ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume X-1/W2-2025
13th International Conference on Mobile Mapping Technology (MMT 2025)
"Mobile Mapping for Autonomous Systems and Spatial Intelligence", 20–22 June 2025, Xiamen, China

A Feature-Driven Approach to Semantic Segmentation in Large-Scale 3D Urban Dataset

Jing Du[1], John Zelek[2], Michael A. Chapman[3], Jonathan Li[1,4*]

[1] Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada - j7du@uwaterloo.ca
[2] Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada - jzelek@uwaterloo.ca
[3] Department of Civil Engineering, Toronto Metropolitan University, Toronto, ON M5B 2K3, Canada - mchapman@torontomu.ca
[4] Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada– junli@uwaterloo.ca

Keywords: Point Cloud, Semantic Segmentation, Novel Class Discovery, Scene Understanding.

Abstract

Urban environments are continually evolving, which presents significant challenges for 3D semantic segmentation systems that must adapt to emerging object categories. In this paper, we address the problem of Novel Class Discovery (NCD) in 3D semantic segmentation for urban scenes. We introduce a feature-driven framework that leverages the Dynamic Multi-level Feature Synthesis Module (D-MFSM) to extract and integrate multi-scale, cross-view structural information from raw urban point clouds. D-MFSM dynamically partitions point clouds via an adaptive grouping mechanism that utilizes a learnable spatial weight vector, and subsequently constructs local neighborhoods by means of an improved farthest point sampling strategy. The extracted local features are then processed by a dual-path adaptive synthesis mechanism and further refined through a novel cross-axis reordering strategy, which together yield comprehensive aggregated feature representations. These representations facilitate robust novel class discovery while maintaining high segmentation accuracy on known classes. Comprehensive evaluations on the DALES dataset demonstrate that the proposed approach yields substantial improvements in segmentation performance across diverse urban scenarios. The proposed framework, therefore, offers a complementary solution to existing methods and contributes to the development of more adaptive and accurate 3D semantic segmentation systems in complex urban settings.

1. Introduction

Rapid urbanization and the continuous evolution of urban infrastructures necessitate robust 3D semantic segmentation methods for applications such as urban planning, infrastructure monitoring, environmental management, and autonomous driving (Guo et al., 2021; Zou et al., 2024; Du et al., 2024). While significant progress has been made in 3D semantic segmentation, the task of Novel Class Discovery (NCD) within this domain remains underexplored despite its high relevance to real-world applications. In urban environments, new object categories may emerge as cities evolve, making it essential for segmentation frameworks to not only maintain performance on known classes but also reliably identify and delineate previously unseen categories from unlabeled data.

Most existing work on NCD has primarily focused on 2D image analysis (Han et al., 2020, 2019; Zhao and Han, 2021; Jia et al., 2021; Zhao et al., 2022; Li et al., 2023), and several pioneering studies have successfully extended these concepts to 3D point clouds (Riz et al., 2023; Du et al., 2025). These innovative approaches employ advanced techniques such as clustering, uncertainty-aware pseudo-labeling, and multi-head segmentation strategies, thereby laying a strong foundation for discovering novel classes in three-dimensional urban data. Their contributions have been instrumental in demonstrating the feasibility of leveraging both labeled and unlabeled data for comprehensive scene understanding in complex environments. Motivated by these seminal works, we propose a complementary method that further harnesses the rich multi-scale and cross-view structural information inherent in urban 3D point clouds.

We propose a novel feature-driven framework that incorporates the Dynamic Multi-level Feature Synthesis Module (D-MFSM). The D-MFSM is designed to extract and integrate detailed structural cues across multiple scales, thereby enhancing segmentation performance for both established and emergent classes in urban settings.Our approach dynamically partitions raw urban point clouds and employs a dual-path adaptive synthesis mechanism to extract and fuse multi-scale, cross-view structural features. Moreover, a cross-axis reordering strategy, combined with linear and non-linear transformations, is introduced to generate comprehensive aggregated feature representations. These representations enable robust novel class discovery within a unified 3D semantic segmentation framework. Figure 1 illustrates the partitioning of the DALES dataset (Varney et al., 2020) into ground truth, base classes, and novel classes. We validate our method on the DALES dataset, a large-scale aerial LiDAR dataset that encompasses diverse urban scenes, and our experiments demonstrate substantial improvements in segmenting both base and novel classes in realistic urban settings.

2. Related work

Hsu et al. (2018) was recognized as the first solution to the novel class discovery problem (Troisemaine et al., 2023). This innovative method utilized learned pairwise similarity predictions, facilitating unsupervised learning across various domains and tasks. By integrating a learnable clustering objective into neural networks, it significantly enhanced cross-domain and cross-task transfer learning, marking a crucial advancement in the field and setting a foundational standard for future research. The study (Han et al., 2019) focused on the challenge of discovering novel object categories in unlabeled image collections, utilizing prior knowledge from related, labeled classes. It modified Deep Embedded Clustering (DEC) for transfer learning, introducing enhancements like a representational bottleneck, tem-

* Corresponding author

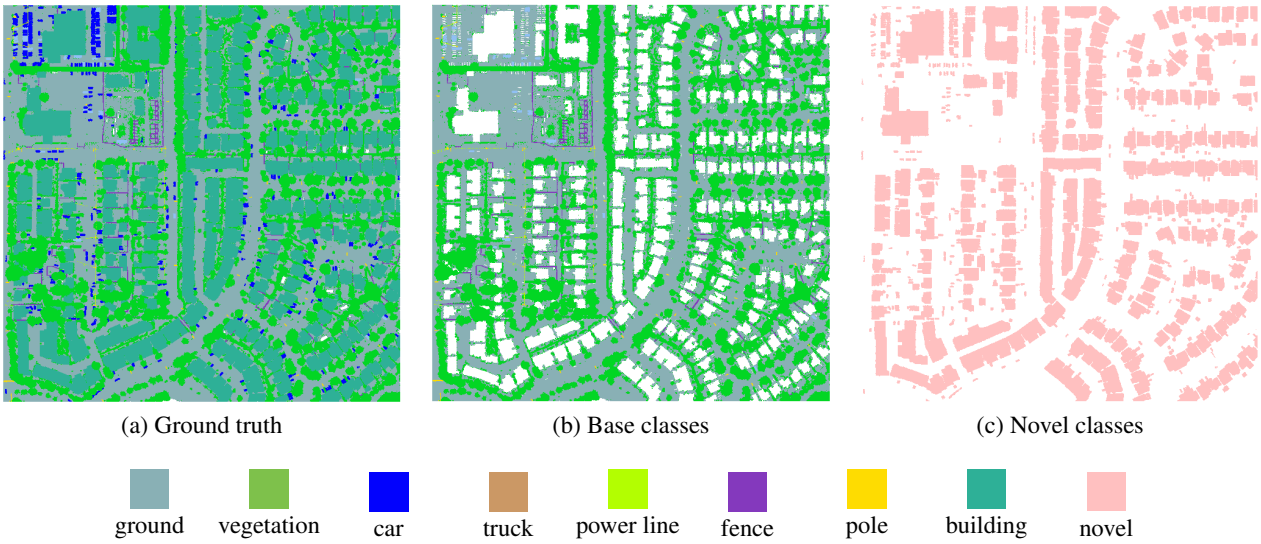

Figure 1. Visualization of the DALES-2[1] dataset partitioning into known (base) and unknown (novel) classes. (a) Ground truth labels; (b) Base (known) classes; (c) Novel (unknown) classes. The bottom panel displays the color legend for each class.

poral ensembling, and consistency to boost clustering effectiveness. Crucially, it devised a method for estimating the number of classes in unlabeled data, using labeled classes as diagnostic tools.

In 2020, Han et al. (2020) introduced a framework for identifying new classes in image datasets without pre-labeled data, using three key strategies: self-supervised learning to train image representation on both labeled and unlabeled data to reduce bias; rank statistics for knowledge transfer in clustering unlabeled images; and joint optimization of an objective function across labeled and unlabeled datasets to enhance classification and clustering performance. Building on this, Zhao and Han (2021) proposed a two-branch learning framework for novel visual category discovery, which groups unlabeled images into semantic partitions. One branch focused on local part-level details, while the other addressed overall characteristics, using dual ranking statistics to transfer knowledge from labeled to unlabeled data and generating robust pseudo labels. Furthermore, mutual knowledge distillation fosters synergy between branches, leveraging both global and local features for effective category discovery. Similarly, Jia et al. (2021) introduced a framework for discovering categories in single- and multi-modal data using both labeled and unlabeled data. This approach extends traditional contrastive learning to include category and cross-modal discrimination, enhancing representation learning. It employed Winner-Take-All (WTA) hashing to generate pseudo labels in shared representation spaces, facilitating robust knowledge transfer. This end-to-end framework combined binary and standard cross-entropy losses, effectively learning feature representations and cluster assignments, and demonstrated superior performance in challenging image and video benchmarks. Additionally, OpenMix (Zhong et al., 2021b) innovatively mixed labeled and unlabeled examples to create a joint label distribution, addressing the challenge of non-overlapping classes. This method enhanced model training by generating credible pseudo-labels and integrating reliable anchors from unlabeled examples, leading to more effective learning of novel classes.

In 2021, Zhong et al. (2021a) developed Neighborhood Contrastive Learning (NCL), which focused on learning distinct representations for clustering by exploiting local neighborhoods in the embedding space. This method gathers pseudo-positive pairs and generates hard negatives by mixing features of labeled and unlabeled data, enhancing the model’s ability to discriminate new classes in unlabeled samples. Additionally, Fini et al. (2021) introduced a Unified Objective (UNO) for novel class discovery, streamlining the process by treating cluster pseudo-labels homogeneously with ground truth labels. This approach simplifies NCD by eliminating the need for multiple objective functions and self-supervised pretraining. Utilizing multi-view, multi-head, and over-clustering strategies, UNO effectively learns powerful representations and discovers new classes using a single cross-entropy loss on both labeled and unlabeled data.

Novel Class Discovery without Forgetting (NCDwF), introduced by Joseph et al. (2022a), focused on identifying new categories in unlabeled data while preserving the accuracy of previously learned classes. It uses pseudolatent representations to reduce forgetting, a mutual-information based regularizer to enhance novel class discovery, and a Known Class Identifier for better generalized inference. This method addressed the need to incrementally identify novel categories while maintaining performance on known ones, making it suitable for dynamic, real-world applications. Additionally, Zhao et al. (2022) proposed Novel Class Discovery in Semantic Segmentation (NCDSS), which segments unlabeled images with new classes using knowledge from labeled, disjoint classes. The Entropy-based Uncertainty Modeling and Self-training (EUMS) method tackled challenges in semantic segmentation, such as distinguishing multiple classes and backgrounds. EUMS uses a saliency model for initial clustering and applies entropy ranking and dynamic reassignment for refining pseudo-labels, improving performance on novel classes by balancing label accuracy and uncertainty. In another approach, Yang et al. (2022) developed dual groups of Compositional Experts (ComEx) to analyze data in an integrated manner, merging global-to-local alignment with novel local-to-local aggregation in pseudo-labeling. This model advances the recognition of both established and new classes. Similarly, Joseph et al. (2022b) introduced ’Spacing Loss’, a loss function designed to improve latent space separability by enforcing spatial distancing between semantically different classes.

The Soft-contrastive All-in-one Network (SAN) (Zang et al.,

2023) specifically enhanced the discovery of new classes in unsupervised domain adaptation, including open-set and universal domain adaptations. SAN's innovation lies in its soft contrastive learning (SCL) loss and All-in-One (AIO) classifier. The SCL loss effectively addressed view-noise issues in data augmentation, improving feature transfer for novel class discovery. The AIO classifier was designed to overcome the overconfidence problem in existing classifiers, thereby increasing the robustness and accuracy in identifying new, previously unseen classes in the target domain. Furthermore, Gu et al. (2023) introduced a novel framework for novel class discovery that emphasizes knowledge transfer from known to novel classes. The approach centers around a class-relation representation, derived from predicted class distributions on known classes. This is complemented by a unique knowledge distillation strategy, which maintains informative class relations during the training of novel classes. Additionally, the framework features a learnable weighting function, allowing adaptive knowledge transfer based on semantic similarities between known and novel classes.

Yang et al. (2023) tackled a new aspect of novel class discovery with "Bootstrapping Your Own Prior (BYOP)", which addressed distribution-agnostic NCD where data come from unknown and potentially imbalanced class distributions. BYOP iteratively estimates class distributions based on model predictions, refining the process with a dynamic temperature technique to enhance prediction sharpness for less-confident samples. This approach results in more accurate pseudo-labels for novel classes in subsequent training iterations and shows superior performance, especially in handling imbalanced class distributions in NCD tasks. Similarly, Li et al. (2023) introduced an innovative approach to NCD using inter-class and intra-class constraints based on symmetric Kullback-Leibler divergence (sKLD). The inter-class sKLD constraint enhanced the separability of classes between labeled and unlabeled data, improving the discriminability of feature representations. Additionally, the intra-class sKLD constraint strengthens the relationship between samples and their augmentations, ensuring training stability. This dual-constraint model significantly outperforms existing NCD methods, proving more effective in learning discriminative and invariant feature representations.

NCD for semantic segmentation has been explored in 2D images; NOPS (Riz et al., 2023) pioneered its application to 3D point clouds. Motivated by inadequate 2D assumptions for multi-class, unbalanced 3D data, NOPS integrated online clustering via Sinkhorn-Knopp, uncertainty-aware pseudo-labeling, class-balanced queuing, and multi-head segmentation. This approach significantly resolved class collapse, offered a new evaluation protocol, and consistently demonstrated robust empirical improvements on several diverse benchmark datasets.

3. Methodology

In this work, we propose the Dynamic Multi-level Feature Synthesis Module (D-MFSM), which is designed to adaptively extract and fuse multi-scale, cross-view structural information from raw urban point clouds. The aggregated features are subsequently utilized in our NCD semantic segmentation framework.

The input point cloud data is represented as $\mathcal{P} \in \mathbb{R}^{B\times N\times 3}$, where B denotes the batch size and N denotes the total number of points in each sample. Each point is defined by its three-dimensional coordinates. Initially, D-MFSM employs a dynamic grouping mechanism. A learnable spatial weight vector $\boldsymbol{\omega} \in \mathbb{R}^3$ is introduced to reweight the original coordinates, thereby modulating the relative importance of each axis. The reweighted point cloud is computed as

$$\tilde{\mathcal{P}} = \mathcal{P} \odot \boldsymbol{\omega}, \tag{1}$$

where $\odot$ denotes elementwise multiplication, and $\tilde{\mathcal{P}}$ is the weighted point cloud. Based on the weighted data, an improved farthest point sampling (FPS) strategy is applied to select a representative set of center points. This yields the set

$$\mathcal{C} = \{\mathbf{c}_j \in \mathbb{R}^3 \mid j = 1, \ldots, G\}, \tag{2}$$

where each center $\mathbf{c}_j$ is a point in $\mathbb{R}^3$ and the set $\mathcal{C}$ serves as the basis for constructing local regions.

For each center $\mathbf{c}_j$, an adaptive neighborhood $\mathcal{N}(\mathbf{c}_j)$ is constructed by performing a nearest-neighbor search on the full point set $\mathcal{P}$. The mechanism dynamically determines the number of neighbors M_j for each group based on the local point distribution. This adaptive neighborhood construction ensures that the local region captures the intrinsic geometric structure of the scene. Subsequently, each local neighborhood $\mathcal{N}(\mathbf{c}_j)$ is reshaped into a tensor and passed through a primary encoder $E(\cdot)$. This encoder, which consists of a series of convolutions, batch normalization, and non-linear activations, transforms the raw 3D coordinates into an initial high-level feature map $\mathbf{F}^{(1)} \in \mathbb{R}^{(B\cdot G)\times C_1\times M}$. C_1 denotes the number of channels in the initial feature representation and M is the fixed group size obtained after dynamic grouping.

To capture the salient and overall characteristics of each local neighborhood, the maximum and mean statistics are computed along the point dimension. These are defined as

$$\mathbf{f}_{\max} = \max_{k=1,\ldots,M} \mathbf{F}^{(1)}(:,k), \tag{3}$$

$$\mathbf{f}_{\text{avg}} = \frac{1}{M}\sum_{k=1}^{M} \mathbf{F}^{(1)}(:,k), \tag{4}$$

where $\mathbf{f}_{\max}$ and $\mathbf{f}_{\text{avg}}$ represent, respectively, the per-group maximum and average feature vectors, capturing salient and overall statistics of the local region. These statistical descriptors are spatially replicated and concatenated with the original feature map to form an enriched feature tensor:

$$\mathbf{F}_c = \text{Concat}\left(\mathbf{f}_{\max} \otimes \mathbf{1},\ \mathbf{f}_{\text{avg}} \otimes \mathbf{1},\ \mathbf{F}^{(1)}\right), \tag{5}$$

where $\otimes\mathbf{1}$ denotes the replication of the vector along the point dimension and $\text{Concat}(\cdot)$ indicates channel-wise concatenation.

The enhanced feature $\mathbf{F}_c$ is then processed by a dual-path adaptive synthesis mechanism. One branch applies a linear mapping $g(\cdot)$ via a 1×1 convolution, while the other branch applies a non-linear mapping $a(\cdot)$ (composed of a sequence of convolutions followed by ReLU and Sigmoid functions). Their outputs are combined through an elementwise product:

$$\mathbf{F}_{\text{D}} = g\big(\mathbf{F}_c\big) \odot \sigma\Big(a\big(\mathbf{F}_c\big)\Big), \tag{6}$$

where $\sigma(\cdot)$ is the Sigmoid function, $\odot$ denotes the Hadamard (elementwise) product, $g(\cdot)$ represents the linear transformation, and $a(\cdot)$ represents the non-linear mapping that generates

dynamic modulation factors.

After this adaptive synthesis, a subsequent transformation unit $h(\cdot)$ further refines the features. A max-pooling operation is then performed over the local points to extract a fixed-length descriptor for each local group:

$$\mathbf{t}_j = \max_{k=1,\dots,M} h\big(\mathbf{F}_{\mathrm{D}}^{(j)}\big)(:,k), \tag{7}$$

where $\mathbf{t}_j$ is the synthesized feature vector for the j-th group and D_e denotes the target feature dimensionality after the transformation. The per-group features are then reassembled into a feature matrix $\mathbf{T} \in \mathbb{R}^{B\times G\times D_e}$. Each row corresponds to one sample in the batch and each column corresponds to the feature vector of one local group. To further incorporate spatial priors, the center of each local group $\mathbf{c}_j$ is fed into a position mapping network $P(\cdot)$ to yield a latent position code:

$$\mathbf{p}_j = P(\mathbf{c}_j) \in \mathbb{R}^{D_e}, \tag{8}$$

where $\mathbf{p}_j$ encodes the spatial context of the j-th group. The collection of all such codes forms a position matrix $P \in \mathbb{R}^{B\times G\times D_e}$.

A novel cross-axis reordering strategy is then applied to both $\mathbf{T}$ and $\mathbf{P}$. For each coordinate axis $a \in \{x, y, z\}$, sorting indices are computed based on the corresponding component of the group centers:

$$\pi_a = \mathrm{SortIndex}\left(\mathbf{c}_{:,a}\right), \tag{9}$$

where π_a denotes the permutation that orders the group centers along axis a. Using these indices, the feature matrix and position matrix are re-sampled:

$$\mathbf{T}_a = \mathrm{Gather}\left(\mathbf{T}, \pi_a\right), \quad \mathbf{P}_a = \mathrm{Gather}\left(\mathbf{P}, \pi_a\right), \tag{10}$$

where $\mathrm{Gather}(\cdot,\cdot)$ represents the operation of reordering the rows of the matrix according to the given indices. The re-sampled representations from the three axes are then concatenated to form a composite feature and position representation:

$$\tilde{\mathbf{T}} = \mathrm{Concat}\left(\mathbf{T}_x, \mathbf{T}_y, \mathbf{T}_z\right), \quad \tilde{\mathbf{P}} = \mathrm{Concat}\left(\mathbf{P}_x, \mathbf{P}_y, \mathbf{P}_z\right), \tag{11}$$

where the concatenation is performed along the group dimension, thereby fusing spatial information from three orthogonal views. The fused features $\tilde{\mathbf{T}}$ and corresponding position codes $\tilde{\mathbf{P}}$ are subsequently processed by a context interweaving unit, denoted as $\mathcal{T}(\cdot,\cdot)$, which is based on self-attention mechanisms. The output of this unit is integrated with $\tilde{\mathbf{T}}$ via a residual connection, followed by layer normalization:

$$\mathbf{F} = \mathrm{LN}\left(\tilde{\mathbf{T}} + \mathcal{T}\big(\tilde{\mathbf{T}}, \tilde{\mathbf{P}}\big)\right), \tag{12}$$

where $\mathrm{LN}(\cdot)$ denotes the layer normalization operation applied to stabilize feature distributions, and $\mathcal{T}(\cdot,\cdot)$ models the inter-group context by leveraging multi-head self-attention.

Finally, a global pooling operation is applied across the reordered group dimension to aggregate the fused information into a single global descriptor:

$$f = \frac{1}{\Lambda}\sum_{j=1}^{\Lambda} \mathbf{F}_{:,j}, \tag{13}$$

where Λ represents the total number of groups after cross-axis reordering, and f is the aggregated global feature vector for each sample. This descriptor is then mapped via a projection function $\phi(\cdot)$ to produce the final implicit feature representation:

$$f^{\mathrm{final}} = \phi(f), \tag{14}$$

where $\phi(\cdot)$ is a learned projection mapping that transforms the global descriptor f into the compact feature space utilized by our NCD semantic segmentation system.

In summary, the D-MFSM operates through a sequence of well-defined stages. It initially partitions the raw point cloud via adaptive sampling and constructs dynamic local neighborhoods. The local regions are then encoded by a primary network, and their statistical extremes and averages are fused to form an enhanced feature tensor. This tensor is further processed through dual-path synthesis, yielding refined local descriptors that are aggregated into a matrix $\mathbf{T}$. Meanwhile, position codes are generated from the group centers and, using a cross-axis reordering strategy based on sorting indices π_x, π_y, and π_z, the module reconstructs multi-view representations $\tilde{\mathbf{T}}$ and $\tilde{\mathbf{P}}$. These representations are then interwoven via a self-attention based context interweaving unit, and the resulting features are globally pooled and projected to obtain the final implicit representation f^{final}. The module integrates dynamic grouping, advanced statistical fusion, dual-path adaptive synthesis, and sophisticated cross-axis reordering with contextual interweaving, thereby constructing a high-level representation that robustly supports subsequent NCD semantic segmentation in complex urban scenes.

4. Experiments

4.1 Dataset

The DALES dataset (Varney et al., 2020) is a large-scale aerial LiDAR dataset. It was collected using an Aerial Laser Scanner (ALS), presenting unique challenges in terms of point cloud resolution and perspective. DALES is classified into eight classes: ground, building, car, truck, pole, power line, fence, and vegetation. Its diversity of scenes, including urban, suburban, rural, and commercial areas, makes it highly relevant for urban planning, land management, and environmental monitoring. This dataset allows for assessing the model's ability to segment and monitor critical urban infrastructure and natural elements over large areas, making it applicable to a wide range of urban management and planning scenarios. For our experiments, we selected the first six scenes from the training set for model training, and the first two scenes from the test set for evaluation.

4.2 Evaluation and visualization results on DALES

The evaluation results on Table 1 showcase the proposed method's ability to generalize across novel and base classes in urban environments, specifically in the context of city planning, infrastructure monitoring, road and traffic management, and land cover mapping. The dataset is split into four configurations (DALES-2^0, DALES-2^1, DALES-2^2, and DALES-2^3), with novel class discovery being a primary focus. The criteria for partitioning novel classes follow NOPS (Riz et al., 2023).

In the DALES-2^0 split, where ground and vegetation are introduced as novel classes, the proposed method demonstrates a significant improvement in segmenting these large-scale urban elements. With an mIoU of 89.74% for ground and 75.81%

Table 1. Novel class discovery results on DALES dataset (%). Pink highlighted values are the novel classes in each split. "Novel" denotes the mIoU of novel classes, "Base" indicates the mIoU of non-novel classes, and "All" shows the mIoU of all classes. We selected the first six scenes ('5080_54435', '5085_54320', '5095_54440', '5095_54455', '5100_54495', '5105_54405') from the training set for model training, and the first two scenes ('5080_54400', '5080_54470') from the test set for evaluation.

Split	Model	Ground	Building	Car	Truck	Pole	Power Line	Fence	Vegetation	Novel	Base	All
DALES-2^0	NOPS	81.50	63.39	53.75	4.75	3.23	15.75	12.76	73.59	77.55	25.61	38.59
	Ours	89.74	75.63	49.57	6.46	4.98	19.66	13.42	75.81	82.78	28.29	41.91
DALES-2^1	NOPS	91.31	71.97	0.00	5.97	6.26	15.97	19.04	82.18	35.99	36.79	36.59
	Ours	92.32	78.82	0.00	13.96	7.65	17.11	26.55	86.56	39.41	40.69	40.37
DALES-2^2	NOPS	92.68	80.81	56.46	7.57	4.89	13.99	15.00	80.20	11.29	54.84	43.95
	Ours	93.04	84.39	62.91	13.81	7.13	17.78	30.42	86.68	22.12	58.66	49.52
DALES-2^3	NOPS	92.33	78.96	57.50	5.42	0.00	13.13	16.98	82.10	6.57	55.55	43.30
	Ours	93.42	85.37	63.24	12.44	5.43	14.59	25.22	86.05	10.01	60.96	48.22

for vegetation, the method surpasses NOPS, which achieved 81.50% and 73.59%, respectively. These results are particularly relevant for land cover mapping and city planning, where accurate segmentation of ground surfaces and vegetation is crucial for evaluating land use and urban expansion. The overall performance in this split, with an mIoU of 41.91%, reflects the method's robustness in handling both natural and constructed elements in the urban landscape.

In the DALES-2^1 split, which introduces buildings and cars as novel classes, the method shows competitive results, achieving a novel class mIoU of 39.41%, an improvement over NOPS's 35.99%. The method's high mIoU of 78.82% for buildings (compared to NOPS's 71.97%) is particularly significant for infrastructure monitoring and urban cartography, where precise identification and segmentation of buildings are necessary for maintaining up-to-date city models. The ability to consistently identify complex structures such as buildings showcases the method's robustness in real-world urban applications, with the overall mIoU in this split reaching 40.37%.

For the DALES-2^2 split, where trucks and fences are novel classes, the proposed method achieves an mIoU of 22.12% for novel classes, compared to NOPS's 11.29%. The improvement in segmenting fences (30.42% mIoU vs. NOPS's 15.00%) is particularly valuable for city planning and road management, where boundaries and barriers need to be accurately detected for safety and infrastructure development. The method's overall mIoU of 49.52% demonstrates its ability to maintain strong performance in diverse urban elements, from vehicles to smaller, more intricate structures like fences.

In the DALES-2^3 split, where poles and power lines are novel categories, the method continues to outperform NOPS in novel class segmentation, achieving an mIoU of 10.01% compared to NOPS's 6.57%. The accurate identification of utility infrastructure such as poles and power lines is crucial for infrastructure monitoring, as it supports the effective maintenance and management of urban utilities.

Figure 2 presents the segmentation results for the 5080_54400 scene from the DALES dataset, comparing the method's performance across the DALES-2^0, DALES-2^1, DALES-2^2, and DALES-2^3 splits. The ground truth (Figure 2a) serves as a reference for evaluating the accuracy of segmenting urban features such as ground, vegetation, cars, trucks, power lines, fences, pole, and building. Across the splits (Figures 2b-e), the method shows solid performance in categories like building and ground, which are crucial for infrastructure management and city planning. However, categories such as poles and power line—which are often difficult to segment due to their small size and sparse distribution—remain challenging. Similarly, Figure 3 displays segmentation results for the 5080_54470 scene. The ground truth (Figure 3a) provides the baseline for comparison, and the results across the splits (Figures 3b-e) illustrate the method's strengths in segmenting larger categories like building and vegetation. The performance on more challenging categories, such as pole and power line, remains limited, reflecting the inherent difficulty in accurately segmenting these smaller features. Despite these challenges, the method demonstrates reliable segmentation of core urban elements, which are essential for applications such as infrastructure monitoring and city planning. Overall, these results demonstrate that while the method excels in segmenting larger, well-defined urban categories, further refinement is needed to improve performance on smaller, more intricate features like poles and power lines.

5. Conclusion

In this paper, we have presented a novel feature-driven framework for Novel Class Discovery (NCD) in 3D semantic segmentation of urban scenes. Our approach centers on the Dynamic Multi-level Feature Synthesis Module (D-MFSM), which dynamically partitions raw point clouds and extracts multi-scale, cross-view structural features through a dual-path adaptive synthesis mechanism. A cross-axis reordering strategy is integrated to effectively fuse spatial information from multiple viewpoints, yielding aggregated feature representations that support the segmentation of both established and emerging classes.

Comprehensive evaluations on the DALES dataset demonstrate that our method significantly improves segmentation performance across diverse urban elements. In particular, the proposed framework achieves enhanced accuracy in segmenting large-scale urban categories and in identifying novel object classes from unlabeled data, thereby providing a robust solution for practical applications in urban planning and infrastructure monitoring.

Future work will focus on further refining the segmentation of smaller and more intricate features, such as utility poles and power lines, which continue to present challenges due to their sparse and complex nature. Additionally, the incorporation of temporal dynamics and further optimization of the feature synthesis process may yield further improvements in overall segmentation performance. Overall, the proposed framework represents a complementary contribution to the advancement of 3D semantic segmentation and novel class discovery in complex urban environments.

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-X-1-W2-2025-27-2025 | © Author(s) 2025. CC BY 4.0 License.

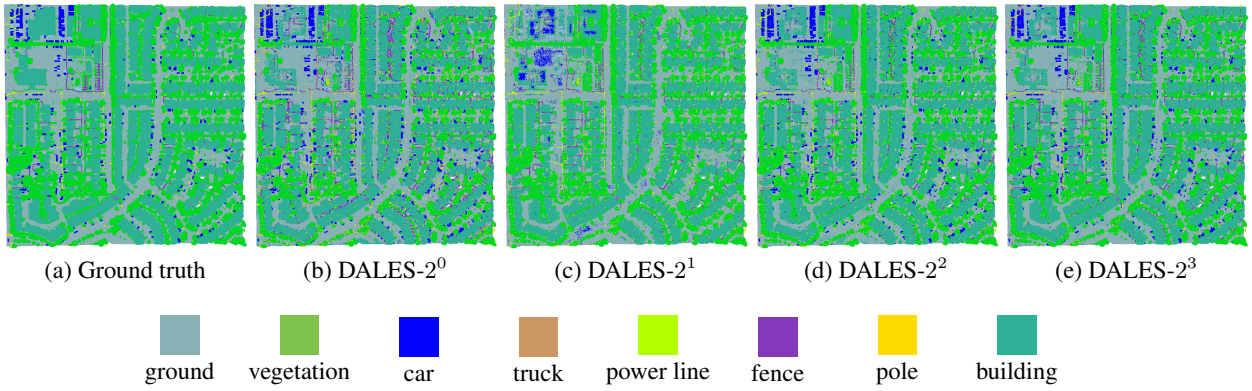

Figure 2. Segmentation results on DALES. (a) Ground truth, (b) DALES-2^0, (c) DALES-2^1, (d) DALES-2^2 and (e) DALES-2^3.

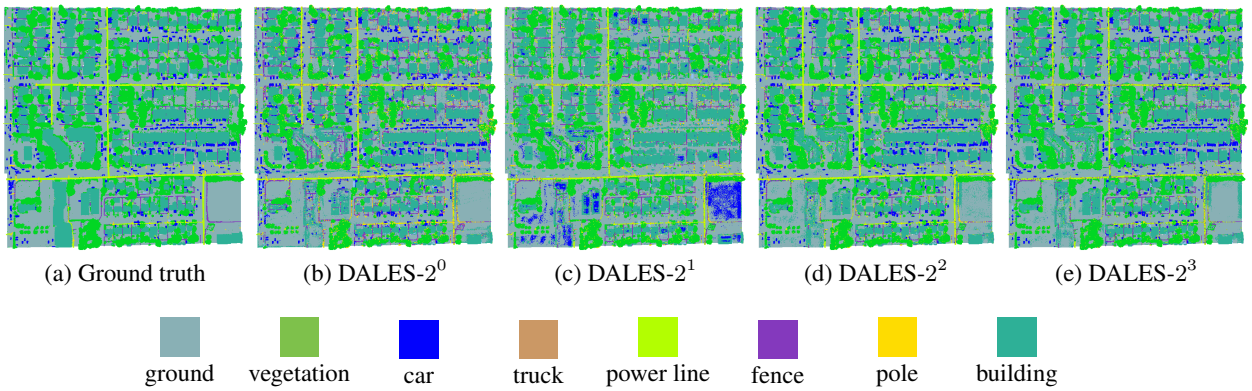

Figure 3. Segmentation results on DALES. (a) Ground truth, (b) DALES-2^0, (c) DALES-2^1, (d) DALES-2^2 and (e) DALES-2^3.

References

Du, J., Xu, L., Ma, L., Gao, K., Zelek, J., Li, J., 2025. 3D semantic segmentation: Cluster-based sampling and proximity hashing for novel class discovery. *ISPRS J. Photogramm. Remote Sens.*, 223, 274-295.

Du, J., Zelek, J., Li, J., 2024. Weather-aware autopilot: Domain generalization for point cloud semantic segmentation in diverse weather scenarios. *ISPRS J. Photogramm. Remote Sens.*, 218, 204-219.

Fini, E., Sangineto, E., Lathuilière, S., Zhong, Z., Nabi, M., Ricci, E., 2021. A unified objective for novel class discovery. *Proc. ICCV*, IEEE, 9264–9272.

Gu, P., Zhang, C., Xu, R., He, X., 2023. Class-relation knowledge distillation for novel class discovery. *Proc. ICCV*, IEEE, 16428–16437.

Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., Bennamoun, M., 2021. Deep Learning for 3D Point Clouds: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43(12), 4338–4364.

Han, K., Rebuffi, S., Ehrhardt, S., Vedaldi, A., Zisserman, A., 2020. Automatically discovering and learning new visual categories with ranking statistics. *Proc. ICLR*, OpenReview.net.

Han, K., Vedaldi, A., Zisserman, A., 2019. Learning to discover novel visual categories via deep transfer clustering. *Proc. ICCV*, IEEE, 8400–8408.

Hsu, Y., Lv, Z., Kira, Z., 2018. Learning to cluster in order to transfer across domains and tasks. *Proc. ICLR*, OpenReview.net.

Jia, X., Han, K., Zhu, Y., Green, B., 2021. Joint representation learning and novel category discovery on single- and multi-modal data. *Proc. ICCV*, IEEE, 590–599.

Joseph, K. J., Paul, S., Aggarwal, G., Biswas, S., Rai, P., Han, K., Balasubramanian, V. N., 2022a. Novel class discovery without forgetting. *Proc. ECCV*, 13684, Springer, 570–586.

Joseph, K. J., Paul, S., Aggarwal, G., Biswas, S., Rai, P., Han, K., Balasubramanian, V. N., 2022b. Spacing loss for discovering novel categories. *Proc. CVPR*, IEEE, 3760–3765.

Li, W., Fan, Z., Huo, J., Gao, Y., 2023. Modeling inter-class and intra-class constraints in novel class discovery. *Proc. CVPR*, IEEE, 3449–3458.

Riz, L., Saltori, C., Ricci, E., Poiesi, F., 2023. Novel class discovery for 3d point cloud semantic segmentation. *Proc. CVPR*, IEEE, 9393–9402.

Troisemaine, C., Lemaire, V., Gosselin, S., Reiffers-Masson, A., Flocon-Cholet, J., Vaton, S., 2023. Novel Class Discovery: an Introduction and Key Concepts. *CoRR*, abs/2302.12028.

Varney, N. M., Asari, V. K., Graehling, Q., 2020. DALES: A large-scale aerial lidar data set for semantic segmentation. *Proc. CVPR Workshops*, IEEE, 717–726.

Yang, M., Wang, L., Deng, C., Zhang, H., 2023. Bootstrap your own prior: Towards distribution-agnostic novel class discovery. *Proc. CVPR*, IEEE, 3459–3468.

Yang, M., Zhu, Y., Yu, J., Wu, A., Deng, C., 2022. Divide and conquer: Compositional experts for generalized novel class discovery. *Proc. CVPR*, IEEE, 14248–14257.

Zang, Z., Shang, L., Yang, S., Wang, F., Sun, B., Xie, X., Li, S. Z., 2023. Boosting novel category discovery over domains with soft contrastive learning and all in one classifier. *Proc. ICCV*, IEEE, 11824–11833.

Zhao, B., Han, K., 2021. Novel visual category discovery with dual ranking statistics and mutual knowledge distillation. *NeurIPS*, 22982–22994.

Zhao, Y., Zhong, Z., Sebe, N., Lee, G. H., 2022. Novel class discovery in semantic segmentation. *Proc. CVPR*, IEEE, 4330–4339.

Zhong, Z., Fini, E., Roy, S., Luo, Z., Ricci, E., Sebe, N., 2021a. Neighborhood contrastive learning for novel class discovery. *Proc. CVPR*, IEEE, 10867–10875.

Zhong, Z., Zhu, L., Luo, Z., Li, S., Yang, Y., Sebe, N., 2021b. Openmix: Reviving known knowledge for discovering novel visual categories in an open world. *Proc. CVPR*, IEEE, 9462–9470.

Zou, P., Zhao, S., Huang, W., Xia, Q., Wen, C., Li, W., Wang, C., 2024. AdaCo: Overcoming Visual Foundation Model Noise in 3D Semantic Segmentation via Adaptive Label Correction. *arXiv preprint arXiv:2412.18255*.