Point Cloud-Based Segmentation of Small Roof Components in Chinese Ancient Architecture

Jianghong Zhao, Xueqing Zhang, Ziyu Liu, Jia Yang, Haiquan Yu

School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing,102616, China zhaojiangh@bucea.edu.cn, 15810486358@163.com, 956828753@qq.com, 5091112@qq.com, 2321483450@qq.com
Key Laboratory for Architectural Heritage Fine Reconstruction & Health Monitoring, Beijing, China

Keywords: Ancient architecture, Architectural cultural heritage, Semantic segmentation of point clouds, Attention mechanism

Abstract

The roofs of ancient Chinese buildings are rich in cultural symbolism, embodying profound historical and artistic significance. To preserve the structural and semantic information of these roof components, this study employs point cloud semantic segmentation, as point clouds effectively capture their authentic geometry and dimensions. To reduce the high cost of manual annotation, we propose a weakly supervised learning approach for point cloud segmentation. However, a significant challenge arises due to the overwhelming presence of roof tiles in the point cloud data, which hinders segmentation performance. Since tiles constitute the majority of the point cloud, smaller architectural components become underrepresented. As a result, when ground truth labels are assigned randomly, the number of labeled points for these smaller elements is insufficient, leading to suboptimal segmentation accuracy. To address this issue, we refine the positional encoding method based on advancements in the attention mechanism, thereby enhancing the model’s ability to focus on small-scale components. Experimental results demonstrate that our approach achieves a 38.61% improvement in mean Intersection over Union (mIoU) compared to SQN, along with a 3.36% increase in overall accuracy (OA). Notably, our method even outperforms certain fully supervised networks in segmentation effectiveness.

1. Introductory

Chinese ancient architecture has a long history of development and is imbued with unique cultural connotations, serving as an important symbol of historical heritage and spiritual cohesion. Typically, it comprises three primary components: the foundation, the structural frame, and the roof. Its most distinctive feature is the prominent large-roof structure, which sets it apart from architecture in other regions and has earned it the designation of “big-roof architecture”(Wang, 2011). The roof styles of ancient Chinese architecture are complex and diverse, primarily comprising the hip roof, gable-and-hip roof, overhanging gable roof, flush gable roof, and pyramidal roof, among others. The roof structure primarily comprises tiles, ridges, ridge beasts, chiwen, and ornamental animals. Beyond serving as primary roof components, these features fulfill both structural and decorative functions by securing roof tiles while enhancing aesthetic appeal. Roof typology served as a critical manifestation of hierarchical systems in ancient Chinese architecture, embodying the sociopolitical ideologies of successive dynasties. Over millennia of architectural evolution, traditional Chinese roof structures and their associated eave decorations have become significant cultural artifacts that preserve historical memory and aesthetic philosophy. However, prolonged exposure to meteorological factors-including solar radiation, precipitation, aeolian erosion, and cryospheric processes-coupled with natural calamities, has resulted in progressive material degradation and structural compromise of these historic roofing systems. To safeguard this architectural heritage, systematic digital documentation of historic Chinese structures has become imperative. Terrestrial Laser Scanning (TLS) technology enables the acquisition of high-fidelity geometric data of architectural elements. Subsequent application of point cloud semantic segmentation algorithms facilitates the precise identification and preservation of intricate component-level details, particularly those pertaining to roof assemblies. This methodological framework establishes a critical foundation for engineering interventions, including structural health diagnostics, material rehabilitation protocols, and digital twin development for historic conservation.

Our contributions are as follows:
Using SQN as the baseline and extending our existing network (SQN-DLA) for roof segmentation, we integrate positional information-specifically, the coordinates of the central point and its neighboring points-into the segmentation process to achieve precise results.

2. Related Work

2.1 Overview of Ancient Roofs

The roof of an ancient Chinese building comprises several components, primarily including the ridge, the roof surface, the eaves, supporting elements, and additional parts(Yuan et al., 2022). The roof ridge, as the highest part of the roof, serves load-bearing, decorative, and waterproofing functions. The main ridge connects the two roof surfaces and is adorned at both ends with decorative ridge beasts and carved cloud motifs. The pendant ridges, located at the four corners of the roof, primarily enhance structural stability and direct rainwater flow. The bump ridges, situated at the roof's turning points, act as connectors between the main ridge and the pendant ridges, reinforcing the roof against wind forces. The roof surface, which constitutes the primary envelope of the structure, provides windproofing and thermal insulation. It is covered with roof tiles and features a drainage ridge strip, with the tile color indicating the building’s intended usage. The eaves, located along the lower edge of the roof, serve to block sunlight and prevent rainwater erosion. In addition, the dougong elements within the eaves also function as load-bearing components, contributing to structural stabilization.

The various parts of the roof have different roles,to preserve the intricate details of these roof components, 3D laser scanning technology is employed to generate point clouds for digital conservation.

2.2 Point cloud semantic segmentation

Currently, with the continuous development of LiDAR technology, point cloud semantic segmentation technology has been widely used in the digital preservation of ancient buildings and disease detection, and the main methods include traditional machine learning (ML), deep learning (DL) and hybrid methods (Zhou et al., 2024). Traditional machine learning relies on manually designed geometric features with regularization algorithms, which are more suitable for segmentation of regular components. Elkhrachy (Elkhrachy ,2017) identifies boundary points by thresholds such as normal vector pinch angle, curvature, etc. Maltezos (Maltezos and Ioannidis, 2018) utilizes Hough transform and RANSAC to fit planar, cylindrical and other geometries, and Zhang Ruiju (Zhang et al., 2020) et al. segmented beams and columns a priori in conjunction with building structures. Qian et al. (Qian et al.,2024) and others iteratively merge neighboring points using normal vector or density as constraints. Wan Fei (Wan et al., 2021) aggregated similar points based on covariance matrix and Euclidean distance. Grilli (Grilli and Remondino, 2019) explored the performance of geometric covariance features under different spherical neighborhood radii. Machine learning relies on manual feature design and neighborhood knowledge, which has insufficient generalization ability for irregular components (e.g., flying eaves and arches) in ancient architecture, making it difficult to achieve automated segmentation. Deep learning solves the problem of relying on rules in traditional methods through end-to-end feature learning, and improves the segmentation ability of complex scenes. PointNet realizes the original point cloud input for the first time; Hu et al. (Hu et al., 2020) RandLA-Net adopts random sampling with local feature aggregation; Zhang et al (Zhang et al., 2021) in MSFA-Net proposed Dual Attention Aggregation Module (DAA) with edge interaction classifier. However, deep learning relies on a large amount of labeled data, and it may take weeks or even months to label a complex point cloud of ancient buildings for semantic segmentation. In order to balance accuracy and efficiency, hybrid methods combine supervised learning methods with geometric features to further improve the segmentation effect. However, fully supervised methods rely on a large amount of labeled data and have a large labeling cost, so weakly supervised learning is more suitable for segmenting point clouds of ancient buildings with a large amount of data, which can achieve close to fully supervised segmentation performance by using less labeled data, and greatly reduces the time and labor cost. In the process of point cloud segmentation of ancient buildings, due to the complex structure of ancient building roofs and the small size of the target components, the existing segmentation methods cannot efficiently and accurately segment the detailed information of ancient building roofs, so in order to take into account the purpose of high efficiency and accurate segmentation, this paper will be improved on the basis of weakly supervised learning.

3. Method

To address the difficulty of segmenting small targets on the roof and to ensure that these targets receive greater attention, we take SQN(Hu et al.,2022) as the baseline and segment the roof on the basis of our existing network (SQN-DLA). To overcome the deficiencies of the positional coding in SQN-DLA (Zhao et al.,2024) for segmenting the roofs of ancient buildings, we add the values of the centroid and the neighboring points, thereby achieving improved roof segmentation.

The overall structure of the DLA is illustrated in Fig. 1. The inputs consist of spatial information and previously learned features. After encoding the spatial information, it is combined with the learned features and passed through a self-attention block to generate local attention features. These local attention features are then concatenated with the original features to form a residual connection, which is subsequently processed by an attention pooling block to yield enhanced local attention features. Finally, these enhanced features are summed with the original features to produce the final spatial attention features. To capture richer information, our input features are encoded using features that include color information.

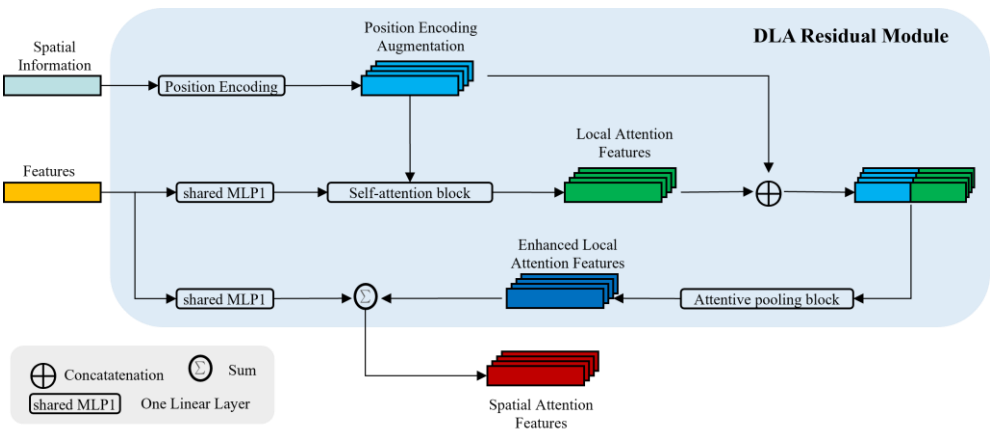

Figure 1 DLA overall framework

The DLA network comprises two components: a point-local feature extractor and a point-feature query network.

First, we integrate the DLA residual module into the point-local feature extraction process. The input consists of N points with xyz coordinates and RGB values. These points are initially

passed through a fully connected layer that increases the feature dimension to 8. The encoder consists of four layers, each comprising a DLA module and a random sampling (RS) operation. After four iterations of DLA and RS, the feature dimensions become 32, 128, 256, and 512, while the number of points is reduced to N/4, N/16, N/64, and N/256, respectively.

Next, at each layer, the xyz coordinates are used to query the neighborhoods of the labeled points. The Euclidean distance between the centroid and its neighboring points is employed as a weight, which is then used to perform trilinear interpolation on the encoded features. Finally, the interpolated features are concatenated and fed into a series of MLPs to directly infer the semantic categories of the points. The predicted points can subsequently be used to generate weak labels.

3.1 Network

From a local perspective, ridge beasts, chiwen, and ornamental animals are connected with ridges and exhibit similar appearances, which leads to these small target classes of point clouds being easily categorized as ridges. We introduce a self-attention mechanism based on the previously studied SQN-DLA [] to better distinguish the classes of point clouds, and we continue on the basis of this network. Since the positions of ridge beasts, ornamental animals, and chiwen in the roofs of ancient buildings are relatively fixed (the chiwen are all on both sides of the main ridge and directly above it, the ornamental animals are basically in the corners of the roof, and ridge beasts are generally located behind the ornamental animals), the information of the center point and its neighboring points is very important and plays a crucial role in the segmentation of the individual categories.

The positional coding method is constructed using centroids, neighborhood points, relative positional distances, and Euclidean distances. Equation (1) is provided below:

$$r_i^k = MLP\left(p_i \oplus p_i^k \oplus \left(p_i - p_i^k\right) \oplus \| p_i - p_i^k \|\right) \quad (1)$$

Where p_i and p_i^k denote the centroid and its neighboring points, respectively, p_i-p_i^k denotes the relative distance between the centroid and the neighboring points, $\|\cdot\|$ represents the Euclidean distance between the computed centroid and its neighboring points, and MLP represents a linear transformation function.

The specific structure of the positional coding is shown in Fig.2:

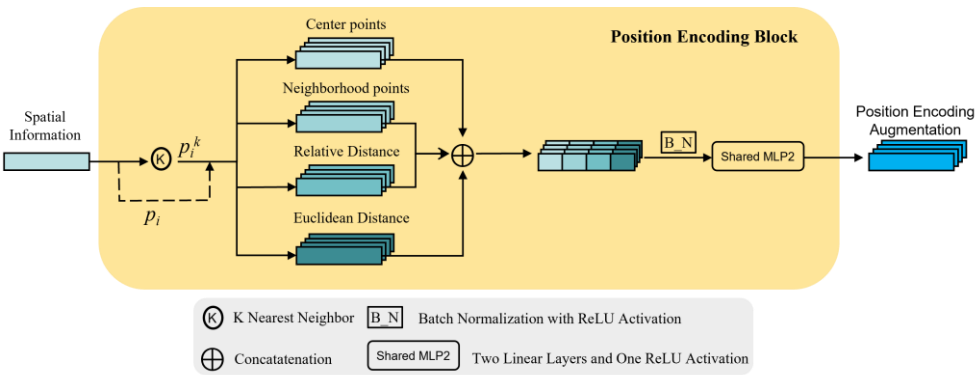

Figure 2 Position encoding method

In this regard, our overall pipeline is shown in Fig. 3. It comprises the SQN-DLA network, with the DLA component incorporating the positional encoding defined in Eq. (1). The input points are downsampled using a grid and then passed through the network, thereby completing the training of the model. The model generated by the network can be used not only for testing but also for generating pseudo-labels on the training set. The pseudo-labels are then employed to train new models.

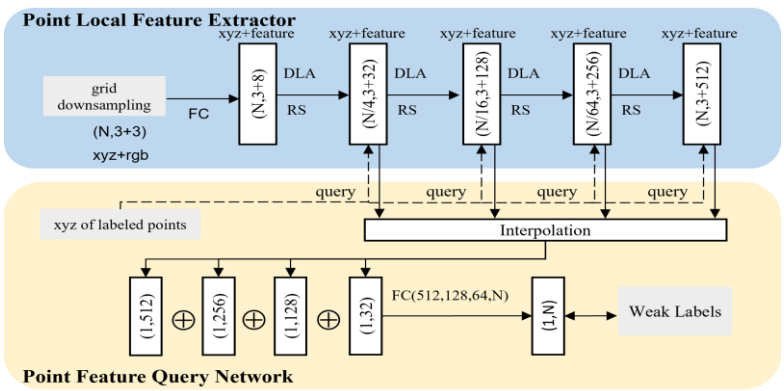

Figure 3 Roof Segmentation Process

4. Experiments

4.1 Dataset

The 3D laser scanning point cloud of the Beiding Niangniang Temple is used as the experimental data in this study. The temple, constructed over 600 years ago, is a significant architectural landmark along Beijing's central axis, possessing considerable cultural heritage value. As a traditional timber structure, it features a well-organized spatial layout that exemplifies the architectural characteristics of the Ming and Qing dynasties.

The roof structure primarily comprises tiles, ridges, ridge beasts, chiwen, and ornamental animals, and is equipped with wires for lightning protection. The dataset includes three roof types-flush gable, overhanging gable, and round gable-hip roofs-as shown in Fig. 4. Since both the round gable-hip and overhanging gable roofs originally feature mountain flowers, these elements were removed to ensure consistency between the training and testing datasets. The round gable-hip roof dataset was then merged with that of the overhanging gable roof, and the two were manually separated to expand the dataset. The input data and corresponding ground truth values are presented in Fig. 4. These roofs pertain to the Niangniang Hall, the Hall of the Heavenly King, and its two side halls. The roof datasets of the Niangniang Hall and its two supporting halls, totaling 2.04 GB, are used for training, while the roof dataset of the Hall of the Heavenly King, at 788 MB, is used for testing.

Our dataset comprises six categories: tiles, ridges, ridge beasts, chiwen, ornamental animals, and wires. All roofs include two ridge types-the main ridge and the pendant ridge. The main ridge is situated between the two chiwen, while the pendant ridge is oriented perpendicular to the main ridge. Roofs with hip structures also feature impinging (forked) ridges that intersect the pendant ridge at a 45° angle externally, and roofs with pediments include an additional ridge positioned below the pediment. These elements are collectively referred to as ridges and are indicated in orange in Fig. 5. On the pendant and impinging (forked) ridges, secondary ridge elements are present and are shown in gray in Fig.5. Ridge beasts and ornamental animals, collectively referred to as walking animals, are distributed along the vertical or bump ridges and are marked in red in Fig. 5. The chiwen, tiles, and wires are depicted in yellow, green, and black, respectively, in Fig. 5.

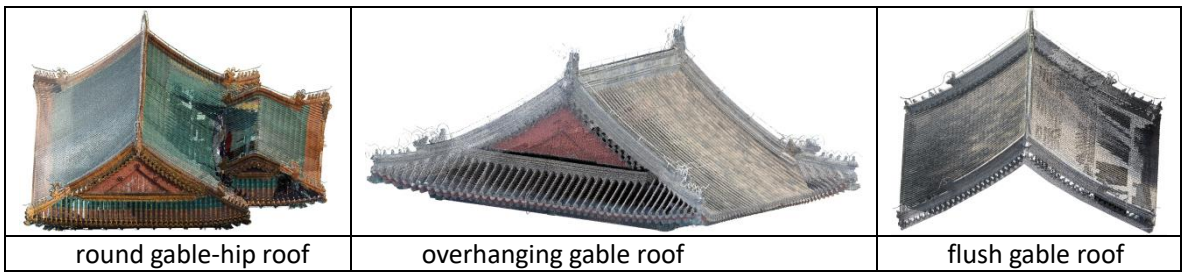

Figure 4 Dataset roof types

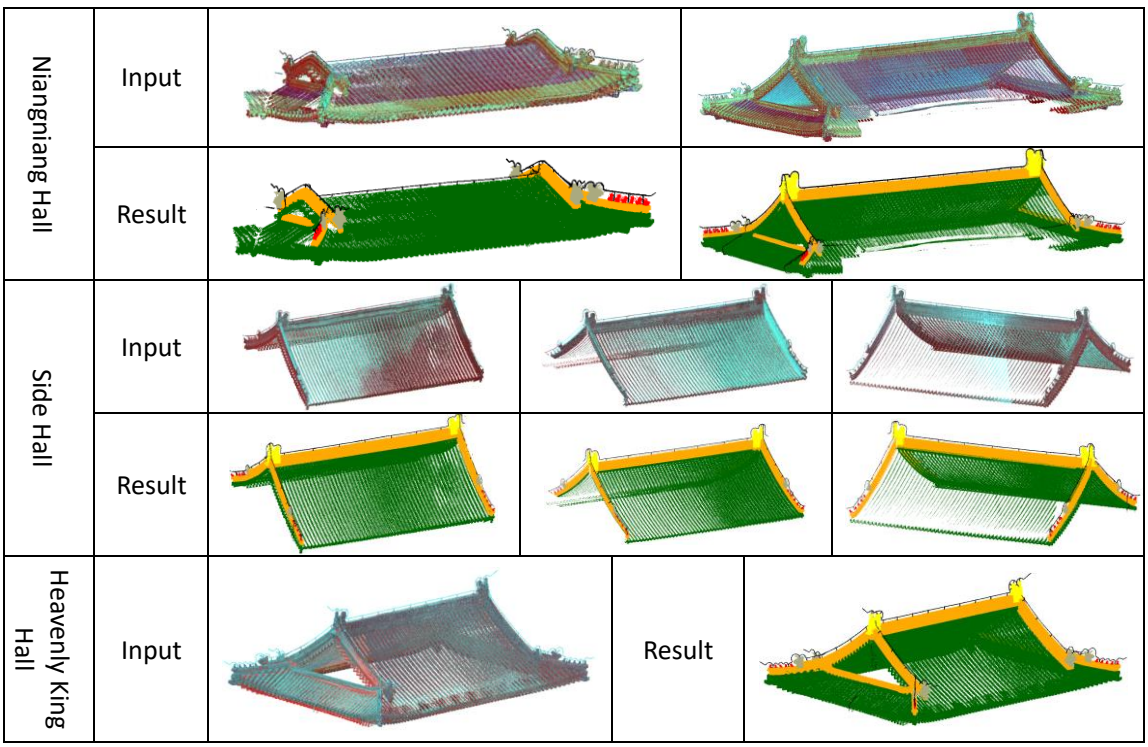

Figure 5 Dataset roof types

Table 1 presents the training set point cloud statistics, from which it is evident that the vast majority of points belong to the tile category. Categories such as ornamental animals, ridge beasts, and wires account for only a very small fraction; notably, ornamental animals comprise less than one percent relative to tiles. Due to computational constraints, it is not feasible to input all point clouds into the network, making down sampling inevitable. Consequently, the number of points representing these minor targets is further reduced or may even vanish, potentially preventing the network from learning sufficient features and leading to poor segmentation performance.

Types	ridges	ridge beasts	chiwen	ornamental animals	tiles	wire
Number of point clouds	1731069	141515	333935	66888	6806225	111891

Table 1 Number of various types of points in China's roof point cloud dataset

4.2 Experimental details

We adopt the data preprocessing method of RandLA-Net (Hu, 2020) to perform grid-based down sampling on the raw data, with a sampling point spacing set to 0.01 meters. For data annotation, we randomly select 0.1% of the points for labelling and conduct end-to-end training based on these annotations. All experiments were performed on an environment equipped with an Intel® Xeon® Platinum 8255C CPU @ 2.50 GHz and an NVIDIA RTX 2080Ti GPU. During training, 40,960 points are randomly sampled from each scene as model input. Owing to the rapid convergence of the loss function, we set the number of training epochs to 60, with an initial learning rate of 0.01 and a decay of 5% after each epoch. Additionally, we configure the nearest neighbour search with a K value of 16 and employ a batch size of 3. Considering that some segmentation targets are small, the pseudo-label generation process may yield relatively high prediction errors for these targets-and the errors in the original network are even higher-which could compromise the fairness of the experiment. Therefore, in this study, we do not perform iterative pseudo-labelling, thereby ensuring the objectivity and consistency of the experimental comparisons.

4.3 Contrast Experiment

To validate the feasibility of our method, we compare it with several state-of-the-art networks from recent years, including fully supervised approaches (RandLA-Net [Hu et al.,2020], BAAF [Qiu et al.,2021]) and weakly supervised methods (SQN [Hu et al.,2022], SQN-DLA [Zhao et al.,2024], PSD [Zhang et al.,2024]).

Table 2 presents the quantitative segmentation results, while Fig.6 provides a qualitative comparison between our method and SQN. The results indicate that our approach outperforms the competing methods in most category IoU, with overall performance surpassing that of the fully supervised methods.

This demonstrates that our method not only significantly reduces time and labor costs but also achieves superior segmentation outcomes. In particular, our method shows enhanced performance in segmenting small targets (such as ornamental animals, ridge beasts, and chiwen). The baseline network struggles to differentiate between the various types of beast ornaments on the roof ridge, resulting in lower segmentation accuracy, and it also exhibits a higher rate of misclassification within the tile category. This is mainly due to the overwhelming proportion of tile points and the difficulty in distinguishing the features of the beast ornaments. Overall, compared with SQN, our method improves the mIoU by 32.9% and the overall accuracy (OA) by 2.15%. The relatively modest increase in OA is primarily attributable to the fact that tiles account for the majority of points and thus exert a greater influence on overall accuracy. However, from the perspectives of mIoU and per-category IoU, our method achieves more precise segmentation of the various beast ornaments on the roof ridge, further validating its effectiveness in the semantic segmentation of complex ancient architectural point clouds.

methods	Labeling ratios	mIoU(%)	ridges	ridge beasts	chiwen	wire	ornamental animals	tiles	OA(%)
SQN	0.1%	45.83	70.20	37.04	2.83	34.51	35.12	95.26	93.28
Randla-net	100%	71.97	73.99	56.96	56.57	86.94	62.37	95.01	94.61
SQN-DLA	0.1%	74.07	79.78	48.44	39.58	87.88	91.63	97.14	96.03
BAAF	100%	73.71	66.23	63.78	61.29	86.84	70.05	94.06	93.99
PSD	1%	38.94	54.81	0	0	86.30	0.02	92.52	91.31
Ours	0.1%	78.73	74.02	74.81	38.99	91.66	93.92	96.79	95.43

Table 2 Various types of segmentation IoU, mIoU and OA of the roof as a whole of the roof of the ancient building

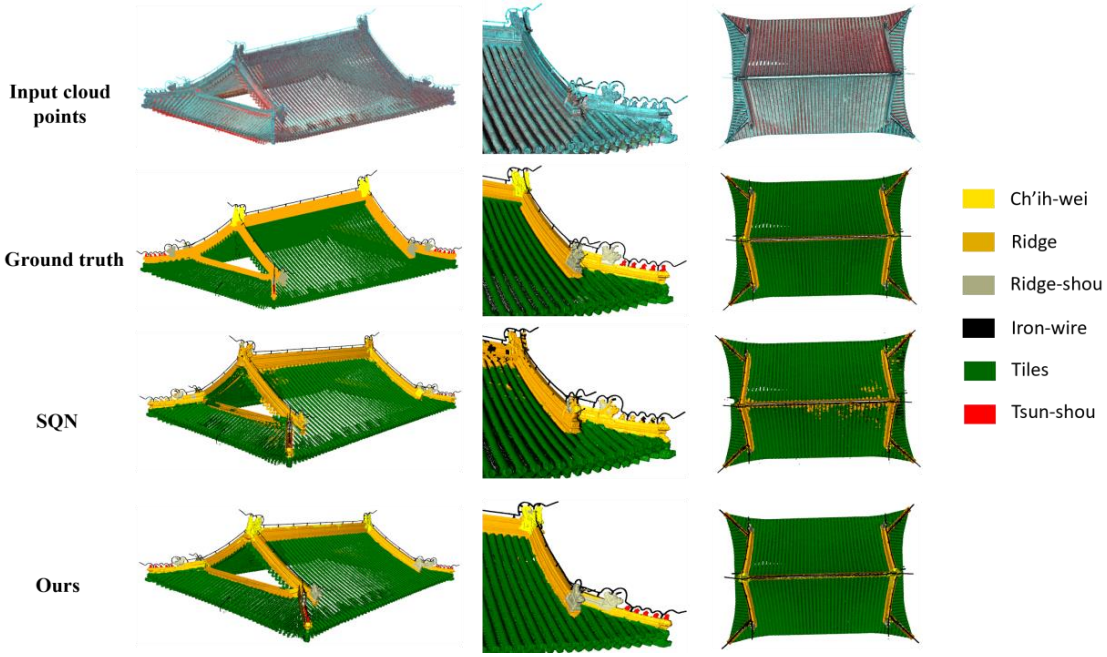

Figure 6 Comparison of multi-view segmentation

4.4 Ablation Experiments

Using SQN as the baseline, we conducted ablation experiments. We introduced a self-attention module and 4-position encoding, and the experimental results are presented in Table 3. Since the baseline already employs 4-position encoding, it serves as the point of comparison. SQN-DLA is presented separately because it uses 2-position encoding; its segmentation performance is shown in the second row of Table 3, which clearly demonstrates that incorporating the self-attention module is indeed necessary relative to the baseline. Subsequently, by adding 4-position encoding on top of the self-attention module, the segmentation performance was further enhanced, as shown in the third row of Table 3. Although the OA decreased slightly, the mIoU improved. As can be clearly seen from Table 2, the drop in OA is mainly due to reduced performance in tile segmentation, whereas the segmentation of ridge beasts and walking animals improved considerably. This indicates that augmenting the positional encoding and integrating it with the attention module can effectively enhance the segmentation of small targets.

baseline	4-position encoding	self-attention module	mIoU(%)	OA(%)
√	√		45.83	93.28
√		√	74.07	96.03
√	√	√	78.37	95.44

Table 3 Roof split ablation experiment

5. Conclusions and outlook

For the refinement of roof components in ancient buildings, we propose a discussion on positional encoding based on our previous research in weakly supervised semantic segmentation. This method not only reduces the cost of point cloud annotation but also enables fine-grained segmentation of each roof component. Notably, most of the primary segmentation targets belong to small-scale architectural elements. Furthermore, we reaffirm the effectiveness of our method (SQN-DLA) in point cloud semantic segmentation of roof components. This study provides an important theoretical and methodological foundation for the future conservation, digital modeling, and mapping of ancient building roofs.

References

Elkhrachy, I.,2017. Feature Extraction of Laser Scan Data Based on Geometric Properties. Journal of the Indian Society of Remote Sensing, 45(1), 1-10.

Grilli, E., & Remondino, F.,2019. Classification of 3D Digital Heritage. Remote Sensing, 11(7), 847. doi.org/10.3390/rs11070847.

Hu, Q., Yang, B., Fang, G., Guo, Y., Leonardis, A., Trigoni, N., & Markham, A.,2022. Sqn: Weakly-supervised semantic segmentation of large-scale 3d point clouds. In *European Conference on Computer Vision* (pp. 600-619). Cham: Springer Nature Switzerland.

Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N., & Markham, A.,2020. RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 11108-11117). Seattle, WA, USA.

Maltezos, E., & Ioannidis, C.,2018. Plane Detection of Polyhedral Cultural Heritage Monuments: The Case of Tower of Winds in Athens. Journal of Archaeological Science: Reports, 19, 562-574.

Qian, Y.H., Wang J.X., Zheng X.T.,2024. A method for single tree segmentation in airborne LiDAR point cloud based on spectral clustering and particle swarm optimization to improve K-means clustering[J]. Journal of Geo-information Science, 26(9): 2177-2191. DOI:10.12082/dqxxkx.2024.240243.

Qiu, S., Anwar, S., & Barnes, N.,2021. Semantic Segmentation for Real Point Cloud Scenes via Bilateral Augmentation and

Adaptive Fusion. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

Wan, F., Liu, Z.X., & Tan, M.,2021. Nanyue Guyidao Lishi Jianzhu Shuzihua Baohu yu VR Changjing Goujian [Digital Preservation and VR Scene Construction of Historical Buildings on the South China Ancient Post Road]. Cehui Tongbao [Bulletin of Surveying and Mapping], (2), 108-111. doi.org/10.13474/j.cnki.11-2246.2021.0054.

Wang, Z.L. ,2011. Lüe Lun Zhongguo Gudai Jianzhu [A Brief Discussion on Ancient Chinese Architecture]. In Y.H. Zhang (Ed.), Gu Jianzhu Mingjia Tan [Perspectives from Masters of Ancient Architecture] (p. 6). Beijing: China Architecture & Building Press.

Yuan, J.L., Li, S.C., & Song, T.,2022. Mu Goujia Gu Jianzhu Wuding yu Wei Huqiang Kangzhen Gouzao Jianding de Tantao [Seismic Evaluation of Roof and Enclosure Walls in Ancient Timber-Frame Buildings]. Dizhen Gongcheng yu Gongcheng Zhendong [Earthquake Engineering and Engineering Vibration], 42(5), 104-109. doi.org/10.13197/j.eeed.2022.0511.

Zhang, R.J., Zhou, X., Zhao, J.H., & Cao, M.,2020. Yi Zhong Gu Jianzhu Dianyun Shuju de Yuyi Fenge Suanfa [A Semantic Segmentation Algorithm for Ancient Building Point Clouds]. Wuhan Daxue Xuebao (Xinxi Kexue Ban) [Geomatics and Information Science of Wuhan University], (5), 753-759. doi.org/10.13203/j.whugis20180428.

Zhang, Y., Li, Z., Xie, Y., Qu, Y., Li, C., & Mei, T.,2021. Weakly Supervised Semantic Segmentation for Large-Scale Point Cloud. Proceedings of the AAAI Conference on Artificial Intelligence, 35(4), 3421-3429.

Zhang, Y., Qu, Y., Xie, Y., Li, Z., Zheng, S., & Li, C.,2021. Perturbed Self-Distillation: Weakly Supervised Large-Scale Point Cloud Semantic Segmentation. Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 15520-15528).

Zhao, J., Yu, H., Hua, X., Wang, X., Yang, J., Zhao, J., & Xu, A.,2024. Semantic Segmentation of Point Clouds of Ancient Buildings Based on Weak Supervision. Heritage Science, 12(1), 232.

Zhou, C., Dong, Y., Hou, M., Ji, Y., & Wen, C.,2024. MP-DGCNN for the semantic segmentation of Chinese ancient building point clouds. Heritage Science, 12, 1-14.