IGARSS 2025 - 2025 IEEE International Geoscience and Remote Sensing Symposium | 979-8-3315-0810-4/25/$31.00 ©2025 IEEE | DOI: 10.1109/IGARSS55030.2025.11243257

# Enhancing Few-Shot Point Cloud Semantic Segmentation via Superpoint Semantics

line 1: 1st Zhiling Wang
line 2: *College of Architecture and Urban Planning*
line 3: *ShenZhen University*
line 4: ShenZhen, China
line 5: 2310324005@email.szu.edu.cn

line 1: 2nd Ruisheng Wang*
line 2: *College of Architecture and Urban Planning*
line 3: *ShenZhen University*
line 4: ShenZhen, China
line 5: ruiswang@szu.edu.cn

line 1: 3rd Yujun Liu
line 2: *College of Architecture and Urban Planning*
line 3: *ShenZhen University*
line 4: ShenZhen, China
line 5: yujunliu2024@mails.szu.edu.cn

line 1: 4th Bo Guo
line 2: *College of Architecture and Urban Planning*
line 3: *ShenZhen University*
line 4: ShenZhen, China
line 5: guobo@szu.edu.cn

line 1: 5th Tsz Nam Chan
line 2: *College of Computer Science and Software Engineering*
line 3: *ShenZhen University*
line 4: ShenZhen, China
line 5: edisonchan2013928@gmail.com

line 1: 6th Yibin Tian
line 2: *College of Mechatronics and Control Engineering*
line 3: *ShenZhen University*
line 4: ShenZhen, China
line 5: ybtian@szu.edu.cn

***Abstract*—Few-shot point cloud semantic segmentation aims to segment novel categories with limited labeled data, making it crucial for remote sensing and computer vision applications. Meta-learning-based methods have demonstrated success in segmenting unseen categories. However, existing approaches struggle with insufficient information and ambiguity when handling small objects or those with similar structures. To address these limitations, we propose SpSeg, an optimization framework that leverages superpoint semantics. SpSeg exploits the observation that semantic primitives of specific categories become increasingly evident at larger point set scales, focusing on semantic primitive information previously overlooked. We first construct initial superpoints and leverage the emerging semantic primitives from expanding superpoints to obtain refined superpoint representations. SpSeg then performs pre-classification on original points based on generated superpoints, providing semantic primitive and classification information for few-shot segmentation. Experimental results demonstrate that our model achieves state-of-the-art (SOTA) performance in few-shot segmentation tasks on the S3DIS dataset, surpassing all previous methods across eight different settings, thereby validating the effectiveness of the model.**

***Keywords—Meta-Learning, Few-shot Learning, Superpoint construction, Point Cloud, Semantic Segmentation***

## I. INTRODUTION

Point cloud semantic segmentation[1-3] holds significant value in computer vision and remote sensing applications. With advancements in remote sensing technology, high-precision point cloud data acquisition has become increasingly common. However, annotating such data remains time-consuming and costly[4], leading to a scarcity of samples for specific categories or scenarios. Few-shot point cloud segmentation enables rapid adaptation to new environments and categories while improving model generalization and practicality through effective segmentation techniques, without relying on large annotated datasets. This progress further accelerates research and applications of point cloud segmentation in remote sensing. Therefore, achieving efficient and accurate point cloud segmentation under few-shot conditions has become a crucial topic in remote sensing research.

Recent advances few-shot point cloud semantic segmentation (FS-PCS) have achieved notable progress. Current frameworks primarily aim to address key challenges to enhance segmentation performance. To mitigate issues such as foreground leakage and sparse point distribution, researchers have proposed a standardized FS-PCS setup[5] and established a corresponding benchmark. Additionally, a correlation optimization segmentation model has been developed, representing the relationship between query points and support set prototypes by computing multi-prototypical correlations for each category. Furthermore, a hyper-correlation augmentation module has been introduced to enhance the effectiveness of these multi-prototypical correlations.

Despite achieving promising results, current few-shot point cloud segmentation frameworks[5] exhibit notable limitations. Direct correlation computation between support and query set prototypes often leads to misclassification among structurally similar categories. Moreover, these methods demonstrate reduced accuracy when segmenting small objects due to limited data availability. Existing approaches fail to fully leverage semantic primitive information during feature extraction, resulting in insufficient model robustness across diverse object categories.

In this paper, we propose a novel optimization frame work for few-shot point cloud semantic segmentation, termed SpSeg. SpSeg enhances few-shot segmentation models by integrating meta-learning methods with a superpoint construction framework that leverages semantic primitives. It capitalizes on the property that, as the point set scale increases, the semantic primitives of specific categories become progressively evident. SpSeg focuses on extracting this semantic primitive information, which is often overlooked in prior methods. Specifically, SpSeg first constructs initial

This work is supported by the National Key Research and Development Program of China (2022YFB2602105), the Key Technological Innovation Program of Ningbo City under Grant No. 2024Z297, and the Technical Service Project for Real-Scene 3D Guangxi Construction under Grant No. KWBD5C2024025.
* Corresponding author.

superpoints based on fundamental point cloud features. During the expansion of these superpoints, it utilizes the progressively revealed semantic primitives to achieve more accurate superpoint classifications. These classifications are then employed for pre-classifying the original points, effectively compensating for the neglect of semantic primitive information in mainstream frameworks, which often directly compute prototype correlations between support and query sets. By emphasizing semantic primitives and superpoint classification, SpSeg significantly improves the accuracy of few-shot point cloud semantic segmentation. Experimental results on the S3DIS dataset demonstrate that SpSeg achieves substantial performance gains over baseline methods.

Our contributions are summarized in the following:

- We introduce SpSeg, a novel framework for few-shot point cloud semantic segmentation that enhances performance under limited data conditions..
- We develop a semantic primitive-based superpoint construction module that effectively leverages previously overlooked semantic information.
- Extensive experiments on S3DIS demonstrate that SpSeg achieves state-of-the-art performance, substantially outperforming existing methods.

## II. METHODOLOGY

### A. Overview

We propose a novel method, SpSeg, which integrates a meta-learning framework with superpoint construction based on semantic primitives to address challenges in few-shot point cloud semantic segmentation.

Figure 1 illustrates the primary framework of the proposed SpSeg method. SpSeg consists of two key modules: 1) A multi-prototype correlation module, which models the relationship between query points and support set category prototypes. This is enhanced by a super-correlation module that improves the effectiveness of multi-prototype correlation computation. 2) A semantic primitive-based superpoint construction and expansion module. Initially, this module constructs superpoints using fundamental information from the point cloud. High-dimensional features are subsequently extracted to iteratively cluster these superpoints, and further refinement is achieved by incorporating semantic primitive information obtained during the clustering process. This iterative process improves segmentation accuracy, especially for small and structurally similar objects, and enhances the overall performance of the segmentation task.

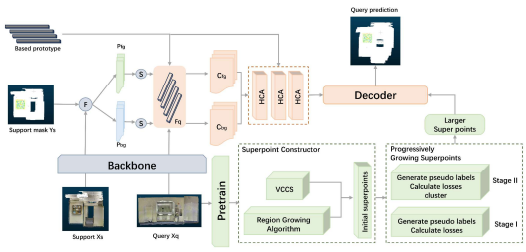


Fig. 1. The main framework of SpSeg.

### B. Multi-Prototypical Correlation

In this section, the model computes cosine similarity between query points and foreground and background prototypes in the support set, generating the multi-prototype correlation between the query and support sets. Figure 2 presents the framework of the multi-prototype correlation computation stage.

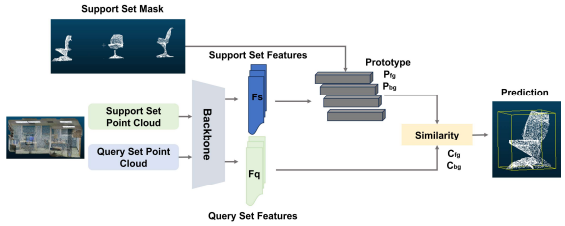


Fig. 2. Multi-prototype correlation calculation framework.

$F_s$ and $F_q$ represent the feature representations of the support and query sets, respectively. $L_s$ corresponds to the xyz coordinates of support points from $X_s$, while $\tilde{Y}_s$ denotes the inverse mask of $Y_s$. $F_{fps}$ refers to the farthest point sampling operation, primarily used to compute the foreground prototype $P_{fg}$ and background prototype $P_{bg}$ for the support set.

$$P_{fg} = F_{clus}(F_s \cdot Y_s,\ S_{fg}),\ S_{fg} = F_{fps}(L_s \cdot Y_s) \quad (1)$$

$$P_{bg} = F_{clus}(F_s \cdot \tilde{Y_s},\ S_{bg}),\ S_{bg} = F_{fps}(L_s \cdot \tilde{Y_s}) \quad (2)$$

The cosine similarity between the features of the query points and the foreground prototype $P_{fg}$ and background prototype $P_{bg}$ is computed, resulting in the relevance matrices $C_{fg} \in R^{N_Q \times N_O}$ and $C_{bg} \in R^{N_Q \times N_O}$.

$$C_{fg} = \frac{F_q \cdot P_{fg}^T}{\|F_q\|\|P_{fg}^T\|}, C_{bg} = \frac{F_q \cdot P_{bg}^T}{\|F_q\|\|P_{bg}^T\|} \quad (3)$$

### C. Semantic Prototype-Based Superpoint Construction and Expansion

We propose a semantic prototype-based superpoint construction and expansion module to address the omission of semantic prototype information when calculating correlations between prototypes in the support and query sets. The module leverages the VCCS (Voxel-based Clustered Superpoint) method along with region-growing algorithms to construct initial superpoints. In the superpoint expansion phase, high-dimensional feature calculations are performed, integrating these features with semantic prototype information. This integration enhances the clustering and expansion of superpoints, resulting in refined classification and more accurate pseudo-labels for each point. Figure 3 presents the framework of the superpoint construction and expansion module based on semantic primitives. This approach improves the model's ability to differentiate between categories, especially in the presence of structural similarities, and facilitates better handling of small objects in the segmentation task.

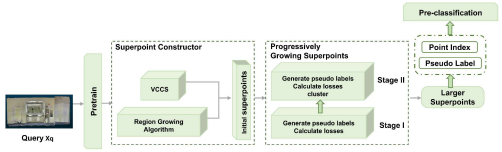


Fig. 3. Framework diagram of superpoint construction and expansion based on semantic primitives.

*1) Initial Superpoint Construction*

In the experiments conducted on the S3DIS dataset, the input point cloud is voxelized into a 5×5×5 cm grid[6]. A set of seed points is uniformly distributed at 50 cm intervals within the voxelized cloud. For each seed point, a spherical region with a 50 cm radius is defined, with the seed point as the center. Its 27 neighbors are then searched, and the distance between each neighbor and the center is calculated as follows:

$$D = \sqrt{\omega_c D_c^2 + \frac{\omega_s D_s}{3R_{seed}^2} + \omega_n D_n} \tag{4}$$

Where Dc，Ds and Dn represent the Euclidean distances in terms of color, spatial position, and normal vector, respectively. The point with the smallest distance is assigned to the superpoint associated with the current center. The newly added point becomes the new center, incorporating additional overlapping points until the spherical boundary is satisfied. In the experiments, the weights $\omega_c$， $\omega_s$ and $\omega_n$ are set to 0.2, 0.4, and 1, respectively. The 50 cm interval controls the size of the overlapping points.

Simultaneously, the region growing algorithm[7] is executed. This algorithm merges points with similar local smoothness, forming clusters of points belonging to the same smooth surface. Smoothness is assessed based on the similarity of point normals. Additionally, the algorithm initiates growth from points with the smallest curvature values, typically found in flat areas, effectively guiding region expansion.

The input point cloud data is initially segmented into multiple regions using supervoxel clustering (VCCS) and region growing algorithms. These regions are subsequently merged to enhance coherence. Specifically, for each region generated by supervoxel clustering, if at least half of its points are covered by a region from the region growing algorithm, all points from the former are merged into the latter. This merging process leads to larger, more coherent regions, which improves the quality of the segmentation and contributes to more accurate semantic predictions.

*2) Superpoint Expansion*

During the superpoint expansion phase, the model employs the K-means[8] algorithm to extract high-dimensional features. This phase consists of two parts: in the first part, the superpoints do not expand, and the model continuously applies the K-means algorithm to extract features, calculate pseudo-labels, and evaluate loss. In the second part, the model continues extracting features, calculating pseudo-labels, and evaluating loss through K-means, while introducing new clustering operations to expand the superpoints and update the pseudo-labels. Through iterative calculations in both phases, superpoints gradually expand, leading to more accurate classification results.

This part takes the point cloud data $P^h$ and its corresponding neural features$F_h \in R^{N\times K}$, where N represents the number of points and K denotes the feature dimension for each point. Additionally, the initial superpoints are provided, represented by the set $\{P_1^{\sim h} \cdots P_{m^0}^{\sim h} \cdots P_{M^0}^{\sim h}\}$. First, we will compute the average neural features of the initial superpoints, denoted as $\{f_1^{\sim h} \cdots f_{m^0}^{\sim h} \cdots f_{M^0}^{\sim h}\}$:

$$f_{m^0}^{\sim h} = \frac{1}{Q}\sum_{q=1}^{Q} f_q^h, \quad f_q^h \in R^{1\times K} \tag{5}$$

Q is the total number of points in the initial superpoint $P_{m^0}^{\sim h}$, and $f_q^h$ is the feature vector of the q[th] point extracted from $F_h$. After obtaining these initial superpoint features, we use the K-means algorithm to group the $M^0$ feature vectors into $M^1$ clusters, where $M^1 < M^0$. Each cluster represents a new, larger superpoint. Ultimately, we obtain the new superpoints:

$$\{P_1^{\sim h} \cdots P_{m^1}^{\sim h} \cdots P_{M^1}^{\sim h}\} \xleftarrow{Kmeans} \{f_1^{\sim h} \cdots f_{m^0}^{\sim h} \cdots f_{M^0}^{\sim h}\} \tag{6}$$

The superpoint expansion step is executed independently on each input point cloud. A smaller $M^1$ results in a more aggressive expansion process.

## III. EXPERIMENTS

### A. Experimental Data and Evaluation Metrics

To evaluate the effectiveness of the proposed few-shot point cloud semantic segmentation method, we use the Stanford 3D Indoor Space Dataset (S3DIS)[9] for experiments. S3DIS is a well-known dataset for indoor scene understanding and 3D semantic segmentation, featuring rich diversity and complexity. It consists of six regions (Region 1 to Region 6), thirteen semantic categories (e.g., ceiling, floor, wall), and eleven scene types (e.g., office, conference room, corridor). We adopt the Mean Intersection over Union (mIoU) as the evaluation metric to assess segmentation performance. mIoU measures model accuracy by computing the overlap between predicted and ground truth results for each category, with higher mIoU values indicating better performance.

### B. Experimental Details

The SpSeg framework is implemented in PyTorch, with all experiments conducted on a single NVIDIA A6000 GPU. The framework uses the first three modules of the Stratified Transformer[10] as the backbone network. Superpoint construction employs the SparseConv[11] architecture, with the encoder based on Res16 and the decoder consisting of four MLP layers, producing 128-dimensional point-wise features. Training is carried out using the AdamW[12] optimizer, with a learning rate of 0.00005 and a weight decay of 0.01.

### C. Quantitative Analysis

The comparison results of our method with four previous approaches on eight benchmarks are shown in Tables 1 and 2. Table 1 presents the model performance on the S3DIS dataset with cvfold set to 0, while Table 2 shows performance with cvfold set to 1.

When cvfold is 0, SpSeg significantly outperforms COSeg[5], with improvements ranging from 2.25% to 4.51%. When cvfold is set to 1, the improvement ranges from 1.22% to 2.98%. SpSeg achieves state-of-the-art results across all eight benchmarks. Additionally, the model demonstrates significant accuracy improvement when the cvfold is 0, owing to the smaller volume and higher structural similarity of categories in the test set. This observation highlights the advantage of SpSeg in effectively handling few-shot classification challenges, particularly in scenarios involving small objects and structurally similar categories. The results underscore the model's ability to maintain high performance under such conditions, confirming its robustness in dealing with challenging segmentation tasks.

TABLE I. RESULTS (%) ON DIFFERENT PROTOTYPES（$S^0$）

| | Method | | | |
|---|---|---|---|---|
| | *AttMPTI[13] (CVPR2021)* | *QGPA[14] (TIP2023)* | *COSeg[5] (CVPR2024)* | ***SpSeg Ours*** |
| 1-way 1-shot | 36.32 | 35.50 | 46.31 | **48.56** |
| 1-way 5-shot | 46.71 | 38.07 | 51.40 | **54.46** |
| 2-way 1-shot | 31.09 | 25.52 | 37.44 | **40.12** |
| 2-way 5-shot | 39.53 | 30.22 | 42.27 | **46.78** |

TABLE II. RESULTS (%) ON DIFFERENT PROTOTYPES（$S^1$）

| | Method | | | |
|---|---|---|---|---|
| | *AttMPTI[13] (CVPR2021)* | *QGPA[14] (TIP2023)* | *COSeg[5] (CVPR2024)* | ***SpSeg (Ours)*** |
| 1-way 1-shot | 38.36 | 35.83 | 48.10 | **49.43** |
| 1-way 5-shot | 42.70 | 39.70 | 48.68 | **51.66** |
| 2-way 1-shot | 29.62 | 26.26 | 36.45 | **37.67** |
| 2-way 5-shot | 32.62 | 32.41 | 38.45 | **40.45** |

### *D. Visual Analysis*

Figure 4 presents visualizations of the generated initial superpoints. (a), (b), and (c) correspond to one region, while (d), (e), and (f) depict another. Specifically, (b) and (e) visualize the original labels, whereas (c) and (f) show pseudo-labels of the initial superpoints. The results indicate that while the initial superpoints provide a coarse segmentation, they lack fine-grained precision, necessitating further refinement in the superpoint expansion module.

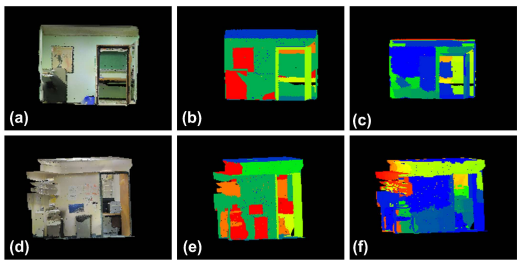


Fig. 4. Visualization of Initial Superpoints.

Figure 5 presents the final visualization results of the model. In (a), (b), (c), and (d), the red, yellow, green, and orange regions indicate the areas to be predicted, while (e), (f), (g), and (h) show the corresponding predictions. The results indicate that the model achieves satisfactory segmentation performance for most objects.

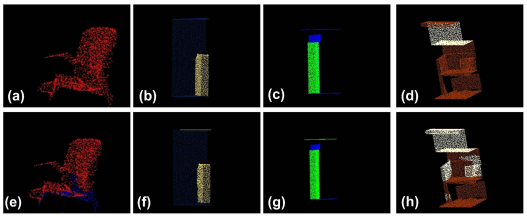


Fig. 5. Visualization of Final Segmentation Results.

## IV. CONCLUSION

In this paper, we proposed SpSeg, a novel few-shot point cloud segmentation framework that integrates semantic primitive-based superpoint construction and pre-classification to enhance segmentation accuracy. SpSeg addresses challenges in segmenting small objects and structurally similar categories by iteratively refining superpoints and leveraging semantic information. Our method significantly outperforms existing approaches on the S3DIS dataset, achieving state-of-the-art performance across eight different settings. While SpSeg improves segmentation robustness, challenges remain in differentiating highly similar categories due to dataset limitations. Future work will explore diverse datasets and adaptive learning mechanisms to further enhance few-shot segmentation in real-world applications.

## REFERENCES

[1] Charles R. Qi, Hao Su, Kaichun Mo and Leonidas J Guibas, Pointnet: Deep learning on point sets for 3d classification and segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 652-660.

[2] Charles R. Qi, Li Yi, Hao Su and Leonidas J Guibas, Pointnet++: Deep hierarchical feature learning on point sets in a metric space[J]. Advances in neural information processing systems, 2017, 30.

[3] Yangyan Li, Rui Bu, Mingchao Sun, et al. Pointcnn: Convolution on x-transformed points[J]. Advances in neural information processing systems, 2018, 31.

[4] Jens Behley, Martin Garbade, Andres Milioto, Jan Quen-zel, Sven Behnke, Cyrill Stachniss, and Jurgen Gall, Semantickitti: A dataset for semantic scene understanding of lidar sequences. In Proceedings of the IEEE/CVF international conference on computer vision, pages 9297–9307, 2019.

[5] Z. An, G. Sun, Y. Liu, F. Liu, Z. Wu, et al. Rethinking few-shot 3D point cloud semantic segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 3996–4006.

[6] Jeremie Papon, Alexey Abramov, Markus Schoeler, and Florentin Worgotter, Voxel Cloud Connectivity Segmentation Supervoxels for Point Clouds. CVPR, 2013. 3, 5, 12.

[7] R. Adams and L. Bischof, Seeded Region Growing. TPAMI, 1994. 3, 5, 12.

[8] S. Huang, Z. Kang, Z. Xu, Deep k-means: a simple and effective method for data clustering, in: Proceedings of the International Conference on Neural Computing for Advanced Applications. Springer, 2020, pp. 272-283.

[9] Iro Armeni, Sasha Sax, Amir R. Zamir and Silvio Savarese, Joint 2D-3D-Semantic Data for Indoor Scene Understand ing. arXiv:1702.01105, 2017. 1, 2, 5, 6, 7, 15.

[10] Xin Lai, Jianhui Liu, Li Jiang, Liwei Wang, Hengshuang Zhao, Shu Liu, Xiaojuan Qi, and Jiaya Jia, Stratified transformer for 3d point cloud segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 8500–8509, 2022. 2, 3, 4, 6, 11.

[11] Benjamin Graham, Martin Engelcke, and Laurens van der Maaten, 3D Semantic Segmentation with Submanifold Sparse Convolutional Networks. CVPR, 2018. 1, 2, 3, 5,6, 7, 12, 13, 14, 15.

[12] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization", Proc. Int. Conf. Learn. Representations, 2018.

[13] Na Zhao, Tat-Seng Chua, and Gim Hee Lee, Few-shot 3d point cloud semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 8873–8882, 2021. 1, 2, 3, 4, 5, 6, 7, 11.

[14] Shuting He, Xudong Jiang, Wei Jiang, and Henghui Ding, Prototype adaption and projection for few-and zero-shot 3d point cloud semantic segmentation. IEEE Transactions on Image Processing, 2023. 1, 2, 3, 4, 5, 6, 7, 11.