Extraction of Pole-like Road Objects from MMS Point Clouds Using Deep Learning and Geometric-Topological Feature Fusion

Shu Su, Masataka Shirai, Hiroyuki Yokota

AERO TOYOTA CORPORATION, Japan - (shuu-so, masataka-shirai, hiroyuki-yokota) @aerotoyota.co.jp

Keywords: Point Clouds, Pole-like Objects, Deep Learning, Geometric-Topological Fusion, Cross-Domain Generalization

Abstract

This paper presents a fusion framework for the automatic extraction of pole-like road objects, including traffic lights, road signs, streetlights, and utility poles, from Mobile Mapping System (MMS) point clouds. The proposed method combines KPConv-based semantic segmentation with geometric-topological reasoning, enabling structural completion and heuristic filtering of nearby clutter without retraining or additional annotated data. The framework was trained on 8 km of manually annotated MMS data collected in the Kinki region of Japan and evaluated on two large-scale datasets: (i) a 26 km MMS dataset from Hokkaido (≈2.53 billion points) acquired using the same LiDAR sensor, and (ii) the Paris-Lille-3D benchmark (France) captured with a different LiDAR sensor. Quantitative evaluation demonstrates that the proposed fusion framework consistently outperforms the KPConv baseline across all datasets, particularly in recall and F_1-score. On the Hokkaido dataset, recall improved from 0.7952 to 0.8924 (+0.0972), and the F_1-score increased from 0.8263 to 0.8689 (+0.0426), reflecting successful reconstruction of lamp tops, signal arms, and previously unseen snow delineator posts (snow poles). On the Paris-Lille-3D benchmark, representing a cross-sensor and cross-domain scenario, recall improved from 0.5109 to 0.6656 (+0.1547), while the F_1-score increased from 0.6230 to 0.7032 (+0.0802). In terms of computational efficiency, the 26 km Hokkaido dataset was processed in under 13 hours on a single NVIDIA Quadro RTX 8000. Overall, these results confirm that the proposed deep- learning-geometry-topology fusion framework achieves high accuracy, robust generalization, and practical scalability for large-scale road-asset mapping and digital-twin generation.

1. Introduction

Recent developments in Mobile Mapping Systems (MMS) have significantly enhanced the ability to capture dense, high-resolution point clouds in complex urban environments. These 3D spatial data now underpin a wide range of applications, including autonomous navigation, urban infrastructure management, and digital-twin construction (Elhashash et al., 2022). Within these urban environments, pole-like road objects—streetlights, traffic signs, signal poles, and utility poles—play a crucial role in ensuring traffic safety and urban asset management (Brenner, 2009; El-Halawany and Lichti, 2013).

However, the automatic and reliable extraction of pole-like objects from MMS point clouds remains an open challenge. In real-world datasets, variations in geometry, occlusion from surrounding structures, and highly non-uniform point-density distributions caused by scanning geometry and sensor characteristics severely affect segmentation robustness (Qi et al., 2017a; Thomas et al., 2019). These factors become especially problematic in large-scale and heterogeneous environments, where generalization across domains and sensors becomes critical.

Previous work on pole-like extraction can be broadly categorized into three groups: (1) image-based, (2) point-cloud-based, and (3) hybrid approaches that combine imagery with LiDAR.

Image-based methods typically rely on texture, color, and shape; however, their performance is often sensitive to lighting conditions and occlusion (Zhang et al., 2018). Hybrid approaches that combine camera imagery and LiDAR point clouds have also been explored for the classification of pole-like roadside objects in MMS data (Mori et al., 2018), and more generally for improving semantic consistency across modalities (Jaritz et al., 2020). However, such approaches require accurate cross-sensor calibration and stable multimodal alignment, which are often difficult to maintain in long-distance MMS acquisitions (Tan et al., 2024). For this reason, this study focuses exclusively on MMS point clouds.

Earlier geometric approaches—including RANSAC-based fitting, region-growing, and density-driven clustering—provided interpretable pipelines and were generally robust to noise (Belton and Lichti, 2006; El-Halawany and Lichti, 2013; Shi et al., 2018). The structural regularity of pole-like objects has also been exploited in classical approaches, such as the use of pole patterns for vehicle localization (Brenner, 2009), further highlighting the importance of geometric and spatial relationships in such objects. However, despite these advantages, these approaches rely heavily on parameter tuning and struggle to differentiate poles from other vertical objects with similar shapes (Cabo et al., 2014). To improve structural completeness, Wang et al. (2021) proposed a multi-scale fusion framework that exploits vertical continuity and structural constraints to improve the completeness of pole-like object segmentation in road scenes. Similarly, Li and Cheng (2022) presented a supervoxel-based approach that explicitly separates vertical pole shafts from their attachments and associates them using spatial correspondence, enabling individual extraction of complex pole-like objects. In addition, Nurunnabi et al. (2023) combined robust PCA-based saliency features with DBSCAN clustering to detect and segment pole-like objects from mobile laser scanning data, demonstrating the continued relevance of density-based clustering in this context.

More recently, machine-learning-based approaches have also been explored for identifying pole-like objects in urban roadway MLS data, helping bridge the gap between purely geometric pipelines and fully deep-learning-based semantic segmentation frameworks (Yadav et al., 2022). In addition, Zhang et al. (2024) proposed Pole-NN, a few-shot learning framework specifically

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-145-2026

designed for pole-like object classification in LiDAR point clouds, highlighting the growing interest in learning-based, pole-specific representations for road infrastructure analysis.

Deep-learning-based networks designed for point clouds—such as PointNet (Qi et al., 2017a), PointNet++ (Qi et al., 2017b), KPConv (Thomas et al., 2019), and RandLA-Net (Hu et al., 2020)—have achieved notable progress in semantic segmentation. Yet, domain shift remains a persistent issue in point cloud semantic segmentation, as models trained in one environment often degrade when applied to different cities, acquisition conditions, or sensing modalities (Jaritz et al., 2020). Small or infrequent classes, such as poles, present an additional challenge because they are typically underrepresented in large-scale urban point clouds and are associated with limited contextual support. More recently, transformer-based architectures for point clouds, including Point Transformer (Zhao et al., 2021), Point Transformer v2 (Wu et al., 2022), and Point Transformer v3 (Wu et al., 2024), have demonstrated strong capability in modeling long-range contextual dependencies and structural relationships in an end-to-end manner. These approaches inherently capture both local geometry and global context, which are particularly beneficial for complex scenes containing thin or elongated structures.

More recently, hybrid pipelines have emerged that combine deep-learning predictions with explicit geometric or topological reasoning. Wen et al. (2019) applied U-Net segmentation followed by geometric clustering; Wang et al. (2023) combined α-shape modelling; and Tardy et al. (2023) incorporated LiDAR segmentation with geometric measurements for road inventory generation. These approaches reflect a growing trend toward "DL for semantics → geometry for structure," improving completeness and structural consistency. However, most existing frameworks focus on planar or surface-type objects, leaving slender, attachment-rich pole structures insufficiently addressed. In addition, large-scale, multi-region evaluations—particularly those involving different acquisition domains and sensor types—also remain limited (Truong-Hong et al., 2022).

To address these limitations, this study proposes a geometry-topology completion framework that integrates KPConv semantic segmentation with post-hoc geometric and topological reasoning. The method reconstructs missing or fragmented pole components (e.g., signal arms, lamp fixtures and region-specific elements such as snow delineator posts (snow poles)), while maintaining semantic consistency. Unlike conventional post-processing filters, the proposed approach performs semantic-structural fusion in a unified reasoning pipeline, yielding interpretable results without retraining or additional annotated data.

Comprehensive experiments were conducted on large-scale MMS data from Kinki (8 km) and Hokkaido (26 km) in Japan, and the Paris-Lille-3D benchmark (France). These datasets differ in terms of geographic environment, urban morphology, and LiDAR sensors (Riegl VQ-450 vs. Velodyne HDL-32E).

The proposed method achieved consistent improvements in recall and F_1-score across all datasets, with only a slight decrease in precision, while processing the 26 km Hokkaido dataset in less than 13 hours on a single GPU—demonstrating both high accuracy and large-scale computational efficiency.

The primary contributions of this study are summarized as follows:

1. A unified semantic-structural fusion framework that combines deep-learning-based segmentation with geometric-topological reasoning for reliable extraction of pole-like road objects from MMS point clouds.
2. Geometric-topological descriptors that enhance vertical continuity, attachment recognition, and structural completion without requiring retraining or additional annotation.
3. Comprehensive cross-domain validation across multiple regions (Kinki, Hokkaido), countries (Japan, France), and LiDAR sensors (Riegl, Velodyne), demonstrating strong adaptability and efficiency for large-scale urban mapping.

2. Methodology

The proposed method combines deep-learning-based semantic segmentation with geometric and topological reasoning to achieve robust extraction of pole-like road objects from MMS point clouds. It combines semantic inference from a pretrained KPConv model with a graph-based geometric completion process. This enables the reconstruction of partially missing structures—such as signal arms, curved lamp brackets, and region-specific structures, such as snow poles—without retraining or additional annotated data.

2.1 Overview of the Processing Pipeline

The overall workflow consists of five main stages:
(1) Deep-learning Semantic Segmentation
(2) Clustering and Buffer Construction
(3) Candidate Selection and Denoising
(4) Graph Construction and Connectivity Analysis
(5) Final Pole Classification

The pipeline begins with semantic inference using KPConv, followed by geometric-topological reasoning that reconstructs discontinuous or occluded structures through graph connectivity and tangent-based propagation. This hybrid approach substantially improves recall and F_1-score, while yielding only a slight reduction in precision compared with methods that rely solely on learning-based techniques.

2.2 Deep-learning Semantic Segmentation

Let the input MMS point clouds be defined as:

$$P = \{(x_i, y_i, z_i)\}_{i=1}^{N}, \quad (1)$$

where N is the number of points.

The KPConv model assigns each point to one of the following five semantic classes:
1. Unclassified (class0)
2. Ground (class1)
3. Guardrail (class2)
4. Power-line (class3)
5. Pole-like (class4)

For geometric-topological reasoning, these outputs are grouped into three sets:

$$C_G = \{1\},\ C_{KP} = \{4\},\ C_{ADJ} = \{0, 2, 3\}, \quad (2)$$

where
$\boldsymbol{C_G}$ contains ground points (class1),

$\boldsymbol{C_{KP}}$ contains the pole-like points (traffic lights, road signs, streetlights, utility poles, etc.; class 4),
$\boldsymbol{C_{ADJ}}$ holds candidate points (classes 0, 2, and 3) that may be promoted to pole-like points through the geometric-topological reasoning described in Sections 2.3-2.6.

2.3 Clustering and Buffer Construction

Points assigned as pole-like ($\mathcal{C}_{\mathrm{KP}}$, Eq.2) are initially grouped using DBSCAN (Ester et al., 1996):

$$\mathrm{DBSCAN}(\varepsilon, \mathrm{minPts}), \tag{3}$$

where ε denotes the neighborhood radius, and minPts denotes the minimum number of points required to form a cluster.

For each cluster C_k, a cylindrical buffer $\mathcal{B}_k$ is created:

$$B_k = \{\, p \mid |p_{xy} - c_{k,xy}| < r_{xy},\ z_{\min(k)} \le z \le z_{\max(k)} \,\}, \tag{4}$$

where r_{xy} is the horizontal buffer radius and the vertical bounds are

$$z_{\min(i)} = \mathrm{z}_{\mathrm{ground}}\ ,\ z_{\max(i)} = \max_{q \in C_i} q_z + z_{\mathrm{buf}}\ , \tag{5}$$

The local ground height z_{ground} is computed as the median value of all points predicted as ground ($\boldsymbol{C_G}$, Eq. 2) within the same buffer footprint. The added vertical margin z_{buf} ensures that small attachments extending above the main shaft—which are otherwise prone to being missed—are successfully extracted. This buffered region defines a spatially bounded range, keeping the subsequent reasoning computationally efficient and stable.

2.4 Candidate Selection and Denoising

Within each buffer $\mathcal{B}_k$ (Eq. 4), we distinguish:
Seed points ($\mathcal{S}_k$): those predicted as pole-like (C_{KP}, Eq.2) by KPConv.
Candidate points ($\mathcal{C}_k$): originating from the adjacent classes (C_{ADJ}, Eq.2).

A candidate point $\mathrm{p}_j \in \mathrm{C}_{ADJ}$ is added to $\mathcal{C}_k$ if it lies sufficiently close to any seed point $\mathrm{p}_i \in \mathcal{S}_k$ and satisfies:

$$|p_{j,xy} - p_{i,xy}| \le r_s\ , \tag{6}$$

where r_s is the search radius. A KD-tree (Bentley, 1975) is used to accelerate these neighborhood queries.

To further eliminate outliers and spatial noise, any point having fewer than min_nb neighbors within a radius defined by edge_max is removed. Subsequently, voxel downsampling (voxel size: voxel) is applied to regularize spatial density and reduce redundancy. This step both accelerates subsequent processing and regularizes the spatial distribution of points for stable graph construction.

The output of Section 2.4 is a cleaned point set ($\mathcal{V}_k$). It consists of the seed points $\mathcal{S}_k$, which KPConv initially classified as pole-like, and the denoised candidate points $\tilde{\mathcal{C}}_k$. This refined set $\mathcal{V}_k$ forms the vertex set for the k-nearest neighbor (kNN) graph (Tenenbaum et al., 2000; Von Luxburg, 2007) constructed in Section 2.5.

2.5 Graph Construction and Connectivity Analysis

Using the refined point set $\mathcal{V}_k$, a KNN graph $\mathcal{G} = (V, E)$ is constructed with k nearest neighbors and maximum edge length $d_{\max}$.

Each vertex $v_i \in V$ stores its local tangent direction t_i, estimated by the first principal component of PCA (Pearson, 1901; Hotelling, 1933) applied to its local neighborhood:

$$t_i = \frac{v_1^i}{|v_1^i|}, \tag{7}$$

where $\mathbf{v}_1^{(i)}$ denotes the dominant eigenvector.

To restore fragmented pole structures caused by occlusion or scanning gaps, Dijkstra's shortest-path algorithm (Dijkstra, 1959) is applied to propagate connectivity from each seed point to all candidate points within $\mathcal{G}$. The edge cost is defined as:

$$\Delta c_{ij} = d_{ij} + \lambda_\theta (1 - |\, \mathbf{t}_i^{\mathrm{T}} \mathbf{t}_j \,|), \tag{8}$$

where
d_{ij} is Euclidean distance between p_i and p_j,
$\mathrm{t}_i, \mathrm{t}_j$ are unit tangent vectors estimated from PCA,
λ_θ penalizes angular discontinuities between connected segments.

The cumulative path cost c_j determines whether a candidate point is geometrically reachable from any seed:

$$R_k = \{p_j \in\ \tilde{C}_k \mid c_j \le p_{\max}\} \tag{9}$$

where $p_{\max}$ is regarded as geometrically reachable and included in the reconstructed pole region.

Thus, Section 2.5 produces the subset $\mathcal{R}_k$ consisting only of candidates that are structurally connected to existing pole-like segments. This reachable set $\mathcal{R}_k$ becomes the sole input to the final classification stage in Section 2.6.

2.6 Final Pole Classification

Section 2.6 evaluates the reachable candidates $\mathcal{R}_k$ and determines whether they should be reclassified as pole-like. For each candidate point p_i, the horizontal and vertical offsets from the cluster center c_k are computed:

$$d_{xy} = \left|\left|p_{i,xy} - c_{k,xy}\right|\right|, d_z = \left|z_i - c_{k,z}\right|, \tag{10}$$

These distances are converted into confidence scores and merged using a sigmoid function, yielding a final plausibility value:

$$\mathrm{s_{xy}} = \exp\left(-\frac{\mathrm{d_{xy}}}{\alpha}\right), \mathrm{s_h} = \exp\ \left(-\frac{\mathrm{d_z}}{\beta}\right), \tag{11}$$

where α, β control the attenuation of horizontal and vertical confidence, respectively, which are combined via a logistic sigmoid:

$$\sigma(u) = \frac{1}{1+e^{-u}}, \tag{12}$$

$$s_{\mathrm{mid}} = \sigma\left(0.5 s_{xy} + 0.5 s_h\right), \tag{13}$$

Points satisfying

$$s_{\text{mid}} \geq \tau, \tag{14}$$

are finally reclassified as pole-like.

This step serves as a heuristic plausibility filter for reachable candidate points and is empirically effective in reducing some false positives caused by nearby clutter, such as overhead cables or signboards. A summary of the key parameters is provided in Table 1.

Section	Processing	Equation No.	Parameter	Value (Unit)	Description
2.3	Clustering and Buffer Construction	3	ε	0.35 m	DBSCAN neighborhood radius for clustering
		3	minPts	8	Minimum points per cluster for DBSCAN
		4	r_{xy}	4.0 m (Hokkaido) 2.0 m (Paris-Lille-3D)	Horizontal buffer radius around cluster centroid
		5	z_{buf}	5.0 m	Vertical extension margin above cluster
2.4	Candidate Selection and Denoising	6	r_s	2.0 m	XY search radius for candidate selection
		—	voxel	0.05 m	Voxel grid cell size for downsampling
		—	min_nb	3	Min neighbors for density filter
		—	edge_max	0.2 m	Neighbor search radius for density filter
2.5	Graph Construction and Connectivity Analysis	—	k	12	Number of neighbors for graph edges
		—	$d_{\max}$	0.15 m	Maximum edge length in kNN graph
		8	λ_θ	0.60	Weight of angular deviation in edge cost
		9	$p_{\max}$	4.0 m (Hokkaido) 2.0 m (Paris-Lille-3D)	Maximum cumulative path cost for reachability
2.6	Final Pole Classification	11	α	2.0 m	Horizontal attenuation
		11	β	3.0 m	Vertical attenuation
		14	τ	0.2 (Hokkaido) 0.1 (Paris-Lille-3D)	Threshold for pole classification

Table 1. Key parameters in the proposed method

3. Experiments and Results

3.1 Description of the Point Clouds

This study evaluates the robustness and cross-domain generalization of the proposed framework using large-scale MMS point cloud datasets collected in different regions and under varying sensor conditions. All Japanese datasets were acquired along roads in the Kinki and Hokkaido regions using the same MMS. To further assess cross-country and cross-sensor generalization, the benchmark Paris-Lille-3D dataset (France) was also employed. Together, these datasets provide a comprehensive basis for evaluating the scalability and adaptability of the proposed method across diverse domains.

3.1.1 Kinki region, Japan — Training dataset:

The training dataset was collected along approximately 8 km of roads in the Kinki region, Japan, covering a mixture of residential streets, suburban corridors, and major arterial routes. The MMS configuration illustrated in Figure 1 consisted of two Riegl VQ-450 laser scanners, four Grasshopper 505GC cameras, and three Ikegami ISD-200HD video cameras, integrated with an Applanix POS/LV 520 GNSS/IMU for accurate positioning. Only the LiDAR point clouds were used in this study.

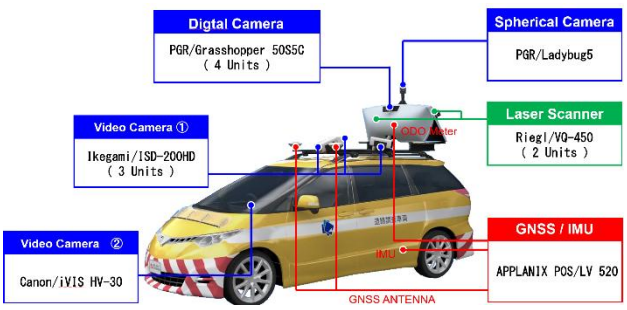

Figure 1. Configuration of the MMS.

The raw dataset originally contained approximately 1,286.2 million points. To remove distant structures beyond the roadway corridor, a 15 m buffer from the road centerline was applied. After trimming, 1,097.1 million points remained. These points were manually annotated into five semantic classes:
1. Unclassified (class0)
2. Ground (class1)
3. Guardrail (class2)
4. Power-line (class3)
5. Pole-like (class4)

The annotated dataset served as the training dataset for KPConv. The class distribution of the annotated training dataset is highly imbalanced. As summarized in Table 2, ground points account for 75.23% of all annotated points, whereas the pole-like class represents only 1.44%. This distribution reflects the inherent class imbalance in large-scale MMS point clouds, where pole-like road objects occupy only a small fraction of the scene. In particular, the pole-like class is highly underrepresented compared with the dominant ground class, which makes the extraction task especially challenging.

Class	Number of points	Proportion (%)
Unclassified	246,638,946	22.48
Ground	825,380,963	75.23
Guardrail	5,137,941	0.47
Power-line	4,145,373	0.38
Pole-like	15,838,120	1.44
Total	1,097,141,343	100.00

Table 2. Class distribution of the annotated training dataset (Kinki region).

3.1.2 Hokkaido region, Japan — Test dataset (Cross-regional and large-scale):

To evaluate intra-sensor and inter-regional robustness, the proposed framework was tested using data collected along approximately 31 km of roads in the Hokkaido region, Japan. Among these, about 5 km of tunnel sections were excluded due to limited LiDAR visibility, resulting in an effective evaluation length of approximately 26 km. The surveyed route extended from the boundary of Hamamasu Ward to the Yawata 1-chome intersection in Ishikari City, as shown in Figure 2.

The raw point cloud contained 2,870.4 million points. Similar to the Kinki dataset, a buffer—20 m in this case—was applied to remove distant off-road features. The cropped dataset comprised 2,529.9 million points.

Compared with the Kinki dataset used for training, the Hokkaido dataset exhibits markedly different environmental and structural characteristics. In particular, Hokkaido is primarily characterized by snow-protection facilities and snow delineator posts (snow poles), which are characteristic of snowy regions in Japan (Figure 3). These snow poles mark road edges during heavy snowfall or whiteout conditions, enabling both vehicle operators and snow-removal teams to judge the roadway alignment when pavement markings are obscured. The presence of such snow poles, which do not appear in the Kinki dataset. This regional difference serves as a valuable test case for assessing the cross-domain generalization of the proposed framework under contrasting environmental and structural conditions.

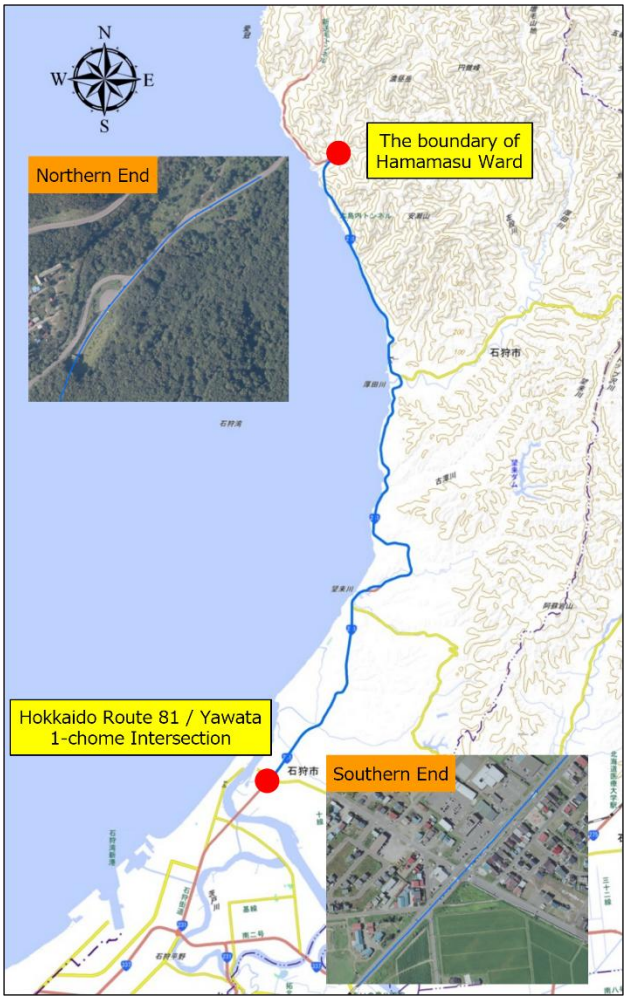

Figure 2. Spatial coverage of the Hokkaido MMS dataset. Background map: © Geospatial Information Authority of Japan (GSI), “GSI Maps”.

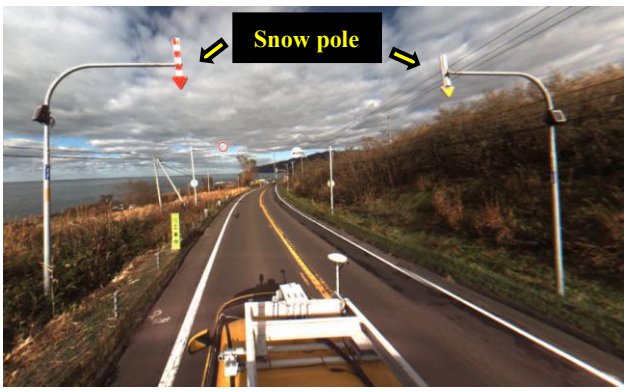

Figure 3. Example of snow pole in Hokkaido (captured by an MMS camera).

3.1.3 Paris-Lille-3D, France — Test dataset (Cross-country and cross-sensor):

To further assess cross-sensor and cross-country generalization, the Paris-Lille-3D dataset (Roynard et al., 2018) was employed. This dataset was acquired in France using a Velodyne HDL-32E laser scanner mounted on a different MMS platform, covering

approximately 1.9 km of urban and suburban roads. Only the training dataset (Paris.ply, Lille1_1.ply, Lille1_2.ply, Lille2.ply) was used for evaluation because the official test dataset (Dijon_9.ply, Ajaccio_2.ply, Ajaccio_57.ply) does not provide publicly available ground-truth annotations.

This dataset provides a strict cross-domain benchmark, combining differences in both sensor type (Riegl vs. Velodyne) and geographical domain (Japan vs. France). A summary of all datasets used for training and evaluation is provided in Table 3.

	Kinki	Hokkaido	Paris-Lille-3D
Country	Japan	Japan	France
LiDAR Sensor	Riegl VQ-450	Riegl VQ-450	Velodyne HDL-32E
Density (points/m²)	5,000 - 7,000	4,000 - 6,000	1,000 - 2,000
Length (km)	8.0	26.0	1.9
Points (Million)	1,097.1	2,529.9	143.1
Purpose	Training	Cross-regional test	Cross-country and Cross-sensor test

Table 3. Datasets used for training and evaluation.

3.2 Experimental Settings

All experiments were performed on a workstation equipped with two Intel Xeon Silver 4214R CPUs, two NVIDIA Quadro RTX 8000 GPUs (48 GB VRAM each), and 768 GB RAM (12 × 64 GB). The software environment included Ubuntu 22.04 LTS, Python 3.6, TensorFlow-GPU 1.12.0, CUDA 9.0, and cuDNN 7.4.

KPConv was implemented using the official source code (https://github.com/HuguesTHOMAS/KPConv). Only 3D spatial coordinates (x, y, z) were used as input features for both the training and testing datasets. The batch size was set to 4, the sampling grid voxel size to 0.1 m, and the kernel radius to 4.0 m. Other hyperparameters followed the original implementation. The network was trained for 500 epochs using the Kinki region dataset, requiring approximately 7 hours and 39 minutes.

The annotated Kinki region dataset was used exclusively for training, and no formal validation split was introduced. The model obtained after 500 training epochs was then used for inference on the Hokkaido and Paris-Lille-3D datasets. This setting was chosen because the main purpose of this study was to evaluate cross-region and cross-sensor generalization under a fixed training condition, rather than to optimize hyperparameters on the source domain. As a supplementary check, qualitative assessment on separate Kinki region data not used for training indicated that the predictions were sufficiently stable for the subsequent cross-domain experiments.

3.3 Evaluation Metrics

Performance was evaluated using precision (P), recall (R), and F_1-score (F_1), computed on a per-point basis following standard 3D semantic segmentation practice (Qi et al., 2017a; Thomas et al., 2019; Hu et al., 2020). The metrics are defined as:

$$P = \frac{TP}{TP+FP}, \quad R = \frac{TP}{TP+FN}, \quad F1 = \frac{2PR}{P+R}, \tag{15}$$

where TP, FP, and FN denote the number of true positive, false positive, and false negative points, respectively.

3.4 Quantitative Results

3.4.1 Hokkaido region — Cross-Regional and Large-Scale Test:

Table 4 summarizes the quantitative evaluation on the Hokkaido dataset. The proposed method yielded a well-balanced improvement over the KPConv baseline. The F_1-score increased from 0.8263 to 0.8689 (+0.0426), primarily due to the recovery of previously omitted pole components such as lamp tops and side-mounted signal arms. This improvement is further reinforced by the reliable reconstruction of regional snow poles, which are prevalent in Hokkaido but absent from the training dataset.

Although precision decreased only marginally from 0.8599 to 0.8466 (-0.0133), recall increased from 0.7952 to 0.8924 (+0.0972), demonstrating that the geometric-topological reasoning effectively complements semantic segmentation. These results indicate that the proposed module successfully compensates for cases where semantic segmentation alone produces fragmented pole structures.

Method	Precision	Recall	F_1-score
KPConv (Baseline)	0.8599	0.7952	0.8263
Ours	0.8466	0.8924	0.8689
Ours - Baseline	-0.0133	+0.0972	+0.0426

Table 4. Quantitative results for the Hokkaido dataset

3.4.2 Paris-Lille-3D — Cross-Country and Cross-Sensor Test:

As shown in Table 5, the proposed framework significantly outperformed the KPConv baseline under a severe domain shift between countries and sensor types. Recall increased from 0.5109 to 0.6656 (+0.1547), while precision decreased slightly from 0.7982 to 0.7454 (-0.0528), resulting in an overall F_1-score improvement from 0.6230 to 0.7032 (+0.0802). These results confirm that the proposed method generalizes robustly across both geographic and sensor modalities. By jointly leveraging semantic inference and geometric-topological reasoning, it adapts effectively to variations in scanning geometry, point density, and urban morphology—demonstrating broad applicability to heterogeneous MMS datasets.

Method	Precision	Recall	F_1-score
KPConv (Baseline)	0.7982	0.5109	0.6230
Ours	0.7454	0.6656	0.7032
Ours - Baseline	-0.0528	+0.1547	+0.0802

Table 5. Quantitative results for the Paris-Lille-3D dataset

3.5 Qualitative Analysis

Figures 4-7 present qualitative comparisons for both the Hokkaido dataset (Japan) and the Paris-Lille-3D dataset (France).

As shown in Figure 4, KPConv often fails to detect the curved L-shaped sections of these poles, capturing only partial vertical shafts. In contrast, the proposed method successfully reconstructs the full L-shaped geometry, including the horizontal arm, and also reduces missed detections for snow-pole arms and lamp tops. These results confirm that geometric-topological reasoning enables robust reconstruction of region-specific structures unseen during training.

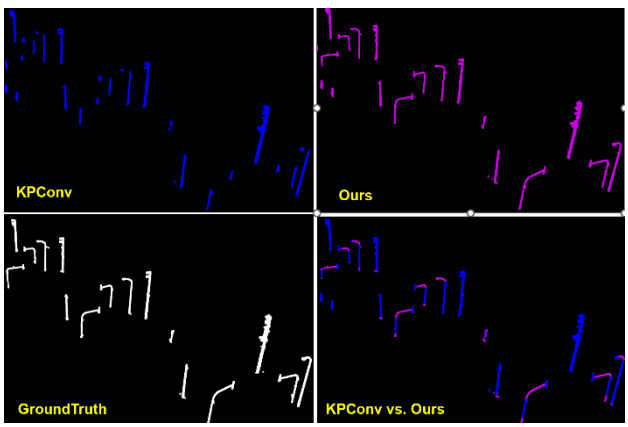

Figure 4. Qualitative comparison on the Hokkaido dataset.

Figure 5 illustrates large-scale operational applicability. The method maintains stable detection quality over 26 km of MMS data, producing consistent extractions that are directly usable for roadside asset inventory and digital-twin generation.

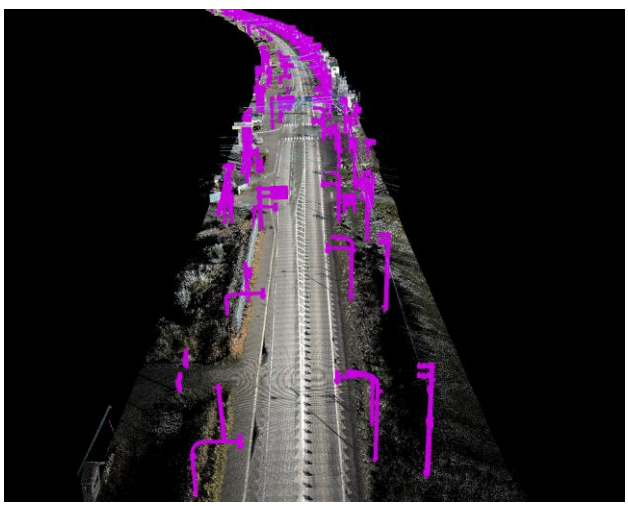

Figure 5. Qualitative result on the Hokkaido dataset.

Figure 6 illustrates qualitative comparisons on the Paris-Lille-3D dataset, which was acquired using a different LiDAR sensor and under distinct urban conditions in France. Unlike the baseline KPConv—which tends to omit upper pole segments and fine attachments—the proposed geometric-topological reasoning successfully reconstructs complete, vertically continuous pole structures, as shown in Figure 6. These comparisons confirm that geometric-topological reasoning enhances structural completeness even under cross-sensor conditions.

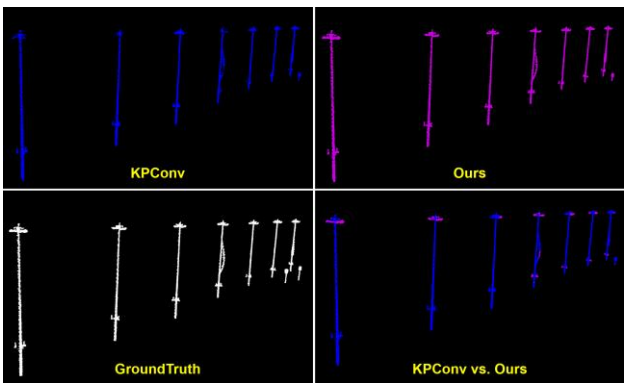

Figure 6. Qualitative comparison on the Paris-Lille-3D dataset.

Figure 7 illustrates both the strengths and limitations of the proposed method. It successfully recovers small pole components missed by KPConv (e.g., lamp heads). However, if KPConv entirely misses a pole, no seeds exist and the pole cannot be reconstructed (green bounding box). Conversely, if KPConv incorrectly infers vegetation stems as poles, the geometric propagation may extend these false positives both upward and downward (orange bounding box).

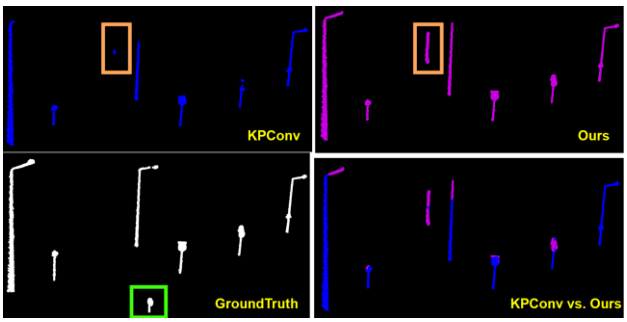

Figure 7. Examples of limitations in the proposed method.

The decrease in precision was primarily associated with local over-segmentation around detected pole-like structures. In most cases, the false positives consisted of partial structures, such as nearby vegetation stems or small adjacent vertical components, that were included during the geometric-topological expansion. The qualitative results indicate that these errors were typically local and partial rather than complete non-pole objects being mislabeled as poles. This suggests that the practical impact of the precision decrease is limited compared with the gain in recall.

4. Discussion

The results demonstrate that the proposed method effectively improves the extraction of pole-like road objects by decoupling geometric-topological completion from network training,

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-145-2026 | © Author(s) 2026. CC BY 4.0 License.

providing a computationally efficient solution for domain adaptation without additional annotation or fine-tuning.

The improvement in recall indicates that fewer true pole-like objects and attached components are missed. In the present study, the main gains were observed in the recovery of structures that are easily omitted by KPConv alone, including lamp tops, signal arms, snow poles, and other thin upper attachments. This improvement is particularly important in practical applications, especially because the decrease in precision was only marginal. Moreover, over-segmentation is generally easier to correct, since false positives can often be removed through subsequent visual inspection or minor editing. Therefore, reducing omissions is especially valuable in practical workflows, and the improved recall directly enhances the completeness and practical usefulness of the extracted results.

Unlike conventional geometric post-filters, the framework fuses learned semantic features with domain-invariant geometric and topological cues, compensating for the limitations of purely data-driven segmentation and achieving more reliable and interpretable results. This post-hoc fusion mechanism improves recall and structural completeness without modifying the underlying KPConv network or requiring extra annotated data.

Transformer-based point cloud networks, including the Point Transformer series, are known to provide stronger contextual modeling in an end-to-end manner and may yield better initial segmentation, especially for thin or elongated structures. In contrast, the main contribution of this study lies not in the choice of a specific backbone network, but in the explicit fusion of deep-learning-based semantic prediction with geometric-topological reasoning. Rather than attempting to learn structural consistency only implicitly, the proposed framework introduces an interpretable geometric-topological reasoning that can complement the outputs of a semantic segmentation model without additional retraining or annotation. In this sense, the added value of the proposed method lies in its ability to explicitly enforce geometric and topological consistency, thereby improving structural completeness in a model-agnostic manner.

4.1 Robustness and Domain Generalization

The robustness of the proposed framework was validated across two datasets that differ in both geographic context and sensor characteristics. Deep learning models generally suffer from performance degradation when applied to domains that differ from the training dataset, especially when encountering sensor-dependent variations or features not present in the training dataset. The geometric-topological reasoning incorporated in our framework mitigates these limitations by compensating for such domain shifts and ensuring consistent structural extraction.

In the post-hoc geometry-topology processing applied after KPConv inference, the method relies on only a few interpretable parameters (Table 1), allowing it to be readily transferred to other MMS datasets and acquisition configurations. Compared with retraining-based domain adaptation techniques, the proposed approach achieves comparable improvements in cross-domain segmentation accuracy with zero additional learning cost, underscoring its efficiency and practicality for large-scale, multi-regional LiDAR applications.

4.2 Computational Efficiency and Scalability

The proposed method demonstrated scalability and computational efficiency in processing large-scale MMS point clouds. As summarized in Table 6, the Hokkaido dataset (≈ 26 km, 2.53 billion points) was processed in less than 13 hours using a single NVIDIA Quadro RTX 8000 GPU, whereas the Paris-Lille-3D dataset (≈ 1.9 km, 143.1 million points) required approximately 1 hour. These results confirm that the method can efficiently handle extensive city-scale datasets, a key requirement for nationwide infrastructure mapping and digital-twin generation.

Section	Description	Hokkaido	Paris-Lille-3D
2.2	KPConv inference	6h45m	0h44m
2.3-2.6	Geometric-Topological	6h06m	0h17m
Total		12h51m	1h01m

Table 6. Total processing time of the proposed pipeline

4.3 Limitations and Future Work

The proposed method serves as a geometric-topological post-processing module that complements the semantic inference of KPConv. While it effectively extracts fragmented or partially missing pole-like structures, it inherently depends on the initial seed regions detected by the baseline network. Consequently, if a pole-like object is entirely missed by KPConv, the proposed method cannot recover it because no valid seeds exist to initiate the geometric-topological growth. In addition, when non-pole structures—such as tree trunks or vegetation stems—are mistakenly classified as poles by KPConv, the subsequent geometric-topological expansion may unintentionally reinforce these false positives. These limitations indicate that the framework performs best when the baseline segmentation provides reliable pole seeds.

Future work will explore the integration of more discriminative appearance features and multi-modal cues (e.g., intensity, color, or image-based priors) to further suppress erroneous growth and enhance robustness across diverse environments. In addition, future work will investigate the use of more advanced point-based semantic segmentation backbones in place of KPConv, including transformer-based architectures such as the Point Transformer series (Zhao et al., 2021; Wu et al., 2022; Wu et al., 2024) as well as other recent state-of-the-art approaches. Since these models are expected to provide stronger semantic predictions and richer contextual representations, especially for thin or elongated structures, combining them with the proposed explicit geometric-topological reasoning framework may further enhance both detection accuracy and structural completeness. This extensibility indicates that the proposed framework is model-agnostic and can serve as a complementary refinement module for a wide range of point cloud semantic segmentation methods.

5. Conclusion

This study proposed an approach that integrates deep learning-based semantic segmentation with explicit geometric and topological reasoning to improve the extraction of pole-like road objects from MMS point clouds. Unlike conventional approaches that rely solely on deep neural networks or require domain-specific retraining, the proposed method performs post-hoc completion without modifying the base KPConv model or requiring additional annotated data.

Comprehensive experiments on large-scale cross-regional datasets demonstrated both quantitative accuracy and

computational scalability. On the Hokkaido dataset (≈ 26 km, 2.53 billion points), the proposed framework achieved an F_1-score of 0.8689, improving by +0.0426 over the KPConv baseline (F_1-score=0.8263). In particular, the method successfully extracted lamp tops, side-mounted signal arms, and snow poles, demonstrating its capability to generalize to previously unseen structural types through geometric-topological reasoning. The entire dataset was processed in approximately 13 hours on a single NVIDIA Quadro RTX 8000 GPU, confirming the framework’s scalability for nationwide and city-scale 3D mapping applications.

Further evaluation on the Paris-Lille-3D dataset, captured with a different LiDAR sensor and acquisition geometry, demonstrated strong cross-country and cross-sensor generalization. Recall increased from 0.5109 to 0.6656 (+0.1547), and the F_1-score improved by +0.0802 relative to the KPConv baseline, demonstrating robustness across heterogeneous sensor modalities and diverse urban environments. Given its efficiency and generality, the proposed framework is well suited for nationwide road-asset management, infrastructure inspection, and digital-twin construction.

Acknowledgements

Among the datasets used in this study, the MMS data collected in Hokkaido were provided by the Hokkaido Regional Development Bureau, Ministry of Land, Infrastructure, Transport and Tourism. We would like to express our sincere gratitude for their support. We also thank the anonymous reviewers for their helpful suggestions and comments.

References

Belton, D., Lichti, D., 2006. Classification and segmentation of terrestrial laser scanner point clouds using local variance information. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 36(5), 44-49.

Bentley, J.L., 1975. Multidimensional binary search trees used for associative searching. Communications of the ACM, 18(9), 509-517.

Brenner, C., 2009. Global localization of vehicles using local pole patterns. Joint Pattern Recognition Symposium. Berlin, Heidelberg: Springer Berlin Heidelberg, 61-70.

Cabo, C., Ordoñez, C., García-Cortés, S., Martínez, J., 2014. An algorithm for automatic detection of pole-like street furniture objects from mobile laser scanner point clouds. ISPRS Journal of Photogrammetry and Remote Sensing, 87, 47-56.

Dijkstra, E.W., 1959. A note on two problems in connexion with graphs. Numerische Mathematik, 1, 269-271.

El-Halawany, S.I., Lichti, D.D., 2013. Detecting road poles from mobile terrestrial laser scanning data. GIScience & Remote Sensing, 50(6), 704-722.

Elhashash, M., Albanwan, H., & Qin, R., 2022. A review of mobile mapping systems: From sensors to applications. Sensors, 22(11), 4262.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD), 226-231.

Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 24, 417-441.

Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Markham, A., 2020. RandLA-Net: Efficient semantic segmentation of large-scale point clouds. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 11108-11117.

Jaritz, M., Vu, T. H., Charette, R. D., Wirbel, E., & Pérez, P., 2020. xmuda: Cross-modal unsupervised domain adaptation for 3d semantic segmentation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 12605-12614.

Li, J., Cheng, X., 2022. Supervoxel-based extraction and classification of pole-like objects from MLS point cloud data. Optics & Laser Technology, 146, 107562.

Mori, Y., Kohira, K., & Masuda, H., 2018. Classification of pole-like objects using point clouds and images captured by mobile mapping systems. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 42, 731-738.

Nurunnabi, A., Sadahiro, Y., Teferle, F. N., Laefer, D. F., & Li, J., 2023. Detection and Segmentation of Pole-like Objects in Mobile Laser Scanning Point Clouds. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 48, 27-34.

Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. Philosophical Magazine, 2(11), 559-572.

Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017a. PointNet: Deep learning on point sets for 3D classification and segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 652-660.

Qi, C.R., Yi, L., Su, H., Guibas, L.J., 2017b. PointNet++: Deep hierarchical feature learning on point sets in a metric space. Advances in Neural Information Processing Systems, 30.

Roynard, X., Deschaud, J.-E., Goulette, F., 2018. Paris-Lille-3D: A large and high-quality ground-truth urban point cloud dataset for automatic segmentation and classification. International Journal of Robotics Research, 37(6), 545-557.

Shi, Z., Kang, Z., Lin, Y., Liu, Y., Chen, W., 2018. Automatic recognition of pole-like objects from mobile laser scanning point clouds. Remote Sensing, 10(12), 1891.

Tan, Z., Zhang, X., Teng, S., Wang, L., & Gao, F., 2024. A review of deep learning-based lidar and camera extrinsic calibration. Sensors, 24(12), 3878.

Tardy, H., Soilán, M., Martín-Jiménez, J.A., González-Aguilera, D., 2023. Automatic road inventory using a low-cost mobile mapping system and a semantic segmentation deep learning model. Remote Sensing, 15(5), 1351.

Tenenbaum, J.B., de Silva, V., Langford, J.C., 2000. A global geometric framework for nonlinear dimensionality reduction. Science 290(5500), 2319-2323.

Thomas, H., Qi, C.R., Deschaud, J.E., Marcotegui, B., Goulette, F., Guibas, L.J., 2019. KPConv: Flexible and deformable convolution for point clouds. Proceedings of the IEEE/CVF International Conference on Computer Vision, 6411-6420.

Truong-Hong, L., Laefer, D.F., et al., 2022. Urban object extraction from MLS point clouds using geometric-contextual constraints. ISPRS Journal of Photogrammetry and Remote Sensing, 184, 38-53.

Von Luxburg, U., 2007. A tutorial on spectral clustering. Statistics and Computing, 17(4), 395-416.

Wang, Y., Wang, J., Li, F., Liu, Z., 2023. Framework for geometric information extraction and digital modeling of roads from LiDAR point clouds. Remote Sensing, 15(3), 576.

Wang, Z., Yang, L., Sheng, Y., & Shen, M., 2021. Pole-Like Objects Segmentation and Multiscale Classification-Based Fusion from Mobile Point Clouds in Road Scenes. Remote Sens. 2021, 13, 4382.

Wen, C., Sun, X., Li, J., Cheng, W., Guo, Y., Habib, A., 2019. A deep learning framework for road marking extraction, classification and completion from mobile laser scanning point clouds. ISPRS Journal of Photogrammetry and Remote Sensing, 147, 178-192.

Wu, X., Jiang, L., Wang, P. S., Liu, Z., Liu, X., Qiao, Y., ... & Zhao, H., 2024. Point transformer v3: Simpler faster stronger. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 4840-4851.

Wu, X., Lao, Y., Jiang, L., Liu, X., Zhao, H., 2022. Point transformer v2: Grouped vector attention and partition-based pooling. Advances in Neural Information Processing Systems, 35, 33330-33342.

Yadav, M., Khan, P., & Singh, A. K., 2022. Identification of pole-like objects from mobile laser scanning data of urban roadway scene. Remote Sensing Applications: Society and Environment, 26, 100765.

Zhang, W., Witharana, C., Li, W., Zhang, C., Li, X., & Parent, J., 2018. Using deep learning to identify utility poles with crossarms and estimate their locations from google street view images. Sensors, 18(8), 2484.

Zhang, Z., Khoshelham, K., & Shojaei, D., 2024. Pole-NN: Few-shot classification of pole-like objects in lidar point clouds. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 10, 333-340.

Zhao, H., Jiang, L., Jia, J., Torr, P. H., Koltun, V., 2021. Point transformer. In Proceedings of the IEEE/CVF international conference on computer vision, 16259-16268.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-145-2026