PAN-SUNET: UTILITY CORRIDOR UNDERSTANDING USING SPATIAL LAYOUT CONSISTENCY

Maryam Jameela[1] and Gunho Sohn[1]*

[1]Department of Earth and Space Science and Engineering, Lassonde School of Engineering
York University, Canada - (maryumja, gsohn)@yorku.ca

KEY WORDS: Computer Vision, Utility Networks, Deep Learning Network, Semantic Segmentation, Spatial Layout Consistency

ABSTRACT:

The article addresses the need for a dependable and efficient computer vision system to examine utility networks with minimal human intervention, given the deteriorating state of these networks. To classify the dense and irregular point clouds obtained from the airborne laser terrain mapping (ALTM) system, which is used for data collection, we suggest a deep learning network named Panoptic-Semantic Utility Network (Pan-SUNet). The proposed network incorporates three networks to achieve voxel-based semantic segmentation and 3D object detection of the point clouds at various resolutions, including object categories in three dimensions, and predicts two-dimensional regional labels to differentiate utility and corridor regions from non-corridor regions. The network also ensures spatial layout consistency in the prediction of the voxel-based 3D network using regional segmentation. By testing the proposed approach on $67km^2$ of utility corridor data with an average density of $5pts/m^2$, the paper demonstrates the effectiveness of the technique. The proposed network outperforms the state-of-the-art baseline network, achieving an F1 score of 94% for the pylon class, 99% for the ground class, 99% for the vegetation class, and 99% for the powerline class. It also shows high performance for 3D object detection for pylon and span achieveing average precision of 99% and 92% respectively.

1. INTRODUCTION

To protect and enhance the economy, it is necessary to ensure the durability and resilience of the utility grid. This requires conducting safe, effective, and precise inspections of the utility network. Traditionally, ground crews have been deployed to conduct these inspections, which involve managing vegetation encroachment and inspecting the physical condition of the infrastructure (Kim and Sohn, 2010). In recent years, unmanned aerial vehicles (UAVs) have emerged as a cost-effective alternative to traditional data acquisition methods, but there remain significant challenges in using UAVs to collect data across entire utility corridors due to strict flying regulations, limited flight time, and constrained spatial coverage. As a result, Airborne Laser Terrain Mapping (ALTM) is still used as the primary data collection platform (Zhou et al., 2019, Pu et al., 2019). However, the process of labeling semantic features in point clouds using visual perception tasks is still challenging, expensive, and prone to errors (Kim and Sohn, 2013). Therefore, there is a significant need to automate post-data acquisition procedures to reduce user involvement and improve efficiency (Jwa et al., 2009, Wang et al., 2017).

Recent advances in deep neural networks (DNNs) have shown significant improvement in computer vision tasks. There have been successful designs of DNNs for semantic segmentation using point clouds, such as PointNet (Qi et al., 2016), PointNet++ (Qi et al., 2017), and KPConv (Thomas et al., 2019), and object detection using point cloud and 2D images. These vision tasks have been integrated into a panoptic framework to better interpret all visual features of the complete scene. However, these networks have not fully exploited the spatial arrangement of infrastructure, especially for utility corridors, nor have they embedded spatial layout consistency for global context. Previous work have not modeled the 3D transmisson line network as span to pylon relationship. Therefore, this research proposes a network with hierarchical spatial regularity that can be generalized for standard layout panoptic segmentation problems and generate a network of
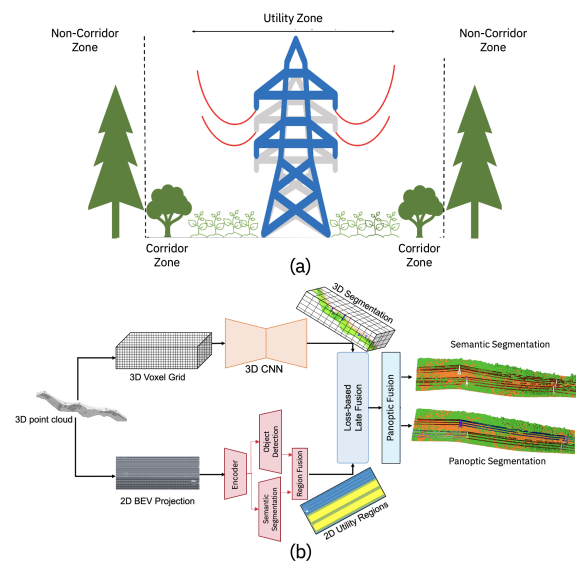

Figure 1: Pan-SUNet is a multi-dimensional and multi-resolution network that imposes the spatial layout consistency (a) through a 2D bird's-eye view (BEV) of utility regions on the outcomes of 3D panoptic segementation via loss-based late fusion and panoptic fusion (b).

object to object relationship.
This research carefully examines utility corridors and unravels spatial layout consistency, identifying hierarchies of regions (utility, corridor and non-corridor) and object classes such as ground, towers, power lines, and vegetation, as shown in figure 1.
The proposed Pan-SUNet extends SUNet (Jameela and Sohn, 2023), which dealt with only two regions, by dividing the task of extracting three regions into a 2D pipeline. This pipeline segments the 2D bird's-eye view (BEV) into two regions (corridor and non-corridor) and detects the pylons and span for the utility region. Three regions are generated by fusing the outputs from segmentation and object detection heads. Three regions fusion

---

*Corresponding author.

facilitates 3D segmentation through a loss-based late fusion module. The work also uses 2D object detection of pylons and spans to project into the third dimension using segmentation results and refine segmentation results based on the detected 3D objects (pylons and spans) and assign them instance id as well.

In the upcoming section, we will delve into the related work on spatial layout consistency, utility corridors, and computer vision tasks. The methodology will provide a detailed explanation of the proposed system, while the experiments will showcase the novelty and effectiveness of the concept.

2. LITERATURE REVIEW

This section discusses the spatial layout consistency in utility corridors and reviews the techniques of semantic segmentation, object detection and panoptic semantic segmentation.

2.1 Spatial Layout

In a number of disciplines, including cognitive science, architecture, and civil engineering, the idea of how items are arranged in a scene and their interrelationship has been investigated. Decision-making is aided by the relationship between items, which gives contextual information. In order to learn global context, it is essential to include spatial recurring patterns. The performance of tasks like railway lane extractions, road lane recognition, and 3D building modelling has been enhanced by the use of spatial arrangement (Jeon and Kim, 2019). Spatial interactions are crucial for identifying small items that can be overlooked, according to previous studies(Rosman and Ramamoorthy, 2011). This serves as the foundation for our investigation, in which we use the network to highlight the significance of embedding spatial consistency.

2.2 Utility Corridor Layout

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 (National Grid Transco UK, Wales, and USA, 11 12 2022). These zones consist of the utility zone, the corridor zone, and the non-corridor zone, each with specific regulations on the size and type of vegetation allowed (Electric Power Research Institute (EPRI), 2012). Existing literature on hierarchical relationships in visual perception, such as detecting human motion(Toshev and Szegedy, 2014), was examined to establish a baseline for the neural network.

2.3 Semantic Segmentation

In recent years, deep learning has played a significant role in semantic segmentation for 3D point clouds, with research focusing on intrinsic, extrinsic, and deep features to classify each point with an enclosing object (Liu et al., 2009). Traditional approaches for utility corridor segmentation have been purely geometric, but they have limitations such as the need for extensive preprocessing and domain expertise (Jung et al., 2020, Jwa et al., 2009). Machine learning algorithms such as support vector machines and decision trees have been used to classify utility objects, but they have limitations when applied to large-scale datasets (Jwa and Sohn, 2010, Wang et al., 2017, Kim and Sohn, 2010, Kim and Sohn, 2013, Pu et al., 2019).

Deep learning, on the other hand, offers the ability to learn features automatically, which has allowed the development of generalizable solutions. Various deep learning-based segmentation networks have been proposed, including PointNet (Qi et al., 2016),

PointNet++ (Qi et al., 2017), KPConv (Thomas et al., 2019), and RandLA(Hu et al., 2019). However, none of these methods have taken advantage of the spatial regularity found in utility infrastructure. Pan-SUNet, an extension of SUNet, has been proposed to address this limitation by fusing regions from the regional networks providing spatial guidelines to improve performance.

2.4 2D Object Detection

Object detection is a process of 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(Redmon and Farhadi, 2018) 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. 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(Lang et al., 2021), such as Oriented R-CNN (Xie et al., 2021) and Rotated Faster R-CNN (Yang et al., 2020), have been developed to use oriented bounding boxes, and have shown comparable performance to horizontal bounding box detectors while reducing background errors.

2.5 Panoptic Segmentation

Our work is 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 (Sirohi et al., 2021), is an extension of a 2D panoptic segmentation network that fuses instance and semantic segmentation across the entire scene. However, there is a lack of fusion from different dimensions in this method, specifically for airborne LiDAR point clouds.

Our research proposes a 2D panoptic segmentation approach that focuses on detecting pylons and spans, as well as segmenting regions, tailored to utility cases. Additionally, it projects the predictions into a 3D pipeline to obtain pylon instance and segmentation labels to model utility corridor.

3. METHODOLOGY

Pan-SUNet is unique novel network which takes multi-dimensional input uses multi-resolution deep learning neural networks that embeds spatial regularities between the regions and objects of interest. It comprises of three different classifiers. A two-dimensional regional prediction network (Ronneberger et al., 2015) and a two-dimensional object detection (Xie et al., 2021) network which collaborate together to generate utility region based segmentation masks and detects objects that constrains and refine the predictions of a three-dimensional network through regional fusion, loss-based late fusion and panoptic fusion.

3.1 3D Pipeline

3D semantic segmentation pipeline is major component of our system design. It can use any existing deep learning semantic segmentation networks as baseline. For our network we have utilized 3D convolution based multi-resolution encoder-decoder network with skip-connections and additive attention module in decoder. It facilities learning process incrementally on different feature maps $\frac{H}{2l} \times \frac{W}{2l} \times \frac{D}{2l} \times 32l$ on resolution level l and aggregate features to output the probability of semantic labels for 3D object classes. Segmentation head generates the confidence score against each class which is then passed through a loss-based late
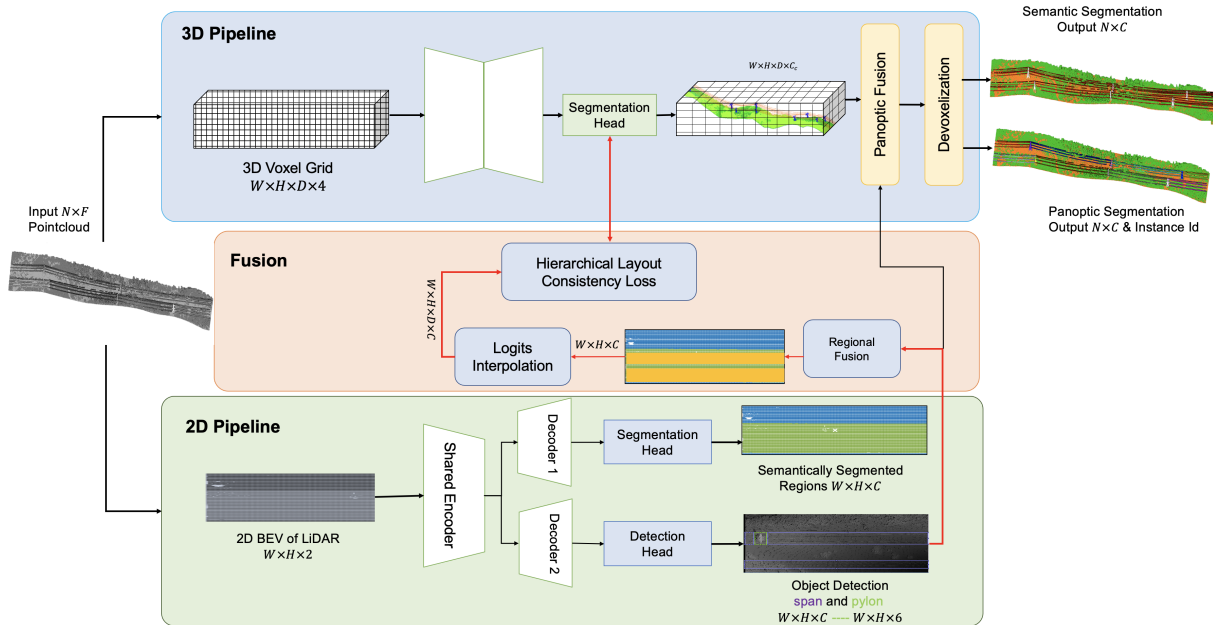
Figure 2: Pan-SUNet generates 3D voxel grids and 2D BEVs from point clouds, combining multi-resolution 3D semantic segmentation and shared encoders for corridor/non-corridor detection, spatial consistency, and label adjustment onto 3D objects.

fusion module to refine and constrain the predictions using spatial layout consistency and back-propagate the loss to better learn deep features.

3.1.1 Panoptic Segmentation The 3D semantic segmentation network's output and object detection in 2D pipeline for pylons and spans are combined by this module that leverages semantic spatial layout consistency. Our work 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 our 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 the figure 2, pylon and span instances are color-coded according to their instance IDs.

3.2 2D BEV 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 our capacity for understanding the larger context. Our three-dimensional segmentation network is only capable of encoding the local context and does not have the coarser scene information required to encode the spatial layout and item global association. In order to solve this problem, we employ a 2D BEV pipeline that fuses the missing data using a loss-based late fusion module. 2D Pipeline consists of two classifiers; a regional semantic segmentation and a two-dimensional object detection.

3.2.1 Regional Semantic Segmentation We can use any simple 2D segmentation network for regional segmentation. In our specific design we are using a U-shaped encoder-decoder which generate 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.
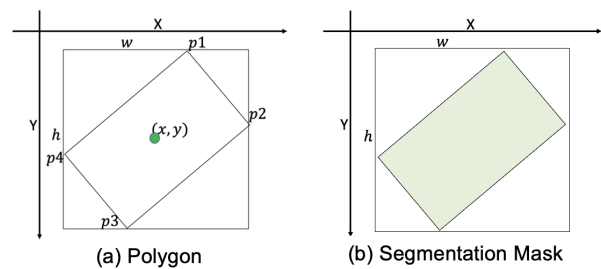

Figure 3: Regional fusion generation of segmentation shows a) polygon generated from step 1 and b) segmentation mask generated from polygon.

3.2.2 Object Detection To take advantage of the regional semantic layout of utility corridors, we added a separate module for 2D object detection with a shared encoder. The utility community divides layouts into three regions: a utility region with pylons and wires, a corridor region with a buffer around the utility region that can contain 3-5 meters tall trees, and a non-corridor region that can have tall trees and buildings. To differentiate between the corridor and non-corridor region can be difficult when non-corridor regions don't have extremely tall trees. Therefore, we merged the corridor and utility regions and performed regional segmentation to generate segmentation masks. We detected span and pylon objects using a shared encoder, which were used to update the regional segmentation mask with the utility corridor class. These regionally fused segmentation masks can improve the prediction of 3D objects of interest. Our 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 and consists of an oriented regional proposal network and a regional classifier and regressor. The regional proposal network takes an encoded feature and 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 spatial location of the object or rotated bounding box. The network is shown in figure 2.

3.3 Fusion

Fusion Module have four key modules, a) regional fusion, b) logits interpolation, c) hierarchical layout consistency loss function, and d) panoptic fusion.

3.3.1 Regional Fusion As discussed in above section we have two classifiers: segmentation classifier outputs a 2D masks with two regional classes of corridor and non-corridor regions. Corridor region also have utility region merged into it. This fusion performs following steps to generate three classes based segmentation mask using predicted oriented bounding boxes and predicted regional segmentation.

- Step 1: Convert oriented bounding boxes to polygons using following equations 1,2,3, and 4 which generates four vertices $p1, p2, p3, p4$.

$$p1 = (x, y - \frac{h}{2}) + (\delta\alpha, 0) \tag{1}$$

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

$$p3 = (x, y + \frac{h}{2}) + (-\delta\alpha, 0) \tag{3}$$

$$p4 = (x - \frac{w}{2}, y) + (0, -\delta\beta) \tag{4}$$

- Step 2: Estimate the segmentation masks from polygons as shown in figure 3.

- Step 3: Fuse the labels from regional segmentation and utility region label generated in previous step as shown in figure 4

3.3.2 Logits Interpolation Module It is a simple module that takes the projection matrix between 3D voxels and 2D BEV and converts the 2D logits for regional network into a 3D $H \times W \times D \times C_P$ by exploring the one-to-many relationship between the two representations.

3.3.3 Loss-based Late Fusion It uses the hierarchical layout consistency loss function as used in SUNet as shown in equation 5 to impose the spatial layout consistency on outcome. This loss function now uses three classes instead of two to impose better spatial regularities.

$$L = \frac{1}{M} \sum_{p=1}^{P} \sum_{c=1}^{C} \sum_{m=1}^{M} (w_p w_c) \times (y_m^{P_c} \times (log(h_\Theta(x_m, c, p)))) \tag{5}$$

3.3.4 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 segemntation results $W \times H \times D \times C$ and adjust the z-axis for 3D object detection of pylon and spans. The semantic segmentation results of pylon and powerline help in estimating 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. These 3D bounding boxes then help in adjust the prediction and misclassification with in 3D box for refining semantic segmentation results as scene table 4.2.
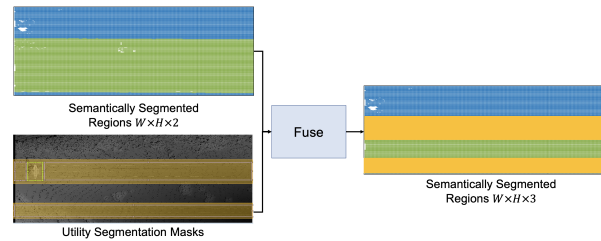

Figure 4: Regional fusion from 2D regional segmentation and utility segmentation mask.

3.4 Voxelization and BEV Projection

A voxel grid is produced by pre-processing the raw point cloud and determining a mean value for all the points that lie within each 3D voxel as the input representation for our segmentation network. Depending on the chosen voxel size, this voxel grid strikes a compromise between efficiency and effectiveness. To make label projection simple, the network also keeps a projection matrix from the voxel grid to the unprocessed point cloud. A bird's eye view (BEV), on the other hand, is a 2D depiction of a 3D point cloud. Our 2D BEV pipeline makes use of a 3D scene's XY-projection, in which each pixel corresponds to a location's points. The best BEV for obtaining global context for regional and item prediction is produced by the XY-projection.Our projection matrix between the 3D voxel grid and the BEV allows for projection compatibility between feature spaces to combine the 2D and 3D predictions, improving the use of spatial layout consistency.

4. EXPERIMENTS AND RESULTS

In our study, we performed a comparative analysis to highlight the importance of the hierarchical layout consistency. We conducted experiments on a test set that assessed the performance of our proposed method on four important classes for the utility industry: ground, pylon, powerline, and vegetation. These classes are crucial for predictive maintenance of utility networks.

4.1 Dataset

We used a Riegl Q560 laser scanner to gather data across a $67km^2$ region in Steamboat Springs, Colorado, in the United States. The acquired data were later split into train and test sets in order to conduct experiments. The first $8km^2$ of the dataset served as testing, and the remaining data served as network training. The collection consisted of 67 non-covering scenes in total, each having millions of focuses and an average thickness of $5pts/m^2$. Using Terrasolid's point cloud handling software (Team, 2023b), we manually labelled the data to produce our groundtruth which require technical and industry experience. The training dataset contained five classes: powerline, low vegetation, pylon, and medium-high vegetation. We combined the low vegetation class with the ground class since low vegetation covered the most of the ground. Also, we named our regional classes based on the literature of utility communities.

4.2 2D Object Detection Groundtruth

To produce object detection labels for pylons and spans as horizontal bounding boxes, the matlab image tool is used; for oriented bounding boxes, Labelme (Wada, 2021) is utilised. Our experiments are conducted on two different types of bounding boxes. The horizontal bounding boxes usually in oriented object cases include alot of background hence we compared the performance using both systems as shown in figure 5.

Methods	Categories	Bounding Box	AP	AR	AP Pylon	AP Span
Faster RCNN	2D Two-stage	Horizontal	97.60	88.40	100.0	96.12
Yolo v3	2D Single stage	Horizontal	97.30	79.21	96.14	92.15
DAFNe	2D Anchor-free Single-stage	Horizontal	95.45	92.25	94.14	92.95
Oriented RCNN	2D Oriented Two-stage	Oriented	**95.40**	**96.91**	**100.0**	90.8
Rotated Faster RCNN	2D Oriented Two-stage	Oriented	**95.40**	**96.91**	**100.0**	**91.0**

Table 1: Comparative study for 2D object detection for pylon and span objects for average precision (AP), average recall (AR), pylon average precision (AP) and span average precision (AP) over IoU=0.5.

Methods	Categories	Bounding Box	AP	AR	AP Pylon	AP Span
Pan-SUNet (ours)	2D Object Detection	Oriented	95.40	96.91	100.0	90.8
Pan-SUNet (ours)	3D Object Detection	Oriented	97.40	95.91	99.0	92.0

Table 2: The performance of Pan-SUNet for detecting 2D pylon and span objects, as well as projecting them into 3D, was evaluated using average precision (AP), average recall (AR), pylon AP, and span AP at IoU=0.5.
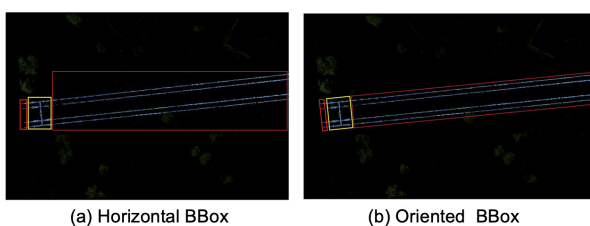

(a) Horizontal BBox   (b) Oriented BBox

Figure 5: Visualizing the bounding boxes of span and pylon a) horizontal bounding boxes with greater background and b) oriented bounding boxes with polygons.

4.3 Experimental Configuration

4.3.1 2D Object Detection To prepare for our experiments, we pre-trained various 2D object detection network and compared the performance. We generated the 2D XY- Projection of point cloud using $640m^2$ grid where each pixel is of $1m^2$. Network's input have three feature channels that represented the normalized standard deviation of elevation between [0-1], normalized standard deviation of elevation values between 0 and 255, and binarized standard deviation of elevation values per pixel. We conducted a detailed comparative study for 2D object detection. We selected state-of-the-art networks to compare performance on horizontal and oriented bounding boxes along two-stage and one stage systems such as faster rcnn, rotated faster rcnn, oriented rcnn, dafne, and yolov3 and selected orieneted rcnn as our baseline due to smoothness and effectiveness of results.

4.3.2 2D Regional Semantic Segmentation We pre-trained our 2D regional prediction network on half of the scenes for global regional prediction of spatial layout in order to prepare for our trials. According to the GPS time of flight line, each scene was separated into four smaller scenes, each of which was then projected onto a 640x640 2D BEV grid with pixels measuring $1m^2$ and three feature channels that represented the normalized standard deviation of elevation between [0-1], normalized standard deviation of elevation values between 0 and 255, and binarized standard deviation of elevation values per pixel. The 2D network's input size was $640 \times 640 \times 640 \times 3$, while the output, which represented the confidence ratings for regional classes, was $640 \times 640 \times 3$. In order to prevent overfitting, we pre-trained the network using K-cross validation with a batch size of 1 and a total of 100 epochs. Random rotation, horizontal flip, and vertical flip were all used to augment the data. Training was done on two GPU RTX 6000 and took between 4-5 hours with inference taking about 30 seconds.

4.3.3 3D Semantic Segmentation We constructed a voxel grid with a size of $640 \times 640 \times 448$ and a voxel size 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. To choose these features, we used the feature engineering study that was presented by SUNet. A confidence score against 3D classes is produced by Pan-SUNet as $32 \times 32 \times 448 \times 5$ (background, pylon, powerline, vegetation, and ground). The final prediction provides a true label based on the greatest confidence score and projects voxel labels on points using the projection matrix. Using two GPU RTX 6000, Pan-SUNet was trained for 100 epochs over the course of 48–60 hours, with inference requiring only a few minutes.

4.4 Evaluation Matrices

4.4.1 Object Detection: Two metrics, Average Precision (AP) and Average Recall (AR), are used to assess how efficiently object detection networks locate and identify items in an image. The area under a precision-recall or recall-precision curve, which accounts for the IoU threshold to separate true positives from false positives, is computed for both metrics.

4.4.2 Pan-SUNet In the paper, the performance of the semantic segmentation model is evaluated using F1 scores for four different classes: vegetation, powerline, pylon, and ground. The F1 score is calculated by taking the harmonic mean of the precision and recall for each class. Recall measures the number of true positives that are correctly identified, while precision measures the proportion of predicted positives that are actually true positives. By considering both precision and recall, the F1 score provides an overall measure of the model's performance that accounts for both accuracy and completeness.

4.5 Results

We conducted compartive study of 2D object detection as shown in table 1 for selecting baseline for 2D object detectuon. Our object detection results in table 2 shows the promising results for 2D and projecting those results onto third dimension. Results indicate comparable performance in 2D and 3D facilitating the modeling of utility corridor through pylon to span relationship. We also evaluated Pan-SUNet's performance for semantic segmentation against various SUNet versions, Attention 3D, and pre-trained RandLA. Our results demonstrated that the regional

Methods	Pylon			Ground			Veg			Powerline		
	Rec	Prec	F1	Rec	Prec	F1	Rec	Prec	F1	Rec	Prec	F1
3D AUNet (baseline)	72.0	84.0	77.0	93.0	94.0	93.0	97.0	97.0	97.0	84.0	98.0	91.0
RandLA	75.0	92.0	83.5	95.0	96.0	95.5	76.0	87.0	81.5	86.5	98.0	91.3
SUNet+MFA+FS	78.0	97.0	87.0	**100.0**	99.0	**100.0**	99.0	99.0	99.0	**99.0**	97.0	98.0
SUNet+MFA	82.0	96.0	89.0	99.0	99.0	99.0	99.0	99.0	99.0	98.0	99.0	99.0
Pan-SUNet-vanilla (ours)	86.0	93.0	90.0	99.0	99.0	99.0	99.0	99.0	99.0	98.0	99.0	99.0
Pan-SUNet + Panoptic Fusion (ours)	90.0	**99.0**	**94.0**	99.0	**99.0**	99.0	**99.0**	**99.0**	**99.0**	98.0	**100.0**	**99.0**

Table 3: Comparative study of Pan-SUNet to SUNet + MFA (multi feature aggregation) module, SUNet + MFA + FS (feature smoothing), Atention UNet and RandLA on recall (Rec), precision (Prec) and F1 score (F1).

fusion module in Pan-SUNet offers significant benefits in terms of higher recall and F1 score for the pylon class in table 3, highlighting the advantages of including spatial arrangement context on a global scale. We used a voxel-based network, which delivers comparable quality performance to point-based networks but with faster inference times (10x faster). Our network outperforms currently available commercial software tools that depend on network maps for predictive analysis or human labeling of powerlines and pylons. Visualization in figure 7 shows the contribution of each stage in the Pan-SUNet framework to model the 3D utility transmission lines infrastructure.

5. CONCLUSION

Our research has shown that ambiguous regions can be given global context by using pre-tasks like object detection. We have also demonstrated that pylon and span objects can be identified using a basic 2D BEV map. This information can then be utilised to extract a utility zone for segmentation, 2D powerline modelling, and investigation of vegetation encroachment. By transforming utility items into utility zones and employing them to impose a hierarchy on the 3D semantic segmentation of objects of interest, our Pan-SUNet network makes use of this module. Our work have made major contribution in convertin 2D object detection into 3D objects and adjusted the misclassification. We intend to concentrate on real-time semantic segmentation of the utility corridor in the future.

ACKNOWELDGEMENT

REFERENCES

Brosamle, M. and Holscher, C., 2007. Architects seeing through the eyes of building users, a qualitative analysis of design cases. 2007, International Conference on Spatial Information Theory (COSIT'07) pp. 8–13.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. and Ronneberger, O., 2016. 3d u-net: Learning dense volumetric segmentation from sparse annotation. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI,2016) 9901, pp. 424–432.

Electric Power Research Institute (EPRI), 2012. Vegetation Management - Best Practices for Transmission Corridors. Electric Power Research Institute.

Guan, H., Yu, Y., Li, J., Ji, Z. and Zhang, Q., 2016. Extraction of power-transmission lines from vehicle-borne lidar data. International Journal of Remote Sensing 37(1), pp. 229–247.

Hackel, T., Savinov, N., Ladicky, L., Wegner, J. D., Schindler, K. and Pollefeys, M., 2017. Semantic3d.net: A new large-scale point cloud classification benchmark. CoRR.

Haldekar, M., Ganesan, A. and Oates, T., 2017. Identifying spatial relations in images using convolutional neural networks. In: 2017 International Joint Conference on Neural Networks (IJCNN), pp. 3593–3600.

Hu, Q., Yang, B., Khalid, S., Xiao, W., Trigoni, N. and Markham, A., 2020. Towards semantic segmentation of urban-scale 3d point clouds: A dataset, benchmarks and challenges. CoRR.

Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N. and Markham, A., 2019. Randla-net: Efficient semantic segmentation of large-scale point clouds. CoRR.

International Society of Arboriculture (ISA), 2017. Guidelines for Vegetation Management Near Power Lines. International Society of Arboriculture.

Isensee, F., Kickingereder, P., Wick, W., Bendszus, M. and Maier-Hein, K. H., 2018. Brain tumor segmentation and radiomics survival prediction: Contribution to the BRATS 2017 challenge. CoRR.

Jameela, M. and Sohn, G., 2023. Spatial layout consistency for 3d semantic segmentation.

Jeon, W.-G. and Kim, E.-M., 2019. Automated reconstruction of railroad rail using helicopter-borne light detection and ranging in a train station. Sensors and Materials 31, pp. 3289.

Jung, J., Che, E., Olsen, M. J. and Shafer, K. C., 2020. Automated and efficient powerline extraction from laser scanning data using a voxel-based subsampling with hierarchical approach. ISPRS Journal of Photogrammetry and Remote Sensing 163, pp. 343–361.

Jwa, Y. and Sohn, G., 2010. A multi-level span analysis for improving 3d power-line reconstruction performance using airborne laser scanning data. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 38, pp. 97–102.

Jwa, Y., Sohn, G. and Kim, H., 2009. Automatic 3d powerline reconstruction using airborne lidar data. IAPRS 38, pp. 105–110.

Kim, H. and Sohn, G., 2010. 3d classification of powerline scene from airborne laser scanning data using random forests. IAPRS 38, pp. 126–132.

Kim, H. and Sohn, G., 2011. Random forests based multiple classifier system for power-line scene classification. Vol. XXXVIII-5/W12.

a) Ground Truth b) Faster RCNN c) Rotated Faster RCNN
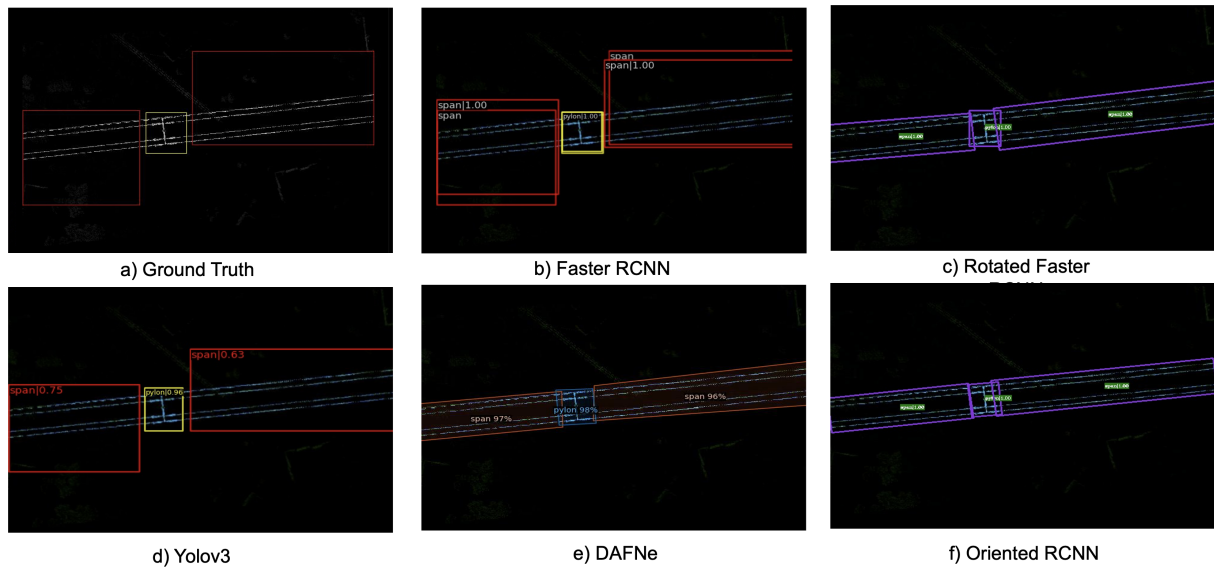
d) Yolov3 e) DAFNe f) Oriented RCNN

Figure 6: Visualization of results of 2D Objects span and pylong; a) Groundtruth, b) Faster RCNN, c) Rotated Faster RCNN, d) Yolo v3, e) DAFNe, and f) Oriented RCNN
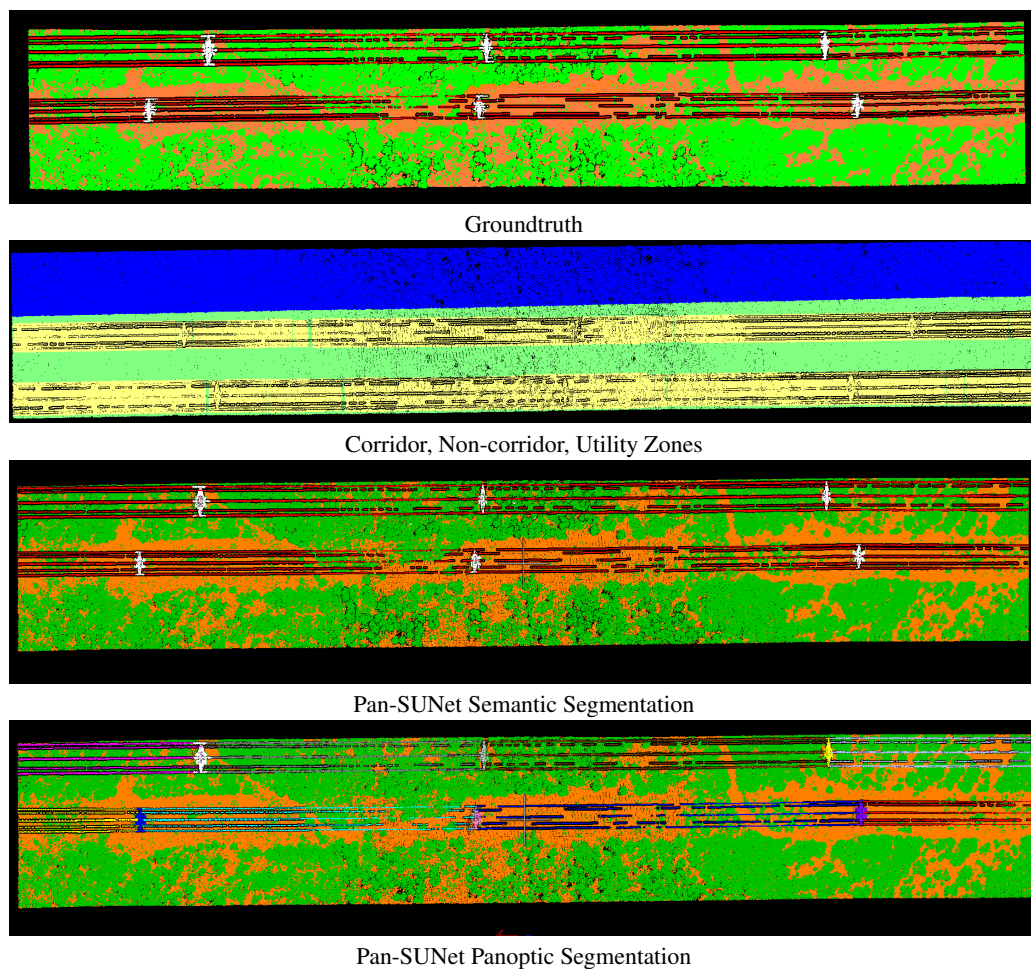

Groundtruth

Corridor, Non-corridor, Utility Zones

Pan-SUNet Semantic Segmentation

Pan-SUNet Panoptic Segmentation

Figure 7: Visualization of results; Groundtruth, Three Regions (Blue: non-corridor, green: corridor, and yellow: utility zone), Utiliy Zone, and Pan-SUNet semantic segmentation and Pan-SUNet panoptic segmentation (ours). Blue: pylon, red: powerline, green: high vegetation and orange: ground

Kim, H. and Sohn, G., 2013. Point-based classification of power line corridor scene using random forests. Photogrammetric Engineering and Remote Sensing 79, pp. 821–833.

Lang, S., Ventola, F. and Kersting, K., 2021. Dafne: A one-stage anchor-free deep model for oriented object detection. CoRR.

Liu, Y., Li, Z., Hayward, R., Walker, R. and Jin, H., 2009. Classification of airborne lidar intensity data using statistical analysis and hough transform with application to power line corridors. In: 2009 Digital Image Computing: Techniques and Applications, pp. 462–467.

Mottaghi, R., Chen, X., Liu, X., Cho, N., Lee, S., Fidler, S., Urtasun, R. and Yuille, A., 2014. The role of context for object detection and semantic segmentation in the wild. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, pp. 891–898. Publisher Copyright: © 2014 IEEE.; 27th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014 ; Conference date: 23-06-2014 Through 28-06-2014.

Nan, Z., Peng, J., Jiang, J., Chen, H., Yang, B., Xin, J. and Zheng, N., 2021. A joint object detection and semantic segmentation model with cross-attention and inner-attention mechanisms. Neurocomputing 463, pp. 212–225.

National Grid Transco UK, Wales, and USA, 11 12 2022. Design guidelines for development near high-voltage overhead lines. https://www.nationalgrid.com/uk/electricity-transmission/publications/design-guidelines-development-near-high-voltage-overhead-lines. Accessed: DD Month YYYY.

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M. C. H., Heinrich, M. P., Misawa, K., Mori, K., McDonagh, S. G., Hammerla, N. Y., Kainz, B., Glocker, B. and Rueckert, D., 2018. Attention u-net: Learning where to look for the pancreas. CoRR.

Pavlakos, G., Zhou, X., Derpanis, K. G. and Daniilidis, K., 2018. Learning to estimate 3d human pose and shape from a single color image. In: Computer Vision and Pattern Recognition, IEEE, pp. 459–468.

Peng, J., Nan, Z., Xu, L., Xin, J. and Zheng, N., 2020. A deep model for joint object detection and semantic segmentation in traffic scenes. In: 2020 International Joint Conference on Neural Networks (IJCNN), pp. 1–8.

Petras, K., ten Oever, S., Jacobs, C. and Goffaux, V., 2019. Coarse-to-fine information integration in human vision. NeuroImage 186, pp. 103–112.

Pu, S., Xie, L., Ji, M., Zhao, Y., Liu, W., Wang, L., Yang, F. and Qiu, D., 2019. Real-time powerline corridor inspection by edge computing of uav lidar data. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-2/W13, pp. 547–551.

Qi, C. R., Su, H., Mo, K. and Guibas, L. J., 2016. Pointnet: Deep learning on point sets for 3d classification and segmentation. CoRR.

Qi, C. R., Yi, L., Su, H. and Guibas, L. J., 2017. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. CoRR.

Redmon, J. and Farhadi, A., 2018. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

Ren, S., He, K., Girshick, R. and Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In: Advances in neural information processing systems, pp. 91–99.

Ronneberger, O., Fischer, P. and Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. CoRR.

Rosman, B. and Ramamoorthy, S., 2011. Learning spatial relationships between objects. I. J. Robotic Res. 30, pp. 1328–1342.

Sirohi, K., Mohan, R., Büscher, D., Burgard, W. and Valada, A., 2021. Efficientlps: Efficient lidar panoptic segmentation. CoRR.

Team, C., 2023a. Cobravision. https://www.cobravision.com. Accessed on March 5, 2023.

Team, T., 2023b. Terrasolid. https://www.terrasolid.com. Accessed on March 5, 2023.

Thomas, H., Qi, C. R., Deschaud, J., Marcotegui, B., Goulette, F. and Guibas, L. J., 2019. Kpconv: Flexible and deformable convolution for point clouds. CoRR.

Toshev, A. and Szegedy, C., 2014. Deeppose: Human pose estimation via deep neural networks. In: Computer Vision and Pattern Recognition, IEEE, pp. 1653–1660.

Wada, K., 2021. labelme. https://github.com/wkentaro/labelme. Accessed on March 5, 2021.

Wang, Y., Chen, Q., Liu, L., Zheng, D., Li, C. and Li, K., 2017. Supervised classification of power lines from airborne lidar data in urban areas. Remote Sensing.

Xie, X., Cheng, G., Wang, J., Yao, X. and Han, J., 2021. Oriented R-CNN for object detection. CoRR.

Yang, J., Huang, Z., Huang, M., Zeng, X., Li, D. and Zhang, Y., 2019. Power line corridor lidar point cloud segmentation using convolutional neural network. In: Z. Lin, L. Wang, J. Yang, G. Shi, T. Tan, N. Zheng, X. Chen and Y. Zhang (eds), Pattern Recognition and Computer Vision, Springer International Publishing, Cham, pp. 160–171.

Yang, S., Pei, Z., Zhou, F. and Wang, G., 2020. Rotated faster r-cnn for oriented object detection in aerial images. In: Proceedings of the 2020 3rd International Conference on Robot Systems and Applications, ICRSA '20, Association for Computing Machinery, New York, NY, USA, p. 35–39.

Zhao, J., Wang, Y., Cao, Y., Guo, M., Huang, X., Zhang, R., Dou, X., Niu, X., Cui, Y. and Wang, J., 2021. The fusion strategy of 2d and 3d information based on deep learning: A review. Remote Sensing.

Zhou, M., Li, K. Y., Wang, J. H., Li, C. R., Teng, G. E., Ma, L., Wu, H. H., Li, W., Zhang, H. J., Chen, J. Y. and Chen, L. S., 2019. Automatic extraction of power lines from uav lidar point clouds using a novel spatial feature. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences IV-2/W7, pp. 227–234.

Zolanvari, S. M. I., Ruano, S., Rana, A., Cummins, A., da Silva, R. E., Rahbar, M. and Smolic, A., 2019. Dublincity: Annotated lidar point cloud and its applications. CoRR.