Pole-NN: Few-Shot Classification of Pole-Like Objects in Lidar Point Clouds

Zezheng Zhang[1], Kourosh Khoshelham[2], Davood Shojaei[1]

[1] Centre for Spatial Data Infrastructures and Land Administration, Department of Infrastructure Engineering, The University of Melbourne, Melbourne, Victoria, Australia

[2] Department of Infrastructure Engineering, The University of Melbourne, Melbourne, Victoria, Australia

E-mails: zezzhang@student.unimelb.edu.au, (shojaeid, k.khoshelham)@unimelb.edu.au

Keywords: Mobile Laser Scanning, Street Furniture, Road Assets, Few-shot Learning, Feature Learning, Classification.

Abstract

In the realm of autonomous systems and smart-city initiatives, accurately detecting and localizing pole-like objects (PLOs) such as electrical poles and traffic signs has become crucial. Despite their significance, the diverse nature of PLOs complicates their accurate recognition. Point cloud data and 3D deep learning models offer a promising approach to PLO localization under varied lighting, addressing issues faced by camera systems. However, the distinct characteristics of different street scenes worldwide require infeasibly extensive training data for satisfactory results because of the nature of deep learning. This prohibitively increases the cost of lidar data capture and annotation. This paper introduces a novel few-shot learning framework for the classification of outdoor point cloud objects, leveraging a minimalistic approach that requires only a single support sample for effective classification. Central to our methodology is the development of Pole-NN, a **N**on-parametric **N**etwork that efficiently distinguishes between various PLOs and other road assets without the need for extensive training datasets traditionally associated with deep learning models. Additionally, we present the Parkville-3D Dataset, an annotated point cloud dataset we have captured and labelled, which addresses the notable scarcity of fine-grained PLO datasets. Our experimental results demonstrate the potential of our approach to utilize the intrinsic spatial relationships within point cloud data, promoting a more efficient and resource-conscious strategy for PLO classification.

1. Introduction

In the rapidly evolving landscape of autonomous systems and smart-city technologies, the accurate recognition and localization of pole-like objects (PLOs) have emerged as a key challenge and research focus. These objects, encompassing utility poles, street lamps, and traffic signs, serve as indispensable components within complex urban environments. They have numerous applications, from vehicle localization in autonomous driving (Dong et al., 2023) to infrastructure maintenance (Cabo et al., 2014) and 5G network planning (Gholampooryazdi et al., 2017). Despite their ubiquity and functional significance, PLOs pose unique challenges in terms of accurate detection and localization because of their diverse forms, varying sizes, and complex surroundings. Using point cloud data is one of the ways to detect and localize PLOs (Luo et al., 2023). Scanned by Light Detection and Ranging (lidar) sensors, point cloud data can provide accurate depth and spatial information that lends itself well to tasks requiring precision localization and identification of structures in the environment. Among all lidar types, Mobile Laser Scanning (MLS) data prove particularly crucial to our research topic, given its ability to capture point cloud data while moving on the road. Its active sensing nature ensures reliable performance under various lighting conditions, overcoming a key limitation of camera-based systems, which can suffer from over or under-exposed imagery in high-contrast scenes and various weather conditions (Yeong et al., 2021). However, previous studies show that the recognition accuracy of PLOs is always relatively low among other street assets on point cloud data (Thomas et al., 2019; Luo et al., 2020; Nie et al., 2022; Boulch et al., 2020).

Most point-cloud PLOs recognition studies still use non-learning-based traditional methods. Those methods highly depend on prior knowledge and hand-crafted features leading to the problem that they cannot widely apply to all types of PLOs (Luo et al., 2023) and can not describe the PLO well enough (Li et al., 2019b). Those problems can be mitigated by applying deep learning techniques, which can extract learned features, enabling a comprehensive representation of the properties intrinsic to PLOs (Plachetka et al., 2021).

Street scenes differ significantly across global locales. As such, current deep-learning methods often require an influx of fresh training samples to produce satisfactory results in unfamiliar new environments. Even though some available datasets capture street scenes in point clouds, they are not ubiquitous because lidar technology is relatively recent, and the cost of sensors remains prohibitively high. A survey review reveals that the primary constraint hindering the widespread application of deep learning on point cloud data is the cost of capturing data (Luo et al., 2023). That presents a unique challenge in applying and developing deep learning techniques in different environments for lidar-based PLO recognition, underlining the need for creative strategies that maximize the utility of data.

Inspired by cutting-edge advancements in few-shot learning research, this paper introduces a novel, training-free method for road object classification. Our method, named Pole-NN, incorporates multiple non-trainable feature extraction components. Those components enable Pole-NN to effectively distinguish pole-like objects from other road entities and further categorize them into specific types of poles. To empirically validate Pole-NN's efficacy, we present experiments focusing on its one-shot learning performance. Remarkably, Pole-NN achieves classification accuracy comparable to that of previous deep learning methods, which traditionally rely on extensive training datasets. This achievement is particularly noteworthy considering that

Pole-NN was provided with only a single example as a support case, starkly contrasting with the data-intensive requirements of preceding models. The principal innovation of Pole-NN lies in its training-free architecture, which empowers it to accurately classify outdoor 3D objects with minimal support samples, demonstrating high classification accuracy even with the use of a single example. This approach marks a significant step forward in the efficient and rapid deployment of object classification models in dynamic outdoor environments.

In addressing the critical gap in the availability of dense, outdoor, street-scene point cloud datasets for PLO recognition, this work also introduces the Parkville-3D dataset[1]. Parkville-3D stands as a dense and accurate dataset, offering a fine-grained semantic segmentation of urban elements. This contribution not only enriches the domain with a much-needed resource but also sets the stage for enhanced performance testing of deep learning models in outdoor environment applications, particularly in the precise recognition of various PLOs.

The structure of the paper is as follows. In Section 2, we will provide a background of previous PLO classification methods. Section 3 introduces our newly devised framework for PLO classification, along with a detailed exposition of a typical implementation. Section 4 details the experimental setup, including evaluations conducted on both publicly available benchmark datasets and custom datasets, created to address the lack of detailed pole-type labels in existing datasets. Finally, Section 5 discusses the outcome and future directions.

2. Related Work

In this section, we will introduce two categories of PLO classification: the Hand-crafted Feature-based Approach and the Learning-based Approach. Subsequently, we will explain the methodologies previously employed in few-shot Learning.

2.1 PLO Classification based on Hand-crafted Features

Yan et al. (2017) employed a shape-matching technique for categorizing PLOs into eight distinct groups. They utilized the Ensemble of Shape Functions (ESF) to extract various shape features such as point distance, area, and angles, represented in a 64-bin histogram. The classification is achieved by comparing the shapes of their histograms. An alternative approach, as described by Hao et al. (2018), leverages the distinctive horizontal point densities characteristic of various poles for classification purposes.

However, as noted in multiple studies (Li et al., 2019a; Kang et al., 2018; Shi et al., 2018; Huang and You, 2015; Li et al., 2018; Wang et al., 2022), devising a comprehensive method capable of classifying a broad spectrum of pole types or their subcategories remains challenging.

In response, there has been a shift towards machine learning methods for classification, employing algorithms such as Random Forest (Li et al., 2019a; Wu et al., 2017; Yan et al., 2017; Yuma et al., 2018; He et al., 2017) and Support Vector Machine (SVM) (Li et al., 2019a; Wu et al., 2017; He et al., 2017). To construct a machine learning-based PLO classification model, Li et al. (2019a) extracted three types of features: size, eigenvalue-based, and radiometric. Those features were then input into various classifiers including SVM, Random Forest (RF), Gaussian Mixture Model (GMM), and Naïve Bayes (NB), with Random Forest demonstrating superior performance. Ferrin et al. (Ferrin et al., 2018) similarly trained a neural network using eigenvalue-based features for pole classification.

Despite these advances, the limits of hand-crafted features have been noted, especially in accommodating diverse scenes and pole characteristics in different geographical locations (Li et al., 2019a; Dong et al., 2023; Ferrin et al., 2018; Wu et al., 2017; Wang et al., 2022), posing a challenge in developing universally applicable hand-crafted feature-based methods.

2.2 PLO Classification based on Learned Features

Seeking a more comprehensive feature extraction approach, previous studies have turned to deep learning networks for processing point cloud data. One of the earliest such endeavours, *PointNet* (Qi et al., 2017a), directly consumed point cloud data. The first end-to-end deep learning pole recognition network was introduced in 2021 (Plachetka et al., 2021), incorporating elements from *PointPillars* (Lang et al., 2019), *VoxelNet* (Zhou and Tuzel, 2018), and concluding with a classification head from *SSD* (Liu et al., 2016). Due to the lack of a public dataset with PLO labels, they created their own dataset for training and evaluation, achieving a classification accuracy of 0.93, excluding billboard poles. While deep learning networks offer improved recognition accuracy, they are heavily reliant on extensive training datasets (Plachetka et al., 2021; Dong et al., 2023). That highlights the urgent necessity to develop methods that can effectively learn from limited training instances.

In summary, the hand-crafted features used in traditional methods cannot describe and differentiate PLOs well enough, while the current deep-learning feature extraction requires a large amount of training data. Therefore, a PLO classification method which ables to extract high-level features but requires a few training samples needs to be developed.

2.3 Few-shot Learning

Few-shot learning holds substantial significance and poses notable challenges in the deep learning field. It aims to solve the problem that traditional deep learning models typically require massive amounts of supervised samples to ensure their generalization capabilities.

Data augmentation is an intuitive method to increase the number of training samples and enhance data diversity. For example, we can crop and add noise to point cloud as basic augmentation operations. Previous studies have designed more complex network models to generate better data, such as encoder-decoder augmentation networks (Schwartz et al., 2018; Chen et al., 2019) and generative adversarial networks (GAN) (Gao et al., 2018; Antoniou et al., 2017). In point cloud data, another viable approach to accomplish few-shot Learning involves initially training a model on a large volume of synthetic data and then fine-tuning it with a limited number of real point cloud scenes (Chitnis et al., 2021; Huang et al., 2023).

Meta-learning, also known as "learning to learn", is a recent subfield of few-shot Learning, where the central idea is to design models that can learn new skills or rapidly adapt to new environments with little input after training. During training, a meta-learning model is exposed to various tasks, each with a small amount of training and testing data. The model is trained not

[1] https://github.com/Cipher-zzz/Parkville-3D

just to perform the inference but to learn a strategy for learning the tasks. That enables the model, after training, to adapt when it encounters a new task (Finn et al., 2017).

Unsupervised learning, a branch of deep learning, is characterized by its ability to learn from data without reliance on human-labelled data. It can significantly reduce the cost of labelling data and thus can be considered a few-shot Learning approach. For instance, an autoencoder can be trained without reliance on any labelled data (Yang et al., 2021). It comprises an encoder and a decoder. The encoder's role is to encode the input into a feature vector, while the decoder's function is to reconstruct the original input from this feature vector. The autoencoder does not necessarily need to comprehend the specifics of the input data, its purpose is to learn how to encapsulate the essence of the input and leverage that distilled feature representation for accurate input reconstruction. Following this process, the well-trained encoder can serve as the backbone for other tasks, such as classification or segmentation. Similar to that self-reconstruction, we can assign the autoencoder to learn from masked input (randomly hide) and ask it to train on the completion task (Yu et al., 2022) or random sample the input and ask it to up-sample back to the original input (Remelli et al., 2019).

Another intuitive and effective few-shot Learning method is metric learning. The key idea of such methods is to classify the unknown sample by its similarity to the learned dataset. Similar pairs of samples can obtain higher similarity scores, while dissimilar pairs receive lower scores. Metric Learning utilizes training datasets to establish a similarity measure and then generalizes this to the test set of tasks. The similarity measure can range from a simple distance measurement (Ye and Guo, 2018; Scott et al., 2018) to complex networks (Koch et al., 2015) or any other viable algorithms which able to estimate the similarity between features.

3. Pole-NN

Rather than relying on extensive training datasets to develop a deep neural network, we focus on a non-parametric methodology capable of extracting and interpreting features from a limited support set instead of a huge amount of training set. Drawing inspiration from the 2D segmentation method SG-One (Zhang et al., 2020), as well as from established metric learning methodologies (Ye and Guo, 2018; Scott et al., 2018), we have conceptualized a feature similarity-based classification framework. Within that framework, we introduce our pole-like objects feature extractor, Pole-NN, which is inspired by the principles of Point-NN (Zhang et al., 2023). This approach signifies a strategic shift towards more efficient and adaptable outdoor point cloud classification in situations where data availability is limited.

Building on the preceding discussion, this section will first provide an overview of our proposed framework. After this introductory exposition, we will delve into a detailed description of the framework, with a particular focus on explaining the rationale behind our selection of several critical components. That will illuminate the strategic decisions that underpin the choice of its key elements, thereby offering a comprehensive understanding of our methodological approach.

3.1 Feature Similarity-based Classification

Figure 1 shows our feature similarity-based classifier which involves two key sections: 'Supporting feature preparation' and 'Inference'. In the 'Supporting feature preparation' phase, each distinct object is passed through a feature extraction unit which extracts its unique feature representation. Those features are then categorized for creating a comprehensive support feature database. It's important to note that 'Supporting Feature Preparation' encompasses a broader concept than the traditional 'Training' phase in deep learning. Specifically, when employing a trainable feature extractor, this phase aligns with conventional 'Training'. Conversely, when leveraging a pre-trained feature extractor or training-free methods, this phase is not 'training' but building up the features that support the classification in the 'Inference' phase. Following the preparation phase, the 'inference' phase utilizes the same feature extraction unit to extract features from the query input. Those features are then compared against the established support feature database. The category of the most similar feature in the database is then defined as the inferred result.

This framework allows for flexibility in the choice of feature extraction and similarity measurement units. The feature extraction component can be many methods ranging from pre-trained models, and training-free models, to even traditional methodologies such as Principal Component Analysis (PCA). The critical requirement is that the chosen method should efficiently interpret the point cloud object and compress this information into a compact feature vector. On the other hand, the similarity measurement unit, which is pivotal in gauging the similarity among feature vectors, can vary from straightforward distance measurements to more intricate network-based algorithms or any other effective algorithms that can accurately estimate the similarity between features.

In line with previous studies, we employ the notations X and Y to represent the number of support cases and query cases, respectively. The C-way K-shot learning task is utilized to define the few-shot learning where C denotes the number of classes, and K represents the count of support samples per category (Lu et al., 2023). The primary objective of this research is to identify a combination of a feature extraction method and a similarity measurement that achieves the best performance. The goal is to attain high-accuracy classification with the least possible reliance on the number of K, irrespective of the size of Y.

3.2 Training-free Feature Extraction

Given the point cloud of a separate outdoor object, our feature extraction approach is designed to produce a feature embedding that effectively encapsulates the characteristics of the captured object. Our objective is for feature embeddings of identical objects to be closely aligned in the feature space, distinctly separated from those of different categories. That separation facilitates the classification of those objects through a feature similarity-based approach, as detailed in Section 3.1.

Considering the variability in the number of points constituting different objects' point clouds, our initial step involves bootstrapping 512 points from each object's point cloud. This random sampling process ensures uniformity in the input size, and the choice of 512 points has been validated for its efficacy in point cloud classification (Dovrat et al., 2019). Subsequently, we employ a hierarchical point cloud encoding strategy, inspired by Pointnet++ (Qi et al., 2017b). The initial encoder is tasked with capturing broad, general features from the point cloud, operating with a full receptive field. The subsequent encoder focuses on mid-level features, utilizing a half-sized receptive field. Since it covers only a portion of the point cloud at

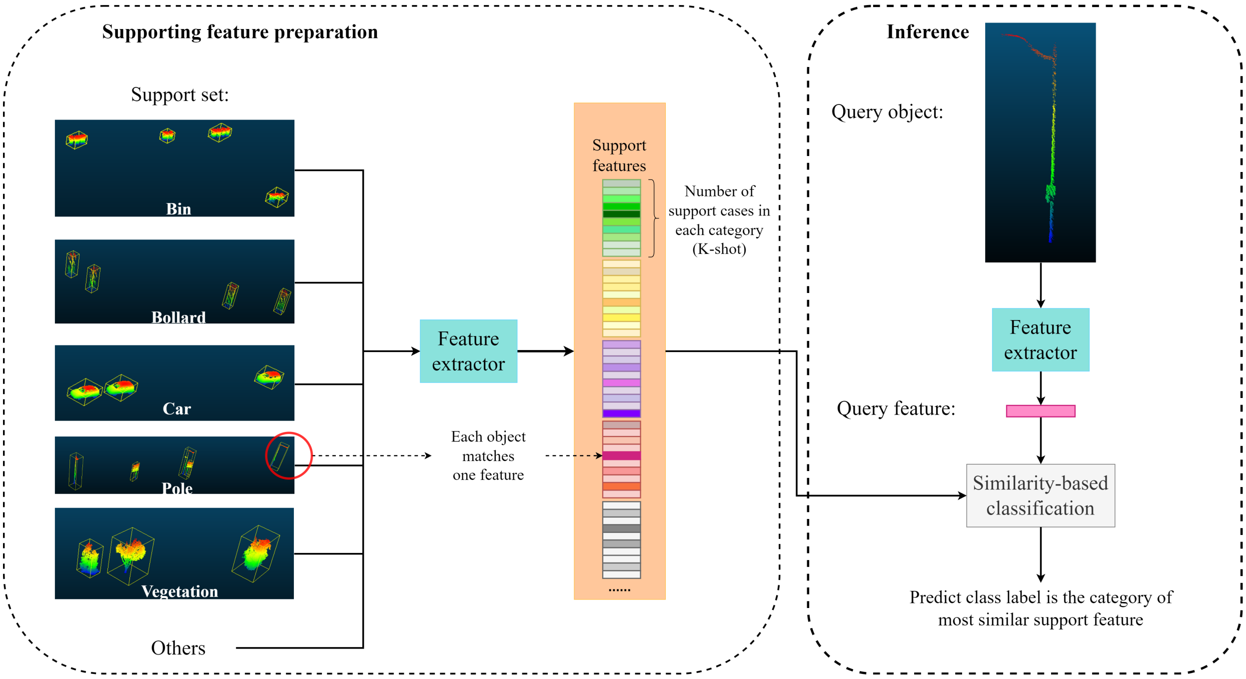

Figure 1. Feature similarity-based classification

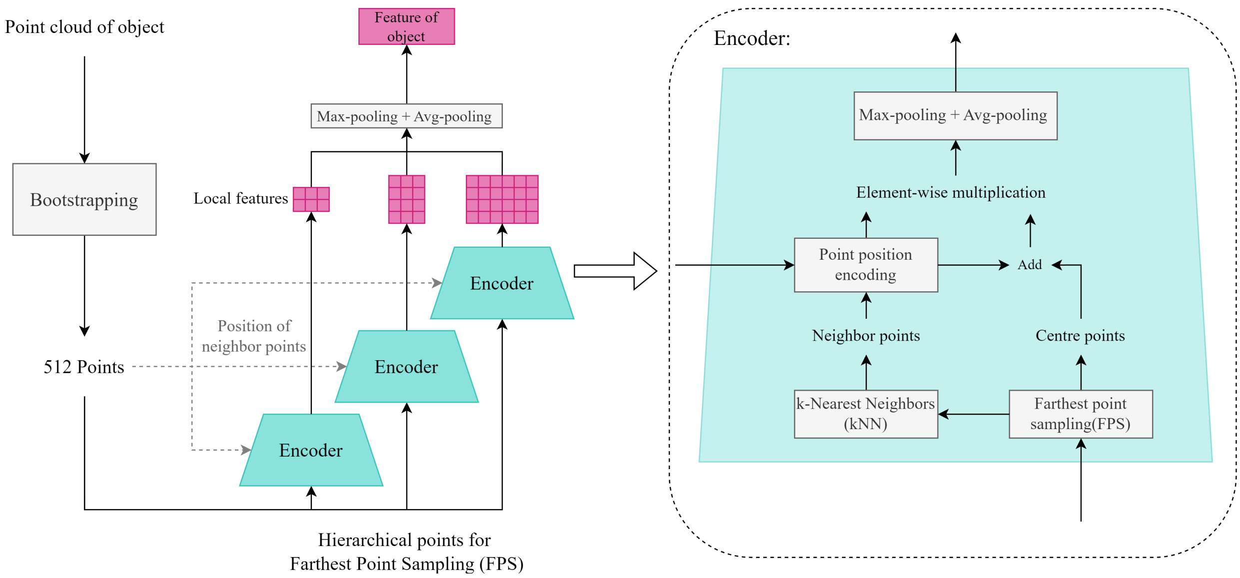

Figure 2. Pole-NN

a time, it operates twice, thereby doubling the size of the locally extracted features compared to the first encoder. The final encoder concentrates on extracting more precise, narrow features. We then aggregate those three multi-level features using both max-pooling and average-pooling to synthesize a comprehensive global feature representation.

3.2.1 Encoder The encoding mechanism in our classifier was adopted the encoder design from Point-NN (Zhang et al., 2023). Based on Pointnet++ (Qi et al., 2017b), Point-NN uses farthest point sampling (FPS) to create a set of key points. For each of those key points, it considered the k nearest neighbours to represent its local features. A notable aspect of Point-NN is its unique approach to positional encoding for 3D point cloud data, mirroring the concept used in the original Transformer architecture (Vaswani et al., 2017). This process involves multiplying the extracted features with a global positional embedding, which aids in representing the inter-relations among all the 3D points. This step is followed by a max-pooling layer, which serves to compress the feature vector, thereby reducing its dimension while retaining critical information. Since Point-NN is a fully training-free network, it effectively addresses the over-fitting problem without needing extensive training data, unlike traditional networks that rely on learnable parameters and struggle with limited training samples. That shows the tremendous potential usage of Point-NN in 3D point cloud few-shot learning tasks.

Our Pole-NN was customized for outdoor point cloud object classification, especially for the PLOs while the original Point-

NN was developed and tested exclusively for small indoor objects. Although it shows great potential in few-shot learning among indoor datasets such as ModelNet40 and ShapeNet, its effectiveness diminishes when applied to outdoor street scene point clouds. It failed to detect the thin but tall poles, indicating that the receptive field of Point-NN is not large enough for large outdoor objects.

Besides, the original datasets used in Point-NN are the point clouds generated from the 3D objects, exhibiting smooth, orderly structures with minimal noise. In contrast, outdoor datasets are inherently noisier, with significant shifts in data due to point cloud registration and vehicular disturbances.

Addressing those challenges, we modified the original Point-NN architecture by reducing the layer of the encoders. That adjustment encourages the model to ignore inessential details, thereby mitigating the effects of noise. Furthermore, by modifying the encoding complexity, we were able to broaden the scope of k-nearest neighbours, effectively extending the receptive field of the model, as the 'k' value is the amount of k-nearest points considered for each key point representation.

3.3 Similarity-based Classification

In the inference stage of our proposed method, we opt for Cosine Similarity as our primary metric for measuring feature similarity. This decision is grounded in its widespread adoption and proven effectiveness in numerous previous studies within the field of few-shot learning (Lu et al., 2023). It measures the cosine of the angle between two non-zero vectors, which can be more informative than comparing the raw data values (Han et al., 2022).

The cosine similarity is calculated by taking the dot product of two vectors and dividing it by the product of their norms (As shown in Formula 1). This results in a value between -1 and 1, where a value of 1 shows the vectors are identical, and a value of -1 shows the vectors are diametrically opposed. A value of 0 indicates that the vectors are unrelated or independent.

$$SIM(\mathbf{a},\mathbf{b}) = \frac{\mathbf{ab}}{\|\mathbf{a}\|\|\mathbf{b}\|} = \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n}(a_i)^2}\sqrt{\sum_{i=1}^{n}(b_i)^2}} \quad (1)$$

4. Experiments and Results

4.1 Experiment Dataset

4.1.1 Parkville-3D Dataset Parkville-3D dataset is a street-scene point cloud dataset we have captured by the Hovermap lidar sensor. The dataset is an invaluable resource for exploring deep learning models that deal with outdoor environments, explicitly focusing on PLO recognition.

The motivation behind creating the Parkville-3D dataset stemmed primarily from the noticeable dearth of dense, outdoor, street-scene point cloud datasets. Existing datasets mainly consist of unregistered frames designed with a specific focus on real-time vehicle recognition, such as Kitti (Fritsch et al., 2013) and nuScenes (Fong et al., 2021). Besides, there is no existing open-sourced dataset that fine-grained classified PLOs.

Our dataset is currently semantically segmented into categories: unclassified, electrical pole, light pole, road sign, vehicle, vegetation, building, and pedestrian. That segmentation was labelled to allow extracting separate street objects and testing the classification performance of our proposed model.

4.1.2 Paris-Lille-3D Dataset The Paris-Lille-3D dataset, part of the NPM3D benchmark suite, offers a rich collection of point cloud data from urban environments in Paris and Lille, France (Roynard et al., 2018). Captured using mobile lidar technology, this dataset encompasses a wide array of urban features including buildings, vehicles, and pole-like objects, making it an invaluable resource for research in 3D object recognition and classification within urban landscapes. Compared with the Parkville-3D dataset, Paris-Lille-3D does not contain subclasses of PLOs but it has bollards and bins where not exist in Melbourne.

4.2 Experiment and Analysis

To evaluate the proposed framework, we implement the model with PyTorch and design two classification tasks with those two datasets.

We first evaluate the performance of differentiating various types of PLOs by an object classification task among 106 point cloud objects in the Parkville-3D Dataset. Figure 3 shows the support set, which contains five support samples provided to Pole-NN for building the support features. Figure 4 presents the 106 testing objects in the query set, that will be classified during the inference stage.

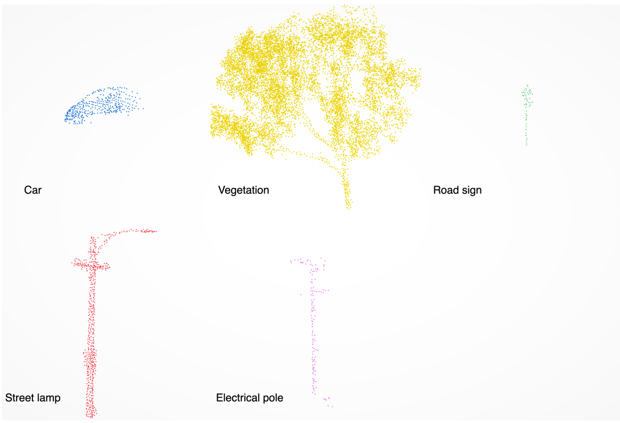

Figure 3. 5-way 1-shot supporting data from Parkville-3D Dataset

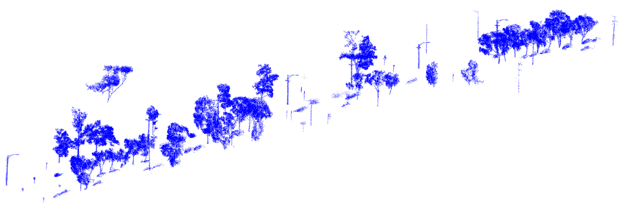

Figure 4. 5-way 1-shot query data from Parkville-3D Dataset

By using t-distributed stochastic neighbour embedding (t-SNE), we visualize the dimension-reduced features extracted by the Pole-NN. As shown in Figure 5, we can clearly see that the three PLO categories stay closer to each other, while cars and vegetation are farther away, which indicates that Pole-NN can learn the general characteristics of objects with different shapes. Table 1 shows the testing result matrix, it can be seen that Pole-NN can distinguish well between PLOs and other road assets.

To verify our findings, we designed a similar classification experiment in the Paris-Lille-3D dataset, which has five different categories and different shapes of the same category objects.

	Parkville-3D Dataset				Paris-Lille-3D Dataset			
	Number	Precision	Recall	F1-Score	Number	Precision	Recall	F1-Score
Pole	41	0.79	0.73	0.75	34	0.47	0.79	0.59
Electrical pole	7	0.54	1.00	0.70	-	-	-	-
Light pole	11	0.50	0.45	0.48	-	-	-	-
Road sign	23	1.00	0.78	0.88	-	-	-	-
Car	31	1.00	1.00	1.00	64	0.93	1.0	0.96
Vegetation	34	1.00	1.00	1.00	31	1.00	0.87	0.93
Bollard	-	-	-	-	81	1.00	0.51	0.67
Bin	-	-	-	-	21	0.54	0.95	0.69
Total	106	0.92	0.90	0.90	231	0.86	0.77	0.78

Table 1. 5-way 1-shot classification results

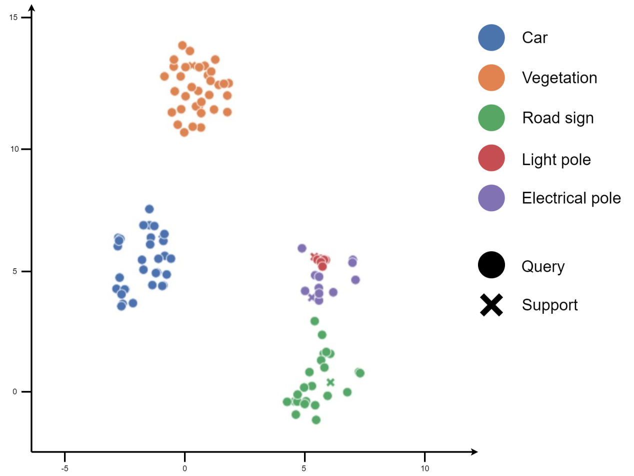

Figure 5. t-SNE visualization of Pole-NN extracted features on Parkville-3D Dataset

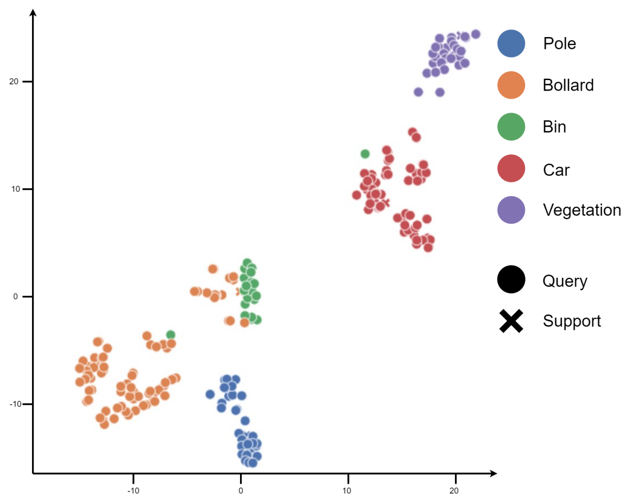

Figure 6. t-SNE visualization of Pole-NN extracted features on Paris-Lille-3D Dataset

As shown in Table 1 and Figure 6, despite a dip in the overall F1-score observed in the second experiment, particularly within the 'Bin' and 'Bollard' categories, the results continue to underscore the effectiveness of our method. The noted decrease in performance is largely due to the sub-sampling process, during which the features of bin and bollard objects became nearly indistinguishable, posing a challenge for accurate classification.

5. Discussions and Future Directions

The experimental outcomes demonstrate that our proposed methodology is capable of effectively and efficiently distinguishing outdoor point cloud objects, demonstrating its ability to classify objects with minimal exposure to supporting samples. That represents a significant advancement over conventional deep learning approaches, which typically rely on extensive data labelling and model re-training or fine-tuning to adapt to new environments. Our framework, by contrast, requires only a single labelled data point as a supporting sample, underscoring its efficiency and practicality.

Furthermore, the findings highlight the rich informational content inherent in the relational information among points. That suggests a deeper understanding of objects can be achieved by leveraging that relational information, rather than relying solely on complex, trainable networks. Our approach, which focuses on extracting positional relationships, prompts a reevaluation of current strategies, advocating for the exploration of point cloud data that moves beyond the application of computationally intensive deep learning models, particularly those adapted from 2D computer vision paradigms.

Looking ahead, it is imperative to test and validate our method across additional datasets featuring varying point densities to further verify its efficacy. Moreover, integrating our framework with other models to facilitate comprehensive object recognition and semantic segmentation represents a crucial avenue for future research. That will not only broaden the applicability of our method but also contribute to the ongoing advancement of point cloud processing techniques.

6. Conclusion

We showcase a framework that classifies outdoor point cloud objects with remarkable efficiency using just a single support sample. We also introduce the Parkville-3D Dataset, a significant contribution to PLO classification, with its detailed semantic segmentation of road assets especially subclass labels of PLOs. The effectiveness of our approach, which leverages the rich spatial relationships within 3D point cloud data, marks a departure from traditional, data-intensive deep learning methods. Looking ahead, we aim to extend our validation across varied datasets and explore the integration of our framework with other methods for comprehensive object recognition and semantic segmentation.

References

Antoniou, A., Storkey, A., Edwards, H., 2017. Data augmentation generative adversarial networks. *arXiv preprint arXiv:1711.04340*.

Boulch, A., Puy, G., Marlet, R., 2020. FKAConv: Feature-kernel alignment for point cloud convolution. *Proceedings of the Asian Conference on Computer Vision*.

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. Publisher: Elsevier.

Chen, Z., Fu, Y., Zhang, Y., Jiang, Y.-G., Xue, X., Sigal, L., 2019. Multi-level semantic feature augmentation for one-shot learning. *IEEE Transactions on Image Processing*, 28(9), 4594–4605. Publisher: IEEE.

Chitnis, S. A., Huang, Z., Khoshelham, K., 2021. Generating Synthetic 3d Point Segments for Improved Classification of Mobile LIDAR Point Clouds. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 139–144.

Dong, H., Chen, X., Särkkä, S., Stachniss, C., 2023. Online pole segmentation on range images for long-term LiDAR localization in urban environments. *Robotics and Autonomous Systems*, 159, 104283.

Dovrat, O., Lang, I., Avidan, S., 2019. Learning to sample. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2760–2769.

Ferrin, A., Larrea, J., Realpe, M., Ochoa, D., 2018. Detection of utility poles from noisy Point Cloud Data in Urban environments. *Proceedings of the 2018 Artificial Intelligence and Cloud Computing Conference*, 53–57.

Finn, C., Abbeel, P., Levine, S., 2017. Model-agnostic meta-learning for fast adaptation of deep networks. *International Conference on Machine Learning*, PMLR, 1126–1135.

Fong, W. K., Mohan, R., Hurtado, J. V., Zhou, L., Caesar, H., Beijbom, O., Valada, A., 2021. Panoptic nuScenes: A Large-Scale Benchmark for LiDAR Panoptic Segmentation and Tracking. *arXiv preprint arXiv:2109.03805*.

Fritsch, J., Kuehnl, T., Geiger, A., 2013. A new performance measure and evaluation benchmark for road detection algorithms. *International Conference on Intelligent Transportation Systems (ITSC)*.

Gao, H., Shou, Z., Zareian, A., Zhang, H., Chang, S.-F., 2018. Low-shot learning via covariance-preserving adversarial augmentation networks. *Advances in Neural Information Processing Systems*, 31.

Gholampooryazdi, B., Hämmäinen, H., Vijay, S., Savisalo, A., 2017. Scenario planning for 5g light poles in smart cities. *2017 Internet of Things Business Models, Users, and Networks*, IEEE, 1–7.

Han, J., Pei, J., Tong, H., 2022. *Data mining: concepts and techniques*. Morgan kaufmann.

Hao, W., Wang, Y., Li, Y., Shi, Z., Zhao, M., Liang, W., 2018. Hierarchical extraction of pole-like objects from scene point clouds. *Optical Engineering*, 57(8), 083106. Publisher: SPIE.

He, H., Khoshelham, K., Fraser, C., 2017. A two-step classification approach to distinguishing similar objects in mobile lidar point clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4, 67–74.

Huang, J., You, S., 2015. Pole-like object detection and classification from urban point clouds. *2015 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 3032–3038.

Huang, R., Peng, S., Takmaz, A., Tombari, F., Pollefeys, M., Song, S., Huang, G., Engelmann, F., 2023. Segment3D: Learning Fine-Grained Class-Agnostic 3D Segmentation without Manual Labels. arXiv:2312.17232 [cs].

Kang, Z., Yang, J., Zhong, R., Wu, Y., Shi, Z., Lindenbergh, R., 2018. Voxel-based extraction and classification of 3-D pole-like objects from mobile LiDAR point cloud data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(11), 4287–4298. Publisher: IEEE.

Koch, G., Zemel, R., Salakhutdinov, R., 2015. Siamese neural networks for one-shot image recognition. *ICML deep learning workshop*, 2, Lille. Issue: 1.

Lang, A. H., Vora, S., Caesar, H., Zhou, L., Yang, J., Beijbom, O., 2019. Pointpillars: Fast encoders for object detection from point clouds. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12697–12705.

Li, F., Lehtomäki, M., Elberink, S. O., Vosselman, G., Kukko, A., Puttonen, E., Chen, Y., Hyyppä, J., 2019a. Semantic segmentation of road furniture in mobile laser scanning data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 154, 98–113. Publisher: Elsevier.

Li, F., Oude Elberink, S., Vosselman, G., 2018. Pole-like road furniture detection and decomposition in mobile laser scanning data based on spatial relations. *Remote Sensing*, 10(4), 531. Publisher: MDPI.

Li, Y., Wang, W., Li, X., Xie, L., Wang, Y., Guo, R., Xiu, W., Tang, S., 2019b. Pole-like street furniture segmentation and classification in mobile LiDAR data by integrating multiple shape-descriptor constraints. *Remote Sensing*, 11(24), 2920. Publisher: MDPI.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A. C., 2016. Ssd: Single shot multibox detector. *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, Springer, 21–37.

Lu, J., Gong, P., Ye, J., Zhang, J., Zhang, C., 2023. A survey on machine learning from few samples. *Pattern Recognition*, 139, 109480.

Luo, H., Chen, C., Fang, L., Khoshelham, K., Shen, G., 2020. MS-RRFSegNet: Multiscale regional relation feature segmentation network for semantic segmentation of urban scene point clouds. *IEEE Transactions on Geoscience and Remote Sensing*, 58(12), 8301–8315. Publisher: IEEE.

Luo, Z., Gao, L., Xiang, H., Li, J., 2023. Road object detection for HD map: Full-element survey, analysis and perspectives. *ISPRS Journal of Photogrammetry and Remote Sensing*, 197, 122–144. Publisher: Elsevier.

Nie, D., Lan, R., Wang, L., Ren, X., 2022. Pyramid architecture for multi-scale processing in point cloud segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17284–17294.

Plachetka, C., Fricke, J., Klingner, M., Fingscheidt, T., 2021. DNN-Based Recognition of Pole-Like Objects in LiDAR Point Clouds. *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, IEEE, 2889–2896.

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.

Remelli, E., Baque, P., Fua, P., 2019. Neuralsampler: Euclidean point cloud auto-encoder and sampler. *arXiv preprint arXiv:1901.09394*.

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. *The International Journal of Robotics Research*, 37(6), 545–557. Publisher: SAGE Publications Sage UK: London, England.

Schwartz, E., Karlinsky, L., Shtok, J., Harary, S., Marder, M., Kumar, A., Feris, R., Giryes, R., Bronstein, A., 2018. Delta-encoder: an effective sample synthesis method for few-shot object recognition. *Advances in Neural Information Processing Systems*, 31.

Scott, T., Ridgeway, K., Mozer, M. C., 2018. Adapted deep embeddings: A synthesis of methods for k-shot inductive transfer learning. *Advances in Neural Information Processing Systems*, 31.

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. Publisher: MDPI.

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.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, \., Polosukhin, I., 2017. Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

Wang, J., Wang, C., Xi, X., Wang, P., Du, M., Nie, S., 2022. Location and Extraction of Telegraph Poles from Image Matching-Based Point Clouds. *Remote Sensing*, 14(3), 433. Publisher: MDPI.

Wu, F., Wen, C., Guo, Y., Wang, J., Yu, Y., Wang, C., Li, J., 2017. Rapid Localization and Extraction of Street Light Poles in Mobile LiDAR Point Clouds: A Supervoxel-Based Approach. *IEEE Transactions on Intelligent Transportation Systems*, 18(2), 292–305. Conference Name: IEEE Transactions on Intelligent Transportation Systems.

Yan, L., Li, Z., Liu, H., Tan, J., Zhao, S., Chen, C., 2017. Detection and classification of pole-like road objects from mobile LiDAR data in motorway environment. *Optics & Laser Technology*, 97, 272–283.

Yang, J., Ahn, P., Kim, D., Lee, H., Kim, J., 2021. Progressive seed generation auto-encoder for unsupervised point cloud learning. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 6413–6422.

Ye, M., Guo, Y., 2018. Deep triplet ranking networks for one-shot recognition. *arXiv preprint arXiv:1804.07275*.

Yeong, D. J., Velasco-Hernandez, G., Barry, J., Walsh, J., 2021. Sensor and sensor fusion technology in autonomous vehicles: A review. *Sensors*, 21(6), 2140. Publisher: MDPI.

Yu, X., Tang, L., Rao, Y., Huang, T., Zhou, J., Lu, J., 2022. Point-bert: Pre-training 3d point cloud transformers with masked point modeling. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 19313–19322.

Yuma, M., Keisuke, K., Hiroshi, M., 2018. Classification of pole-like objects using point clouds and images captured by mobile mapping systems.

Zhang, R., Wang, L., Wang, Y., Gao, P., Li, H., Shi, J., 2023. Parameter is not all you need: Starting from non-parametric networks for 3d point cloud analysis. *arXiv preprint arXiv:2303.08134*.

Zhang, X., Wei, Y., Yang, Y., Huang, T. S., 2020. Sg-one: Similarity guidance network for one-shot semantic segmentation. *IEEE Transactions on Cybernetics*, 50(9), 3855–3865. Publisher: IEEE.

Zhou, Y., Tuzel, O., 2018. Voxelnet: End-to-end learning for point cloud based 3d object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4490–4499.