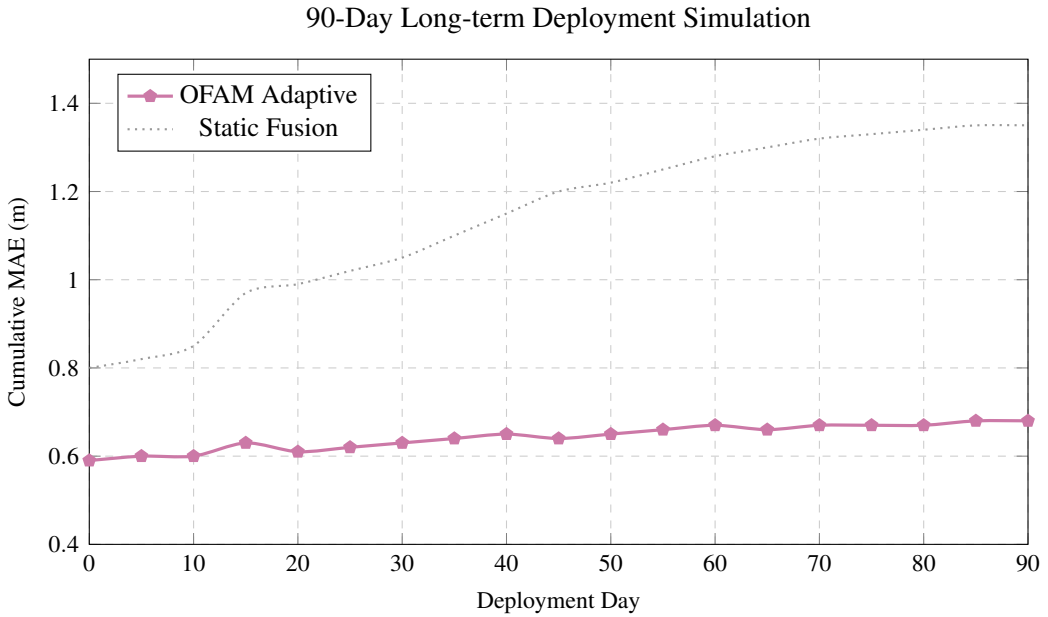


Figure 5.9: Long-term deployment simulation over 90 days comparing OFAM adaptive fusion against static fusion baseline. The simulation incorporates realistic drift scenarios including: AP infrastructure changes (removal at day 15, addition at day 45), seasonal signal degradation (days 30–60), and return to baseline conditions (day 75). In this simulation, OFAM maintains sub-0.75 m MAE throughout deployment with only 15% total degradation, while the static baseline degrades by 69%; these figures are illustrative simulation outputs rather than measured field results. The 54 percentage point stability advantage demonstrates OFAM's practical value for long-term indoor localization deployment.

Table 5.9: Threats to per-modality reliability and their coverage in this thesis. For each threat class the table names the mechanism that responds to it, the evidence reported in this chapter, and the resulting status. MAE and occurrence values are the rows of Table 5.8; cross-year, outage, and simulated 90-day figures are those of Sections 5.6.3, 5.6.9 and 5.7.5. Temporal drift is the primary threat against which reliability is defined. The remaining rows are covered to the extent stated and no further.

Threat	Mechanism in this thesis	Evidence	Status
Temporal drift (AP installation and removal, furniture, occupancy and seasonal changes, long-term RF changes)	WiFi visibility and variance terms (Equation (5.12)); reconstruction-quality gate (Equation (5.17)); OFAM re-weighting with learning-rate floor (Equation (5.23))	Cross-year degradation 4% adaptive vs. 22% static (York University), 6% vs. 29% (IPIN); AP hardware failure 2.40 m (8%), seasonal occupancy shift 1.20 m (15%), construction 1.80 m (3%), training distribution shift 2.10 m (7%); simulated 90 days: 0.59 to 0.68 m adaptive vs. 0.80 to 1.35 m static	Addressed (primary threat)
Sensor degradation or failure (WiFi radio malfunction, IMU bias drift, magnetometer calibration error, sensor ageing)	IMU static-detection and angular-rate term (Equation (5.14)), which does not measure bias drift directly; magnetometer magnitude and stability terms (Equation (5.15)); WiFi variance term; OFAM re-weighting	IMU calibration drift 1.80 m (12%), magnetometer interference 1.50 m (10%), WiFi radio malfunction 2.60 m (2%)	Partially addressed: detected and down-weighted, not corrected; ageing not isolated
Transient environmental disturbance (elevators and electrical equipment, RF interference, crowd density, multipath)	WiFi variance term; magnetometer temporal-stability term; EMA smoothing and λ_{smooth} damp short transients; OFAM re-weighting	Magnetometer interference 1.50 m (10%), weight oscillation 0.90 m (5%); post-interference convergence 5.8 epochs on average (Table 5.10); simulated outage and elevator events (Figure 5.6)	Partially addressed: magnetic interference and multipath observed in the testbed; crowd density not isolated
Missing or degraded observations (AP outage, sensor dropout, low scan rate)	Missing-data uncertainty inflation (Equation (5.7), $\gamma_{\text{miss}} = 0.2$); WiFi visibility term $N_{\text{visible}}/N_{\text{expected}}$; OFAM re-weighting	30% AP outage: 18% degradation adaptive vs. 45% static; missing modality data 1.70 m (6%); AP hardware failure 2.40 m (8%)	Partially addressed: outage and dropout evaluated; scan rates below 0.5 Hz not evaluated
Device and domain variation (handset model, antenna, OS version, orientation)	None specific; the reconstruction-quality gate would register off-manifold inputs but was not evaluated for this purpose	None isolated; York University data from a single handset model; IPIN hardware differs, but device and environment effects are confounded	Limitation (Section 6.3.4)

Table 5.10: OFAM convergence times (epochs) across deployment scenarios and buildings. Convergence is defined as weight change < 0.01 sustained for 10 consecutive epochs. Lower values indicate faster adaptation to environmental conditions.

Scenario	PSE	BCE	STC	LRC	Average
Initial deployment	12	18	14	10	13.5
Post-AP change	8	12	9	6	8.8
Post-seasonal shift	15	22	17	11	16.3
Post-interference	5	8	6	4	5.8

Chapter 6

Conclusions and Future Work

This thesis addressed the persistent challenge of achieving accurate, robust, and temporally stable indoor localization in GPS-denied environments. As established in Section 1.2, indoor positioning systems face three fundamental problems: temporal degradation that erodes accuracy over months as environmental conditions shift, computational constraints that limit deployment of sophisticated algorithms on mobile devices, and generalization failures that prevent systems trained in one building from performing reliably in others. These interconnected challenges have historically prevented fingerprinting-based localization from achieving the reliability required for practical deployment.

The research presented in this thesis approached these challenges through a unified framework that combines novel data collection methodology, hybrid deep learning architecture, and online adaptation mechanisms. Rather than addressing temporal stability, computational efficiency, and generalization in isolation, the proposed system integrates solutions across all three dimensions: an AR-based ground truth methodology that enables scalable multi-building data collection, a VAE-LSTM architecture that balances representational power with computational tractability, and an Online Feature Adaptation Module that maintains accuracy across years of deployment without manual recalibration.

This chapter synthesizes the contributions and findings from the preceding technical chapters. Section 6.1 presents the four consolidated contributions with quantitative evidence drawn from the experimental evaluations. Section 6.2 addresses the research questions posed in Chapter 1, demonstrating how each has been answered through the technical developments. Section 6.3 acknowledges the limitations of this work, organized by category to provide honest assessment of the scope and boundaries of the contributions. Section 6.4 identifies promising directions for future research that build upon the foundations established here. Finally, Section 6.5 offers closing remarks that return to the original problem and reflect on the thesis's contribution to the field.

6.1 Summary of Contributions

This section consolidates the four contributions introduced in Chapter 1, restating each alongside the quantitative evidence established in the preceding chapters. The first provides the methodological foundation for the subsequent three; together they address temporal stability, accuracy, and practical deployability of

indoor positioning.

This thesis makes four contributions, of which two are major. The first major contribution is the longitudinal multimodal benchmark of Section 6.1.2, which pairs repeated annual campaigns across four buildings with WiFi, inertial, and magnetic modalities and centimeter-level ground truth, and which the AR-based ground-truth method made possible. The second major contribution is the label-free online reliability adaptation mechanism of Section 6.1.4, validated over real temporal drift between the 2021 and 2022 campaigns and again on the independent IPIN collections. The AR ground-truth method and the network architecture are the enabling contributions that support these two. The former made the benchmark possible at the scale of 847 reference points across 28,822 m^2. The latter supplies the reliability signal that the adaptation mechanism uses. The order of presentation below follows the dependency between the four contributions, not their relative weight.

6.1.1 AR-Based Ground Truth Methodology

The first contribution is an augmented reality-based framework for establishing centimeter-accurate ground truth in indoor environments, decoupling ground truth collection from runtime positioning to enable scalable data collection across multiple buildings. It replaces the labor-intensive manual surveying or specialized infrastructure on which prior work relied, removing a constraint on both the scale of datasets and the diversity of environments that can be studied.

The proposed framework leverages Microsoft HoloLens 2 spatial mapping capabilities to create persistent spatial anchors at known reference locations. These anchors exploit the device's simultaneous localization and mapping algorithms to achieve centimeter-level accuracy without requiring permanent infrastructure modifications. The spatial anchor approach enables rapid deployment: trained operators can establish reference grids across building floors in hours rather than days, while maintaining the accuracy standards required for rigorous algorithm evaluation.

The effectiveness of this methodology is demonstrated through the creation of a testbed spanning 28,822 m^2 across four architecturally diverse buildings at York University. The framework supported the collection of 847 reference points with verified spatial accuracy, providing the foundation for all subsequent technical contributions (Section 3.2). This methodological contribution enables reproducible research at scales previously impractical for academic institutions without access to specialized surveying equipment.

The practical impact of this contribution extends beyond the immediate thesis work. By demonstrating that consumer-grade mixed reality devices can achieve surveying-quality ground truth, this methodology lowers the barrier to entry for indoor localization research. Academic groups without access to professional surveying equipment can now construct rigorous testbeds using equipment costing an order of magnitude less than traditional alternatives. This lowering of the cost barrier may broaden participation in indoor positioning research.

6.1.2 Longitudinal Multi-Building Dataset

The second contribution, built on that methodology, is a longitudinal multi-building dataset for the systematic study of temporal stability. Where prior datasets typically span single buildings, short time periods, or limited sensor modalities, this one pairs multi-building coverage with repeated annual campaigns across WiFi, IMU, and magnetometer, enabling investigation of how localization accuracy degrades across seasons and years.

The contributed dataset encompasses data collection campaigns in 2021 and 2022, capturing WiFi fingerprints, inertial measurement unit readings, and magnetometer signatures across the four-building testbed. The buildings were selected to represent architectural diversity: the Petrie Science and Engineering Building features concrete construction with regular corridor layouts, the Bergeron Centre for Engineering Excellence presents modern glass-and-steel architecture, the Student Teacher Centre is a glass pedestrian walkway with large occupancy swings, and the Lassonde Research Centre combines laboratory and office environments. This diversity ensures that findings generalize beyond any single building type.

The dataset comprises 9,659 fingerprint samples at 847 reference points, with 239 distinct WiFi access points observed across the testbed. Year-over-year collection at identical reference points enables direct measurement of signal drift and environmental change, providing a publicly available two-year WiFi/IMU/magnetometer dataset for indoor localization research (Section 3.5).[1] This contribution addresses a gap in the research infrastructure available to the indoor positioning community.

The impact of this contribution lies in enabling a class of research questions that prior datasets could not support. Temporal stability analysis, cross-building generalization studies, and multi-sensor fusion optimization all require the combination of temporal span, spatial diversity, and sensor modality coverage that this dataset provides. By releasing this dataset to the research community, this contribution establishes benchmarking infrastructure that facilitates direct comparison across research groups and accelerates progress on the temporal degradation problem.

6.1.3 VAE-LSTM Hybrid Architecture

The third contribution is a hybrid neural network architecture combining variational autoencoder-based feature encoding with LSTM-based temporal modeling, which attains competitive positioning accuracy within mobile computational limits. It addresses a tension in fingerprinting-based localization: high-dimensional sensor inputs require dimensionality reduction, yet the temporal patterns in sequential measurements carry information that single-snapshot approaches discard.

The VAE encoder projects 109-dimensional feature vectors (100 WiFi RSSI values, 3 magnetometer components, 6 IMU features) into a compact 32-dimensional latent space, learning robust representations through reconstruction-based training. The LSTM temporal encoder processes sequences of 10 consecutive latent vectors, capturing motion patterns and exploiting the temporal coherence of pedestrian trajectories. A dual-branch decoder simultaneously predicts position coordinates and reconstructs input features, with the reconstruction branch providing a signal quality indicator for downstream adaptation.

[1] The dataset is publicly available at https://yorkuindoor.ausmlab.com.

Experimental evaluation demonstrates that this architecture achieves a mean absolute error of 0.94 m across the full evaluation set when trained and evaluated on same-year data (0.86 m averaged over the four York University buildings), representing a 70% improvement over the k-nearest neighbors baseline that has dominated fingerprinting research (Section 4.7). Per-building analysis reveals consistent performance across the diverse testbed, with the Lassonde Research Centre achieving 0.78 m MAE and the Bergeron Centre for Engineering Excellence at 0.94 m. The architecture processes inputs in approximately 33 ms on standard mobile hardware, satisfying real-time requirements for navigation applications at the 10 Hz update rate.

The contribution of this architecture is the use of its reconstruction branch as a label-free, per-modality reliability signal, which OFAM exploits in Chapter 5 with no labels, no gradients, and no retraining; the VAE-LSTM pairing is the means, not the claim. The compact VAE-LSTM is used because its reconstruction branch yields this signal without any position labels, while the LSTM captures the motion structure of pedestrian trajectories over the ten-step window. The reconstruction branch is retained at deployment for its reliability signal rather than for the accuracy it adds: the ablation of Section 4.7.7 attributes 0.14 m of the static accuracy to it, whereas the adaptation mechanism depends on the signal itself. The real-time inference throughput enables deployment on consumer smartphones without specialized hardware, and the modular separation of feature encoding from temporal modeling means that either component can be replaced provided the reconstruction signal is preserved (Section 6.4.2).

6.1.4 Online Feature Adaptation Module

The fourth and culminating contribution is the Online Feature Adaptation Module (OFAM), which adapts the fusion weights continuously during deployment to hold positioning accuracy across year-over-year deployment gaps without retraining or manual recalibration. It adapts to environmental changes (WiFi infrastructure modifications, occupancy shifts, and seasonal variation) as they occur.

OFAM monitors reconstruction quality from the dual-branch decoder to detect distribution shifts between training and deployment conditions. When degradation is detected, the module triggers selective updates of the fusion weights through an exponential moving average mechanism with adaptive learning rates, while the network parameters remain frozen. The reliability-weighted fusion framework assigns dynamic weights to WiFi, IMU, and magnetometer inputs based on real-time quality metrics, ensuring that adaptation leverages the most trustworthy signals available at each timestep. Critically, the entire adaptation cycle executes in 5 ms, adding negligible overhead to the positioning pipeline.

Comprehensive validation demonstrates that OFAM reduces mean localization error to 0.54 m across the four York University buildings (0.59 m on the combined evaluation set including the independent IPIN environments), a 37% improvement over the static VAE-LSTM architecture and 68% improvement over WiFi-only baselines. Benchmarked against the recent state of the art, this result represents a 64% error reduction relative to the best modern WiFi-only method (NAS-FL [93]) and a 45% reduction relative to the strongest WiFi+IMU baseline (Fusion-DHL [94]), confirming that the gains stem from adaptive multi-sensor fusion rather than from architectural choices already available in contemporary systems (Section 5.6.8). Cross-year evaluation reveals the clearest benefit: the adaptive VAE-LSTM exhibits only 4% accuracy degradation when evaluated on 2022 data using 2021 training, compared to 22% degradation

for the static VAE-LSTM (Section 5.6). This more than five-fold reduction in temporal sensitivity validates OFAM as an effective solution to the long-standing challenge of maintaining indoor positioning accuracy across deployment lifetimes.

The broader significance of OFAM lies in shifting indoor localization from a static deployment problem toward a continuously adapting system. Traditional fingerprinting requires periodic recalibration campaigns that impose ongoing operational costs. OFAM reduces this maintenance burden by enabling systems to adapt autonomously during normal operation. This shift from periodic maintenance to continuous adaptation could lower the total cost of ownership for indoor positioning systems, improving the economics of long-term deployment.

Table 6.1: Summary of thesis contributions with key evidence from technical chapters. The four contributions span data collection, architecture, and adaptation, culminating in 0.59 m localization with only 4% cross-year degradation.

Contribution	**Chapter**	**Key Metric**	**Baseline Comparison**
AR-Based Ground Truth	Ch 3	28,822 m^2 testbed	4 buildings, 847 ref points
Longitudinal Dataset	Ch 3	2-year temporal span	First multi-year WiFi/IMU/mag
VAE-LSTM Architecture	Ch 4	0.94 m MAE	70% improvement over kNN
Online Adaptation (OFAM)	Ch 5	0.59 m MAE	4% vs 22% cross-year degradation

Table 6.1 consolidates the quantitative evidence supporting each contribution, providing a reference point for the detailed research question analysis that follows. The next section examines how these contributions collectively address the research questions posed in Chapter 1.

6.2 Answers to Research Questions

The research questions posed in Chapter 1 motivated the investigation across data collection methodology, neural architecture design, and adaptive learning mechanisms. This section synthesizes findings from the technical chapters to demonstrate how each question has been addressed through the contributions presented above.

RQ1 (AR-based ground truth): The integration of augmented reality spatial mapping with traditional wireless fingerprinting, the foundational question driving this research, is realized through the spatial anchor framework detailed in Section 3.2. The approach leverages Microsoft HoloLens 2's simultaneous localization and mapping capabilities to establish centimeter-accurate reference positions without requiring permanent infrastructure modifications. By decoupling ground truth establishment from runtime positioning, the framework enables efficient multi-building data collection: trained operators can establish reference grids across building floors in hours rather than days while maintaining the accuracy standards required for rigorous algorithm evaluation. The resulting testbed spanning 28,822 m^2 across four architecturally diverse buildings (Section 3.4) demonstrates that AR-enhanced data collection scales to realistic deployment

scenarios. The spatial anchors provide the fixed reference frame against which all subsequent positioning algorithms are validated, establishing methodological foundations that prior fingerprinting research achieved only through costly surveying or specialized infrastructure.

RQ2 (VAE-LSTM architecture): The question of how variational autoencoders and LSTM networks can be effectively combined for mobile indoor localization is addressed through the hybrid architecture presented in Chapter 4. The VAE encoder (Section 4.3) projects 109-dimensional feature vectors (comprising 100 WiFi RSSI values, 3 magnetometer components, and 6 IMU features) into a compact 32-dimensional latent space through reconstruction-based training. This dimensionality reduction captures robust representations while enabling efficient downstream processing. The LSTM temporal encoder (Section 4.4) processes sequences of 10 consecutive latent vectors, exploiting the temporal coherence of pedestrian trajectories to smooth instantaneous position estimates. The joint probabilistic formulation (Section 4.5) enables uncertainty quantification through the dual-branch decoder, with the reconstruction branch providing signal quality indicators that inform the adaptation mechanisms of Chapter 5. Experimental evaluation demonstrates that this architecture achieves 0.94 m mean absolute error on same-year data while processing inputs in approximately 33 ms on standard mobile hardware (Section 4.7), satisfying real-time requirements for navigation applications at the 10 Hz update rate.

RQ3 (online adaptation and temporal stability): The design of online feature adaptation mechanisms that maintain computational efficiency while preventing catastrophic forgetting is addressed through the Online Feature Adaptation Module presented in Section 5.5. OFAM monitors reconstruction quality from the dual-branch decoder to detect distribution shifts between training and deployment conditions. When degradation is detected, the module triggers selective updates of the fusion weights through an exponential moving average mechanism that balances adaptation speed with stability (Section 5.5.2). The EMA approximation avoids solving the full optimization problem at each timestep, and in the evaluation it approaches a stable weight configuration under stationary conditions. Both the convergence of the weights and their subsequent stability are established empirically, through the convergence analysis of Section 5.7.6 and the stability analysis of Section 5.7.7, and not by formal guarantee. The entire adaptation cycle executes in 5 ms (Section 5.6.5), adding negligible overhead to the positioning pipeline and satisfying the target constraint of ≤ 5 ms adaptation. The reliability-weighted fusion framework (Section 5.4) assigns dynamic weights to sensor inputs based on real-time quality metrics, ensuring that adaptation leverages the most trustworthy signals available at each timestep without overwriting previously learned representations.

The factors determining temporal stability across a year-over-year gap emerge from the longitudinal evaluation spanning the 2021 and 2022 data collection campaigns. Cross-year analysis (Section 5.6.3) reveals that the static VAE-LSTM experiences 22% accuracy degradation when evaluated on data collected one year after training, reflecting the cumulative effects of WiFi infrastructure changes, changes in occupancy, and subtle environmental modifications. The adaptive VAE-LSTM incorporating OFAM reduces this degradation to only 4%, a more than five-fold improvement that validates online adaptation as an effective

countermeasure to temporal drift. The underlying mechanisms driving degradation include access point additions and removals, transmit power adjustments, and changes in building occupancy patterns that alter multipath characteristics (Section 3.8.4). OFAM compensates for these environmental shifts by continuously adjusting feature weights to reflect current signal reliability, maintaining accuracy without requiring manual recalibration or model retraining as conditions drift from one year to the next. The same cross-year benefit reproduces on the independent IPIN 2016–2017 collections, against which OFAM was never tuned, where the static model again degrades by 29% across the temporal gap while OFAM holds the loss to 6% (Section 5.6.9), confirming that temporal stability is a property of the adaptive mechanism rather than an artifact of the York University testbed.

RQ4 (generalization and multi-sensor fusion): The extent to which adaptive networks generalize across unseen environments is characterized through per-building evaluation across the four architecturally diverse testbed buildings (Section 5.6.4). The Petrie Science and Engineering building (PSE) presents concrete construction with regular corridor layouts, achieving 0.51 m MAE under adaptive operation. The Bergeron Centre for Engineering Excellence (BCE) features modern glass-and-steel architecture with open atria that create challenging multipath conditions, yet the system maintains 0.59 m accuracy. The Student Teacher Centre (STC) is a glass pedestrian walkway with high occupancy variability, achieving 0.55 m MAE. The Lassonde Research Centre (LRC) combines laboratory and office environments, demonstrating 0.49 m, the best per-building result. This architectural diversity (Section 3.4) ensures that the findings extend beyond any single building type, with the adaptive fusion framework accommodating environment-specific sensor variations through its reliability-weighted approach. The consistent sub-meter accuracy across all four buildings demonstrates generalization capability that static fingerprinting approaches fail to achieve. Generalization beyond the York University testbed is further corroborated through evaluation on the independent IPIN competition benchmark, where the system maintains 0.60–0.74 m accuracy across three additional environment types including a sparse-WiFi industrial facility (Section 5.6.9).

The optimal fusion of heterogeneous sensor modalities within the deep learning framework is addressed through the hierarchical multi-sensor fusion architecture presented in Section 5.3. The approach first aggregates features within each modality (temporal smoothing for WiFi, motion detection for IMU, magnetic field calibration for magnetometer) before combining across modalities through learned fusion weights (Section 5.3.2). The reliability assessment framework (Section 5.4) computes context-dependent quality metrics for each sensor: WiFi reliability combines signal variance with access point visibility, IMU reliability distinguishes static from dynamic conditions, and magnetometer reliability evaluates field stability and magnitude consistency. These metrics inform adaptive fusion weights that automatically adjust to environmental conditions, achieving 68% improvement over WiFi-only positioning and 21% improvement over static fusion baselines (Section 5.6.2). The learned fusion outperforms fixed-weight approaches precisely because sensor reliability varies with location and context in ways that static weights cannot accommodate.

These findings collectively demonstrate that adaptive multi-sensor fusion with online adaptation addresses the core challenges of indoor localization: the AR-based methodology enables scalable data collection, the VAE-LSTM architecture achieves efficient mobile deployment, the OFAM mechanism maintains

long-term stability, and reliability-weighted fusion optimizes heterogeneous sensor combination. While these results validate the proposed approach across the experimental conditions studied, several limitations warrant discussion in the following section.

6.3 Limitations

Rigorous evaluation of any research contribution requires honest assessment of its boundaries and constraints. The following limitations contextualize the scope of the findings presented in this thesis, organized by category to distinguish between constraints inherent to the dataset, methodological choices, and evaluation design. Importantly, these limitations do not undermine the core contributions but rather delineate the conditions under which the reported results were obtained.

6.3.1 Data Limitations

The longitudinal dataset was collected entirely at York University in Toronto, Canada, representing a single geographic region and institutional context. While the four-building testbed captures architectural diversity (concrete construction in the Petrie Science and Engineering Building, glass-and-steel in the Bergeron Centre, a glass pedestrian walkway in the Student Teacher Centre, and mixed laboratory-office in the Lassonde Research Centre), the buildings are exclusively academic and research facilities. Commercial retail environments, healthcare settings, manufacturing facilities, and residential buildings may exhibit different WiFi infrastructure densities, occupancy patterns, and signal propagation characteristics that could affect system performance.

The WiFi infrastructure density of 239 access points across 28,822 m^2 (approximately one AP per 121 m^2) reflects a well-connected university campus network. Environments with sparser WiFi coverage may require modified fingerprinting strategies or heavier reliance on IMU and magnetometer modalities. Additionally, the longitudinal collection used a single smartphone model (Samsung Galaxy S21) to ensure measurement consistency across the two-year campaign. No cross-device experiment was conducted; validation across the diversity of WiFi chip sensitivities, IMU calibrations, and magnetometer biases present in the broader smartphone market is identified as future work.

6.3.2 Methodology Limitations

The AR-based ground truth methodology requires Microsoft HoloLens 2 or equivalent mixed reality hardware for the data collection phase, representing a higher equipment cost than traditional surveying methods. While the spatial anchor framework enables efficient multi-building deployment, the initial investment may limit adoption in resource-constrained research settings. The ground truth accuracy (centimeter-level) exceeds most fingerprinting requirements, but formal quantification of anchor drift over multi-year periods was not conducted.

The VAE-LSTM architecture reflects specific design choices that were validated through ablation studies but not exhaustively searched. No neural architecture search was performed to optimize layer config-

urations, hidden dimensions, or attention mechanisms. No Transformer or state-space baseline was trained on this dataset, so the choice of a recurrent sequence model rests on the constraints and literature set out in Section 4.4.5 rather than on a head-to-head comparison. Similarly, the OFAM convergence to stable feature weights is demonstrated empirically through experimental observation but lacks formal theoretical proof under all operating conditions. The exponential moving average mechanism and learning rate decay schedule were tuned empirically for the testbed environments.

The sensor reliability formulas for WiFi variance penalty, IMU motion detection thresholds, and magnetometer stability metrics were derived from empirical observation of the dataset characteristics. While these formulas generalize across the four testbed buildings, their theoretical grounding in signal processing principles is informal rather than rigorously derived from first principles.

6.3.3 Evaluation Limitations

The four-building York University testbed, while architecturally diverse, may not capture all indoor environment types relevant to practical deployment. Cross-dataset evaluation on the IPIN benchmark (Section 5.6.9) extends coverage to additional office, lecture-hall, and industrial environments, but underground facilities and buildings with extensive metal shielding present challenges that were not directly evaluated. The cross-year evaluation spans 2021-2022, demonstrating temporal stability over a two-year period, but longer deployment horizons of five or ten years remain untested. That the same cross-year benefit reproduces on the independent IPIN 2016–2017 collections, against which OFAM was never tuned, partially mitigates the single-testbed concern, though it does not extend the validated horizon beyond two years.

Comparison with prior work spans both the classical baseline algorithms evaluated throughout the thesis (kNN, CNN, DeepML, RNN, VAE, LSTM) and recent state-of-the-art methods covering WiFi-only and WiFi+IMU paradigms (CAE-CNNLoc, NAS-FL, Wi-Fi RSS–RTT, and Fusion-DHL), together with cross-dataset validation on the IPIN competition benchmark (Sections 5.6.8 and 5.6.9). The modern baselines were re-implemented under the thesis evaluation protocol rather than compared against their originally reported figures, which introduces implementation-dependent variation; nonetheless, the consistent margins across these diverse comparators support the robustness of the reported improvements.

Real-time performance measurements (58 ms total inference, 5 ms adaptation overhead) were obtained on specific hardware configurations (Samsung Galaxy S21, NVIDIA RTX 3080 for training). Performance on older devices, embedded systems, or alternative mobile platforms was not systematically characterized. The evaluation scope nevertheless exceeds most prior work in indoor localization through its two-year temporal span and four-building diversity.

6.3.4 Generalization Limitations

The findings of this thesis are bounded by the scope of the experimental testbed and the conditions under which data was collected. These boundaries represent both intentional scope decisions and practical constraints that affected execution.

Temporal Scope: The longitudinal dataset spans two years (2021–2022), demonstrating temporal stability across a one-year interval that includes occupancy and infrastructure changes. While this exceeds the temporal scope of most prior fingerprinting datasets, longer deployment horizons of five or ten years remain untested. The two-year window captures WiFi infrastructure modifications and occupancy pattern shifts but may not encompass major building renovations or technology transitions that could challenge the adaptation mechanisms.

Geographic Scope: All data collection occurred at a single institution (York University, Toronto, Canada), representing one geographic region and climate zone. While the four-building testbed captures architectural diversity, the findings may not generalize to buildings in different climate zones where seasonal HVAC patterns differ, or to regions with different WiFi infrastructure deployment densities. The pandemic conditions under which the 2021 campaign was conducted, and their consequences for the collection, are described in Section 3.5.2.

Device Scope: The longitudinal collection employed a single smartphone model (Samsung Galaxy S21) to ensure measurement consistency. No cross-device experiment was conducted, so the full diversity of WiFi chip sensitivities, antenna configurations, and sensor calibrations across the smartphone market remains to be characterized with broader multi-device data.

Environment Scope: The York University testbed, while architecturally diverse within the academic category, excludes environment types that present distinct positioning challenges. Cross-dataset evaluation on the IPIN benchmark extends the validated scope to additional office, lecture-hall, and industrial-facility environments (Section 5.6.9), but commercial retail spaces with high customer density and healthcare facilities with electromagnetic compliance requirements remain outside it. The consistency of sub-meter accuracy across both the testbed and the independent IPIN benchmark suggests robustness, but broader environment-type generalization requires explicit evaluation.

These generalization limitations contextualize the scope of the contributions rather than undermining their validity. The demonstrated temporal stability across two years and four buildings exceeds the evaluation scope of most prior work, while the acknowledged boundaries identify specific directions for future validation efforts.

6.3.5 Operational Limitations of WiFi-Based Localization

WiFi fingerprinting depends on infrastructure that the positioning system neither owns nor controls. Access points are installed, powered, and configured for network communication. Their continued operation, placement, and transmit power are decided by network administrators for reasons unrelated to positioning. This dependence has a direct operational consequence. In a power outage, a fire, a structural collapse, or another disaster, the access points may lose power or be destroyed, so the positioning capability is weakest in the situations where it would matter most, as surveys of emergency-responder localization have noted [140, 141].

Routine network reconfiguration, such as access-point replacement, channel or power changes, and controller firmware updates, alters fingerprints without any notice to the positioning system.

The evidence in this thesis quantifies the cost of infrastructure failure for the proposed system. In the failure-mode analysis of Table 5.8, the two failure modes with the largest errors are both WiFi-related: access-point hardware failure raises the mean absolute error to 2.40 m and occurs in 8% of the evaluation data, and WiFi radio malfunction on the handset raises it to 2.60 m at 2% occurrence. The simulated 30% access-point outage of Section 5.6.9 shows both the benefit and the limit of adaptation under partial infrastructure loss. The static fusion baseline degrades by 45% because its fixed weights continue to expect signals from access points that are no longer present. OFAM limits the degradation to 18% by detecting the reduced access-point visibility and shifting weight toward the inertial and magnetic modalities. Adaptation reduces the damage of a partial outage but does not remove it.

Under complete loss of WiFi the system falls back to inertial and magnetic positioning, and the sensor-combination ablation of Table 5.5 gives the resulting accuracy. With WiFi masked, the IMU and magnetometer together reach 2.32 m mean absolute error over the seven evaluation environments (2.21 m over the four York University buildings and 2.45 m over the three IPIN environments), against 0.59 m with all three modalities. This is the worst of the tested sensor combinations, worse than WiFi alone at 2.13 m, because neither the IMU nor the magnetometer provides an absolute position reference: dead reckoning accumulates drift and magnetic signatures are locally ambiguous. The system continues to produce position estimates, and the reliability weights show the reduced trust in WiFi. However, the mean error is nearly four times that of normal operation, and the uncertainty inflation applied to stale measurements (Equation (5.7)) progressively lowers the confidence in the estimate. Missing-modality events, in which a sensor stream is temporarily unavailable, have a mean error of 1.70 m at 6% occurrence in Table 5.8.

Two conclusions follow for deployment. First, the graceful degradation demonstrated in Chapter 5 is a property of the fusion design and not a substitute for infrastructure. It bounds the loss during transient or partial outages, which is the case the 30% outage test represents, but it does not sustain sub-meter accuracy without WiFi. Second, for safety-critical uses such as emergency response, evacuation guidance, or first-responder tracking, the system should not be the sole positioning source. An infrastructure-free fallback that does not depend on the building's power and network should be provided, for example inertial dead reckoning constrained by a floor plan or portable radio anchors carried by the responders themselves. The application should also expose the WiFi reliability weight so that users are warned when the estimate rests on the inertial and magnetic modalities alone. Neither fallback was implemented or evaluated in this thesis.

6.3.6 Security Considerations for Deployment

Chapter 1 placed security analysis outside the scope of this thesis, and no adversarial evaluation was performed. A system that adapts online to its own sensor inputs nevertheless has an attack surface that differs from that of a static fingerprinting system, and a deployment must account for it. This subsection sets out the threat model for a deployed instance of the proposed system, stating for each threat what the current design already provides and what it does not. The additions named below are engineering measures for a deployment. None of them has been implemented or tested in this work.

Rogue or spoofed access points and RSSI manipulation. An attacker who broadcasts the identifiers of legitimate access points from a different location, or who raises or lowers the transmit power of access points under their control, alters the WiFi fingerprint and can displace the position estimate [142]. Deep-learning fingerprinting models have been shown to be sensitive to such perturbations, and countermeasures have been proposed [114]. The current design offers partial and incidental resistance: the WiFi reliability metric of Section 5.4 responds to increased RSSI variance and to changes in access-point visibility, and the fusion weights then shift toward the inertial and magnetic modalities, as they do during a genuine outage. This response is statistical, not cryptographic. A spoofed environment that is internally consistent and low in variance would not be flagged, and the network was not trained against adversarial inputs. Access-point authentication, which requires support from the network infrastructure, and an anomaly detector on the fingerprint itself are additions for deployment.

Jamming. Deliberate jamming of the WiFi band denies the fingerprint entirely and is, from the system's point of view, indistinguishable from a complete access-point outage [142]. The consequences are those described in Section 6.3.5: the reliability weight on WiFi falls, positioning continues on the inertial and magnetic modalities, and the mean error rises toward the 2.32 m of the IMU-plus-magnetometer configuration. The design detects the condition and degrades rather than failing silently, but it has no countermeasure. Only simulated outages, not deliberate jamming, were evaluated.

Poisoning of the online adaptation. Because OFAM adjusts its fusion weights from the sensor streams it observes, an attacker who can manipulate those streams, for example by combining access-point spoofing with a magnetic disturbance near the user, could attempt to drive the weights toward a modality of their choosing. The current design bounds the damage such an attack can do in three ways. The softmax projection of Equation (5.24) keeps the weights on the probability simplex, so every modality retains a strictly positive weight and no weight can grow without bound. The learning rate decays to its floor of $\eta_{\min} = 0.05$, so once the initial adaptation has passed, each update moves a weight by at most about 0.05 (Equation (5.23)), and the weights cannot be moved abruptly. And the network parameters are frozen at deployment, so the adaptation acts only on the $K = 3$ fusion weights and cannot corrupt the learned representation. The worst outcome is a mis-weighting of the modalities, not a corrupted model. The stability analysis of Section 5.7.7 reports recovery within 30 epochs after the extreme disturbances injected into the simulation. Those disturbances were faults rather than adversaries, and a sustained manipulation that keeps the reliability metrics consistently biased was not tested. A watchdog that monitors the weight trajectory and the reliability metrics and falls back to fixed fusion weights when they leave their empirically observed ranges is the natural engineering addition. It was not implemented.

Confidentiality of the fingerprint database and the anchor cloud. Two data stores require protection. The fingerprint database used for training encodes the access-point topology and floor layout of the buildings, and the spatial anchors stored in Azure Spatial Anchors encode visual features of their interiors (Sec-

tion 3.2.2). The current design provides two safeguards. Inference and adaptation run entirely on the handset, so no position estimate or raw sensor stream leaves the device in normal operation. Chapter 1 relied on this local processing to address privacy, and privacy-preserving localization more generally rests on it [113]. Access to the anchor store is mediated by the account-based access control of the Azure Spatial Anchors service [121], and the anchor set is needed only during ground-truth collection, not at runtime. The design does not protect the distribution of the trained model and its normalization statistics to devices, or any telemetry a deployment might collect. Transport encryption and access control for those channels are deployment additions.

Tampering with the client. A user or attacker with control of the handset can modify the application, replay recorded sensor streams, or feed synthetic ones, and thereby produce arbitrary position reports to any service that trusts the device. The current design does not address this and assumes an honest client. Device attestation and application integrity checks, available on current mobile platforms, are the deployment measures that would raise the cost of such tampering.

In summary, the design already limits what a compromised sensor stream can do to the adaptation, through the simplex projection, the decaying learning rate, and the frozen network, and it keeps location data on the device. It does not authenticate the radio environment, it was not tested against adversarial inputs, and it assumes an honest client. A security evaluation covering these threats is required before deployment in any setting where position reports carry consequences.

6.4 Future Work

The contributions of this thesis establish foundations upon which future research can build. The following directions emerge directly from the findings and limitations identified in the preceding chapters, organized by the thesis contributions they extend.

6.4.1 Extending the Dataset Contribution

The longitudinal dataset's current scope of four buildings and two years of collection provides a foundation that future work could extend along several dimensions. Cross-device data collection using multiple smartphone models would enable investigation of device-invariant fingerprinting techniques, addressing a limitation acknowledged in Section 6.3.1. The single-device collection protocol ensured measurement consistency for temporal analysis but leaves cross-device generalization as an open question.

Environment diversity represents another extension opportunity. The current academic building testbed, while architecturally varied, excludes commercial retail, healthcare, and industrial environments that present distinct challenges: high customer density affecting WiFi propagation, strict electromagnetic compliance requirements, and metallic machinery creating signal shadowing. Future data collection campaigns in these environments would test the generalization of patterns observed in the academic testbed.

6.4.2 Extending the VAE-LSTM Architecture

The hybrid architecture's design reflects specific choices validated through ablation studies but not exhaustively searched. Neural architecture search techniques could potentially identify more efficient configurations that maintain accuracy while reducing computational overhead. The current 32-dimensional latent space and 10-timestep sequence length were empirically tuned; systematic exploration might reveal environment-specific optima.

Additional sensors. If the collection campaigns were repeated with current hardware, several measurements that were unavailable or immature in 2021 and 2022 would be added to the input vector. WiFi fine timing measurement, standardized as round-trip-time ranging in IEEE 802.11mc and refined in 802.11az, gives access-point distances that are less sensitive to transmit-power changes than RSSI [34, 35]. The traces in this thesis predate its wide availability, which is why the RSS–RTT baseline of Section 5.6.8 was evaluated in RSS-only mode [36]. Bluetooth 5.1 direction finding provides angle-of-arrival measurements from multi-antenna beacons [98]. Ultra-wideband ranging, now present in consumer smartphones, provides centimeter-accurate distances where anchor infrastructure exists [22]. A barometer would give a direct floor cue in the multi-floor buildings of the testbed, and the visual-inertial odometry exposed by smartphone AR frameworks would supply a drift-corrected motion estimate in place of raw inertial dead reckoning. The VAE encoder's modular design facilitates appending such features to the input vector, and the reliability-weighted fusion generalizes directly to weighting each new modality by its own quality indicator, for example anchor or beacon visibility for the ranging modalities. Each addition would require a reliability metric of the kind developed in Section 5.4.

Sequence model. The LSTM was chosen for the reasons set out in Section 4.4.5, and no Transformer baseline was trained on this dataset. The modern alternatives are a state-space model such as Mamba [100] or a compact Transformer encoder [13, 96]. State-space models suit this application because they offer linear-time streaming inference with a fixed-size recurrent state, the property that motivated the LSTM. They also model selective long-range dependence, which a ten-step window does not exploit but a longer one might. A head-to-head comparison at matched parameter count on the longitudinal dataset is the natural next experiment. Any replacement must retain the reconstruction branch: OFAM depends on the per-modality reconstruction error as its label-free reliability signal, so a sequence model that discards the generative decoder would remove the mechanism on which the adaptation rests.

Uncertainty estimation. The thesis uses the latent variance and the reconstruction error as its uncertainty indicators because both are already computed for OFAM at no extra cost (Section 4.5.4). The resulting estimate correlates with error ($r = 0.73$) but was not evaluated for coverage or calibration. Two in-network alternatives and a post-hoc conformal wrapper would give the estimate a stronger footing. A heteroscedastic covariance head would predict a position covariance in a single forward pass [103]. Evidential regression

would do the same while separating aleatoric from epistemic uncertainty [110]. A conformal wrapper around the existing estimate or either alternative would convert the point estimate into a region with a finite-sample coverage guarantee, as recent indoor-positioning work has begun to do [107]. The conformal wrapper is the cheapest route, since it requires only a held-out calibration set and no change to the trained network.

6.4.3 Extending the OFAM Contribution

The OFAM convergence to stable feature weights is demonstrated empirically but lacks formal theoretical guarantees. Building on online convex optimization literature, future work could derive regret bounds characterizing OFAM performance relative to an oracle knowing optimal weights in hindsight. Such analysis would strengthen confidence in long-term deployment and inform principled hyperparameter selection.

Privacy-preserving adaptation in multi-user scenarios represents a compelling extension. The current single-user OFAM design performs adaptation locally on the user's device. Federated learning approaches could enable collaborative model improvement across users without centralizing location data, addressing privacy concerns while leveraging collective experience. The reliability-weighted framework may facilitate privacy-preserving aggregation by weighting contributions based on signal quality rather than user identity.

The reliability signal itself is a third direction. OFAM currently consumes the reconstruction error together with hand-designed per-modality statistics whose empirical derivation is noted in Section 6.3.2. An output layer that predicts its own uncertainty, as described in Section 6.4.2, would provide a learned per-modality uncertainty that could replace or complement those statistics, and the watchdog fallback to fixed fusion weights identified in Section 6.3.6 would give the adaptation a bounded failure mode under sustained sensor manipulation. Both preserve the label-free and gradient-free character of the present mechanism.

6.4.4 Deployment and Scaling Considerations

Transitioning from research prototype to commercial deployment introduces practical considerations. Scalability to multi-building enterprise deployments requires efficient fingerprint database management, potentially involving hierarchical indexing that reduces computational complexity as reference point counts grow. The demonstrated accuracy (0.54 m MAE across the York University buildings, 0.59 m on the combined evaluation set) and temporal stability (4% cross-year degradation) position the system competitively for applications in retail analytics, museum navigation, and emergency response coordination. For the last of these, the infrastructure-free fallback recommended in Section 6.3.5 and the security measures identified in Section 6.3.6 are required.

Fifth-generation cellular networks provide dedicated positioning reference signals and, through wide bandwidth and multi-antenna base stations, timing and angle measurements that could complement or partially substitute for WiFi fingerprinting [31, 32]; Section 2.2.3 explains why they were not used here. The evidence does not support 5G as simply superior. A 2026 comparison of commercial 5G and WiFi channel state information found that 5G was more stable over time and generalized better in regression-based localization, whereas WiFi was more discriminative over short intervals and lower in localization

cost [33]. The two carriers thus have complementary strengths, and WiFi remains the more ubiquitous indoor infrastructure. The relevant direction is therefore fusion rather than replacement. A 5G modality would enter the reliability-weighted framework with its own reliability metric, and the longitudinal methodology of Chapter 3 would show whether its signals drift on a different time scale from WiFi, which would require a modified adaptation schedule.

6.5 Closing Remarks

This thesis began with a fundamental problem: indoor localization systems, despite decades of research, continue to suffer from temporal degradation that erodes accuracy over months, computational constraints that limit mobile deployment of sophisticated algorithms, and generalization failures that prevent systems from performing reliably across diverse environments. These interconnected challenges have historically prevented fingerprinting-based localization from achieving the reliability required for practical, long-term deployment.

The research presented here addressed each of these challenges through an integrated approach. Temporal degradation is countered by the Online Feature Adaptation Module, which maintains accuracy through continuous adaptation of the fusion weights, reducing cross-year performance loss from 22% to 4%. Computational efficiency is achieved through the VAE-LSTM hybrid architecture, which completes inference in 58 ms with only 5 ms of adaptation overhead, within the 100 ms budget for real-time 10 Hz operation on standard mobile hardware. Generalization across environments is demonstrated through consistent submeter accuracy (0.49–0.59 m) across four architecturally diverse buildings, enabled by reliability-weighted sensor fusion that accommodates environment-specific signal characteristics. The AR-based ground truth methodology that underpins this work establishes a scalable data collection framework spanning 28,822 m^2 with centimeter-level accuracy.

The integration of AR-based ground truth, VAE-LSTM architecture, and online adaptation provides a unified framework that advances the practical viability of indoor positioning systems. The methodologies, dataset, and architectural patterns presented in this thesis establish a foundation upon which future research can build. As indoor localization moves from laboratory demonstrations toward widespread deployment, adaptive systems that learn and improve over time, rather than degrading, will prove essential. This work demonstrates that such adaptation is achievable within the computational constraints of mobile devices, bringing the vision of reliable indoor positioning closer to reality.

Bibliography

[1] MarketsandMarkets Research, "Indoor location market by component, technology, application, and region—global forecast to 2029." MarketsandMarkets Research Report, 2024. Report ID: TC 2498. Accessed December 2024.

[2] J. Torres-Sospedra, A. R. Jiménez, A. Moreira, T. Lungenstrass, W.-C. Lu, S. Knauth, G. M. Mendoza-Silva, F. Seco, A. Pérez-Navarro, M. J. Nicolau, A. Costa, F. Meneses, J. Farina, J. P. Morales, W.-C. Lu, H.-T. Cheng, S.-S. Yang, S.-H. Fang, Y.-R. Chien, and Y. Tsao, "Off-line evaluation of mobile-centric indoor positioning systems: The experiences from the 2017 IPIN competition," *Sensors*, vol. 18, no. 2, p. 487, 2018.

[3] R. Mautz, *Indoor Positioning Technologies*. Habilitation thesis, ETH Zurich, Zurich, Switzerland, 2012.

[4] International Organization for Standardization and International Electrotechnical Commission, Geneva, Switzerland, *Information technology — Real time locating systems — Test and evaluation of localization and tracking systems*, 2016. ISO/IEC 18305:2016, first edition, 2016-11-01.

[5] R. Want, A. Hopper, V. Falcão, and J. Gibbons, "The active badge location system," *ACM Transactions on Information Systems*, vol. 10, no. 1, pp. 91–102, 1992.

[6] A. Ward, A. Jones, and A. Hopper, "A new location technique for the active office," *IEEE Personal Communications*, vol. 4, no. 5, pp. 42–47, 1997.

[7] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," in *Proceedings IEEE INFOCOM 2000*, vol. 2, pp. 775–784, 2000.

[8] M. Youssef and A. Agrawala, "The Horus WLAN location determination system," in *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services (MobiSys)*, pp. 205–218, 2005.

[9] K. Wu, J. Xiao, Y. Yi, M. Gao, and L. M. Ni, "FILA: Fine-grained indoor localization," in *Proceedings of IEEE INFOCOM 2012*, pp. 2210–2218, 2012.

[10] Apple Inc., "iBeacon." https://developer.apple.com/ibeacon/, 2013. Introduced at the Apple Worldwide Developers Conference, June 2013. Accessed: 2026-09-07.

[11] J. Torres-Sospedra, R. Montoliu, A. Martínez-Usó, J. P. Avariento, T. J. Arnau, M. Benedito-Bordonau, and J. Huerta, "UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems," in *2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 261–270, 2014.

[12] X. Wang, L. Gao, and S. Mao, "CSI-based fingerprinting for indoor localization: A deep learning approach," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 763–776, 2017.

[13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[14] H. Xue, W. Jiang, C. Miao, Y. Yuan, F. Ma, X. Ma, Y. Wang, S. Yao, W. Xu, A. Zhang, and L. Su, "DeepFusion: A deep learning framework for the fusion of heterogeneous sensory data," in *Proceedings of the 20th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, pp. 151–160, 2019.

[15] D. Ungureanu, F. Bogo, S. Galliani, P. Sama, X. Duan, C. Meekhof, J. Stühmer, T. J. Cashman, B. Tekin, J. L. Schönberger, *et al.*, "HoloLens 2 research mode as a tool for computer vision research," 2020.

[16] T. Zhang, D. Zhang, G. Wang, Y. Li, Y. Hu, Q. Sun, and Y. Chen, "RLoc: Towards robust indoor localization by quantifying uncertainty," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 7, no. 4, pp. 1–28, 2023.

[17] J. Ott, J. Pirkl, M. Stahlke, T. Feigl, and C. Mutschler, "Radio foundation models: Pre-training transformers for 5G-based indoor localization," *arXiv preprint arXiv:2410.00617*, 2024.

[18] N. S. Bhatia and K. Obraczka, "Transforming decoder-only transformers for accurate WiFi-telemetry based indoor localization," *arXiv preprint arXiv:2505.15835*, 2025.

[19] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 6, pp. 1067–1080, 2007.

[20] L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil, "LANDMARC: Indoor location sensing using active RFID," *Wireless Networks*, vol. 10, no. 6, pp. 701–710, 2004.

[21] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2568–2599, 2019.

[22] A. Alarifi, A. Al-Salman, M. Alsaleh, A. Alnafessah, S. Al-Hadhrami, M. A. Al-Ammar, and H. S. Al-Khalifa, "Ultra wideband indoor positioning technologies: Analysis and recent advances," *Sensors*, vol. 16, no. 5, p. 707, 2016.

[23] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: part I," *IEEE Robotics & Automation Magazine*, vol. 13, no. 2, pp. 99–110, 2006.

[24] J. Carmigniani, B. Furht, M. Anisetti, P. Ceravolo, E. Damiani, and M. Ivkovic, "Augmented reality technologies, systems and applications," *Multimedia Tools and Applications*, vol. 51, no. 1, pp. 341–377, 2011.

[25] R. T. Azuma, "A survey of augmented reality," *Presence: Teleoperators & Virtual Environments*, vol. 6, no. 4, pp. 355–385, 1997.

[26] A. Yassin, Y. Nasser, M. Awad, A. Al-Dubai, R. Liu, C. Yuen, R. Raulefs, and E. Aboutanios, "Recent advances in indoor localization: A survey on theoretical approaches and applications," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 1327–1346, 2017.

[27] G. Pau, F. Arena, Y. E. Gebremariam, and I. You, "Bluetooth 5.1: An analysis of direction finding capability for high-precision location services," *Sensors*, vol. 21, no. 11, p. 3589, 2021.

[28] Z. Farid, R. Nordin, and M. Ismail, "Recent advances in wireless indoor localization techniques and system," *The Scientific World Journal*, vol. 2013, p. 185138, 2013.

[29] S. O. H. Madgwick, A. J. L. Harrison, and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," in *IEEE International Conference on Rehabilitation Robotics (ICORR)*, pp. 1–7, 2011.

[30] T. Roos, P. Myllymäki, H. Tirri, P. Misikangas, and J. Sievänen, "A probabilistic approach to WLAN user location estimation," *International Journal of Wireless Information Networks*, vol. 9, no. 3, pp. 155–164, 2002.

[31] S. Dwivedi, R. Shreevastav, F. Munier, J. Nygren, I. Siomina, Y. Lyazidi, D. Shrestha, G. Lindmark, P. Ernström, E. Stare, S. M. Razavi, S. Muruganathan, G. Masini, Å. Busin, and F. Gunnarsson, "Positioning in 5G networks," *IEEE Communications Magazine*, vol. 59, no. 11, pp. 38–44, 2021.

[32] Y. Ruan, L. Chen, X. Zhou, Z. Liu, X. Liu, G. Guo, and R. Chen, "iPos-5G: Indoor positioning via commercial 5G NR CSI," *IEEE Internet of Things Journal*, vol. 10, no. 10, pp. 8718–8733, 2023.

[33] Y. Ruan, X. Zhou, Z. Liu, X. Liu, R. Chen, and L. Chen, "Learning-based analysis of 5G and WiFi CSI for indoor localization: Feature stability, model generalization, and performance trade-offs," *Neurocomputing*, vol. 677, p. 133114, 2026.

[34] M. Ibrahim, H. Liu, M. Jawahar, V. Nguyen, M. Gruteser, R. Howard, B. Yu, and F. Bai, "Verification: Accuracy evaluation of WiFi fine time measurements on an open platform," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking (MobiCom)*, pp. 417–427, 2018.

[35] IEEE, New York, NY, USA, *IEEE Standard for Information Technology–Telecommunications and Information Exchange between Systems Local and Metropolitan Area Networks–Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*

Amendment 4: Enhancements for Positioning, 2023. IEEE Std 802.11az-2022, approved 3 December 2022, published 3 March 2023.

[36] X. Feng, K. A. Nguyen, and Z. Luo, "A Wi-Fi RSS-RTT indoor positioning model based on dynamic model switching algorithm," *IEEE Journal of Indoor and Seamless Positioning and Navigation*, vol. 2, pp. 151–165, 2024.

[37] K. Kaemarungsi and P. Krishnamurthy, "Properties of indoor received signal strength for WLAN location fingerprinting," in *Proceedings of the 1st Annual International Conference on Mobile and Ubiquitous Systems (MobiQuitous)*, pp. 14–23, 2004.

[38] G. M. Mendoza-Silva, P. Richter, J. Torres-Sospedra, E. S. Lohan, and J. Huerta, "Long-term WiFi fingerprinting dataset for research on robust indoor positioning," *Data*, vol. 3, no. 1, p. 3, 2018.

[39] I. Silva, C. Pendão, J. Torres-Sospedra, and A. Moreira, "Overcoming radio map degradation in Wi-Fi-based positioning systems," in *2023 13th International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–6, 2023.

[40] J. Ma, X. Wang, Y. Wei, and Z. Zhang, "WiFi sensing with channel state information: A survey," *ACM Computing Surveys*, vol. 52, no. 3, pp. 1–36, 2019.

[41] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "SpotFi: Decimeter level localization using WiFi," in *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication (SIGCOMM)*, pp. 269–282, 2015.

[42] J. Xiao, K. Wu, Y. Yi, and L. M. Ni, "FIFS: Fine-grained indoor fingerprinting system," in *Proceedings of the 21st International Conference on Computer Communications and Networks (ICCCN)*, pp. 1–7, 2012.

[43] X. Wang, L. Gao, S. Mao, and S. Pandey, "PhaseFi: Phase fingerprinting for indoor localization with a deep learning approach," in *2015 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, 2015.

[44] X. Wang, L. Gao, and S. Mao, "BiLoc: Bi-modal deep learning for indoor localization with commodity 5GHz WiFi," *IEEE Access*, vol. 5, pp. 4209–4220, 2017.

[45] H. Chen, Y. Zhang, W. Li, X. Tao, and P. Zhang, "ConFi: Convolutional neural networks based indoor Wi-Fi localization using channel state information," *IEEE Access*, vol. 5, pp. 18066–18074, 2017.

[46] F. Wang, W. Gong, and J. Liu, "On spatial diversity in WiFi-based human activity recognition: A deep learning-based approach," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2035–2047, 2019.

[47] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, pp. 1180–1189, 2015.

[48] I. Silva, C. Pendão, and A. Moreira, "Collection of a continuous long-term dataset for the evaluation of Wi-Fi-fingerprinting-based indoor positioning systems," *Sensors*, vol. 22, no. 22, p. 8585, 2022.

[49] S. He and S.-H. G. Chan, "Wi-Fi fingerprint-based indoor positioning: Recent advances and comparisons," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 466–490, 2016.

[50] H. Chen, Y. Zhang, W. Li, X. Tao, and P. Zhang, "WiFi fingerprinting indoor localization using local feature-based deep LSTM," *IEEE Systems Journal*, vol. 14, no. 2, pp. 3001–3010, 2020.

[51] Y. Fan and H. Sun, "Environmental causality calibration: Advancing WLAN RF fingerprinting for precise indoor localization," *PLOS ONE*, vol. 19, no. 2, p. e0297108, 2024.

[52] N.-Y. Liang, G.-B. Huang, P. Saratchandran, and N. Sundararajan, "A fast and accurate online sequential learning algorithm for feedforward networks," *IEEE Transactions on Neural Networks*, vol. 17, no. 6, pp. 1411–1423, 2006.

[53] H. Zou, X. Lu, H. Jiang, and L. Xie, "A fast and precise indoor localization algorithm based on an online sequential extreme learning machine," *Sensors*, vol. 15, no. 1, pp. 1804–1824, 2015.

[54] H. Zou, M. Jin, H. Jiang, L. Xie, and C. J. Spanos, "WinIPS: WiFi-based non-intrusive indoor positioning system with online radio map construction and adaptation," *IEEE Transactions on Wireless Communications*, vol. 16, no. 12, pp. 8118–8130, 2017.

[55] P. Bolliger, "Redpin – adaptive, zero-configuration indoor localization through user collaboration," in *Proceedings of the First ACM International Workshop on Mobile Entity Localization and Tracking in GPS-less Environments (MELT)*, pp. 55–60, 2008.

[56] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen, "Zee: Zero-effort crowdsourcing for indoor localization," in *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking (MobiCom)*, pp. 293–304, 2012.

[57] A. M. Rajab and B. Wang, "Autonomous and reliable fingerprint map maintenance for indoor positioning system," *Computer Networks*, vol. 251, p. 110626, 2024.

[58] S. J. Pan, V. W. Zheng, Q. Yang, and D. H. Hu, "Transfer learning for WiFi-based indoor localization," in *Proceedings of the AAAI 2008 Workshop on Transfer Learning for Complex Tasks*, AAAI Press, 2008. AAAI Technical Report WS-08-13.

[59] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[60] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7167–7176, 2017.

[61] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, D. Hassabis, C. Clopath, D. Kumaran, and R. Hadsell, "Overcoming catastrophic forgetting in neural networks," *Proceedings of the National Academy of Sciences*, vol. 114, no. 13, pp. 3521–3526, 2017.

[62] D. Wang, E. Shelhamer, S. Liu, B. Olshausen, and T. Darrell, "Tent: Fully test-time adaptation by entropy minimization," in *International Conference on Learning Representations (ICLR)*, 2021.

[63] J. Liang, D. Hu, and J. Feng, "Do we really need to access the source data? Source hypothesis transfer for unsupervised domain adaptation," in *Proceedings of the 37th International Conference on Machine Learning (ICML)*, vol. 119 of *Proceedings of Machine Learning Research*, pp. 6028–6039, PMLR, 2020.

[64] Y. Sun, X. Wang, Z. Liu, J. Miller, A. A. Efros, and M. Hardt, "Test-time training with self-supervision for generalization under distribution shifts," in *Proceedings of the 37th International Conference on Machine Learning (ICML)*, vol. 119 of *Proceedings of Machine Learning Research*, pp. 9229–9248, PMLR, 2020.

[65] X. Li, Y. Wang, and K. Khoshelham, "A robust and adaptive complementary kalman filter based on mahalanobis distance for ultra wideband/inertial measurement unit fusion positioning," *Sensors*, vol. 18, no. 10, p. 3435, 2018.

[66] S. Zhong, L. Cheng, H. Yuan, and X. Li, "Adaptive Kalman filter fusion positioning based on Wi-Fi and vision," *Sensors*, vol. 25, no. 3, p. 671, 2025.

[67] M. Abbas, M. Elhamshary, H. Rizk, M. Torki, and M. Youssef, "WiDeep: WiFi-based accurate and robust indoor localization system using deep learning," in *Proceedings of the IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 1–10, 2019.

[68] F. Potortì, S. Park, A. R. Jiménez Ruiz, P. Barsocchi, M. Girolami, A. Crivello, S. Y. Lee, J. H. Lim, J. Torres-Sospedra, F. Seco, R. Montoliu, G. M. Mendoza-Silva, M. d. C. Pérez Rubio, C. Losada-Gutiérrez, F. Espinosa, and J. Macias-Guarasa, "Comparing the performance of indoor localization systems through the EvAAL framework," *Sensors*, vol. 17, no. 10, p. 2327, 2017.

[69] B. A. Abdalla, H. S. Maghdid, and A. T. Sabir, "A multi-modal dataset for hybrid indoor positioning using Wi-Fi RSS, embedded inertial sensors, and CCTV images," *Data in Brief*, vol. 64, p. 112370, 2025. Available online December 2025.

[70] M. Ammad, A. Michler, P. Schwarzbach, J. Ninnemann, H. Ußler, and O. Michler, "Descriptor: A hybrid indoor and indoor-outdoor positioning multi-technology dataset (HYMN)," *arXiv preprint arXiv:2604.20349*, 2026.

[71] G. M. Mendoza-Silva, P. Richter, J. Torres-Sospedra, E. S. Lohan, and J. Huerta, "Long-term Wi-Fi fingerprinting dataset and supporting material," 2020. Version 2.2, dataset.

[72] G. C. La Delfa, J. Prieto, S. Monteleone, H. Rafique, M. Palesi, and D. Patti, "Survey of smartphone-based datasets for indoor localization: A machine learning perspective," *Internet of Things*, vol. 34, p. 101753, 2025.

[73] R. Mahony, T. Hamel, and J.-M. Pflimlin, "Nonlinear complementary filters on the special orthogonal group," *IEEE Transactions on Automatic Control*, vol. 53, no. 5, pp. 1203–1218, 2008.

[74] Z. Chen, H. Zou, H. Jiang, Q. Zhu, Y. C. Soh, and L. Xie, "Fusion of WiFi, smartphone sensors and landmarks using the Kalman filter for indoor localization," *Sensors*, vol. 15, no. 1, pp. 715–732, 2015.

[75] S. Yao, S. Hu, Y. Zhao, A. Zhang, and T. Abdelzaher, "DeepSense: A unified deep learning framework for time-series mobile sensing data processing," in *Proceedings of the 26th International Conference on World Wide Web (WWW)*, pp. 351–360, 2017.

[76] F. Alhomayani and M. H. Mahoor, "Deep learning methods for fingerprint-based indoor positioning: a review," *Journal of Location Based Services*, vol. 14, no. 3, pp. 129–200, 2020.

[77] T. Qin, P. Li, and S. Shen, "VINS-Mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.

[78] A. Poulose and D. S. Han, "Hybrid indoor localization using IMU sensors and smartphone camera," *Sensors*, vol. 19, no. 23, p. 5084, 2019.

[79] W. Zhang, K. Liu, W. Zhang, Y. Zhang, and J. Gu, "Deep neural networks for wireless localization in indoor and outdoor environments," *Neurocomputing*, vol. 194, pp. 279–287, 2016.

[80] K. S. Kim, S. Lee, and K. Huang, "A scalable deep neural network architecture for multi-building and multi-floor indoor localization based on Wi-Fi fingerprinting," *Big Data Analytics*, vol. 3, no. 4, 2018.

[81] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[82] A. Nessa, B. Adhikari, F. Hussain, and X. N. Fernando, "A survey of machine learning for indoor positioning," *IEEE Access*, vol. 8, pp. 214945–214965, 2020.

[83] X. Feng, K. A. Nguyen, and Z. Luo, "A survey of deep learning approaches for WiFi-based indoor positioning," *Journal of Information and Telecommunication*, vol. 6, no. 2, pp. 163–216, 2022.

[84] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[85] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[86] M. T. Hoang, B. Yuen, X. Dong, T. Lu, R. Westendorp, and K. Bhatt, "Recurrent neural networks for accurate RSSI indoor localization," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10639–10651, 2019.

[87] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724–1734, 2014.

[88] N. D. Lane, S. Bhattacharya, P. Georgiev, C. Forlivesi, L. Jiao, L. Qendro, and F. Kawsar, "DeepX: A software accelerator for low-power deep learning inference on mobile devices," in *Proceedings of the 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pp. 1–12, 2016.

[89] Z. Zhou, X. Chen, E. Li, L. Zeng, K. Luo, and J. Zhang, "Edge intelligence: Paving the last mile of artificial intelligence with edge computing," *Proceedings of the IEEE*, vol. 107, no. 8, pp. 1738–1762, 2019.

[90] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[91] D. J. Rezende, S. Mohamed, and D. Wierstra, "Stochastic backpropagation and approximate inference in deep generative models," in *Proceedings of the 31st International Conference on Machine Learning (ICML)*, pp. 1278–1286, 2014.

[92] A. Kargar-Barzi, E. Farahmand, N. T. Chatrudi, A. Mahani, and M. Shafique, "An edge-based WiFi fingerprinting indoor localization using convolutional neural network and convolutional auto-encoder," *arXiv preprint arXiv:2303.03699*, 2023.

[93] W. Liu, H. Jiang, R. Li, and Z. Deng, "NAS-FL: Fingerprint localization method based on automatically designed neural network architecture," in *Proc. Work-in-Progress Papers, 14th International Conference on Indoor Positioning and Indoor Navigation (IPIN-WiP 2024)*, vol. 3919 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2024.

[94] S. Herath, S. Irandoust, B. Chen, Y. Qian, P. Kim, and Y. Furukawa, "Fusion-DHL: WiFi, IMU, and floorplan fusion for dense history of locations in indoor environments," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5677–5683, IEEE, 2021.

[95] A. Poulose, J. Kim, and D. S. Han, "A sensor fusion framework for indoor localization using smartphone sensors and Wi-Fi RSSI measurements," *Applied Sciences*, vol. 9, no. 20, p. 4379, 2019.

[96] Y. Tay, M. Dehghani, D. Bahri, and D. Metzler, "Efficient transformers: A survey," *ACM Computing Surveys*, vol. 55, no. 6, pp. 1–28, 2022.

[97] S. Masrur, J.-F. Cheng, A. R. Khamesi, and I. Guvenc, "Transforming indoor localization: Advanced transformer architecture for NLOS dominated wireless environments with distributed sensors," *arXiv preprint arXiv:2501.07774*, 2025.

[98] W. Wu, D. Zhou, L. Shen, Z. Zhao, C. Li, and G. Q. Huang, "TransAoA: Transformer-based angle of arrival estimation for BLE indoor localization," *IEEE Transactions on Instrumentation and Measurement*, vol. 74, pp. 1–12, 2025.

[99] H. Rafique, D. Patti, M. Palesi, and G. C. La Delfa, "Fusing visuals with magnetic signals to improve indoor localization using vision transformer," in *2024 14th International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–6, 2024.

[100] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," *arXiv preprint arXiv:2312.00752*, 2023.

[101] J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei, "Scaling laws for neural language models," *arXiv preprint arXiv:2001.08361*, 2020.

[102] J. Gawlikowski, C. Tasber, M. Ali, J. Lee, M. Humt, J. Feng, A. Kruspe, R. Triebel, P. Jung, R. Roscher, M. Shahzad, W. Yang, R. Bamler, and X. X. Zhu, "A survey of uncertainty in deep neural networks," *Artificial Intelligence Review*, vol. 56, no. Suppl 1, pp. 1513–1589, 2023.

[103] A. Kendall and Y. Gal, "What uncertainties do we need in Bayesian deep learning for computer vision?," in *Advances in Neural Information Processing Systems*, vol. 30, pp. 5574–5584, 2017.

[104] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proceedings of the 33rd International Conference on Machine Learning*, pp. 1050–1059, 2016.

[105] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in *Advances in Neural Information Processing Systems*, vol. 30, pp. 6402–6413, 2017.

[106] R. Guan, A. Zhang, M. Li, and Y. Wang, "Measuring uncertainty in signal fingerprinting with Gaussian processes going deep," in *Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–8, 2021.

[107] Z. Zhou, H. Peng, and H. Long, "Conformal prediction for indoor positioning with correctness coverage guarantees," *arXiv preprint arXiv:2505.01810*, 2025.

[108] J. Ma, X. Wang, J. Zhang, S. Mao, S. C. G. Periaswamy, and J. Patton, "Navigating uncertainty: Ambiguity quantification in fingerprinting-based indoor localization," in *2024 IEEE Annual Congress on Artificial Intelligence of Things (AIoT)*, pp. 123–128, 2024.

[109] X. Feng, K. A. Nguyen, and Z. Luo, "Robust indoor positioning with hybrid WiFi RTT-RSS signals," *Sensors*, vol. 26, no. 1, p. 284, 2026.

[110] A. Amini, W. Schwarting, A. Soleimany, and D. Rus, "Deep evidential regression," in *Advances in Neural Information Processing Systems*, vol. 33, pp. 14927–14937, 2020.

[111] J. Xiao, Z. Zhou, Y. Yi, and L. M. Ni, "A survey on wireless indoor localization from the device perspective," *ACM Computing Surveys*, vol. 49, no. 2, pp. 1–31, 2016.

[112] E. S. Lohan, J. Torres-Sospedra, H. Leppäkoski, P. Richter, Z. Peng, and J. Huerta, "Wi-Fi crowdsourced fingerprinting dataset for indoor positioning," *Data*, vol. 2, no. 4, p. 32, 2017.

[113] A. Konstantinidis, G. Chatzimilioudis, D. Zeinalipour-Yazti, P. Mpeis, N. Pelekis, and Y. Theodoridis, "Privacy-preserving indoor localization on smartphones," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 11, pp. 3042–3055, 2015.

[114] S. Tiku and S. Pasricha, "Overcoming security vulnerabilities in deep learning-based indoor localization frameworks on mobile devices," *ACM Transactions on Embedded Computing Systems*, vol. 18, no. 6, pp. 1–24, 2019.

[115] P. Davidson and R. Piché, "A survey of selected indoor positioning methods for smartphones," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 1347–1370, 2017.

[116] J. Kunhoth, A. Karkar, S. Al-Maadeed, and A. Al-Ali, "Indoor positioning and wayfinding systems: a survey," *Human-centric Computing and Information Sciences*, vol. 10, no. 1, p. 18, 2020.

[117] A. Ahmad, P. Claudio, A. Alizadeh Naeini, and G. Sohn, "Wi-Fi RSS fingerprinting for indoor localization using augmented reality," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. V-4-2020, pp. 57–64, 2020.

[118] J. Li, B. Yang, D. Chen, N. Wang, G. Zhang, and H. Bao, "Survey and evaluation of monocular visual-inertial SLAM algorithms for augmented reality," *Virtual Reality & Intelligent Hardware*, vol. 1, no. 4, pp. 386–410, 2019.

[119] P. Hübner, K. Clintworth, Q. Liu, M. Weinmann, and S. Wursthorn, "Evaluation of HoloLens tracking and depth sensing for indoor mapping applications," *Sensors*, vol. 20, no. 4, p. 1021, 2020.

[120] M. Weinmann, M. A. Jäger, S. Wursthorn, B. Jutzi, M. Weinmann, and P. Hübner, "3D indoor mapping with the Microsoft HoloLens: Qualitative and quantitative evaluation by means of geometric features," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. V-1-2020, pp. 165–172, 2020.

[121] Microsoft Corporation, "Azure Spatial Anchors documentation." https://docs.microsoft.com/azure/spatial-anchors/, 2023. Accessed: 2023-10-15.

[122] J. D. Gutiérrez, A. R. Jiménez, F. Seco, F. J. Álvarez, T. Aguilera, J. Torres-Sospedra, and F. Melchor, "Getsensordata: An extensible android-based application for multi-sensor data registration," *SoftwareX*, vol. 19, p. 101186, 2022.

[123] R. Harle, "A survey of indoor inertial positioning systems for pedestrians," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 1281–1293, 2013.

[124] A. Poulose, O. S. Eyobu, and D. S. Han, "An indoor position-estimation algorithm using smartphone IMU sensor data," *IEEE Access*, vol. 7, pp. 11165–11177, 2019.

[125] I. Ashraf, S. Hur, and Y. Park, "MagIO: Magnetic field strength based indoor-outdoor detection with a commercial smartphone," *Micromachines*, vol. 9, no. 10, p. 534, 2018.

[126] Y. Li, Y. Zhuang, H. Lan, P. Zhang, X. Niu, and N. El-Sheimy, "Self-contained indoor pedestrian navigation using smartphone sensors and magnetic features," *IEEE Sensors Journal*, vol. 16, no. 19, pp. 7173–7182, 2016.

[127] H. Fu, C. Li, X. Liu, J. Gao, A. Celikyilmaz, and L. Carin, "Cyclical annealing schedule: A simple approach to mitigating KL vanishing," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 240–250, 2019.

[128] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[129] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Machine Learning*, vol. 8, no. 3–4, pp. 229–256, 1992.

[130] C. P. Burgess, I. Higgins, A. Pal, L. Matthey, N. Watters, G. Desjardins, and A. Lerchner, "Understanding disentangling in β-VAE," *arXiv preprint arXiv:1804.03599*, 2018.

[131] D. J. Rezende and S. Mohamed, "Variational inference with normalizing flows," in *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, vol. 37 of *Proceedings of Machine Learning Research*, pp. 1530–1538, PMLR, 2015.

[132] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, "Normalizing flows for probabilistic modeling and inference," *Journal of Machine Learning Research*, vol. 22, no. 57, pp. 1–64, 2021.

[133] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Advances in Neural Information Processing Systems*, vol. 33, pp. 6840–6851, 2020.

[134] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

[135] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform manifold approximation and projection for dimension reduction," *arXiv preprint arXiv:1802.03426*, 2018.

[136] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222–2232, 2017.

[137] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994.

[138] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[139] L. Bottou, "Online learning and stochastic approximations," in *On-Line Learning in Neural Networks* (D. Saad, ed.), pp. 9–42, Cambridge University Press, 1998.

[140] C. Fischer and H. Gellersen, "Location and navigation support for emergency responders: A survey," *IEEE Pervasive Computing*, vol. 9, no. 1, pp. 38–47, 2010.

[141] J. Rantakokko, J. Rydell, P. Strömbäck, P. Händel, J. Callmer, D. Törnqvist, F. Gustafsson, M. Jobs, and M. Grudén, "Accurate and reliable soldier and first responder indoor positioning: multisensor systems and cooperative localization," *IEEE Wireless Communications*, vol. 18, no. 2, pp. 10–18, 2011.

[142] N. O. Tippenhauer, K. B. Rasmussen, C. Pöpper, and S. Čapkun, "Attacks on public WLAN-based positioning systems," in *Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services (MobiSys)*, pp. 29–40, 2009.

Appendix A

Supplementary Material

This appendix provides supplementary material that complements the main text. Section A.1 presents detailed pseudocode for the key algorithms developed in this thesis, including complete implementation details that would interrupt the flow of the main chapters. Section A.2 documents the sensor data formats and fingerprint structure used throughout the experiments. Section A.3 provides comprehensive hyperparameter configurations for reproducibility, Section A.4 summarizes per-building and per-year dataset statistics, and Section A.5 records the two collection campaigns building by building together with the hardware and software inventory.

A.1 Algorithm Pseudocode

This section presents complete pseudocode for the primary algorithms developed in this thesis. While the main chapters describe the mathematical formulations and design rationale, this appendix provides implementation-level detail suitable for reproduction.

A.1.1 VAE-LSTM Training Procedure

Algorithm 3 presents the complete training procedure for the VAE-LSTM architecture described in Chapter 4. The algorithm includes all preprocessing steps, loss computation, and optimization details.

Algorithm 3 VAE-LSTM Training Procedure. Includes per-building normalization, reparameterization-based sampling, and early stopping with a composite loss balancing position accuracy, KL divergence, and reconstruction quality.

Require: Training dataset $\mathcal{D} = \{(\mathbf{S}_n, \mathbf{x}_n)\}_{n=1}^{N}$ with sensor sequences and ground truth positions

Require: Hyperparameters: learning rate $\eta = 10^{-3}$, batch size $B = 128$, sequence length T, latent dimension d_z

Require: Loss weights: β_1 (KL), β_2 (reconstruction)

Ensure: Trained model parameters θ^*

Initialize:

  $\theta_{\text{enc}} \leftarrow$ Xavier initialization for encoder weights

  $\theta_{\text{lstm}} \leftarrow$ Orthogonal initialization for LSTM weights

  $\theta_{\text{dec}} \leftarrow$ Xavier initialization for decoder weights

  $\theta \leftarrow \{\theta_{\text{enc}}, \theta_{\text{lstm}}, \theta_{\text{dec}}\}$

Preprocess training data:

for each building $b \in \{1,2,3,4\}$ **do**

  Compute per-building mean μ_b and std σ_b for sensor features

  Normalize: $\mathbf{s}_{n,b} \leftarrow (\mathbf{s}_{n,b} - \mu_b)/\sigma_b$

end for

Zero-pad WiFi features to dimension $N_w = 100$

Stage 1 — VAE pretraining ($E_1 = 50$ epochs, reconstruction branch only):

for epoch $e = 1$ to E_1 **do**

  $\beta_1^{(e)} \leftarrow \min(1, e/\tau_{\text{anneal}}) \cdot \beta_1$ ▷ KL annealing, $\tau_{\text{anneal}} = 10$

  for each batch $\mathcal{B} \subset \mathcal{D}$ of size $B = 128$ **do**

    $\mu_z, \log\sigma_z^2 \leftarrow \text{Encoder}(\mathbf{S})$; $\mathbf{z} \leftarrow \mu_z + \sigma_z \odot \varepsilon$, $\varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ ▷ Reparameterization

    $\hat{\mathbf{s}} \leftarrow \text{ReconstructionHead}(\mathbf{z})$ ▷ Position branch and LSTM inactive in Stage 1

    $\mathcal{L}_{\text{stage1}} \leftarrow \beta_1^{(e)} \mathcal{L}_{\text{KL}} + \beta_2 \mathcal{L}_{\text{rec}}$ ▷ No position supervision

    Update $\{\theta_{\text{enc}}, \theta_{\text{dec}}\}$ via Adam ($\eta = 10^{-3}$); clip $\|\nabla\| \leftarrow \min(\|\nabla\|, 1.0)$

  end for

end for

Stage 2 — end-to-end finetuning ($E_2 = 100$ epochs, full loss):

for epoch $e = 1$ to E_2 **do**

  $\beta_1^{(e)} \leftarrow \min(1, e/\tau_{\text{anneal}}) \cdot \beta_1$ ▷ KL annealing continues

  for each batch $\mathcal{B} \subset \mathcal{D}$ of size $B = 128$ **do**

    ▷ Forward: encode → reparameterize → LSTM over T steps → dual-branch decoder

    $\mathbf{z} \leftarrow \mu_z + \sigma_z \odot \varepsilon$; $\mathbf{h}_{\text{temporal}} \leftarrow \text{LSTM}(\mathbf{z}_{1:T})$

    $\hat{\mathbf{x}} \leftarrow \text{PositionHead}(\mathbf{h}_{\text{temporal}})$; $\hat{\mathbf{s}} \leftarrow \text{ReconstructionHead}(\mathbf{h}_{\text{temporal}})$

    $\mathcal{L}_{\text{pos}} \leftarrow \frac{1}{|\mathcal{B}|} \sum_{n \in \mathcal{B}} \|\hat{\mathbf{x}}_n - \mathbf{x}_n\|_2^2$ ▷ Position MSE

    $\mathcal{L}_{\text{KL}} \leftarrow \frac{1}{2} \sum_{j=1}^{d_z} (\mu_j^2 + \sigma_j^2 - \log\sigma_j^2 - 1)$ ▷ KL divergence

    $\mathcal{L}_{\text{rec}} \leftarrow \frac{1}{|\mathcal{B}|} \sum_{n \in \mathcal{B}} \|\hat{\mathbf{s}}_n - \mathbf{s}_n\|_2^2$ ▷ Reconstruction MSE

    $\mathcal{L}_{\text{stage2}} \leftarrow \mathcal{L}_{\text{pos}} + \beta_1^{(e)} \mathcal{L}_{\text{KL}} + \beta_2 \mathcal{L}_{\text{rec}}$

    Update all parameters θ via Adam ($\eta = 10^{-3}$); clip $\|\nabla\| \leftarrow \min(\|\nabla\|, 1.0)$

  end for

  Compute validation MAE $\mathcal{L}_{\text{val}}$ on held-out set

  if $\mathcal{L}_{\text{val}}$ has not improved for 10 epochs **then**

A.1.2 OFAM Weight Update Algorithm

Algorithm 4 presents the complete Online Feature Adaptation Module (OFAM) weight update algorithm described in Chapter 5. This algorithm runs continuously during deployment to adapt sensor fusion weights based on real-time reliability assessment.

Algorithm 4 OFAM Online Weight Update. Computes per-sensor reliability scores from signal stability, motion dynamics, and magnetic field consistency, gates each score by the dual-branch decoder's reconstruction quality, then applies an exponential moving average update with softmax projection in constant time.

Require: Sensor measurements at time t: WiFi RSSI $\mathbf{r}_t$, IMU $(\mathbf{a}_t, \omega_t)$, magnetometer $\mathbf{m}_t$

Require: Decoder reconstruction $\hat{\mathbf{s}}_t$ with per-modality sub-vectors $\hat{\mathbf{s}}_t^{(i)}$; training-set mean reconstruction errors ε_i and modality dimensions D_i

Require: Previous fusion weights $w_{\text{WiFi}}(t-1), w_{\text{IMU}}(t-1), w_{\text{mag}}(t-1)$

Require: Calibration constants $\sigma^2_{\text{ref}} = 25$ dBm^2, N_{expected}, $\tau_{\text{static}} = 0.1$ m/s^2, $\kappa = 0.5$ rad^{-1}s, $\gamma_{\text{mag}} = 0.1\ (\mu\text{T})^{-2}$, $\|\mathbf{m}_{\text{ref}}\|$

Require: Learning-rate schedule $\eta_t = \eta_0 \exp(-t/\tau) + \eta_{\min}$ with $\eta_0 = 0.3$, $\tau = 100$, $\eta_{\min} = 0.05$

Ensure: Updated weights $w_{\text{WiFi}}(t), w_{\text{IMU}}(t), w_{\text{mag}}(t)$

Step 1: WiFi reliability (Equation (5.12))

$\sigma^2_{\text{RSSI}}(t) \leftarrow \frac{1}{N_{\text{ap}}} \sum_j \text{Var}\big(\{r_j^{(k)}\}_{k=t-W+1}^{t}\big)$ ▷ RSSI variance over $W = 30$ scans

$\rho_{\text{WiFi}} \leftarrow \exp\big(-\sigma^2_{\text{RSSI}}(t)/\sigma^2_{\text{ref}}\big) \cdot \min\big(1, N_{\text{visible}}(t)/N_{\text{expected}}\big)$

Step 2: IMU reliability (Equation (5.14))

if $\|\mathbf{a}_t - \mathbf{g}\| < \tau_{\text{static}}$ **then** ▷ $\mathbf{g} = [0,0,9.81]^\top$ m/s^2; device static

    $\rho_{\text{IMU}} \leftarrow 1.0$

else

    $\rho_{\text{IMU}} \leftarrow 0.8 \cdot \exp(-\kappa \|\omega_t\|)$ ▷ Decay with angular velocity

end if

Step 3: Magnetometer reliability (Equation (5.15))

$\rho_{\text{mag}} \leftarrow \min(1, \|\mathbf{m}_t\|/\|\mathbf{m}_{\text{ref}}\|) \cdot \max\big(0,\ 1 - \gamma_{\text{mag}} \sigma^2_{\text{mag}}(t)\big)$

Step 4: Reconstruction-quality gating (Equations (5.16) and (5.17))

for $i \in \{\text{WiFi}, \text{IMU}, \text{mag}\}$ **do**

    $e_i \leftarrow \frac{1}{D_i} \big\|\mathbf{s}_t^{(i)} - \hat{\mathbf{s}}_t^{(i)}\big\|_2^2; \quad q_i \leftarrow \exp(-e_i/\varepsilon_i)$ ▷ Distribution-shift detection

end for

Step 5: Normalize gated reliability (Equation (5.18))

$\rho_t \leftarrow [\rho_{\text{WiFi}}\, q_{\text{WiFi}}, \rho_{\text{IMU}}\, q_{\text{IMU}}, \rho_{\text{mag}}\, q_{\text{mag}}]; \quad \tilde{\rho}_t \leftarrow \rho_t / \sum_j \rho_j(t)\, q_j(t)$

Step 6: EMA update toward normalized reliability (Equation (5.22))

for $i \in \{\text{WiFi}, \text{IMU}, \text{mag}\}$ **do**

    $w_i(t) \leftarrow (1 - \eta_t)\, w_i(t-1) + \eta_t\, \tilde{\rho}_i(t)$

end for

Step 7: Softmax projection onto the simplex (Equation (5.24))

$w_i(t) \leftarrow \exp(w_i(t)) / \sum_j \exp(w_j(t))$ for all i ▷ Smooth, strictly positive weights

return $w_{\text{WiFi}}(t), w_{\text{IMU}}(t), w_{\text{mag}}(t)$

A.1.3 Sensor Fusion Position Estimation

Algorithm 5 presents the complete sensor fusion algorithm that combines individual sensor position estimates using the OFAM-adapted weights. This algorithm is called at each timestep during localization.

Algorithm 5 Weighted Sensor Fusion for Position Estimation. Combines per-modality position estimates using the OFAM-adapted weights through a reliability-weighted average, with stale-measurement handling for missed sensor updates.

Require: Per-modality position contributions $\hat{\mathbf{x}}_t^{(\text{WiFi})}, \hat{\mathbf{x}}_t^{(\text{IMU})}, \hat{\mathbf{x}}_t^{(\text{mag})}$

Require: OFAM weights $w_i(t)$ with $\sum_i w_i(t) = 1$ and $w_i(t) \geq 0$ (Algorithm 4)

Ensure: Fused position estimate $\hat{\mathbf{x}}_t$

1: **Stale-measurement handling** (Equation (5.7))

2: **for** each modality i with no fresh update for Δ steps **do**

3: $\quad \hat{\mathbf{x}}_t^{(i)} \leftarrow \hat{\mathbf{x}}_{t-\Delta}^{(i)}$; $\quad$ inflate uncertainty by $(1 + \gamma_{\text{miss}}\Delta)$ $\quad \triangleright\ \gamma_{\text{miss}} = 0.2$

4: **end for**

5: **Decision-level weighted fusion** (Equation (5.9))

6: $\hat{\mathbf{x}}_t \leftarrow \sum_{i=1}^{K} w_i(t) \cdot \hat{\mathbf{x}}_t^{(i)}$ $\quad \triangleright$ Reliability-weighted average over $K = 3$ modalities

7: **return** $\hat{\mathbf{x}}_t$

A.2 Sensor Data Formats

This section documents the data formats used in the AR-based indoor localization dataset collected for this thesis. Complete format specifications enable reproducibility and facilitate integration with other indoor positioning systems.

A.2.1 WiFi Fingerprint Format

Each WiFi fingerprint record contains RSSI measurements from all visible access points at a given location and timestamp. Table A.1 describes the field structure, and Table A.2 provides a sample record.

Table A.1: WiFi fingerprint record format. Each record pairs a timestamped scan of visible access point RSSI values with building-local coordinates, enabling reproducible fingerprint database construction.

Field	Type	Description
`timestamp`	int64	Unix timestamp in milliseconds
`building_id`	string	Building identifier (PSE, BCE, STC, LRC)
`floor`	int	Floor number (0 = ground floor)
`x_local`	float32	X coordinate in building-local frame (meters)
`y_local`	float32	Y coordinate in building-local frame (meters)
`scan_count`	int	Number of APs detected in this scan
`bssid_list`	string[]	List of AP MAC addresses
`rssi_list`	int8[]	Corresponding RSSI values (dBm)

Table A.2: Sample WiFi fingerprint record from PSE building. The eight visible access points and their RSSI values illustrate the typical scan density observed in academic corridors.

Field	Value
`timestamp`	1634567890123
`building_id`	PSE
`floor`	3
`x_local`	45.23
`y_local`	12.87
`scan_count`	8
`bssid_list`	[00:1A:2B:3C:4D:01, 00:1A:2B:3C:4D:02, ...]
`rssi_list`	[-45, -52, -67, -71, -73, -78, -82, -85]

The WiFi fingerprints are stored in JSON format with one record per line, enabling efficient streaming processing:

```
{"timestamp":1634567890123,"building_id":"PSE","floor":3,
 "x_local":45.23,"y_local":12.87,"scan_count":8,
 "bssid_list":["00:1A:2B:3C:4D:01","00:1A:2B:3C:4D:02",...],
 "rssi_list":[-45,-52,-67,-71,-73,-78,-82,-85]}
```

A.2.2 IMU Data Format

Inertial measurement unit data comprises accelerometer and gyroscope readings at high frequency (100 Hz). Table A.3 describes the field structure.

Table A.3: IMU record format. Accelerometer and gyroscope readings are sampled at 100 Hz, providing the motion signals used for step detection and heading estimation in the fusion pipeline.

Field	Type	Description
`timestamp`	int64	Unix timestamp in milliseconds
`acc_x`	float32	Accelerometer X-axis (m/s^2)
`acc_y`	float32	Accelerometer Y-axis (m/s^2)
`acc_z`	float32	Accelerometer Z-axis (m/s^2)
`gyro_x`	float32	Gyroscope X-axis (rad/s)
`gyro_y`	float32	Gyroscope Y-axis (rad/s)
`gyro_z`	float32	Gyroscope Z-axis (rad/s)
`accuracy`	int	Android sensor accuracy level (0–3)

Sample IMU reading during normal walking:

```
{"timestamp":1634567890100,"acc_x":0.23,"acc_y":-0.15,
 "acc_z":9.78,"gyro_x":0.012,"gyro_y":-0.008,
 "gyro_z":0.145,"accuracy":3}
```

A.2.3 Magnetometer Data Format

Magnetometer readings capture the local magnetic field vector, sampled at 50 Hz. Table A.4 describes the field structure.

Table A.4: Magnetometer record format. Three-axis magnetic field readings at 50 Hz capture location-specific disturbances used as auxiliary fingerprint features by the VAE encoder.

Field	Type	Description
`timestamp`	int64	Unix timestamp in milliseconds
`mag_x`	float32	Magnetic field X-axis (μT)
`mag_y`	float32	Magnetic field Y-axis (μT)
`mag_z`	float32	Magnetic field Z-axis (μT)
`accuracy`	int	Calibration status (0=unreliable, 3=fully calibrated)

Sample magnetometer reading in PSE building (steel-frame construction exhibits characteristic disturbances):

```
{"timestamp":1634567890050,"mag_x":23.4,"mag_y":-12.8,
 "mag_z":41.2,"accuracy":2}
```

A.2.4 AR Ground Truth Format

Ground truth positions are obtained from the MRTK visual-inertial odometry system, providing centimeter-level accuracy when tracking is stable. Table A.5 describes the field structure.

Table A.5: AR ground truth record format. Position, orientation, and tracking confidence fields are registered to building-local coordinates via the spatial anchor procedure described in the main text.

Field	Type	Description
`timestamp`	int64	Unix timestamp in milliseconds
`x`	float32	X position in AR session frame (meters)
`y`	float32	Y position in AR session frame (meters)
`z`	float32	Z position (height) in AR session frame
`qx, qy, qz, qw`	float32	Orientation quaternion
`tracking_state`	enum	TRACKING, PAUSED, or STOPPED
`confidence`	float32	MRTK tracking confidence (0–1)

AR ground truth records are aligned to the building-local coordinate system through the anchor-based registration procedure described in Section 3.2.2.

A.2.5 Synchronized Multi-Modal Record

For model training, sensor streams are synchronized to produce unified multi-modal records at 0.5 Hz. Table A.6 shows the combined format.

Table A.6: Synchronized multi-modal record format. Sensor streams are aligned to 0.5 Hz reference timestamps, producing the 109-dimensional input vectors consumed by the VAE-LSTM architecture.

Field	Type	Description
`timestamp`	int64	Reference timestamp (0.5 Hz)
`building_id`	string	Building identifier
`x_gt, y_gt`	float32	Ground truth position (meters)
`wifi_rssi`	float32[100]	Zero-padded RSSI vector
`imu_acc`	float32[3]	Mean accelerometer over 2s window
`imu_gyro`	float32[3]	Mean gyroscope over 2s window
`mag`	float32[3]	Mean magnetometer over 2s window
`track_quality`	float32	AR tracking confidence

A.3 Hyperparameter Configurations

This section provides complete hyperparameter configurations for all models evaluated in this thesis, enabling exact reproduction of experimental results.

A.3.1 VAE-LSTM Hyperparameters

Table A.7 provides the complete hyperparameter configuration for the VAE-LSTM architecture.

Table A.7: Complete VAE-LSTM hyperparameter configuration. The 32-dimensional latent space, 10-step sequence length, and composite loss weights were selected through cross-validated grid search on the 2021 training partition.

Category	Parameter	Value
Architecture	Input dimension	109 (100 WiFi + 9 IMU/Mag)
	Encoder hidden layers	[256, 128, 64]
	Latent dimension	32
	LSTM hidden size	128
	LSTM layers	2
	Decoder hidden layers	[128, 64]
	Position output dimension	2
Training	Optimizer	Adam
	Learning rate	1×10^{-3}
	Adam momentum	$(0.9, 0.999)$
	Batch size	128
	Stage 1 epochs (VAE pretrain)	50
	Stage 2 epochs (finetune)	100
	Early stopping patience	10 epochs
Loss Weights	Position loss weight	1.0
	KL divergence weight (β_1)	0.1
	Reconstruction weight (β_2)	0.5
	KL annealing	Linear warmup, 10 epochs
Regularization	Dropout rate	0.3
	LSTM dropout	0.3
	Gradient clipping	1.0
	Weight decay	1×10^{-5}
Sequence	Sequence length	10
	Stride	1
	Sampling rate	0.5 Hz

A.3.2 OFAM Hyperparameters

Table A.8 provides the complete OFAM configuration used during online adaptation.

Table A.8: Complete OFAM hyperparameter configuration. Weight bounds, reliability thresholds, and convergence criteria govern the online adaptation behavior validated in the cross-year experiments.

Category	Parameter	Value
Weight Initialization	Initial WiFi weight	0.55
	Initial IMU weight	0.30
	Initial Mag weight	0.15
Observed weight ranges	WiFi weight range	[0.35, 0.75]
	IMU weight range	[0.15, 0.45]
	Mag weight range	[0.08, 0.28]
Adaptation Rate	Base learning rate (η_0)	0.3
	Minimum learning rate ($\eta_{\min}$)	0.05
	Decay time constant (τ)	100 timesteps
WiFi Reliability	Reference RSSI variance (σ^2_{ref})	25 dBm^2
	RSSI window size (W)	30 scans
	Expected AP count (N_{expected})	from radio map
IMU Reliability	Static threshold (τ_{static})	0.1 m/s^2
	Angular-velocity decay (κ)	0.5 rad^{-1}s
Magnetometer Reliability	Reference field ($\|\mathbf{m}_{\text{ref}}\|$)	50 μT
	Variance sensitivity (γ_{mag})	0.1 (μT)$^{-2}$
Convergence	Weight change threshold	0.01
	Consecutive stable epochs	10

A.3.3 Baseline Model Hyperparameters

For completeness, Table A.9 summarizes the hyperparameter configurations used for baseline comparison methods in Chapter 3.

Table A.9: Baseline model hyperparameter configurations. Each model was tuned independently on the validation partition to ensure fair comparison against the proposed VAE-LSTM architecture.

Model	Key Hyperparameters
k-NN	$k = 5$, distance = Euclidean, weights = distance-weighted
Random Forest	trees = 100, max_depth = 20, min_samples_split = 5, min_samples_leaf = 2
SVM (RBF)	$C = 10$, $\gamma = 0.1$, cache_size = 2000 MB
MLP	layers = [256, 128, 64], activation = ReLU, dropout = 0.3, optimizer = Adam, lr = 0.001
LSTM	hidden = 128, layers = 2, dropout = 0.2, sequence = 10
GRU	hidden = 128, layers = 2, dropout = 0.2, sequence = 10

A.4 Dataset Statistics

This section provides detailed statistics for the AR-based indoor localization dataset used in this thesis.

A.4.1 Per-Building Summary

Table A.10 provides comprehensive statistics for each building in the four-building testbed.

Table A.10: Detailed per-building dataset statistics. Floor area, reference points, WiFi infrastructure, sample counts, and magnetic environment characteristics highlight the architectural diversity across the four-building testbed. Floor area, reference points, AP counts, and sample counts are consistent with Tables 3.2 and 3.3.

Statistic	PSE	BCE	STC	LRC
Spatial Coverage				
Total floor area (m^2)	12,544	13,800	702	1,776
Reference points	312	289	87	159
WiFi Infrastructure				
Unique APs observed	89	67	45	38
Mean APs per scan	8.3	6.2	11.4	5.8
Mean RSSI (dBm)	−62.4	−65.1	−58.7	−67.3
Sample Counts				
Samples (2021)	1,842	1,523	428	687
Samples (2022)	2,156	1,687	512	824
Total samples	3,998	3,210	940	1,511
Magnetic Environment				
Mean field magnitude (μT)	48.2	51.4	46.8	49.7
Field std deviation (μT)	12.3	8.7	15.2	9.1

A.4.2 Temporal Distribution

Data collection was organized into two campaigns, in 2021 and 2022, to capture year-over-year environmental change. Table A.11 summarizes the per-building sample counts for each campaign.

Table A.11: Per-building fingerprint sample counts for the 2021 and 2022 collection campaigns, totaling 9,659 samples across the four-building testbed, consistent with Table 3.3.

Building	2021	2022	Total
PSE	1,842	2,156	3,998
BCE	1,523	1,687	3,210
STC	428	512	940
LRC	687	824	1,511
Total	4,480	5,179	9,659

A.4.3 Train/Test Splits

For cross-year evaluation, the dataset is split by collection campaign, as described in Section 3.8.1. Forward cross-year validation trains on the 2021 campaign and tests on the 2022 campaign:

- **Training set**: 2021 campaign (4,480 samples, 46.4%)
- **Test set**: 2022 campaign (5,179 samples, 53.6%)

Hyperparameters are selected by cross-validation within the training set, following the trajectory-disjoint 70/30 within-year protocol of Section 3.8.1. This temporal split ensures that test-set performance reflects the model's ability to generalize across time, capturing the effects of environmental drift that are central to this thesis.

A.5 Data Collection Campaign Log

This section records the two collection campaigns at the level of individual buildings and lists the hardware and software used for ground-truth acquisition, sensor logging, and model training. Table A.12 gives one row per building and campaign. The collection windows are approximate. They are derived from the campaign schedule in Sections 3.5.1 and 3.5.2, not from per-session logs. Each six-week campaign ran from early September to mid-October, visited the buildings in the order PSE, BCE, STC, LRC, and spent 2 to 4 collection days in each building. Every session was staffed by two operators, one wearing the HoloLens 2 and one carrying the Android handset. The same HoloLens operator served in both campaigns, while Android operators were rotated. The 2022 campaign opened with the anchor re-verification phase described in Section 3.5, in which stored anchors were re-localized before fingerprint collection resumed. Sample counts are those of Table A.11.

Table A.12: Data collection campaign log by building and campaign. The approximate windows are derived from the campaign schedule (six weeks from early September to mid-October, buildings visited in the order PSE, BCE, STC, LRC, with 2 to 4 collection days per building) rather than from session logs; weeks are counted within the named month. Operators are listed by role: each session comprised one HoloLens operator and one Android operator. Sample counts are those of Table A.11.

Building	Approximate window	Operators	Samples
2021 campaign			
PSE	September, weeks 1–2	HoloLens, Android	1,842
BCE	September, weeks 3–4	HoloLens, Android	1,523
STC	October, week 1	HoloLens, Android	428
LRC	October, week 2	HoloLens, Android	687
2022 campaign			
PSE	September, weeks 1–2	HoloLens, Android	2,156
BCE	September, weeks 3–4	HoloLens, Android	1,687
STC	October, week 1	HoloLens, Android	512
LRC	October, week 2	HoloLens, Android	824

Table A.13 lists the hardware and software components. The sensor sampling rates and the synchronization tolerance are those reported in Section 3.3; the anchor validation design is that of Section 3.2.3; the

training hardware is that reported in Chapter 4.

Table A.13: Hardware and software inventory for ground-truth acquisition, sensor logging, and model training.

Component	Role	Version or configuration
Ground truth		
Microsoft HoloLens 2	Spatial mapping, anchor creation, and operator guidance	–
Custom MRTK anchor application	Anchor placement, CAD registration, and re-localization	Built on MRTK and Azure Spatial Anchors
Azure Spatial Anchors	Cloud persistence of anchors across sessions and years	Cloud service
Leica TS15 total station	Independent survey of the anchor validation points	80 points, 3 sessions, 240 placements
Sensor acquisition		
Samsung Galaxy S21	Android handset for WiFi, IMU, and magnetometer logging	–
GetSensorData	Sensor logging application	WiFi 0.5 Hz, IMU 100 Hz, magnetometer 50 Hz
NTP synchronization	Timestamp alignment between HoloLens and Android streams	Within 50 ms
Model training		
NVIDIA RTX 3080	Offline training of the VAE-LSTM and the baseline models	Single GPU
Deep learning framework	Model implementation and training	PyTorch