Robust Network For Segmenting Roof Planes From Sparse Point Clouds

Hang Fu[1,2],Cheng Zeng[1,3*]
[1]Key Laboratory of Intelligent Sensing System and Security, Hubei University, Ministry of Education, Wuhan, China
[2]School of Computer and Information Engineering, Hubei University, Wuhan, China
[3]School of Artificial Intelligence, Hubei University, Wuhan , China;
fuhang@stu.hubu.edu.cn, *zc@hubu.edu.cn

***Abstract*—The plane is the fundamental characteristic for describing the shape of polyhedral buildings. Roof plane segmentation of airborne LiDAR point clouds is an important step in 3D building model reconstruction. Existing methods are either non-automatic and efficient traditional methods, or hindered by the lack of datasets for roof plane segmentation and therefore fail to achieve good training for deep learning methods. To address the above issues, we proposes an end-to-end network called PUTr for roof plane segmentation of sparse point clouds. This network integrates a Point-cloud Upsampling module and Transformer to ensure effective generalization even when trained on low-density datasets, while also capturing dependency relationships within point cloud neighborhoods. Our model overcomes the low automation and manual intervention of existing methods, and the limited generalization of deep learning due to scarce high-quality datasets. Evaluations also demonstrate its effectiveness in roof plane segmentation of laser point cloud data.**

***Index Terms*—roof plane segmentation, building plane**

I. Introduction

Roof plane segmentation is inherently complex due to the irregular and unordered nature of point clouds, which lack any connectivity information and do not provide semantic features at the lower level [1-3]. Existing roof plane segmentation methods still primarily rely on traditional techniques such as region growing [4-5]. Some researchers have improved seed point selection or enhanced plane extraction efficiency. However, while region growing algorithms are robust against noise and irregular point cloud regions, they can lead to over-segmentation or under-segmentation of roof plane boundaries. Similarly, model fitting techniques [6-7] have been refined to mitigate issues like excessive false planes and long processing times, yet these methods still struggle with point clouds characterized by high noise levels and irregular roof shapes. Additionally, Li Li et al. [8] applied hierarchical clustering to iteratively merge adjacent planes and then refine them to extract high-quality roof planes. However, this clustering-based approach is prone to erroneous clustering in uneven or noisy point clouds and is not suitable for large datasets [9-10]. In summary, the use of these traditional methods is increasingly constrained, and they seem outdated in the current landscape dominated by deep learning.

With the advancement of deep learning in point cloud processing, deep learning models have gradually been applied to tasks such as point cloud classification, segmentation, and 3D reconstruction, yielding promising results in recent years [11] . The deep learning-based plane segmentation network RoofNet has been proposed [12]. However, it is worth noting that compared to modern deep learning models like Transformers, RoofNet has a relatively simple structure and limited representation learning capabilities. In addition, the scarcity of sufficient datasets for roof plane segmentation in point clouds makes it challenging to find high-quality data, significantly hindering progress [13]. Although RoofN3D [14] provides segmentation of individual roofs, its average point density is only around 5 points/square meter, which is much lower than the standard point density captured by mainstream LiDAR scanning devices (10-12 points/square meter). This limitation leads researchers to rely on non-standard public datasets or self-generated datasets for training, which undermines the persuasiveness of their segmentation results and hinders their practical applications.

DOI reference number: 10.18293/SEKE2024-073

To address the aforementioned challenges, we propose an end-to-end roof plane segmentation network that integrates a Point-cloud Upsampling module and Transformer (PUTr). The primary contributions of this research can be summarized as follows:

1) Through the specialized upsampling module introduced in our proposed PUTr model, we have conducted reasonable adaptive upsampling on RoofN3D, the currently available public dataset for roof plane segmentation, to achieve standard point density. This enables the network models trained using the RoofN3D dataset to possess a certain level of generalization and improves the situation of the lack of available public datasets in the field of plane segmentation.

2) An improved end-to-end roof plane segmentation network called PUTr is proposed by us. By incorporating a Transformer with channel-wise attention into roof plane segmentation, we enhance the ability to capture crucial features, thereby improving the accuracy of roof plane segmentation.

II. Approach

This research can be considered the first application of Transformers to point cloud roof plane segmentation, augmented with a adaptive upsampling module, enabling efficient processing and generalization of roof plane segmentation. Firstly, we introduce the upsampling module and Transformer module in our network. Finally, we present the overall architecture of PUTr for handling the task of roof plane segmentation, as illustrated in Figure 1.

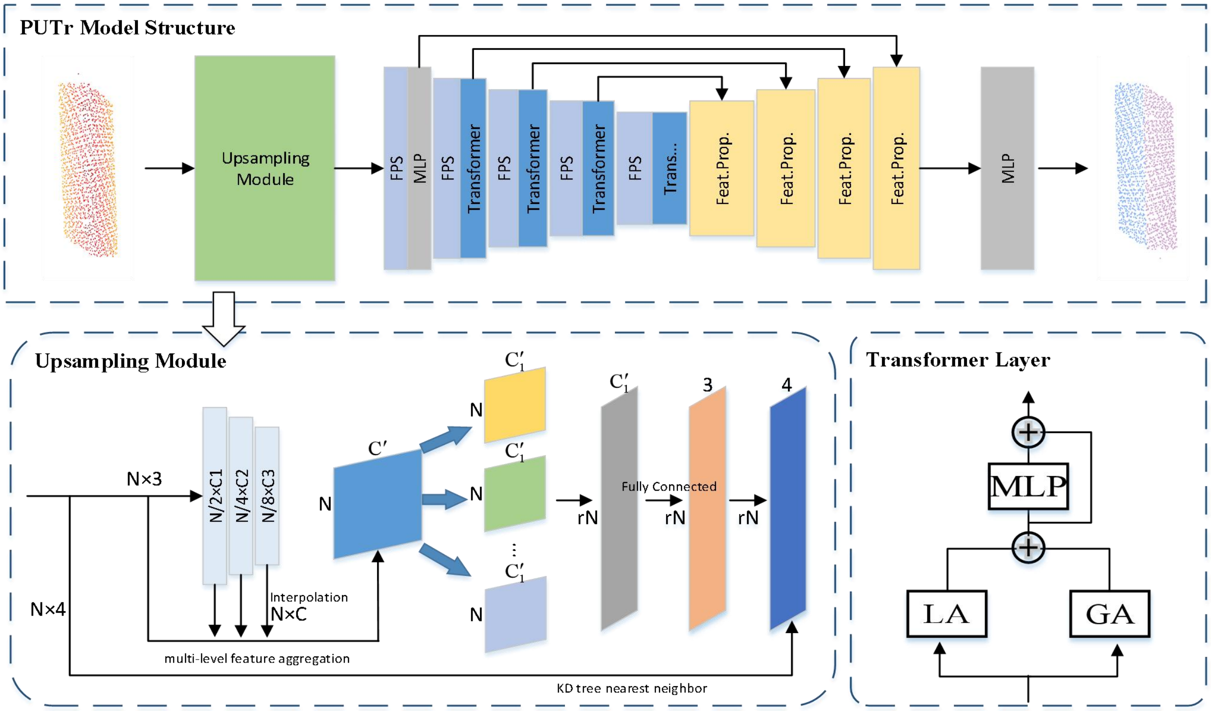

Figure 1. The Overall Architecture Of PUTr

A. Upsampling module

In the field of computer vision, researchers have utilized deep learning-based super-resolution reconstruction algorithms to restore low-resolution images to their corresponding high-resolution counterparts, addressing issues such as image blur and low quality in image acquisition.

Our upsampling module is influenced by [15] and improved upon accordingly.Instead of opting for a fixed upsampling rate, we attempted adaptive upsampling of the RoofN3D dataset. This approach allows for adaptive upsampling based on the current input of the original point cloud, enabling it to approximate standard point cloud density. Consequently, the trained point cloud roof plane segmentation model exhibits generality.The upsampling module takes a low-density $N \times 3$ point set as input. Initially, it employs the group feature extraction proposed by PointNet++ to obtain three-dimensional features, denoted as $N/2 \times C1$, $N/4 \times C2$, and $N/8 \times C3$. Subsequently, the interpolation method from PointNet++ is used to restore the three multidimensional features to the original point's feature dimension and aggregate them, resulting in a feature dimension of $N \times C^{'}$. Then, the feature expansion is performed on f using separable convolution. Since points and features are interchangeable, the number of expanded features in the expansion space is equivalent to the number of points, resulting in the conversion of the expanded tensor $N \times rC_1^{'}$ to a new feature $f^{'}$ of size $rN \times C_1^{'}$, where r represents the adaptively learned upsampling rate based on the original point cloud density. Finally, a uniform and dense point cloud of size $rN \times 3$ is obtained through fully connected layers.

Then, we incorporate KD-tree-based neighborhood search during upsampling, along with the original point cloud's label data. This approach allows us to compute the magnitude of each class label within the neighborhood and determine the label of the current augmented point based on the relative quantities of each class label.Examples of the results after applying the upsampling module can be seen in Figure 2.

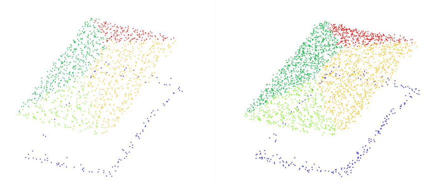

Figure 2. Four-slope roof with upsampling module results.

B. Transformer Layer

Due to the inherently irregular nature of point clouds, which are embedded in metric space, attention mechanisms are particularly suitable for processing point clouds. To ensure different attention weights between queries and key points for each channel, we utilize the Channel Attention (CA) method to build local attention and global attention, forming Transformer layers. This allows the model to capture not only short-distance local information but also long-distance global information. The formula for Channel Attention (CA) is as follows:

$$\mathbb{A}_{i,c} = \frac{\exp(\mathcal{M}'([\mathcal{R}'(f_1, f_2); \phi(x_i)]/\tau)_c)}{\sum \exp(\mathcal{M}'([\mathcal{R}'(f_1, f_2); \phi(x_i)]/\tau)_c)} \tag{1}$$

$$CA(x_i, f_i) = \sum \mathbb{A}_{i,c} \odot \mathcal{M}([\mathcal{R}(f_1, f_2); \phi(x_i)]) \tag{2}$$

Where x_i and f_i represent the position coordinates and corresponding features of the query point. $\phi(x_i)$ normalizes the position coordinates, and f_1, f_2 undergo linear transformations on the features. $\mathcal{R}$ represents the relationship function between the two transformed features (e.g., $f_1 - f_2$), and $\mathcal{M}$ denotes the mapping function. $\mathbb{A}_{i,c}$ is the channel-wise attention, where c represents the channel index, and τ denotes the temperature coefficient.

Local Attention We utilize spherical queries to capture local neighborhoods, where local point groups are formed by anchor points and points queried within the spherical neighborhood. Local attention can be defined as:

$$y_i^{LA} = CA\left(x_i, f_i, \{x_j\}_{j\in\mathcal{G}_i}, \{f_j\}_{j\in\mathcal{G}_i}\right) \quad (3)$$

Where $\mathcal{G}_i$ represents the local group captured by anchor point x_i. y_i^{LA} is the output feature vector of local attention, similarly employing channel-wise attention.

Global Attention We also consider utilizing attention mechanisms to capture global features. However, directly applying attention mechanisms to the original points without any processing would incur a significant computational cost, affecting the efficiency of the network. Therefore, we first perform uniform downsampling on the point cloud, representing global information with fewer points without compromising performance. Global attention can be defined as:

$$y_i^{GA} = CA\left(x_i, f_i, \{\delta_s\}_{s=1}^{S}, \{\eta_s\}_{s=1}^{S}\right) \quad (4)$$

Where S represents the set of sampled points, δ_s and η_s denote the position information and corresponding features of the current sampled point, and y_i^{GA} is the output global attention feature vector.

C. *Network architectures*

The entire network consists of two components: an upsampling module and a backbone network based on Transformer blocks. To address the issue of low point density in the RoofN3D dataset and to enhance the model's universality, we preprocess the dataset using the upsampling module to obtain a dataset with standard point density. Subsequently, the data is fed into the backbone network based on the U-Net design. The encoder mainly consists of Transformer blocks. To enhance representational capacity, the input data is first processed through an MLP. The entire process undergoes five stages of downsampling, where furthest point sampling (FPS) is utilized to obtain sampled points, which are then fed into MLP or Transformer blocks. The decoder is coupled with the symmetric encoder, combining interpolated point features with the corresponding level encoder features provided by skip connections. This combined data undergoes processing through linear layers, batch normalization, and ReLU activation. Finally, through trilinear interpolation, the low-level features are mapped to the high-level features for feature expansion.

Finally, to enhance the model's ability to capture representations, we first apply max pooling or average pooling to the low-level features learned by the backbone network, allowing for better capture of global context information. Then, we use an MLP to map the learned features to the probability distribution of each point belonging to a specific plane, thereby achieving end-to-end roof plane segmentation directly.

III. Experiments and Analysis

A. *Experimental environment and data*

The experiments were conducted on the RoofN3D dataset. We filtered the dataset to include only roof buildings with more than 700 points, covering 11,180 roofs. With sufficient training data, we ensured that the shape details of the roof buildings were preserved. Finally, the training, validation, and test sets consisted of 9,503, 1,118, and 559 samples, respectively. The model was trained using a smoothed cross-entropy loss function. For evaluation metrics, we used the mean Intersection over Union (mIoU) and Overall Point-wise Accuracy (OA).

B. *Experimental Result*

The network is evaluated on the RoofN3D dataset for the task of building roof plane segmentation. Since there are relatively few network models specifically designed for roof plane segmentation based on point cloud deep learning, we consider adaptively improving network models for other point cloud segmentation tasks and conduct comparative experiments. The results of the comparative experiments are as follows.

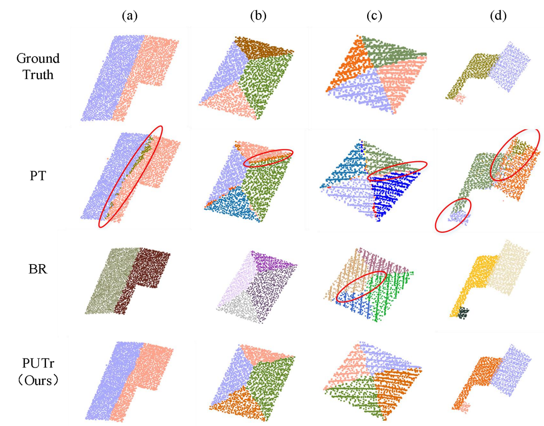

Figure 3. Visual comparison of PointTransformers (PT), Relabeling-based (BR), and the proposed method (PUTr) in the task of roof plane segmentation.

TABLE I. The results of roof plane segmentation on the RoofN3D dataset

Method	OA(%)	mIoU(%)
Convpoint[16]	86.9	80.3
PT	89.2	83.5
BR	91.4	83.2
PUTr（ours）	**93.8**	**86.9**

Figure 3 presents the visual comparison results between ground truth and those generated by different methods. Due to the presence of a small number of unfiltered background points in the dataset, we filtered out noise to better demonstrate the segmentation results. We observed that directly applying methods designed for other tasks to roof plane segmentation, such as PT, does not yield satisfactory segmentation results. There tends to be a certain degree of over-segmentation and under-segmentation, particularly concentrated at boundary intersections. Meanwhile, the traditional method BR [8] exhibits suboptimal performance in handling pyramid roofs, and furthermore, when dealing with large sample datasets, manual parameter adjustments are frequently required, leading to reduced efficiency. Additionally, Table I presents the quantitative results of different methods for roof plane segmentation in point cloud buildings. Our proposed method, which integrates advanced network architecture with unique modules, demonstrates superior representation and generalization capabilities. It achieves strong segmentation results on the RoofN3D dataset, with an OA of 93.8% and an mIoU of 86.9%, outperforming other methods.

C. *Ablation Studies*

We have conducted some ablation experiments to examine specific design decisions in PUTr and to facilitate further discussion. These studies are still based on the RoofN3D dataset for roof plane segmentation.

TABLE II. ABLATION EXPERIMENT: WITH/WITHOUT UPSAMPLING MODULE AND UPSAMPLING RATE R

Upsampling Module	Upsampling Rate(r)	OA(%)	mIoU(%)
Without	-	90.1	84.7
	2	91.9	85.5
With	3	94.2±0.2	89.6
	self-adaptation	**94.4±0.1**	**89.6**

Upsampling Module. First, we investigated the choice of the upsampling module. To validate the impact of the upsampling module on the generality of our network, we selected 207 high-density roofs with more than 2048 points from RoofN3D as test samples. The results are shown in Table Ⅱ. It can be observed that the performance improves when the upsampling module is used compared to when it is not, indicating that the model's generalization ability is enhanced. This enables the model to effectively handle modern high-density point cloud data. When we replace the fixed upsampling rate with an adaptive upsampling rate tailored to the current input sample, we effectively avoid issues associated with low-density training samples (r=2) and unnecessary upsampling (r=3). This enhances our model's versatility and efficiency, allowing it to effectively handle high-density point clouds captured by modern scanning devices.

TABLE III. ABLATION EXPERIMENT: ATTENTION TYPES

Attention Types	OA(%)	mIoU(%)
MLP	82.1	76.8
MLP+Pooling	87.2	80.6
Standard Att.	89.2	83.5
Ours	**93.8**	**86.9**

Attention Types. Now, we investigate the type of attention mechanism in the PUTr network. The results are shown in Table Ⅲ, where we compare four different scenarios. Compared to MLP and the standard attention mechanism, we can observe that our channel attention is more expressive. This indicates that, compared to MLP or standard attention mechanisms, channel-wise attention is more powerful in representing point clouds.

IV. CONCLUSION

In this paper, we propose PUTr, an end-to-end network for handling sparse point cloud roof plane segmentation, which solves the problems of low automation and reliance on manual intervention of existing roof plane segmentation methods, and the lack of generality of deep learn-based methods due to scarcity of data sets, showing versatility. In the future, we will focus on simplifying complex roofs by decomposing them into simpler components, aiming to enhance generalization capabilities when handling intricate roof structures. Additionally, we hope for the availability of more high-quality, multi-class roof building datasets in the future, which will enable us to further improve our method and enhance its capability to handle segmentation of complex roofs with multiple categories.

ACKNOWLEDGMENT

This research was funded by The Major Program(JD) of Hubei Province (2023BAA018).

REFERENCES

[1] Wang, X. and S. Ji. "Roof plane segmentation from lidar point cloud data using region expansion based l 0 gradient minimization and graph cut." IEEE Journal of Selected Topics in Applied Earth Observations Remote Sensing 14 (2021): 10101-16.

[2] Liu, K., H. Ma, L. Zhang, X. Liang, D. Chen and Y. Liu. "Roof segmentation from airborne lidar using octree-based hybrid region growing and boundary neighborhood verification voting." IEEE Journal of Selected Topics in Applied Earth Observations Remote Sensing 16 (2023): 2134-46.

[3] Gilani, S. A. N., M. Awrangjeb and G. Lu. "Segmentation of airborne point cloud data for automatic building roof extraction." GIScience remote sensing 55 (2018): 63-89.

[4] Jingya, X., L. Li, G. Ye, W. Jing and Y. Jian. "A method for roof plane segmentation by supervoxel and regional growth(in chinese)." Journal of Geomatics (2021):

[5] Vo, A.-V., L. Truong-Hong, D. F. Laefer and M. Bertolotto. "Octree-based region growing for point cloud segmentation." ISPRS journal of photogrammetry remote sensing 104 (2015): 88-100.

[6] Xu, J. and J. Li. "Optimal ransac method for segmentation of complex building roof planes." Geomatics Information Science of Wuhan University 48 (2023): 1531-37.

[7] Tarsha-Kurdi, F., T. Landes and P. Grussenmeyer. "Hough-transform and extended ransac algorithms for automatic detection of 3d building roof planes from lidar data." Presented at ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007, 2007. 36, 407-12.

[8] Li, L., J. Yao, J. Tu, X. Liu, Y. Li and L. Guo. "Roof plane segmentation from airborne lidar data using hierarchical clustering and boundary relabeling." Remote Sensing 12 (2020): 1363.

[9] Chen, H., M. Liang, W. Liu, W. Wang and P. X. Liu. "An approach to boundary detection for 3d point clouds based on dbscan clustering." Pattern Recognition 124 (2022): 108431.

[10] Chen, H., T. Xie, M. Liang, W. Liu and P. X. Liu. "A local tangent plane distance-based approach to 3d point cloud segmentation via clustering." Pattern Recognition 137 (2023): 109307.

[11] Park, J., S. Lee, S. Kim, Y. Xiong and H. J. Kim. "Self-positioning point-based transformer for point cloud understanding." Presented at Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023. 21814-23.

[12] Zhang, C. and H. Fan. "An improved multi-task pointwise network for segmentation of building roofs in airborne laser scanning point clouds." The Photogrammetric Record 37 (2022): 260-84.

[13] Kölle, M., D. Laupheimer, S. Schmohl, N. Haala, F. Rottensteiner, J. D. Wegner and H. Ledoux. "The hessigheim 3d (h3d) benchmark on semantic segmentation of high-resolution 3d point clouds and textured meshes from uav lidar and multi-view-stereo." ISPRS Open Journal of Photogrammetry Remote Sensing 1 (2021): 100001.

[14] Wichmann, A., A. Agoub, M. J. T. I. A. o. t. P. Kada, Remote Sensing and S. I. Sciences. "Roofn3d: Deep learning training data for 3d building reconstruction." 42 (2018): 1191-98.

[15] Yu, L., X. Li, C.-W. Fu, D. Cohen-Or and P.-A. Heng. "Pu-net: Point cloud upsampling network." Presented at Proceedings of the IEEE conference on computer vision and pattern recognition, 2018. 2790-99.

[16] Boulch, A. "Convpoint: Continuous convolutions for point cloud processing." Computers Graphics 88 (2020): 24-34