

**INTEGRATIVE AND MULTI-SCALE DEEP LEARNING FOR 3D POINT
CLOUD TRANSMISSION CORRIDOR SCENE SEGMENTATION: NOISE
FILTERING, ATTENTION-FUSED FEATURE INTEGRATION, AND
PANOPTIC NETWORK**

MARYAM JAMEELA

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN EARTH AND SPACE SCIENCE

YORK UNIVERSITY
TORONTO, ONTARIO

NOVEMBER 2023

© Maryam Jameela, 2023

Abstract

LiDAR technology plays a crucial role in mapping natural and built environments across various civic and military applications. It enables the acquisition of high-density 3D point clouds with pulse repetition frequencies ranging from 100Hz to 2MHz. However, the increased overlap with atmospheric points has posed challenges in noise filtering and 3D point cloud quality. This dissertation proposes the Noise Seeking Attention Network (NSANet), a novel solution integrating psychological theories of feature integration and attention engagement. NSANet achieves a 4.10% increase in F1-Score and a 7.30% improvement in recall by employing multiscale context, global physical priors, and local spatial attention for noise filtering, surpassing previous techniques.

The study explores the relationship between vegetation height and plantation guidelines, identifying spatial layout consistency in utility layouts and transmission objects. This insight drives the development of three semantic analysis approaches: Semantic Utility Network (SUNet), Fusion-Semantic Utility Network (Fusion-SUNet), and Panoptic-Semantic Utility Network (Pan-SUNet). Encouraged by the performance improvements of SUNet and Fusion-SUNet, Pan-SUNet achieves outstanding results, boasting a 94% F1-Score for pylons, 99% for ground, vegetation, and powerlines, and demonstrating high precision in 3D object detection.

Experiments conducted on Teledyne Optech’s Galaxy T1000 dataset, which features diverse voltage transmission lines, validate the effectiveness of Pan-SUNet, particularly when combined with the RandLA baseline. Significant improvements are observed, including an increase from 80% to 85% in F1-Score for pylons, 98% to 99% for ground, 93% to 97% for powerline, 75% to 78.3% for other objects, 86% to 88% for buildings, while maintaining 98.2%

for high vegetation and 93% for medium vegetation.

The key contribution of this research is a significant advancement beyond basic object classification, as it not only identifies the class of an object but also distinguishes between different instances of the same class. This instance segmentation is critical for utility network modeling and simulation.

The research emphasizes the importance of external cues such as contextual reasoning, spatial cognition, and physical priors for multiscale fusion in scene understanding systems. Moreover, the study’s adaptability to integrate proposed contributions into existing networks enhances their overall performance, making them network-agnostic.

November 5th, 2023

To Hasan and Aaira!

Acknowledgements

Science and this program have profoundly strengthened my connection with Allah like nothing else in my life. I am deeply grateful for the opportunities, context, and logic it provided, enabling me to explore the intricate details of nature.

Everything I have achieved and possess in my life during last five years is a direct result of the guidance and mentorship of Dr. Gunho Sohn. I wouldn't be the person I am today if he hadn't challenged me with countless 'Whys' on a daily basis.

I extend my humble gratitude to my supervisory committee, Dr. Costas Armenakis, and Dr. Hui Jiang, for their unwavering support over the five years. I also appreciate the valuable time dedicated by external and internal examiners, Dr. Jinha Jung, Dr. Voula Pantazopoulou, and Dr. Pirathayini Srikantha.

In the academic realm, I greatly admire Dr. Jungwon Kang, Dr. Connie Ko, Dr. Mojgan Jadidi, Dr. Yujia Zhang, and Dr. Yeonjeong Jeong. I am thankful for their consistent encouragement.

Professionally, I want to acknowledge Burns Foster, Alvin Poernomo, Hamdy Elsyed, and Zhongwei Huang.

On a personal level, this journey wouldn't have begun without the sincere support of Dr. Muhammad Usman and Aneeqa Ali, who not only took my passion for graduate studies seriously but also helped me recognize the opportunities before me.

I am profoundly grateful for Aman Ullah Usmani, who has been my constant and safe haven in this otherwise chaotic world. He provided me with a space to express my frustrations, vents, and challenges with the complexities of science and the world at large.

I am blessed with incredible friends, but these four individuals – Aleena Hassan Qureshi, Mahya Joderi Rad, Dr. Maryam Parsian, and Zahra Arjmandi – stood by me during my lowest moments, providing unwavering support.

My lab mates, Jacob Yoo, Hyungju Lee, Kunwoo Park, Kang Zhao, Kivanc Babacan, David Recchia, Ali Hassan, and Afnan Ahmad, played a crucial role in my journey. Their constant support kept me going.

I also want to express my gratitude to friends from home – Ali Aosjah Mohammad, Azeema Batool, and Adeena Mirza – who, despite the time difference, were always there for me.

Lastly, but certainly not least, I owe my deepest gratitude to my parents and my brother. They are my pillars of strength. I wouldn't have lasted five days, let alone five years, without their encouragement and support at every step of my journey.

Table of Contents

Abstract	ii
Acknowledgements	v
Table of Contents	vii
List of Tables	xii
List of Figures	xv
Abbreviations	xxii
1 Introduction	1
2 Literature Review	13
2.1 Airborne Laser Scanning: Pulse in Air (PIA)	14
2.1.1 SPIA	17
2.1.2 MPIA	17
2.2 Deep Learning in Computer Vision	19
2.2.1 Physics-based Learning: Domain-specific External Cues	20
2.2.2 Multiscale Deep Learning	20
2.2.3 3D Semantic Segmentation	21
2.2.4 2D Object Detection	22
2.2.5 Panoptic Segmentation	23

2.3	Data Augmentation	23
2.3.1	Simulations	24
2.3.2	Deep Learning	25
2.4	Noise Filtering Techniques	25
2.4.1	Statistical and Geometrical Noise Filtering	26
2.4.2	Cluster-based Noise Filtering	26
2.4.3	Machine Learning based Noise Filtering	27
2.4.4	Deep Learning based Noise Filtering	27
2.5	Multiscale Feature Fusion	28
2.5.1	Attention	28
2.6	Multimodal Fusion	32
2.7	Utility Scene Understanding	32
2.7.1	Deep Learning Spatial Decision Making for Utility Network	33
2.7.2	Utility Powerline Network	34
2.7.3	Utility Vegetation Management	37
2.7.4	Historical Context of High-Voltage Transmission Line Maintenance in Canada	38
2.7.5	Computational Techniques in Utility Infrastructure Analysis	39
2.8	Summary	44
3	Simulation-based Data Augmentation using Physical Priors	45
3.1	Overview	45
3.2	Methodology	47
3.2.1	P2-Simulation	48
3.2.2	GM-Simulation	49
3.2.3	Noise Filtering Neural Network	51
3.3	Experiment	52
3.3.1	Dataset	53
3.3.2	Experimental Configuration	53

3.3.3	Evaluation Metrics	54
3.4	Results and Discussion	54
3.4.1	Noise Simulation	54
3.4.2	Noise Filtering Neural Network	55
3.4.3	Simulation-based Data Augmentation	56
3.5	Summary	65
4	Noise Filtering Deep Neural Network using Dual-Attention	66
4.1	Overview	66
4.1.1	Data Analysis	68
4.1.2	Physical Priors	69
4.2	Methodology	70
4.2.1	3D Noise Filtering Neural Network	71
4.2.2	Physical Priors Estimation Module	72
4.2.3	Dual Attention Module	74
4.2.4	Voxelization	76
4.3	Experiment	76
4.3.1	Dataset	77
4.3.2	Experimental Configuration	78
4.4	Results and Discussion	79
4.4.1	NSANet Ablation Study	79
4.4.2	Feature Engineering	80
4.4.3	Loss Functions	81
4.4.4	Attention Ablation Study	83
4.4.5	Comparative Study	83
4.4.6	Error Analysis	85
4.5	Summary	85

5	Multiscale Deep Learning for Utility Network Segmentation	89
5.1	Overview	91
5.2	SUNet-Semantic Utility Network	92
5.2.1	3D Semantic Segmentation Pipeline	94
5.2.2	2D BEV Pipeline	97
5.2.3	Voxelization and BEV Projection	98
5.2.4	Absolute and Relative Elevation Module	99
5.2.5	Loss-based Late Fusion	99
5.3	Fusion-SUNet	101
5.3.1	Fusion	101
5.4	Pan-SUNet: Panoptic Semantic Utility Network	105
5.4.1	Panoptic Segmentation	106
5.4.2	2D BEV Modified Pipeline	107
5.4.3	Fusion	109
5.5	Summary	111
6	Deep Learning Evaluation Framework for Utility Network Segmentation	112
6.1	Dataset	112
6.1.1	Riegl Q560	114
6.1.2	Teledyne Optech Galaxy T1000	114
6.1.3	Qualitative Comparision of Two Datasets	114
6.2	Experimental Configuration	119
6.2.1	SUNet and Fusion-SUNet	119
6.2.2	Pan-SUNet	123
6.3	Evalaution Metrics	126
6.3.1	Semantic Segmentation	126
6.3.2	Object Detection	126
6.3.3	Ablation Studies	127
6.4	Summary	128

7	Experimental Results and Discussion for Utility Network Segmentation	129
7.1	SUNet Performance Evaluation on Riegl Dataset	129
7.1.1	Analysis of Loss Functions and Elevation Features	130
7.1.2	Module-based Ablation Study	132
7.2	Fusion-SUNet Performance on Riegl Dataset	137
7.2.1	Recall Analysis	137
7.2.2	Precision Analysis	138
7.2.3	F1-score Analysis	138
7.3	Pan-SUNet Performance Evaluation on Riegl Dataset	140
7.3.1	2D Object Detection Results	140
7.3.2	2D Regional Segmentation Results	144
7.3.3	3D Semantic Segmentation Results	144
7.3.4	3D Object Detection Results	149
7.4	Pan-SUNet Performance: Teledyne Galaxy T1000	149
7.4.1	2D Object Detection Results	151
7.4.2	2D Regional Segmentation Results	153
7.4.3	3D Semantic Segmentation Results	156
7.4.4	3D Object Detection Results	164
7.5	Summary	167
8	Conclusions and Future Work	168
8.1	Summary of Key Findings	168
8.2	Strengths and Contributions	171
8.3	Limitations and Prospects of Future Directions:	173
8.4	Concluding Remarks	176
	Bibliography	177

List of Tables

3.1	Comparison of 3D U-Net for noise filtering with SVM, Denoising Autoencoder, and PointNet over Recall, Precision and F1-score.	56
3.2	3D U-Net results for noise class using with and without augmentation for individual test scenes.	61
4.1	Ablation study of modules for NSANet to compare the performance over recall, precision, and F1-score.	79
4.2	Ablation study on the importance of feature engineering for optimized NSANet-1 over recall, precision, and F1-score; Occ: No. of Occupancy Points, MZ: Mean Elevation, INS: Intensity, R: Range and NR: Number of Returns. . . .	80
4.3	Ablation study on the importance of loss function over Recall, Precision, F1-score; CE: Cross Entropy Loss, FL: Focal Loss, and WCE: Weighted Cross-Entropy	82
4.4	Ablation Study of global, local and dual attention modules about DDA-module in Figure no. 4.6 to compare performance advantage over Recall, Precision, and F1-score.	82
4.5	Comparison of NSANet for noise filtering with SVM, Denoising Autoencoder, 3D U-Net, PointNet++, and PointNet over Recall, Precision, and F1-score for noise class.	83
4.6	Ablation Study based on voxel post-processing module and compared performance over Recall, Precision, and F1-score.	85

6.1	Comparison of Teledyne and Riegl dataset.	113
7.1	Comparative analysis of SUNet ablation: evaluating the impact of loss-based late fusion on F1-score performance with WCE and HLC considerations. . .	130
7.2	SUNet ablation study assessing feature engineering impact on F1-score: evaluating occupancy points, absolute and relative elevation, and number of returns.	131
7.3	SUNet ablation study examining the impact of MFA and FS on Recall. . . .	132
7.4	SUNet ablation study examining the impact of MFA and FS on Precision. .	132
7.5	SUNet ablation study examining the impact of MFA and FS on F1-scores. .	132
7.6	Comparative study of Fusion-SUNet for Riegl dataset on Recall.	139
7.7	Comparative study of Fusion-SUNet for Riegl dataset on Precision.	139
7.8	Comparative study of Fusion-SUNet for Riegl dataset on F1-score.	139
7.9	Comparative study for 2D object detection for pylon and span objects for average precision (AP) and average recall (AR) over IoU=0.5.	140
7.10	Class-specific evaluation for 2D object detection of pylon and span objects for average precision (AP) over IoU=0.5.	141
7.11	The results of regional segmentation results for corridor class in the Riegl dataset.	144
7.12	Riegl dataset's ablation study of Pan-SUNet performance comparison for Recall.	147
7.13	Riegl dataset's ablation study of Pan-SUNet to compare performance on Precision.	147
7.14	Riegl dataset's ablation study of Pan-SUNet to compare performance on F1-score.	148
7.15	Riegl dataset's cross network comparative study with Pan-SUNet on F1-score.	148
7.16	Comparative performance matrices of Pan-SUNet for 2D and 3D object detection on Riegl dataset: highlighting overall AP, AR, and class-wise AP for pylon and span detection at IoU=0.5.	149
7.17	Comparative study for 2D object detection for Teledyne dataset for pylon and span objects for average precision (AP) and average recall (AR) over IoU=0.5.	151
7.18	Comparative study for 2D object detection for Teledyne dataset for pylon and span objects for pylon average precision (AP) and span average precision (AP).	153

7.19	Comparative evaluation of 3D regional segmentation using Teledyne Optech’s Galaxy T1000 and Riegl datasets.	156
7.20	Recall metrics comparison for Pan-SUNet with ablation setups: three baline networks, HLC, Panoptic fusion integration, 2D regional segmentation and datasets.	157
7.21	Precision metrics comparison for Pan-SUNet with ablation setups: three baline networks, HLC, Panoptic fusion integration, 2D regional segmentation and datasets.	157
7.22	F1-score metrics comparison for Pan-SUNet with ablation setups: three baline networks, HLC, Panoptic fusion integration, 2D regional segmentation and datasets.	158
7.23	Teledyne dataset’s compartive study with Pan-SUNet on F1-score.	163
7.24	Comparative performance matrices of Pan-SUNet for 2D and 3D object detection on Teledyne Optech’s dataset: highlighting overall AP, AR, and class-wise AP for pylon and span detection at IoU=0.5.	167

List of Figures

1.1	Overview of 3D Integrative and Multiscale Deep Learning System.	5
2.1	Airborne LiDAR.	15
2.2	Pulse in air zones with point cloud and noise on transition point. R_{max} shows the range a pulse travels before next one is fired.	16
2.3	(a) SPIA zone shows one pulse sent and received before next is fired and (b) MPIA fires second pulse before first is returned.	18
2.4	Utility Powerline Network.	35
2.5	Utility transmission zones' width by BC Hydro.	36
2.6	Utility transmission zones.	37
3.1	3D Point Cloud from Teledyne Optech's Galaxy T1000; Red Points represent Noise.	46
3.2	Parametric Generative Model: A Simulation workflow which explains that P2-Simulation only uses LiDAR physics, and GM-Simulation uses Gaussian function and LiDAR physics to generate noise.	47
3.3	3D U-Net Architecture is a baseline multi-resolution encoder and decoder with skip-connection, which is used to benchmark the performance of simulation-based augmentation and real noise-based filtering results.	52
3.4	GM-Simulation Synthetic vs Real data; red: noise points and blue: regular points.	55

3.5	Visualization comparison of results for Test Scene 1; a) Ground truth, b) 3D U-Net (w/o aug.) and c) 3D U-Net (aug.). Red points are noise, and blue points are regular non-noise objects.	57
3.6	Visualization comparison of results for Test Scene 2; a) Ground truth, b) 3D U-Net (w/o aug.) and c) 3D U-Net (aug.). Red points are noise, and blue points are regular non-noise objects.	58
3.7	Visualization comparison of results for Test Scene 3; a) Ground truth, b) 3D U-Net (w/o aug.) and c) 3D U-Net (aug.). Red points are noise, and blue points are regular non-noise objects.	59
3.8	Visualization comparison of results for Test Scene 4; a) Ground truth, b) 3D U-Net (w/o aug.) and c) 3D U-Net (aug.).	60
3.9	Comparison of the distribution of a) real noise from Figure 3.4a) and b) GM-Simulation synthetic noise from Figure 3.4b).	61
3.10	Two voxels from one scene, a) sparse noise far from the terrain, b) complex noise overlapping with bushes; red: noise points, blue: regular points and green: centroid.	62
3.11	Normal distribution of noise points for voxels from Figure 3.10.	62
3.12	Normal Distribution of regular points for voxels from Figure 3.10.	63
3.13	Raw point cloud of $20m^3$ relatively far from terrain containing only noise, a) sparse noise, b) dense noise relatively close to the terrain than a); red: noise points and green: centroid.	63
3.14	Normal distribution of noise points for voxels from Figure 3.13.	64
4.1	a) SPIA is a traditional technology which matched the signal sent by a transmitter (S) and received by a receiver(R) efficiently but was expensive, b) MPIA technology improved the cost and speed up the data collection though it is more prone to atmospheric points which can be seen overlapping with objects of interest in c), d) The proposed NSANet uses physical priors and spatial attention to filter noise from the point cloud.	67

4.2	The three noise types: (1) Complex, (2) Sparse Noise, and (3) Global Systematic.	70
4.3	NSANet-1 based on AET. It uses the Physical Priors Estimation (PPE) module to generate a PIA proximity probability map converted into a voxel grid and downsample by a factor of 2. The multi-resolution encoder-decoder uses a Decoder Dual-Attention (DDA) module to eliminate noise based on local and global contexts. The Voxel Post-processing (VPP) module eliminates borderline misclassification of noise points.	71
4.4	Physical priors estimation module workflow.	72
4.5	Physical priors calculated using algorithm no. 1, Higher the proximity to PIA transition-zone, high probability is of point to be part of a global systematic noise pattern as shown with red color.	74
4.6	Dual attention module calculates attention coefficient based on the local and global feature map. Local feature map from the encoder x_l and the previous layer of the decoder or encoder h_{l-1} and global feature maps comes from physical priors g_r , which are calculated based on the global positioning of each point in the point cloud from the PPE module.	76
4.7	NSANet-2 adapted FIT by employing attention module in encoder as well. The first variation contains only local attention in encoder layers called the encoder single attention (ESA) module.	77
4.8	NSANet-3 adapted FIT by employing attention module in encoder as well. This variation contains dual attention in encoder layers called encoder dual attention (EDA) module.	78
4.9	3D scene with sparse noise spread across scene shows four different scenes comparing the SOTA performance of baseline with weighted cross-entropy loss and NSANet-1 variation employed AET and weighted cross-entropy loss. Results reflect NSANet's superior performance. Red shows noise points and white reflects non-noise points.	86

4.10	3D scene with dense noise pattern shows four different scenes comparing the SOTA performance of baseline with weighted cross-entropy loss and NSANet-1 variation employed AET and weighted cross-entropy loss. Results reflect NSANet’s superior performance. Red shows noise points and white reflects non-noise points.	87
5.1	The three proposed utility scene understanding approaches and the fusion modules. a) SUNet contributes loss-based late fusion, b) Fusion-SUNet builds on SUNet and proposed deep fusion module, and c) Pan-SUNet builds on SUNet’s loss-based late fusion using regional segmentation (RS) and incorporates 2D Object detection (OD) for panoptic fusion.	90
5.2	SUNet is a multi-dimensional and multi-resolution network that imposes the spatial layout consistency (a) through 2D utility regions on the outcomes of 3D segmentation network via loss-based late fusion (b).	93
5.3	The overall architecture of SUNet: Point cloud is preprocessed into voxel grid and BEV. These are separately processed by a multi-resolution 3D semantic segmentation pipeline and 2D BEV pipeline for regional prediction to impose the spatial layout consistency on 3D objects through loss-based late fusion.	94
5.4	The overall architecture of Fusion-SUNet: Point cloud is preprocessed into voxel grid and BEV. These are separately processed by a multi-resolution 3D semantic segmentation pipeline and 2D BEV pipeline for regional prediction to impose the spatial layout consistency on 3D objects through feature-based and loss-based late fusion.	102
5.5	A-Fusion module to impose spatial layout consistency through the fusion of features, also known as middle fusion.	104

5.6	The Overall architecture of Pan-SUNet: 3D voxel grid and 2D BEV is generated from point cloud. 3D Voxel grid is convoluted through multi-resolution 3D semantic segmentation. 2D BEV uses a shared encoder for segmentation regional prediction of corridor and non-corridor classes. The detection head produces span and pylon objects that can later fuse regional masks to impose the spatial layout consistency on 3D objects through the hierarchical layout consistency loss function. Panoptic fusion fuses the 2D Object and semantic segmentation to project onto 3D objects and readjusts semantic segmentation labels.	106
5.7	Regional fusion generation of segmentation shows a) polygon generated from step 1 and b) segmentation mask generated from polygon.	107
5.8	Regional fusion from 2D regional segmentation and utility segmentation mask.	109
5.9	Panoptic Fusion: Conversion of the 2D Bounding boxes objects to 3D objects.	110
6.1	Riegl and Teledyne Optech sensors used to acquire dataset.	113
6.2	Elevation-based visual comparison between Riegl and Teledyne Optech Datasets.	115
6.3	Comparative visualization of echo signatures for High Vegetation, Powerline, and Pylon classes in Riegl and Teledyne Optech datasets.	116
6.4	Elevation profile visualization for Pylon, Powerline, and High Vegetation features in Riegl and Teledyne Optech datasets.	118
6.5	Comparative visualization across density, echo, and elevation attributes in Teledyne Optech Galaxy T1000 and Riegl Q560 datasets.	120
6.6	Class map visualization: comparative analysis of Riegl and Teledyne Optech datasets.	121
6.7	Comparative visualization of span and pylon detection: horizontal vs. oriented bounding boxes.	123
6.8	Visualization of input feature channels for the 2D regional pipeline.	125

7.1	Comparative Visualization: Ground Truth vs. SUNet vs. SUNet with MFA - Classifying Pylons (blue), Powerlines (red), Vegetation (green), and Ground (dark green).	134
7.2	Error Analysis: SUNet versus SUNet with MFA - Classifying Pylons (blue), Powerlines (red), Vegetation (green), and Ground (dark green).	135
7.3	Visualization of results of 2D objects span and pylon; a) Ground truth, b) Faster RCNN, and c) YOLO v3.	142
7.4	Visualization of results of 2D objects span and pylon; a) Ground truth, b) Rotated Faster RCNN, c) Oriented RCNN, and d) DAFNe.	143
7.5	Visualization of Pan-SUNet's segmentation performance on the Riegl dataset: ground truth and three Regions (blue: non-corridor, green: corridor, and yellow: utility zone).	145
7.6	Comparative results of ground truth vs Pan-SUNet's semantic segmentation and Pan-SUNet's Panoptic segmentation results for utility corridor scene: blue: pylon, red: powerline, green: high vegetation and orange: ground, and Pan-SUNet Panoptic segmentation.	146
7.7	Noise removal process in Teledyne Optech T1000 dataset using NASNet. . .	150
7.8	Teledyne dataset Visualization of results of 2D objects span and pylon for high voltage span and pylon.	152
7.9	Teledyne dataset Visualization of results of 2D objects span and pylon for overlapping wires.	154
7.10	Teledyne dataset Visualization of results of 2D objects span and pylon for high voltage transmission and distribution spans and pylon.	155
7.11	Performance comparison of wires of different voltage in Teledyne Optech dataset. The results reflect the lower performance on low voltage powerlines.	159
7.12	Performance comparison of pylons of different heights in Teledyne Optech dataset. The results reflect the lower performance on small size pylons. . . .	160

7.13	Visualization showing the problematic areas of low voltage powerline being misclassified while high voltage powerline are correctly classified.	161
7.14	Visualization of high voltage powerline and tall steel pylons reflecting the ease of segmentation for Pan-SUNet irrespective of baseline.	162
7.15	Comparison of the performance between the building class and low voltage powerline highlights the challenge in generalization. This disparity reflects the degradation in performance and misclassification issues, especially for buildings with flat surfaces.	163
7.16	F1-score comparison for object detection on Teledyne vs. Riegl datasets with varied point densities.	164
7.17	Comparative misclassification analysis of 230kV voltage powerline and medium wood pylons by different 3D segmentation models.	165
7.18	Instance segmentation of span objects and pylons detected by Pan-SUNet with unique ID coloring.	166

Abbreviations

AET Attentional Engagement Theory

A-Fusion Attention Fusion

AI Artificial Intelligence

ALS Airborne Laser Scanning

ALTM Airborne Laser Terrain Mapping

ARE Absolute and Relative Elevation

BEV Bird's Eye View

CNN Convolution Neural Network

DDA Decoder Dual Attention

DCNN Deep Convolution Neural Network

EDA Encoder Dual Attention

ESA Encoder Single Attention

FIT Feature Integration Theory

FS Feature Smoothing

Fusion-SUNet Fusion Semantic Utility Network

GAN Generative Adversarial Network

GM-Simulation Gaussian Model-based Simulation

GPS Global Positioning System

GPU Graphics Processing Unit

GNSS Global Navigation Satellite System

HLC Hierarchical Layout Consistency Loss

Hz Hertz

IMU Inertial Measurement Unit

LiDAR Light Detection and Ranging

MFA Multi-resolution Feature Aggregation

MHz Mega Hertz

MLP Multi-Layer Perceptrons

MPIA Multiple Pulses in Air

NERC North American Electric Reliability Corporation

NSANet Noise-Seeking Attention Deep Neural Network

OD Object Detection

Pan-SUNet Panoptic Semantic Utility Network

P2-Simulation Physical priors-based Simulation

PIA Pulses in Air

PPE Physical Priors Estimation

PRF Pulse Repetition Frequency

RS Regional Segmentation

SFOV Scan Field of View

SLAM Simultaneous Localization and Mapping

SOTA State-Of-The-Art

SPIA Single Pulses in Air

SUNet Semantic Utility Network

UAV Unmanned Aerial Vehicle

VAE Variational Autoencoder

VPP Voxel Post-processing

3D Three-dimensional

2D Two-dimensional

Chapter 1

Introduction

Motivation

LiDAR (Light Detection and Ranging) technology, particularly when deployed via airborne platforms, has become integral to mapping urban and natural landscapes. Advances in electro-optical engineering has facilitated high Pulse Repetition Frequencies (PRF), ranging from 100Hz to 2MHz, resulting in denser 3D point clouds and enabling more detailed analyses [1].

Utility infrastructure, encompassing vast and complex network maintenance and modernization efforts, faces numerous challenges. Conventional methods are manual, labor-intensive and prone to errors, leading to increased costs and reduced accuracy. There is a compelling need for automation, and deep learning presents a transformative opportunity, especially when applied to Airborne LiDAR Scanning (ALS) data, to enhance efficiency and reduce operational costs [2].

Yet, transitioning to fully automated analysis of utility networks—capable of processing dense, large-scale 3D point clouds—poses significant challenges. Deep learning offers a promising solution with advancements in denoising [3, 4], segmentation [5, 6, 7, 8, 9], and object detection [10, 11, 12, 13, 14], which could streamline semantic analysis and enhance predictive maintenance. However, integrating these deep learning solutions into a complete

utilitynetwork management system is an intricate task currently underway.

The application of deep learning to outdoor, large-scale utility environments is further complicated by the lack of specialized datasets and the need for deep learning architectures to manage the complexity of such settings. Research has mainly focused on pixel-based or point-based semantic analysis, which does not fully address the relational dynamics among the components of utility powerline networks [15, 16, 17], critical for thorough network analysis.

Effective management of utility network assets calls for an in-depth understanding of the infrastructure, necessitating detailed insights into the assets’ identity, location, and condition. This level of understanding is essential for monitoring operations and managing risks. Beyond recognizing points, management requires instance-based modeling and regional analysis to differentiate individual components, such as spans between pylons, and to classify zones, including utility, corridor, and non-corridor zones, which are fundamental for strategic decision-making and proactive maintenance.

This research aims to refine the post-data acquisition phase of utility network management by developing state-of-the-art deep learning techniques for noise filtering, semantic segmentation, and object detection. By harnessing ALS point clouds, the research seeks to convert vast data sets into actionable insights, thus enhancing utility infrastructure management.

The thesis introduces multiscale, integrative deep learning models that incorporate attention feature fusion, domain embedding, and panoptic segmentation, all of which are expected to increase the efficiency and accuracy of ALS data analysis. The anticipated impact on infrastructure maintenance and risk assessment is substantial, setting the stage for the evolution of smarter, more reliable utility management systems.

Problem Domain

In the past, numerous 3D scene semantics systems were developed using a combination of statistical, geometrical, and machine learning methods for sparse 3D point cloud outdoor scenes captured with airborne LiDAR [18, 19, 20, 21].

However, traditional methods have proven ineffective and inefficient when dealing with dense 3D point clouds contaminated with clustered, sparse, and systematic noise [22, 23, 24, 3]. These conventional approaches demonstrate limitations in achieving predictive maintenance through segmentation and object detection.

Recognizing the shortcomings of existing frameworks and inspired by an extensive new study on deep learning, there has been a burgeoning interest in exploring its potential in scene semantic and risk analysis. Deep learning with computer vision has shown promise for performing post-data acquisition analysis effectively [7, 6, 5, 9, 8].

By employing convolutional methods, deep learning can extract differentiating features and decode this information for segmentation, classification, and detection, eliminating the need for manual crafting and selection of features for learning [25]. However, existing deep learning methods encounter challenges in incorporating domain knowledge or physical priors, ensuring standard spatial consistency, and extending local and global context through attention to make decisions based on elevated confidence for different classes.

Attention phenomena have been integrated into computer vision to address these challenges, drawing inspiration from psychological and biological concepts. Attention-based convolutional neural networks have not fully explored the potential of various psychological theories and guiding methods for decision-making through external cues embedded in attention. These external cues encompass domain knowledge, hierarchical spatial consistency, and contextual reasoning [26, 27, 28].

An integrative and multiscale deep learning solution can facilitate the cleaning of the point cloud and be used for specific industrial applications, particularly utility powerline network semantics and object-to-object risk analysis, as explored in this research. Maintenance of the utility powerline network is a crucial infrastructure management responsibility due to heavy reliance on electricity for daily activities [29]. Since the 2003 blackout, power outages have encouraged policymakers to focus on maintaining reliable electricity delivery to consumers [30]. This increased demand for infrastructure maintenance and managing vegetation around utility corridors to prevent encroachment-induced power outages [31, 32]. The lack of a

transmission line predictive maintenance system using deep learning is a major limitation, primarily because of utility companies' unavailability of a benchmark dataset and adaption.

The problem domain concerning utility powerline networks primarily focuses on understanding the relationships among objects and regions within and around the network. The study of these relationships is motivated by the right-of-way or wire zone regions and the guidelines provided by transmission companies regarding vegetation height in the area [32]. By delving deeper into the application of attention and refining the utilization of external cues, the research aims to enhance the capabilities of the deep learning-based semantic analysis pipeline. Panoptic segmentation has been utilized in autonomous driving applications but has not been employed for utility datasets. The potential of panoptic segmentation can facilitate object-to-object risk analysis and utility network modeling. These advancements can potentially revolutionize the risk and semantic analysis of 3D scenes, leading to more accurate and efficient predictive maintenance and other applications across diverse fields.

Research Goals and Objectives

This research strives to harness the power of external and domain-oriented cues to create an advanced artificial intelligence pipeline for post-data acquisition analysis, specifically tailored for handling large-scale dense 3D point clouds. The dissertation encompasses five major objectives, all centered around the common theme of integrative and multiscale deep learning, feature and multimodal fusion, attention, and domain embedding, focusing on spatial layout consistency. The thesis aims to significantly enhance the efficiency and accuracy of analyzing complex 3D scenes by integrating attentional phenomena.

Our research will explore how external cues drawn from the vast domain knowledge available can effectively inform the AI pipeline's decision-making process. By incorporating attention mechanisms inspired by how humans prioritize and process information, this research seeks to enable the AI system to focus on relevant details and make more informed judgments, as shown in Figure 1.1.

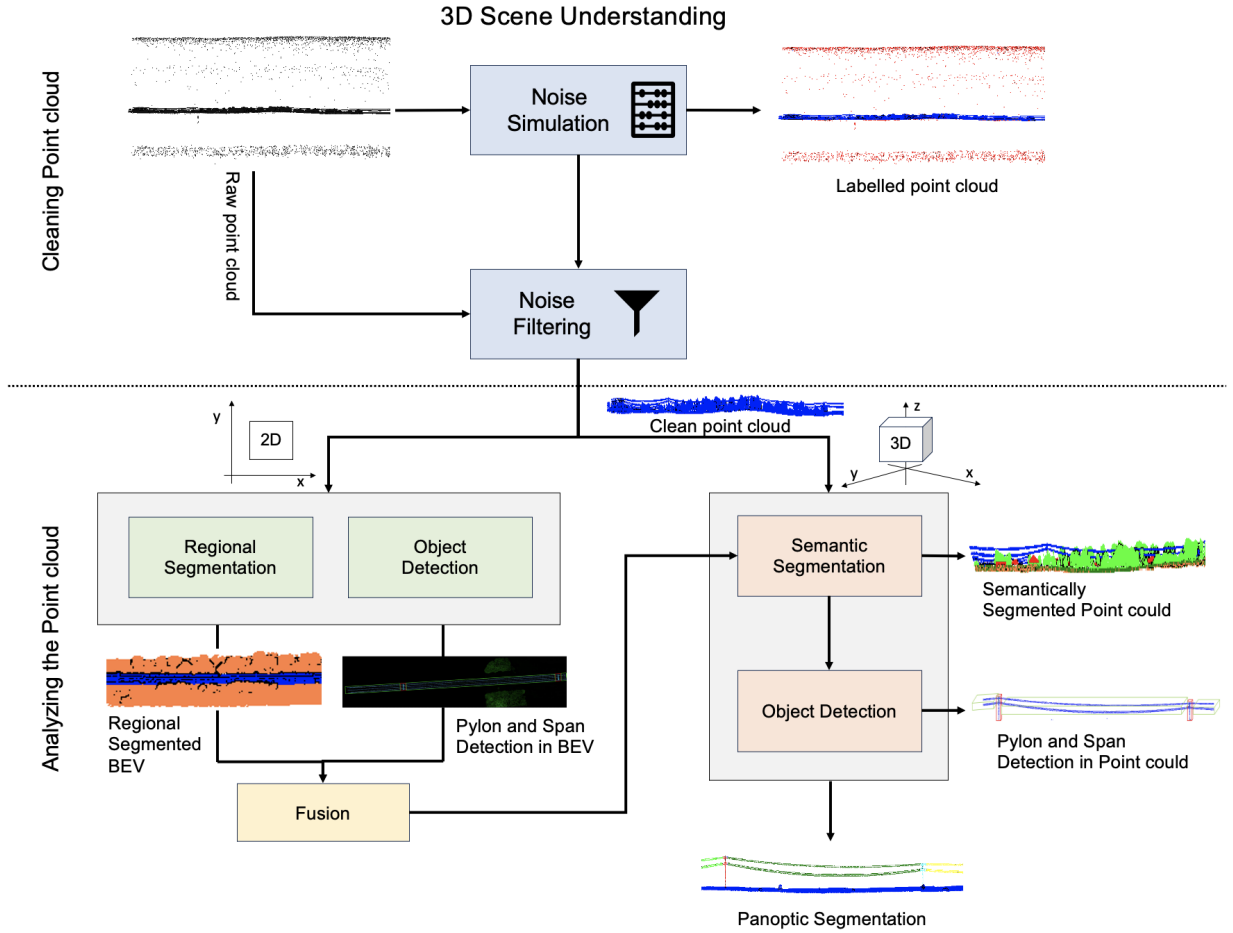


Figure 1.1: Overview of 3D Integrative and Multiscale Deep Learning System.

Additionally, this dissertation recognizes the critical importance of spatial layout consistency in the scene understanding process. This thesis will delve into methods that ensure the AI pipeline maintains coherence in spatial relationships, leading to a more coherent and reliable risk and semantic analysis.

The research has following major objectives.

- **Noise Simulation:** The primary and first objective of this research is to clean the raw point cloud data effectively and to gain a comprehensive understanding of the noise that compromises the perceived quality of the scene and replicate it through simulation.

To achieve this, the thesis explores the innovative concept of modern airborne physical technology known as multiple pulses in the air (MPIA), which offers potential solutions to previous technological limitations of a single pulse in the air (SPIA). Through the simulation process, research can replicate corrupted points' distribution and processing, enabling us to filter out noise and augment the dataset. This augmented data will train a neural network specifically designed for noise filtering.

- **Noise Filtering Deep Neural Network:** The thesis gained valuable insights into different noise categories by conducting noise simulations considering data distribution, density, and localization. The main goal is to effectively filter out noise while preserving the objects of interest, resulting in a clean and accurate point cloud. A deep neural network that explores noise filtering using extended multiscale contextual features and draws from the physical priors studied during noise simulation is employed to achieve this.
- **Utility Semantic Segmentation Network:** One of the primary goals of this research is to enable predictive maintenance for utility transmission and distribution infrastructure. The demand for a robust solution to address issues like vegetation encroachments, unforeseen events, and component anomalies is substantial. To tackle this challenge, I developed an advanced computer vision system that excels in segmenting utility corridors while ensuring consistency in spatial layout. This system fuses the voxels with pixels to impose spatial layout consistency and integrate contextual reasoning. By achieving accurate segmentation, it can effectively monitor and maintain the infrastructure, reducing the risk of outage and improving overall reliability.
- **Multi-task Joint 2D Regional Segmentation and Object Detection:** The extraction of spatial layout consistency is achieved by considering the broader context of the powerline corridor scene. To accomplish this, a joint network is used for regional labeling, categorizing the powerline corridor into three regions: corridor, non-corridor, and utility. One of the main objectives of the research was to create a system that can accurately detect pylons and span objects, effectively converting them into utility

regions, segments, and objects. The developed system serves a dual purpose. Firstly, it provides a foundation for integrating spatial layout into the 3D segmentation and object detection network, enhancing the overall scene understanding capabilities. Secondly, it can function as a standalone system for regional segmentation, allowing it to be independently applied in various applications. The system contributes significantly to understanding and maintaining utility infrastructure through span-to-span risk analysis by accurately detecting and segmenting pylons and spans. Moreover, incorporating spatial layout consistency adds a crucial contextual dimension to the analysis, enabling more informed decision-making and improving the efficiency of the entire process. With its versatility and effectiveness, this system holds great potential in advancing research and practical applications in powerline corridor analysis and management.

- **Panoptic Semantic Utility Network:** An additional crucial objective is to seamlessly fuse 2D segmentation and object detection into the 3D segmentation process using a powerful deep convolutional neural network. This network is specifically designed to detect pylons and spans, facilitating object-to-object risk analysis and complementing the semantic segmentation efforts. By achieving this multiscale object, layout, and feature fusion, research establishes a robust foundation for creating accurate 3D models of the transmission network. This advancement can significantly enhance the overall understanding and management of the infrastructure, enabling more efficient and effective maintenance practices.

Research Contributions

- *Noise Simulation [33]:* This thesis contributes to understanding the physical priors or domain knowledge for building a noise simulation for 3D point clouds. It conducts a quantitative analysis to assess the feasibility of simulation for data augmentation. Additionally, qualitative analysis is performed to understand the distribution of noise points and categorize noise into three types, forming the foundation for designing a noise-filtering neural network.

- *Noise Filtering Neural Network [34]*: This thesis chapter introduces a dual attention-based noise-filtering neural network based on physical priors. It consolidates the idea of combining global physical priors and local spatial context for multiscale feature fusion to filter noise and produce a clean point cloud for semantic analysis. These physical priors provide domain knowledge embedded into the network for a better understanding of the global context.
- *Spatial Layout Consistency for 3D Semantic Segmentation [35]*: This research contribution is the foundation for multiscale layout, voxel-to-region, and feature fusion to enforce domain knowledge of spatial layout consistency in utility corridors. This research proposes a Semantic Utility Network (SUNet), which conducts semantic analysis on 3D objects and 2D regions to analyze the relationship between regions and objects better. The quantitative analysis demonstrates performance improvement in baseline methods, resulting in better segmentation of pylons, powerlines, and vegetation to avoid encroachment.
- *Fusion-SUNet: Spatial Layout Consistency for 3D Semantic Segmentation[36]* : Fusion-SUNet extends SUNet by incorporating feature fusion from multiple scales: 2D regions and 3D voxels. These features are fused using a deep fusion module called attention fusion. This work embeds spatial layout consistency found in the utility domain at the feature level, leading to a performance boost across classes, as reflected in the results.
- *Pan-SUNet: Utility Corridor Understanding using Spatial Layout Consistency*: This contribution is divided into two sections. The first section focuses on enhancing the 2D Bird’s Eye View (BEV) pipeline, which performs regional segmentation and conducts 2D object detection of spans and pylons. This pipeline outputs all three utility transmission zones and pylon and span objects. These 2D objects and regions are fused into a voxel 3D network to embed domain-specific spatial layout consistency. It refines the semantic analysis and produces 3D object detection from 2D detections. A quantitative analysis is conducted to compare the performance improvements on various baselines for

Teledyne and Riegl datasets. Additionally, this thesis performs a detailed quantitative analysis to observe the impact of voltage on the segmentation of powerlines and the height of pylons on the performance.

Thesis Outline

This chapter introduces the research with the motivation to build a 3D scene understanding pipeline for a large-scale 3D point cloud. It discussed a module-by-module pipeline to automate post-data acquisition analysis. The outline of this thesis for the rest of the document is organized as follows:

- Chapter 2 delves into the current framework of airborne laser scanning technologies, focusing on noise issues, data augmentation through simulations, noise filtering techniques, attention frameworks, and incorporating attention fusion theories derived from psychology. It also elaborates on physics-based learning, spatial layout considerations, utility corridor layout, and semantic segmentation methods discussed in the existing literature. Our literature review comprehensively summarizes the previous research contributions to developing a scene understanding pipeline.
- Chapter 3 focuses on comprehending the characteristics of noise, the significance of physical priors, and the role of simulation-based data augmentation for noise filtering. Additionally, this research conducts baseline experiments to evaluate the effectiveness of noise removal techniques. The thesis introduces two simulations based on physical priors and a Gaussian model and analyze their impact on noise generation and subsequent noise filtering processes.
- Chapter 4 presents a unique approach to noise filtering, incorporating physical priors fused through dual attention into a deep neural network. This chapter delves into the experiments design of ablation studies and thoroughly assesses the performance of our proposed system, comparing it with state-of-the-art methods.

- Chapter 5 elaborates the pipeline’s second phase involves the development of a semantic utility network that relies on spatial layout consistency. This stage focuses on studying the spatial patterns observed in utility infrastructure and then applying this knowledge to segment the 3D point cloud data more broadly. It also discusses an extension of the previous approach, which fused the spatial layout consistency as a middle deep fusion module to extract the advantage of spatial regularities better. The third approach, Pan-SUNet, integrates 2D object detection into 2D regional segmentation and uses it to impose spatial layout consistency for 3D segmentation. It segments the third zone by detecting pylons and spans from utility zones and uses it to expand the regional spatial layout consistency further. Later, it uses 2D object detection to project into 3D using segmentation results through a panoptic fusion module and refine segmentation results further. It completes the pipeline with a major contribution of 3D semantic segmentation and byproducts of pylon and span as 3D instances and objects.
- In Chapter 6, datasets, experimental configurations, and evaluation metrics are discussed. This chapter provides all the necessary information to study the ablation and comparative studies designed to evaluate semantic segmentation, object detection, and regional segmentation. It elaborates on the variables that impact the performance of these computer vision tasks within a dataset.
- Chapter 7 presents and analyzes the results collected from the ablation and comparative studies designed in the previous chapter across Teledyne and Riegl datasets. It offers a detailed qualitative and quantitative analysis of potential contributions across datasets.
- Finally, Chapter 8 concludes our research study. It addresses the findings and limitations. It also discusses future work.

Publications

- *Pan-SUNet: Utility Corridor Understanding using Spatial Layout Consistency* – Publication in ISPRS, Geospatial Week (GSW), 2023 - Cairo, Egypt.
- *Fusion-SUNet: Spatial Layout Consistency for 3D Semantic Segmentation*[36] – Publication in PCV, IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR), 2023 - Vancouver, BC, Canada.
- *Human Vision Based 3D Point Cloud Semantic Segmentation of Large-Scale Outdoor Scenes*[37] – Collaborative Publication in PCV, IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR), 2023 - Vancouver, BC, Canada.
- *Spatial Layout Consistency for 3D Semantic Segmentation* [35] – Publication in PRRS, International Conference on Pattern Recognition, 2022 - Montreal, QB, Canada.
- *AI for Noise Classification in Large-scale Point cloud* – Abstract for Canadian Symposium on Remote Sensing, 2021 - Yellowknife, NWT, Canada.
- *Simulation-based Data Augmentation using Physical Priors*[33] – Publication in ISPRS 2020 Congress, Nice, France: Noise filtering neural network and pilot experiments for understanding noise in airborne LIDAR.
- *Noise Seeking Attention* [34] – Publication submitted for IEEE Remote Sensing Transaction Journal Dual attention-based noise filtering neural network based of physical priors .
- *Pan-SUNet: Utility Corridor Understanding using Spatial Layout Consistency* – Publication in progress for ISPRS Journal of Photogrammetry and Remote Sensing.
- *EyeNet: Fast and Efficient Semantic segmentation of large-scale point clouds via human vision inspired approach* – Collaborative Publication in progress for ISPRS Journal of Photogrammetry and Remote Sensing.

Presentations

- *Deep Learning Workshop* – Instructor for Workshop in Canadian Symposium on Remote Sensing, 2023 - Yellowknife, NWT, Canada.
- *3MT: Three Minute Thesis Presentation* – Second Prize in Canadian Symposium on Remote Sensing, Yellowknife, NWT, Canada in 2023.
- *Research Evaluation Presentation* – Second Prize in Summer Research Evaluation Conference, 2023 - Department of Earth and Space Science and Engineering, Lassonde School of Engineering, York University, Toronto, ON, Canada.
- *Deep Learning Workshop: 3D Point cloud* – Instructor for Workshop in Canadian Symposium on Remote Sensing, 2021 Yellowknife, NWT, Canada.
- *GeoAI: Geospatial Artificial Intelligence* – Presentation for Marine Technology Society, 2020 - Fredericton, NB, Canada.
- *A Large-scale Noise Filtering of Bathymetric Point Cloud using Deep Neural Network* – Presentation for Canadian Hydrography Conference, 2020 - Quebec City, QB, Canada.

Chapter 2

Literature Review

Chapter 2 establishes the theoretical foundation for this thesis, offering a thorough review of the pertinent literature in three distinct segments. The chapter discusses the importance of airborne laser scanning and data acquisition technologies in producing 3D point cloud data. It then delves into an extensive examination of deep learning in computer vision, which is the central theme of the thesis. The analysis covers various sub-disciplines of deep knowledge and the visual tasks they encompass to construct a comprehensive integrative deep learning framework.

Next, data enhancement techniques are explored, including noise and diverse data augmentation strategies. The chapter focuses on the current methods for noise filtration, with particular emphasis on incorporating physical priors as directive cues at a global scale. It thoroughly explores the effectiveness of deep neural networks in reducing noise and demonstrates how assimilating physical priors as attention-guiding mechanisms can significantly enhance the efficiency of noise filtering.

The final section of this chapter is devoted to understanding utility scenes and discussing the application of deep learning for spatial decision-making in the context of utility corridor layouts. Topics such as utility powerline infrastructure, historical outage data, and vegetation management are discussed in detail. This knowledge underpins the aim to develop predictive maintenance systems for utility infrastructures, essential for maintaining their reliability and

durability.

2.1 Airborne Laser Scanning: Pulse in Air (PIA)

Airborne LiDARs are active sensors emitting pulses and generating backscattered signal reflections. Over the past few decades, remote sensing technology has progressed, and airborne LiDAR surveys have become more common. They are used to plan for natural and human disasters, which can be prevented with the aid of this technology. In contrast to other data acquisition methods and tools, airborne LiDARs are more efficient, faster, and accurate due to the technology’s maturity and the following vital characteristics: (1) their ability to capture an exact and highly dense point cloud, (2) LiDARs can record discrete or full waveform of the backscattered signal, map ground surfaces, and detect low vegetation even in forests and heavily populated urban areas, (3) the sensors have no blind zones, (4) surveys are faster, making them an economical data acquisition method, and (5) extreme weather or environmental conditions have a relatively lower effect on the resultant mapping [38].

Point clouds’ digitization using full or discrete waveform has advantages and disadvantages. Full waveform digitization captures vertical structures in a time series dataset, while discrete waveform only captures peaks. Since the thesis only utilizes discrete waveform-based LiDAR, this section will focus solely on airborne LiDARs that are based on discrete waveform technology [38].

The airborne LiDAR system is a complex multi-sensor device with various essential components, such as a reflectorless laser sensor, an AIMU (airborne inertial measuring unit), a GPS (global positioning system), data storage, and a power supply. As the system carrier, such as an aircraft or plane, flies over the area of interest, the laser sensor emits pulses toward the surface. These pulses travel at the speed of light and reflect to the sensor, depending on the surface’s reflectivity. This enables the system to create a highly accurate 3D map of the area of interest, which can be used for various applications in fields such as surveying, mapping, and environmental management.

After receiving the reflected energy, the sensor calculates a 3D point based on the distance traveled by the pulse, the sensor's position, and the beam's direction. To accurately track the LiDAR's position, the system uses trajectory information from a GNSS (global navigation satellite system) receiver, which provides altitude. Additionally, the AIMU records the LiDAR's position using pitch, roll, and yaw angles. The LiDAR's signal is deflected using an internal mirror, and the mirror's position is recorded for every laser pulse shot, as shown in Figure 2.1. Finally, the range between the target and the sensor is calculated using the Euclidean distance (Equation-2.1). This equation is used for calculating range when we have sensor location and 3D points of target for understanding the noise points.

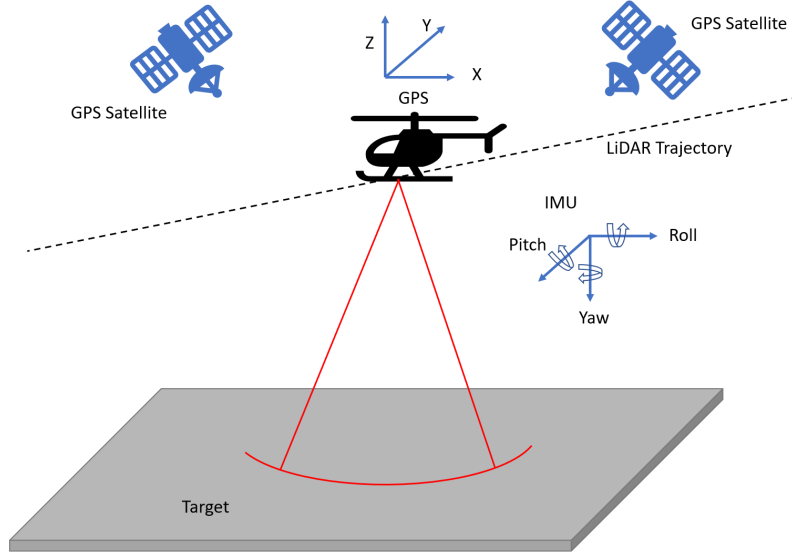


Figure 2.1: Airborne LiDAR.

$$R = \sqrt{(x_p - x_l)^2 + (y_p - y_l)^2 + (z_p - z_l)^2} \quad (2.1)$$

where x_p, y_p, z_p = coordinates of the points
 x_l, y_l, z_l = coordinates of the sensor

Advancements in LiDAR technology have enabled a high Pulse Repetition Frequency

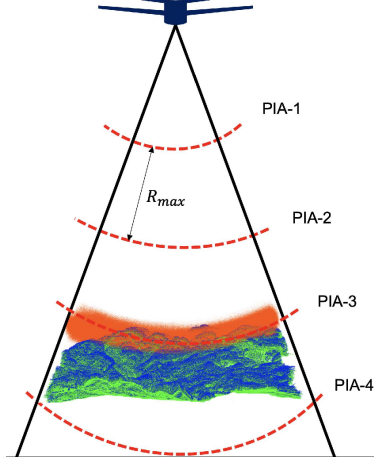


Figure 2.2: Pulse in air zones with point cloud and noise on transition point. R_{max} shows the range a pulse travels before next one is fired.

(PRF) of over half a million Hz. PRF represents the ability to fire the pulses within one second, calculating how many pulses can travel with the speed of light (c) to the maximum unambiguous range R_{max} in one second, as shown in Equation 2.2.

$$PRF = \frac{c}{2 \times R_{max}} \quad (2.2)$$

It means if the PRF is 1000kHz than R_{max} will be 149.8 meters as shown in Equation 2.3.

$$R_{max} = \frac{c}{2 \times PRF} \quad (2.3)$$

This enhances survey efficiency, allowing data acquisition with fewer flight lines and a wider Scan Field of View (SFOV). The SFOV is the angle at which LiDAR emits the pulses, set based on the target application and sensor range. Different acquisition techniques have been used for effective, efficient, and fast surveys. In airborne lidar systems, the term "Pulse in Air (PIA) zone" refers to a maximum unambiguous range where the laser pulse propagates through the air before reaching the ground or target surface, and the next beam is yet to be fired, as shown in Figure 2.2.

The PIA serves as the basis for acquisition technology, allowing the measurement of how many pulses can be in the air at a given time or before the next one is fired. The number of total PIA zones depends on Equation 2.4.

$$PIA = \frac{PRF/(c/2)}{Altitude/\cos(SFOV)} \quad (2.4)$$

The conventional LIDAR technology performs the survey using a Single Pulse In Air (SPIA) zone with a lower PRF.

Single Pulse In Air (SPIA) is a traditional technology that was discarded due to its inability to match the high PRF. Multiple Pulses In Air (MPIA) technology has been developed, surpassing the limitations of traditional SPIA technology [39]. This technology ensures the firing of multiple pulses before receiving the last one, making it efficient and fast, as depicted in Figure 2.2. This thesis largely focuses on the 3D point cloud acquired using MPIA technology for noise filtering and utility semantic analysis.

2.1.1 SPIA

SPIA is a conventional method for data acquisition in airborne LiDAR, which is cost-effective at low PRF. SPIA involves discrete data acquisition, firing the next pulse only when the previous one is received. However, at higher PRF, SPIA technology struggles to keep up due to its limitation of allowing only one pulse in the air at a time. This drawback can lead to missing or corrupted data points, especially at higher altitudes or higher PRF speeds where the speed of light is not fast enough to return the pulse before the next one is fired [40]. Figure 2.3(a) illustrates that at a certain time, only one pulse is in the air, such as S2 is not fired before R1 is received. In this system, a single pulse is fired, and the transmitter waits for a time interval of $1/PRF$ before shooting the next pulse and receiving the previous signal before it. These systems are slow, costly, and have only one operational range per flight line. Moreover, they generate less detailed data and have a relatively lower point density [41].

2.1.2 MPIA

In contrast to SPIA technology, MPIA technology operates more efficiently by not waiting for a pulse to be fully reflected and matched before firing the next one, as illustrated in

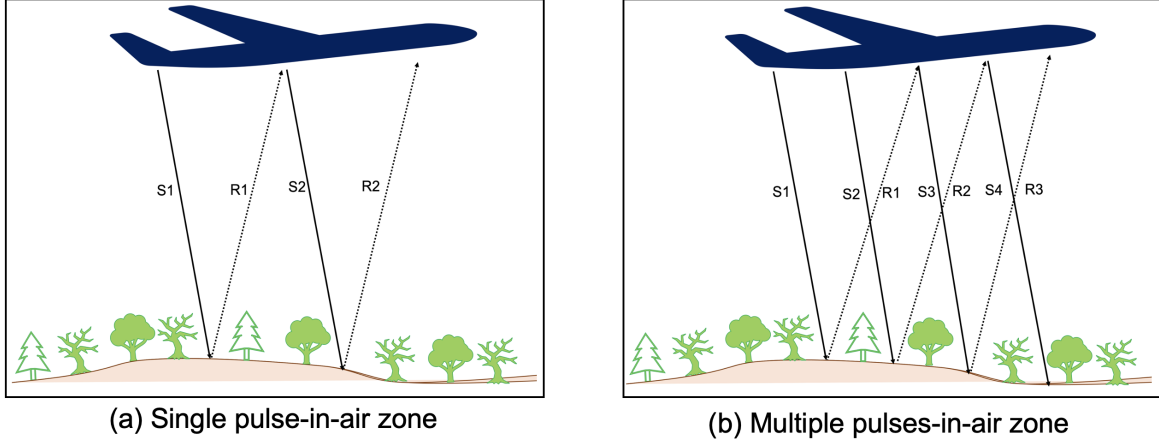


Figure 2.3: (a) SPIA zone shows one pulse sent and received before next is fired and (b) MPIA fires second pulse before first is returned.

Figure 2.3. However, this efficiency has a drawback—the laser pulse pairs across PIA zones can result in atmospheric points being projected closer to or overlapping with objects of interest. Consequently, the point cloud obtained through MPIA technology may suffer from compromised scene perception due to noise.

MPIA technology is highly configurable, allowing intensity and PRF adjustments to acquire more detailed, high-density point clouds. This results in fast, cost-effective, gate-less sensors that do not generate blind zones. These sensors provide multiple operational ranges for a single flight. As illustrated in Figure 2.3(b), at a certain time, there could be multiple pulses in the air, such as S1 and S2, before R1 is reflected. The use of MPIA technology in LIDAR systems can be affected by atmospheric points caused by environmental factors such as dust, fog, snow, or rain particles and incorrect projections of objects of interest, causing systematic noise during range ambiguity resolution. When the laser beam is deflected or scattered by these atmospheric particles, it can result in misinterpretation of signals on the receiver as objects of interest, generating noise [39].

Increasing the PRF has a detrimental effect as it narrows the PIA zones and increases the difficulty of differentiating noise points from objects of interest. Consequently, the projection of these noise points can give the impression that they are closer to the objects of interest

and occur during the transition between PIA zones, as depicted in Figure 2.2. This proximity makes distinguishing the noise points from the objects of interest difficult. Resolving range ambiguity is crucial in the post-processing analysis of acquired data. Range ambiguity arises when the returned pulses overlap on the receiver.

The noise points are projected onto the ground at the closest possible range to address range ambiguity. This is necessary because noise points detected at longer ranges can falsely appear closer to the ground.

Consequently, when the noise points are projected, they can give the illusion of being closer to the transition zones between different PIA areas. This phenomenon occurs because the range ambiguity resolution process places the noise points at ranges closer to the ground, creating the perception that they are adjacent to the transition zones. Figure 2.2 depicts the compromised quality of perception due to noise [4].

2.2 Deep Learning in Computer Vision

Artificial intelligence, particularly deep learning, has become pervasive daily, influencing diverse fields such as autonomous vehicles, healthcare, remote sensing, and anomaly detection. While machine learning relies on manual feature crafting, deep learning, the primary focus of this thesis, requires minimal human intervention and leverages neural networks inspired by the human brain. Deep learning excels in computer vision tasks, such as scene analysis, risk assessment, vegetation management, and powerline anomaly detection, benefiting from its ability to interpret visual signals without intricate feature engineering [42, 43, 6, 8, 15, 44, 45, 46].

Despite its advantages, deep learning introduces abstraction and potential biases from training data, especially in high-dimensional diverse datasets. However, its efficiency in handling such datasets, coupled with minimal human intervention, makes it a powerful tool for computer vision tasks. This thesis aims to develop an integrative deep learning system, emphasizing the theoretical foundations of scene understanding.

2.2.1 Physics-based Learning: Domain-specific External Cues

Deep learning has significantly improved classification, object detection, recognition, and segmentation. However, integrating physical priors and domain knowledge into deep learning models, known as physics-based learning, has shown great potential for solving some problems [47]. These domain-specific external cues play a critical role in human intelligence, enabling individuals to make decisions even in entirely unfamiliar environments within the same domain. Human intelligence leverages a vast contextual global knowledge base, effectively allowing us to adapt and make decisions in new environments. It’s crucial to recognize that much of this valuable information is readily accessible to us, which can significantly enhance the precision of segmentation results compared to relying solely on available datasets.

P^3VAE is a generative model integrating deep learning with physics modeling to learn a high-dimensional latent space. It has shown improvements for semantic segmentation by regularizing the feature space [48]. A self-supervised dehazing algorithm also used physical priors to generate a lightweight network that outperformed existing systems [3].

In various fields, domain-specific external cues have proven invaluable. They have been employed in medical imaging to normalize data, aiding clinicians in understanding tumors [49]. Similarly, in physics-based studies, these cues validate predictions, ensuring they remain within valid ranges of values [3, 47]. Despite various physics-based learning models that target complex scientific problems, there is a research gap in utilizing physical priors for semantic segmentation and denoising of 3D point clouds [50]. These findings strongly advocate integrating utility corridor layout consistency as domain-specific external cues in deep learning-based scene understanding systems.

2.2.2 Multiscale Deep Learning

Multiscale deep learning is a sub-branch of deep learning that focuses on fusing features from different scales and receptive fields to better incorporate context for decision-making. Various techniques have been introduced to extract features with a broader contextual field [51].

This section explores the importance of multiscale convolutional neural networks and fusion in addressing the need for multilevel context. Previous studies in semantic segmentation and object detection for 2D images [52, 53, 54, 55], RGBD images [56, 57], and 3D point clouds [8, 7, 6, 15, 5] have demonstrated the significance of aggregating features from different resolutions to encode discriminating features effectively. These features shape the output, enabling the system to handle objects of various sizes within the larger scene context [58].

Artificial Intelligence (AI) aims to mimic human cognition, and humans often base decisions on prior knowledge and spatial context to understand scenes [28]. This thesis focuses on embedding domain-specific external cues, physical priors from the global scene, and spatial layout consistency. Therefore, multiscale deep learning shows immense potential to enhance semantic analysis.

Deep learning in computer vision encompasses various tasks, including classification, object detection, semantic segmentation, and panoptic segmentation. This research focuses on semantic segmentation, 2D object detection, and panoptic segmentation, and the following section will discuss them in detail.

2.2.3 3D Semantic Segmentation

Utility scene understanding involves semantic segmentation, a crucial task for classifying and segmenting points or pixels within 3D point clouds in utility applications. Deep learning has significantly influenced 3D point cloud semantic segmentation, enabling automatic feature learning for generalizable solutions[59]. While voxel-based approaches like 3D U-Net [9] and 3D Attention U-Net [56] have shown comparable performance, limitations such as quantization error and resource-intensive voxel grids make them less ideal [60, 61, 62]. SPConv addressed these issues, optimizing voxel-based approaches [63].

Point-based networks like PointNet [7] and PointNet++ [6], along with kernel-based approaches like KPConv [5], PointConv [64], PointCNN [65], and SpiderConv [66], provide alternatives with improved handling of unordered point clouds. RandLA stands out as a state-of-the-art network, excelling in robustly handling point clouds through random sampling

[8]. Powerline-Net targets high voltage transmission line semantic segmentation, addressing local features and downsampling [15]. However, it focuses on a specific powerline type (500kV) and lacks consideration of spatial layout. Notably, none of these methods fully leverage the spatial layout inherent in utility infrastructure.

2.2.4 2D Object Detection

Object detection is classifying and locating objects in a given scene. Traditional techniques have been replaced by deep learning-based neural networks, with two main types of architectures: single-stage [14] and two-stage detectors. While single-stage detectors are computationally efficient, they can be less accurate than two-stage detectors due to the coarse-to-fine process [14, 67].

Most object detection networks can only handle horizontal bounding boxes, which is problematic for real-world scenes with non-horizontal objects. To address this, rotated object detection methods [13], such as Oriented R-CNN [11], RoI Transformer [12], and Rotated Faster R-CNN [10], have been developed to use oriented bounding boxes, and have shown comparable performance to horizontal bounding box detectors while reducing background errors.

Multi-task learning and panoptic segmentation are examples of fusing and integrating different computer vision tasks into one. There are shared encoders to learn various tasks for multi-task learning [68]. However, panoptic fusion is relatively different with segmentation and instance segmentation branches. Few papers have focused on object detection and employed connectivity analysis to get instance segmentation [69]. However, there is still a lack of fusing the 2D objects with 3D segmentation results to get 3D object detection. This thesis explored existing panoptic segmentation literature and identified the limitations that can be modified to handle multiscale object fusion.

2.2.5 Panoptic Segmentation

This research work is also based on panoptic segmentation, which separates scenes into "stuff" (e.g., ground, sky, vegetation) and "things" (e.g., objects like cars, pylons, and spans) using a combination of semantic and object detection. EfficientLPS [70], is an extension of a 2D panoptic segmentation network that fuses instance and semantic segmentation across the entire scene. PUPS also performs panoptic segmentation analysis but doesn't use object detection, for instance, segmentation [71]. However, it reflects encouraging performance against panoptic evaluation metrics. CPSeg is cluster-free panoptic segmentation, which uses a 2D range view. This works well on autonomous driving and mobile mapping point cloud [72]. However, there is a lack of fusion from different dimensions in this method, specifically for airborne LiDAR point clouds.

Extensive research has explored combining different semantic segmentation networks with object detection and cluster analysis to leverage the panoptic segmentation pipeline [69]. This research encourages further exploration, specifically expanding semantic segmentation and 2D object detection within the panoptic segmentation framework. The article evaluated performance using mobile mapping point cloud data. Although this research primarily concentrates on 3D object detection and semantic segmentation, leveraging 3D object detection results to convert them into instances.

This thesis relates to a 2D panoptic segmentation approach that focuses on detecting pylons and spans and segmenting regions tailored to utility cases. Additionally, it projects the predictions into a 3D pipeline to obtain pylon and span segmentation objects to model utility corridors.

2.3 Data Augmentation

Deep learning sophisticated algorithms require a larger labeled dataset for effective training. This research explored various augmentation techniques to address the need for a labeled dataset. Data augmentation generates additional labeled data by applying transformations

to existing labeled data. This section will discuss the recent work related to 3D data augmentation. Data augmentation increases the volume and diversity of the dataset. This section is divided into two significant subsections, simulation-based and deep learning-based data augmentation [73].

2.3.1 Simulations

Different simulations are developed to address the data scarcity for large-scale point cloud using virtual reality and gaming. One such simulation generates synthetic scenes from the game-based environment [74]. The user can customize these scenes, which can help boost the performance of neural network training on a synthetic dataset. The results showed performance gain on a semantic segmentation task. The research in this article [75] proposed a hybrid technique that used real data to make simulated data more realistic using CycleGAN. It also shows that simulated data is not usually very realistic, especially in off-the-shelf open-source simulations for autonomous driving scenes. Commercial tool Blender’s 3D sensor simulation plugin generates a large-scale 3D point cloud called SynthCity [76].

Multiple other data augmentation techniques are based on class weights for imbalanced class distribution, or random duplication of points proposed by [77] and [7]. Research work for autonomous driving vehicles that deal with noise due to adverse weather conditions proposed an augmentation model for fog and rain [78]. It utilizes a distance matrix, intensity matrix, extinction coefficient, and point scattering rate to augment rain and fog. The augmented dataset trains the proposed neural network architecture and increases overall performance. However, it also caused ambiguity between rain and fog classes due to their inherently similar nature. The noise augmentation simulation research can leverage this literature to develop a simulation to generate atmospheric points based on the principles of MPIA and increase PRF to reflect the uniqueness of modern sensor noise.

2.3.2 Deep Learning

Two approaches of deep learning are often used for data augmentation: generative adversarial network (GAN) and adversarial training networks. PC-GAN is a novel approach to generating synthetic 3D object point cloud [79]. PC-GAN proposed a generic framework to use the underlying data distribution to create a point cloud for 3D models. Due to its usage of local latent variables to understand the neighborhood and spatial location of the point and global latent variables to interpret the object’s overall shape, it’s unsuitable for large-scale point cloud generation.

The research in [80] has proposed an autoencoder that produces a latent space that effectively increases the performance accuracy of GAN networks for 3D point cloud objects. PointFlow learns the distribution of shapes and points and uses the invertible parametrized transformation to learn from these distributions and generate a model for synthetic data [81]. This approach increases the accuracy of generating a 3D object point cloud. Adversarial examples are objects that look like real objects with few perturbations. These examples are generated using deep learning or simple geometrical manipulation to create synthetic datasets. GAN networks proposed to generate the synthetic 3D object point cloud are relevant for object detection and recognition [82]. Existing frameworks lack the generation of synthetic data for large-scale 3D point cloud. They also deal with convergence problems and struggle to produce high-resolution output.

2.4 Noise Filtering Techniques

The presence of noise points can corrupt the signal and distort results. The techniques employed for noise filtering can be broadly categorized into four main groups: statistical and geometrical methods, clustering-based approaches, machine learning techniques, and deep learning algorithms.

2.4.1 Statistical and Geometrical Noise Filtering

The statistical filtering techniques used for removing noise from 3D point clouds and images include univariate location and scale estimation, low-pass filters, mean filters, median filters, principal component analysis, multivariate location, and covariance estimation, and spatial filters [83, 84, 85, 86, 87, 88, 89]. While these filters have been useful for cleaning data collected using the SPIA technology, they are ineffective and expensive for large-scale 3D point clouds with high noise density and complexity.

Geometrical techniques have traditionally been used for outlier detection in 3D point clouds, such as density [24], distance [90], graph [91, 92, 93, 94], and probability-based methods. These techniques were mostly employed for denoising 3D meshes, models, and sparse point clouds. However, they have also been transformed into voxelization pipelines, and density or distance-based filters have been applied to remove noise [23]. Bipartite graphs for approximating errors (noise) and manifold-to-manifold distance-based methods for outlier detection have shown good performance [95].

Nonetheless, geometrical methods are ineffective and expensive for large-scale 3D point clouds with clustered and complex noise. For instance, the research work in [96] deals with a large-scale point cloud that contains single-echo reflection values. It has dealt with multi-sensor-based point cloud to remove reflection-based noise. However, it doesn't work for MPIA-based LiDAR due to multiple returns.

2.4.2 Cluster-based Noise Filtering

It includes methods like the nearest neighbor, radius outlier removal, mean-shift clustering, K-Means, and DBScan [97, 98, 83, 99]. However, these methods require significant memory to construct clusters that adequately preserve object points while removing outliers and corrupted points. These work well for datasets with low overlap with objects of interest but can be resource-intensive when covering each point and performing recursive clustering. Furthermore, these techniques are commonly used for 3D point cloud models rather than

large-scale scenes. While some modifications have been made to run these techniques on GPUs for a performance boost, they may still not be suitable for handling large-scale scenes effectively.

2.4.3 Machine Learning based Noise Filtering

The emergence of machine learning has brought a significant impact on data analysis techniques, including the cleaning of 3D point clouds. Linear regression, support vector machine, random forest, and other methods have been employed by various communities to filter or eliminate noise [22, 100]. These techniques have proven successful in preserving the object points while removing obvious noise. However, the trade-off between maintaining scene quality and removing noise depends on the application domain’s analysis. One of the challenges of using these techniques is the need for feature crafting and selection, which requires a thorough understanding of the dataset and domain. In recent years, researchers in noise filtering have explored the use of convolutional neural networks (CNNs) after introducing deep learning for scene understanding.

2.4.4 Deep Learning based Noise Filtering

Deep learning in computer vision has significantly improved segmentation, classification, reconstruction, and detection. Convolution neural networks extract features from input data and detect decision-making features. They have shown performance boosts in denoising, surface reconstruction, densification, noise classification, and filtering.

PointCleanNet [4], PointFilter [101], Neural Project Denoising (NPD) [102], MaskNet [103], and CNN-based normal estimation [104] for smoothing have improved filtering noise using various techniques through deep neural networks. Recently, transformer-based encoder-decoder has been used to filter noise within point clouds containing single-echo reflection values [105]. It converts the point cloud into range images and removes noise. The work is efficient for single-echo reflective sensors but couldn’t generate range images for airborne lidar using MPIA. Previous research is promising and shows potential for noise-filtering 3D point

clouds.

The limitation of the previous research work is the lack of a dedicated network for a large-scale 3D point cloud. These methods have not shown comparable performance on complex noise. These methods have also not utilized the physical priors of the dataset and fused them into the network as attention for filtering noise points.

The research on noise filtering neural networks is critically inspired by the role of physical priors from airborne LiDAR data collection [33]. These physical priors provide information about the flight trajectory and offer a global context [106]. When fused with local spatial context, these physical priors provide a global scale that, when integrated, can help filter noise. The next section will discuss attention mechanisms for multiscale feature fusion.

2.5 Multiscale Feature Fusion

Feature fusion typically serves as a central module in feature aggregation. These modules necessitate embedding and fusing features from different scales to integrate context seamlessly into the deep learning network. Rather than simply concatenating features from different scales, these fusion modules capitalize on attention [107]. This approach enables the selecting of significant features for enhanced highlighting and performance improvement in the given task. To gain a deeper insight into feature fusion using attention, the following section outlines the attention mechanisms and psychological theories behind fusing features more effectively incorporating context.

2.5.1 Attention

This research explored using attention to embed physical priors into a neural network. Attention is a cognitive phenomenon that allows humans to focus on certain objects while ignoring others selectively. In deep learning, attention is a two-stage process where the network first extracts features from the entire input and then emphasizes important details to improve the accuracy of tasks such as classification, object detection, segmentation, and

tracking [108]. Attention has been widely studied and adapted from human vision into computer vision. In fact, "Attention Is All You Need" [26] was the first publication to integrate attention mechanisms into natural language processing and sparked the use of attention in other deep learning tasks.

Over time, researchers have developed and studied various types of attention mechanisms, such as single-head, multi-head, additive, multiplicative, spatial, channel, and fusion attention [108]—the questions of where and what to attend to have been major research areas in this field. For example, spatial self-attention has improved performance in segmentation and object detection [56]. Additionally, additive self-attention has been found to work well for ambiguous boundaries between objects of interest. Psychological theories of attention guide this thesis. It utilizes a three-stage paradigm to embed physical priors and domain-based spatial layouts into a neural network for filtering noise and understanding large-scale 3D point clouds.

- A context-independent extraction of low-level features.
- Modeling relationship and analysis of low-level features are based on the computer vision task.
- Focus on relevant regions based on earlier stages of extracting high-level features for the final task [27].

Research focuses on understanding this fusion of attention theories to propose the most appropriate embedding technique of physical priors into the deep learning model.

Attentional Engagement Theory

Attentional Engagement Theory (AET), as introduced by Duncan and Humphrey in 1989[109], outlined the attention process into two distinct stages: the pre-attentive stage, where features are extracted without focal attention, and the selective attention stage, where these features are intensely and focused upon to facilitate decision-making.

In computer vision and deep learning, this theory can be translated into a two-part model where, initially, an attention-free encoder processes features from the input data. This corresponds to the pre-attentive stage of AET, where features are detected without attentional focus. Such an encoder would scan the entire dataset, identifying potential interest features without determining their significance or role.

Subsequently, the selective attention stage of AET is mirrored in the form of an attention-based decoder. This model component is designed to selectively concentrate on the features encoded by the first stage, applying intense focus to discern patterns, make connections, and ultimately assist decision-making. This could manifest as a network that, having encoded various features, now applies a 'filter' to decide which features are most pertinent to the task, such as segmenting specific structures from a medical image or identifying objects within a complex scene.

The principles of AET are exemplified in deep learning architectures like the Attention 3D U-Net[56]. These networks leverage attention mechanisms to enhance the feature selection process during the decoding phase, ensuring that the network pays 'attention' to the most informative parts of the data for accurate segmentation or detection.

Incorporating AET into deep learning models provides a framework that can handle complex, noisy datasets by mimicking the human attention process—initially surveying and extracting a broad array of information, then honing in on the most critical features for the task. This allows for more precise and relevant feature utilization in object detection and segmentation tasks.

Feature Integration Theory

According to Feature Integration Theory (FIT), proposed by Treisman in 1980 [110], the visual system processes individual features like color, shape, and size in a parallel and pre-attentive manner. However, integrating these features into a coherent object perception requires focused attention.

In the scope of this thesis, the theory is operationalized through the design of an attention-

based encoder and decoder framework. The encoder emulates the pre-attentive stage, capturing and encoding disparate features independently. These features are processed in parallel, akin to the initial stage of visual processing, where the brain registers various attributes of objects.

Conversely, the decoder embodies the attentive phase of FIT. It employs an attention mechanism to focus selectively on certain encoded features, integrating them to construct a complete and recognized object or scene. This selective attention in neural networks is modeled after the cognitive process where attention binds disparate features into a singular, identifiable form.

To apply this to a deep learning model, one could employ an attention mechanism that echoes the selective focus of attention. For example, convolutional layers in the attention-based encoder could detect a wide array of features within input data. At the same time, the decoder, through an attention mechanism, would prioritize and fuse these features based on their relevance to the task—such as enhancing the clarity of 3D point clouds by filtering out noise.

By integrating this selective focus into the neural network, it could be trained to identify and highlight features that are critical for specific outcomes, thereby enhancing the model’s ability to discern and prioritize information in complex data sets, much like the human visual system’s proficiency in focusing on salient features in a visual scene.

Channel Attention and Spatial Attention

Multiple types of attention are used to fuse features or generally apply to the available set of deep features. Various research studies have employed two of the most significant attention types. Spatial attention, utilized by additive and multiplicative methods as shown in [56] and [26], and channel attention highlights different advantages and cater to diverse needs when fusing features. Channel-based attention ensures that significant features are emphasized for fusion, while spatial attention focuses on highlighting important areas for fusion. Combined, they ensure that spatial locations with notable features carry higher weightage in fusion. has

used this combination citecbam, LANet,bagnet, proving useful. However, fusing features from 2D BEV images and 3D voxels to impose a global spatial layout context has not been explored extensively.

2.6 Multimodal Fusion

Another crucial fusion method extensively employed by various research groups involves leveraging multi-modality. Multimodal fusion combines the input from two modalities to take advantage of spatial and contextual views for better decision-making. Studies have focused on integrating multiple views [111, 60], such as 2D camera images and 3D point clouds [112], as well as different representations of 3D point clouds [113], such as 2D BEV, range images [114], or spherical images [115]. These representations are fused at the feature level from different branches and integrated to enhance the learning of features for various computer vision tasks related to scene understanding. Research in this field has effectively utilized different views and global contexts. However, a limitation of this work lies in incorporating a spatial global layout to provide contextual reasoning for a deeper understanding of questions such as 'Where is the object?' 'what is it next to?' and 'What is it?' [28]. These challenges are addressed by comprehending the standard spatial layout of everyday scenes and integrating it with the finer object-level details of the scene in this dissertation.

2.7 Utility Scene Understanding

Utility scene understanding is the second stage of the integrative system proposed in this thesis. It is important to understand that the 3D point cloud needs to be semantically segmented, and the object-to-object relationship needs to be extracted to interpret the visual scene. Utility scene understanding is the key application of this thesis. The next section will discuss deep learning for spatial decision-making and utility scenes.

2.7.1 Deep Learning Spatial Decision Making for Utility Network

Deep learning has emerged as a transformative technology that can interpret complex 3D scenes accurately. It is particularly useful for tasks that require understanding spatial dynamics within an environment. The key to advancing deep learning in this domain lies in its ability to analyze the spatial relationships between objects and the regions they occupy. This understanding is critical for contextualizing scene elements accurately, which is essential for scene perception.

This study explores the nuances of spatial relationships and the context-aware interpretation of visual scenes. It draws inspiration from spatial cognition, contextual reasoning, and spatial layout principles to lay the theoretical groundwork for applying deep learning in spatial decision-making tasks. Traditionally studied in psychology and cognitive science, these cognitive concepts are now being transposed into artificial intelligence. This will enable deep learning models to navigate and interpret spatial information more effectively, akin to human-like understanding.

Spatial Cognition and Utility Corridors

Spatial cognition, a subset of cognitive psychology, focuses on comprehending one’s environment, encompassing location, orientation, and actions. This field aims to grasp these variables concerning oneself and others in the surrounding environment. It represents the pinnacle of human cognitive ability, enabling individuals to navigate their surroundings and reach their destinations [28]. Spatial cognition is studied for applications such as Simultaneous Localization and Mapping (SLAM) and autonomous vehicles [116]. However, the integration of this specific principle for solving semantic segmentation problems is currently limited and can be used to understand the region-to-object relationship between utility transmission zones and utility objects.

Cognitive processes are intricately linked to context. Humans’ ability to reach a destination or understand their surroundings is connected to a global knowledge base. Much of humans’

understanding of the location of objects is rooted in their contextual placement [117]. Visual experiences are influenced by the location of objects organized within regions, shaping seamless daily cognitive activities [28]. This contextual understanding is pivotal when integrating spatial cognition into an AI-based scene understanding pipeline. The embedding of deep learning-based contextual reasoning can be achieved through multimodal fusion. In this thesis, a 2D view of utility transmission zones or regions can provide a holistic context to locate the utility objects better.

Electric hydro companies around the world follow guidelines for setting utility transmission zones, which are designed to address safety concerns related to infrastructure, residential areas, and vegetation [118]. These zones consist of the type-I: utility zone, type-II: the corridor zone, and type III: the non-corridor zone, each with specific regulations on the size and type of vegetation allowed based on the voltage of the transmission lines[119].

2.7.2 Utility Powerline Network

The powerline network begins by generating electricity in power production plants and transmitting it through a system of high-voltage and low-current powerlines to the distribution system as shown in Figure 2.4. The distribution system converts the electricity for appropriate usage and delivers it to consumers. This system, responsible for safely transmitting and reliably providing electric power to maintain uninterrupted electricity and avoid power outages, comprises many components [120]. The following sections will discuss these components individually CPUCTransmission.

Transmission Substations

This component is located next to the utility transmission lines. Usually, the electric power generated from power plants enters the transmission substation through phase-3 power lines, where a transformer or switch is used to generate high-voltage electricity. High-voltage electricity is used for transmission due to its inherent nature of not losing energy due to resistance. Most transmission lines operate from 69 kV to 700 kV, depending on the country

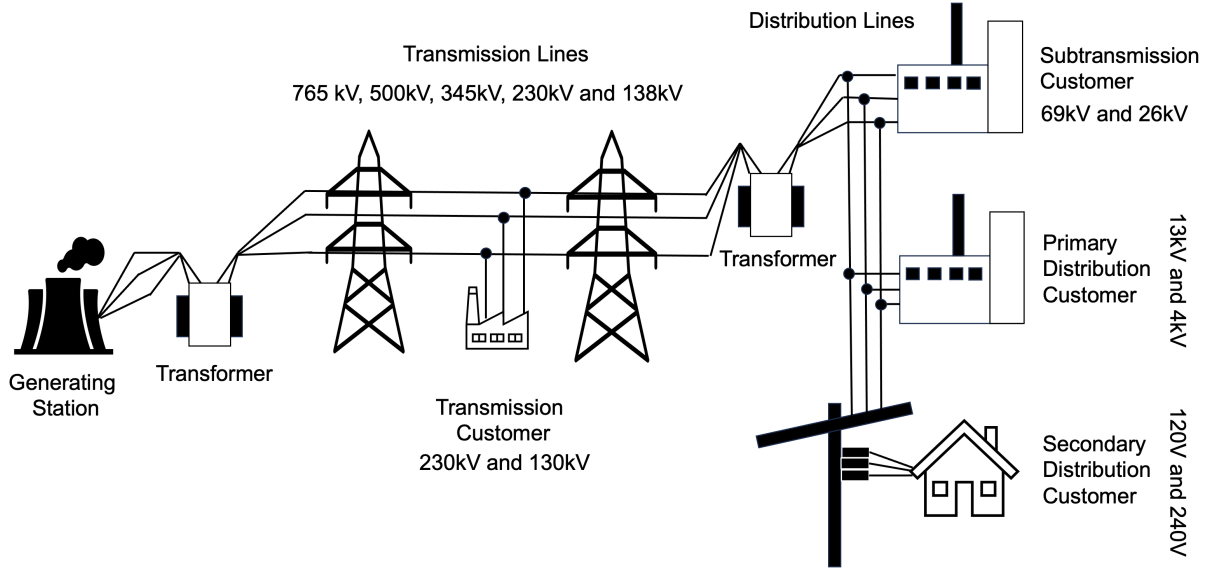


Figure 2.4: Utility Powerline Network.

and infrastructure support [121].

Utility Transmission Lines and Utility Zone

As discussed in the previous section, utility transmission lines transmit electricity with high voltage over long distances. The higher the voltage, the longer the distance it can travel without losing energy. Utility companies typically install these transmission lines in clear environments to maintain and adhere to the right-of-way or wire corridor standards. The width and size of the utility zone, wire zone, or right-of-way depend on the voltage of the transmission line, the height of pylons, and regulations in place by NERC and utility companies.

Utility transmission lines usually have three conductor wires grouped, often combined with a ground or earth wire to shield against lightning strikes. These Phase-1, Phase-2, and Phase-3 wires ensure safe high-voltage transmission. Typically, utility transmission zone widths are as follows: 500 kV - 69 meters, 287 kV - 45 meters, 230 kV - 30 meters, 138 kV - 20 meters, and 69 kV - 10 meters, as shown in Figure 2.5 taken from BC Hydro [122]. BC Hydro follows these guidelines [123] and HydroOne. This transmission corridor is designated

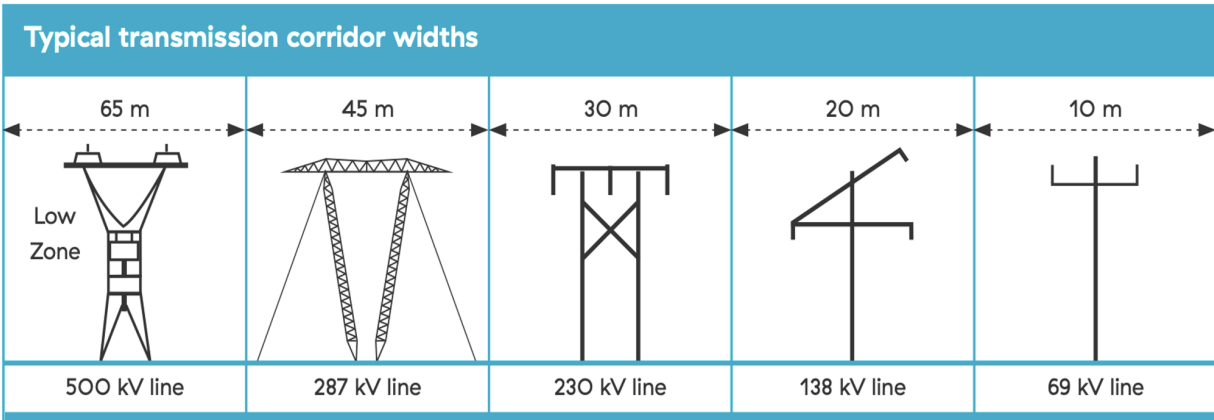


Figure 2.5: Utility transmission zones' width by BC Hydro.

to ensure that the vegetation in the area remains at the lowest height to avoid vegetation encroachment. The corridor zone of a buffer zone can contain medium-height vegetation, and tall-zone or noncorridor can have high vegetation and buildings, as shown in Figure 2.6.

Distribution Substations

This facility converts high-voltage electric power to high current and distribution voltage to deliver it to end consumers. These facilities step down the voltage to various levels depending on the end consumers of these distribution lines [121].

Utility Distribution Lines and Wire Zone

These powerlines transmit data to end consumers and deliver electric power to secondary users such as residential consumers. These lines are then divided into residential or service wires. The distribution powerlines have guidelines for wire zones and non-corridor zones as well. The guidelines state that within 5 meters on both sides of the distribution network, pylon is a low vegetation area that can only contain trees up to 5 meters tall. The corridor zone can contain trees up to 12 meters tall, and there is a non-corridor zone of 10 meters away from the pole for buildings and tall vegetation [122]. These guidelines and regulations are employed by most countries worldwide, with slight adjustments based on country-specific

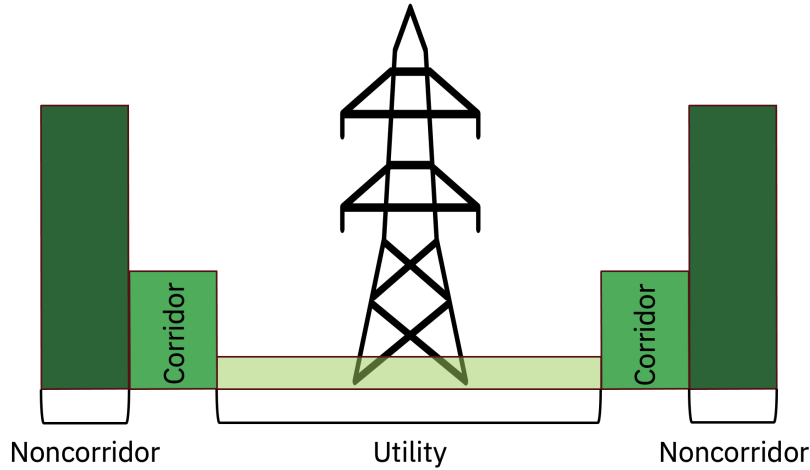


Figure 2.6: Utility transmission zones.

environmental and terrain conditions.

Understanding the utility powerline network emphasizes the importance of utility, corridor, and non-corridor zones. These zones are identified in Figure 5.2 a and show the relationship of vegetation with the voltage and height of the powerline and pylons. These regions and their relationship with objects have encouraged the imposition of spatial layout consistency for better segmentation of utility scenes.

2.7.3 Utility Vegetation Management

The crucial step in avoiding power outages is to manage vegetation around utility transmission corridors. Utility companies spend billions of dollars maintaining vegetation and ensuring reliable electricity delivery to consumers. After the most outrageous blackout in 2003, NERC established policies and guidelines to ensure vegetation management. Customer-related outages on the distribution network due to vegetation encroachment or contact still average 1 in 5 in Canada. Vegetation management techniques include stump removal, root management, and the identification of 'danger trees,' in addition to reducing tree-related outages through the pruning of back trees, underbrush, and vines [29, 30].

Over the last decade, most utility companies have started using LiDAR and cameras for surveying and inspecting utility transmission line corridors. These surveys are faster and provide datasets that can later be used to identify vegetation encroachment and powerline anomalies [31, 32]. Canada is the second largest country by area, and managing and maintaining the aging infrastructure is crucial. LiDAR has enabled faster surveys, but the risk analysis and vegetation management must be automated. These techniques are important, as manually inspecting the collected 3D point cloud data is labor-intensive and fatigue-inducing.

It is possible to semantically analyze the 3D point cloud and satellite imagery for span-to-span risk analysis and interpretation of utility regions along utility objects and vegetation [124, 125, 126, 127]. This vegetation can further be analyzed through temporal semantic analysis to observe the growth of various trees and avoid vegetation-based power outages [44].

In addition to being essential to operations, proper vegetation management reduces the risk of wildfires starting because of electrical lines by clearing slash or vegetation from right-of-way.

2.7.4 Historical Context of High-Voltage Transmission Line Maintenance in Canada

The network of high-voltage transmission lines in Canada, extending over 166,000 km, is a testament to the vastness and complexity of the national energy infrastructure. Maintaining such an expansive system is paramount for the energy sector, necessitating periodic replacement of critical components like pylons and powerlines as they wear out. Vigilant monitoring to identify and rectify any potential disruptions, including routine vegetation management, is also a part of this crucial maintenance routine. The largest power outage in North American history, affecting over 50,000 people on August 14th, 2003, was largely attributed to a failure in communication between utility companies and inadequate vegetation maintenance. The blackout lasted 30 hours, with a state of emergency declared until August 22nd. While some outages have been historically linked to increased demand during heatwaves natural disasters

like hurricanes and storms, the vegetation-induced outages highlight the necessity of regular surveillance, tree trimming, and infrastructure inspections to prevent future incidents [128, 129].

2.7.5 Computational Techniques in Utility Infrastructure Analysis

This section reviews various computational techniques, from machine learning to deep learning, and their applications in utility infrastructure analysis, such as classification, detection, and reconstruction of utility corridors and components.

Utility Corridor Classification and Object Detection

Utility corridor classification and object detection are key areas for utility applications. These systems support utility infrastructure management and monitoring. One such method involves combining statistical and geometrical techniques for scene classification, which utilizes Markov Random Field (MRF) for utility objects and RANSAC for building classification [21].

However, machine learning algorithms such as Support Vector Machines [17] and decision trees [18] have been used to classify utility objects. One significant paper proposed using random forests for utility scene classification, as detailed in [130]. This research extracted 21 features from 3D voxels of the point cloud. Various combinations of features have been employed, including density-based, eigenvalue-based, height-based, echo-based, convex-hull-based, vertical profile-based, Hough transformation features, and Principal Component Analysis (PCA)-based features.

Another machine learning approach, presented in [20], utilized an ensemble-like classifier trained with different sets of features. One set was extracted using methods defined in [130], while the other set employed PCA-based features. These two random forest models collectively performed point-based classification and demonstrated improved accuracy compared to [130]. The extension of the research work published in [20] is presented in [18]. This research extended the use of supervised classification employing 21 features with a wide variety of experiments better to handle the balanced class distribution in the dataset. The study also

compared the importance of features toward the classification results to understand the model’s sensitivity. The strength of this research lies in its ability to generate a point-based classification of the 3D point cloud using a machine learning-based supervised algorithm [18]. Another piece of literature generates labeled point clouds for wires, pylons, and other classes using random forest [131].

One of the key limitations identified by previous research work in this review is the influence of the ground class on the prediction, causing an imbalance. The research article proposed a solution that used the JointBoost classification model [19]. It utilizes fewer features than other machine learning-based supervised classifiers and selects the important features from the given input for classifying the scene. By measuring the height differences, the proposed classifier adopted a step-off count to avoid errors in labeling building surfaces as ground. This module improves performance and reduces misclassification of building rooftops. Another module called graph cut segmentation is included to deal with cases where the foreground is not tightly connected to the background. Although it is a one-shot supervised method without any intermediate steps, it still heavily relies on handcrafted features and an offline process of setting a threshold for step-off to avoid errors in the ground class.

The features used in machine learning methods are not learnable and are limited by the researchers’ knowledge. The strength of existing methods lies in understanding the relationship between span and pylon and identifying the use of wire direction for pylon localization, inspiring this thesis to utilize these relationships.

Powerline-Net [15] is a deep learning-based high-voltage transmission line semantic segmentation network proposed to address local features and downsampling approaches for power lines. It has shown good performance on the in-house dataset presented in the paper. However, it only deals with one type of high-voltage power line (500kV). DCPLD-Net is another deep learning network that generates 2D cross-sections and uses a diffusion-based deep neural network to detect power lines and classify everything else as a non-wire class. The limitation of the proposed system is its use of only 2D cross-sections and detection of just power lines [132]. Another deep learning method extracts the power line by filtering 90%

Utility Network Model Reconstruction

Automatic 3D powerline reconstruction is a topic of prior research in the utility community that facilitates powerline inspection and monitoring to prevent vegetation encroachment and anomaly detection. A technique proposed in [133] for 3D powerline extraction and modeling involves minimum linkage hierarchical clustering and RANSAC. This method uses hierarchical clustering to group candidate points for powerlines and RANSAC for robust model fitting by iteratively selecting subsets of points.

This is followed by another reconstruction technique for 3D modeling of wires using catenary curve embedded local affine modeling. Various researchers also adopted this method to incorporate the catenary curve formed by the overhead powerlines. This approach uses affine transformations to preserve powerline points' collinearity and ratios of distances [134]. A key piece of research in this domain proposed a voxel-based classification method and piece-wise reconstruction of 3D powerlines, as detailed in [125].

The research work presented in the paper [135] extends the approach in [125], addressing reconstruction errors found at different levels of spans, especially in overlapping regions where inner spans and closely located spans are difficult to differentiate. The multi-level span analysis mitigates these challenges by considering the relationship of powerlines with points of attachment to pylons. Another voxel-based hierarchical subsampling method employed a series of statistical filters to extract the ground and then powerline candidate points, successfully clustering and labeling powerlines. A polynomial mathematical model was used to reconstruct the powerlines fully [127].

Eigen-analysis of the point distribution tensor is proposed for 3D powerline reconstruction [136]. It computes eigenvalues and eigenvectors of a tensor derived from the point distribution of powerlines. This method helps understand the distribution and orientation of points in 3D space, which is valuable in 3D modeling and reconstructing powerlines.

All the proposed utility powerline reconstruction models have suggested techniques for acquiring the point cloud in good weather conditions. This model is proposed for reconstructing 3D powerlines affected by wind, causing noise in the point cloud. The method

calculates parameters to adjust the wind-related noise in the dataset and reconstructs the 3D powerline for utility network modeling and management [137].

These methods represent diverse strategies for reconstructing 3D point cloud data models, each with unique mathematical and computational approaches. However, there are limitations due to the conventional classification methods used in these techniques, which result in reconstruction errors.

Utility Network Asset Management

Utility asset management is the process utility companies employ to optimize, maintain, and improve their services and physical assets. Infrastructure such as power plants, substations, transformers, meters, pylons, powerlines, and other devices required to provide services to clients are included in these assets. To guarantee utility services' reliability, security, safety, and effectiveness, effective utility asset management incorporates various procedures and tactics [138]. Among the crucial facets of utility asset management are maintenance of asset inventory, condition assessment of assets, repair and replacement planning, compliance with regulations, risk management, cost optimization, and data analysis. The asset management facets, when combined with the latest technological updates, can reduce operational costs by 10-20% and enhance the overall reliability of the network [2]. Utility corridor classification, object detection, and utility network model reconstruction support asset management.

One such method, proposed in the research article by Kala et al. [139], employs Faster-RCNN on optical images to detect utility powerline network assets such as transformers, insulators, IACM, IATCT, pylons, and supports. Detecting these components aids in generating reports indicating what needs fixing or replacement. However, the method's limitation lies in its inability to completely distinguish between IACM, IATCT, and insulators due to their structural similarity. Nevertheless, this object detection method significantly contributes to improved asset management.

Another technique is based on historical asset conditioning records, using K-Means clustering to predict asset repairs and failures. This method does not directly utilize images

or point clouds but can utilize information retrieved from imagery or 3D point clouds collected and analyzed for monitoring and inspection [140].

Utility asset management also involves managing assets to better cope with adverse weather conditions and environmental catastrophes. MIT and National Grid USA researchers have developed predictive models using weather forecasts, asset conditions, network conditions, historical damages due to weather, and power outage records. This system aids in maintaining the reliable delivery of electricity during storms and extreme weather conditions [141].

Utility Vegetation Management Analysis

As previously discussed, vegetation management is a key operation for maintaining clearance in utility corridors and ensuring the reliable delivery of electricity. It involves classifying vegetation in utility scenes and predicting tree species to guide risk management based on annual tree growth for planning trimming to avoid encroachments [31, 32].

A key research article focuses on observing single trees within the right-of-way to detect growth and dangerous trees that might lean onto transmission lines, causing outages or blackouts. This research utilizes features such as lines, volume, clusters, 3D buffers of points, and the overall tree shape for detecting tree genera. It also develops geometric tree metrics to predict the direction of tree growth and overlap with corridor zones. This information aids in vegetation analysis [142]. Another article proposes a hybrid ensemble method for tree genera classification using 3D point clouds acquired through ALS [143]. This classification method uses random forest and replaces the base classifier with various models such as Linear Discriminant Analysis (LDA), SVM, and k-nearest Neighbors (kNN). The results show improvements in both 3D and vertical profile classifiers, enabling the detection of tree genera like pine, poplar, and maple. These tree detection methods, combined with the geometric tree metrics proposed in [142], can assist in clearing the corridor zone.

In addition to machine learning techniques, deep learning methods for single tree detection have been proposed in the last decade. One method converts tree point clouds from forestry datasets, generates voxelization, and performs classification using Deep Belief Net (DBN) due

to its superior performance with small datasets [144]. Another technique uses voxel-based 3D CNN to classify and detect single trees from ALS point clouds of urban scenes [145]. The MBNet deep learning approach first filters the terrain from the point cloud and then utilizes multichannel features with the point cloud to individually detect trees in the forest. This approach has shown promising results. Existing 3D object detection networks such as VoxelNet [146], Part-A² Net [147], SECOND [148], and PointPillar [149] can also be trained on tree benchmarks for the detection of single trees.

These asset and vegetation management techniques provide a solid foundation for the semantic and panoptic segmentation of 3D point clouds. This thesis can aid in generating asset management reports, further enhancing decision-making processes.

2.8 Summary

This chapter provides an extensive overview of data collecting and aerial laser scanning methods, emphasizing the issues that lead to complicated noise in 3D scene reconstructions. It highlights the shortcomings of existing approaches, especially the underutilization of physical priors, and suggests incorporating physical priors into deep learning systems to mitigate noise. The chapter sets the stage for a detailed exploration of proposed methods in subsequent chapters, emphasizing the necessity of analyzing utility scenes in large 3D point cloud datasets. In addition to filling gaps in previous research, it describes the utility powerline network, its constituent parts, and the significance of vegetation management. It also paves the path for developing a deep learning-based network with learnable characteristics to use spatial relationships in utility scene segmentation.

Chapter 3

Simulation-based Data Augmentation using Physical Priors

This chapter discusses the methodology of noise simulation based on physical priors aimed at understanding the noise generation in airborne LiDAR. The chapter presents a case study of sensor noise, simulation-based data augmentation, and the performance evaluation of various deep neural networks for noise filtering. It concludes by assessing simulation-based data augmentation on the 3D U-Net deep neural network.

3.1 Overview

This section addresses the issue of noise in LiDAR data and its impact. Airborne LiDAR captures the real world, and unwanted elements, such as sensor noise or adverse weather conditions like fog, rain, or snow, can compromise the perceived quality of the scene.

A drawback of MPIA technology is that atmospheric points are more likely to overlap with objects of interest, as depicted in Figure 3.1, which shows a 3D point cloud captured using Teledyne Optech’s Galaxy T1000 airborne LiDAR. While atmospheric returns have always existed, they could be effectively filtered out with simple height filters or nearest-neighbor algorithms at lower PRFs. However, when atmospheric points mixed with more

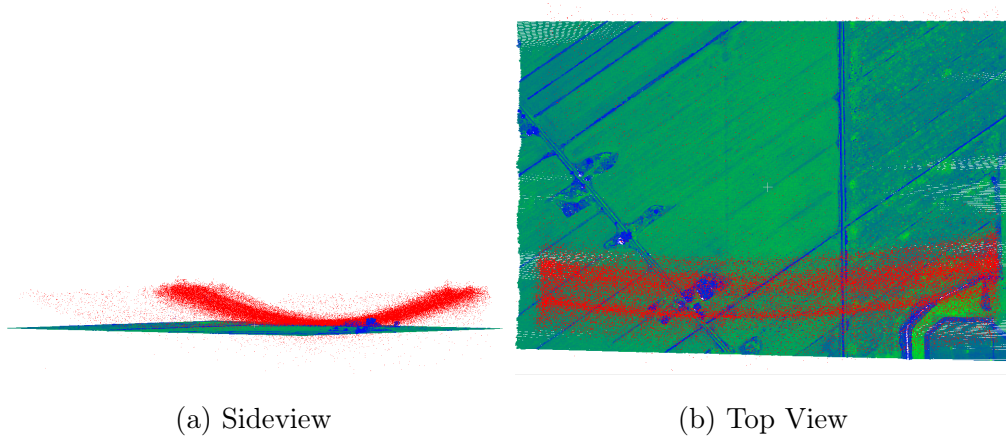


Figure 3.1: 3D Point Cloud from Teledyne Optech’s Galaxy T1000; Red Points represent Noise. .

sophisticated algorithms are required objects of interest, more sophisticated algorithms result in denoising 2D images using statistical, machine learning, and deep learning methods [25]. Some sophisticated deep learning methods, such as PointCleanNet [4], have addressed sparse noise points for meshes. However, limited research exists on noise filtering specifically for adverse weather conditions in autonomous driving systems, where sensor noise is not considered [78, 150].

Developing a deep neural network for noise filtering requires diverse annotated datasets, which can be labor-intensive and expensive to obtain. To overcome this, the simulations were designed to replicate the noise to analyze and augment the dataset for training the noise-filtering neural network. Noise filtering utilized physical priors and Gaussian generative functions to generate synthetic noise with varying densities. They also incorporated the MPIA technology to estimate the distance R_{OBS} traveled by each pulse beyond the maximum range before the next beam was fired [106]. The proximity of each point to its respective PIA zone, determined by R_{OBS} , can indicate the point belonging to the systematic noise pattern.

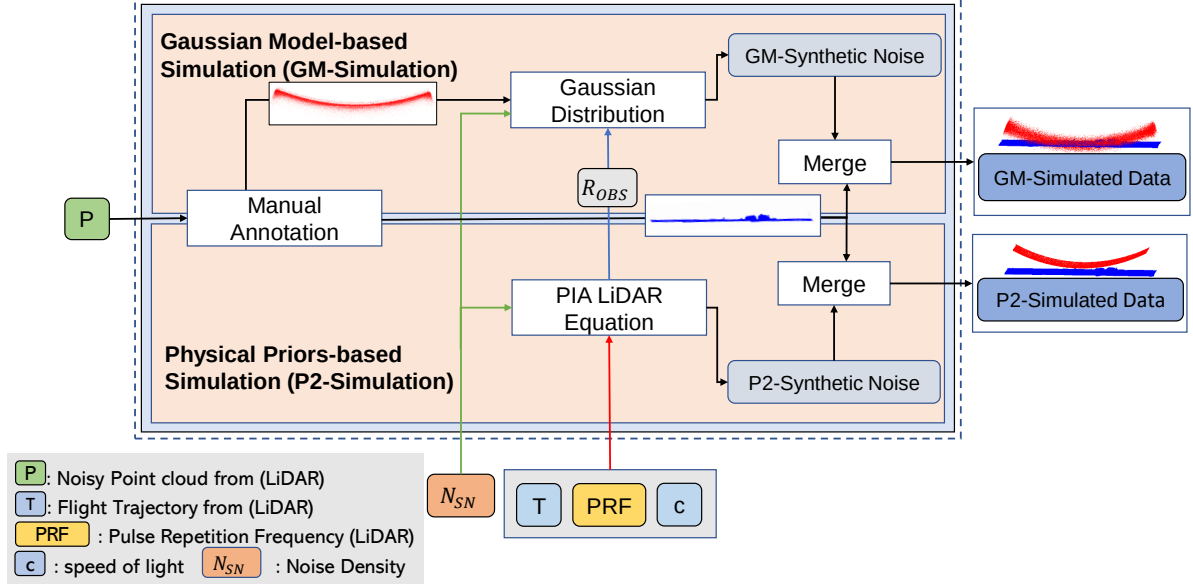


Figure 3.2: Parametric Generative Model: A Simulation workflow which explains that P2-Simulation only uses LiDAR physics, and GM-Simulation uses Gaussian function and LiDAR physics to generate noise.

3.2 Methodology

In this section, the proposed methodology for synthetic noise simulations is explored. The fundamental concepts of LiDAR scanning will be explained initially to ensure a comprehensive understanding of the approach. Following that, two methods will be introduced: physical priors-based simulation and Gaussian model-based simulation, both aimed at generating synthetic noise.

Figure 3.2 illustrates the workflow of the simulations. Physical Priros-based Simulation (P2-Simulation) and Gaussian Gaussian Model-based Simulation (GM-Simulation) use the PIA LiDAR equation to estimate physical priors. However, GM-Simulation also uses real noise points' normal point distribution function.

3.2.1 P2-Simulation

The simulation based on physical priors assumes that atmospheric points' travel time is shorter than the target data's. The simulation model uses PRF, the speed of light c , regular data points P_{REG} , noise density D_{SN} , and the maximum sensing range R_{max} from Equation 2.3 to generate simulated data as output. Next is the step-by-step process of developing P2-simulated data.

- Step 1: A noisy raw point cloud P acquired from LiDAR is manually annotated as P_A by domain experts into noise and non-noise points.
- Step 2: In this step, P2-Simulation subsamples N_{SN} non-noise or regular points, denoted as P_{REG} , from the annotated point cloud P_A based on the noise density D_{SN} chosen for simulation.
- Step 3: P2-Simulation uses R_{max} and the flight trajectory T to determine the position of the LiDAR (l) (x_l, y_l, z_l) at a specific time when the laser beam is emitted at the speed of light, denoted as c , and with the sensor's configured PRF.
- Step 4: The round-trip travel time is calculated for each laser pulse from emission to hitting the target point and returning to the receiver. This calculation applies to all subsampled points and utilizes the specific equation referenced as Equation 3.1. For example, should the range R_{REG} for subsampled point be 100 meters, then $t = 0.067 \mu s$.

$$t = \frac{2 \times R_{REG}}{c} \quad (3.1)$$

- Step 5: Simulation introduces pseudo-random errors to the travel time t of subsampled non-noise points to generate travel times for synthetic noise points, denoted as t_{SN} , simulating a mismatch on the receiver.
- Step 6: The range R_{SN} is calculated using t_{SN} for synthetic noise points P_{P2-SN} using Equation 3.2.

$$R_{SN} = \frac{c \times t_{SN}}{2} \quad (3.2)$$

- Step 7: Synthetic noise points P_{P2-SN} are generated by adding R_{SN} to the respective LiDAR positions of subsampled regular points mentioned in Step 2.
- Step 8: P_{P2-SN} is then merged with all regular points P_{REG} to generate P2-Simulated data.

The synthetic noise generated by P2-Simulation is projected onto the transition point between two PIA zones. However, it does not have the same point distribution of points as real noise. The reason P2-Simulation generates synthetic noise points that are dissimilar to real noise points is due to the sole use of travel time, disregarding the distance R_{OBS} a laser beam travels beyond the maximum range to hit the target and the possibility of multiple pulses in the air simultaneously. To address this challenge, GM-Simulation is proposed in the next section.

3.2.2 GM-Simulation

For the second simulation, the Gaussian generative model is chosen to overcome the limitations of P2-Simulation and incorporate range constraints. This information pertains to the distance a laser beam covers beyond R_{max} to reach the target point. The simulation model factors in the Pulse Repetition Frequency (PRF), the speed of light c , regular data points P_{REG} , flight trajectory T , and noise density D_{SN} , generating simulated data.

To calculate the physical priors, the PIA LiDAR equation is utilized, determining $PIA = \lceil \frac{R_{RN}}{R_{max}} \rceil$, and the distance R_{OBS} traveled by a pulse beyond the maximum range R_{max} . PIA provides information about the assigned PIA zone for a regular point. These physical priors are then employed in the generative noise functions to estimate the Gaussian parameters. The following outlines the step-by-step process for generating GM-simulated data.

- Step 1: A noisy raw point cloud P acquired from LiDAR was manually annotated as P_A by domain experts into noise and non-noise points.
- Step 2: In this step, GM-Simulation subsamples N_{SN} non-noise or regular points,

denoted as P_{REG} , from the annotated point cloud P_A based on the noise density D_{SN} chosen for simulation.

- Step 3: GM-Simulation utilizes the flight trajectory T to determine the position of the LiDAR (x_l, y_l, z_l) at a specific time when the laser beam is emitted at the speed of light, denoted as c , and with the sensor's configured PRF.
- Step 4: Simulation calculates R_{max} using Equation 2.3 from the speed of light c and PRF.
- Step 5: To simulate synthetic noise similar to real noise, the simulation extracts real noise points P_{RN} from the annotated point cloud P_A .
- Step 6: The range for real noise points R_{RN} is calculated using Euclidean distance based on the LiDAR position information for subsampled regular points obtained in Step 3.
- Step 7: To compute the normal distribution of real noise points, the simulation requires the respective R_{OBS} . This involves determining PIA zones for real noise points, calculated using Equation 3.3. Utilizing PIA_{RN} zones, R_{OBS} is calculated for real noise points using Equation 3.4.

$$PIA_{RN} = \lceil \frac{R_{RN}}{R_{max}} \rceil \quad (3.3)$$

$$R_{OBS} = R_{RN} \mod R_{max} \quad (3.4)$$

- Step 8: The simulation calculates the mean μ_{Obs} and standard deviation σ_{Obs} of R_{OBS}^{RN} for the real noise points.
- Step 9: The simulation utilizes the μ_{Obs} and σ_{Obs} obtained from the previous step, along with N_{SN} from Step 1, to provide parameters to the Gaussian generative function in Equations 3.5 and 3.6. This calculation determines the range for synthetic noise points

R_{OBS}^{SN} .

$$R_{OBS}^{SN} \sim \mathcal{N}(\mu, \sigma^2). \quad (3.5)$$

$$\mathcal{N}(\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (3.6)$$

- Step 10: R_{OBS}^{SN} is employed to calculate the range for all synthetic noise points using the PIA zones of subsampled regular points selected in Step 2 and R_{max} determined by Equation 3.7. To ensure that the PIA zone for noise points is not the same as the PIA zone for regular points, they are projected in zones above the regular points' zones without compromising the point distribution.

$$R_{SN} = \{(PIA_{REG} - 1) \times R_{max}\} + R_{OBS}^{SN} \quad (3.7)$$

- Step 11: The simulation added R_{SN} to the respective LiDAR positions for subsampled regular points from Step 2 and generated GM-Simulated synthetic noise points P_{GM-SN} .
- Step 12: Lastly, GM-Simulated noise points P_{GM-SN} merged with the non-noise points P_{REG} of the annotated point cloud.

3.2.3 Noise Filtering Neural Network

An experimental study is conducted to observe the effects of noise filtering on 3D point cloud data, employing the 3D U-Net deep neural network as the baseline. The 3D U-Net architecture is chosen for its larger receptive field, performance efficiency, and state-of-the-art performance in semantic segmentation across various medical imaging datasets. Figure 3.3 depicts the architecture of the 3D U-Net, showcasing its capability to extract multi-resolution features and decode them back to the full resolution using skip connections.

The study gained a comprehensive understanding of the characteristics of LiDAR noise by analyzing the results of experiments and simulations. Additionally, the performance of the 3D U-Net was compared with other methods, including support vector machine (SVM), denoising autoencoder (DAE), and PointNet.

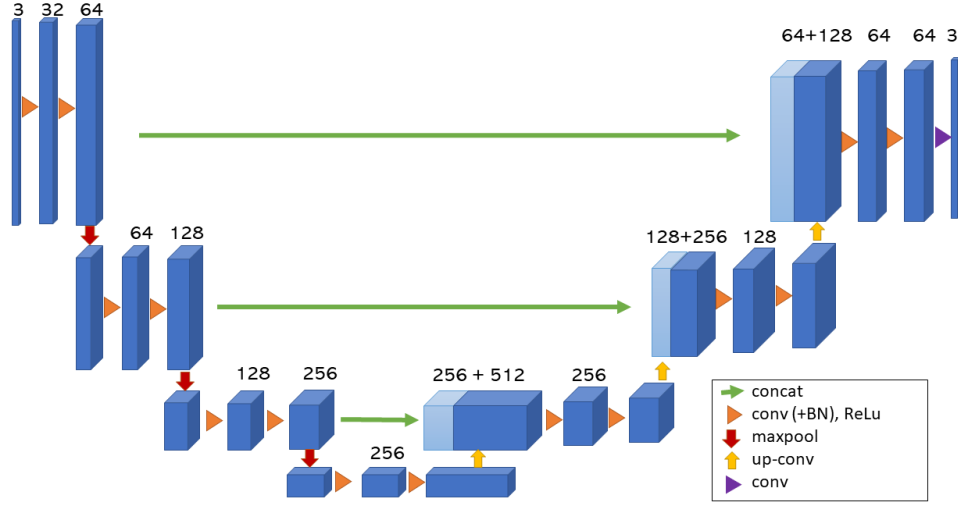


Figure 3.3: 3D U-Net Architecture is a baseline multi-resolution encoder and decoder with skip-connection, which is used to benchmark the performance of simulation-based augmentation and real noise-based filtering results.

To train the 3D U-Net, a substantial and diverse dataset was required. To address this, an augmented dataset was generated using simulation-based techniques. The 3D U-Net was trained both with and without the augmented dataset to assess any differences in performance. The training data, specifically synthetic data from the GM-Simulation, was selected as it provides a fair basis for comparison due to its resemblance to real datasets, as illustrated in Figure 3.2.

3.3 Experiment

The experimental study involved generating synthetic noise using simulations and verifying its accuracy by calculating the distribution for the observed range. Furthermore, additional experiments were configured to evaluate the performance of noise-filtering techniques.

In the first experiment, the 3D U-Net model was employed, and its superior performance in noise filtering was observed compared to other methods, such as SVM, DAE, and PointNet.

Finally, the study verified the effectiveness of simulation-based data augmentation and its role in enhancing noise-filtering techniques.

3.3.1 Dataset

The selected dataset was collected by Teledyne Optech’s Galaxy T1000 laser scanner, covering an area of $13km^2$. For experiments, the collected data was divided into train and test sets. Specifically, the first $4km^2$ of the dataset was allocated for testing purposes, while the remaining data was used to train the network.

The dataset comprised 14 distinct scenes that did not overlap, containing an average of over five million points, with a density of approximately $30points/m^2$. To establish the ground truth, Teledyne Optech employed Terrasolid point cloud processing software [151] to label the data by generating vertical profiles. Vertical profiles let the labeler view the point cloud as the 2D cross-section, which aids the labeling process, especially for noise points. These labels were developed based on extensive domain knowledge and technical expertise.

In the training dataset, two classes, noise, and non-noise, allow us to train binary baseline networks effectively.

3.3.2 Experimental Configuration

Each scene within the dataset is divided into four sub-scenes based on the GPS time of the flight line. A voxel grid with an input grid size of $64m \times 64m \times 64m$ is generated for each sub-scene. The batch size for processing these grids is 64, and each voxel has a size of $1m^3$. The grid features include the mean absolute elevation and the number of occupied points selected through a comprehensive feature engineering study detailed in the subsequent section.

The 3D U-Net produces outputs of size $64m \times 64m \times 64m \times 2$, representing confidence scores for the two 3D classes: noise and non-noise. The final prediction is determined by assigning the true label based on the highest confidence score. Additionally, post-processing

techniques are employed to address borderline misclassifications and enhance the overall accuracy of the predictions.

The training process for the 3D U-Net involves training on two GPU RTX 6000 units for a total of 90 epochs. The training duration typically ranges between 48 and 60 hours, while the inference phase, during which predictions are generated, takes approximately 2 to 3 minutes.

3.3.3 Evaluation Metrics

Precision, recall, and F1-score are selected to evaluate the network’s performance as noise is in the minority class. The precision determines the total number of predicted noise or non-noise out of the total instances of the class. Recall measures correctly predicted samples out of the total predicted sample of the specific class. The F1-score shows the balance between precision and recall. These matrices defined in the following equations 3.8, 3.9, and 3.10.

$$Recall = \frac{TP}{TP + FN} \quad (3.8)$$

$$Precision = \frac{TP}{TP + FP} \quad (3.9)$$

where TP is true positive, FN is false negative, and FP is false positive.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3.10)$$

3.4 Results and Discussion

3.4.1 Noise Simulation

The experiments conducted on sensor noise simulation and its use for data augmentation in noise filtering have produced intriguing results. Specifically, the GM-Simulation approach successfully generated synthetic noise that resembles real noise, as illustrated in Figure 3.4.

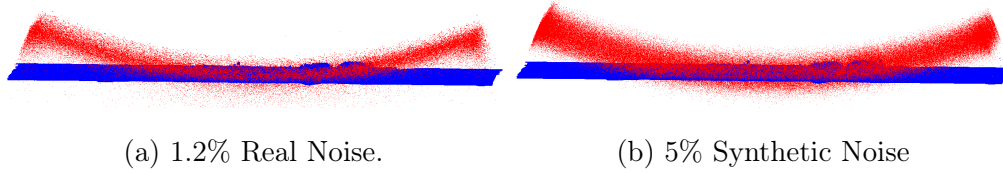


Figure 3.4: GM-Simulation Synthetic vs Real data; red: noise points and blue: regular points.

To further validate these findings, the experiment calculated the probability density function for the noise, as depicted in Figure 3.9. The results confirm a similar spatial distribution of noise over the range R_{OBS} , further supporting the accuracy and effectiveness of the simulation-based approach.

3.4.2 Noise Filtering Neural Network

The comparison study on noise filtering observed that the 3D U-Net model achieved the highest recall and F1-score for the noise class while also demonstrating comparable precision. These metrics play a crucial role in effectively addressing noise filtering problems. Recall measures the ability to correctly identify noise points among the correctly predicted points, while the F1-score provides a balanced assessment of the completeness and correctness of the predicted noise points. The results presented in Table 3.1 show the performance improvement of 31.2% over the SVM benchmark for recall, 37.4% for precision, and 34.0% on the F1-score. These results shed light on the challenges involved in noise filtering compared to semantic segmentation; hence, the point-based semantic segmentation network fails to match the performance of voxel-based 3D U-Net.

The main reason behind the superior performance of the 3D U-Net model over other methods is its larger receptive field, which enables effective context learning to capture the global systematic noise pattern inherent in sensor noise. In contrast, other techniques may struggle to achieve comparable results due to their limited receptive field.

Methods	Noise		
	Recall	Precision	F1-score
SVM	46.8%	54.0%	50.1%
DAE [152]	53.6%	99.9%	69.7%
PointNet [7]	68.3%	98.2%	80.5%
3D U-Net [9]	77.9%	91.4%	84.1%

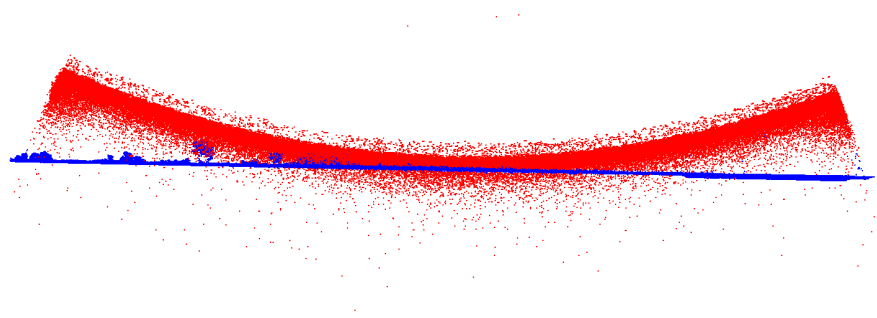
Table 3.1: Comparison of 3D U-Net for noise filtering with SVM, Denoising Autoencoder, and PointNet over Recall, Precision and F1-score.

3.4.3 Simulation-based Data Augmentation

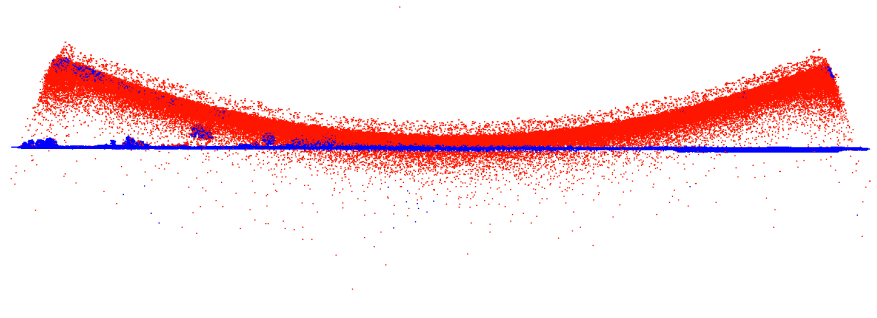
An additional experiment was conducted using the 3D U-Net model to investigate the impact of data augmentation on network learning. The network was trained using a combination of synthetic and real data and then tested on four real test scenes, as depicted in Figures 3.5, 3.6, 3.7, and 3.8. Training with synthetic data generated by P2-Simulation was excluded as it produced noise primarily at the transition point between two PIA zones, lacking complex and random overlaps with objects of interest. This would have led to an unfair comparison.

Empirically, the network trained with a combination of real and synthetic data outperformed the vanilla training approach in terms of precision (improving by 4.4% for test scene 2, 26.5% for test scene 3, and 28.7% for test scene 4) and F1-score (improving by 2.0% for tests scene 2, 15.8% for test scene 3, and 18.9% for test scene 4) for three test scenes. It demonstrated comparable performance in terms of recall, except for test scene 1, which shows the decrease of recall with augmentation by 14.3%, as shown in Figure 3.5. Test scene 1 had denser noise than the other scenes, and the training process might have been affected by the bias introduced due to the similar noise density throughout the synthetic dataset.

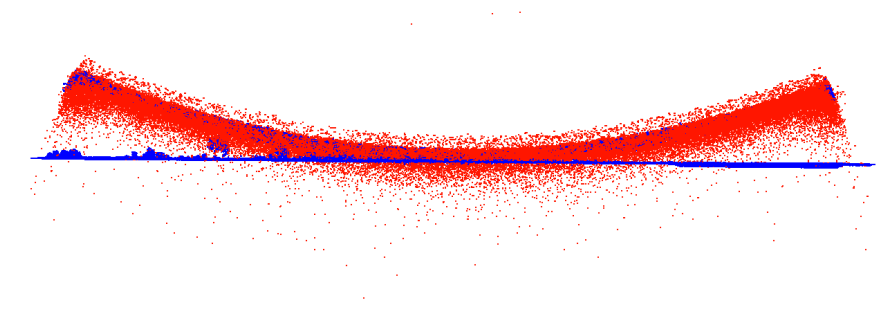
Conversely, test scene 4 performed better in all three metrics for the augmentation experiment, as shown in Table 3.2. The visualization of the results highlighted problematic areas where noise and regular objects overlapped or noise points formed clusters resembling complex objects like trees or bushes. These findings motivated further investigation into the



a) Ground truth

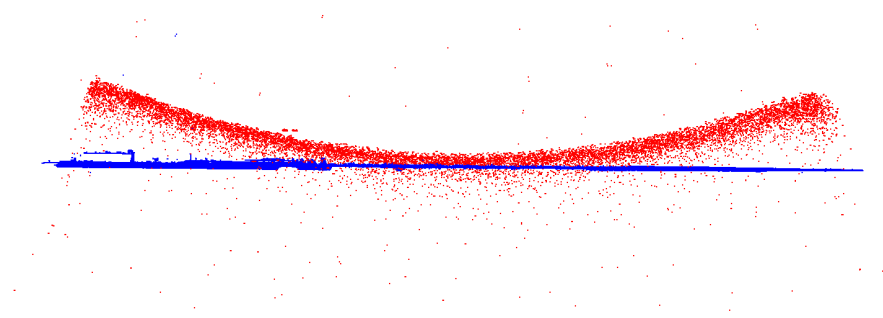


b) 3D U-Net (w/o aug.)

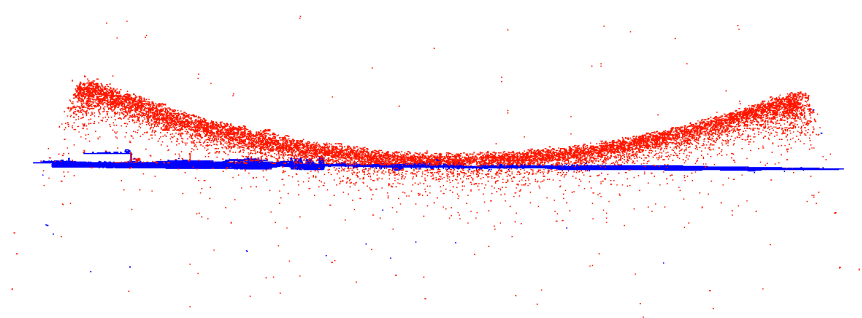


c) 3D U-Net (aug.)

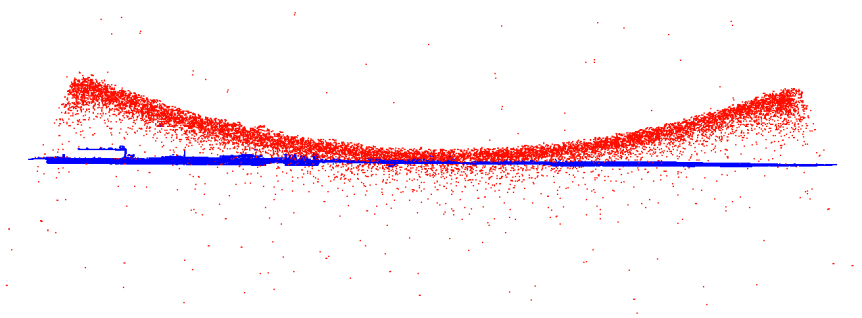
Figure 3.5: Visualization comparison of results for Test Scene 1; a) Ground truth, b) 3D U-Net (w/o aug.) and c) 3D U-Net (aug.). Red points are noise, and blue points are regular non-noise objects.



a) Groud truth

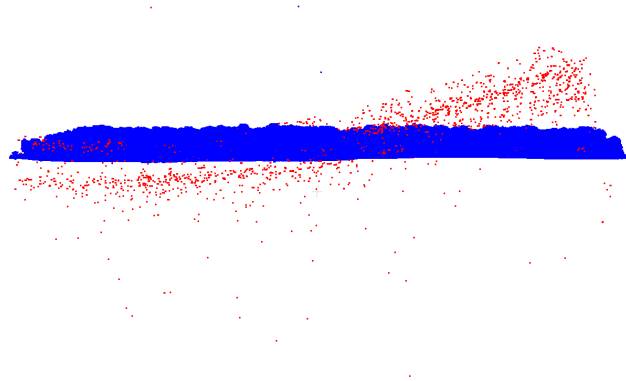


b)3D U-Net (w/o aug.)

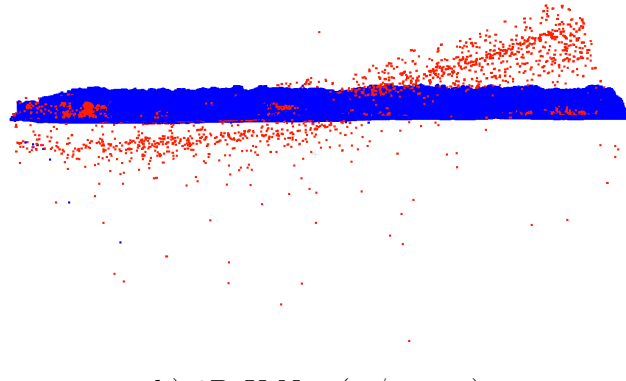


c) 3D U-Net (aug.)

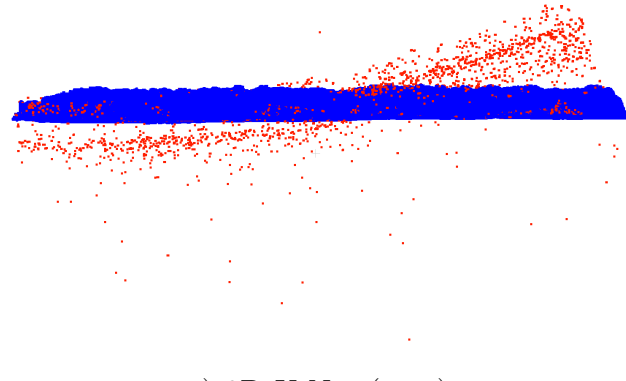
Figure 3.6: Visualization comparison of results for Test Scene 2; a) Ground truth, b) 3D U-Net (w/o aug.) and c) 3D U-Net (aug.). Red points are noise, and blue points are regular non-noise objects.



a) Ground truth

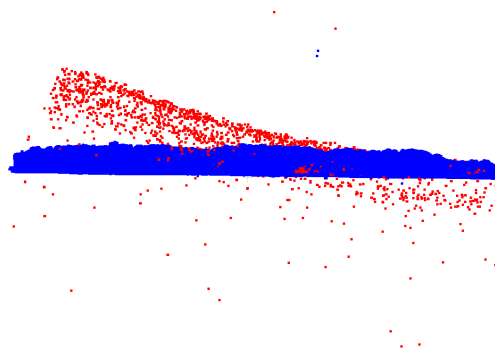


b) 3D U-Net (w/o aug.)

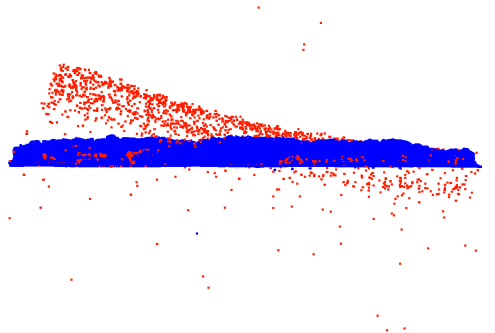


c) 3D U-Net (aug.)

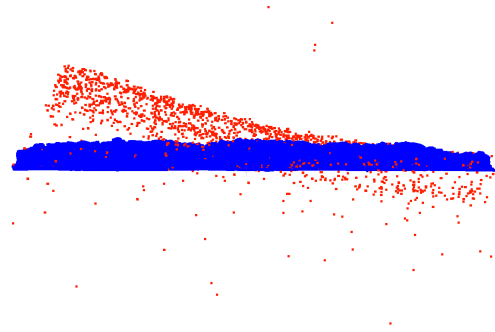
Figure 3.7: Visualization comparison of results for Test Scene 3; a) Ground truth, b) 3D U-Net (w/o aug.) and c) 3D U-Net (aug.). Red points are noise, and blue points are regular non-noise objects.



a) Ground truth



b) 3D U-Net (w/o aug.)



c) 3D U-Net (aug.)

Figure 3.8: Visualization comparison of results for Test Scene 4; a) Ground truth, b) 3D U-Net (w/o aug.) and c) 3D U-Net (aug.).

Scenes	Noise (w/o aug.)			Noise (aug.)		
	Recall	Precision	F1-score	Recall	Precision	F1-score
Test Scene 1	93.8%	96.5%	95.1%	79.5%	96.5%	87.2%
Test Scene 2	93.7%	90.7%	92.1%	93.1%	95.1%	94.1%
Test Scene 3	43.2%	17.6%	25.0%	38.8%	44.1%	40.8%
Test Scene 4	47.1%	23.5%	34.1%	53.8%	52.2%	53.0%

Table 3.2: 3D U-Net results for noise class using with and without augmentation for individual test scenes.

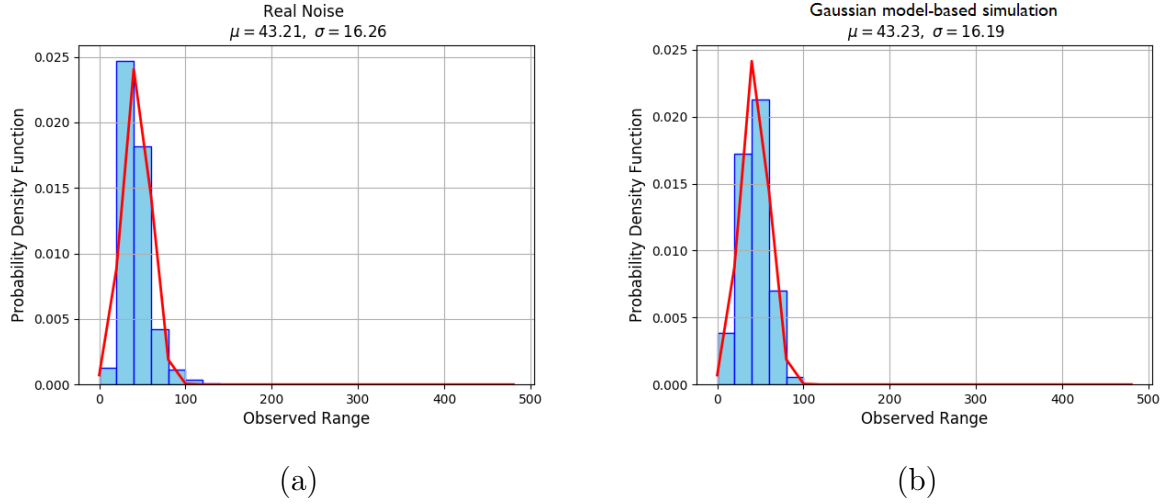


Figure 3.9: Comparison of the distribution of a) real noise from Figure 3.4a) and b) GM-Simulation synthetic noise from Figure 3.4b).

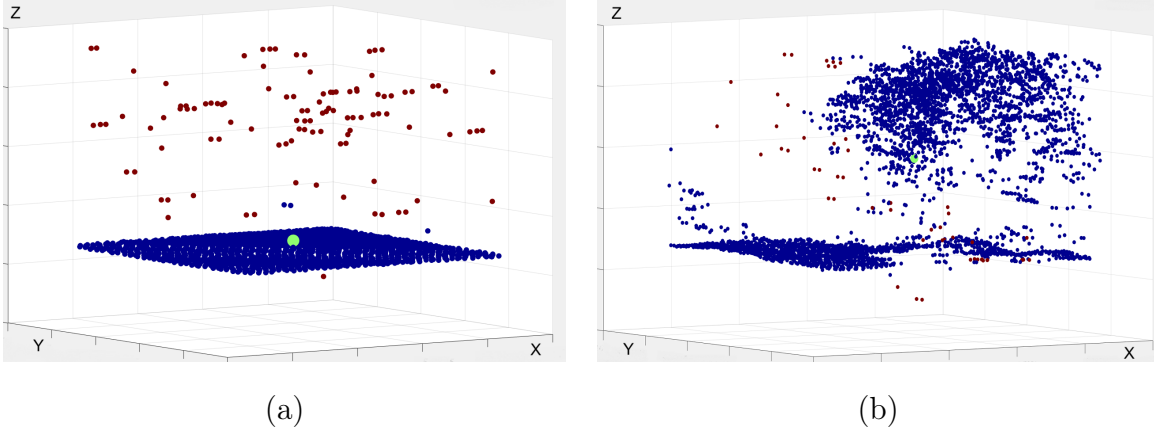


Figure 3.10: Two voxels from one scene, a) sparse noise far from the terrain, b) complex noise overlapping with bushes; red: noise points, blue: regular points and green: centroid.

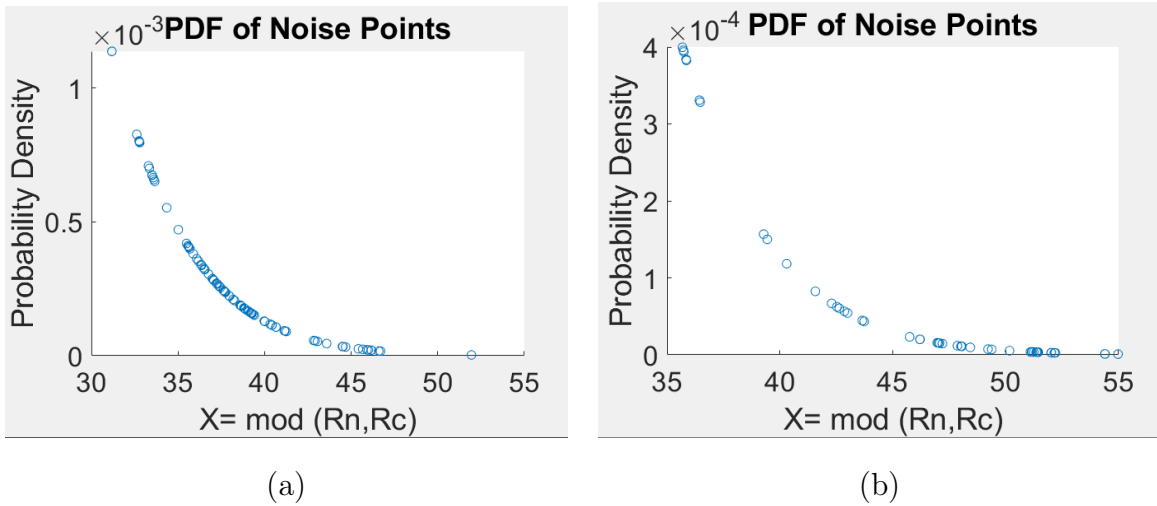
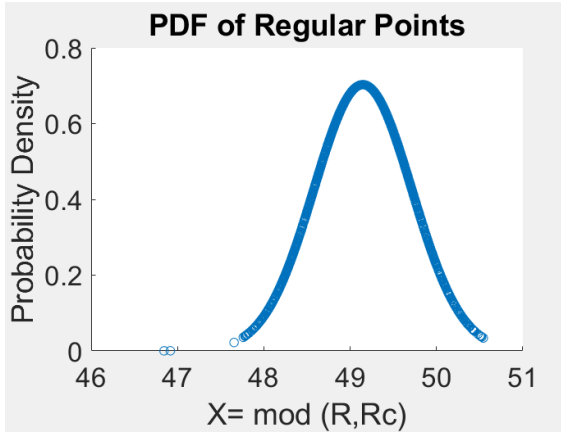
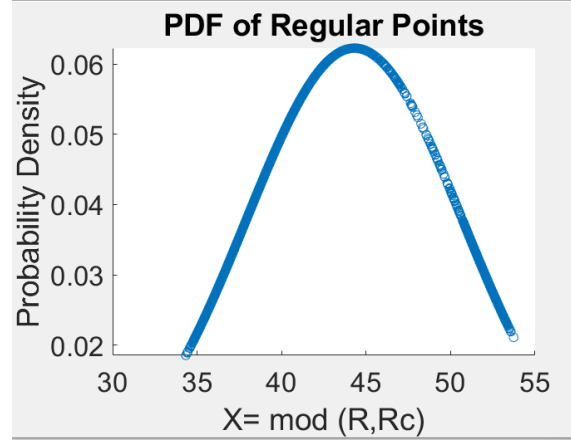


Figure 3.11: Normal distribution of noise points for voxels from Figure 3.10.

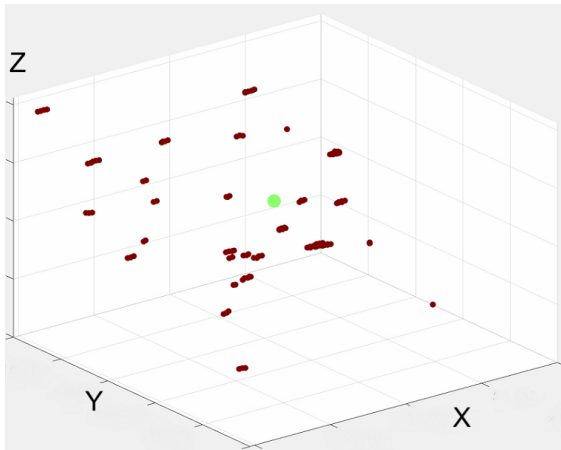


(a)

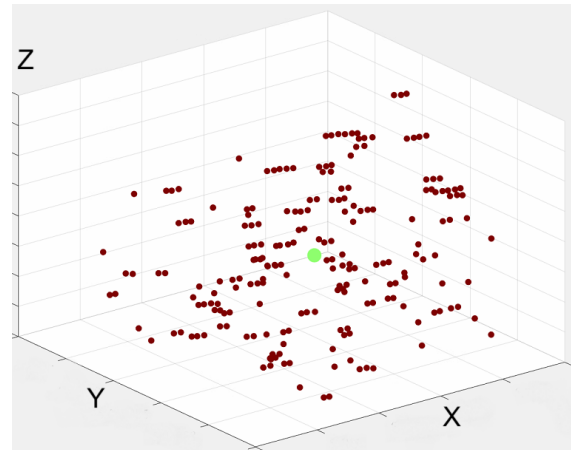


(b)

Figure 3.12: Normal Distribution of regular points for voxels from Figure 3.10.



(a)



(b)

Figure 3.13: Raw point cloud of $20m^3$ relatively far from terrain containing only noise, a) sparse noise, b) dense noise relatively close to the terrain than a); red: noise points and green: centroid.

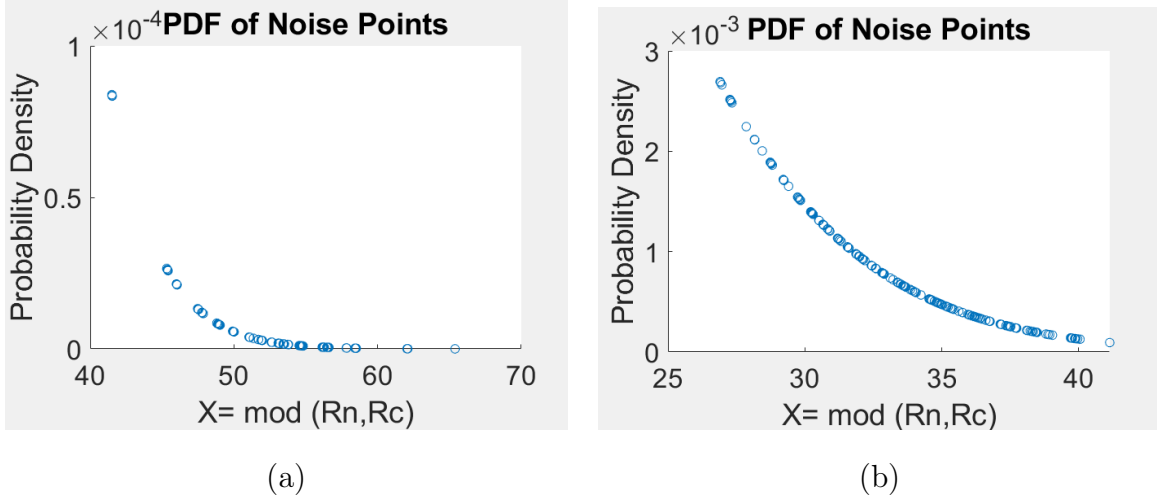


Figure 3.14: Normal distribution of noise points for voxels from Figure 3.13.

underlying global and local spatial distribution of points to understand the issues at hand better.

Exploring the issue further, the global spatial distribution of noise for all scenes was calculated, confirming the assumption that noise was uniformly distributed throughout the dataset. The local spatial distribution of noise points was examined to gain further clarity. Voxel cubes of size $20m^3$ were generated, ten times the input voxel size, to accommodate the limitations of storage memory resources.

Figure 3.10 depicts two cases of noise and regular objects overlapping within a voxel. The graphs in Figure 3.11 illustrate the probability density of noise points at R_{OBS} for the voxels shown in Figure 3.10(a) and 3.10(b). The probability of a random point being noise is lower in Figure 3.10(b) compared to Figure 3.10(a) due to larger overlaps and the position of the centroid. The graphs in Figure 3.12 show that the probability of a regular point being within the observed range interval of $[34-47]$ m is 0 for Figure 3.12(a), while for Figure 3.12(b), it is lowest at 34m and highest within the interval of $[42-46]$ m around the transition point of the PIA zone. This discrepancy is due to the distribution of points in the voxel being very close to noise.

Two additional voxels containing noise but with different densities were examined. The

normal distribution of points in Figure 3.13 indicates that the overall probability is higher for voxel Figure 3.13(b) due to its higher density. At the same time, it is lower for voxel Figure 3.13(a).

Based on these observations, the noise present in the dataset exhibits various characteristics such as complexity, randomness, and a global systematic pattern. Research categorized it into three major types: type I - complex noise, type II - sparse noise, and type III - systematic noise. This categorization will help compare the generalization and performance of the neural network concerning different noise types in future stages of the research.

3.5 Summary

This study entails a comprehensive investigation into the advancements of airborne LiDAR systems and their applications in data acquisition for scientific and commercial purposes. The research reveals that incorporating physical priors and the spatial distribution of points enables effective simulation of synthetic noise. The study presents the potential of simulation-based data augmentation to enhance noise filtering performance in specific scenarios. Moreover, analyzing the overall and local distribution of noise points provides valuable insights into the outcomes achieved through noise filtering techniques. Notably, differentiating noise types into type I (complex noise), type II (sparse noise), and type III (systematic noise) emerges as a crucial factor in these analyses. In the next chapter, these findings will serve as the foundation for developing an innovative deep neural network explicitly designed for noise filtering featuring a physical priors-based attention module.

Chapter 4

Noise Filtering Deep Neural Network using Dual-Attention

This chapter proposed a physics-based deep neural network for noise filtering in 3D point cloud data. The algorithm incorporates physical priors and employs a deep learning network to eliminate noise from the 3D point cloud. The network is influenced by data analysis and physical priors studies, crucial for understanding noise overlapping with objects of interest. The chapter studies two vision theories, attentional engagement theory (AET) and feature integration theory (FIT), guiding the incorporation of global-scale physical priors and spatial local attention into the network. Experiments using a domain-based dual attention module demonstrate the effectiveness of the proposed approach in removing noise from 3D point cloud data.

4.1 Overview

This section reveals the problem of noise filtering in 3D point c. An abstract overview of the proposed approach is shown in Figure 4.1. Experimental data shows variable noise density exhibiting characteristics like complex objects (i.e., trees and bushes). Furthermore, the global systematic noise pattern is localized on the transition point between two PIA

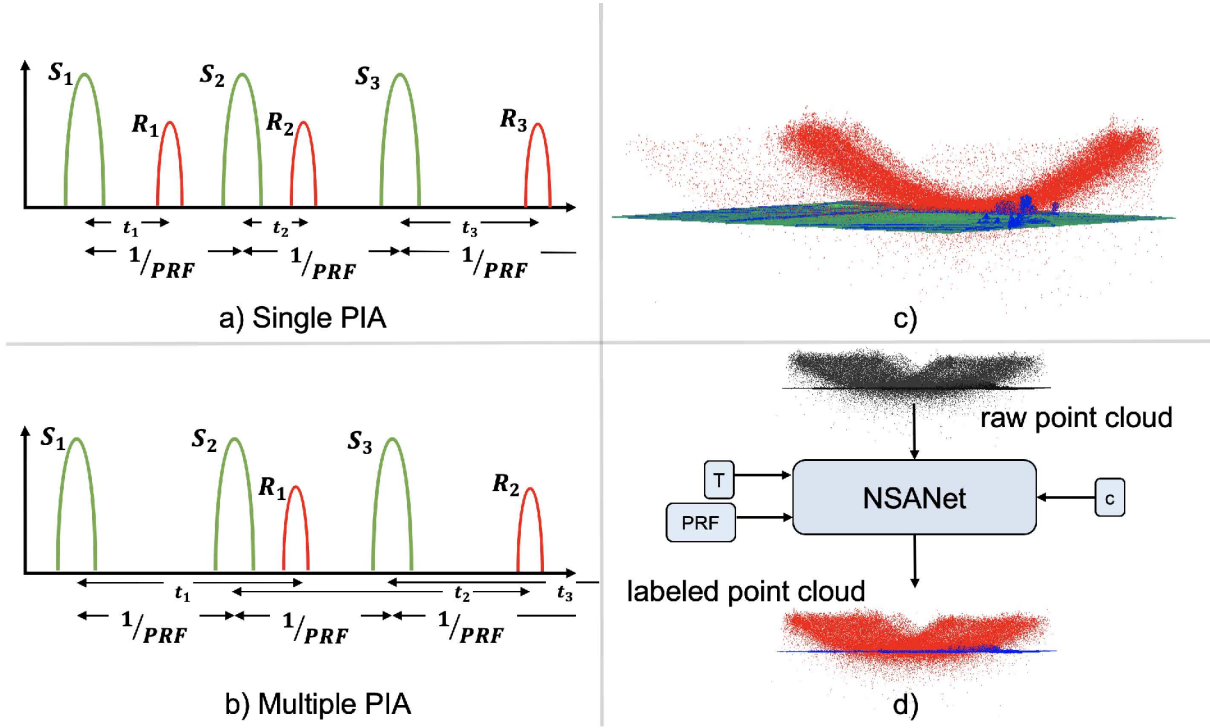


Figure 4.1: a) SPIA is a traditional technology which matched the signal sent by a transmitter (S) and received by a receiver(R) efficiently but was expensive, b) MPIA technology improved the cost and speed up the data collection though it is more prone to atmospheric points which can be seen overlapping with objects of interest in c), d) The proposed NSANet uses physical priors and spatial attention to filter noise from the point cloud.

zones and overlaps with objects of interest. Another major characteristic is imbalanced class distribution, i.e., noise points making up (1-3)% of the complete scene.

While the atmospheric returns always subsisted at lower PRF, it could be eliminated with simple height filtering or nearest-neighboring algorithms [25]. However, more sophisticated filtering algorithms are required for noise overlapping with objects of interest. Deep convolutional neural networks (CNNs) have surpassed traditional computer vision techniques in recent times [7, 6, 9, 56] which develop a large-scale noise classification deep neural network that efficiently and robustly filters noise from the point cloud.

This noise classification system integrates spatial contexts through attention mechanisms inspired by human attention phenomena. The goal is to focus on relevant objects by assigning more weight to relevant features, utilizing additive and multiplicative attention. Psychologically informed by the feature integration theory (FIT) and attentional engagement theory (AET), the network is designed to incorporate attention fusion at both encoding and decoding stages (FIT) and only at the decoding stage (AET).

The Noise-Seeking Attention Deep Neural Network (NSANet) proposed system is a novel prior-based dual-attention noise-filtering neural network. The chapter will discuss data analysis, physical priors, methodology, experimental configuration, and results. The network uses the physical properties of sensors and mapping principles to filter out noise. The design process involved two crucial studies: (i) data analysis and (ii) physical priors.

4.1.1 Data Analysis

The noise simulation study in the previous chapter identified the essential characteristics of a dataset. The following sections will discuss the findings and key features of the study:

- **Density:** Determining the varying density of noise points across a scene is vital to network design because it can be challenging to distinguish between foreground and noise. The noise points are not uniformly distributed and can appear in clusters, resembling complex objects like trees or bushes, while others are sparsely dispersed.

- **Localization:** The dataset also reveals another distinguishing characteristic of noise points: their localization. As mentioned, many noise points are projected at the transition point between two PIA zones. These atmospheric points exhibit a systematic global pattern that can be either dense or sparse, depending on the environmental conditions of the surveying area. Interestingly, this recurring noise pattern can be close to the objects of interest in some cases, while in others, it is found below the terrain.
- **Distribution:** The PIA zones are narrower because of the higher PRF, which results in atmospheric points overlapping with the foreground. This phenomenon is particularly noticeable in areas with trees, bushes, buildings, power lines, and occasionally even the terrain. Despite this, the atmospheric points account for only a small percentage (1-3%) of the overall foreground in the dataset.

4.1.2 Physical Priors

An extensive analysis to investigate various aspects of noise data points, such as their distribution, projection near the PIA transition zone, clustering with objects of interest, and how they affect the perceived quality of the scene, is conducted in the previous chapter [33]. The range ambiguities resolution of noise points results in projecting noise on the transition between PIA zones. Hence, the proposed network leverages this physical information, including PIA, PRF, and flight trajectory, to calculate the proximity of each point to the PIA transition zone. This proximity data is then used to generate a heatmap that indicates the probability of a point being a noise. The heatmap serves as a physical prior or external cue, enabling the network to effectively incorporate global-scale context and filter out noise. Based on these studies, the thesis has identified three noise types, as illustrated in Figure 4.2. This categorization motivated us to propose a design integrating physical priors to filter out noise.

- **Complex Noise:** This type of noise shares similar characteristics with trees and bushes and often appears in clusters that overlap with objects of interest.
- **Global Systematic Noise:** This noise type exhibits a dense or sparse distribution and

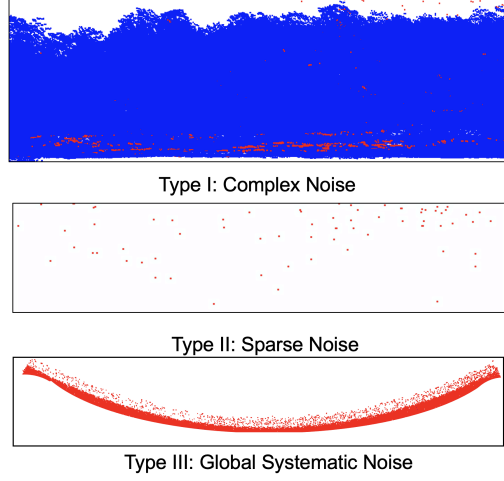


Figure 4.2: The three noise types: (1) Complex, (2) Sparse Noise, and (3) Global Systematic.

follows a parabolic systematic pattern localized at the transition point between PIA zones refer to Figure 2.2.

- Sparse Noise: It consists of outlier and sparse noise points located far away from the terrain and objects of interest.

4.2 Methodology

NSANet is a physical prior-based deep neural network for noise filtering that employs a multi-resolution approach and multiscale feature fusion dual attention mechanism. The network is designed to address all three types of noise mentioned earlier. To reduce ambiguity in distinguishing between noise and non-noise objects, It develops local spatial and global scale physical priors-based attention, inspired by [56]. The research highlights the significance of embedding the attention module in the encoder and decoder blocks based on the AET and FIT feature fusion theories.

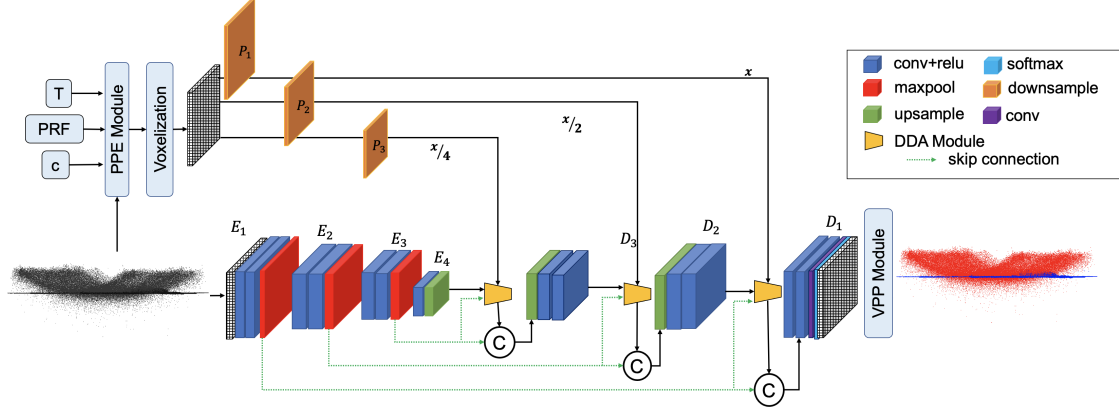


Figure 4.3: NSANet-1 based on AET. It uses the Physical Priors Estimation (PPE) module to generate a PIA proximity probability map converted into a voxel grid and downsample by a factor of 2. The multi-resolution encoder-decoder uses a Decoder Dual-Attention (DDA) module to eliminate noise based on local and global contexts. The Voxel Post-processing (VPP) module eliminates borderline misclassification of noise points.

4.2.1 3D Noise Filtering Neural Network

The NSANet architecture, depicted in Figure 4.3, is a U-shaped multi-resolution encoder-decoder that utilizes a three-dimensional voxel grid to learn deep features for noise labeling and filtering [9]. Research opted for a voxel-based architecture based on a comparative study, which demonstrated that 3D voxel-based networks outperform point-based networks (such as PointNet [7] and PointNet++ [6]) in terms of efficiency and performance for all types of noise. The network encodes four feature maps of different resolutions, denoted by E_1 , E_2 , E_3 , and E_4 . The first few layers extract essential low-level features, while deeper layers extract more sophisticated features. E_n decodes feature map D_n using local spatial additive attention between D_{n-1} and E_n , followed by multiplicative attention using the physical prior probability map P_n . This process is followed by two $3 \times 3 \times 3$ convolution operations, batch normalization, and relu activation. The output consists of three feature maps, D_1 , D_2 , D_3 , each with dimensions $\frac{H}{2l} \times \frac{W}{2l} \times \frac{D}{2l} \times 32l$, where l represents the number of layers. These feature maps are trained to focus on all three types of noise at multiple receptive fields,

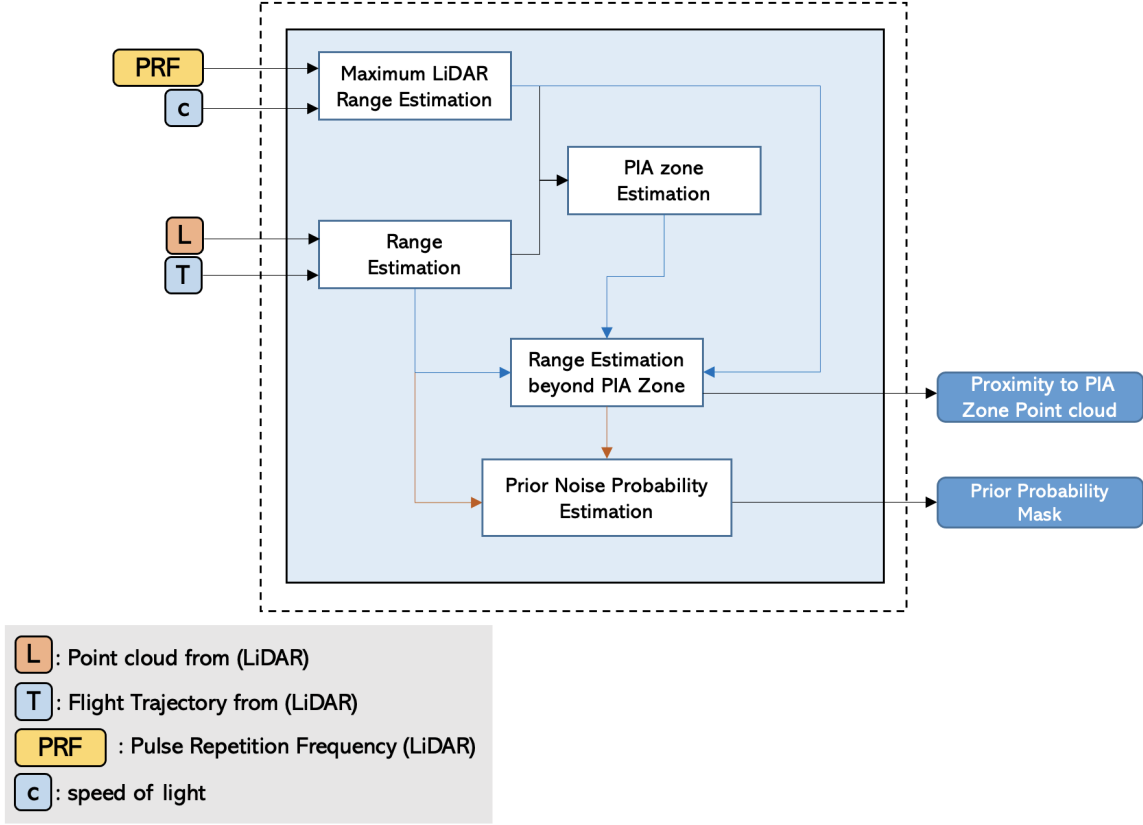


Figure 4.4: Physical priors estimation module workflow.

integrate global receptive fields through global attention, and subsequently remove noise points.

4.2.2 Physical Priors Estimation Module

The attention module in the proposed network requires a physical prior to estimating global attention accurately. It utilizes the speed of light (c) and PRF to estimate the maximum LIDAR range (R_{max}). Additionally, the system takes the minimum and maximum GPS time (P_{mingt} , P_{maxgt}) recorded from the 3D point cloud to determine the position (T_{match}) of the sensor during a specific flight time. The trajectory of the flight (T_m) and 3D point clouds (P_n) are used to estimate the range of each shot (R_n). Then, R_{max} and R_n are used to calculate

Algorithm 1 Physical Priors Estimation Module.

Input: $PRF, c, P_n, T_l, N, N_T, n = 1, 2, 3, \dots, N, l = 1, 2, 3, \dots, N_T, P_{mingt}, P_{maxgt}$ N = Total no. of raw input points N_T = Total no. of shots in flight trajectory**Output:** $Prob$: Physical Priors

$$R_{MAX} = \frac{c}{2 \times PRF}$$

while $l \leq N_T$ **do**| $T_{match} = T_l \leq P_{mingt}$ and $T_l \geq P_{maxgt}$ **end****while** $n \leq N$ **do**

| $T_m = P_{gt} - T_{match} < 1e - 2$

| $R_n = \sqrt{(P_n - T_m)^2}$

| $PIA_n = \text{ceil}(\frac{R_n}{R_{max}})$

| $R_{obs}^{[n]} = R_n \text{mod} R_{max}$

| $Prob_n = R_{obs} / R_{max}$

end

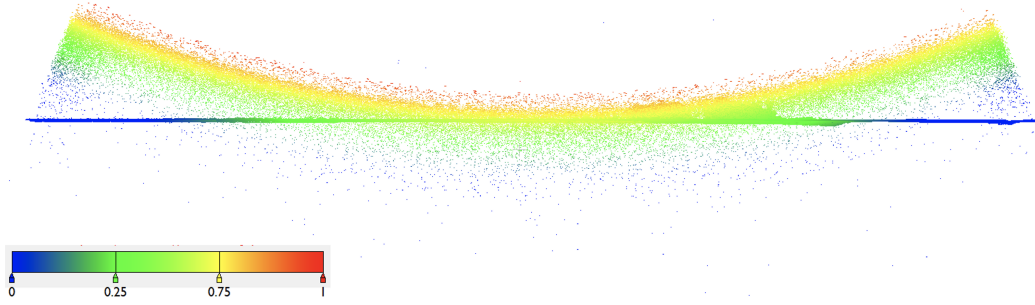


Figure 4.5: Physical priors calculated using algorithm no. 1, Higher the proximity to PIA transition-zone, high probability is of point to be part of a global systematic noise pattern as shown with red color.

their respective PIA zones (PIA_n). Furthermore, PIA zones, R_n , and R_{max} are utilized to compute the observed range (R_{obs}) beyond PIA zones, which is then translated into a probability map based on the distance from the transition point ($Prob_n$) as the workflow shows in Figure no. 4.4. The closer a point is to the transition between PIA zones, the higher the probability it is part of the parabolic systematic noise pattern, as shown in Figure no. 4.5.

4.2.3 Dual Attention Module

Shallow machine learning architectures or DAEs [152] can easily label sparse noise. However, complex noise often exhibits similar characteristics to objects of interest, making it difficult to distinguish between the two due to the distribution of points and shape variance. This ambiguity can lead to an overpopulation of false positives. We use local additive attention gates to address this issue, which handles spatial ambiguities, reduces misclassification, and suppresses irrelevant features when labeling complex noise. Global systematic noise patterns have been one of the most challenging problems in noise filtering. However, after careful study, we discovered that they only occur close to the PIA transition zone. Hence, the physical prior's heatmap from the PPE module embedded this information as global multiplicative attention. This attention mechanism focuses on global systematic noise patterns.

This novel attention model distinguishes itself from existing attention mechanisms by seamlessly integrating both additive and multiplicative attention, thereby facilitating the imposition of local and global contexts. This model incorporates a broader context, unlike conventional approaches that solely compute similarity maps through skip connections and hidden layers. This extended context is advantageous for mitigating ambiguous boundaries in local regions and activating pertinent global regions, contributing to a more comprehensive understanding of the data.

$$q_{add}^l = \sigma_1(W_x x_i^l + W_H x_i^l + b_H) \quad (4.1)$$

These attention gates take input from the previous layer of the decoder (D_{l-1}) for D_l as shown in Equation 4.1 $W_H x_i^l$ represent the weight of decoder tensor h_{l-1} and b_H is bias.

$$q_{mul}^l = (q_{add}^l \times W_p x_i^l) \quad (4.2)$$

The encoder feature map of the same level (E_2) as shown in Equation 4.1 denoted as $W_x x_i^l$ and the prior probability map of the same level (P_2) presented in Equation 4.2 as $W_p x_i^l$. To calculate the local spatial attention to target complex noise, Equation 4.1 is used to perform the additive operation, and then σ_1 is used to reflect ReLu operation $\sigma_1(x) = \max(0, x)$. The output q_{add}^l is then used to generate multiplicative attention of global physical priors based-attention using Equation 4.2.

$$q_{add_mul}^l = \varphi^T(q_{mul}^l) + b_\varphi \quad (4.3)$$

An mlp with $1 \times 1 \times 1$ generated a dual attention tensor with bias b_φ in Equation 4.3.

$$\alpha_i^l = \sigma(q_{add_mul}^l(x_i^l, g_r^l, h_{l-1}; \Theta_i)) \quad (4.4)$$

The output $q_{add_mul}^l$ for each layer l is then normalized using sigmoid operation $\sigma_2(x) = \frac{1}{1+e^{(-x)}}$. This operation generates an attention coefficient (α), which is used to take an element-wise summation with the feature map E_2 , as shown in Figure 4.6 and Equation 4.4.

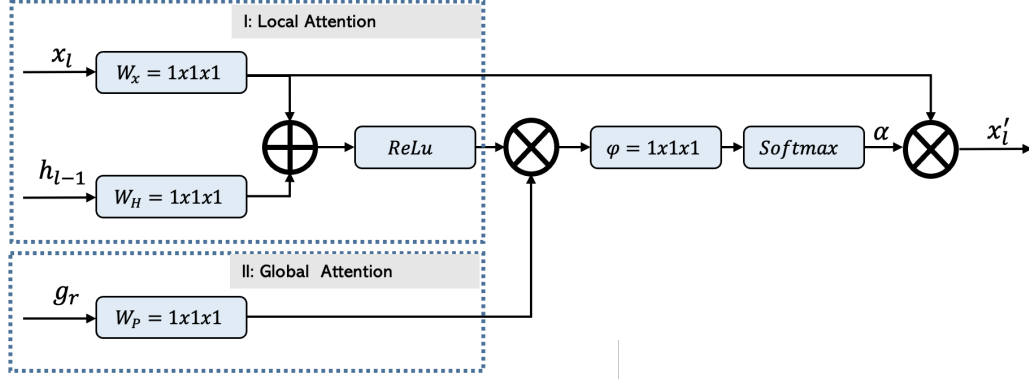


Figure 4.6: Dual attention module calculates attention coefficient based on the local and global feature map. Local feature map from the encoder x_l and the previous layer of the decoder or encoder h_{l-1} and global feature maps comes from physical priors g_r , which are calculated based on the global positioning of each point in the point cloud from the PPE module.

4.2.4 Voxelization

The network uses a voxel grid representation to prepare the 3D point cloud for input into the NSANet architecture. Raw point cloud data is preprocessed onto the voxel grid, and a mean is calculated to represent all the points within each 3D voxel. The network also maintains a projection matrix from the voxel grid to the raw point cloud, making it easy to project labels. Choosing an appropriate voxel size provides a trade-off between efficiency and effectiveness.

4.3 Experiment

The research conducted experiments to validate hypotheses. An ablation study compared NSANet-1, NSANet-2, and NSANet-3 with a baseline, maintaining two input features (occupancy points per voxel, mean-Z) and using a weighted cross-entropy (WCE) loss function. Subsequently, another experiment focused on feature engineering, revealing that a smaller set of features improves performance quantitatively and in terms of efficiency compared to manually designed geometric features.

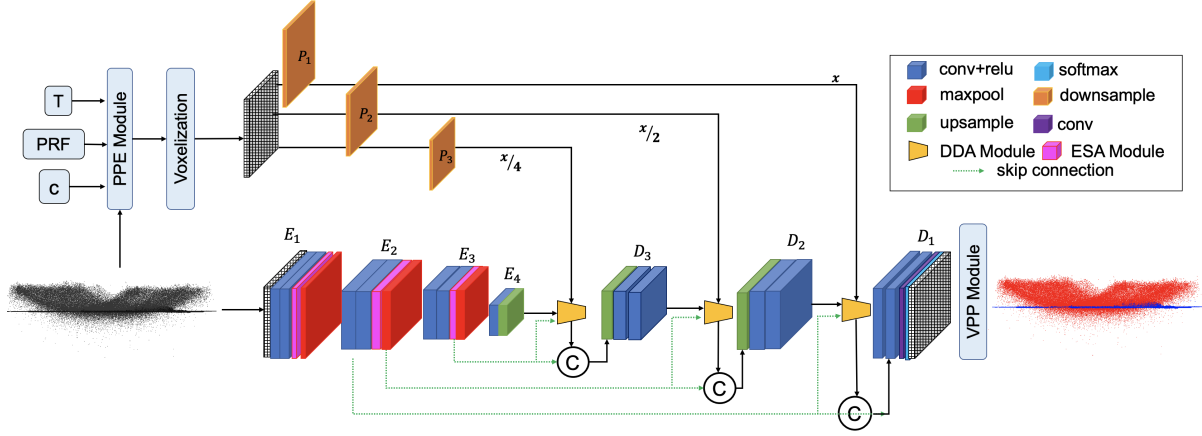


Figure 4.7: NSANet-2 adapted FIT by employing attention module in encoder as well. The first variation contains only local attention in encoder layers called the encoder single attention (ESA) module.

After identifying the optimal NSANet configuration and features input, an ablation study determined the optimal loss and attention configuration. A comparative analysis supported the choice of a voxel-based network for noise-filtering neural networks, with experiments evaluating test sets designed to assess the noise class.

4.3.1 Dataset

Teledyne Galaxy T1000 laser scanner collected data, covering an area of $13km^2$. The collected data was subsequently divided into train and test sets for experiments. Specifically, the first $4km^2$ of the dataset was allocated for testing purposes, while the remaining data was utilized to train the network.

Within the dataset, 14 distinct scenes did not overlap, encompassing over five million points on average, with a density of approximately $30(points/m^2)$. To establish the ground truth, Teledyne Optech employed Terrasolid point cloud processing software to label the data [151]. These labels were generated based on extensive domain knowledge and technical expertise.

The training dataset has two classes, noise, and non-noise, that effectively train the

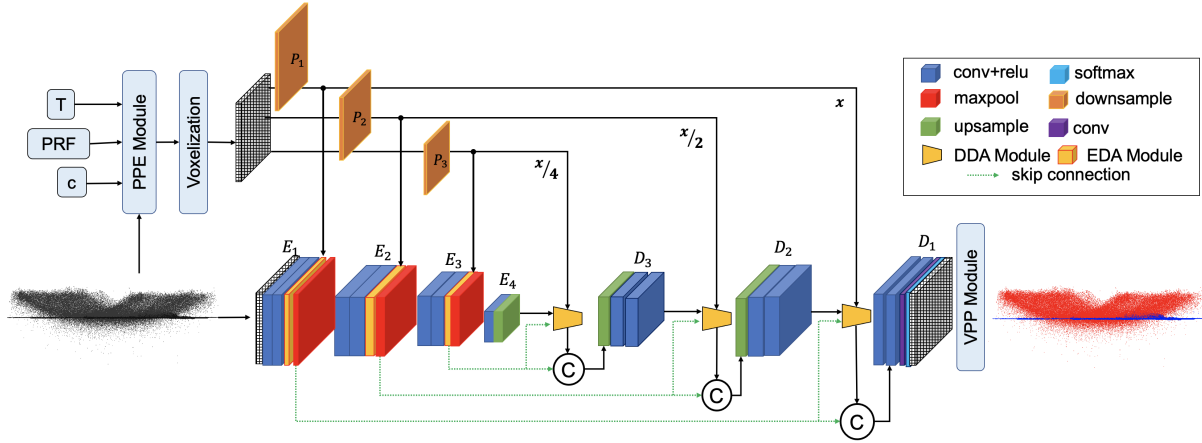


Figure 4.8: NSANet-3 adapted FIT by employing attention module in encoder as well. This variation contains dual attention in encoder layers called encoder dual attention (EDA) module.

NSANet network.

4.3.2 Experimental Configuration

The dataset undergoes a 70%-30% split for training and testing, with the training set further divided into training and validation sets (80-20%). Sub-scenes based on the GPS time of the flight line are generated within each scene and a voxel grid of $64m \times 64m \times 64m$ is created for each sub-scene. Features include mean absolute elevation (MZ) and the number of occupied points (Occ.). NSANet outputs confidence scores for noise and non-noise classes, with the final prediction determined by the highest confidence score. Post-processing techniques enhance accuracy.

NSANet training involves 90 epochs on two GPU RTX 6000 units, utilizing the Adam optimizer with a learning rate of $\alpha = 0.0001$ and a 0.8 decay rate for every 100,000 steps through a staircase learning rate schedule. Inference takes 2 to 3 minutes, and precision, recall, and F1-score assess performance. PointNet and PointNet++ use a $1m^3$ partition size, training for 90 epochs on two GPU RTX 6000 units with a learning rate of $\alpha = 0.001$ and a 0.7 decay rate for every 200,000 steps via a staircase learning rate schedule. SVM

Methods	Fusion	Attention	Features		Loss	Recall	Precision	F1-score
			Occ.	MZ	WCE			
Baseline	X	X	✓	✓	✓	77.9%	91.4%	84.1%
NSANet-1	AET	DDA	✓	✓	✓	87.3%	92.9%	90.0%
NSANet-2	FIT	ESA+DDA	✓	✓	✓	83.3%	93.6%	88.2%
NSANet-3	FIT	ESA+DDA	✓	✓	✓	92.1%	76.6%	85.6%

Table 4.1: Ablation study of modules for NSANet to compare the performance over recall, precision, and F1-score.

training involves voxelization, grouping 100 points using voxels, and extracting features like the number of neighboring points, mean elevation, and geometric features (Omnivariance, Eigentropy, Anisotropy, Planarity, Linearity, Surface Variation, Sphericity, and Verticality). SVM is trained with a soft-margin RBF kernel of gamma=0.05.

4.4 Results and Discussion

4.4.1 NSANet Ablation Study

In the ablation study, research introduced a dual attention module in both the encoder and decoder based on the thesis’s interpretation of FIT and AET. Through a series of experiments, it validated that the implementation of NSANet-1, as depicted in Figure 4.3, which focuses on utilizing attention during the decoding stage, outperformed the variations of NSANet-2 and NSANet-3 illustrated in Figures 4.7 and 4.8. This outcome confirms that the network effectively leverages the principles of attention theory as exemplified by AET, thus experimentally validating its usability. The results presented in Table 4.1 highlight that incorporating attention mechanisms during the feature interpretation stage to filter noise is a superior strategy for addressing this problem. The performance improvement for AET-based dual attention fusion shows a balanced performance with an overall increase of 5.9% in F1-Score over baseline. However, NSANet-2 and NSANet-3, which are variations of

Methods	Features					Fusion	Recall	Precision	F1-score
	Occ.	MZ	INS	NR	R				
NSANet-1	✓	✗	✗	✗	✗	AET	72.5%	85.4%	78.9%
NSANet-1	✓	✓	✗	✗	✗	AET	87.3%	92.9%	90.0%
NSANet-1	✓	✓	✓	✓	✓	AET	87.3%	83.5%	85.5%

Table 4.2: Ablation study on the importance of feature engineering for optimized NSANet-1 over recall, precision, and F1-score; Occ: No. of Occupancy Points, MZ: Mean Elevation, INS: Intensity, R: Range and NR: Number of Returns.

FIT, deliver a high precision of 93.6% and high recall of 92.1% and fail to show a balanced performance across all metrics. Hence, NSANet-1 with AET fusion is selected for the next experiments. It also validates the human vision principle of paying attention to the decoding stage instead of the data extraction stage.

4.4.2 Feature Engineering

The input features for deep learning networks are crucial for learning. Initially, the experiment is configured to select a palette of features, including Occ. and MZ, for the NSANet-1 ablation study. The mean elevation for each voxel was calculated based on the z values of the points within the voxel, enhancing the network’s ability to learn the vertical geometry of the scene (Table 4.2). For the proposed NSANet-1, which incorporates physical priors and spatial attention, training was conducted using a single channel, surpassing the performance of the baseline, achieving a recall of 72.5%, precision of 85.4%, and F1-score of 78.9%. Subsequently, MZ was added to the feature channels alongside (Occ.), which improves the performance on recall by 14.8% and precision by 7.5%, and F1-score by 11.1%. The final results in Table 4.2 highlight the importance of absolute elevation as a discriminating feature, especially due to the distinctive noise distribution in the elevation dimension. Thus, the feature combination of mean elevation and occupancy points will be used in all NSANet-1 experiments.

4.4.3 Loss Functions

The research employed the Cross-Entropy (CE) loss, which is commonly used for segmentation problems.

$$CE = [-\sum_{i=1}^C y_i \log(p_i)] \quad (4.5)$$

In equation 4.5, y represents the true label, p is the predicted probability for the positive class in binary classification, and p_i is the predicted probability for class i in multiclass classification. C is the number of classes.

$$WCE = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C w_c \cdot y_{i,c} \cdot \log(p_{i,c}) \quad (4.6)$$

where, N represents the total number of samples, C denotes the number of classes, w_c signifies the weight assigned to class c , $y_{i,c}$ is the true label for sample i and class c , and $p_{i,c}$ is the predicted probability for sample i and class c . However, this specific problem necessitated using a WCE loss as shown in equation 4.6. This was due to the significant class imbalance in the dataset, making it crucial to assign appropriate weights. The introduction of class weights proved highly beneficial and led to a substantial improvement in performance, as demonstrated in Table 4.3. The results show the performance for the noise class was best when WCE is used with the recall of 87.3%, precision of 92.9%, and 90.0% F1-score. To explore alternative optimal loss functions, a few tests were conducted using Focal Loss (FL) as shown in equation 4.7 to assess its impact on performance.

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (4.7)$$

Where p_t is the predicted probability of the true class, α_t is the focal loss balancing factor, and γ is the focusing parameter. Additionally, the specific distribution and characteristics of the noise and non-noise classes in data have not been effectively addressed by focal loss exhibiting recall decrease by 11.1%, precision by 2.2%, and F1-score by 6.6%. Therefore, WCE loss significantly outperformed FL, confirming its superiority in this problem domain.

Methods	Loss			Fusion	Recall	Precision	F1-score
	CE	FL	WCE				
NSANet-1	✓	✗	✗	AET	72.1%	90.7%	81.4%
NSANet-1	✗	✓	✗	AET	76.2%	90.7%	83.4%
NSANet-1	✗	✗	✓	AET	87.3%	92.9%	90.0%

Table 4.3: Ablation study on the importance of loss function over Recall, Precision, F1-score; CE: Cross Entropy Loss, FL: Focal Loss, and WCE: Weighted Cross-Entropy

Methods	DDA-Module		Fusion	Recall	Precision	F1-score
	local	global				
NSANet-1	✓	✗	AET	80.5%	92.0%	86.2%
NSANet-1	✗	✓	AET	85.2%	91.6%	88.4%
NSANet-1	✓	✓	AET	87.3%	92.9%	90.0%

Table 4.4: Ablation Study of global, local and dual attention modules about DDA-module in Figure no. 4.6 to compare performance advantage over Recall, Precision, and F1-score.

Methods	input	Recall	Precision	F1-score
SVM (benchmark)	segment	46.7%	54.0%	50.1%
DAE [152]	voxel	53.5%	99.9%	69.7%
PointNet [7]	point	68.3%	98.2%	80.5%
PointNet++ [6]	point	57.6%	90.4%	74.0%
3D U-Net [9]	voxel	77.9%	91.4%	84.1%
Att-U-Net [56]	voxel	80.5%	92.0%	86.2%
Multiview 2D-U-Net	views	80.0%	92.9%	85.9%
NSANet-1	voxel	87.3%	92.9%	90.0%

Table 4.5: Comparison of NSANet for noise filtering with SVM, Denoising Autoencoder, 3D U-Net, PointNet++, and PointNet over Recall, Precision, and F1-score for noise class.

4.4.4 Attention Ablation Study

Furthermore, another ablation study assessed the significance of global, local, and dual attention modules. The findings presented in Table 4.4 demonstrate that integrating physical priors yields a relatively higher performance improvement of 85.2% recall, 91.6% precision, and 88.4% F1-score compared to using only local attention, which shows performance decrease of 4.71% recall, 0.4% precision, and 2.2% F1-score. However, the dual attention module plays a critical role in effectively eliminating noise from the point cloud, facilitating comprehensive scene understanding with an F1-score of 90.0%. These experiments have effectively validated the importance of physics-based deep learning approaches for computer vision tasks.

4.4.5 Comparative Study

Most importantly, a comparative study conducted based on an extensive literature review, considering various models, including Support Vector Machines (SVM), Denoising Autoencoder (DAE), PointNet, PointNet++, 3D U-Net, Attention 3D U-Net, and Multi-view-based 2D U-Net. The results are in Table no. 4.5 of each model and are summarized as follows:

- SVM was trained to balance preserving objects of interest and eliminating noise points. It achieved a 50.1% F1-score for the noise class.
- Denoising autoencoder (DAE), which focuses on denoising reconstruction, outperformed SVM with a relatively higher F1-score of 69.7% for the noise class.
- PointNet [7], a pioneering network that directly processes 3D point cloud data, performed better than DAE but struggled with complex noise and lacked global context, multi-resolution, and local contextual features and shows the F1-score of 80.5%.
- PointNet++ [6], was also tested but did not yield satisfactory results (F1-score 74.0%).
- The 3D U-Net model with WCE loss was selected as a baseline. It utilized a multi-resolution encoder-decoder structure to incorporate more context and exhibited exponentially improved performance on the noise class by exhibiting a recall of 77.9% and an F1-score of 84.1%.
- It further compared the performance of Attention 3D U-Net and Multi-view-based 2D U-Net. The 2D U-Net, which took two different views as input, showed comparable performance to the Attention 3D U-Net by exhibiting the F1-score of 85.9% and 86.2%. The 2D U-Net has a limitation as the view selection can be tedious and dependent on the dataset. However, Attention 3D U-Net exhibited aggressive noise classification, misclassifying non-noise as noise, while the 3D U-Net with cross-entropy loss demonstrated generally poor performance.

The proposed system, which incorporates global and local context through attention mechanisms embedded in the network, is concluded to work effectively. This approach led to significant performance enhancements, as demonstrated in Table 4.4. Unlike other networks with better precision but lower recall and F1-scores, NSANet-1 achieved a balanced performance across all metrics. SVM and 3D UNet have been used for noise filtering. Still, comparative studies have shown more conservative performance, which was why NSANet-1

Methods	VPP	Fusion	Features		Recall	Precision	F1-score
			Occ.	MZ			
NSANet-1	✗	AET	✓	✓	87.30%	92.90%	90.0%
NSANet-1	✓	AET	✓	✓	98.90%	96.10%	97.5%

Table 4.6: Ablation Study based on voxel post-processing module and compared performance over Recall, Precision, and F1-score.

was proposed to improve the performance of filtering noise points without compromising the objects of interest to a greater extent.

4.4.6 Error Analysis

The research has yielded remarkable advancements in noise filtering, surpassing the state-of-the-art (SOTA) methods, as illustrated in Figures 4.9 and 4.10. It is important to highlight that this network draws inspiration from semantic segmentation and denoising research. Filtering noise from large-scale 3D point cloud data captured by airborne LiDAR remains a challenging problem with limited existing solutions.

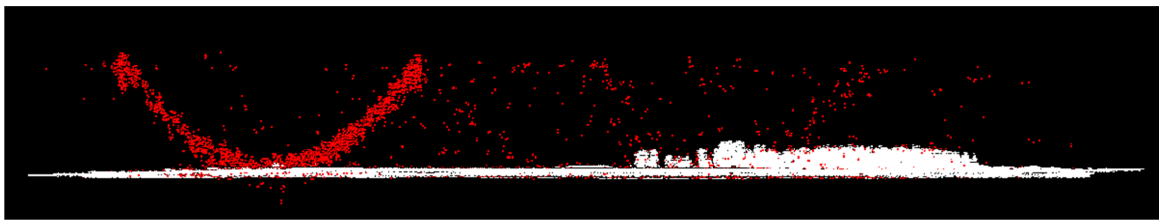
However, there are certain scenarios where the proposed network encounters difficulties in filtering noise, particularly in cases where the distribution of objects of interest closely resembles noise or overlaps with it. These challenging instances serve as hard examples and contribute to a performance gap. To address this, a voxel-based post-processing (VPP) module has been incorporated into the final design, which enhances the recall performance by 10% by considering the confidence of neighboring voxels as shown in Table 4.6.

4.5 Summary

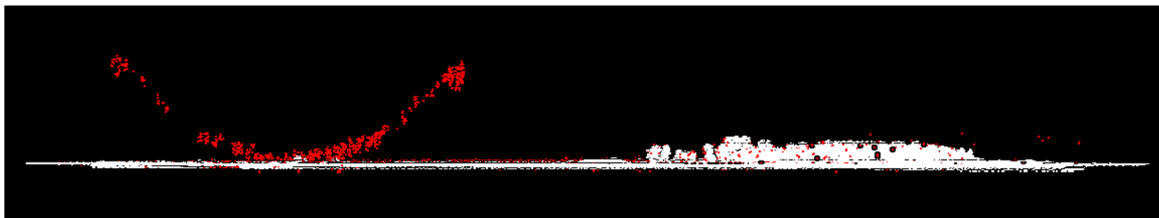
This work introduced an innovative dual-attention NSANet model for noise filtering in large-scale point clouds from airborne LiDAR scanners. The model effectively addresses challenges like overlapping objects, systematic global patterns, and sparse noise by leveraging physical



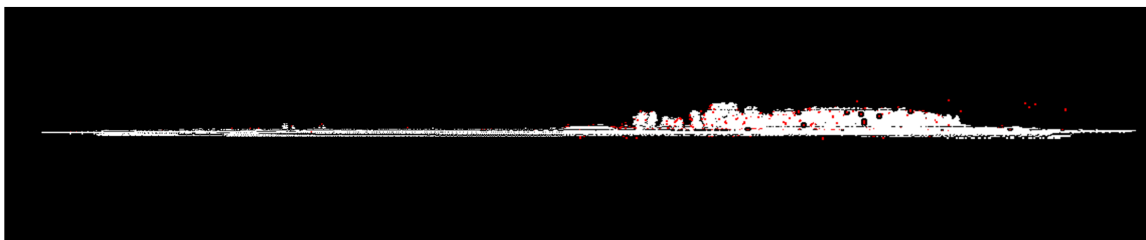
Groundtruth



Baseline



NSANet-1



NSANet-1 + VPP

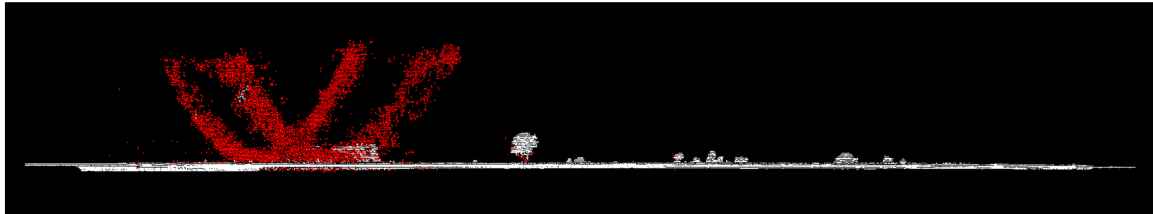
Figure 4.9: 3D scene with sparse noise spread across scene shows four different scenes comparing the SOTA performance of baseline with weighted cross-entropy loss and NSANet-1 variation employed AET and weighted cross-entropy loss. Results reflect NSANet's superior performance. Red shows noise points and white reflects non-noise points.



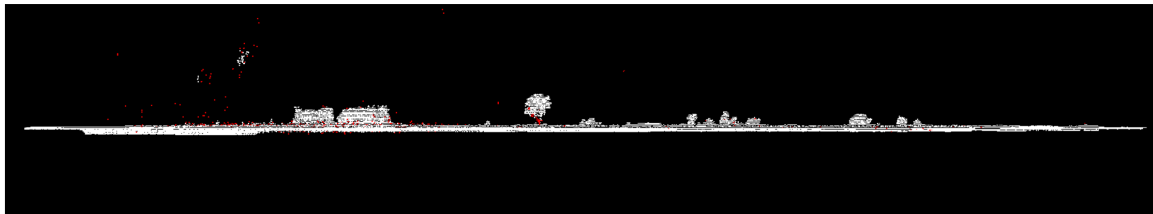
Groundtruth



Baseline



NSANet-1



NSANet-1 + VPP

Figure 4.10: 3D scene with dense noise pattern shows four different scenes comparing the SOTA performance of baseline with weighted cross-entropy loss and NSANet-1 variation employed AET and weighted cross-entropy loss. Results reflect NSANet’s superior performance. Red shows noise points and white reflects non-noise points.

priors and local spatial attention. Experimental studies confirm the model’s ability to enhance scene understanding by incorporating global context. The results validate the two-stage-based attentional engagement theory, demonstrating improved noise filtering performance with integrated physical priors. NSANet consistently outperforms other models across evaluation metrics, particularly in handling airborne LiDAR data. Future chapters will extend the focus to scene understanding for utility corridors using neural networks and further advance physical priors learning.

Chapter 5

Multiscale Deep Learning for Utility Network Segmentation

This chapter introduces the second stage of the scene understanding pipeline, discussing three approaches for utility scene understanding. The framework focuses on utility semantic segmentation, emphasizing spatial layout consistency as an external cue to enhance deep learning-based semantic segmentation. The aging infrastructure of utility networks requires reliable computer vision systems for inspection and minimizes human intervention. The airborne laser terrain mapping (ALTM) system is favored for its rapid data collection, foliage penetration, and accurate ranging. However, classifying dense and irregular point clouds from ALTM remains a bottleneck, hindering efficiency and accuracy. To address this challenge, this research proposes a Semantic Utility Network (SUNet), a deep convolutional neural network (DCNN) for voxel-based semantic segmentation of ALTM point clouds. SUNet enforces spatial layout consistency on segmentation results from voxel-based and regional approaches. The DCNN combines hierarchical context for spatial layout through multiscale and multimodal fusion. This methodology extends to Fusion-SUNet and Pan-SUNet, as shown in Figure 5.1. Fusion-SUNet focuses on incorporating spatial layout consistency at a deeper level for utility network segmentation. It fuses features to impose spatial layout consistency, classifying dense point clouds from ALTM. The architecture achieves voxel-based

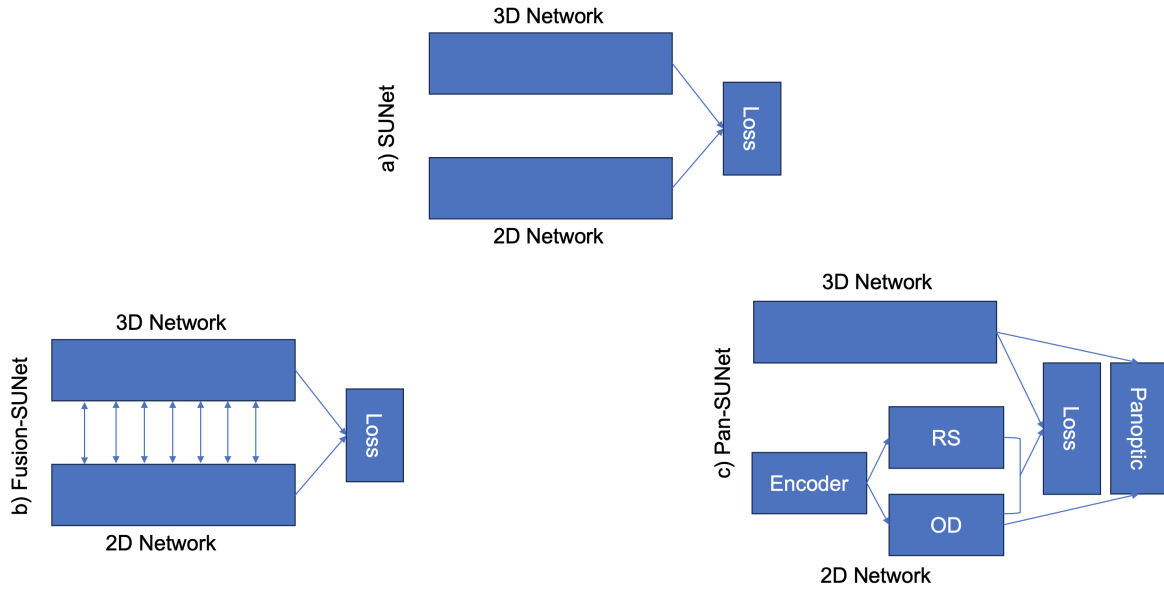


Figure 5.1: The three proposed utility scene understanding approaches and the fusion modules. a) SUNet contributes loss-based late fusion, b) Fusion-SUNet builds on SUNet and proposed deep fusion module, and c) Pan-SUNet builds on SUNet’s loss-based late fusion using regional segmentation (RS) and incorporates 2D Object detection (OD) for panoptic fusion.

semantic segmentation at multi-resolution in three dimensions (3D) and distinguishes corridor regions from non-corridors in two dimensions (2D). Spatial layout consistency is ensured using spatial and channel-based attention, fusing regional and 3D segmentation features. Pan-SUNet performs 3D semantic segmentation and object detection using 2D regional segmentation and object detection through the power panoptic fusion module. It enhances 3D semantic segmentation and 3D object detection, benefiting the utility community in segmenting corridors. PanSUNet predicts 2D regional labels, differentiating utility and corridor regions from non-corridor areas, ensuring spatial layout consistency in 3D network predictions.

5.1 Overview

The section discusses the importance of conducting safe, efficient, and accurate inspections for utility network management to ensure grid stability and resilience [130, 153, 16, 18, 142, 143]. While traditional methods involve ground crews for tasks like managing vegetation encroachment, there is a growing trend in using unmanned aerial vehicles (UAVs) to reduce data collection costs [153, 16]. However, challenges such as strict regulations, limited operation time, and spatial coverage hinder the full use of UAVs. Airborne Laser Terrain Mapping (ALTM) is a central data acquisition platform, providing dense point clouds with survey-grade accuracy for utility infrastructure representation [18]. Despite these advancements, the manual labeling point clouds with semantic attributes remains a laborious and error-prone task, prompting the need for automation.

The urgency to automate maintenance processes is emphasized, particularly to prevent vegetation-related electricity outages in aging and lengthy infrastructures [154]. Regulatory obligations, such as those set by the North American Electric Reliability Corporation (NERC), mandate utility transmission companies to comply with standards and avoid unforeseen disasters and vegetation encroachment [29, 30]. This has led to research efforts in exploring traditional and deep learning methods, with Deep Convolutional Neural Networks (DCNN)

emerging as a powerful tool in computer vision [5]. The thesis highlights the success of DCNNs in semantic segmentation using point clouds but notes a gap in utilizing spatial relationships for utility corridors and embedding spatial layouts for multiscale contexts. As a response to this limitation, the thesis proposes a network that addresses hierarchical spatial consistency, aiming for generalizability to standard layout segmentation problems [7, 6, 8, 63, 5, 65, 64, 66]. This thesis thoroughly investigates the utility corridor dataset. Firstly, the study unraveled spatial layout consistency. It identifies the hierarchy between regions (corridor and non-corridor) and object classes such as ground, towers, powerlines, and vegetation as shown in Figure 5.2 [130]. Secondly, it utilizes the hierarchical spatial layout consistency by combining multimodal input. Finally, it explores the science of human vision and information extraction for decision-making as a hierarchy. Spatial cognition is one of the key concepts driving this research in semantic utility networks. It involves high-level cognitive processes in the human brain that influence decisions regarding the representation of location, orientation, and action in relation to oneself, others, objects, or the environment. The spatial layout of objects significantly affects the human visual experience as they are organized into regions. Spatial layout consistency plays a crucial role in this cognition, addressing questions such as the object’s location, neighboring elements, and identity. By understanding these representations, one can gain insights into a 3D scene based on its layout. Consequently, research leverages this concept to enhance understanding of the utility transmission layout and improve scene comprehension [28]. This chapter will detail all three approaches individually and elaborate on the modules.

5.2 SUNet-Semantic Utility Network

SUNet is a multimodal and multi-resolution network designed to capture the spatial consistency between the layout and objects of interest. It comprises two networks: a two-dimensional regional segmentation network [52] and a three-dimensional segmentation network. The 2D regional prediction network predicts the regional labels, distinguishing corridor regions

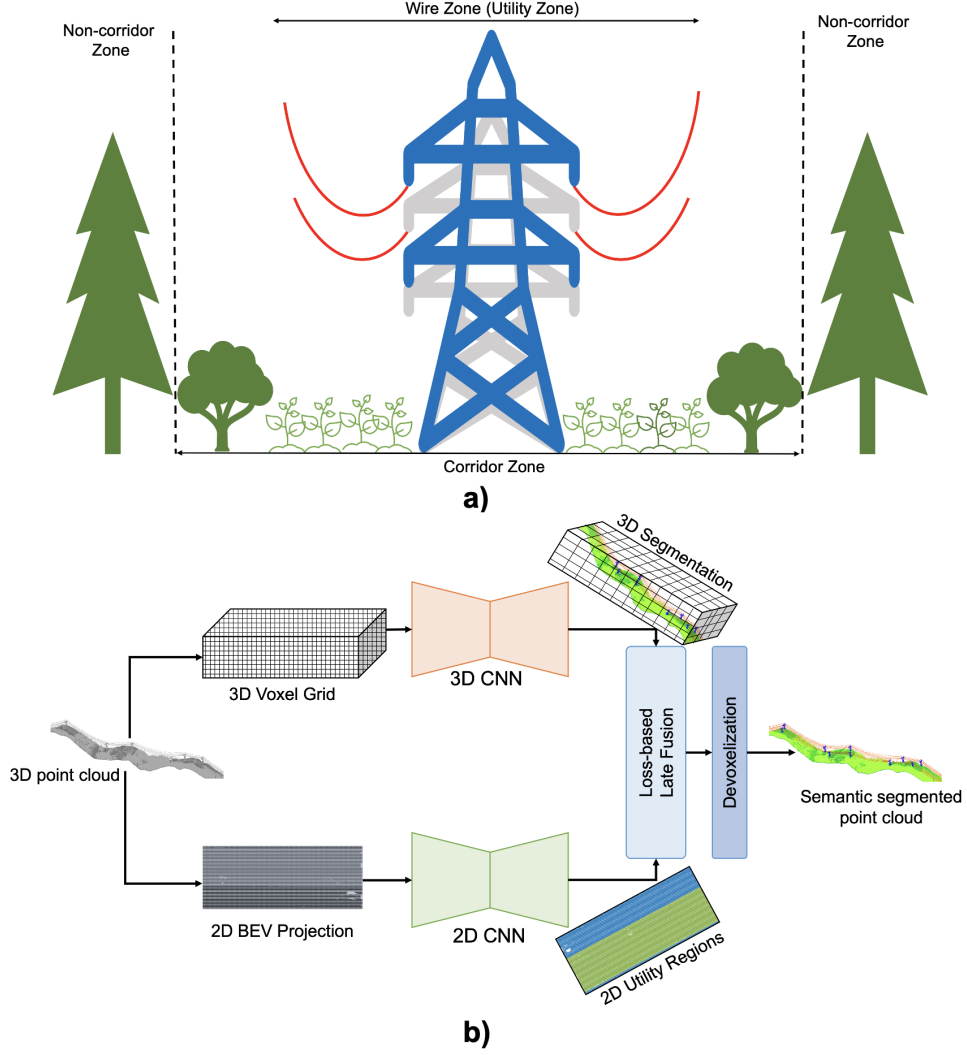


Figure 5.2: SUNet is a multi-dimensional and multi-resolution network that imposes the spatial layout consistency (a) through 2D utility regions on the outcomes of 3D segmentation network via loss-based late fusion (b).

from non-corridor. This prediction network imposes constraints on the predictions of the 3D network through loss-based late fusion. By combining the predictions of both networks, SUNet achieves improved performance in capturing the spatial layout consistency and accurately classifying objects of interest, as shown in Figure 5.2.

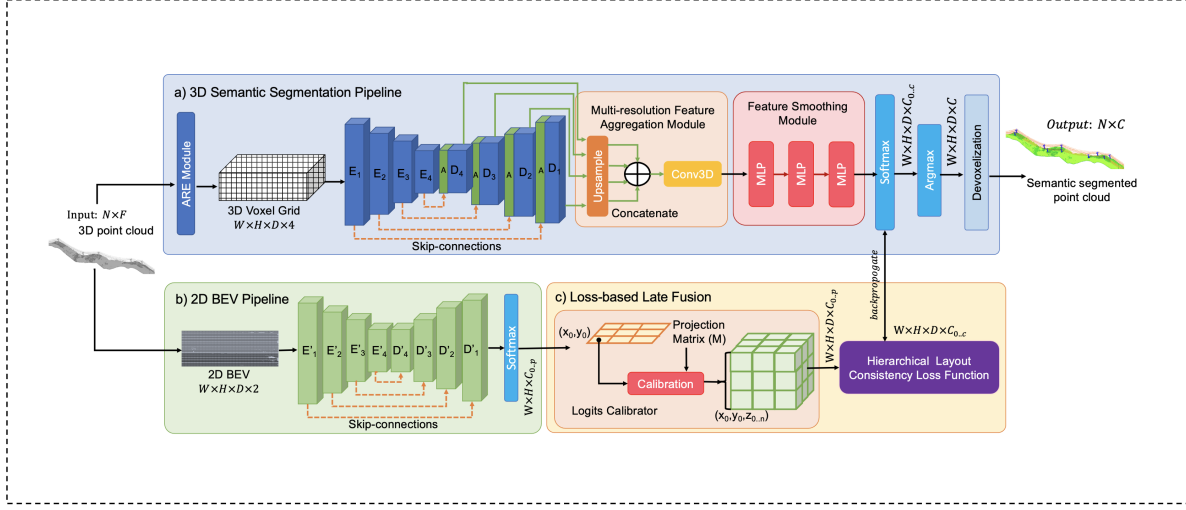


Figure 5.3: The overall architecture of SUNet: Point cloud is preprocessed into voxel grid and BEV. These are separately processed by a multi-resolution 3D semantic segmentation pipeline and 2D BEV pipeline for regional prediction to impose the spatial layout consistency on 3D objects through loss-based late fusion.

5.2.1 3D Semantic Segmentation Pipeline

The overall network architecture of SUNet is depicted in Figure 5.3. The 3D segmentation network follows a U-shaped multi-resolution encoder-decoder structure. It operates on a 3D voxel grid to extract features and predict semantic labels for object classes [56]. The network consists of four levels of feature extraction, denoted as E_1 , E_2 , E_3 , and E_4 , as shown in Figure 5.3. Each level constructs an output feature map D_l through a series of operations. These operations include additive attention applied on the previous output feature map D_{l-1} and the current feature map E_l , followed by two $3 \times 3 \times 3$ convolution operations, batch normalization, and relu activation. The result is four output feature maps $\{D_1, D_2, D_3, D_4\}$ with dimensions $\frac{H}{2^l} \times \frac{W}{2^l} \times \frac{D}{2^l} \times 32^l$ where (H: height, W: width, and D: Depth) for output D of decoder layer at level l . These feature maps represent the deep multi-resolution output, which provides a multi-receptive field for segmenting objects and areas of different sizes.

The feature maps D_1 , D_2 , D_3 , and D_4 are then passed through a multi-resolution feature

aggregation module to incorporate knowledge and context from all levels. This module helps capture comprehensive information from different scales and resolutions, enhancing the overall segmentation performance.

The final part of the pipeline is the feature smoothing module, which aims to eliminate noisy and cluttered predictions. It refines the predictions and delivers a confidence score for each class. This score is then utilized by the loss-based late fusion module, which refines and constrains the predictions using spatial layout consistency through 2D global scale regional predictions. By incorporating spatial layout consistency, the network learns to capture the relationships between the layout and object classes, leading to improved segmentation accuracy. The loss-based late fusion module also allows the network to back-propagate the loss and refine the learning of deep features.

Additive Attention Module

In the multi-resolution encoder-decoder architecture of SUNet, downsampling operations can sometimes lead to an overpopulation of false positives, especially for small objects that exhibit significant shape variance. An additive attention gate is incorporated into the network to address this issue. This attention gate module helps to alleviate localization problems and reduce misclassification by selectively attending to relevant information [26, 56].

$$q_{att}^l = \varphi(\sigma_1(E_l + D_{l-1})) + b_\varphi \quad (5.1)$$

where σ_1 is used to reflect ReLu operation $\sigma_1(x) = \max(0, x)$ which takes input from the previous layer of the decoder, D_{l-1} , and the encoder feature map of the same level, E_l as shown in Equation 5.1.

$$\alpha_i^l = \sigma(q_{att}^l(E_l, D_{l-1}; \Theta_{att})) \quad (5.2)$$

The additive attention q_{att}^l in Equation 5.2 learns multiple attention coefficients for each class using $\sigma_2(x) = \frac{1}{1+e^{(-x)}}$ and produces activated feature maps that are pruned to suppress irrelevant information. It used Θ_{att} as paramaterized function for linear transformation using

mlp $1 \times 1 \times 1$.

$$E_l = E_l \cdot \alpha_i^l \quad (5.3)$$

Then, Equation 5.3 uses attention coefficients, denoted as $\alpha_i \in [0, 1]$, and i is a spatial dimension, which is then element-wise summed with the feature map E_l .

Using the attention coefficients, the attention gate selectively highlights important features and suppresses less relevant information, improving small objects' localization and classification accuracy. This mechanism allows the network to focus on the most discriminative and informative features, mitigating the false positives and enhancing the overall segmentation performance.

Multi-resolution Feature Aggregation Module

The proposed module performs feature map aggregation by upsampling and concatenating the feature maps from $\{ D_1, D_2, D_3, D_4 \}$. This aggregation process combines the information from multiple levels of the decoder. To refine and consolidate the aggregated features, a $1 \times 1 \times 1$ convolution operation is applied.

$$Y'_l = f(D_l, \beta) \quad (5.4)$$

where Y'_l is upsampled tensor at level l . f is a function for upsampling, transpose 3D convolution which takes β as an upsampling factor to give all decoder feature maps of the same size for aggregation as shown in Equation 5.4.

$$Y_{agg} = Y'_1 \parallel Y'_2 \parallel \dots \parallel Y'_l \quad (5.5)$$

Feature aggregation operation or concatenation $\parallel$ followed by $1 \times 1 \times 1$ helps generate a feature map Y'_{agg} that minimizes the errors introduced during the upsampling process of the decoder, particularly for the prediction of minority classes. This module enhances the overall segmentation performance by effectively integrating the multi-resolution information and ensures more accurate predictions, even for small and less represented object classes.

Feature Smoothing Module

The optional module in the architecture consists of three multi-layer perceptron (MLP) operations. Its purpose is to smooth the feature maps by evenly distributing the mean and variances of the features.

$$Y_{fs}^n = (W^{(n)}Y_{agg}^{(n-1)} + b^{(n)}) \quad (5.6)$$

where n is the number of MLP layers in the feature smoothing model, which takes input from the previous layer $n - 1$ and calculates its weight W and bias b as shown in Equation 5.6. By incorporating this module, the aim is to reduce misclassifications by a minimal margin. The smoothing operation helps to refine the feature representations and improve the overall consistency of the predictions, leading to more accurate and reliable segmentation results. Lastly, it performs softmax operation σ to get the probability of prediction against each class c using the following Equation 5.7.

$$\sigma(Y_{fs}^i) = \frac{e^{Y_{fs}^i}}{\sum_{j=1}^C e^{Y_{fs}^j}} \quad (5.7)$$

where i is spatial dimension and j is one of the semantic segmentation classes C .

5.2.2 2D BEV Pipeline

Integrating a larger receptive field and global context into a network poses a significant challenge for semantic segmentation. Understanding the spatial layout and global correlation of objects is crucial for humans to make accurate judgments about scene semantics. However, the 3D segmentation network is limited to capturing local context and cannot encode the scene’s coarser details and global context.

The 2D Bird’s Eye View (BEV) pipeline, which incorporates spatial layout and global correlation information through a loss-based late fusion module that acts as a multimodal fusion, is introduced to address this limitation. The 2D network, depicted in Figure 5.3 as a regional semantics block [52], is a multi-resolution encoder-decoder architecture. It consists

of four feature maps, namely $\{E'_1, E'_2, E'_3, E'_4\}$, which are used to construct four output maps $\{D'_1, D'_2, D'_3, D'_4\}$.

This 2D network takes the BEV representation of a complete 3D point cloud scene as input and predicts the probability tensor of regional classes with a shape of $W \times H \times C_p$. By leveraging the BEV representation, this network can capture the regional semantics of the scene and incorporate the spatial layout and global context information for more informed and accurate predictions.

5.2.3 Voxelization and BEV Projection

The input representation for the segmentation network is a voxel grid, which involves preprocessing the raw point cloud data. The raw point cloud is organized onto the voxel grid, and the mean value is calculated to represent all the points within each 3D voxel. Additionally, the network maintains a projection matrix that maps the voxel grid to the raw point cloud, facilitating easy label projection.

Using a voxel grid as input balances efficiency and effectiveness, with the voxel size selection influencing this trade-off. It allows us to efficiently process and analyze the point cloud data while capturing essential spatial information.

On the other hand, the Bird's Eye View (BEV) is a 2D representation of the 3D point cloud scene. In the 2D BEV pipeline, XY coordinates of the 3D scene are projected onto a 2D plane, where each pixel represents the points within the corresponding location. This XY projection provides an optimal BEV representation for extracting the global context required for regional and object predictions.

A projection matrix, denoted as M , is created to enable fusion and effective utilization of spatial layout consistency, which establishes compatibility between the BEV and the 3D voxel grid feature spaces. This projection matrix allows us to seamlessly combine the 2D and 3D predictions, facilitating a more comprehensive understanding of the scene and enhancing the accuracy of the segmentation results.

5.2.4 Absolute and Relative Elevation Module

The data analysis showed non-homogeneous areas where points belonging to the same classes could be on different elevations. Hence, local pillar-based normalization is calculated from Equation 5.9, which uses the minimum elevation of each column and the maximum elevation of the scene for normalization and calculates over the whole of each data point i of a whole point cloud of N points contains three dimensions $\mathbb{R}^3 = \{(x, y, z) \mid x, y, z \in \mathbb{R}\}$.

Global Normalization:

$$Absolute\ Elevation_i = 1 + \frac{z_i}{z_{max_global}} \quad (5.8)$$

where z is third dimension of 3D point cloud $\mathbb{R}^3$ and i represent the i th data point. z_{max_global} is maximum value of elevation or z dimension in the 3D point cloud in dataset.

Local Normalization:

$$Relative\ Elevation_i = \frac{(z_i - z_{min_local})}{(z_{max_scene} - z_{min_local})} \quad (5.9)$$

where z is third dimension of 3D point cloud $\mathbb{R}^3$ and i represent the i th data point, z_{min_local} is minimum value elevation in a pillar of size of $1m^2$ in the 3D point cloud and z_{max_scene} in maximum elevation of particular scene in a dataset.

5.2.5 Loss-based Late Fusion

This is a key module that is further divided into two sub-modules. The following sections will discuss this in detail.

Logits Calibrator

This module is a straightforward calibration that takes the projection matrix between 3D voxels and 2D BEV and converts the 2D logits for regional classes of shape $W \times H \times C_p$ to 3D representation $W \times H \times D \times C_p$ by exploring the one-to-many relationship as shown in Equation 5.10.

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} \quad (5.10)$$

X and Y are the original 3D voxel coordinates of size W and H. Z is the coordinate along the depth D axis (not considered in the projection). After the orthographic projection, x and y are the 2D grid coordinates of sizes W and H.

Hierarchical Layout Consistency Loss Function

The proposed loss function plays a crucial role in the late fusion module, allowing us to leverage the hierarchy of object classes in different regions (corridor and non-corridor) to penalize misclassifications. It serves as a key element for fusing context from both local and global scales. The loss function takes the 3D predictions generated by the 3D pipeline. It refines them by incorporating global spatial layout information extracted from the 2D pipeline, bridging the gap regarding the larger receptive field. The mathematical representation of the loss function can be seen in Equation 5.11

$$HLC = \frac{1}{M} \sum_{p=1}^P \sum_{c=1}^C \sum_{m=1}^M (w_p w_c) \times (y_m^{P_c} \times (\log(h_{\Theta}(x_m, c, p)))) \quad (5.11)$$

Where M is the number of samples, C is the number of semantic segmentation classes, w_c is the weight for the semantic segmentation class c , P is the number of regional segmentation classes, w_p is the weight for the regional segmentation class p , $y_m^{P_c}$ is the true label for sample m in semantic segmentation class c and regional segmentation class p , and $\log(h_{\Theta}(x_m, c, p))$ is the predicted probability for sample m in semantic segmentation class c and regional segmentation class p . To achieve hierarchical layout consistency, the loss function HLC evaluates the correlation between the target class $y_m^{P_c}$ and the predicted class for x_m , weighted by w_p and w_c as shown in Equations 5.12 and 5.13 respectively.

$$w_p = \frac{N}{P \cdot n_p} \quad (5.12)$$

Where P is the total number of regional segmentation classes, n_p represents the number of samples belonging to class p , and N is the total number of samples in the dataset.

$$w_c = \frac{N}{C \cdot n_c} \quad (5.13)$$

Where C is the total number of semantic segmentation classes, n_c represents the number of samples belonging to class c , and N is the total number of samples in the dataset.

This weighting helps reduce the loss for improved predictions.

Incorporating the global spatial layout information into the loss function can encourage more accurate predictions by penalizing misclassifications and promoting better alignment between the predicted and target classes. This enables the network to utilize local and global context for semantic segmentation tasks effectively.

5.3 Fusion-SUNet

As discussed in the previous section, Fusion-SUNet is an extension of SUNet. Fusion-SUNet is a multi-scale and multi-resolution network that incorporates spatial layout consistency between the layout and objects of a scene on a deep feature level. It comprises two networks: a 2D regional prediction network [52] that constrains the predictions of a 3D network through multimodal and multiscale deep feature fusion based module and loss-based late fusion as shown Figure 5.4. The network architecture enables multiscale and multimodal fusion at different depths, allowing it to incorporate regional information from the 2D network to improve the 3D network’s prediction accuracy [155, 156]. The next section will only discuss the novel multimodal Attention Fusion (A-Fusion) module as the baseline of SUNet remains the same.

5.3.1 Fusion

The fusion pipeline has two key modules: Feature-based middle fusion and loss-based late fusion. SUNet has already proved the importance of loss-based late fusion. This section will introduce feature-based fusion.

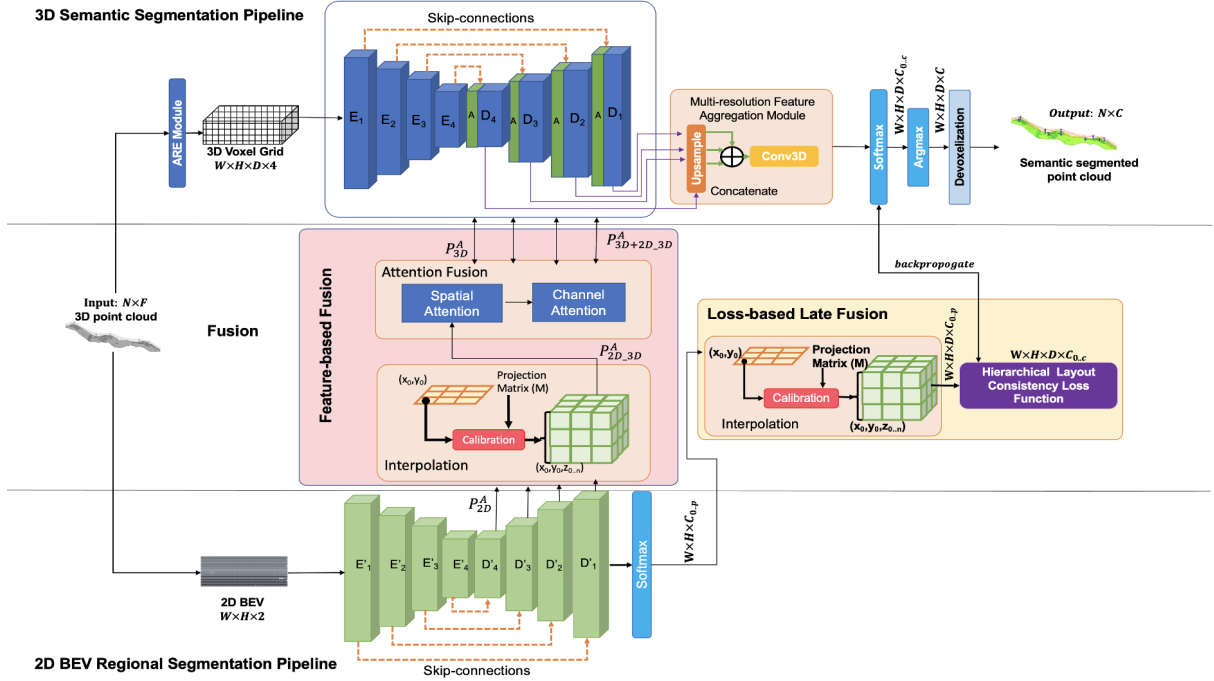


Figure 5.4: The overall architecture of Fusion-SUNet: Point cloud is preprocessed into voxel grid and BEV. These are separately processed by a multi-resolution 3D semantic segmentation pipeline and 2D BEV pipeline for regional prediction to impose the spatial layout consistency on 3D objects through feature-based and loss-based late fusion.

Feature Interpolation Module

It is a simple module that takes the projection matrix between 3D voxels and 2D BEV and converts the 2D features map for the regional network's layers of shape $\frac{H}{2l} \times \frac{W}{2l} \times 32l$ into a 3D representation of shape $\frac{H}{2l} \times \frac{W}{2l} \times \frac{D}{2l} \times 32l$ represented as P_{2D_3D} by exploring the one-to-many relationship between the two representations.

A-Fusion Module

The module is divided into two attention modules: spatial attention and channel-based attention, as shown in Figure 5.5. The spatial attention mechanism generates an attention-focused feature map by interpolating a 2D regional network to a 3D feature map and concatenating it with the 3D feature map from the additive attention map output to provide a feature map for a decoder level in the 3D network. The spatial attention mechanism imposes layout constraints on the features to facilitate learning objects of interest by taking advantage of spatial consistency at deeper levels.

- The A-Fusion module takes input from the feature interpolation module $P_{2D_3D_l}$ and the output of the additive attention module from the 3D segmentation pipeline $P_{3D_l}^A$.
- The attention fusion module performs spatial attention using both feature maps, using element-wise summation and ReLU activation σ_1 as the final step, followed by element-wise multiplication with the additive attention map output $P_{3D_l}^A$ as shown in Equation 5.14.

$$P_{2D_3D_l}^A = \sigma_1(P_{2D_3D_l} \oplus P_{3D_l}^A) \quad (5.14)$$

- The channel attention map is generated from the interpolated 2D to 3D regional feature map $P_{2D_3D_l}$ using max-pooling *max* and average *avg* pooling operations from Equations 5.15 and 5.16, followed by a shared MLP and element-wise summation to generate a 1D channel attention map of size $1 \times 1 \times 1 \times 32l$ after a sigmoid operation

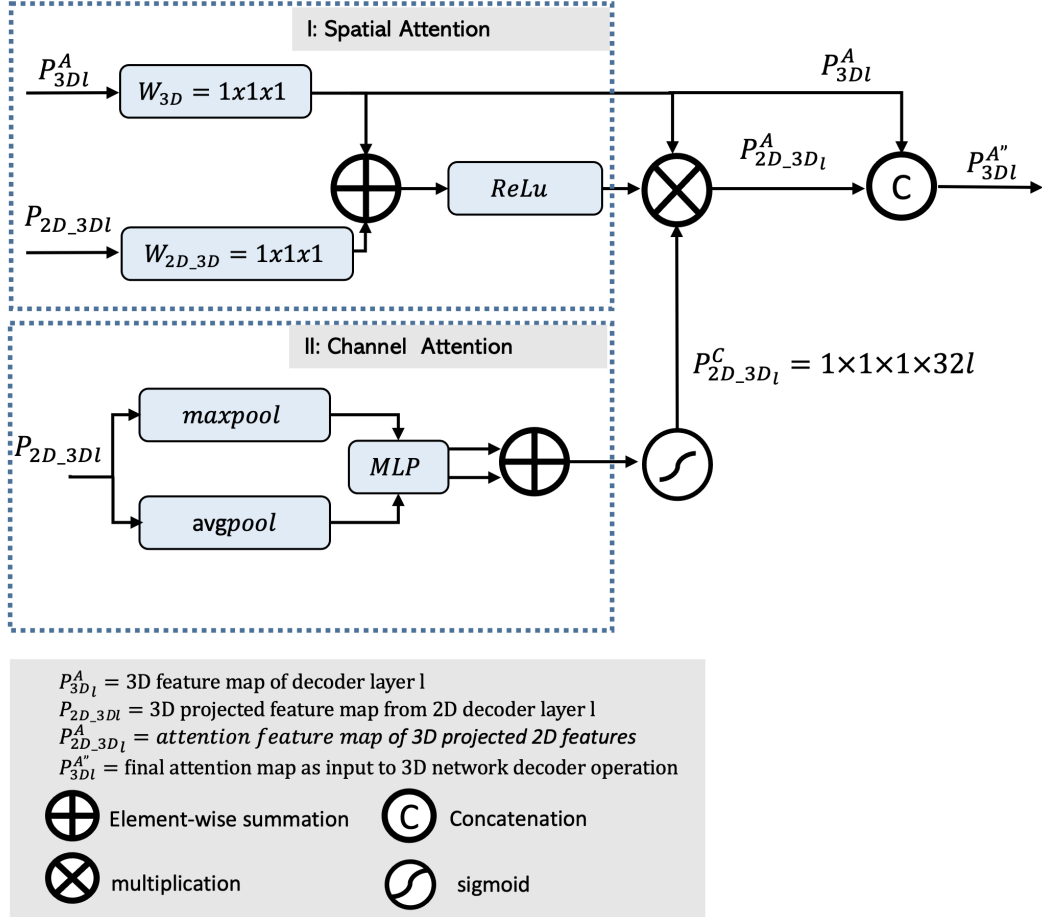


Figure 5.5: A-Fusion module to impose spatial layout consistency through the fusion of features, also known as middle fusion.

$\sigma_2(x) = \frac{1}{1+e^{-x}}$ as shown in Equation 5.17.

$$P_{2D_3D_{max_l}} = mlp(\max(P_{2D_3D_l i:k:(i+1) \cdot k-1, j:k:(j+1) \cdot k-1})) \quad (5.15)$$

$$P_{2D_3D_{avg_l}} = mlp((P_{2D_3D_l i:k:(i+1) \cdot k-1, j:k:(j+1) \cdot k-1})) \quad (5.16)$$

where i, j is the value at position i, j in the output matrix $P_{2D_3D_l}^C$ and $k \times k$ is the size of the window for max and average pooling operations.

$$P_{2D_3D_{l i,j}}^C = \sigma_2(P_{2D_3D_{max_l}} \oplus P_{2D_3D_{avg_l}}) \quad (5.17)$$

- The output of the channel attention mechanism $P_{2D_3D_l}^C$ is multiplied element-wise with the output of the spatial attention map $P_{2D_3D_l}^A$ and later concatenated with the feature map $P_{3D_l}^A$ from the 3D network as shown in Equation 5.18.

$$P_{3D_l}^{A''} = (P_{2D_3D_l}^C \odot P_{2D_3D_l}^A) \parallel P_{3D_l}^A \quad (5.18)$$

- The final output of the feature fusion module is denoted as $P_{3D_l}^{A''}$ and is sent to the 3D network for decoder operations involving two 3 x 3 x 3 convolution batch normalization, ReLU, and upsampling.

5.4 Pan-SUNet: Panoptic Semantic Utility Network

This introduction has discussed that Pan-SUNet takes SUNet as a baseline to expand further the scene understanding pipeline for the utility transmission corridor. This section will introduce Pan-SUNet, a unique novel network that takes multimodal input and uses multi-resolution deep-learning neural networks that embed spatial regularities between the regions and objects of interest. It comprises three different classifiers. A 2D regional prediction network [52] and a 2D object detection [11] network collaborate to generate utility region-based segmentation masks and detect objects that constrain and refine the predictions of

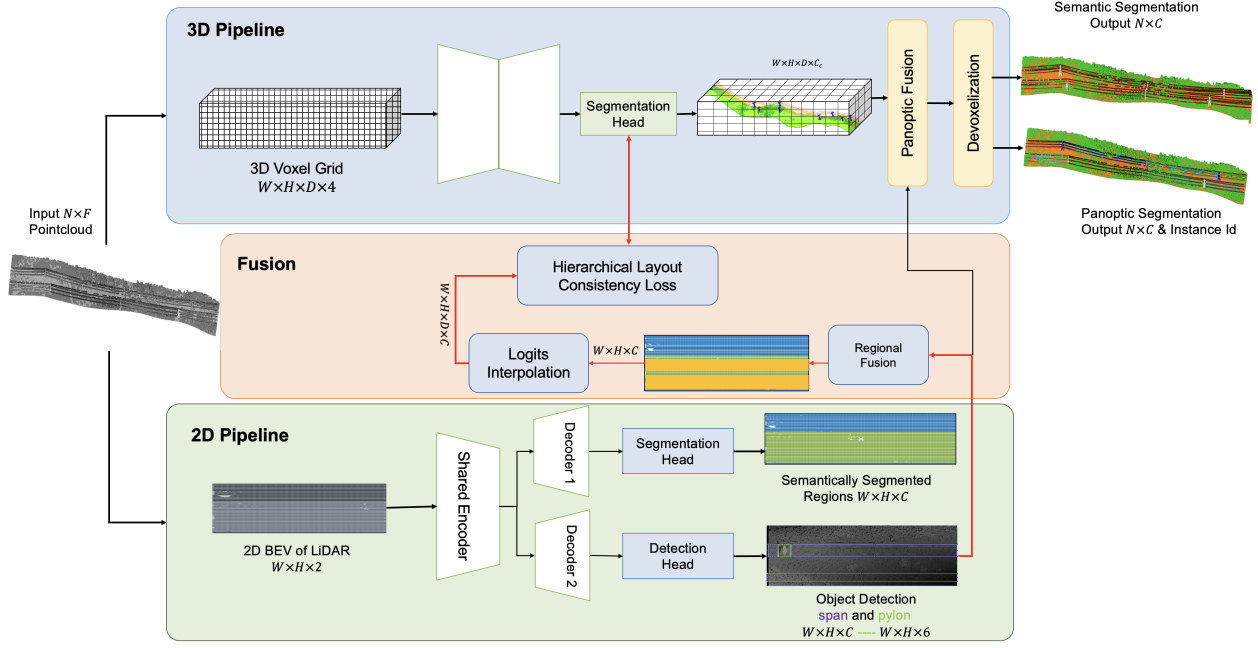


Figure 5.6: The Overall architecture of Pan-SUNet: 3D voxel grid and 2D BEV is generated from point cloud. 3D Voxel grid is convoluted through multi-resolution 3D semantic segmentation. 2D BEV uses a shared encoder for segmentation regional prediction of corridor and non-corridor classes. The detection head produces span and pylon objects that can later fuse regional masks to impose the spatial layout consistency on 3D objects through the hierarchical layout consistency loss function. Panoptic fusion fuses the 2D Object and semantic segmentation to project onto 3D objects and readjusts semantic segmentation labels.

a 3D network through regional fusion, loss-based late fusion, and panoptic fusion as shown in Figure 5.6. The next sections will discuss panoptic segmentation’s contribution in detail: BEV modified pipeline, regional fusion, and panoptic fusion modules, as the baseline of SUNet remains the same.

5.4.1 Panoptic Segmentation

This module combines the 3D semantic segmentation network’s output and object detection in the 2D pipeline for pylons and spans, which leverages semantic spatial layout consistency.

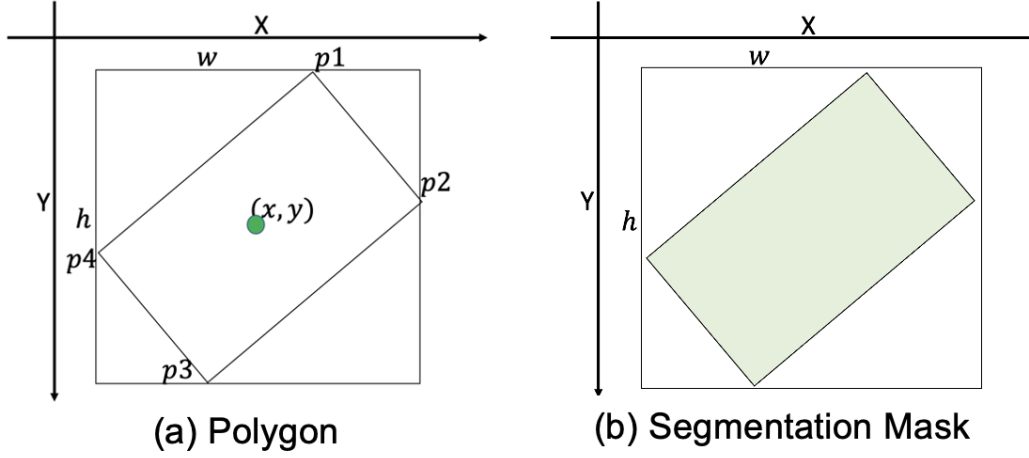


Figure 5.7: Regional fusion generation of segmentation shows a) polygon generated from step 1 and b) segmentation mask generated from polygon.

This approach focuses on utility regions and key objects like pylons and spans, which are detected in 2D BEV and projected onto 3D voxel space using a simple process. The Panoptic segmentation module estimates the height of pylons and spans based on the 3D semantic segmentation prediction, enabling the conversion of 2D bounding boxes to 3D by adding the z-axis. While this work emphasizes semantic segmentation, panoptic segmentation generates instance IDs for pylons and spans and fine-tunes segmentation results by adjusting misclassifications within 3D bounding boxes. In Figure 5.6, pylon and span instances are color-coded according to their object instance IDs.

5.4.2 2D BEV Modified Pipeline

Making spatially accurate predictions requires integrating a broader receptive field and global context into a semantic segmentation network, which is a considerable problem. Human scene semantic perception strongly depends on their capacity for understanding the larger context. The 3D segmentation network can only encode the local context and does not have the global scene information required to encode the spatial layout and the object’s regional relationship. To solve this problem, a 2D BEV pipeline that fuses the missing data using a

loss-based late fusion module is utilized. The 2D pipeline consists of two classifiers: a regional semantic segmentation and a two-dimensional object detection.

Regional Semantic Segmentation

Any simple 2D segmentation network can be used for regional segmentation. In this specific design, a U-shaped encoder-decoder is used, which generates a regional class probability of shape $W \times H \times C_p$ and takes Bird’s Eye View (BEV) representation of a complete 3D point cloud. Regional segmentation and object detection share an encoder to learn shared features.

Object Detection

To leverage the regional semantic structure of utility corridors, research introduced an additional module for 2D object detection. Previous work has configured utility transmission zones based on corridor and non-corridor regions. This choice was due to the lack of discriminating features between corridor and non-corridor regions, as vegetation height wasn’t diverse in the available dataset. This research work attempts to devise a methodology to segment utility regions from the 2D BEV (Bird’s Eye View). The proposed methodology utilizes off-the-shelf 2D object detection networks to detect pylons and spans, which represent the utility region [14, 10, 67, 13, 11, 12]. Hence, this part of the Pan-SUNet utilizes the three-region-based spatial layout consistency. The state-of-the-art orientated object detection network is used to detect span and pylon objects. It consists of a two-stage object detection network that uses a shared encoder, an oriented regional proposal network, and a regional classifier and regressor. The regional proposal network takes an encoded feature. It generates a regional proposal vector $(x, y, h, w, \delta\alpha, \delta\beta)$ with a rotated RoI that is passed to stage II, where the R-CNN produces a classification label and regress spatial location of the object or rotated bounding box [11, 10]. By incorporating these regionally fused segmentation masks of three utility transmission zones, the network’s ability to predict 3D objects that are of interest in the utility transmission zones is enhanced.

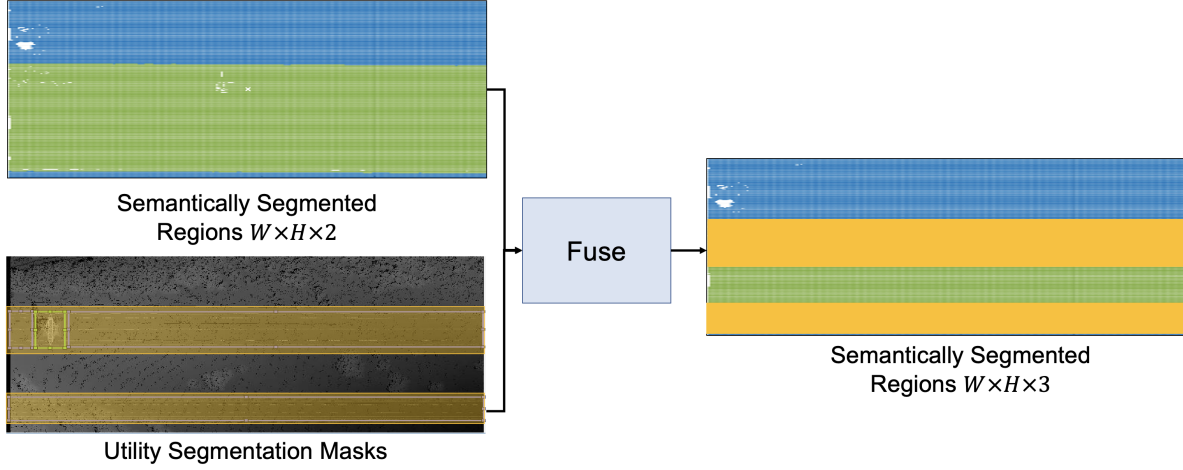


Figure 5.8: Regional fusion from 2D regional segmentation and utility segmentation mask.

5.4.3 Fusion

The Fusion has four key modules: a) regional fusion, b) logit interpolation, c) hierarchical layout consistency loss function, and d) panoptic fusion.

Regional Fusion

As discussed in the above section, this network has two classifiers: the segmentation classifier outputs a 2D mask with two regional classes of corridor and non-corridor regions. Corridor regions also have utility regions merged into them. This fusion performs the following steps to generate implicit three classes based 2D regional segmentation mask using predicted oriented bounding boxes of pylons and span along predicted regional segmentation.

- Step 1: Convert oriented bounding boxes to polygons using the following Equations 5.19, 5.20, 5.21, and 5.22, which generates four vertices $p1, p2, p3, p4$.

$$p1 = (x, y - \frac{h}{2}) + (\delta\alpha, 0) \quad (5.19)$$

$$p2 = (x + \frac{w}{2}, y) + (0, \delta\beta) \quad (5.20)$$

$$p3 = (x, y + \frac{h}{2}) + (-\delta\alpha, 0) \quad (5.21)$$

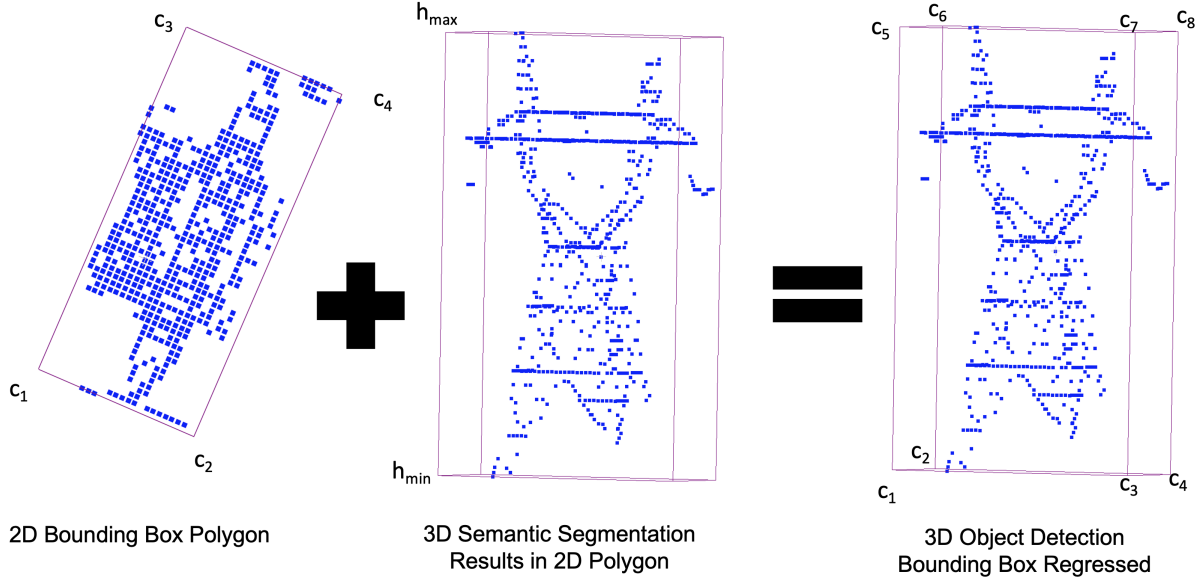


Figure 5.9: Panoptic Fusion: Conversion of the 2D Bounding boxes objects to 3D objects.

$$p4 = (x - \frac{w}{2}, y) + (0, -\delta\beta) \quad (5.22)$$

- Step 2: Estimate the segmentation masks from polygons as shown in Figure 5.7.
- Step 3: Fuse the labels from regional segmentation and utility region generated in the previous step as shown in Figure 5.8

Panoptic Fusion

The module takes input from 2D object detection of pylon and span $W \times H \times C$ and $W \times H \times 6$ and use the 3D semantic segmentation results $W \times H \times D \times C$ and adjust the z-axis for implicit 3D object detection of pylon and spans. The semantic segmentation results of pylon and powerline help estimate the minimum and maximum height of pylon points and powerline points to adjust the 2D bounding boxes. There are four corners $p1, p2, p3, p4$ and height information z_{min} and z_{max} to estimate the 3D bounding boxes as shown in Figure 5.9. These 3D bounding boxes then help adjust the prediction and misclassification within the 3D box for refining semantic segmentation. These 3D objects are assigned the instances ID based on their unique bounding boxes for a better semantic analysis of utility network modeling.

5.5 Summary

This chapter proposes three approaches to understanding the 3D utility transmission scene from the 3D point cloud. The first approach is SUNet, which embeds spatial layout information through loss-based late fusion. Additionally, it proposed and developed the design of Fusion-SUNet, which builds over the SUNet baseline. Fusion-SUNet aims to generalize the proposed network to fuse features on a deeper layer. Meanwhile, Pan-SUNet expands into 3D object detection and instance segmentation guided by panoptic semantic segmentation and proposed panoptic fusion while building on SUNet. The next chapter will discuss experimental configurations for all the proposed approaches and examine the dataset used for the experiments in detail.

Chapter 6

Deep Learning Evaluation Framework for Utility Network Segmentation

This chapter will discuss the two datasets used for experiments and outline the experimental configurations. It will elaborate on the design of experimental settings ablation studies for all three approaches.

6.1 Dataset

The experiment used datasets acquired through ALTM (Airborne Laser Terrain Mapping) from two sensors shown in Figure 6.1. These 3D point clouds were collected over utility transmission lines to detect powerlines, pylons, high veg, medium veg, ground, and buildings. The first dataset was acquired using Riegl, and the other was from Teledyne Optech’s sensor.

Table 6.1 shows the comparison between the dataset acquired by Riegl Q560 and Teledyne Optech Galaxy T1000 with various factors such as density, powerline voltage, available classes, coverage area, terrain characteristics, environment, location, and segmentation difficulty. These variables reveal the high point density and diversity of transmission lines, presenting challenges for subsequent scene analysis and computer vision tasks such as semantic segmentation and object detection.



(a) Riegl Q560



(b) Teledyne Optech Galaxy T1000

Figure 6.1: Riegl and Teledyne Optech sensors used to acquire dataset.

Varriables	Riegl	Teledyne Optech
Density	$5pts/m^2$	upto $15pts/m^2$
Voltage	230kV	Variable
Wires Types	Transmission	Transmission, Distribution, Residential
Classes	4	7
Coverage Area	67km	84km
Terrain	Non-Homogenous	Relatively Flat
Environment	High Vegetated Area	Mix Vegetations
Locations	North America	Europe
Segmentation Difficulty	Medium-High	High

Table 6.1: Comparison of Teledyne and Riegl dataset.

6.1.1 Riegl Q560

Riegl Q560 laser scanner to collect the data across a $67km$ region in Steamboat Springs, Colorado, in the United States. The acquired data were later split into train and test sets to conduct experiments. The first $14km$ of the dataset served as a testing increment, and the remaining data served as network training and validation increments. The collection consisted of 67 non-overlapping scenes, each with two million points and an average density of $5points/m^2$. It contains four classes: ground, high vegetation, pylon, and powerlines.

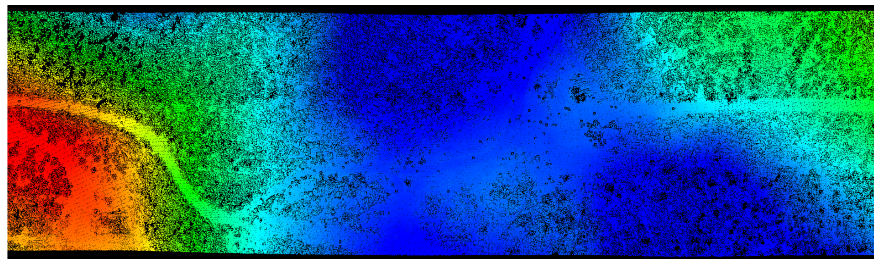
6.1.2 Teledyne Optech Galaxy T1000

Teledyne Optech’s Galaxy T1000 airborne LiDAR scanner is utilized by the Network Mapping Group (NMGroup) to survey transmission corridors throughout the United Kingdom. NMG surveyed a transmission corridor that was $84km$ long. The acquired data contained the noise points, but the research did not include quantitative analysis due to the lack of ground truth labels. The data was cleaned using NSANet-1 and voxel-based post-processing. The acquired cleaned data was divided into training and testing sets, with approximately $20km$ selected for testing increment. The average density of each point cloud is up to $(15points/m^2)$.

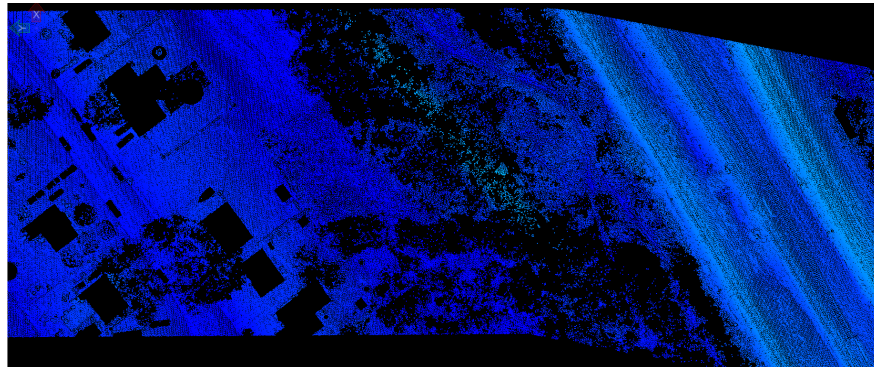
6.1.3 Qualitative Comparision of Two Datasets

Figure 6.2 presents a visual comparison of two datasets, with a focus on the elevation aspect of the ground. The figure is split into two parts, (a) and (b), each depicting a color-coded elevation map where the color gradient from red to blue represents a transition from high to low elevation. In subfigure (a), the Riegl dataset showcases a diverse topography with a smooth elevation gradient depicted by colors ranging from red (higher elevation) to blue (lower elevation), offering potential applications in geological analysis.

Subfigure (b) reveals the Teledyne Optech dataset, featuring a distinctive linear pattern in the red-to-blue gradient, suggesting consistent, flat terrain or abrupt elevation changes, such as a riverbed or man-made features.

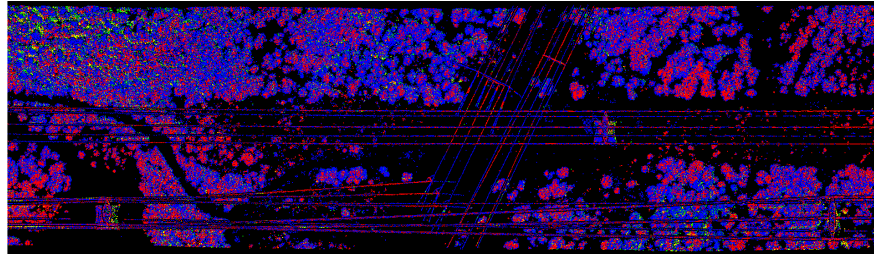


(a) Riegl: Red to blue represent high to low elevation

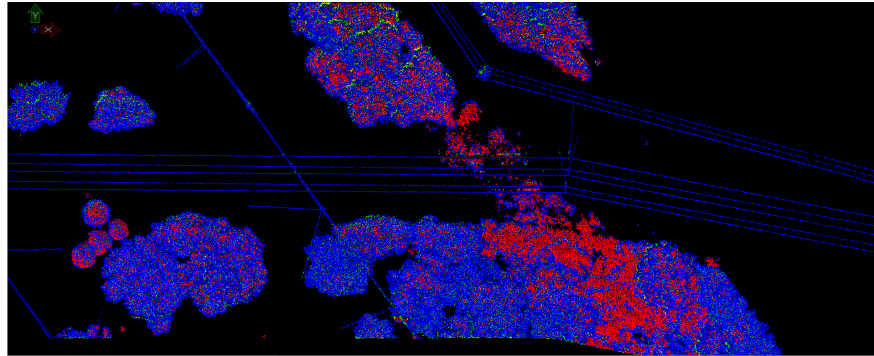


(b) Teledyne Optech: Red to blue represent high to low elevation

Figure 6.2: Elevation-based visual comparison between Riegl and Teledyne Optech Datasets.



(a) Riegl: Red: Only Echo, Blue: First of Many, Green Intermediate, Yellow: Last of Many



(b) Teledyne Optech: Red: Only Echo, Blue: First of Many, Green Intermediate, Yellow: Last of Many

Figure 6.3: Comparative visualization of echo signatures for High Vegetation, Powerline, and Pylon classes in Riegl and Teledyne Optech datasets.

Figure 6.3 provides a comparative visualization of echo classification from Riegl and Teledyne Optech LiDAR datasets, color-coded to represent single and multiple echoes. In subfigures (a) and (b), red indicates single echoes, blue denotes the first of many echoes, green represents intermediate echoes, and yellow signifies the last of multiple echoes.

In Figure 6.3(a), the Riegl dataset showcases a complex environment with a notable presence of intermediate and last echoes. Figure 6.3(b) illustrates the Teledyne Optech dataset, displaying a mix of single and multiple echoes with a distribution pattern suggesting variations in canopy density or surface complexity. This comparison aids in understanding how different LiDAR systems interpret scenes, particularly in distinguishing features like vegetation density, terrain roughness, and object height.

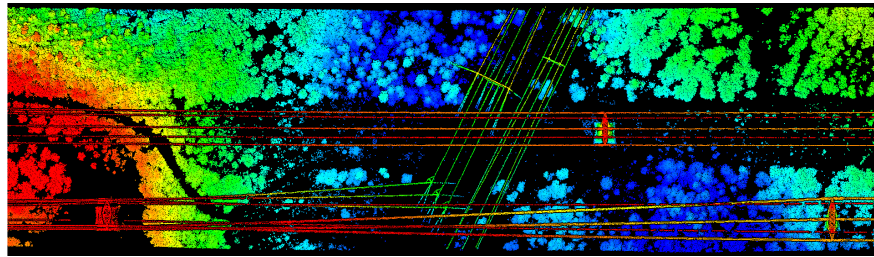
Figure 6.4 visually compares elevation data captured by Riegl and Teledyne Optech LiDAR sensors, focusing on the vertical aspect of pylons, powerlines, and high vegetation. In subfigures (a) and (b), a color gradient from red to blue signifies a range from high to low elevation.

In Figure 6.4(a), the Riegl dataset displays a vivid spectrum where warmer colors represent high elevations and cooler colors denote lower elevations. This visual cue aids in interpreting the terrain's verticality, especially for infrastructure mapping of pylons and powerlines. Figure 6.4(b) shows the Teledyne Optech dataset using the same color scheme to depict elevation differences. The warmer colors highlight high-elevation features against the cooler backdrop of lower terrain.

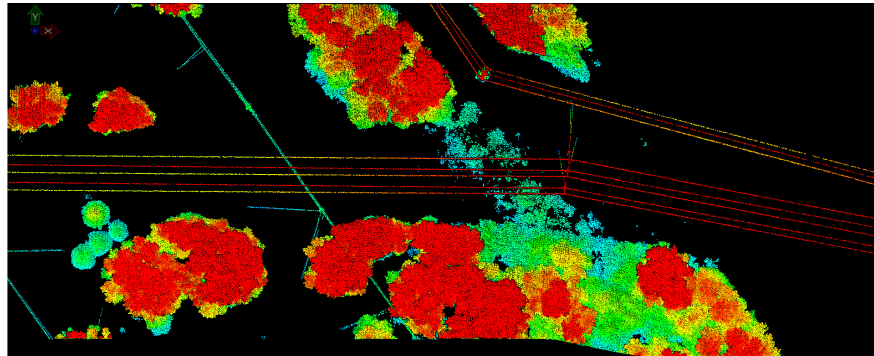
This visual analysis emphasizes elevation disparities within each dataset. It showcases the sensors' ability to capture the vertical structure of the environment, which is crucial for applications like powerline network management, forestry, and urban development.

Figure 6.5 presents a comparative visualization of Teledyne Optech Galaxy T1000 and Riegl Q560 LiDAR datasets, emphasizing point density, echo type, and elevation.

The top row compares point density, with the Riegl Q560 on the left showing well-defined linear patterns denoting high density and the Teledyne Optech Galaxy T1000 on the right exhibiting a seemingly more uniform distribution.



(a) Riegl: Red to blue represent high to low elevation



(b) Teledyne Optech: Red to blue represent high to low elevation

Figure 6.4: Elevation profile visualization for Pylon, Powerline, and High Vegetation features in Riegl and Teledyne Optech datasets.

The middle row illustrates echo patterns, with the Riegl dataset dominated by first and only echoes, while the Teledyne dataset displays a more complex pattern, including intermediate and last echoes.

The bottom row provides an elevation map, with the Riegl data exhibiting clear elevation along infrastructure lines and the Teledyne dataset showcasing a rich topography with elevation variations.

Figure 6.6 visually compares classified data from Riegl and Teledyne Optech LiDAR systems, revealing distinctions in class granularity. The Riegl image focuses on powerlines (red), pylons (white), and ground (orange), while the Teledyne image introduces more classes, including medium vegetation (dark green), buildings (yellow), and other features (blue). This contrast suggests that Teledyne Optech may have a higher classification capability or a different processing algorithm discerning a broader range of features.

6.2 Experimental Configuration

This section discusses the architectural designs and experimental setups for the deep learning networks introduced in Chapter 5: SUNet, Fusion-SUNet, and Pan-SUNet. It outlines the data preprocessing steps, such as input data representation and feature channel organization, and discusses the selection of voxel and grid sizes. Additionally, it covers the computational and memory resources utilized for the experiments and describes the methodologies employed for ground truth preparation.

6.2.1 SUNet and Fusion-SUNet

Incremental Learning with K-fold Cross-Validation

This involves a combination of incremental learning to train the network on a large dataset using increments and k-cross validation for fine-tuning to prevent overfitting. This technique is applied to train regional and 3D semantic segmentation networks in this thesis. The dataset is divided into three increments of equal size. One increment is reserved for testing, and

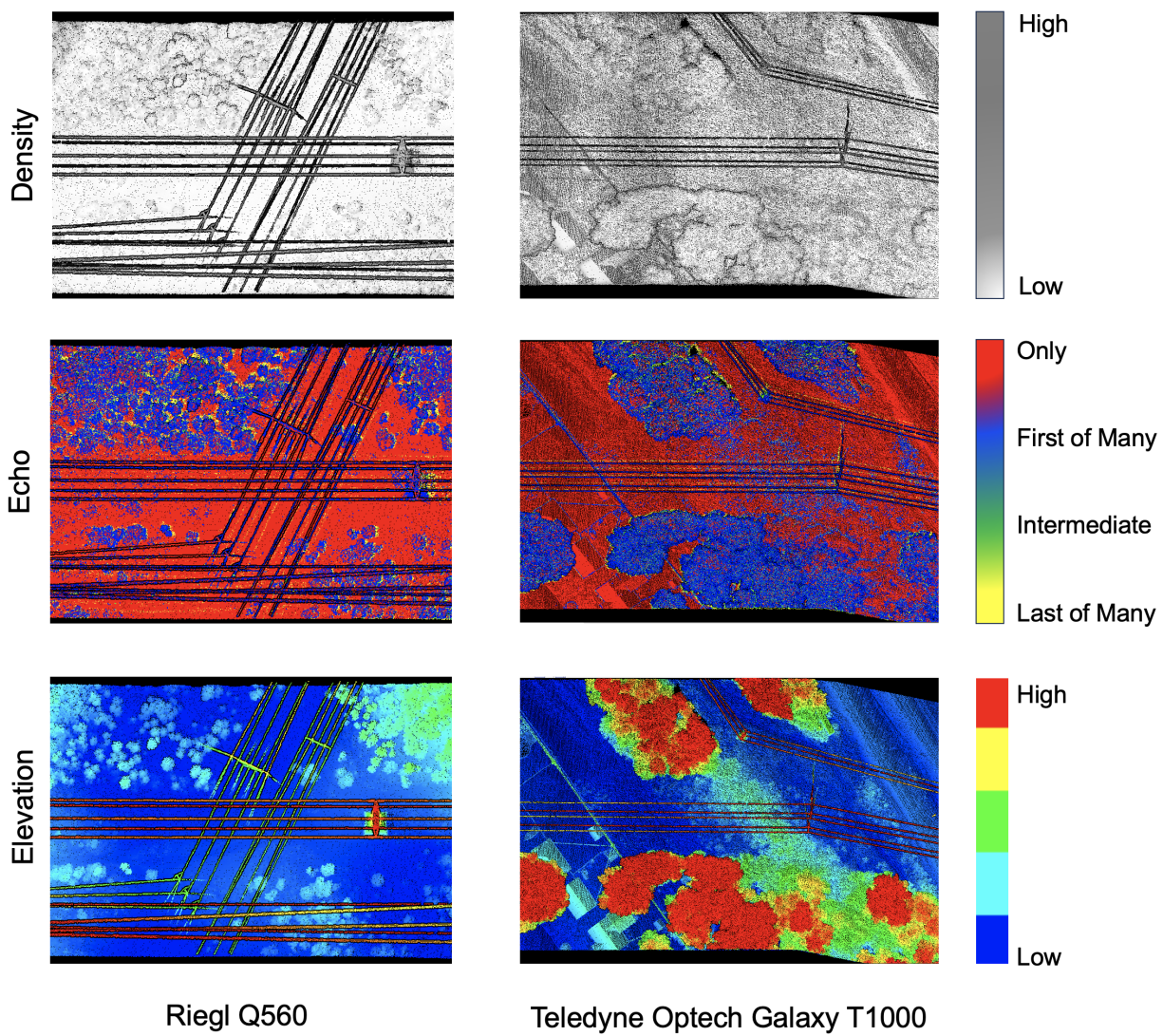
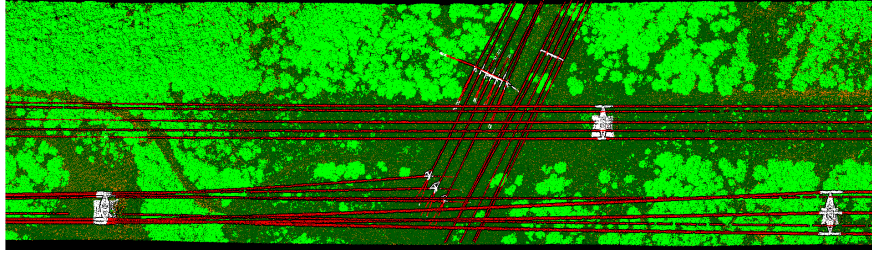
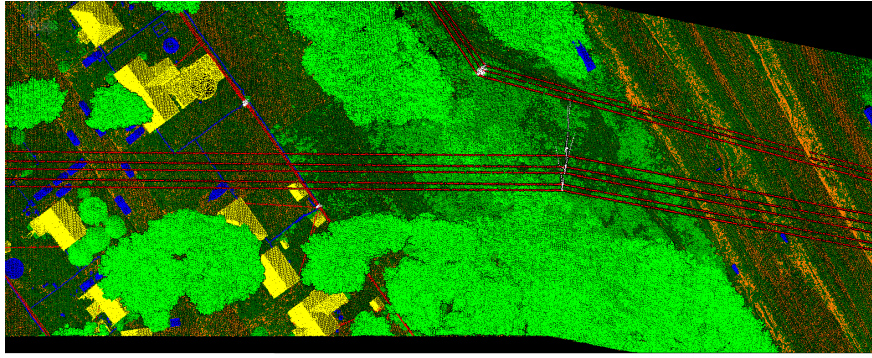


Figure 6.5: Comparative visualization across density, echo, and elevation attributes in Teledyne Optech Galaxy T1000 and Riegl Q560 datasets.



(a) Riegl Classes; Red: Powerline, White: Pylon, Orange: Ground, Green: High Veg



(b) Teledyne Optech Classes; Red: Powerline, White: Pylon, Orange: Ground, Green: High Veg, Dark Green: Medium Veg, Yellow: Building, Blue: Others

Figure 6.6: Class map visualization: comparative analysis of Riegl and Teledyne Optech datasets.

the other two increments are used for training. Each training increment undergoes $K=3$ cross-validation for fine-tuning a neural network with an 80-20%

2D Regional Segmentation

For pretraining the *2D regional prediction network* in the context of global spatial layout prediction, the experiments utilized half of the available scenes. Each scene underwent subdivision into four smaller subscenes using the GPS timestamps associated with the flight lines. Subsequently, these subscenes were mapped onto a 2D bird's eye view (BEV) grid, each with dimensions of 640×640 , where each pixel represents an area of $1m^2$.

Within this grid, two feature channels were designated: the mean elevation and the count of occupancy points per pixel, resulting in an input dimensionality for the 2D network of $640 \times 640 \times 2$. The network’s output corresponds to a size of $640 \times 640 \times 3$, representing the confidence scores for the regional classes.

The network underwent pretraining with a batch size of 1 over 100 epochs, applying incremental learning with k-fold cross-validation for fine-tuning a neural network to mitigate the risk of overfitting. Data augmentation strategies were employed to enhance the robustness of the model, including horizontal and vertical flips, along with random rotations.

The training was executed on a dual GPU setup, utilizing RTX 6000 cards using the Adam optimizer with a learning rate of $\alpha = 0.001$ and a 0.85 decay rate for every 10,000 steps through a staircase learning rate schedule. The training duration ranged between 4 to 5 hours, and inference on this pre-trained model was achieved in approximately 30 seconds, demonstrating the network’s efficiency.

3D Semantic Segmentation

The voxel grid is generated over each subscene with dimensions of $640 \times 640 \times 448$, where each voxel represents a volume of $1m^3$. Each batch is fed with the maximum elevation data from the entire scene to give the network a comprehensive vertical context. The feature channels comprise absolute and relative elevation metrics, as detailed in the methodology and illustrated in Figure 6.4, the number of occupancy points, and the number of returns. A feature engineering study was conducted to select these specific features, which will be expounded upon in the subsequent section. The output from both approaches is a tensor of dimensions $32 \times 32 \times 448 \times 5$, signifying the confidence scores for 3D class predictions, which include background, pylon, powerline, vegetation, and ground. For the Teledyne Optech Galaxy T1000 dataset, the output dimensionality is $32 \times 32 \times 448 \times 8$. The final prediction phase assigns the true class label based on the highest confidence score and projects voxel labels onto the points using the projection matrix. Training for these networks was conducted using two NVIDIA RTX 6000 GPUs over 100 epochs, utilizing the Adam optimizer with a

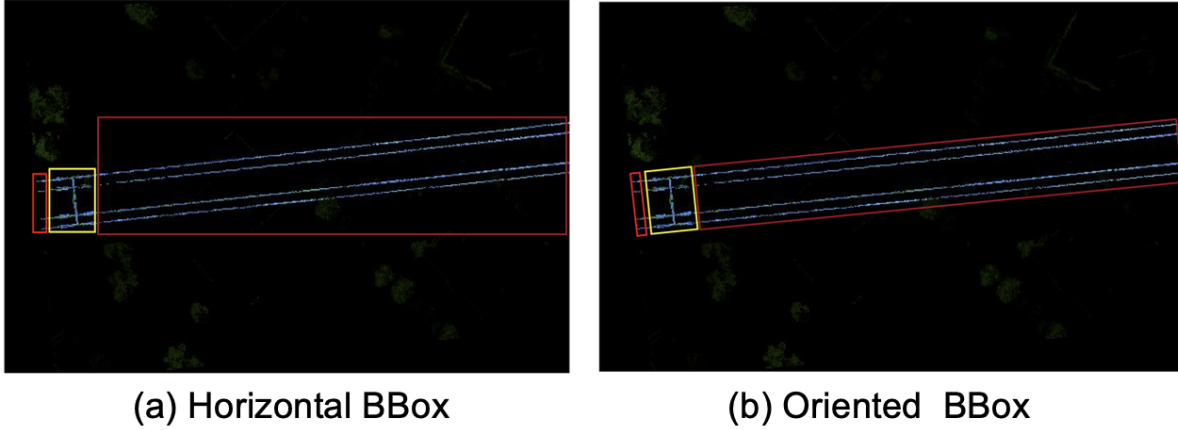


Figure 6.7: Comparative visualization of span and pylon detection: horizontal vs. oriented bounding boxes.

learning rate of $\alpha = 0.0005$ and a 0.80 decay rate for every 100,000 steps through a staircase learning rate schedule. The training time spanned between 48 and 60 hours, with inference taking approximately 2 to 3 minutes for each model.

6.2.2 Pan-SUNet

2D Object Detection Ground truth

To generate labels for object detection of pylons and spans, two distinct annotation tools were employed depending on the orientation of the bounding boxes. Horizontal bounding boxes were created using the MATLAB Image Processing Toolbox, while oriented bounding boxes were annotated with the Labelme tool [157]. The experiments conducted compared the effectiveness of these two bounding box types. Horizontal bounding boxes tend to encompass more background in cases involving oriented objects, which can influence detection performance. Consequently, a comparative analysis was conducted to assess the impact of bounding box orientation on model performance, with results illustrated in Figure 6.7.

To set the stage for experiments, a range of 2D object detection networks were pre-trained and evaluated. A 2D XY-projection of the point cloud is rasterized on a $640m^2$ grid, with

each grid pixel representing $1m^2$. The networks' inputs comprised three feature channels: the normalized standard deviation of elevation scaled between [0-1], the normalized standard deviation of elevation values quantized between 0 and 255, and a binarized representation of elevation standard deviation values per pixel.

A comprehensive comparative study was undertaken to examine the feasibility and performance of 2D object detection. This study selected state-of-the-art networks for this evaluation, comparing their effectiveness with horizontal versus oriented bounding boxes, and contrasting two-stage systems like Faster R-CNN [67] and Rotated Faster R-CNN [10], with one-stage systems such as Oriented R-CNN [11], DAFNe [13], and YOLOv3 [14].

2D Regional Pipeline

The 2D regional prediction network is pre-trained on half of the available scenes for the global, regional prediction of spatial layout as part of the experimental preparation. Each scene is segmented into four smaller scenes according to the GPS time associated with the flight line. These are then projected onto a 640×640 2D BEV grid, where each pixel corresponds to $1m^2$. The network's input features three channels: the normalized standard deviation of elevation scaled between [0-1], the normalized standard deviation of elevation values ranging between 0 and 255, and the binarized standard deviation of elevation values per pixel, as shown in Figure 6.8. The input size for the 2D network is $640 \times 640 \times 3$, and the output size, representing the confidence scores for the regional classes, is $640 \times 640 \times 3$. To avoid overfitting, the network is pre-trained with a batch size of 1 for 100 epochs, employing K-fold cross-validation. Data augmentation techniques are utilized for random rotation, horizontal flip, and vertical flip. The training is conducted on two RTX 6000 GPUs, using the Adam optimizer with a learning rate of $\alpha = 0.001$ and a 0.85 decay rate for every 10,000 steps through a staircase learning rate schedule. The training time ranges between 4 to 5 hours, and inference takes approximately 30 seconds.

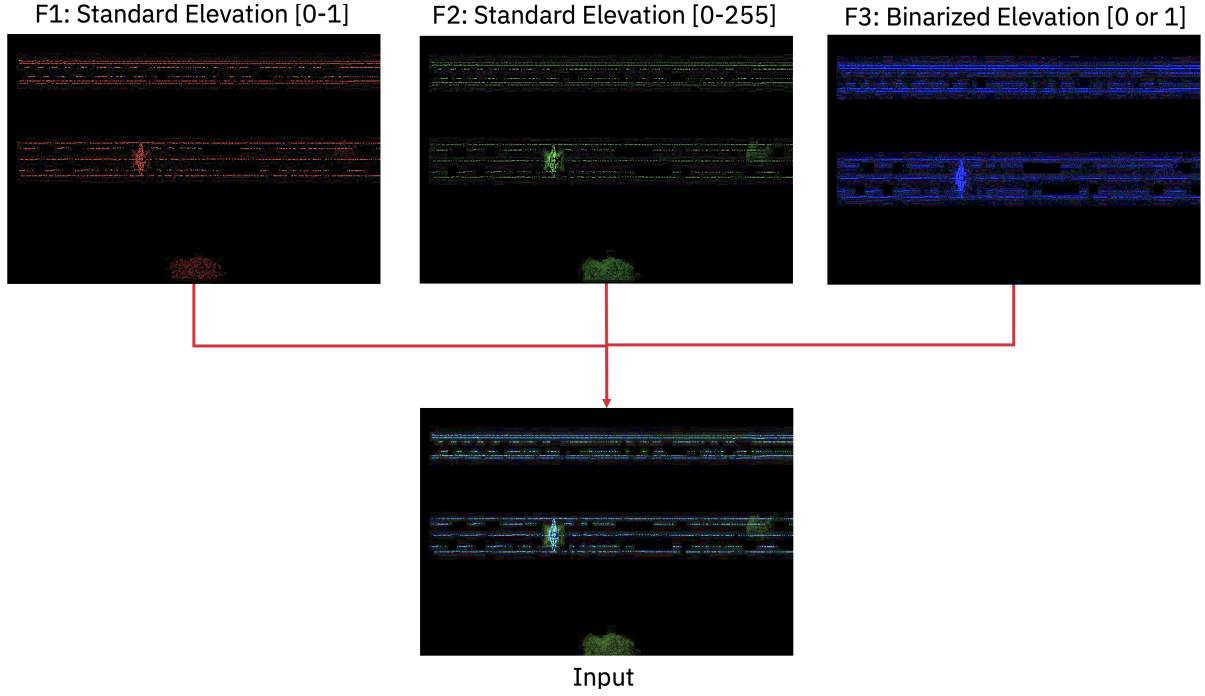


Figure 6.8: Visualization of input feature channels for the 2D regional pipeline.

3D Semantic Segmentation

A voxel grid is generated of size $640 \times 640 \times 448$ and a voxel of $1m^3$ over each subscene. Each batch $32 \times 32 \times 448 \times 4$ contained the highest elevation of the entire picture to give the network a comprehensive look and help it cope with vertical context more effectively. The number of returns, the number of occupancy points, and the absolute and relative elevation were all included in the feature channels. The feature engineering study presented by SUNet is utilized to choose these features. Pan-SUNet produces a confidence score against 3D classes as $32 \times 32 \times 448 \times 5$ (background, pylon, powerline, vegetation, and ground) and $32 \times 32 \times 448 \times 8$ for T1000's dataset. The final prediction provides a true label based on the greatest confidence score and projects voxel labels on points using the projection matrix and using two GPU RTX 6000, utilizing the Adam optimizer with a learning rate of $\alpha = 0.0005$ and a 0.80 decay rate for every 100,000 steps through a staircase learning rate schedule. Pan-SUNet was trained for 100 epochs over 48–60 hours, with inference requiring only a few

minutes.

6.3 Evalaution Metrics

6.3.1 Semantic Segmentation

Precision, recall, and F1-Score are the key performance indicators to assess the effectiveness of the network. Precision quantifies the proportion of accurately predicted instances out of all the cases classified as a particular class. Recall, however, captures the ratio of correctly predicted cases to the total number of actual instances of that class. The F1-Score provides a harmonic mean of precision and recall, indicating their balance.

6.3.2 Object Detection

The success of object detection networks in locating and identifying objects within an image is gauged using Average Precision (AP) and Average Recall (AR). AP is derived from the area under the precision-recall curve and considers the Intersection over the Union (IoU) threshold, which discriminates between true positives and false positives. Similarly, AR is calculated from the area under the recall-precision curve, factoring in the same IoU threshold. These metrics provide a comprehensive view of a model's detection performance.

$$AP = \sum_{k=0}^{k=n-1} [Recall(k) - Recall(k+1)] * Precision(k) \quad (6.1)$$

$$IoU = \frac{AreaOfOverlap}{AreaOfUnion} \quad (6.2)$$

6.3.3 Ablation Studies

SUNet

This research undertook a sequence of ablation experiments to substantiate the individual contributions of various components. The impact of each was scrutinized by systematically removing elements related to loss functions, features, and architectural modules. The investigations focused on the Riegl dataset, benchmarking the results against the capabilities of current leading-edge networks. Key design decisions were informed by an exhaustive performance assessment, providing a granular understanding of the factors influencing network efficiency.

Fusion-SUNet

The ablation study conducted a comparative analysis of the deep attention fusion module against leading networks, specifically within the context of the Riegl dataset.

Pan-SUNet

This section delves into the comparative performance evaluation of three foundational models for 3D semantic segmentation: 3D U-Net [9], 3D Attention U-Net [56], and RandLA [8]. The analysis spans datasets acquired from Riegl and Teledyne Optech sensors, focusing on their application in utility regions and the integration of various fusion modules. Furthermore, the study undertakes a comprehensive feasibility analysis to determine the most suitable 2D object detection network tailored to these specific datasets. It also benchmarks the capabilities of the latest object detection networks, assessing their effectiveness on data from Riegl and Teledyne.

6.4 Summary

This chapter describes the datasets used in this research and highlights the differences between data obtained from two sensor technologies. It examines the point density, terrain complexity, surrounding environment, and voltage levels of transmission lines. The discussion also covers the challenges these variables pose to the analytical processes involved in scene interpretation. In addition, the chapter provides information on the experimental setups and evaluation metrics used to assess performance, explaining the experimental design and its impact on the results. The next chapter will expand on this discussion by providing a detailed quantitative and qualitative analysis, further demonstrating the methods introduced in the previous chapter.

Chapter 7

Experimental Results and Discussion for Utility Network Segmentation

This chapter will present the results from all three approaches. It will discuss the reported results and analyze the observations. This chapter’s qualitative and quantitative analysis will summarize the second stage of the utility scene understanding pipeline. The next few sections will be divided by proposed approaches across datasets and ablation studies, as discussed in the previous chapter.

7.1 SUNet Performance Evaluation on Riegl Dataset

Chapter 7 presents the results obtained from the three deep-learning networks for segmenting utility corridor point clouds explained in Chapter 5. The goal is to perform a comprehensive and rigorous ablation analysis to evaluate the network’s performance under various factors through quantitative and qualitative techniques. The chapter is divided into sections focusing on each deep learning network and comparing their performance across different datasets. In addition, the chapter explores the detailed findings from ablation studies, as introduced earlier, to provide a more in-depth understanding of each network model’s capabilities. This analytical approach promises to highlight the relative effectiveness of the models and provide

Ablation Setup		2D Regions(2 Classes)	Dataset	F1-score			
WCE	HLC			Pylon	Powerline	High Veg	Ground
✓	✗	✓	R	82.0 %	91.0%	97.0%	94.0%
✗	✓	✓	R	84.0%	98.0%	99.0%	99.0%

Table 7.1: Comparative analysis of SUNet ablation: evaluating the impact of loss-based late fusion on F1-score performance with WCE and HLC considerations.

a comprehensive synthesis of our findings, thus guiding future research initiatives in the field.

7.1.1 Analysis of Loss Functions and Elevation Features

Loss Functions

Table 7.1 showcases an ablation study conducted on the SUNet model to investigate its performance variations under different loss functions, represented by the F1-score metric. The assessment revolves around two primary loss functions: WCE and HLC.

The results indicate that using solely the WCE loss function, the model posts an F1-score of 82% for Pylon, 91% for Powerline, 97% for High Vegetation, and 94% for Ground. In contrast, when adopting the HLC loss function, a discernible enhancement in performance across all categories emerges. To be specific, the F1-score registers (2% improvement) for Pylon , (7% improvement) for Powerline , achieves (2% improvement) for High Vegetation, and (5% improvement) for Ground .

This analysis shows the most substantial performance gains using HLC in the Powerline, High Vegetation, and Ground categories. However, while Pylon detection also benefits from the HLC loss function, the improvement is relatively modest.

This result suggests that the HLC Loss appears to be more effective than the WCE for the SUNet model in enhancing its F1-score across all tasks on the Riegl dataset.

Ablation Setup				Fusion	Dataset	F1- Score			
Occ.	AE	NR	RE			Pylon	Powerline	High Veg	Ground
✓	✓	✓	✗	HLC	R	77.0%	91.0%	97.0%	93.0%
✓	✓	✓	✓	HLC	R	84.0%	98.0%	99.0%	99.0%

Table 7.2: SUNet ablation study assessing feature engineering impact on F1-score: evaluating occupancy points, absolute and relative elevation, and number of returns.

Feature Engineering

In this ablation study, the efficiency of the SUNet model was examined through a progressive feature engineering analysis, which incorporates a combination of occupancy points, absolute elevation, and the number of returns. These features were selected to enhance the model’s input data, influenced by methodologies from [18] that were pivotal in developing absolute and relative elevation modules to contend with non-homogeneous surfaces effectively. Non-homogeneous terrains present challenges in classification tasks, often causing powerline points at lower elevations to be misclassified as ground.

When assessing the absolute performance, as depicted in Table 7.2, the SUNet model demonstrates outstanding performance, especially when ARE module is engaged. The model attains remarkably high F1-scores in the classification of High Vegetation and Ground (99% or both) upon incorporating all selected features, reflecting near-perfect classification accuracy. Furthermore, the model exhibits a robust F1-score for Powerline classification (98%).

Comparatively, excluding the RE feature decreases performance, as the F1-score drops to 77% for Pylons, 91% for Powerlines, and 93% for ground. This relative gain underlines the critical role of RE in the model’s ability to discriminate between classes effectively.

The study outcomes, particularly the improvement with the ARE module, underscore the utility of tailored feature engineering in enhancing deep learning networks.

7.1.2 Module-based Ablation Study

In a module-based ablation study, experiments are conducted by incorporating different combinations of modules into the SUNet framework. Table 7.3 presents an ablation study focusing on the recall metric for a SUNet model by varying the use of MFA and FS across different object classes.

Ablation Setup		Features				Fusion	Dataset	Recall			
MFA	FS	Occ.	AE	NR	RE			Pylon	Powerline	High Veg	Ground
$\times$	$\times$	✓	✓	✓	✓	HLC	R	72.0%	84.0%	97.0%	83.0%
✓	✓	✓	✓	✓	✓	HLC	R	78.0%	99.0%	99.0%	100.0%
✓	$\times$	✓	✓	✓	✓	HLC	R	82.0%	98.0%	99.0%	99.0%

Table 7.3: SUNet ablation study examining the impact of MFA and FS on Recall.

Ablation Study		Features				Fusion	Dataset	Precision			
MFA	FS	Occ.	AE	NR	RE			Pylon	Powerline	High Veg	Ground
$\times$	$\times$	✓	✓	✓	✓	HLC	R	84.0%	98.0%	97.0%	94.0%
✓	✓	✓	✓	✓	✓	HLC	R	97.0%	97.0%	99.0%	99.0%
✓	$\times$	✓	✓	✓	✓	HLC	R	96.0%	99.0%	99.0%	99.0%

Table 7.4: SUNet ablation study examining the impact of MFA and FS on Precision.

Ablation Study		Features				Fusion	Dataset	F1-score			
MFA	FS	Occ.	AE	NR	RE			Pylon	Powerline	High Veg	Ground
$\times$	$\times$	✓	✓	✓	✓	HLC	R	77.0%	91.0%	97.0%	93.0%
✓	✓	✓	✓	✓	✓	HLC	R	87.0%	98.0%	99.0%	100.0%
✓	$\times$	✓	✓	✓	✓	HLC	R	89.0%	99.0%	99.0%	99.0%

Table 7.5: SUNet ablation study examining the impact of MFA and FS on F1-scores.

The table displays three configurations of the SUNet model, which show the impact of including MFA and FS on the recall rates for all classes. Without MFA and FS, the first configuration serves as a baseline, yielding lower recall rates across the board, notably 72% for Pylons and 83% for Ground. Including MFA and FS in the second configuration shows a significant improvement in recall rates, particularly for the Powerline and Ground categories, which peak at 99% and 100%, respectively. This indicates a perfect recall for the Ground class, implying that all relevant ground instances in the dataset were correctly identified.

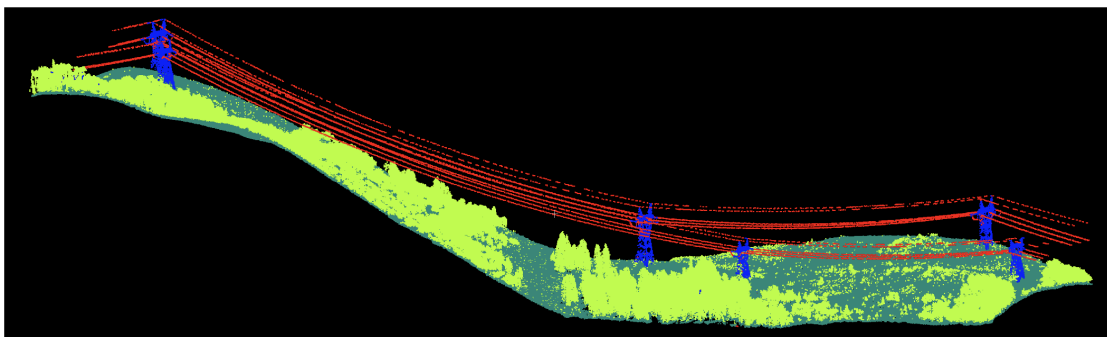
The third configuration, where MFA is enabled, but FS is disabled, reveals a subtle effect on the recall rates. While the recall for Powerline remains high at 98% there is a slight drop compared to the configuration where both MFA and FS are used. However, this configuration still outperforms the baseline for all classes, demonstrating the effectiveness of MFA alone. Notably, High Vegetation consistently shows high recall across all configurations, suggesting that the vegetation-related features are well-captured by the SUNet model irrespective of MFA or FS usage.

In summary, MFA is a critical component for the SUNet model’s ability to recall instances across the evaluated classes correctly. While FS also contributes to performance enhancement, particularly ensuring a perfect recall in the Ground class, its absence does not drastically reduce the model’s efficiency thanks to the robustness provided by MFA.

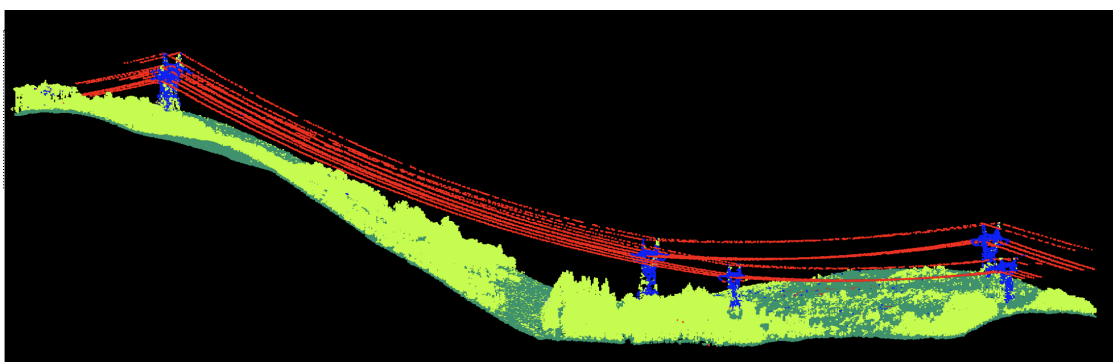
Table 7.4 and Table 7.5 showcase the influence of MFA and FS on the precision and F1-score metrics, respectively, for the SUNet model across four different classes: Pylon, Powerline, High Vegetation, and Ground.

The precision metric is assessed in Table 7.4. The baseline configuration without MFA and FS presents lower precision values for the Pylon and Ground categories, with 84% and 94%, respectively. Introducing MFA and FS yields a remarkable improvement in precision for the Pylon class, indicating a relative gain of approximately 15%. Furthermore, Ground class precision also showed a 5% relative improvement. Interestingly, for Powerline and High Veg, the precision is already high in the baseline and remains consistent across all configurations, suggesting that MFA and FS have a more pronounced effect on classes with lower baseline precision scores.

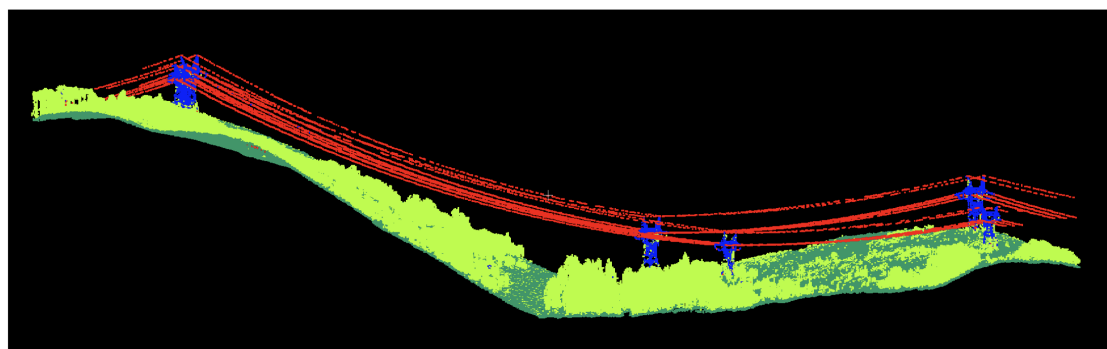
Table 7.5 focuses on the F1-score, a harmonic mean of precision and recall. The baseline (without MFA and FS) shows lower F1-scores for all classes, with the Pylon class at 77%. Implementing MFA and FS results in substantial relative gains for Pylon and Powerline, with increases to 87% and 98% in their F1-scores, translating to relative improvements of approximately 13% and 7%, respectively. The High Veg and Ground classes exhibit the highest F1-scores in all configurations, indicating robust model performance for these



Ground truth

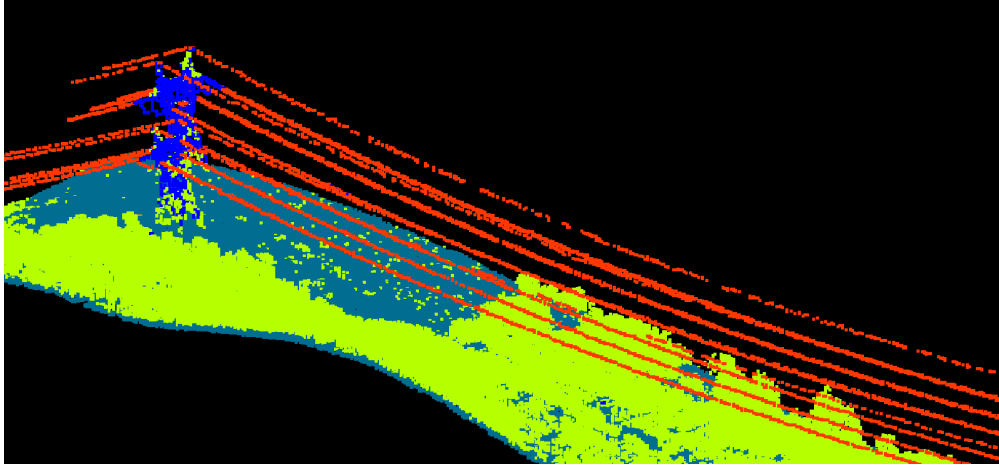


SUNet

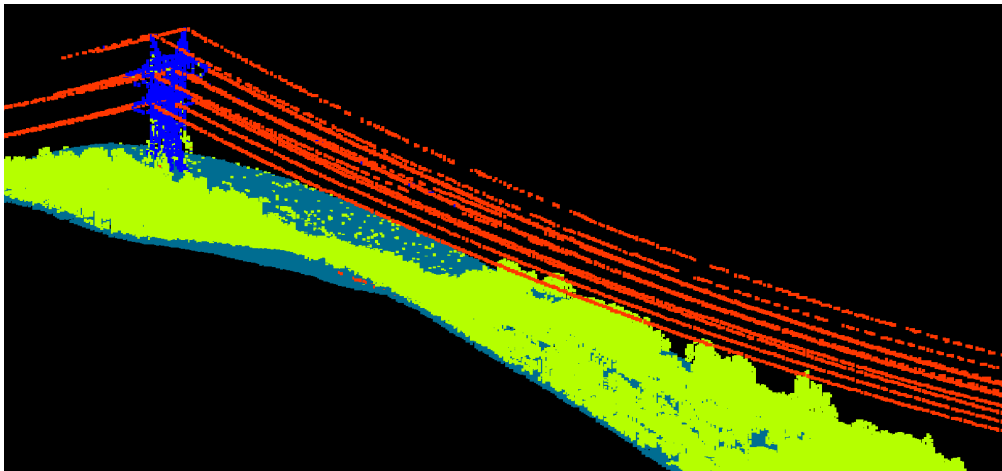


SUNet + MFA

Figure 7.1: Comparative Visualization: Ground Truth vs. SUNet vs. SUNet with MFA - Classifying Pylons (blue), Powerlines (red), Vegetation (green), and Ground (dark green).



a) SUNet



b) SUNet + MFA

Figure 7.2: Error Analysis: SUNet versus SUNet with MFA - Classifying Pylons (blue), Powerlines (red), Vegetation (green), and Ground (dark green).

categories, with the Ground class reaching a perfect score of 100% when both MFA and FS are applied.

The third configuration in each table, where MFA is used without FS, demonstrates that MFA alone contributes significantly to the precision and F1-score improvements. However, the combination of MFA and FS maximizes performance, especially for the Pylon and Ground classes, suggesting that FS provides an additional boost to the model’s performance.

These tables indicate that MFA plays a crucial role in enhancing the model’s precision and F1-score. At the same time, FS refines these metrics, leading to the most pronounced improvements in categories that initially had lower scores.

Figures 7.1 and 7.2 visually compare segmentation results and errors between the SUNet model and the SUNet model with the MFA module.

In Figure 7.1, the color-coded images represent ground truth, SUNet, and SUNet with MFA. The ground truth demonstrates an ideal scenario, while the SUNet image exhibits discrepancies, particularly around pylons and powerlines. The SUNet + MFA results show improved segmentation, with fewer fragmentation issues, closely resembling the ground truth. This suggests that MFA positively contributes to the model’s performance.

Figure 7.2 provides an error visualization, comparing SUNet’s performance with and without the MFA module. The first image (a) shows errors in the original SUNet model, with pronounced blue and red areas indicating misclassifications of pylons and powerlines. Lower parts of pylons are still misclassified as trees, and isolated points are labeled as powerlines. These errors occur because of the structural similarity between trees and pylons regarding point density and data distribution. The second image (b), after integrating MFA, demonstrates reduced errors, indicating improved precision in classifying critical infrastructure components.

These visualizations effectively illustrate the positive impact of the MFA module on SUNet’s accuracy in segmenting and classifying scene elements, particularly in reducing errors for critical infrastructure components like pylons and powerlines.

7.2 Fusion-SUNet Performance on Riegl Dataset

This section aims to evaluate the performance of Fusion-SUNet in comparison to state-of-the-art networks. The study seeks to understand the impact of attention-based multiscale feature fusion in enforcing spatial layout consistency. As discussed in Section 6.3.3, the baseline network for Fusion-SUNet was the Attention 3D U-Net [56]. The comparative study used Riegl data as input for Fusion-SUNet and compared it to other state-of-the-art networks.

7.2.1 Recall Analysis

Experiments were conducted to compare Fusion-SUNet with various versions of SUNet [35], Attention 3D U-Net [56], and pre-trained RandLA [8], which was fine-tuned on the Riegl data. Table 7.6 presents the recall measure of Fusion-SUNet in the Riegl dataset.

At the center of this discussion is the modular design of the tested networks. The networks have MFA, FS, Occ, AE, RE, NR, and HLC. The integration of HLC and the novel A-Fusion module set Fusion-SUNet apart.

With an impressive 94% recall rate, Fusion-SUNet proved to be a more effective pylon detector than Attention 3D U-Net, the baseline model, by a significant 22%.

Fusion-SUNet’s 98% recall in powerline identification somewhat beats that of SUNet, its closest competition, highlighting the added importance of high-level context knowledge. This emphasizes how increased context comprehension has a subtle but significant impact on complex object identification tasks.

Fusion-SUNet consistently performs well in the ‘High Vegetation’ category, suggesting that while it excels in vegetation detection, the integration of A-Fusion may not result in significant improvements in this specific category.

Ground classification recall rates, which are constantly at 99%

Fusion-SUNet is a network that combines advanced modules and features to improve the detection of pylons while still accurately detecting powerlines and vegetative occlusions. It demonstrates its reliability and potential as a model for LiDAR-based aerial surveillance and

analysis.

7.2.2 Precision Analysis

Table 7.7 provides a comparative analysis of Fusion-SUNet’s performance in terms of precision on the Riegl dataset. Regarding pylon identification, Fusion-SUNet achieves a precision of 92% which matches the best performance among other configurations. This proves that while Fusion-SUNet has fewer false negatives in identifying pylons, as seen in the recall rates, it does so with high accuracy, maintaining a low false-positive rate.

For powerline detection and identifying ‘High Vegetation’ objects, Fusion-SUNet displays a precision of 99% indicating its ability to recognize vegetative obstructions and powerlines accurately is comparable to the top-performing models. This further supports its usefulness when differentiating between vegetation and critical infrastructure is crucial.

The ground classification precision of Fusion-SUNet is 99% similar to the performance of other networks, indicating its ability to identify the ground accurately.

7.2.3 F1-score Analysis

Table 7.8 provides a comparative analysis of Fusion-SUNet’s performance in terms of F1-score on the Riegl dataset. Fusion-SUNet attains an F1-score of 93% for pylons, reflecting a balanced and robust performance in detecting the presence of pylons and minimizing incorrect identifications. In powerline classification, the F1-score for Fusion-SUNet is perfect at 100%, indicating an exceptional balance in identifying powerlines accurately while avoiding misclassification. This perfect score mirrors that of the SUNet model without the HLC module, again showing that additional complexity in the model does not necessarily translate to improved performance in all categories. For High Veg objects, the F1-score is a high 99% for Fusion-SUNet, reaffirming its high precision and recall for vegetative features. The ground detection’s F1-score of 99% remains consistent across all models, including Fusion-SUNet, confirming that this task has been effectively optimized in current model configurations.

Fusion-SUNet’s consistent high accuracy is highlighted by the precision and F1-score tables, particularly in powerline detection, where the addition of complex features and modules does not significantly improve precision in every case but maintains an overall efficient performance that shows dependable semantic segmentation within LiDAR datasets.

Ablation Setup	Modules		Features				Fusion		Recall			
	MFA	FS	Occ.	AE	NR	RE	HLC	A-Fusion	Pylon	Powerline	High Veg	Ground
Att 3D U-Net	X	X	✓	✓	✓	✓	X	X	72.0%	84.0%	97.0%	93.0%
RandLA	X	X	X	✓	✓	✓	X	X	75.0%	86.5%	76.0%	95.0%
SUNet	✓	✓	✓	✓	✓	✓	✓	X	78.0%	99.0%	99.0%	100.0%
SUNet	✓	X	✓	✓	✓	✓	✓	X	82.0%	98.0%	99.0%	99.0%
Fusion-SUNet	✓	✓	✓	✓	✓	✓	✓	✓	94.0%	98.0%	99.0%	99.0%

Table 7.6: Comparative study of Fusion-SUNet for Riegl dataset on Recall.

Ablation Setup	Modules		Features				Fusion		Precision			
	MFA	FS	Occ.	AE	NR	RE	HLC	A-Fusion	Pylon	Powerline	High Veg	Ground
Att 3D U-Net	X	X	✓	✓	✓	✓	X	X	84.0%	98.0%	97.0%	94.0%
RandLA	X	X	X	✓	✓	✓	X	X	92.0%	98.5%	87.0%	96.0%
SUNet	✓	✓	✓	✓	✓	✓	✓	X	97.0%	97.0%	99.0%	99.0%
SUNet	✓	X	✓	✓	✓	✓	✓	X	96.0%	99.0%	99.0%	99.0%
Fusion-SUNet	✓	✓	✓	✓	✓	✓	✓	✓	92.0%	99.0%	99.0%	99.0%

Table 7.7: Comparative study of Fusion-SUNet for Riegl dataset on Precision.

Ablation Setup	Modules		Features				Fusion		F1-score			
	MFA	FS	Occ.	AE	NR	RE	HLC	A-Fusion	Pylon	Powerline	High Veg	Ground
Att 3D U-Net	X	X	✓	✓	✓	✓	X	X	77.0%	91.0%	97.0%	93.0%
RandLA	X	X	X	✓	✓	✓	X	X	83.5%	91.3%	81.5%	95.5%
SUNet	✓	✓	✓	✓	✓	✓	✓	X	87.0%	98.0%	99.0%	100.0%
SUNet	✓	X	✓	✓	✓	✓	✓	X	89.0%	99.0%	99.0%	99.0%
Fusion-SUNet	✓	✓	✓	✓	✓	✓	✓	✓	93.0%	99.0%	99.0%	99.0%

Table 7.8: Comparative study of Fusion-SUNet for Riegl dataset on F1-score.

7.3 Pan-SUNet Performance Evaluation on Riegl Dataset

This section will extensively compare the Pan-SUNet’s performance with the Riegl dataset.

7.3.1 2D Object Detection Results

The assessment for determining the most suitable baseline for 2D object detection, essential for accurately modeling utility corridors, is underpinned by the data presented in Tables 7.9 and 7.10. The tables quantitatively compare the performance of different detection methods, focusing on AP and AR for detecting pylon and span structures using Riegl Data.

Table 7.9 offers a holistic view of the object detection performance, reporting both AP and AR for an IoU threshold of 0.5.

In contrast, Table 7.10 provides a class-specific performance analysis, presenting the AP metric for pylons and spans separately. This class-specific evaluation is crucial for the research focus on modeling the utility corridor through the pylon-to-span relationship, as it directly reflects the network’s precision in detecting the critical components of the infrastructure.

Ablation Setup	Category	Bounding Box	AP	AR
Faster RCNN	2D Two-stage	Horizontal	97.6%	88.4%
YOLO v3	2D Single stage	Horizontal	97.3%	79.2%
DAFNe	2D Anchor-free Single-stage	Horizontal	95.4%	92.2%
Orinted RCNN	2D Oriented Two-stage	Oriented	95.4%	96.9%
Rotated Faster RCNN	2D Oriented Two-stage	Oriented	95.4%	96.9%

Table 7.9: Comparative study for 2D object detection for pylon and span objects for average precision (AP) and average recall (AR) over IoU=0.5.

A closer examination of Tables 7.9 and 7.10 reveals that Oriented RCNN and Rotated Faster RCNN both achieve a perfect AP score of 100% for pylon detection, which is notably superior relative to the 97.6% and 97.3% AP achieved by Faster RCNN and YOLO v3

Ablation Setup	Category	Bounding Box	Average Precision	
			Pylon	Span
Faster RCNN	2D Two-stage	Horizontal	100.0%	96.1%
YOLO v3	2D Single-stage	Horizontal	96.1%	92.2%
DAFNe	2D Anchor-free Single-stage	Horizontal	94.1%	92.9%
Oriented RCNN	2D Oriented Two-stage	Oriented	100.0%	90.8%
Rotated Faster RCNN	2D Oriented Two-stage	Oriented	100.0%	91.0%

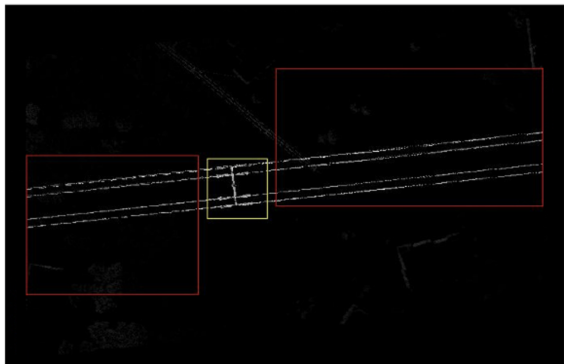
Table 7.10: Class-specific evaluation for 2D object detection of pylon and span objects for average precision (AP) over IoU=0.5.

respectively. However, when considering span detection, Oriented RCNN’s performance exhibits a slight decrease in AP (90.8%) as compared to the 91.0% of Rotated Faster RCNN, suggesting a marginally better alignment with the oriented nature of the spans by the latter.

This quantitative assessment is supported by Figures 7.4c and 7.4d, where the Oriented RCNN and Rotated Faster RCNN depict a more accurate representation of object orientation compared to their horizontal counterparts.

The limitations of horizontal bounding boxes are quantitatively evident: there is a discernible drop in the AR from 96.1% with Rotated Faster RCNN to 88.4% with Faster RCNN (Table 7.9). This decline highlights the ineffectiveness of horizontal boxes in scenarios demanding precise orientation sensitivity, which is further exemplified visually in Figures 7.3 and 7.4b. However, Oriented detectors do not show the highest AP and AR. It shows the best balance between AP and AR. Also, the oriented box is well aligned with our purpose to focus on span and pylon, minimizing the inclusion of background information while aligning it with the network’s orientation. It will enrich the fusion.

A combination of quantitative and visual analyses strongly supports the adoption of Oriented RCNN for future enhancements in Pan-SUNet’s 2D object detection capabilities. This network not only achieves top-tier recall, a critical measure of detection robustness, but also exhibits high precision in object localization.



a) Ground Truth

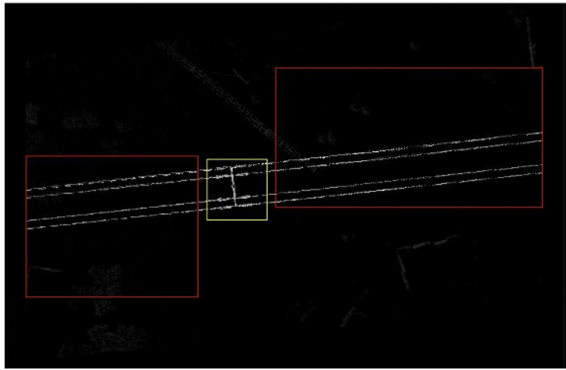


b) Faster RCNN

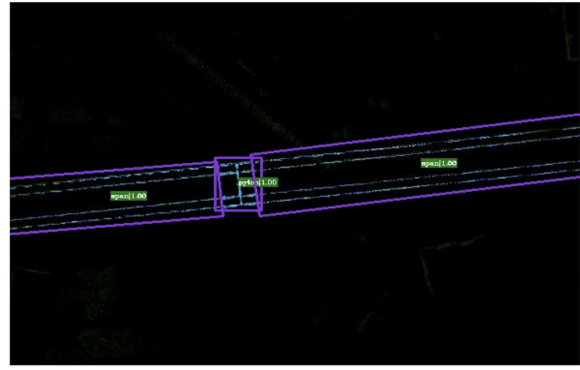


c) Yolov3

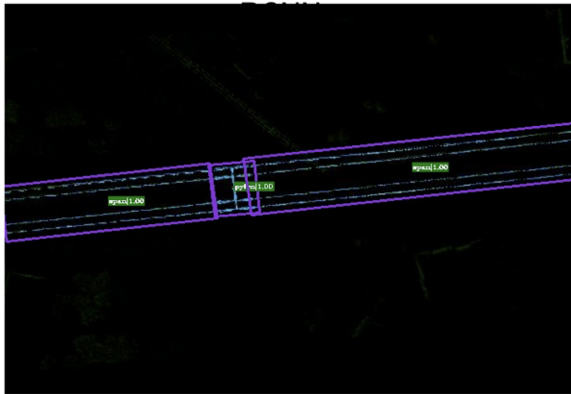
Figure 7.3: Visualization of results of 2D objects span and pylon; a) Ground truth, b) Faster RCNN, and c) YOLO v3.



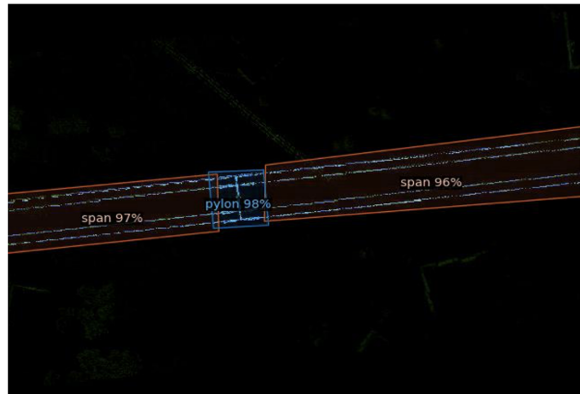
a) Ground Truth



b) Rotated Faster RCNN



c) Oriented RCNN



d) DAFNe

Figure 7.4: Visualization of results of 2D objects span and pylon; a) Ground truth, b) Rotated Faster RCNN, c) Oriented RCNN, and d) DAFNe.

Methods	Dataset	Corridor		
		Precision	Recall	F1-score
2D Regional Segmentation	R	86%	97%	91%

Table 7.11: The results of regional segmentation results for corridor class in the Riegl dataset.

7.3.2 2D Regional Segmentation Results

Our spatial hierarchy framework includes a 2D regional semantic segmentation specifically designed to integrate human-like spatial awareness and contextual reasoning into the proposed deep learning networks, including the Pan-SUNet. This segmentation approach is particularly useful for delineating corridors, which is crucial for analyzing and vectorizing powerlines and pylons.

The segmentation method used in this study efficiently identifies corridor regions within the Riegl dataset. It demonstrates both high processing speed and accuracy. The performance metrics provided indicate the effectiveness of this model. It achieves a precision rate of 86%, meaning that a high proportion of predicted corridors are correct. The network’s recall rate is an impressive 97%, showing that it can correctly identify nearly all relevant corridor features within the dataset. The F1-score is 91%, indicating that the model performs well overall for the corridor class by balancing precision and recall.

Table 7.11 consolidates the performance indicators for the corridor class in the Riegl dataset, segmented by region. The presented data in the table confirms that the segmentation network is suitable for independent use, demonstrating its operational effectiveness in this specialized application.

7.3.3 3D Semantic Segmentation Results

Figure 7.5 illustrates the ground truth and Pan-SUNet’s capability to distinguish between corridor, non-corridor, and utility zones, showing the contrast between the actual classifications and Pan-SUNet’s predictions. Similarly, Figure 7.6 provides a comparative visualization of

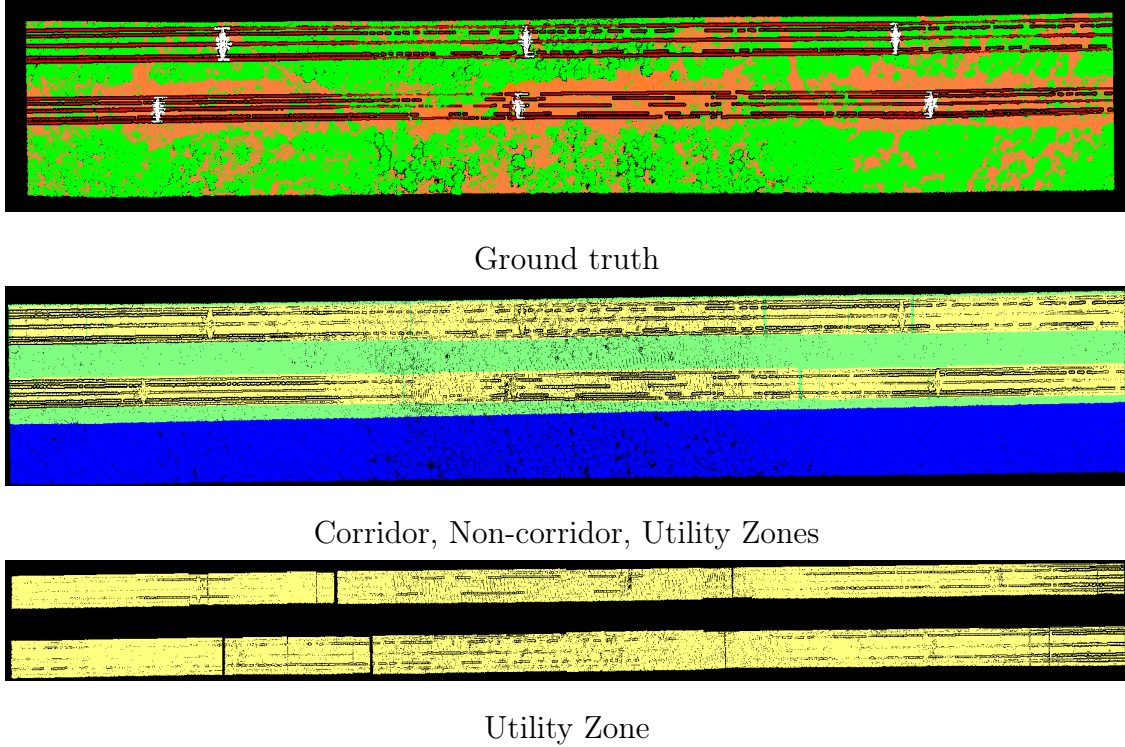
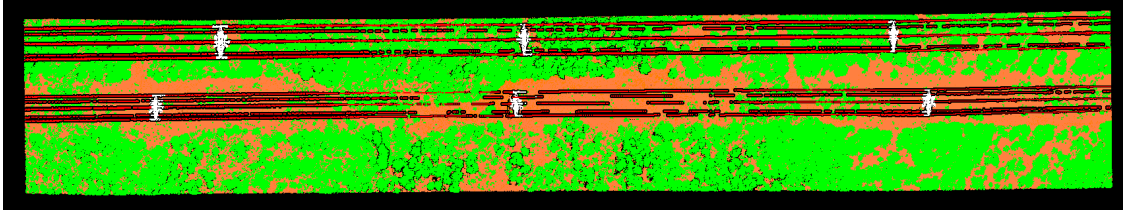


Figure 7.5: Visualization of Pan-SUNet’s segmentation performance on the Riegl dataset: ground truth and three Regions (blue: non-corridor, green: corridor, and yellow: utility zone).

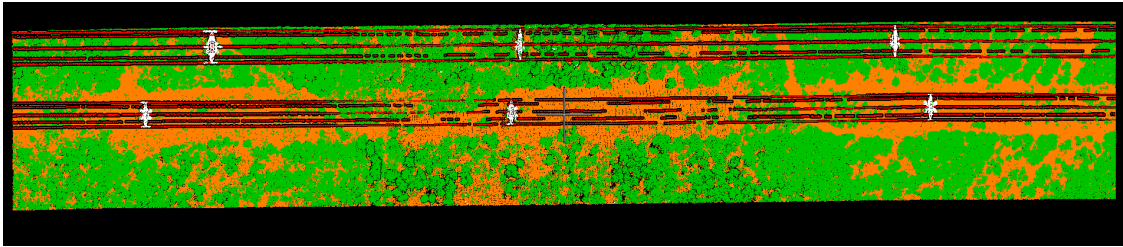
the ground truth versus Pan-SUNet’s semantic and panoptic segmentation abilities, effectively showcasing the model’s precision in identifying pylons, powerlines, high vegetation, and ground.

Ablation Study

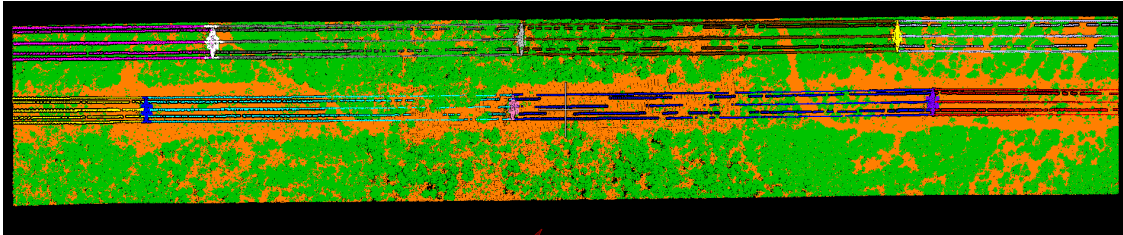
A comprehensive ablation study was conducted to understand the impact of different network architectures and fusion techniques on the Riegl dataset. The study compared the baseline models - 3D U-Net, Attention 3D U-Net, and pre-trained RandLA - with enhanced versions featuring loss-based and panoptic fusion. The results, as shown in Tables 7.12, 7.13, and 7.14, demonstrate that these improvements led to noticeable gains in performance. It is worth noting that RandLA, which has only one voltage wire and pylon context, achieved performance improvements that were very similar to those of the Attention 3D U-Net baseline



Ground truth.



Pan-SUNet Semantic Segmentation.



Pan-SUNet Panoptic Segmentation.

Figure 7.6: Comparative results of ground truth vs Pan-SUNet’s semantic segmentation and Pan-SUNet’s Panoptic segmentation results for utility corridor scene: blue: pylon, red: powerline, green: high vegetation and orange: ground, and Pan-SUNet Panoptic segmentation.

model.

In terms of numbers, using Attention 3D U-Net with fusion strategies resulted in a significant improvement. The F1-score for the pylon class increased by 19% and for powerlines by 8.5% when compared to the baseline 3D U-Net without fusion, as shown in Table 7.14. Moreover, the precision for pylons increased from 84% to 99%, indicating the proposed network’ enhanced accuracy.

	Ablation Settings					Recall			
	Baseline	Fusion		2D Regions	Dataset	Pylon	Powerline	High Veg	Ground
		HLC	Panoptic						
Pan-SUNet	3D U-Net	✗	✗	0	R	70.0%	84.0%	94.0%	93.0%
	Att- 3D U-Net	✗	✗	0	R	72.0%	84.0%	97.0%	93.0%
	RandLA	✗	✗	0	R	75.0%	86.5%	77.0%	95.0%
	3D U-Net	✓	✗	3	R	74.0%	85.0%	94.0%	93.0%
	Att- 3D U-Net	✓	✗	3	R	86.0%	98.0%	99.0%	99.0%
	RandLA	✓	✗	3	R	77.0%	95.5%	95.0%	97.0%
	3D U-Net	✓	✓	3	R	80.0%	95.0%	99.0%	98.0%
	Att- 3D U-Net	✓	✓	3	R	90.0%	98.0%	99.0%	99.0%
	RandLA	✓	✓	3	R	89.0%	97.5%	97.0%	98.0%

Table 7.12: Riegl dataset’s ablation study of Pan-SUNet performance comparison for Recall.

	Ablation Settings					Precision			
	Baseline	Fusion		2D Regions	Dataset	Pylon	Powerline	High Veg	Ground
		HLC	Panoptic						
Pan-SUNet	3D U-Net	✗	✗	0	R	80.0%	97.0%	90.0%	92.0%
	Att- 3D U-Net	✗	✗	0	R	84.0%	98.0%	97.0%	94.0%
	RandLA	✗	✗	0	R	92.0%	99.0%	87.0%	96.0%
	3D U-Net	✓	✗	3	R	90.0%	97.0%	90.0%	95.0%
	Att- 3D U-Net	✓	✗	3	R	93.0%	99.0%	99.0%	99.0%
	RandLA	✓	✗	3	R	94.0%	99.0%	98.0%	99.0%
	3D U-Net	✓	✓	3	R	96.0%	97.0%	95.0%	97.0%
	Att- 3D U-Net	✓	✓	3	R	99.0%	100.0%	99.0%	99.0%
	RandLA	✓	✓	3	R	98.0%	99.0%	98.0%	99.0%

Table 7.13: Riegl dataset’s ablation study of Pan-SUNet to compare performance on Precision.

Comparative Study

Further comparative analysis evaluated the performance of Pan-SUNet for the principal task of semantic segmentation against other SUNet variants, including Attention 3D U-Net and RandLA. For instance, the implementation of Attention 3D U-Net with 3 region fusion

	Ablation Settings					F1-score			
	Baseline	Fusion		2D Regions	Dataset	Pylon	Powerline	High Veg	Ground
		HLC	Panoptic						
Pan-SUNet	3D U-Net	$\times$	$\times$	0	R	75.0%	90.5%	92.0%	92.5%
	Att- 3D U-Net	$\times$	$\times$	0	R	77.0%	91.0%	97.0%	93.0%
	RandLA	$\times$	$\times$	0	R	83.5%	91.3%	81.5%	95.5%
	3D U-Net	$\checkmark$	$\times$	3	R	82.0%	91.0%	92.0%	94.0%
	Att- 3D U-Net	$\checkmark$	$\times$	3	R	86.0%	98.0%	99.0%	99.0%
	RandLA	$\checkmark$	$\times$	3	R	85.5%	97.25%	96.0%	98.0%
	3D U-Net	$\checkmark$	$\checkmark$	3	R	88.0%	95.0%	99.0%	97.0%
	Att- 3D U-Net	$\checkmark$	$\checkmark$	3	R	94.0%	99.0%	99.0%	99.0%
	RandLA	$\checkmark$	$\checkmark$	3	R	93.0%	98.2%	97.0%	98.0%

Table 7.14: Riegl dataset’s ablation study of Pan-SUNet to compare performance on F1-score.

Methods	Dataset	2D Regions	F1-score			
			Pylon	Powerline	High Veg	Ground
3D U-Net	Riegl	0	75.0%	90.5%	92.0%	92.5%
Att- 3D U-Net	Riegl	0	77.0%	91.0%	97.0%	93.0%
RandLA	Riegl	0	83.5%	91.3%	81.5%	95.0%
SUNet	Riegl	2	89.0%	99.0%	99.0%	99.0%
Fusion-SUNet	Riegl	2	93.0%	99.0%	99.0%	99.0%
Pan-SUNet	Riegl	3	94.0%	99.0%	99.0%	99.0%

Table 7.15: Riegl dataset’s cross network comparative study with Pan-SUNet on F1-score.

increased the F1-score for pylons by 19%, indicating improved detection of these critical infrastructure components.

Pan-SUNet’s voxel-based approach was benchmarked against point-based networks, revealing that while the former offers competitive performance, it does so with an increase in inference speed as point-based RandLA takes 400s to process 1km strip of point cloud and voxel-based network takes 50s. Table 7.13 highlights this point, highlighting that the voxel-based network not only matches but often surpasses the precision of point-based methods—achieving a near-perfect precision of 100% for powerlines when enhanced with Attention 3D U-Net and fusion, a testament to its accuracy.

Moreover, existing commercial software that relies on network maps or manual refinement and labeling for pylon and powerlines, Pan-SUNet showcases exemplary performance in autonomously identifying powerlines and pylons with minimal reliance on external data

Methods	Category	Bounding Box	AP	AR	Average Precision	
					Pylon	Span
Pan-SUNet	2D	Oriented	95.4%	96.9%	100.0%	90.8%
Pan-SUNet	3D	Oriented	97.4%	95.9%	99.0%	92.0%

Table 7.16: Comparative performance matrices of Pan-SUNet for 2D and 3D object detection on Riegl dataset: highlighting overall AP, AR, and class-wise AP for pylon and span detection at IoU=0.5.

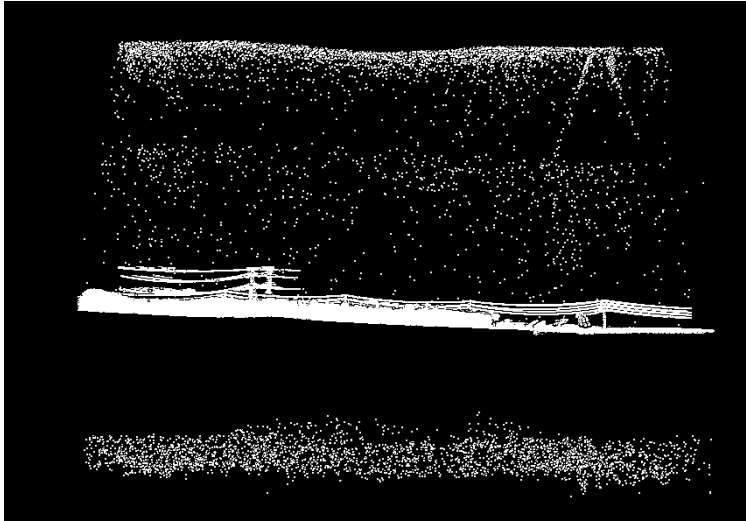
inputs as shown in Table 7.15. In summary, the experimental results of Pan-SUNet show the efficiency and effectiveness for 3D semantic segmentation.

7.3.4 3D Object Detection Results

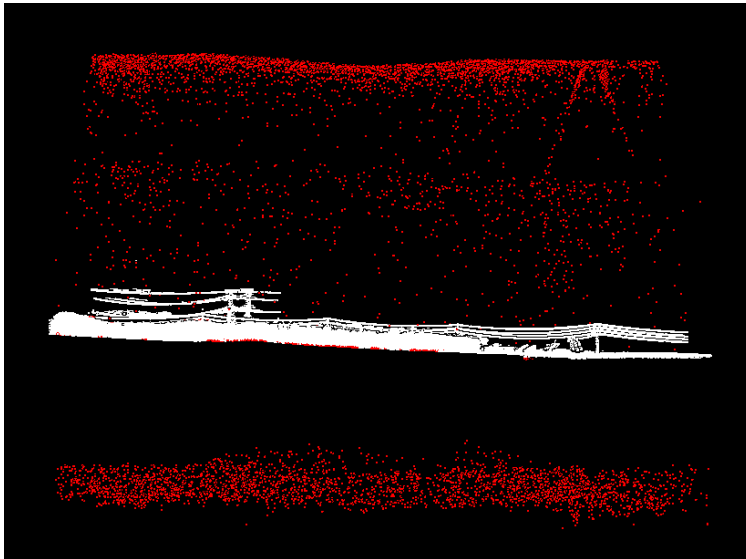
Using Pan-SUNet for object detection makes it easier to transform 2D detections into useable 3D spatial models, which is important for improving utility infrastructure predictive maintenance plans and transmission corridor analysis in-depth. Performance study in Table 7.16 shows that Pan-SUNet can adapt to a more complicated 3D spatial context, as evidenced by its remarkable AP of 97.4% for pylons, and its remarkable precision in 2D object identification, especially for pylons, with an AP of 95.4%. Pan-SUNet’s balanced results validate its effectiveness in semantic segmentation and object detection tasks, even though 3D span detection performance slightly declined. This puts it in a position to be further tested with Teledyne Optech’s more challenging dataset to improve performance in complicated settings.

7.4 Pan-SUNet Performance: Teledyne Galaxy T1000

In this section, a new dataset acquired by Teledyne Optech’s Galaxy T1000, explained in 6.1.2, has been used for evaluating Pan-SUNet’s performance. This dataset contains considerable amounts of noises induced by the MPIA, which has been removed by applying the NSANet



(a) Initial point cloud with ambient noise.



(b) Noise points detection highlighted by NSANet, marked with a red color.



(c) Cleaned point cloud after NSANet noise reduction.

Figure 7.7: Noise removal process in Teledyne Optech T1000 dataset using NASNet.

Methods	Category	Bounding Box	AP	AR
Rotated Faster RCNN	2D Oriented Two-stage	Oriented	65.8%	75.5%
RoI Transformer	2D Oriented Two-stage	Oriented	72.3%	76.7%
Orinted RCNN	2D Oriented Two-stage	Oriented	85.6%	88.0%

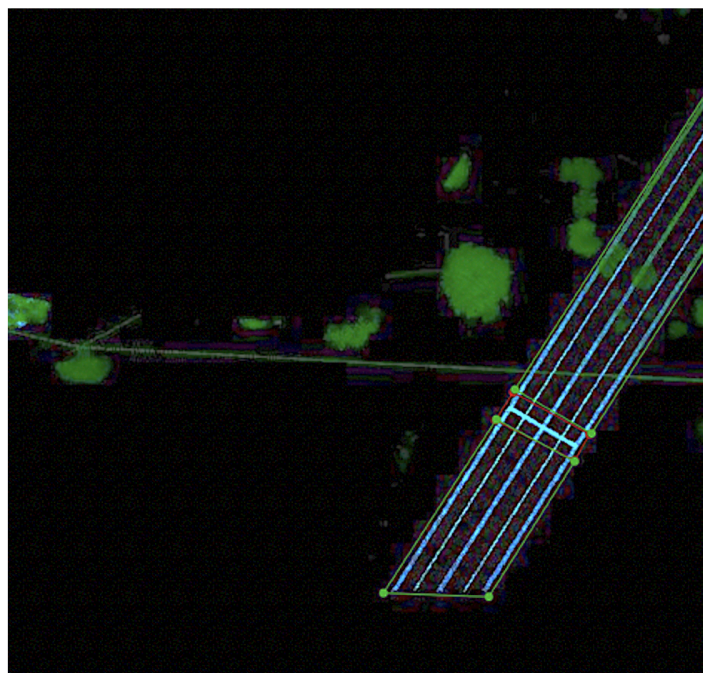
Table 7.17: Comparative study for 2D object detection for Teledyne dataset for pylon and span objects for average precision (AP) and average recall (AR) over IoU=0.5.

proposed in Chapter 4. Figure 7.7 shows the raw dataset that includes noise points, and the classification of noise points by NSANet, marked in red. The figure also presents a final filtered dataset after noise removal, in which only the object points are preserved. The Galaxy dataset, which was free of any noise, was used as input for evaluating the performance of Pan-SUNet.

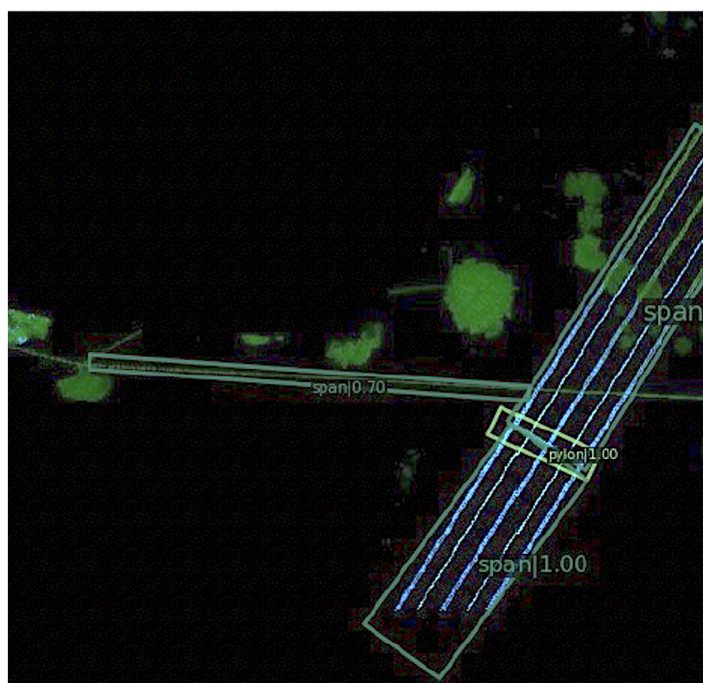
7.4.1 2D Object Detection Results

Based on the analysis performed, it has been determined that using oriented bounding boxes is the most effective method for detecting pylons and spans in electrical transmission corridors within the Riegl dataset. To confirm this, a detailed comparative study was conducted using data compiled by the Network Mapping (NM) Group, a U.K.-based geospatial service company specializing in utility network asset management. Three state-of-the-art-oriented object detection networks were chosen based on their performance on the DOTA benchmark to compare their performance to Pan-SUNet. The analytical findings are summarized in Tables 7.17 and 7.18, and visual validations of detection efficiency are depicted in Figures 7.8, 7.9, and 7.10.

With an AP of 85.6% and AR of 88.0%, the Oriented RCNN performs better in pylon detection than both the Rotated Faster RCNN and the ROI Transformer (Table 7.17). Furthermore, Oriented RCNN performs better than Rotated Faster RCNN, this is roughly 19.8% higher AP for pylons and 19.6% higher AP for spans (Table 7.18). The superior



Groundtruth



Prediction

Figure 7.8: Teledyne dataset Visualization of results of 2D objects span and pylon for high voltage span and pylon.

Methods	Category	Bounding Box	Average Precision	
			Pylon	Span
Rotated Faster RCNN	2D Oriented Two-stage	Oriented	69.7%	61.5%
RoI Transformer	2D Oriented Two-stage	Oriented	78.2%	66.4%
Oriented RCNN	2D Oriented Two-stage	Oriented	89.5%	81.1%

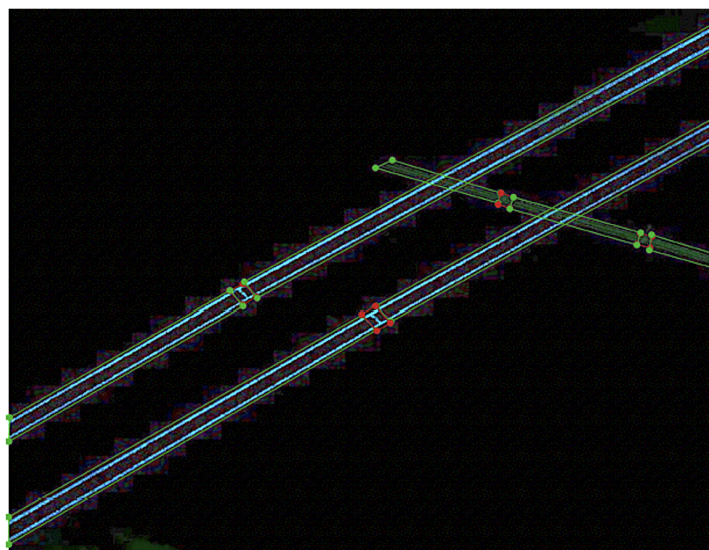
Table 7.18: Comparative study for 2D object detection for Teledyne dataset for pylon and span objects for pylon average precision (AP) and span average precision (AP).

performance of Oriented RCNN is further supported by visual comparisons in Figure 7.10, particularly in difficult settings. Oriented RCNN, chosen as the baseline, shows success in reliably detecting objects in complicated circumstances, despite difficulties in Pan-SUNet’s performance on the Teledyne dataset. This establishes its position for further trials inside the Pan-SUNet framework for utility corridors.

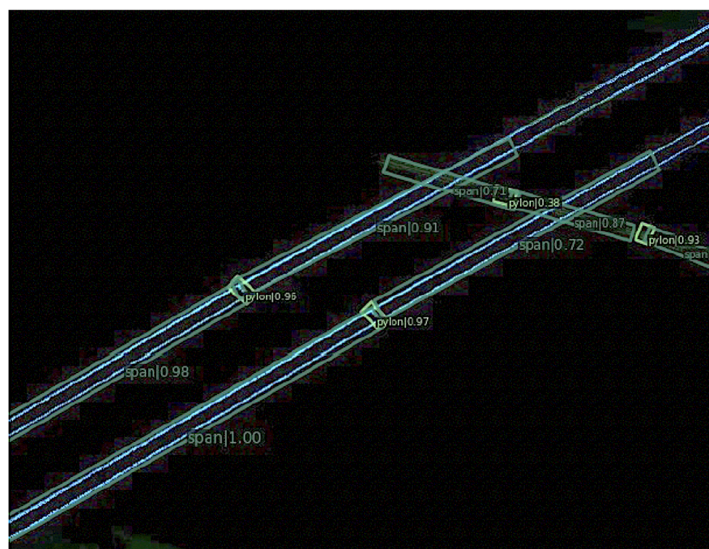
7.4.2 2D Regional Segmentation Results

In this section, the performance of the Teledyne Optech Galaxy T1000 and the Riegl Q560 is compared for 2D regional segmentation tasks. Table 7.19 shows that the Teledyne Optech Galaxy T1000 outperforms the Riegl Q560 (88.1%) with a significantly higher overall accuracy (OA) of (95.0%), almost reaching the maximum score. Teledyne Optech’s dataset performs better than Riegl’s dataset due to fewer instances of high vegetation in non-corridor zones, which do not confuse the classification model with similar structural features to pylons. Therefore, the corridor class has a more balanced precision and recall, leading to increased OA.

Teledyne Optech Galaxy T1000 segmentation results for the corridor class are shown in Table 7.19. Pan-SUNet’s efficiency is demonstrated with precision of 92.0%, recall of 89.0%, and an F1-score of 91.0%, indicating a strong balance between precision and recall and highlighting the model’s potential for accurate corridor space delineation. These results highlight the usefulness of the Teledyne Optech Galaxy T1000 in 2D regional segmentation,

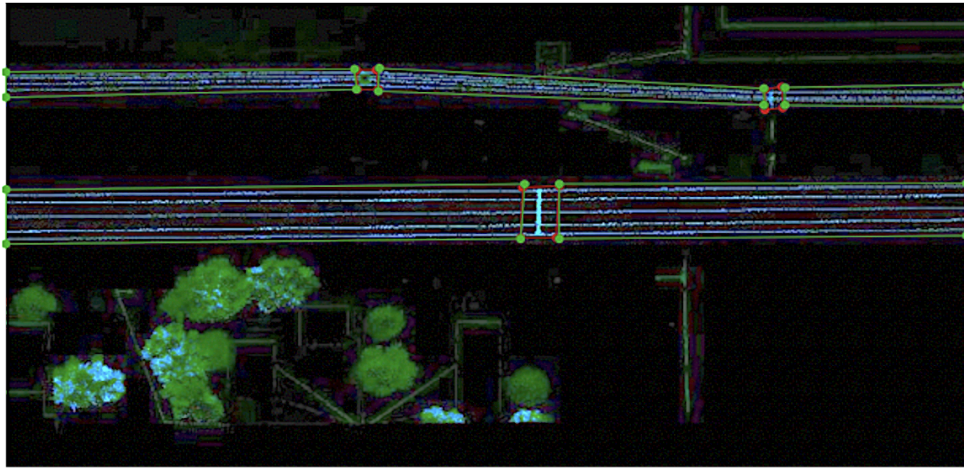


Groundtruth

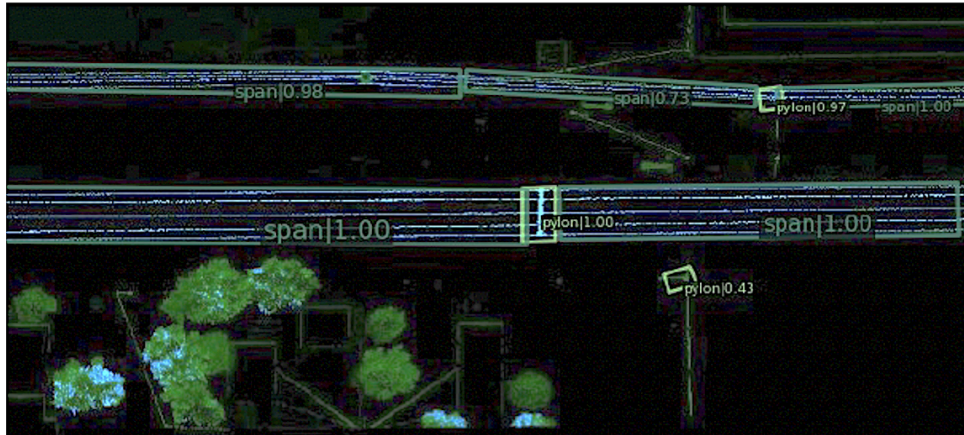


Prediction

Figure 7.9: Teledyne dataset Visualization of results of 2D objects span and pylon for overlapping wires.



Groundtruth



Prediction

Figure 7.10: Teledyne dataset Visualization of results of 2D objects span and pylon for high voltage transmission and distribution spans and pylon.

Methods	Dataset	OA	Corridor		
			Precision	Recall	F1-score
2D Regional Segmentation	T	95.0%	92.0%	89.0%	91.0%
2D Regional Segmentation	R	88.1%	86.0%	97.0%	91.0%

Table 7.19: Comparative evaluation of 3D regional segmentation using Teledyne Optech’s Galaxy T1000 and Riegl datasets.

which is essential for applications involving precise mapping and provides enhanced precision for later phases of data processing.

7.4.3 3D Semantic Segmentation Results

Accurate 3D semantic segmentation of utility corridors is essential for efficient management and upkeep of powerline networks. Research employing the Pan-SUNet and Riegl datasets has demonstrated encouraging outcomes in semantic segmentation. The Teledyne Optech Galaxy T1000 dataset introduces complexity because of the different transmission line voltages from low-voltage distribution lines (4kV-12kV) to high-voltage transmission lines (275kV-400kV), while the Riegl Q560 dataset provides a simpler scenario with constant voltage levels (230kV), posing difficulties in segmenting smaller pylons and lower voltage lines that demand accuracy and detail. The next part examines Pan-SUNet’s 3D semantic segmentation outcomes using the Teledyne Optech Galaxy T1000 dataset and contrasts them with the dataset from Riegl Q560.

Ablation Study

In this section, the impact of integrating HLC and Panoptic Fusion on the performance metrics of Pan-SUNet is assessed. This thesis evaluated its influence on recall, precision, and F1-score against established models such as 3D U-Net, Attention U-Net, and RandLA. You can find the results in Tables 7.20, 7.21, and 7.22.

- Impact on Recall: Baseline models show variable performance across classes, with the 3D U-Net having 66% recall for Pylons and 50% for Buildings. Integration of HLC

Pan-SUNet	Ablation Settings					Recall						
	Baseline	Fusion		2D Regions	Dataset	Pylon	Powerline	Hig Veg	Medium Veg	Others	Ground	Building
		HLC	Panoptic									
Pan-SUNet	3D U-Net	✗	✗	0	T	66.0%	97.0%	83.0%	92.0%	59.0%	98.0%	50.0%
	Att-3D U-Net	✗	✗	0	T	61.0%	94.0%	91.0%	85.0%	37.0%	98.0%	39.0%
	RandLA	✗	✗	0	T	78.0%	96.0%	96.0%	97.0%	83.0%	98.0%	78.0%
	3D U-Net	✓	✗	3	T	66.0%	97.0%	91.0%	92.0%	63.0%	98.0%	50.0%
	Att-3D U-Net	✓	✗	3	T	61.0%	94.0%	97.0%	85.0%	60.0%	98.0%	41.0%
	RandLA	✓	✗	3	T	78.0%	96.0%	96.0%	97.0%	83.0%	98.0%	78.0%
	3D U-Net	✓	✓	3	T	66.0%	97.0%	91.0%	92.0%	63.0%	98.0%	50.0%
	Att-3D U-Net	✓	✓	3	T	86.0%	94.0%	97.0%	85.0%	60.0%	98.0%	41.0%
	RandLA	✓	✓	3	T	83.0%	96.0%	96.0%	97.0%	83.0%	98.0%	78.0%

Table 7.20: Recall metrics comparison for Pan-SUNet with ablation setups: three baline networks, HLC, Panoptic fusion integration, 2D regional segmentation and datasets.

Pan-SUNet	Ablation Settings					Precision						
	Baseline	Fusion		2D Regions	Dataset	Pylon	Powerline	Hig Veg	Medium Veg	Others	Ground	Building
		HLC	Panoptic									
Pan-SUNet	3D U-Net	✗	✗	0	T	32.0%	87.0%	80.0%	77.0%	43.0%	98.0%	83.0%
	Att-3D U-Net	✗	✗	0	T	27.0%	65.0%	92.0%	88.0%	27.0%	98.0%	87.0%
	RandLA	✗	✗	0	T	84.0%	90.0%	95.0%	91.0%	71.0%	99.0%	97.0%
	3D U-Net	✓	✗	3	T	58.0%	96.0%	91.0%	77.0%	39.0%	98.0%	83.0%
	Att-3D U-Net	✓	✗	3	T	50.0%	96.0%	95.0%	88.0%	22.0%	98.0%	88.0%
	RandLA	✓	✗	3	T	86.0%	93.0%	95.0%	91.0%	74.0%	99.0%	97.0%
	3D U-Net	✓	✓	3	T	76.0%	96.0%	99.0%	77.0%	39.0%	98.0%	83.0%
	Att-3D U-Net	✓	✓	3	T	77.0%	96.0%	97.0%	88.0%	22.0%	98.0%	88.0%
	RandLA	✓	✓	3	T	86.0%	97.0%	95.0%	91.0%	74.0%	99.0%	97.0%

Table 7.21: Precision metrics comparison for Pan-SUNet with ablation setups: three baline networks, HLC, Panoptic fusion integration, 2D regional segmentation and datasets.

marks improvement, and with Panoptic Fusion, significant recall increases are observed, such as in Attention U-Net for Pylons, rising to 86%, and in RandLA increasing from 78% to an enhanced 83%.

Pan-SUNet	Ablation Settings					F1-score						
	Baseline	Fusion		2D Regions	Dataset	Pylon	Powerline	Hig Veg	Medium Veg	Others	Ground	Building
		HLC	Panoptic									
	3D U-Net	✗	✗	0	T	43.0%	91.0%	81.5%	84.0%	50.0%	98.0%	62.0%
	Att-3D U-Net	✗	✗	0	T	37.0%	77.0%	91.5%	86.0%	37.0%	98.0%	54.0%
	RandLA	✗	✗	0	T	80.0%	93.0%	98.0%	93.0%	75.0%	98.0%	86.0%
	3D U-Net	✓	✗	3	T	62.0%	96.0%	91.0%	84.0%	48.0%	98.0%	63.0%
	Att-3D U-Net	✓	✗	3	T	55.0%	95.0%	96.0%	86.0%	32.0%	98.0%	56.0%
	RandLA	✓	✗	3	T	82.0%	95.0%	98.0%	93.0%	78.0%	98.0%	88.0%
	3D U-Net	✓	✓	3	T	71.0%	96.0%	95.0%	84.0%	48.0%	98.0%	63.0%
	Att-3D U-Net	✓	✓	3	T	81.0%	95.0%	97.0%	86.0%	32.0%	98.0%	56.0%
	RandLA	✓	✓	3	T	85.0%	97.0%	98.2%	93.0%	78.3%	98.0%	88.0%

Table 7.22: F1-score metrics comparison for Pan-SUNet with ablation setups: three baline networks, HLC, Panoptic fusion integration, 2D regional segmentation and datasets.

- Impact on Precision: For baseline models without HLC and Panoptic Fusion, precision metrics differ significantly; 3D U-Net for Pylons posts 32%. Results are improved by including HLC, and major improvements are observed with Panoptic Fusion. For example, Attention 3D U-Net’s precision for Pylons increases from 27% to 77%, and RandLA for Others climbs from 71% to 74%.
- Impact on F1-score: 3D U-Net shows a low F1-score of 43% for pylon. However, when HLC is included, the performance of the models improves. With the addition of Panoptic Fusion, the models show substantial increases in F1-score. For example, Attention 3D U-Net’s F1-score in the Pylon class improves from 37% to 81%. RandLA also benefits significantly, particularly in the Others class, where F1-score increases from 75% to 78.3%.

According to quantitative assessments, all models perform much better when both HLC and Panoptic Fusion are used. Notable improvements are seen in recall for Pylons and precision for Others, especially in Attention U-Net and RandLA models. The relative metric

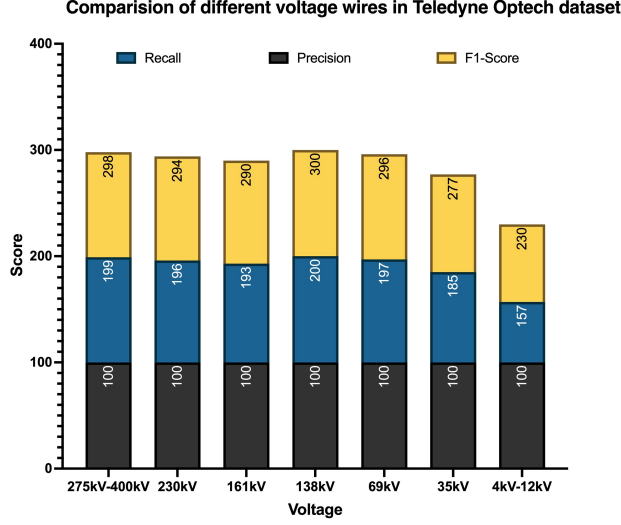


Figure 7.11: Performance comparison of wires of different voltage in Teledyne Optech dataset. The results reflect the lower performance on low voltage powerlines.

studies, which provide remarkable percentage point gains in precision and F1-scores across all models, further highlight the efficacy of these integrations.

This thorough evaluation demonstrates how using Panoptic Fusion and HLC can significantly improve segmentation tasks and improve the performance of cutting-edge segmentation networks using Teledyne Optech datasets.

Qualitative Analysis

In this study, a qualitative analysis of the performance of Pan-SUNet is conducted, which has been augmented with RandLA and various attention mechanisms. Our focus has been on different utility and urban infrastructure elements, and we have specifically explored the detection and classification of voltage power lines of varying sizes, the size of pylons, and the segmentation of buildings in urban scenes.

- **Robustness in High-Voltage Wire Detection:** As seen in Figure 7.14, the network demonstrates remarkable accuracy in recognizing high-voltage power lines with three-phase wires and a ground wire, retaining accuracy even when the ground wire is close

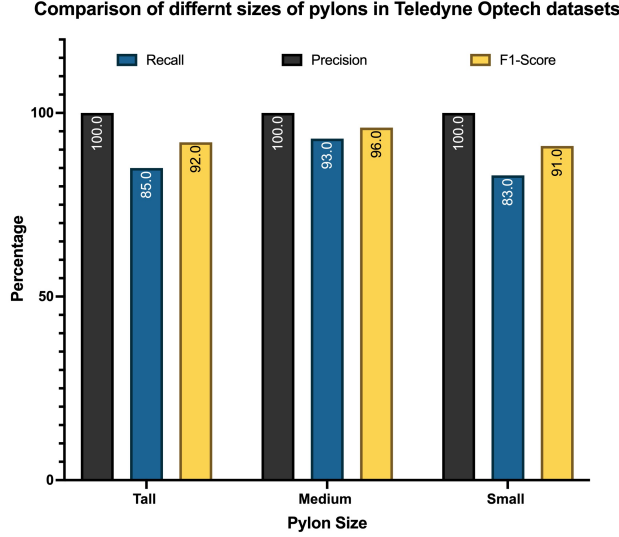


Figure 7.12: Performance comparison of pylons of different heights in Teledyne Optech dataset. The results reflect the lower performance on small size pylons.

to the ground. This demonstrates the network’s strong segmentation capabilities for high-voltage power lines.

- **Challenges in Low-Voltage Wire and Small Pylon Detection:** As Figure 7.13 shows, the network performs noticeably worse when it comes to low-voltage lines (4kV to 12kV) and small pylon structures. This graphic depiction correlates with the numerical data in Figures 7.11 and 7.12, providing information about the network’s difficulties in precisely handling these complex infrastructure elements.
- **Generalization Issues with Building Segmentation:** Building segmentation presents a difficulty to the network, particularly for structures with flat surfaces, as Figure 7.15 illustrates. An examination in comparison with other urban features, such as dense vegetation, highlights the necessity of enhancing the network’s capacity for generalization, underscoring the significance of more accurately differentiating building structures in intricate urban environments.

The comprehensive evaluation results highlight the strengths and weaknesses of Pan-

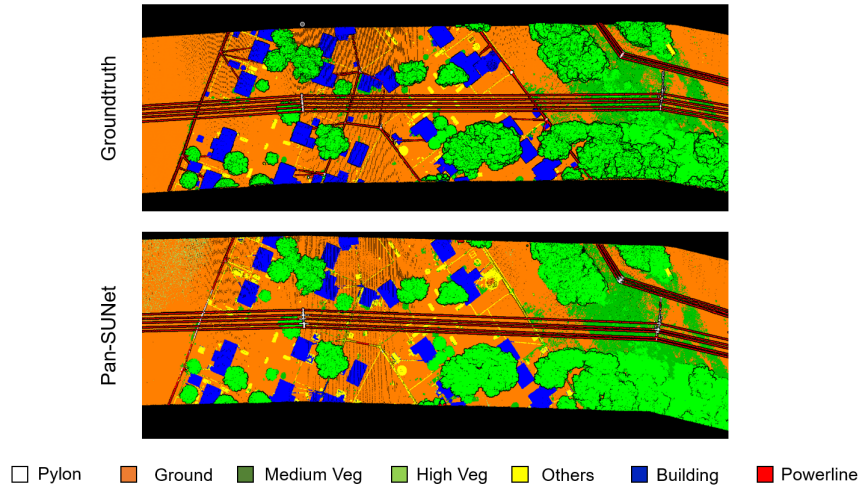


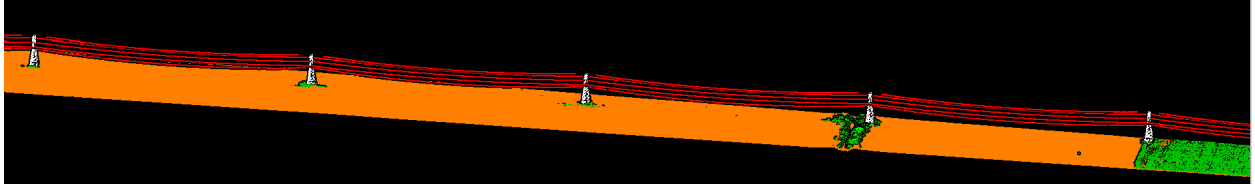
Figure 7.13: Visualization showing the problematic areas of low voltage powerline being misclassified while high voltage powerline are correctly classified.

SUNet’s operations. The detection of high-voltage wires is found to be a strong feature, while the identification of low-voltage wires and small pylons, and accurate segmentation of buildings are areas that require further improvement. The insights gained from the evaluation demonstrate the potential of Pan-SUNet in critical applications and emphasize the need to enhance its performance in analyzing urban infrastructure under different scenarios.

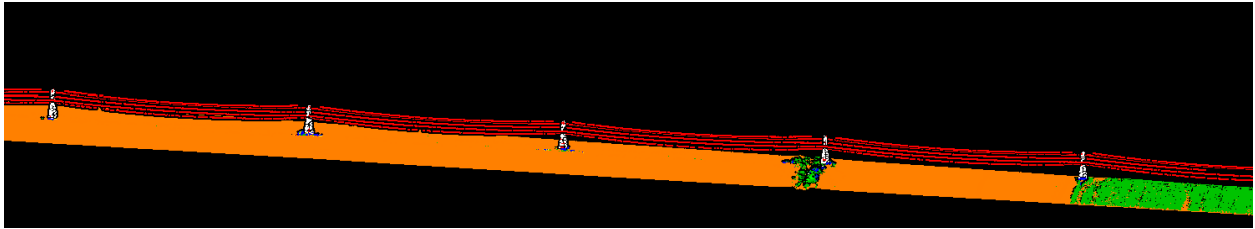
Comparative Study

The results of Pan-SUNet for the Teledyne dataset were compared with 3D U-Net, Attention 3D U-Net, RandLA, and Pan-SUNet with the Riegl dataset. This comparison highlighted the performance disparities and the challenges posed by the complexity of the Teledyne dataset as shown in Figure 7.13. The study also demonstrated that dealing with the diverse range of transmission voltages in high-density point clouds presents significant difficulties for voxel-based networks as shown in Figure 7.15. Consequently, a point-based network as a baseline is essential as shown in Table 7.23.

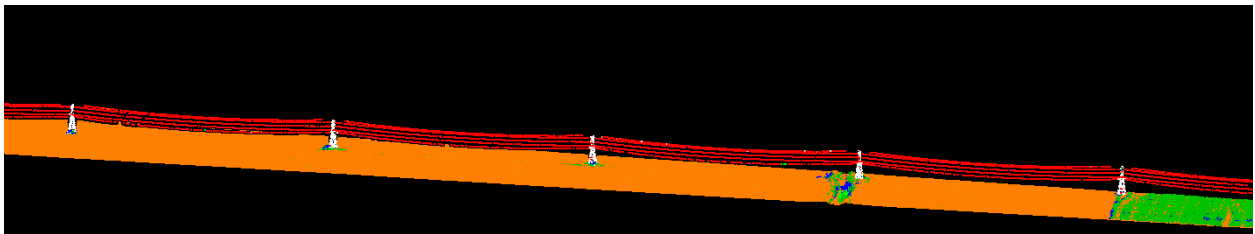
Furthermore, this research illustrated that Panoptic fusion and segmentation fusion have the potential to enhance the outcomes of any existing baseline as shown in Figure 7.18. This finding underscores the network-agnostic capabilities of the proposed system.



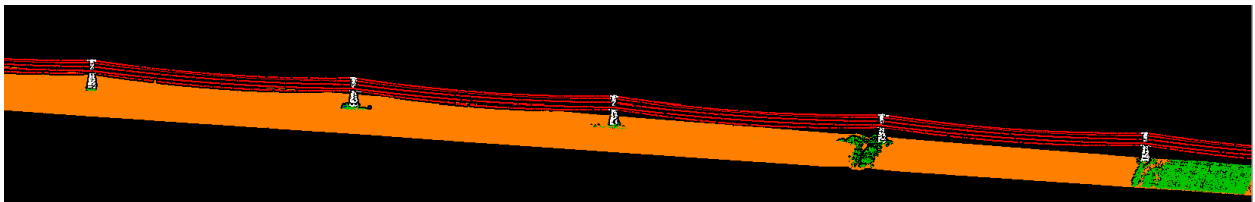
(a) Ground truth



(b) Results of Pan-SUNet with 3D U-Net as baseline.



(c) Results of Pan-SUNet with Attention 3D U-Net as baseline.



(d) Results of Pan-SUNet with RandLA as baseline.

Figure 7.14: Visualization of high voltage powerline and tall steel pylons reflecting the ease of segmentation for Pan-SUNet irrespective of baseline.

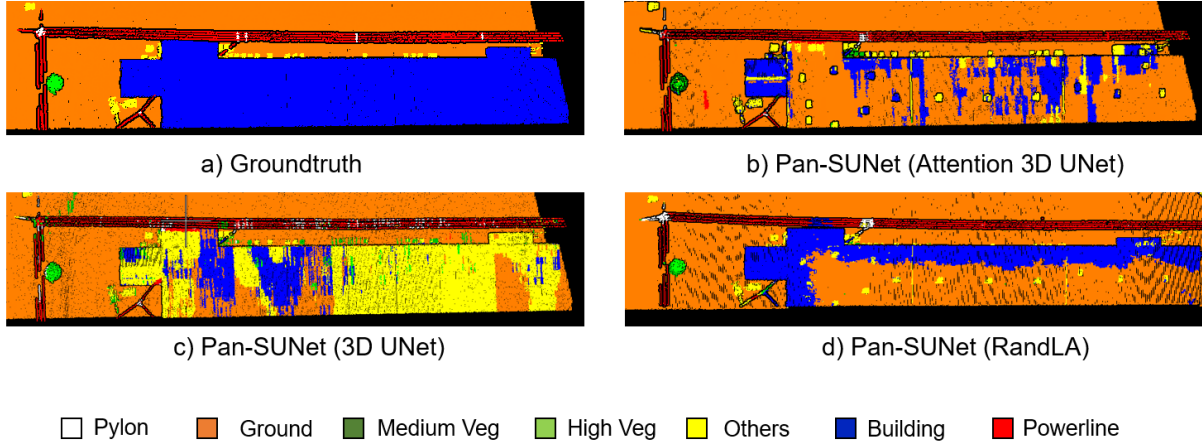


Figure 7.15: Comparison of the performance between the building class and low voltage powerline highlights the challenge in generalization. This disparity reflects the degradation in performance and misclassification issues, especially for buildings with flat surfaces.

Methods	Dataset	2D Regions	F1-score						
			Pylon	Powerline	Hig Veg	Medium Veg	Others	Ground	Building
3D U-Net	Teledyne	0	43.0%	91.0%	81.5%	84.0%	50.0%	98%	62.0%
Att- 3D U-Net	Teledyne	0	37.0%	77.0%	91.5%	86.0%	37.0%	98.0%	54.0%
RandLA	Teledyne	0	80.0%	93.0%	98.2%	93.0%	75.0%	98.0%	86.0%
Pan-SUNet	Teledyne	3	85.0%	97.0%	98.2%	93.0%	78.3%	98.0%	88.0%
Pan-SUNet	Riegl	3	94.0%	99.0%	99.0%	-	-	99.0%	-

Table 7.23: Teledyne dataset's comparative study with Pan-SUNet on F1-score.

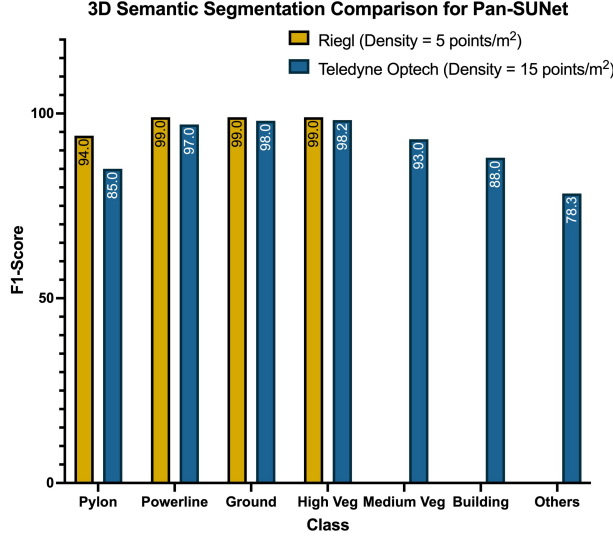
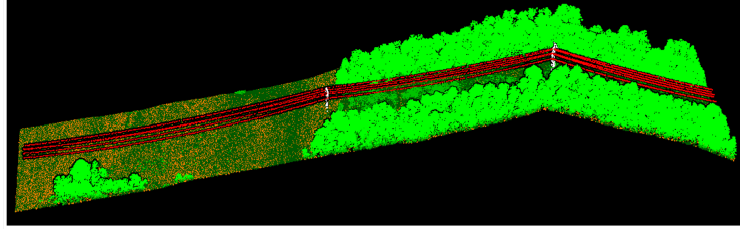


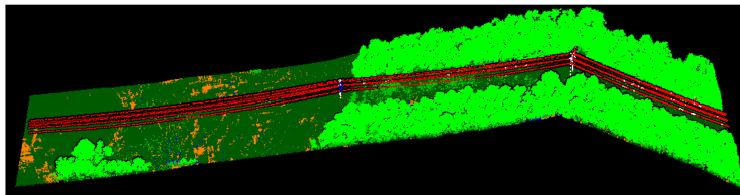
Figure 7.16: F1-score comparison for object detection on Teledyne vs. Riegl datasets with varied point densities.

7.4.4 3D Object Detection Results

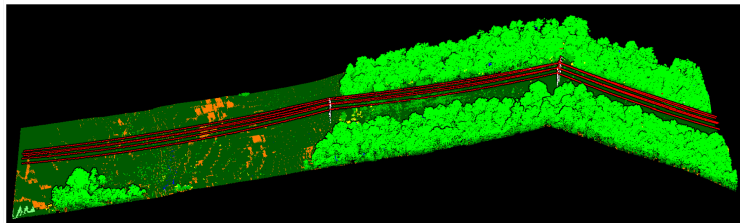
A comprehensive comparison evaluated Pan-SUNet’s effectiveness using Teledyne and Riegl datasets against baselines, including 3D U-Net, Attention 3D U-Net, RandLA, and different Pan-SUNet variations. The analysis highlighted relative performance strengths and challenges in processing complex, high-density point clouds from the Teledyne dataset (Figure 7.16). Quantitative data underscored Pan-SUNet’s superior performance in both 2D and 3D oriented bounding box contexts, revealing marked improvements in AP for pylons and span detections compared to other baselines. In 2D, Pan-SUNet achieved an AP of 89.5% for pylon detection and 81% for span detection, slightly improving to 89.5% and 84% in 3D, indicating sophisticated advancements introduced by 3D orientation. The comparison highlights the struggle of voxel-based network approaches in handling intricate details of transmission voltage variations in high-density point clouds, advocating for the superiority of point-based networks, as evident in the quantitative results in Table 7.24. Additionally, the study confirms the beneficial impact of integrating panoptic fusion and segmentation fusion techniques, visually represented in Figures 7.17 and 7.6. These fusion approaches



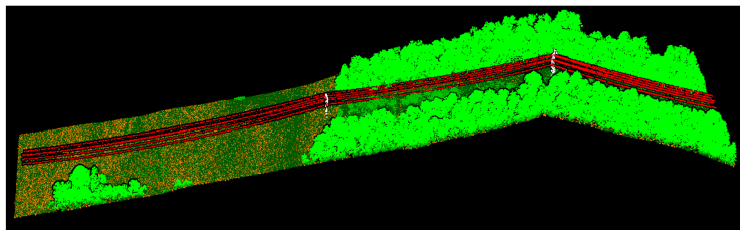
(a) Ground truth



(b) Pan-SUNet detection with 3D U-Net baseline for powerline and pylon segmentation

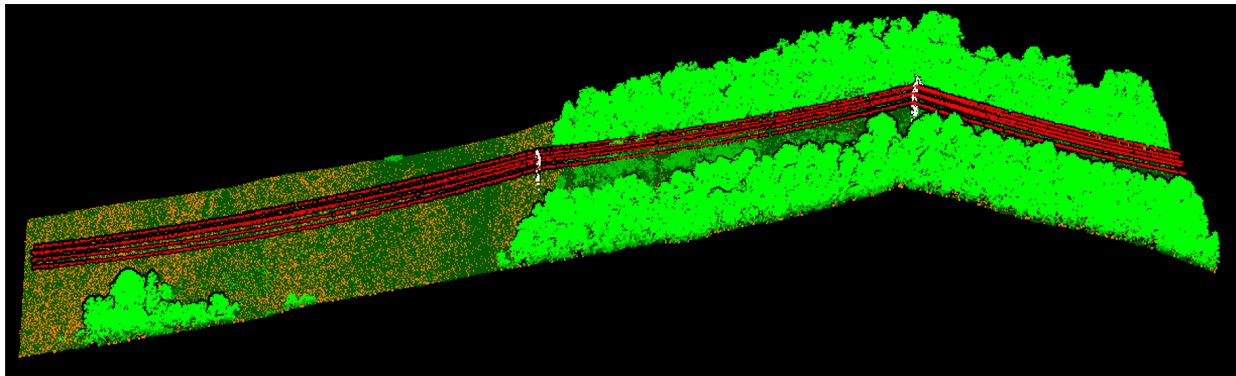


(c) Enhanced Pan-SUNet detection with Attention 3D U-Net baseline

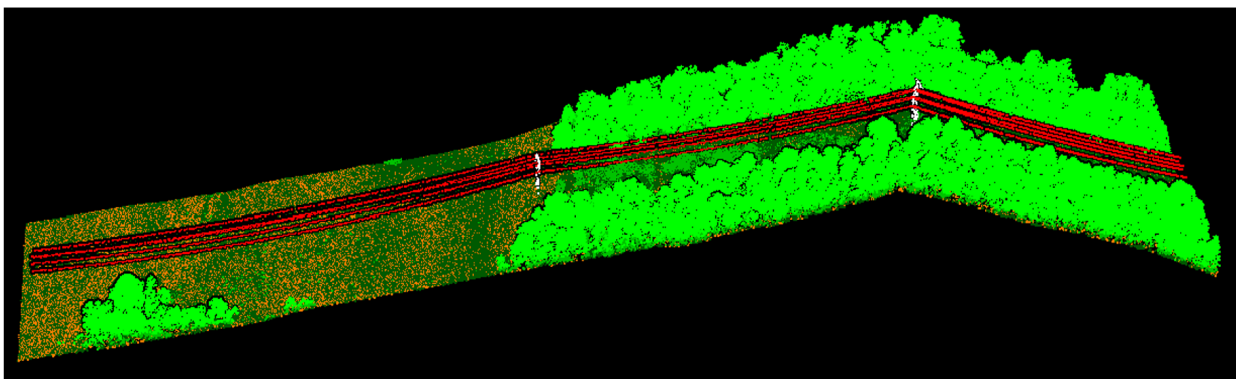


(d) RandLA-Net enhanced powerline and pylon segmentation performance

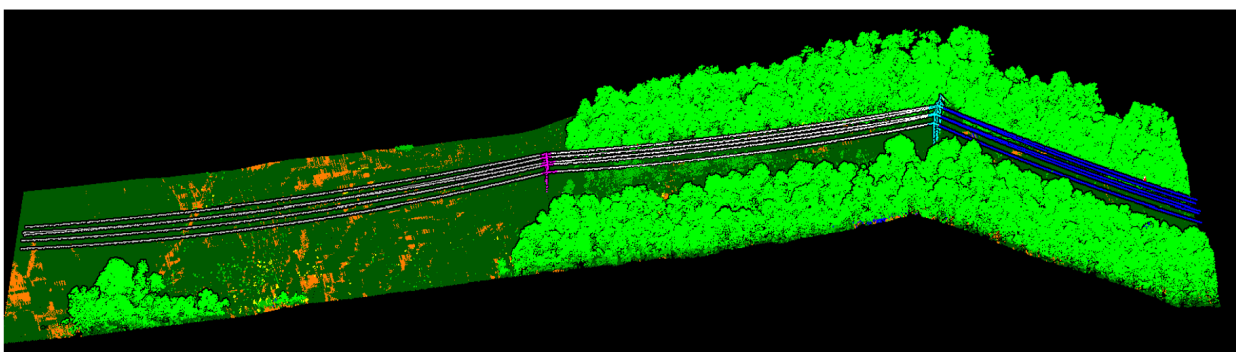
Figure 7.17: Comparative misclassification analysis of 230kV voltage powerline and medium wood pylons by different 3D segmentation models.



(a) Ground truth



(b) Pan-SUNet segmentation output for powerline and pylon identification



(c) Visualization of Span Objects and Pylons through Unique Instance ID Segmentation

Figure 7.18: Instance segmentation of span objects and pylons detected by Pan-SUNet with unique ID coloring.

Methods	Category	Bounding Box	AP	AR	Average Precision	
					Pylon	Span
Pan-SUNet	2D	Oriented	85.6%	88.0%	89.5%	81.0%
Pan-SUNet	3D	Oriented	87.6%	88.0%	89.5 %	84.0%

Table 7.24: Comparative performance matrices of Pan-SUNet for 2D and 3D object detection on Teledyne Optech’s dataset: highlighting overall AP, AR, and class-wise AP for pylon and span detection at IoU=0.5.

demonstrate network-agnostic benefits, showcasing their broad applicability in diverse point cloud segmentation tasks. The instance segmentation demonstrated in Figure 7.6, where each span object and pylon is assigned a unique instance ID, represents a significant advancement beyond basic object classification, crucial for utility network modeling and simulation.

7.5 Summary

In Chapter 7, a detailed exploration of experimental outcomes using two distinct datasets reveals the significance of SUNet’s loss-based late fusion technique for spatial coherence in layout, laying the groundwork for further investigations. The developed Fusion-SUNet and Pan-SUNet methods advance the approach, with evaluation on the Riegl dataset indicating their significant improvement in segmentation tasks. Particularly, Pan-SUNet outperforms Fusion-SUNet in object detection for spans and pylons from both 3D and 2D perspectives, establishing it as the preferred methodology. Comparative analysis with RandLA on the Teledyne dataset demonstrates Pan-SUNet’s high efficiency and effectiveness, highlighting areas for potential improvement in powerline voltage levels and pylon dimensions for future research.

Overall, this chapter provides a comprehensive understanding of the approaches’ strengths and limitations, serving as a solid foundation for future innovations, with Pan-SUNet representing a significant breakthrough in precise object detection and segmentation.

Chapter 8

Conclusions and Future Work

8.1 Summary of Key Findings

This thesis has proposed and developed an integrative and multiscale deep learning network for transmission corridor noise filtering, scene segmentation, and instance span-span analysis. This section will outline the key findings of each contribution.

Noise Filtering Deep Convolutional Neural Network: NSANet effectively filters the noise from a 3D point cloud using physical priors. The AET fusion theory outperforms variations of FIT in recall (87.30%), precision (92.90%), and F1-score (90.0%). Addressing the imbalance in noise points, NSANet utilizes the WCE loss function, surpassing CE by 8.7% and FL by 6.6%. Mean elevation (MZ) proves vital for noise filtering, demonstrating high performance in recall, precision, and F1-score (87.30%, 92.90%, and 90.0%, respectively). Multiscale feature fusion of physical priors enhances performance, surpassing local spatial attention by 2.2%. These findings conclude that multiscale feature fusion, facilitated through attention mechanisms, plays a pivotal role in developing noise filtering networks, and physical priors have a critical role in resolving range ambiguity.

Semantic Utility Deep Convolutional Neural Network: SUNet uses multimodal fusion to offer 2D regional prediction and 3D semantic segmentation for utility applications. Scene classification is improved by using utility-object relationships and a consistent spatial

arrangement. F1-score gains were observed across classes (Pylon 2.0%, Powerline 7.0%, High Veg 2.0%, Ground 5.0%) during validation using the HLC loss function. The ARE module corrected powerline misclassifications into the ground due to non-homogeneous terrain surfaces (Pylon 7.0%, Powerline 7.0%, High Veg 2.0%, Ground 6.0%). Small item misclassifications were decreased using the MFA module (Pylon 87.0% to 89.0% and Powerline 98.0% to 99.0%). However, the misclassification at the pedestal of the pylon persists. Based on these observations, it was determined that multimodal and multiscale fusion improves network performance, which benefits applications related to object detection and scene semantics.

Fusion-Semantic Utility Deep Convolutional Neural Network: Fusion-SUNet is an improvement upon SUNet, aiming to classify objects and regions in 3D point clouds. Fusion-SUNet enhances spatial layout consistency through deep attention fusion. The A-Fusion module performs deep attention multimodal fusion between 2D regions at different scales and 3D voxels. Combined with HLC, this fusion significantly improves performance in the Pylon class (from 77% to 93%) while maintaining other classes at a 99.0% F1-score. These observations indicate that multimodal multiscale fusion of spatial layout can help the network understand the relationships between objects and regions. Theoretically, it emphasizes the significance of spatial layout consistency fused through channel and spatial attention, facilitating scene understanding for specialized layout-based 3D scene applications.

Utility Regional Segmentation Deep Convolutional Neural Network: The 2D regional prediction network aims to generate point-based regional predictions for SUNet, Fusion-SUNet, and Pan-SUNet. This network maintains consistency and produces corridor and non-corridor regional segmentation for Riegl and Teledyne datasets. The experiments demonstrate an accuracy of 88.1% and 95.0% for Riegl and Teledyne datasets, respectively. These results highlight the potential of extracting utility transmission zones to further model utility powerline networks. Utilizing these findings, multimodal spatial layout fusion can be applied to enhance the accuracy of overall point, object, and span-span instance segmentation.

Panoptic Semantic Segmentation Utility Deep Convolutional Neural Network: Pan-SUNet makes object and region-based categorization possible, including instanced

3D object detection, which fulfills the thesis’s ultimate goal. Combining multimodal and multiscale fusion, it highlights spatial arrangement consistency while identifying pylons and span objects in three dimensions. Regional fusion is used to divide these items into three zones. Panoptic segmentation is then used to transfer the objects from 2D to 3D space and assign unique IDs. Experiments were conducted on two datasets: Riegl, containing high voltage (230kV) powerlines, and Teledyne, containing powerlines with variable voltages (400kV-4kV).

Pan-SUNet Findings for Riegl Dataset: Oriented RCNN is used for 2D object detection of Pylons and Spans in Pan-SUNet experiments using the Riegl dataset because of its higher performance (95.4% AP and 96.91% AR). Three regions are supplied to the HLC loss function by combining the results of object detection and 2D regional segmentation. Three baselines (3D U-Net, Attention 3D U-Net, and RandLA) show improved semantic segmentation outcomes; Attention 3D U-Net obtained the best F1-score (Pylon 86.0%, Powerline 99.0%, Ground 99.0%, and High Veg 99.0%). Pan-SUNet is robust on the high-voltage powerline Riegl dataset.

Furthermore, Pan-SUNet presents a panoptic fusion module that translates 2D object detection into 3D space, attaining implicit 3D object identification with an AP of 97.4% and AR of 95.9%. Attention 3D U-Net performs best at this step, refining semantic segmentation predictions and assigning unique instance IDs for 3D objects (Pylon 94.0%, Powerline 99%, Ground 99.0%, and High Veg 99.0%). Beyond point-based semantic segmentation, the qualitative analysis produces instance-based span-span analysis.

Pan-SUNet Findings for Teledyne Optech Dataset: Experiments on the Teledyne dataset using Pan-SUNet are noteworthy for their various features, such as a more even landscape, variable voltage powerlines, and a wide range of item sizes. The reason for the 85.9% AP and 88.0% AR achieved by Oriented RCNN for pylons and spans is the variability of the span object across the dataset. Combining these findings with regional segmentation gives Three regions into the HLC loss function. With the RandLA baseline leading the way in all classes (Pylon 82.0%, Powerline 95.0%, High Veg 98.2%, Medium Veg 93.0%, Ground 98.0%,

Building 88%, Others 78.3%), F1-scores improve everywhere. For medium-to-tall pylons and high-voltage powerlines, Pan-SUNet is reliable. Its panoptic fusion module achieves 87.60% and 88.0% AP by projecting detections into 3D space. Final Pan-SUNet findings for the Teledyne dataset show consistent performance for additional classes, with F1-scores of 97.0% for Powerline and 85.0% for Pylon. The qualitative analysis yields instance-based span-span analysis, point-based semantic segmentation, and 3D object detection.

Pan-SUNet Cross Dataset Findings: The trials demonstrated enhanced performance for both datasets across all baselines when panoptic fusion and HLC were used. Qualitative examination revealed difficulties in low voltage datasets, especially for small pylons, but confirmed robustness for high voltage powerlines. Surface similarities between rooftops and topography led to misclassification overlaps that also caused issues for buildings and other classes.

These results showed that panoptic segmentation may generalize effectively for high-voltage powerline corridor data irrespective of sensor and point density. Nevertheless, because of imbalances in the dataset and the intrinsic difficulty of detecting small objects, it had trouble with small pylons and low-voltage powerlines. Regardless of sensors and point density, Pan-SUNet’s multimodal and multiscale fusion of layout consistency improves scene categorization for span-span analysis, providing insightful information suitable to many utility applications, particularly high-voltage powerlines.

8.2 Strengths and Contributions

The thesis presents several notable strengths and contributions that mark significant advancements in the field of remote sensing and utility management:

Advancement in ALS Noise Characterization and Simulation: In this thesis, a novel approach to simulating noise has been introduced. The focus of this technique is to analyze point distribution within 3D point clouds, and it is particularly useful in addressing the problem of imbalanced class distribution, which is a common challenge in ALS sensors.

The proposed method sets the foundation for improving range ambiguity resolution, thus providing a sophisticated approach to enhance data analysis in ALS applications.

Pioneering Noise Filtration with Neural Networks: One of the notable contributions of this thesis is the development of a neural network that filters noise. The network incorporates physical equations derived from sensor flight trajectory data, resulting in cleaner point clouds that are not restricted to a particular sensor. This enhances the flexibility of the network, making it compatible with various ALS systems. Additionally, the DDA module that has been innovatively developed is network-agnostic, allowing easy integration with any existing neural network architecture.

Network-Agnostic Strategies for Utility Scene Interpretation: This thesis proposes three network-agnostic solutions to enhance the understanding of utility scenes. These adaptable methodologies can use any state-of-the-art point-based or voxel-based semantic segmentation network to improve its performance, which makes it robust for any utility scene understanding application to support asset management.

Enhancing 3D Segmentation Accuracy through Spatial Layout Consistency and 2D Integration: This research introduces a new method for segmenting utility scenes by region, which is a significant step in extracting utility objects and assessing risks more efficiently. This method is particularly effective at identifying spans and pylons in 2D views, which helps to better model utility zones and improve the accuracy of locating utility infrastructure components. Specifically, this research presents an innovative approach that combines 2D regional segmentation and object detection to refine the accuracy of 3D segmentation processes. By maintaining spatial layout consistency, the method leverages the structural regularities observed in 2D utility scene analyses to inform and enhance the depth perception required for 3D segmentation.

Innovation in 3D Object Detection and Instance Segmentation: The thesis represents a significant advancement in 3D object detection and instance segmentation, surpassing the basic level of object classification. These advancements are crucial for distinguishing individual instances within the same object category, providing a detailed view of components

necessary for precise utility network modeling and simulation. This increased detail sets the stage for more intricate simulations and lays a solid foundation for subsequent utility management applications.

The thesis significantly contributes by bridging the gap between academic research and practical application in remote sensing. The ongoing integration of this research into beta launches for utility understanding pipelines and the upcoming production launch in Teledyne Optech’s LMS will provide substantial assistance to network mapping and utility companies for risk analysis and vegetation management. This demonstrates the potential of the research for practical applications with widespread impact.

8.3 Limitations and Prospects of Future Directions:

This thesis advances the application of deep learning in 3D scene analysis, focusing on utility networks. However, it is crucial to identify limitations to highlight potential areas for improvement and further research.

Challenges in Semantic Detail and Classification Accuracy: This research emphasizes the difficulties CNNs encounter in classifying objects of varying sizes, especially when data is sparse. The networks often misrepresent delicate structures, like low-voltage wires or complex pylons, and struggle to differentiate between large areas with similar textures, such as building roofs versus terrain. Upscaling techniques, including stable diffusion or PU-GAN, are highlighted as promising methods to refine semantic segmentation, aiding models in recognizing finer details [158, 159]. Future research will aim to explore and potentially invent new CNN architectures with dynamic, receptive fields tailored to efficiently process variations in object sizes and densities typical in utility management scenarios. This approach sets a clear agenda for advancing CNNs in complex classification tasks, ensuring that future models are precise and adaptable to the evolving demands of utility management.

Scope of Semantic Categorization: The study highlights challenges in utility management semantic categorization with CNNs, emphasizing constraints on specific training data needs.

Despite methodological robustness, resource-intensive dataset curation and data-sharing limitations hinder adaptability in the utility sector. Adaptive learning models, like generative AI for synthetic data [160, 161, 162, 163], knowledge distillation [164, 165, 166], and few-shot learning [167, 168], offer promise for data-scarce scenarios.

The novel concept of prompt engineering could reshape semantic segmentation, allowing CNNs to process diverse tasks through intuitive language prompts [169]. These advancements may significantly enhance CNN’s flexibility in utility management, addressing evolving challenges efficiently.

Constraints of Single-Modal Sensing: The study’s limitation to using only airborne lidar point clouds as data input is a significant constraint. By relying solely on this single sensing modality, the study overlooks the potential advantages of combining data from other sensing modalities, such as UAS (Unmanned Aerial Systems) and satellite imagery. Incorporating these complementary data sources could provide additional contextual cues and increase the overall robustness of the model.

Multi-modal sensing can enrich the input data and provide a more comprehensive feature set for CNNs to learn from. For example, satellite imagery could add a macro perspective of the area, including landscape changes over time, while UAS could offer high-resolution images that reveal surface textures and colors of the objects. These additional data modalities can provide CNNs with a more holistic understanding of the scene, leading to better classification and detection performance.

Generalization and Training Data Challenges: The ability of models to perform well in diverse environments is still limited due to the learning method’s supervised nature and the training data’s specificity. The issue of generalization in CNNs is primarily tied to their reliance on large amounts of labeled training data, which must represent the real-world variability the models are expected to handle. When a CNN is trained on data that is too specific or not diverse enough, its ability to perform accurately across different environments or datasets is compromised—a phenomenon known as overfitting. However, the transferability of the model on various point densities is a domain issue. Therefore, domain adaption and

generalized domain crafting can be a solution. Generative models, such as Stable Diffusion or Generative Adversarial Networks (GANs), can create new, synthetic examples that expand the diversity of the training dataset, helping the CNN to learn more generalized features [170, 171, 159]. Future work could potentially overcome these hurdles by integrating a broader range of learning paradigms and generative techniques to bolster CNN’s ability to adapt and perform well across various environments.

Geometric Representation and Decision-Making: The thesis recognizes the existing discrepancy between CNN-based semantic segmentation and the geometric modeling essential for managing utility networks. Comprehensive utility network management is crucial for making operational decisions and assessing risks. While CNNs are great at identifying and categorizing image data, they cannot translate these categorizations into the geometric models that are vital for simulations. These simulations guide maintenance and emergency responses within utility networks. Future work must focus on developing CNNs that recognize utility components and output data in a format that can be easily integrated with CAD or GIS software. This will result in models that support effective decision-making processes.

Real-Time Processing for Utility Management: Applying deep learning models to utility network management is difficult because of the limited time between offline and online processing, especially in real-time situations. Intense computation impedes real-time decision-making, particularly when dealing with intricate models. The discussed CNNs need to be streamlined using techniques like model pruning and quantization because they are not optimized for real-time analysis. Computations can be accelerated via hardware accelerators. Utility management requires the adaptation of CNNs for speed without sacrificing accuracy; this may require making trade-offs in model complexity or creating new models specifically designed for real-time processing. For 3D scene analysis to be used practically in utility network maintenance, certain obstacles must be overcome.

8.4 Concluding Remarks

This thesis presents novel approaches that enhance utility management by addressing difficulties in semantic segmentation with ALS point clouds for utility scene understanding in remote sensing. Novel noise simulation, NSANet filtering network, and multi-scale, multi-modal point cloud segmentation enhance the state of the art.

Three deep learning architectures—SUNet, Fusion-SUNet, and Pan-SUNet—are presented in this work and intended for utility scene segmentation for ALS point clouds. They reliably classify infrastructure objects in 2D and 3D areas through multi-scale and multi-modal feature fusion, supporting utility asset management and decision-making.

While significant, the thesis recommends further studies, such as expanding methods to other sensing technologies, strengthening AI’s function in remote sensing for autonomous operations, and using results outside the utility sector. This study establishes the groundwork for developing AI and remote sensing, opening the door to more sophisticated utility management systems. The methods and insights guide further study and the creation of sophisticated infrastructure management systems.

Bibliography

References

- [1] Naser El-Sheimy. “An Overview of Mobile Mapping Systems”. In: *International Conference for Global Spatial Data Infrastructure (GSDI-8)* (Apr. 2005), pp. 1–14. URL: <https://api.semanticscholar.org/CorpusID:15467939>.
- [2] Anjan Asthana et al. “A new approach to advanced analytics in utility asset management”. In: *McKinsey’s Electric Power & Natural Gas Practice* (2023). URL: <https://www.mckinsey.com/industries/electric-power-and-natural-gas/our-insights/a-new-approach-to-advanced-analytics-in-utility-asset-management#>.
- [3] Gwangjin Ju et al. “Self-Supervised Dehazing Network Using Physical Priors”. In: *Asian Conference on Computer Vision (ACCV 2022)*. Springer Nature Switzerland, 2023, pp. 290–305. URL: https://doi.org/10.1007/978-3-031-26313-2_18.
- [4] Marie Julie Rakotosaona et al. “PointCleanNet: Learning to Denoise and Remove Outliers from Dense Point Clouds”. In: *Computer Graphics Forum* 39.1 (2020), pp. 185–203. ISSN: 14678659. URL: <https://doi.org/10.1111/cgf.13753>.
- [5] Hugues Thomas et al. “KPConv: Flexible and Deformable Convolution for Point Clouds”. In: *Computing Research Repository* abs/1904.08889 (2019). URL: <http://arxiv.org/abs/1904.08889>.

- [6] Charles Ruizhongtai Qi et al. “PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space”. In: *Computing Research Repository* abs/1706.02413 (2017). URL: <http://arxiv.org/abs/1706.02413>.
- [7] Charles Ruizhongtai Qi et al. “PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation”. In: *International Conference on Computer Vision and Pattern Recognition (CVPR,2017)* (2017). URL: <https://doi.org/10.1109/CVPR.2017.16>.
- [8] Qingyong Hu et al. “RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds”. In: *Computing Research Repository* abs/1911.11236 (2019). URL: <http://arxiv.org/abs/1911.11236>.
- [9] Özgün Çiçek et al. “3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI,2016)*. LNCS 9901 (2016), pp. 424–432. URL: https://doi.org/10.1007/978-3-319-46723-8_49.
- [10] Sheng Yang et al. “Rotated Faster R-CNN for Oriented Object Detection in Aerial Images”. In: *International Conference on Robot Systems and Applications, 2020*. Chengdu, China: Association for Computing Machinery, 2020, pp. 35–39. URL: <https://doi.org/10.1145/3402597.3402605>.
- [11] Xingxing Xie et al. “Oriented R-CNN for Object Detection”. In: *Computing Research Repository* abs/2108.05699 (2021). URL: <https://arxiv.org/abs/2108.05699>.
- [12] Jian Ding et al. “Learning RoI Transformer for Detecting Oriented Objects in Aerial Images”. In: *Computing Research Repository* abs/1812.00155 (2018). arXiv: 1812.00155. URL: <http://arxiv.org/abs/1812.00155>.
- [13] Steven Lang, Fabrizio Ventola, and Kristian Kersting. “DAFNe: A One-Stage Anchor-Free Deep Model for Oriented Object Detection”. In: *Computing Research Repository* abs/2109.06148 (2021). URL: <https://arxiv.org/abs/2109.06148>.
- [14] Joseph Redmon and Ali Farhadi. “YOLOv3: An Incremental Improvement”. In: *Computing Research Repository* abs/1804.02767 (2018).

- [15] Hao Yu et al. “Deep-Learning-Based Semantic Segmentation Approach for Point Clouds of Extra-High-Voltage Transmission Lines”. In: *Remote Sensing* 15.9 (2023). URL: <https://www.mdpi.com/2072-4292/15/9/2371>.
- [16] S. Pu et al. “Real-Time Powerline Corridor Inspection by Edge Computing OF UAV LiDAR Data”. In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-2/W13 (June 2019), pp. 547–551. URL: <https://doi.org/10.5194/isprs-archives-XLII-2-W13-547-2019>.
- [17] Yanjun Wang et al. “Supervised Classification of Power Lines from Airborne LiDAR Data in Urban Areas”. In: *Remote Sensing* 9.8 (2017). ISSN: 2072-4292. DOI: 10.3390/rs9080771. URL: <https://www.mdpi.com/2072-4292/9/8/771>.
- [18] Heungsik Kim and Gunho Sohn. “Point-based Classification of Power Line Corridor Scene Using Random Forests”. In: *Photogrammetric Engineering and Remote Sensing* 79 (Sept. 2013), pp. 821–833. URL: <https://doi.org/10.14358/PERS.79.9.821>.
- [19] Bo Guo et al. “Classification of airborne laser scanning data using JointBoost”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 100 (2015). High-Resolution Earth Imaging for Geospatial Information, pp. 71–83. URL: <https://www.sciencedirect.com/science/article/pii/S0924271614001105>.
- [20] Heungsik Kim and Gunho Sohn. “Random Forests based multiple classifier system for power-line scene classification”. In: vol. XXXVIII-5/W12. Aug. 2011, pp. 253–258. URL: <https://doi.org/10.5194/isprsarchives-XXXVIII-5-W12-253-2011>.
- [21] Gunho Sohn, Yoonseok Jwa, and Heungsik Brian Kim. *Automated Powerline Scene Classification and Reconstruction Utilizing Airborne LiDAR Data*. Melbourne, Australia: International Society for Photogrammetry and Remote Sensing (ISPRS), Aug. 2012. URL: <https://doi.org/10.5194/isprsannals-I-3-167-2012>.
- [22] Mustafa Zeybek. “Inlier Point Preservation in Outlier Points Removed from the ALS Point Cloud”. In: *Journal of the Indian Society of Remote Sensing* (June 2021), pp. 2347–2363. URL: <https://doi.org/10.1007/s12524-021-01397-4>.

- [23] Shan Yong Hua et al. “Denoising algorithm of airborne LIDAR point cloud based on 3D grid”. In: *International Journal of Signal Processing, Image Processing and Pattern Recognition* 10 (Mar. 2017), pp. 85–92. URL: <https://api.semanticscholar.org/CorpusID:65370971>.
- [24] Faisal Zaman, Ya-Ping Wong, and Boon-Yian Ng. “Density-based Denoising of Point Cloud”. In: *Computing Research Repository* abs/1602.05312 (2016). arXiv: 1602.05312. URL: <http://arxiv.org/abs/1602.05312>.
- [25] Bhawna Goyal et al. “Image denoising review: From classical to state-of-the-art approaches”. In: *Information Fusion* 55 (2020), pp. 220–244. URL: <https://doi.org/10.1016/j.inffus.2019.09.003>.
- [26] Ashish Vaswani et al. “Attention Is All You Need”. In: *Computing Research Repository* abs/1706.03762 (2017). URL: <http://arxiv.org/abs/1706.03762>.
- [27] Yehezkel Yeshurun. “Attentional Mechanisms in Computer Vision”. In: 1997. URL: <https://api.semanticscholar.org/CorpusID:16298147>.
- [28] Sarah H. Creem-Regehr. “Spatial Cognition (Category: Others)”. In: *Encyclopedia of Neuroscience*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 3796–3799. URL: https://doi.org/10.1007/978-3-540-29678-2_5542.
- [29] North American Electric Reliability Corporation. *Reliability Principles for Bulk Power Systems*. 2023. URL: https://www.nerc.com/pa/Stand/Resources/Documents/Reliability_Principles.pdf.
- [30] North American Electric Reliability Corporation. *Reliability Standard FAC-003-3 - Vegetation Management*. 2014. URL: <https://www.nerc.com/pa/Stand/Reliability%20Standards/FAC-003-3.pdf#search=vegetation%20standards>.
- [31] I Ituen, G Sohn, and A Jenkins. “A case study: workflow analysis of powerline systems for risk management”. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 37.PART B3B (2008), pp. 331–336. URL: https://www.isprs.org/proceedings/XXXVII/congress/3b_pdf/66.pdf.

- [32] Ima Ituen and Gunho Sohn. “The Way Forward: Advances in Maintaining Right-of-Way of Transmission Lines”. In: *Geomatica* 64.4 (2010), pp. 451–462. URL: <https://cdnsiencepub.com/doi/abs/10.5623/geomat-2010-0047>.
- [33] Maryam Jameela et al. “Simulation-Based Data Augmentation with Physical Priors for Noise Filtering in Deep Neural Networks”. In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLIII-B2-2020 (Aug. 2020), pp. 247–254. URL: <https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-247-2020>.
- [34] Maryam Jameela and Gunho Sohn. “NSANet: Noise Seeking Attention Network”. In: *arXiv e-prints*, arXiv:2302.13392 (Feb. 2023), arXiv:2302.13392.
- [35] Maryam Jameela and Gunho Sohn. “Spatial Layout Consistency for 3D Semantic Segmentation”. In: *Pattern Recognition, Computer Vision, and Image Processing. ICPR 2022 International Workshops and Challenges: 2022, Part III*. Springer-Verlag, 2023, pp. 386–400. URL: https://doi.org/10.1007/978-3-031-37731-0_29.
- [36] Maryam Jameela, Gunho Sohn, and Sunghwan Yoo. “Fusion-SUNet: Spatial Layout Consistency for 3D Semantic Segmentation”. In: *Proceedings IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2023, pp. 6568–6576. URL: <https://doi.org/10.1109/CVPRW59228.2023.00698>.
- [37] Sunghwan Yoo et al. “Human Vision Based 3D Point Cloud Semantic Segmentation of Large-Scale Outdoor Scenes”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. June 2023, pp. 6577–6586. URL: <https://doi.ieeecomputersociety.org/10.1109/CVPRW59228.2023.00699>.
- [38] Mathias Lemmens. *Geo-Information: Technologies, Applications and the Environment*. Berlin, Germany: Springer Science & Business Media, 2011, pp. XXII, 350. URL: <https://doi.org/10.1007/978-94-007-1667-4>.

- [39] Daryl Hartsell, Paul LaRocque, and Jeffrey Tripp. “Galaxy: a new state of the art airborne lidar system”. In: *Proceedings Remote Sensing Technologies and Applications in Urban Environments*; 10008 (2016), 10008U. URL: <https://doi.org/10.1117/12.2241841>.
- [40] K Clint Slatton et al. “Airborne laser swath mapping: achieving the resolution and accuracy required for geosurficial research”. In: *Geophysical Research Letters* 88.17 (2007), pp. 193–200. URL: <https://doi.org/10.1029/2007GL031939>.
- [41] Gordon Petrie. “Mobile Mapping Systems: An Introduction to the Technology”. In: *Geoinformatics* 13 (Jan. 2010), pp. 32–43. URL: <https://api.semanticscholar.org/CorpusID:140086225>.
- [42] Christian Janiesch, Patrick Zschech, and Katharina Heinrich. “Machine learning and deep learning”. In: *Electronic Markets* 31 (2021), pp. 685–695. URL: <https://doi.org/10.1007/s12525-021-00475-2>.
- [43] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. URL: <http://www.deeplearningbook.org>.
- [44] Bulent Ayhan et al. “Vegetation Detection Using Deep Learning and Conventional Methods”. In: *Remote Sensing* 12.15 (2020). URL: <https://www.mdpi.com/2072-4292/12/15/2502>.
- [45] Yifei Zhang and Chen Ling. “A strategy to apply machine learning to small datasets in materials science”. In: *Computational Materials* 4 (2018), p. 25. URL: <https://doi.org/10.1038/s41524-018-0081-z>.
- [46] Niall O’Mahony et al. “Deep Learning vs. Traditional Computer Vision”. In: *Advances in Intelligent Systems and Computing*. Vol. 943. Springer, 2019, pp. 128–144. URL: https://link.springer.com/chapter/10.1007/978-3-030-17795-9_10.
- [47] Yunhao Ba, Guangyuan Zhao, and Achuta Kadambi. “Blending Diverse Physical Priors with Neural Networks”. In: *Computing Research Repository* abs/1910.00201 (2019). URL: <http://arxiv.org/abs/1910.00201>.

- [48] Romain Thoreau et al. *p³ VAE: a physics-integrated generative model. Application to the semantic segmentation of optical remote sensing images*. 2023. arXiv: 2210.10418. URL: <https://doi.org/10.48550/arXiv.2210.10418>.
- [49] Marie Kloenne et al. “Domain-specific cues improve robustness of deep learning-based segmentation of CT volumes”. In: *Scientific Reports* 10.1 (July 2020). URL: <https://doi.org/10.1038/s41598-020-67544-y>.
- [50] Jared Willard et al. “Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems”. In: *ACM Comput. Surv.* 55.4 (2022), pp. 1–37. URL: <https://doi.org/10.1145/3514228>.
- [51] Elizar Elizar et al. “A Review on Multiscale-Deep-Learning Applications”. In: *Sensors* 22.19 (2022). URL: <https://www.mdpi.com/1424-8220/22/19/7384>.
- [52] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *Computing Research Repository* abs/1505.04597 (2015). URL: <http://arxiv.org/abs/1505.04597>.
- [53] Xi Mo and Xiangyu Chen. “Realtime Global Attention Network for Semantic Segmentation”. In: *Computing Research Repository* abs/2112.12939 (2021). URL: <https://arxiv.org/abs/2112.12939>.
- [54] Shice Liu et al. “See and Think: Disentangling Semantic Scene Completion”. In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. NIPS’18. Montréal, Canada: Curran Associates Inc., 2018, pp. 261–272. URL: <https://dl.acm.org/doi/10.5555/3326943.3326968>.
- [55] Tsung-Yi Lin et al. “Feature Pyramid Networks for Object Detection”. In: *Computing Research Repository* abs/1612.03144 (2016). URL: <http://arxiv.org/abs/1612.03144>.
- [56] Ozan Oktay et al. “Attention U-Net: Learning Where to Look for the Pancreas”. In: *Computing Research Repository* abs/1804.03999 (2018). URL: <http://arxiv.org/abs/1804.03999>.

- [57] Fabian Isensee et al. “Brain Tumor Segmentation and Radiomics Survival Prediction: Contribution to the BRATS 2017 Challenge”. In: *Computing Research Repository* abs/1802.10508 (2018). URL: <http://arxiv.org/abs/1802.10508>.
- [58] Mandar Haldekar, Ashwinkumar Ganesan, and Tim Oates. “Identifying spatial relations in images using convolutional neural networks”. In: *International Joint Conference on Neural Networks (IJCNN)*. 2017, pp. 3593–3600. URL: <https://doi.org/10.1109/IJCNN.2017.7966308>.
- [59] Yuee Liu et al. “Classification of Airborne LIDAR Intensity Data Using Statistical Analysis and Hough Transform with Application to Power Line Corridors”. In: *Digital Image Computing: Techniques and Applications*. 2009, pp. 462–467. URL: <https://doi.org/10.1109/DICTA.2009.83>.
- [60] Yan Wang et al. “An Improved MVCNN for 3D Shape Recognition”. In: *IEEE International Conference on Emergency Science and Information Technology (ICESIT)*. 2021, pp. 469–472. URL: <https://doi.org/10.1109/ICESIT53460.2021.9696941>.
- [61] Zongwei Zhou et al. “UNet++: A Nested U-Net Architecture for Medical Image Segmentation”. In: *Computing Research Repository* abs/1807.10165 (2018). URL: <http://arxiv.org/abs/1807.10165>.
- [62] Zeyu Hu et al. “VMNet: Voxel-Mesh Network for Geodesic-Aware 3D Semantic Segmentation”. In: *Computing Research Repository* abs/2107.13824 (2021). URL: <https://arxiv.org/abs/2107.13824>.
- [63] Benjamin Graham. “Spatially-sparse convolutional neural networks”. In: *Computing Research Repository* abs/1409.6070 (2014). URL: <http://arxiv.org/abs/1409.6070>.
- [64] Wenxuan Wu, Zhongang Qi, and Fuxin Li. “PointConv: Deep Convolutional Networks on 3D Point Clouds”. In: *Computing Research Repository* abs/1811.07246 (2018). URL: <http://arxiv.org/abs/1811.07246>.

- [65] Yangyan Li et al. “PointCNN”. In: *Computing Research Repository* abs/1801.07791 (2018). URL: <http://arxiv.org/abs/1801.07791>.
- [66] Yifan Xu et al. “SpiderCNN: Deep Learning on Point Sets with Parameterized Convolutional Filters”. In: *Computing Research Repository* abs/1803.11527 (2018). URL: <http://arxiv.org/abs/1803.11527>.
- [67] Shaoqing Ren et al. “Faster R-CNN: Towards real-time object detection with region proposal networks”. In: *Advances in neural information processing systems*. 2015, pp. 91–99. URL: <https://doi.ieeecomputersociety.org/10.1109/TPAMI.2016.2577031>.
- [68] Michael Crawshaw. “Multi-Task Learning with Deep Neural Networks: A Survey”. In: *Computing Research Repository* abs/2009.09796 (2020). URL: <https://arxiv.org/abs/2009.09796>.
- [69] Binbin Xiang et al. “A Review of panoptic segmentation for mobile mapping point clouds”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 203 (2023), pp. 373–391. URL: <https://www.sciencedirect.com/science/article/pii/S092427162300223X>.
- [70] Kshitij Sirohi et al. “EfficientLPS: Efficient LiDAR Panoptic Segmentation”. In: *Computing Research Repository* abs/2102.08009 (2021). URL: <https://arxiv.org/abs/2102.08009>.
- [71] Shihao Su et al. *PUPS: Point Cloud Unified Panoptic Segmentation*. 2023. arXiv: 2302.06185. URL: <https://doi.org/10.48550/arXiv.2302.06185>.
- [72] Enxu Li et al. “CPSeg: Cluster-free Panoptic Segmentation of 3D LiDAR Point Clouds”. In: *Computing Research Repository* abs/2111.01723 (2021). URL: <https://arxiv.org/abs/2111.01723>.
- [73] Connor Shorten and Taghi M. Khoshgoftaar. “A survey on Image Data Augmentation for Deep Learning”. In: *Journal of Big Data* 6.60 (2019), pp. 1–48. URL: <https://doi.org/10.1186/s40537-019-0197-0>.

- [74] Xiangyu Yue et al. “A LiDAR Point Cloud Generator: from a Virtual World to Autonomous Driving”. In: *International Conference on Multimedia Retrieval (ICMR)* (2018), pp. 458–464. URL: <https://doi.org/10.1145/3206025.3206080>.
- [75] Ahmad El Sallab et al. “LiDAR Sensor modeling and Data augmentation with GANs for Autonomous driving”. In: *ICML Workshop on AI for Autonomous Driving* (2019). URL: <http://arxiv.org/abs/1905.07290>.
- [76] David Griffiths and Jan Boehm. “SynthCity: A large scale synthetic point cloud”. In: *Computing Research Repository* abs/1907.04758 (2019). URL: <http://arxiv.org/abs/1907.04758>.
- [77] David Griffiths and Jan Boehm. “Weighted Point Cloud Augmentation for Neural Network Training Data Class-Imbalance”. In: *Computing Research Repository* abs/1904.04094 (2019). URL: <http://arxiv.org/abs/1904.04094>.
- [78] Robin Heinzler et al. “CNN-based Lidar Point Cloud De-Noising in Adverse Weather”. In: *Computing Research Repository* (Dec. 2019). URL: <http://dx.doi.org/10.1109/LRA.2020.2972865>.
- [79] Chun-Liang Li et al. “Point Cloud GAN”. In: *Computing Research Repository* abs/1810.05795 (2018). URL: <http://arxiv.org/abs/1810.05795>.
- [80] Panos Achlioptas et al. “Representation Learning and Adversarial Generation of 3D Point Clouds”. In: abs/1707.02392 (2017). URL: <http://arxiv.org/abs/1707.02392>.
- [81] Guandao Yang et al. “PointFlow: 3D Point Cloud Generation with Continuous Normalizing Flows”. In: *International Conference on Computer Vision (ICCV, 2019)* (2019). URL: <https://doi.org/10.1109/ICCV.2019.00464>.
- [82] Dong Wook Shu, Sung Woo Park, and Junseok Kwon. “3D Point Cloud Generative Adversarial Network Based on Tree Structured Graph Convolutions”. In: *International Conference on Computer Vision (ICCV, 2019)* (2019). URL: <https://doi.org/10.1109/ICCV.2019.00396>.

- [83] O. Schall, A. Belyaev, and H.-P. Seidel. “Robust filtering of noisy scattered point data”. In: *Proceedings Eurographics/IEEE VGTC Symposium Point-Based Graphics, 2005*. 2005, pp. 71–144. DOI: <https://doi.org/10.1109/PBG.2005.194067>.
- [84] Sven Ronnback and Ake Wernersson. “On filtering of laser range data in snowfall”. In: *2008 4th International IEEE Conference Intelligent Systems*. Vol. 2. 2008, pp. 17-33-17–39. URL: <https://doi.org/10.1109/IS.2008.4670551>.
- [85] Nicholas Charron, Stephen Phillips, and Steven L. Waslander. “De-noising of Lidar Point Clouds Corrupted by Snowfall”. In: *15th Conference on Computer and Robot Vision (CRV)*. 2018, pp. 254–261. URL: <https://doi.org/10.1109/CRV.2018.00043>.
- [86] Peter Rousseeuw and Mia Hubert. “Robust statistics for outlier detection”. In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 1 (2011). URL: <https://doi.org/10.1002/widm.2>.
- [87] Bochang Zou, Huadong Qiu, and Yufeng Lu. “Point Cloud Reduction and Denoising Based on Optimized Downsampling and Bilateral Filtering”. In: *IEEE Access* PP (July 2020), pp. 1–1. URL: <https://doi.org/10.1109/ACCESS.2020.3011989>.
- [88] Chaohui Lv and Min Li. “Point Cloud Denoising Algorithm Based on Noise Classification”. In: *2020 International Conference on Culture-oriented Science Technology (ICCST)*. 2020, pp. 123–127. URL: <https://doi.org/10.1109/ICCST50977.2020.00029>.
- [89] E. Mattei and A. Castrodad. “Point Cloud Denoising via Moving RPCA: MRPCA”. In: *Computer Graphics Forum* 36 (Nov. 2016). URL: <https://doi.org/10.1111/cgf.13068>.
- [90] Q. Kang, G. Huang, and S. Yang. “Gross Error Elimination in Point Cloud Data Using KD-Tree-Based Method”. In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-3* (Apr. 2018), pp. 719–722. URL: <https://doi.org/10.5194/isprs-archives-XLII-3-719-2018>.

- [91] Yann Schoenenberger, Johan Paratte, and Pierre Vanderghenst. “Graph-based denoising for time-varying point clouds”. In: *The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*. 2015, pp. 1–4. URL: <https://doi.org/10.1109/3DTV.2015.7169366>.
- [92] Chinthaka Dinesh et al. *Fast 3D Point Cloud Denoising via Bipartite Graph Approximation and Total Variation*. 2018. arXiv: 1804.10831. URL: <https://doi.org/10.48550/arXiv.1804.10831>.
- [93] Chinthaka Dinesh, Gene Cheung, and Ivan V. Bajic. *3D Point Cloud Denoising via Bipartite Graph Approximation and Reweighted Graph Laplacian*. 2018. URL: <https://doi.org/10.48550/arXiv.1812.07711>.
- [94] Julie Digne and Carlo de Franchis. “The Bilateral Filter for Point Clouds”. In: *Image Processing On Line* 7 (Oct. 2017), pp. 278–287. URL: <https://doi.org/10.5201/ipol.2017.179>.
- [95] Wei Hu et al. “Dynamic Point Cloud Denoising via Manifold-to-Manifold Distance”. In: *IEEE Transactions on Image Processing* 30 (2021), pp. 6168–6183. URL: <http://dx.doi.org/10.1109/TIP.2021.3092826>.
- [96] Rui Gao et al. “Reflective Noise Filtering of Large-Scale Point Cloud Using Multi-Position LiDAR Sensing Data”. In: *Remote Sensing* 13.16 (2021). URL: <https://www.mdpi.com/2072-4292/13/16/3058>.
- [97] Muhammad Abeer Irfan and Enrico Magli. “3D Point Cloud Denoising Using a Joint Geometry and Color k-NN Graph”. In: *2020 28th European Signal Processing Conference (EUSIPCO)* (2021), pp. 585–589. URL: <https://api.semanticscholar.org/CorpusID:229357741>.
- [98] Yutao Wang and Hsi-Yung Feng. “Outlier detection for scanned point clouds using majority voting”. In: *Computer-Aided Design* 62 (2015), pp. 31–43. URL: <https://www.sciencedirect.com/science/article/pii/S0010448514002413>.

- [99] Yao Duan et al. “Low-complexity point cloud denoising for LiDAR by PCA-based dimension reduction”. In: *Optics Communications* 482 (2021), p. 126567. ISSN: 0030-4018. DOI: <https://doi.org/10.1016/j.optcom.2020.126567>. URL: <https://www.sciencedirect.com/science/article/pii/S0030401820309858>.
- [100] Yousra Regaya, Fodil Fadli, and Abbes Amira. “Point-Denoise: Unsupervised Outlier Detection for 3D Point Clouds Enhancement”. In: *Multimedia Tools Appl.* 80.18 (2021), pp. 28161–28177. URL: <https://doi.org/10.1007/s11042-021-10924-x>.
- [101] Dongbo Zhang et al. “Pointfilter: Point Cloud Filtering via Encoder-Decoder Modeling”. In: *Computing Research Repository* abs/2002.05968 (2020). arXiv: 2002.05968. URL: <https://arxiv.org/abs/2002.05968>.
- [102] Chaojing Duan, Siheng Chen, and Jelena Kovacevic. “3D Point Cloud Denoising via Deep Neural Network based Local Surface Estimation”. In: *Computing Research Repository* abs/1904.04427 (2019). URL: <http://arxiv.org/abs/1904.04427>.
- [103] Vinit Sarode et al. “MaskNet: A Fully-Convolutional Network to Estimate Inlier Points”. In: *Computing Research Repository* abs/2010.09185 (2020). arXiv: 2010.09185. URL: <https://arxiv.org/abs/2010.09185>.
- [104] Rajat Sharma et al. “Point Cloud Upsampling and Normal Estimation using Deep Learning for Robust Surface Reconstruction”. In: *Computing Research Repository* abs/2102.13391 (2021). URL: <https://arxiv.org/abs/2102.13391>.
- [105] Rui Gao et al. “Reflective Noise Filtering of Large-Scale Point Cloud Using Transformer” In: *Remote Sensing* 14.3 (2022). URL: <https://www.mdpi.com/2072-4292/14/3/577>.
- [106] Russell B Roth and James Lowry Thompson. “Practical Application of Multiple Pulse in Air (MPIA) Lidar in Large-area Surveys”. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS)* XXXVII.B1 (2008). URL: <https://api.semanticscholar.org/CorpusID:14870371>.

- [107] Hang Gong et al. “Multiscale Information Fusion for Hyperspectral Image Classification Based on Hybrid 2D-3D CNN”. In: *Remote Sensing* 13.12 (2021). URL: <https://www.mdpi.com/2072-4292/13/12/2268>.
- [108] Meng-Hao Guo et al. “Attention Mechanisms in Computer Vision: A Survey”. In: *Computing Research Repository* abs/2111.07624 (2021). URL: <https://arxiv.org/abs/2111.07624>.
- [109] John Duncan and Glyn Humphreys. “Beyond the search surface: Visual search and attentional engagement.” In: *J Exp Psychol Hum Percept Perform* (1992). URL: <https://doi.org/10.1037//0096-1523.18.2.578>.
- [110] Anne M. Treisman and Garry Gelade. “A feature-integration theory of attention”. In: *Cognitive Psychology* 12.1 (1980), pp. 97–136. URL: <https://www.sciencedirect.com/science/article/pii/0010028580900055>.
- [111] Alexandre Boulch et al. “SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks”. In: *Computers & Graphics* 71 (2018), pp. 189–198. URL: <https://www.sciencedirect.com/science/article/pii/S0097849317301942>.
- [112] Yaodong Cui et al. “Deep Learning for Image and Point Cloud Fusion in Autonomous Driving: A Review”. In: *IEEE Transactions on Intelligent Transportation Systems* 23.2 (2022), pp. 722–739. URL: <https://doi.org/10.1109/TITS.2020.3023541>.
- [113] Haoxuan You et al. “PVRNet: Point-View Relation Neural Network for 3D Shape Recognition”. In: (2018). URL: <https://api.semanticscholar.org/CorpusID:54437351>.
- [114] Andres Milioto et al. “RangeNet ++: Fast and Accurate LiDAR Semantic Segmentation”. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2019, pp. 4213–4220. URL: <https://doi.org/10.1109/IROS40897.2019.8967762>.

- [115] Bichen Wu et al. “SqueezeSeg: Convolutional Neural Nets with Recurrent CRF for Real-Time Road-Object Segmentation from 3D LiDAR Point Cloud”. In: *Computing Research Repository* abs/1710.07368 (2017). URL: <http://arxiv.org/abs/1710.07368>.
- [116] Gordon Wyeth and Michael Milford. “Spatial cognition for robots”. In: *IEEE Robotics Automation Magazine* 16.3 (2009), pp. 24–32. URL: <https://doi.org/10.1109/MRA.2009.933620>.
- [117] Fausto Giunchiglia. “Contextual Reasoning”. In: *Epistemologia* 16 (Sept. 1997). URL: https://www.researchgate.net/publication/2824890_Contextual_Reasoning.
- [118] National Grid Transco UK, Wales, and USA. *Design Guidelines for Development near High-Voltage Overhead Lines*. Accessed: 02 March 2023. 2022.
- [119] Electric Power Research Institute (EPRI). *Vegetation Management - Best Practices for Transmission Corridors*. Electric Power Research Institute, 2012.
- [120] L. M. Faulkenberry and W. Coffer. *Electrical Power Distribution and Transmission*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
- [121] Alex Roderick. *Utility Power Transmission and Distribution Systems*. 2023. 2021. URL: <https://eepower.com/technical-articles/utility-power-transmission-and-distribution-systems/#>.
- [122] BC Hydro. *Planting Near Powerlines: Safety Guidelines*. 2019. URL: <https://www.bchydro.com/content/dam/BCHydro/customer-portal/documents/corporate/safety/planting-near-powerlines.pdf>.
- [123] BC Hydro. *Right of Way Clearing Guidelines*. 2017. URL: <https://www.bchydro.com/content/dam/BCHydro/customer-portal/documents/corporate/safety/row-guidelines-2017.pdf>.
- [124] Haiyan Guan et al. “Extraction of power-transmission lines from vehicle-borne lidar data”. In: *International Journal of Remote Sensing* 37.1 (2016), pp. 229–247. URL: <https://doi.org/10.1080/01431161.2015.1125549>.

- [125] Yoonseok Jwa, Gunho Sohn, and Heungsik Kim. “Automatic 3d powerline reconstruction using airborne lidar data”. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38 (Jan. 2009), pp. 105–110. URL: <https://api.semanticscholar.org/CorpusID:111424537>.
- [126] Jisheng Yang et al. “Power Line Corridor LiDAR Point Cloud Segmentation Using Convolutional Neural Network”. In: *Pattern Recognition and Computer Vision*. Cham: Springer International Publishing, 2019, pp. 160–171. URL: https://doi.org/10.1007/978-3-030-31654-9_14.
- [127] Jaehoon Jung et al. “Automated and efficient powerline extraction from laser scanning data using a voxel-based subsampling with hierarchical approach”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 163 (2020), pp. 343–361. URL: <https://doi.org/10.1016/j.isprsjprs.2020.03.018>.
- [128] TERRA FRMA. *Major Power Outage Cases: Biggest Blackouts in U.S. History*. 2022. URL: <https://terrafrma.com/blogs/news/major-power-outage-cases-biggest-blackouts-in-us-history>.
- [129] Graham Flack and Kevin M. Kolevar. *Final Report on the Implementation of the Task Force Recommendations*. Natural Resources Canada and U.S. Department of Energy, 2006. URL: <https://www.ieso.ca/en/Corporate-IESO/Media/Also-of-Interest/Blackout-2003>.
- [130] Heungsik Kim and Gunho Sohn. “3D Classification of Powerline Scene from Airborne Laser Scanning Data using Random Forests”. In: *PCV: ISPRS Technical Commission III Symposium on Photogrammetry Computer Vision and Image Analysis* 38 (Sept. 2010), pp. 126–132. URL: <http://dx.doi.org/10.13140/2.1.1757.4409>.
- [131] C. Askit et al. “Extraction of Point Cloud-Based Information for Powerline Corridors”. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-4/W6-2022 Geoinformation Week 2022 “Broad-*

- ening Geospatial Science and Technology*". Johor Bahru, Malaysia (online), Nov. 2022. URL: <https://doi.org/10.5194/isprs-archives-XLVIII-4-W6-2022-41-2023>.
- [132] Chi Chen et al. "DCPLD-Net: A diffusion coupled convolution neural network for real-time power transmission lines detection from UAV-Borne LiDAR data". In: *International Journal of Applied Earth Observation and Geoinformation* 112 (2022), p. 102960. URL: <https://www.sciencedirect.com/science/article/pii/S1569843222001558>.
- [133] Thomas Melzer and Christian Bries. "Extraction and Modeling of Power Lines from ALS Point Clouds". In: (Jan. 2004). URL: https://www.researchgate.net/publication/250009282_Extraction_and_Modeling_of_Power_Lines_from_ALS_Point_Clouds.
- [134] R.A. McLaughlin. "Extracting transmission lines from airborne LIDAR data". In: *IEEE Geoscience and Remote Sensing Letters* 3.2 (2006). Cited by: 133, pp. 222–226. DOI: 10.1109/LGRS.2005.863390. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-33646407580&doi=10.1109%2fLGRS.2005.863390&partnerID=40&md5=0af183c1a0ba1689c3ce8ce58b86ded0>.
- [135] Yoonseok Jwa and Gunho Sohn. "A multi-level span analysis for improving 3d power-line reconstruction performance using airborne laser scanning data". In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38 (Sept. 2010), pp. 97–102. URL: https://www.isprs.org/proceedings/xxxviii/part3/a/pdf/97_XXXVIII-part3A.pdf.
- [136] Marcel Ritter and Werner Bengel. "Reconstructing power cables from LIDAR data using Eigenvector streamlines of the point distribution tensor field". In: *Journal of WSCG* 20.3 (2012). Cited by: 21, pp. 223–230. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84867181767&partnerID=40&md5=0840413d1a32582ebf76c95a964935b3>.

- [137] Yoonseok Jaw and Gunho Sohn. “Wind adaptive modeling of transmission lines using minimum description length”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 125 (2017), pp. 193–206. URL: <https://www.sciencedirect.com/science/article/pii/S0924271616301253>.
- [138] Odion Omonfoman. *Electricity Distribution Companies – The Challenges and Way Forward*. Published on Premium Times, January 4, 2016. Reading Time: 10 mins read. 2016. URL: <https://opinion.premiumtimesng.com/2016/01/04/electricity-distribution-companies-the-challenges-and-way-forward-by-odion-omonfoman/?tztc=1>.
- [139] Jules Raymond Kala et al. “Assets Management on Electrical Grid Using Faster-RCNN”. In: *Annals of Operations Research* 308 (2022), pp. 307–320. URL: <https://doi.org/10.1007/s10479-020-03650-4>.
- [140] Ming Dong. “Combining Unsupervised and Supervised Learning for Asset Class Failure Prediction in Power Systems”. In: *IEEE Transactions on Power Systems* 34.6 (2019), pp. 5033–5043. URL: <https://doi.org/10.1109/TPWRS.2019.2920915>.
- [141] Mallikarjun Angalakudati et al. “Improving emergency storm planning using machine learning”. In: *2014 IEEE PES T&D Conference and Exposition*. 2014, pp. 1–6. URL: <https://doi.org/10.1109/TDC.2014.6863406>.
- [142] Connie Ko, Tarmo K Rimmel, and Gunho Sohn. “Mapping Tree Genera Using Discrete LiDAR and Geometric Tree Metrics”. In: *Bosque (Valdivia)* 33.3 (2012). URL: https://scielo.conicyt.cl/scielo.php?script=sci_arttext&pid=S0717-92002012000300015.
- [143] Connie Ko et al. “Hybrid Ensemble Classification of Tree Genera Using Airborne LiDAR Data”. In: *Remote Sensing* 6.11 (2014), pp. 11225–11243. URL: <https://www.mdpi.com/2072-4292/6/11/11225>.

- [144] Xinhuai Zou et al. “Tree Classification in Complex Forest Point Clouds Based on Deep Learning”. In: *IEEE Geoscience and Remote Sensing Letters* 14.12 (2017), pp. 2360–2364. URL: <https://doi.org/10.1109/LGRS.2017.2764938>.
- [145] Stefan Schmohl, Alejandra Narváez Vallejo, and Uwe Soergel. “Individual Tree Detection in Urban ALS Point Clouds with 3D Convolutional Networks”. In: *Remote Sensing* 14.6 (2022). URL: <https://www.mdpi.com/2072-4292/14/6/1317>.
- [146] Yin Zhou and Oncel Tuzel. “VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection”. In: *Computing Research Repository* abs/1711.06396 (2017). URL: <http://arxiv.org/abs/1711.06396>.
- [147] Shaoshuai Shi et al. “Part-A² Net: 3D Part-Aware and Aggregation Neural Network for Object Detection from Point Cloud”. In: *Computing Research Repository* abs/1907.03670 (2019). URL: <http://arxiv.org/abs/1907.03670>.
- [148] Yan Yan, Yuxing Mao, and Bo Li. “SECOND: Sparsely Embedded Convolutional Detection”. In: *Sensors* 18.10 (2018). URL: <https://www.mdpi.com/1424-8220/18/10/3337>.
- [149] Alex H. Lang et al. “PointPillars: Fast Encoders for Object Detection from Point Clouds”. In: *Computing Research Repository* abs/1812.05784 (2018). URL: <http://arxiv.org/abs/1812.05784>.
- [150] Leo Stanislas et al. “Airborne Particle Classification in LiDAR Point Clouds Using Deep Learning”. In: *International Conference on Field and Service Robotics (FSR, 2019)* (2019), pp. 1–14. URL: <https://eprints.qut.edu.au/133596/>.
- [151] Terrasolid. *Terrasolid*. Accessed: 2022-11-11.
- [152] Alessandro Palla, David Moloney, and Luca Fanucci. “Fully convolutional denoising autoencoder for 3D scene reconstruction from a single depth image”. In: *International Conference on Systems and Informatics (ICSAI,2017)* (Nov. 2017), pp. 566–575. URL: <https://doi.org/10.1109/ICSAI.2017.8248355>.

- [153] M. Zhou et al. “Automatic Extraction of Power lines from UAV LiDAR Point Clouds using a Novel Spatial Feature”. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* IV-2/W7 (2019), pp. 227–234. DOI: 10.5194/isprs-annals-IV-2-W7-227-2019. URL: <https://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/IV-2-W7/227/2019/>.
- [154] Liza Reed et al. “Expanding Transmission Capacity: Examples of Regulatory Paths for Five Alternative Strategies”. In: *The Electricity Journal* 33.6 (2020), p. 106770. ISSN: 1040-6190. DOI: <https://doi.org/10.1016/j.tej.2020.106770>. URL: <https://www.sciencedirect.com/science/article/pii/S1040619020300622>.
- [155] Jie Hu, Li Shen, and Gang Sun. “Squeeze-and-Excitation Networks”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2018. URL: <https://doi.org/10.48550/arXiv.1709.01507>.
- [156] Sanghyun Woo et al. “CBAM: Convolutional Block Attention Module”. In: *Computing Research Repository* abs/1807.06521 (2018). URL: <http://arxiv.org/abs/1807.06521>.
- [157] Kentaro Wada. *labelme*. Accessed on March 5, 2021.
- [158] Ruihui Li et al. “PU-GAN: a Point Cloud Upsampling Adversarial Network”. In: *IEEE International Conference on Computer Vision (ICCV)*. 2019. URL: <https://doi.org/10.48550/arXiv.1907.10844>.
- [159] Zhendong Wang et al. *Diffusion-GAN: Training GANs with Diffusion*. 2023. URL: <https://doi.org/10.48550/arXiv.2206.02262>.
- [160] Shuwen Peng et al. “Point-Based Multilevel Domain Adaptation for Point Cloud Segmentation”. In: *IEEE Geoscience and Remote Sensing Letters* 19 (2022), pp. 1–5. URL: <https://doi.org/10.1109/LGRS.2020.3037702>.
- [161] Liangwei Shi et al. “DFAN: Dual-Branch Feature Alignment Network for Domain Adaptation on Point Clouds”. In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), pp. 1–12. URL: <https://doi.org/10.1109/TGRS.2022.3171038>.

- [162] Peng Jiang and Srikanth Saripalli. “LiDARNet: A Boundary-Aware Domain Adaptation Model for Point Cloud Semantic Segmentation”. In: *IEEE International Conference on Robotics and Automation (ICRA)*. 2021, pp. 2457–2464. URL: <https://doi.org/10.1109/ICRA48506.2021.9561255>.
- [163] Huan Luo et al. “Domain Adaptation for Object Classification in Point Clouds via Asymmetrical Siamese and Conditional Adversarial Network”. In: *IEEE Geoscience and Remote Sensing Letters* 19 (2022), pp. 1–5. URL: <https://doi.org/10.1109/LGRS.2022.3208589>.
- [164] Linfeng Zhang et al. “PointDistiller: Structured Knowledge Distillation Towards Efficient and Compact 3D Detection”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2023, pp. 21791–21801. URL: <https://doi.org/10.48550/arXiv.2205.11098>.
- [165] Yuenan Hou et al. “Point-to-Voxel Knowledge Distillation for LiDAR Semantic Segmentation”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2022, pp. 8479–8488. URL: <https://doi.org/10.48550/arXiv.2206.02099>.
- [166] Yue Wang et al. “Multi-Frame to Single-Frame: Knowledge Distillation for 3D Object Detection”. In: *Computing Research Repository* abs/2009.11859 (2020). URL: <https://arxiv.org/abs/2009.11859>.
- [167] Chen Ye, Hao Zhu, Bei Zhang, et al. “A Closer Look at Few-Shot 3D Point Cloud Classification”. In: *International Journal of Computer Vision* 131 (Mar. 2023), pp. 772–795. URL: <https://doi.org/10.1007/s11263-022-01731-4>.
- [168] Charu Sharma and Manohar Kaul. “Self-Supervised Few-Shot Learning on Point Clouds”. In: *Advances in Neural Information Processing Systems*. Ed. by H. Larochelle et al. Vol. 33. Curran Associates, Inc., 2020, pp. 7212–7221. URL: https://proceedings.neurips.cc/paper_files/paper/2020/file/50c1f44e426560f3f2cdcb3e19e39903-Paper.pdf.

- [169] J White et al. *A prompt pattern catalog to enhance prompt engineering with ChatGPT*. 2023. URL: <https://doi.org/10.48550/arXiv.2302.11382>.
- [170] Robin Rombach et al. *High-Resolution Image Synthesis with Latent Diffusion Models*. 2021. URL: <https://doi.org/10.48550/arXiv.2112.10752>.
- [171] Prafulla Dhariwal and Alex Nichol. “Diffusion Models Beat GANs on Image Synthesis”. In: *Computing Research Repository* abs/2105.05233 (2021). URL: <https://arxiv.org/abs/2105.05233>.
- [172] Da Li Yu. “A Novel Surface Reconstruction Method for Noisy Cloud Points Based on Support Vector Machine”. In: *Materials, Mechatronics and Automation IV*. Vol. 950. Advanced Materials Research. Trans Tech Publications Ltd, July 2014, pp. 139–144. URL: <https://doi.org/10.4028/www.scientific.net/AMR.950.139>.
- [173] Xavier Roynard, Jean Emmanuel Deschaud, and François Goulette. “Paris-Lille-3D: A large and high-quality ground-truth urban point cloud dataset for automatic segmentation and classification”. In: *International Journal of Robotics Research* 37.6 (2018), pp. 545–557. URL: <https://doi.org/10.1177/0278364918767506>.
- [174] Benoit Brummer and Christophe De Vleeschouwer. “Natural Image Noise Dataset”. In: *New Trends in Image Restoration and Enhancement workshop and challenges In conjunction with CVPR, 2019* (2019). arXiv: 1906.00270. URL: <http://arxiv.org/abs/1906.00270>.
- [175] Abdelrahman Abdelhamed, Stephen Lin, and Michael S. Brown. “A High-Quality Denoising Dataset for Smartphone Cameras”. In: *IEEE International Conference of Computer Vision and Pattern Recognition (CVPR,2018)* (2018), pp. 1692–1700. ISSN: 10636919. DOI: 10.1109/CVPR.2018.00182.
- [176] Demetrios Gatziolis and Hans-Erik Anderse. “A Guide to LIDAR Data Acquisition and Processing for the Forests of the Pacific Northwest”. In: *United States Department of Agriculture, Forest Service, Pacific Northwest Research Station, General Technical*

- Report, PNW-GTR-768* (2008). URL: https://www.fs.fed.us/pnw/pubs/pnw_gtr768.pdf.
- [177] Felix Rohrbach. “An Introduction to LiDAR”. In: *english, personal blog* (Feb. 2015). URL: <https://felix.rohrba.ch/en/2015/an-introduction-to-lidar/>.
 - [178] “Incorporating Prior Domain Knowledge into Deep Neural Networks”. In: *International Conference on Big Data* (2018). URL: <https://doi.org/10.1109/BigData.2018.8621955>.
 - [179] Emmanuel De Bézenac, Arthur Pajot, and Patrick Gallinari. “Deep Learning for Physical Processes Incorporating Prior Scientific Knowledge.” In: *Journal of Statistical Mechanics: Theory and Experiment* (2019). URL: <https://doi.org/10.48550/arXiv.1711.07970>.
 - [180] Qingyong Hu et al. “Towards Semantic Segmentation of Urban-Scale 3D Point Clouds: A Dataset, Benchmarks and Challenges”. In: *Computing Research Repository abs/2009.03137* (2020). URL: <https://arxiv.org/abs/2009.03137>.
 - [181] S. M. Iman Zolanvari et al. “DublinCity: Annotated LiDAR Point Cloud and its Applications”. In: *Computing Research Repository abs/1909.03613* (2019). URL: <http://arxiv.org/abs/1909.03613>.
 - [182] Martin Brosamle and Christoph Holscher. “Architects Seeing through the Eyes of Building Users, A Qualitative Analysis of Design Cases”. In: *2007, International Conference on Spatial Information Theory (COSIT’07)* (Jan. 2007), pp. 8–13. URL: https://www.sfbtr8.spatial-cognition.de/papers/SFB_TR%208%20Rep%20014-09_2007.pdf.
 - [183] Wang-Gyu Jeon and Eui-Myoung Kim. “Automated Reconstruction of Railroad Rail Using Helicopter-borne Light Detection and Ranging in a Train Station”. In: *Sensors and Materials* 31 (Oct. 2019), p. 3289. URL: <https://doi.org/10.18494/SAM.2019.2433>.

- [184] Benjamin Rosman and Subramanian Ramamoorthy. “Learning spatial relationships between objects”. In: *I. J. Robotic Res.* 30 (Oct. 2011), pp. 1328–1342. URL: <https://doi.org/10.1177/0278364911408155>.
- [185] Timo Hackel et al. “Semantic3D.net: A new Large-scale Point Cloud Classification Benchmark”. In: *Computing Research Repository* abs/1704.03847 (2017). URL: <http://arxiv.org/abs/1704.03847>.
- [186] Kirsten Petras et al. “Coarse-to-fine information integration in human vision”. In: *NeuroImage* 186 (2019), pp. 103–112. URL: <https://doi.org/10.1016/j.neuroimage.2018.10.086>.
- [187] Jizhi Peng et al. “A Deep Model for Joint Object Detection and Semantic Segmentation in Traffic Scenes”. In: *International Joint Conference on Neural Networks (IJCNN)*. 2020, pp. 1–8. URL: <https://doi.org/10.1109/IJCNN48605.2020.9206883>.
- [188] Zhixiong Nan et al. “A joint object detection and semantic segmentation model with cross-attention and inner-attention mechanisms”. In: *Neurocomputing* 463 (2021), pp. 212–225. URL: <https://www.sciencedirect.com/science/article/pii/S0925231221012157>.
- [189] Roozbeh Mottaghi et al. “The role of context for object detection and semantic segmentation in the wild”. English. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 2014, pp. 891–898. URL: <https://doi.org/10.1109/CVPR.2014.119>.
- [190] Jianghong Zhao et al. “The Fusion Strategy of 2D and 3D Information Based on Deep Learning: A Review”. In: *Remote Sensing* 13.20 (2021). URL: <https://www.mdpi.com/2072-4292/13/20/4029>.
- [191] International Society of Arboriculture (ISA). *Guidelines for Vegetation Management Near Power Lines*. International Society of Arboriculture, 2017.

- [192] A. Toshev and C. Szegedy. “DeepPose: Human Pose Estimation via Deep Neural Networks”. In: *Computer Vision and Pattern Recognition*. IEEE. 2014, pp. 1653–1660. URL: <https://doi.org/10.1109/CVPR.2014.214>.
- [193] G. Pavlakos et al. “Learning to Estimate 3D Human Pose and Shape from a Single Color Image”. In: *Computer Vision and Pattern Recognition*. IEEE. 2018, pp. 459–468. URL: <https://doi.org/10.48550/arXiv.1805.04092>.
- [194] Cobravisision Team. *Cobravisision*. Accessed on March 5, 2023.
- [195] Lei Ding, Hao Tang, and Lorenzo Bruzzone. “Improving Semantic Segmentation of Aerial Images Using Patch-based Attention”. In: *Computing Research Repository* abs/1911.08877 (2019). URL: <http://arxiv.org/abs/1911.08877>.
- [196] Zhengtan Wang and Yuan Zheng. “BAGNet: Branch Attention Guided Decoder for Semantic Segmentation”. In: *Proceedings of 6th International Conference on Computer Science and Artificial Intelligence*. CSAI ’22. Beijing, China: Association for Computing Machinery, 2023, pp. 106–110. ISBN: 9781450397773. URL: <https://doi.org/10.1145/3577530.3577547>.
- [197] California Public Utilities Commission. *Transmission System Components: Fact Sheet*. Year. URL: https://ia.cpuc.ca.gov/Environment/info/aspen/cltp/archive/Files_8_26_14/_3TransmissionComponentsFactSheet.pdf.
- [198] Peter Rieger and Andreas Ullrich. “Resolving range ambiguities in high-repetition rate airborne lidar applications”. In: *Electro-Optical Remote Sensing, Photonic Technologies, and Applications V*. Vol. 8186. Proceedings of SPIE Security + Defence, Prague, Czech Republic. SPIE. 2011, 81860A. DOI: 10.1117/12.898551. URL: <https://doi.org/10.1117/12.898551>.
- [199] Zhipeng Luo et al. “Detection of Individual Trees in UAV LiDAR Point Clouds Using a Deep Learning Framework Based on Multichannel Representation”. In: *IEEE Transactions on Geoscience and Remote Sensing* 60 (2022), pp. 1–15. URL: <https://doi.org/10.1109/TGRS.2021.3130725>.

- [200] Danesh Shokri et al. “An enhanced descriptor extraction algorithm for power line detection from point clouds”. In: *Journal of Applied Remote Sensing* 17.3 (May 2023), p. 036504. URL: <https://doi.org/10.1111/1745-5871.12604>.
- [201] Diogo Restolho Mateus Marçalo Lavado. “Detection of Power Line Supporting Towers via Interpretable Semantic Segmentation of 3D Point Clouds”. MA thesis. FCT: Departamento de Informática, Oct. 2022. URL: <http://hdl.handle.net/10362/145186>.
- [202] Qing Dong Wenbo Zhao and Zhengli Zuo. “A point cloud segmentation method for power lines and towers based on a combination of multiscale density features and point-based deep learning”. In: *International Journal of Digital Earth* 16.1 (2023), pp. 620–644. URL: <https://doi.org/10.1080/17538947.2023.2168770>.
- [203] Hongcan Guan et al. “UAV-lidar aids automatic intelligent powerline inspection”. In: *International Journal of Electrical Power & Energy Systems* 130 (2021), p. 106987. URL: <https://www.sciencedirect.com/science/article/pii/S0142061521002271>.
- [204] Alexey Dosovitskiy et al. “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale”. In: *Computing Research Repository* abs/2010.11929 (2020). URL: <https://arxiv.org/abs/2010.11929>.
- [205] Chuan Zhao et al. “ALS Point Cloud Classification With Small Training Data Set Based on Transfer Learning”. In: *IEEE Geoscience and Remote Sensing Letters* 17.8 (2020), pp. 1406–1410. URL: <https://doi.org/10.1109/LGRS.2019.2947608>.
- [206] Wouter M. Kouw. “An introduction to domain adaptation and transfer learning”. In: *Computing Research Repository* abs/1812.11806 (2018). URL: <http://arxiv.org/abs/1812.11806>.
- [207] Novi Patricia and Barbara Caputo. “Learning to Learn, from Transfer Learning to Domain Adaptation: A Unifying Perspective”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2014. URL: <https://doi.org/10.1109/CVPR.2014.187>.