EEI-NET: EDGE-ENHANCED INTERPOLATION NETWORK FOR SEMANTIC SEGMENTATION OF HISTORICAL BUILDING POINT CLOUDS

Yaqian Xue[1], Ruiju Zhang[123], Jian Wang[14*], Jianghong Zhao[123], Lei Pang[1]

1 School of Geomantic and Urban Information, Beijing University of Civil Engineering and Architecture, Beijing 102616, China
2 Engineering Research Center of Representative Building and Architectural Heritage Database, Ministry of Education, Beijing 102616, China
3 Beijing Key Laboratory for Architectural Heritage Fine Reconstruction & Health Monitoring, Beijing 102616, China
4 Institute of Science and Technology Development, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

*Corresponding author.
E-mail addresses: 3324767399@qq.com

KEY WORDS: Semantic segmentation, Historical architecture point cloud, EEI, EIC, EEI-Net

ABSTRACT:

In recent years, the conservation research of historical buildings and cultural relics has received a lot of attention from the state and the people, which not only provides a deeper understanding of their historical value and cultural significance, but also promotes the expansion of conservation research to the three-dimensional level. In this context, the semantic segmentation of historical building components is particularly important, which can provide basic support for various historical building applications, such as research and study of historical buildings, repair and protection, and 3D fine reconstruction, etc. However, most of the current methods for semantic segmentation of point clouds of historical buildings suffer from the problems of not being able to fully exploit the local neighborhood information of point clouds and poor edge segmentation. Therefore, we propose a new deep learning semantic segmentation-based approach, which we call EEI-Net. It is an end-to-end deep neural network in which we designed an edge enhancement interpolation (EEI) module and an edge interaction classifier (EIC). The edge enhancement interpolation module performs edge enhancement interpolation by fusing multi-layer features between the encoder and decoder. The edge interaction classifier enables the interaction of edge information through information transfer between individual nodes. EEI-Net incorporates contextual features and better preserves and enhances the edge information of the point cloud. We conduct experiments on the constructed historical architecture dataset, and the results show that the proposed EEI-Net has better performance.

1. INTRODUCTION

Chinese culture has a long history, among which our historical architectural art style is exquisite and scientific, which not only shows the wisdom crystallization of our ancient working people, but also is a valuable heritage wealth. However, with the passage of time and the impact of the environment, only a few ancient wooden structures have been preserved today (Chun et al., 2015). This makes it particularly urgent to strengthen the research work on the conservation of historical buildings in China.

With the rapid development of 3D sensors, the conservation of historical buildings is gradually evolving from the traditional way to the digital way. Using 3D point cloud technology, we can obtain more realistic dimensions and various architectural details of historical buildings and provide precise geometric coordinates (X, Y, Z) in the form of millions of points, which has become one of the most effective ways to document the shape of cultural heritage (Yang et al., 2020). Airborne laser scanners (ALS) (Elsner et al., 2018), mobile laser scanning (MLS) (Zhang et al., 2019), terrestrial laser scanning (TLS) (Zhu et al., 2021) and unmanned aerial vehicle (UAV) photogrammetry (Poli and Caravaggi, 2013) have become the most popular methods for collecting urban 3D point clouds, which can be applied to both indoor and outdoor scenes. In this paper, we construct a traditional historical building point cloud dataset by using MLS, which provides the real coordinates of historical buildings compared to images and is not affected by

lighting conditions and image distortion. In addition, the dataset can show more clearly the spatial information between each component of historical buildings.

In recent years, researchers have proposed many segmentation methods for 3D point clouds, and the irregular and disordered structure of 3D point clouds has become one of the biggest challenges for 3D feature extraction and further semantic segmentation (Xie et al., 2020; Bello et al., 2020; Cheng et al., 2021; Chen et al., 2021a). Therefore, researchers are exploring and investigating more effective feature extraction methods.

Traditional point cloud semantic segmentation methods are based on machine learning methods, such as support vector machines, random forests, etc. These methods use hand-designed features to identify semantic information in the point cloud. While these methods perform well on specific tasks, traditional point cloud semantic segmentation methods have difficulty producing good results on larger and more complex datasets due to the limitations of hand-designed features. As a result, researchers have turned their attention to deep learning-based approaches. Currently, most point cloud feature extraction methods and their corresponding semantic segmentation methods can be classified into three types: point cloud projection-based methods (Milioto et al., 2019; Lyu et al., 2020), voxel-based methods (Le et al., 2018; Meng et al., 2019) and point-based methods (Triess et al., 2020). Since both projection-based and voxel-based methods may lose information during projection or voxelization, most researchers

have focused on point-based methods (Chen et al., 2021b; Qian et al., 2021; Qian et al., 2022). Among them, PointNet (Qi et al., 2017a) as well as PointNet++ (Qi et al., 2017b) are considered as pioneer works. PointNet, the pioneering method, runs directly on the point cloud, but because its point-by-point features are learned individually from each point in PointNet, it ignores the local contextual information between points. Based on this, this work has been extended in various ways to enhance the acquisition of local information from point clouds. For example, DGCNN(Wang et al., 2019) proposes an edge convolution (EdgeConv) for learning edge features, which is used to extract features of centroids and edge vectors of centroids and K nearest neighbors (KNN) points. While these methods achieve better performance in semantic segmentation, most of them are limited to very small 3D point clouds input to the network and do not scale directly to larger point clouds. Thus, RandLA-Net (Hu et al., 2020) was proposed in order to adequately accommodate large-scale point cloud scenarios. It uses an efficient point cloud random sampling strategy and local spatial location encoding, which can achieve high segmentation accuracy and processing speed, but also leads to the possibility that he may lose critical contextual information during reduced sampling. Therefore, most researchers turn to new structures of backbone networks and work on different ways to improve the semantic segmentation accuracy of point clouds.

In recent years, attention-based approaches have flourished (Feng et al., 2020), and most of the different backbones employ different types of attention mechanisms, which automatically learn important local features by assigning larger weights to key information. Existing attention mechanisms include include channel attention, spatial attention, self-attention and multi-attention, etc. In particular, self-attention has shown excellent performance in image analysis (Hu et al., 2019) and point cloud processing (Guo et al., 2021). In point cloud processing (Guo et al., 2021), the self-attentive mechanism is used to establish the relationship between the centroid and all points in the global space. As for the historical architectural heritage, we want to completely divide the upper part of the windows, square and columns, especially their edges, from the walls, so we need to pay more attention to learning the local neighborhood features.

More semantic segmentation methods based on deep learning have also been proposed in the field of historical architectural heritage. Dong et al. (Ji et al., 2021) modified DGCNN to segment MQDOA roofs. Francesca Matrone et al. (Matrone et al., 2020) compared machine learning methods with deep learning methods for large 3D artifact classification, synthesized the advantages of both methods, and proposed a cultural heritage point cloud semantic segmentation architecture,

DGCNN-Mod+3Dfeat, that incorporates the advantages of both methods. Pierdicca et al. (Pierdicca et al., 2020) proposed an improvement by adding meaningful features to the DGCNN, such as normal vectors and colors, but the framework was unable to evaluate the accuracy performance of the acquisition technique and continued improvement is needed. Although all these improvements improve the segmentation accuracy of the point cloud of historical buildings, the accuracy of the column components is lower compared to the other components, and all of them have the problem of unsatisfactory acquisition of local features and edge information.

Our major contributions to this work include:
(1) In order to improve the performance of semantic segmentation of 3D point clouds, we propose a new semantic segmentation method that attempts to redesign the framework structure of semantic segmentation of 3D point clouds. EEI-Net manages to extract differentiated semantic features and predict smoother results.
(2) We design an Edge Enhancement Interpolation Encoder-Decoder (EEI-ED) for efficient feature extraction by reducing the semantic gap between the encoder and decoder through edge enhancement interpolation. With this design, multilayer features from the encoder and features in the decoder are edge-enhanced interpolated. The interaction capability of multi-layer features is enhanced.
(3) We designed an Edge interactive classifier (EIC), which enhances the context-awareness of points through information transfer between nodes for better label prediction.

2. METHODS

In this paper, we propose an edges enhancement interaction network for semantic segmentation of 3D point clouds, which consists of Edge Enhancement Interpolation Encoder-Decoder (EEI-ED) and Edge interactive classifier (EIC). The model uses a U-shaped encoding-decoding structure, where the encoder reduces the number of points by a factor of ×4 through Random Sample (RS) and extends the dimensionality of each point feature in five consecutive layers through the Local feature Aggregate (LFA) (Hu et al., 2020) module, and the corresponding decoder increases the number of points through upsampling (US) and compresses the dimensionality of point features in five connected layers using mlp. To handle the semantic gap between encoder and decoder, we design an edge enhancement interpolation module in the decoder to obtain more discriminative features. Its structure diagram is shown in Figure 1:
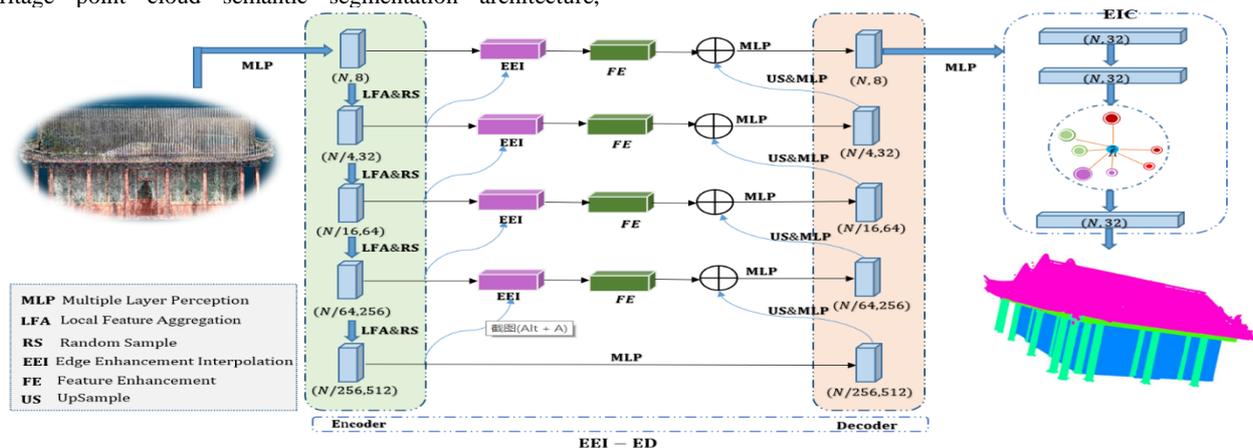
Figure 1. Network architecture of the EEI-Net.

2.1 Edge enhancement interpolation module

The purpose of edge enhancement interpolation is to enhance the information exchange ability of multi-layer features between the encoder and decoder. For convenience, we denote the

features of the decoder layer corresponding to F_e^l as F_d^l. As shown in Figure 2, for F_d^l in the decoder, we augment it with F_e^l and F_e^{l+1} from the encoder.
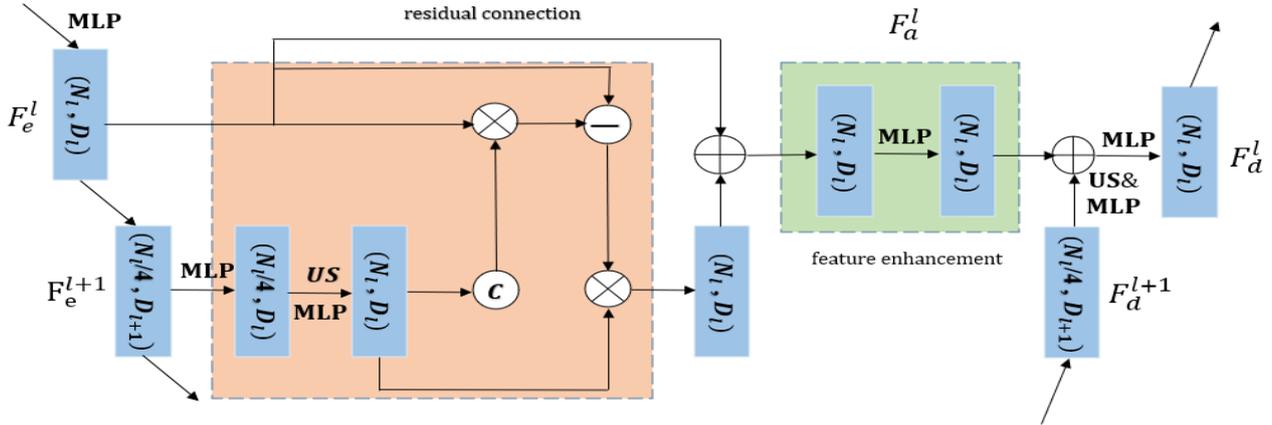
Figure 2. Edge enhancement interpolation module.

We first reapply to a shared MLP on F_e^{l+1}, compress its dimension to D1, weighted interpolation of the points of the intermediate features, then obtain the output features using the convolution operation, fuse and enhance the features of different layers by point-by-point multiplication, then calculate the difference between the fused features and the output features of the jth-1st layer encoder using the absolute difference, and finally multiply the high-resolution features with the above feature fusion differences element-by-element to enhance the expressiveness of F_e^l. We introduce more high-frequency information while keeping the feature space consistent, and then add the enhanced features f_enhance and the original features f_encoder_list[-j-2] to form the final enhanced feature representation. In general, it utilizes the edge-aware upsampling module and edge enhancement module of the deep neural network, thus allowing better preservation and enhancement of the edge information of the point cloud and improving the segmentation quality. The full operation can be expressed as follows:

$$F_a^l = F_e^l + \left(F_e^l \odot US\left(M(F_e^{l+1})\right) - F_e^l\right) \odot US(M(F_e^{l+1})) \quad (1)$$

where F_a^l is the modulation feature between the lth layer encoder and the lth layer decoder, and since this method uses multiple layers of features from the encoder and features in the decoder for edge-enhanced interpolation, the process is called edge-enhanced interpolation (EEI). Edge-enhanced interpolation captures the interaction information of edge multilayer features

and achieves local contrast enhancement of the features using convolution, weighted interpolation and residual concatenation. To further attenuate the effect of spatial indistinguishability, we process the intermediate features using a Feature Enhancement (FE) module, which is implemented with a simple shared MLP. Finally, we fuse the modulated encoder features with the decoder features. This edge enhancement interpolation mechanism can be expressed as:

$$F_d^l = \begin{cases} M\left(F_d^{l+1}\right) & if \quad l = 0 \\ M\left(F_e^l\right) & if \quad l = 5 \\ M\left(FE\left(F_a^l\right) + M(US(F_d^{l+1}))\right) & otherwise \end{cases} \quad (2)$$

2.2 Edge interactive classifier

In previous work, the classifier generates point-by-point semantic labels individually through mlp implemented by a fully connected layer. A fully connected layer consists of a linear exchange and a nonlinear activation function, however, the nonlinear activation function leads to inconsistent neighbors in the prediction. For this purpose we introduce the edge interaction module, which uses the neighboring node information of each node, based on the information transfer from the current node to each neighboring node, thus enabling edge-to-edge interaction. Edge interactive classifier module is illustrated in Figure 3:
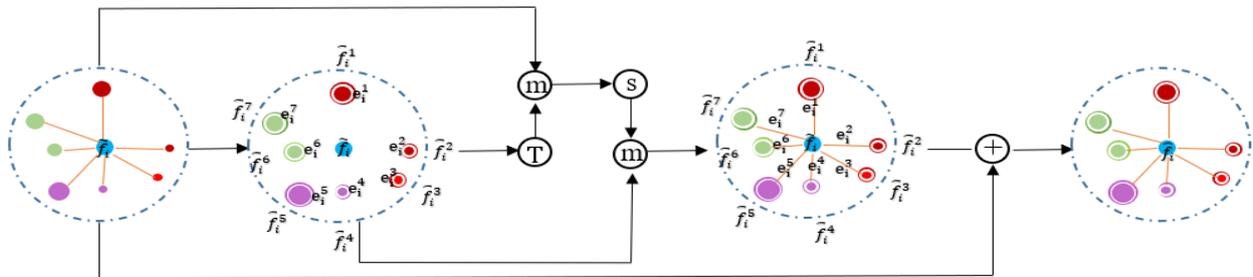
Figure 3. The schematic diagram of the edge enhancement interpolation module.

The point cloud is denoted as $\widetilde{P} = \{x_n, y_n, z_n, \widetilde{f}_n\}_{n=1}^{N}$, where x_n, y_n, z_n denote the coordinates of the nth point and $\widetilde{f}_n \in R^D$ represents its features as the output of the previous layer of the local aggregation module, from which the feature maps of neighboring nodes are aggregated. For vertex $\widetilde{f}_i$, obtain its k nearest vertex features $\{\widetilde{f}_i^k\}_{k=1}^{K}$, obtain edge features by matching them with feature maps, and edge features as:

$$e_i^j = ReLU(W \cdot \widetilde{f}_i^j) \qquad (3)$$

where W is the learnable weight and ReLU denotes the ReLU activation function. e_i^j is the edge feature between the i-th point and its j-th neighboring point. The attention matrix is then computed for each node by computing its attention matrix over the original features and edge features to calculate the relationship between each node and its neighboring nodes:

$$a_j^i = (e_i^j)^T \cdot e_k^j \qquad (4)$$

Further calculate its attention weight matrix. The edge-to-edge interaction is achieved by multiplying it by f_neigh to calculate the neighbor representation of each node and obtain the information transfer from the current node to each neighboring node, and this process is:

$$f_i' = \sum_j f_j^i \cdot E_j \qquad (5)$$

where f_i' denotes the aggregated features of node i, E_j denotes the feature matrix of neighbor node j, and f_j^i denotes the attention coefficient between node i and neighbor node j. Finally, we perform maximum pooling, select the most important features among them, and add them with the results of the previous convolutional layers to form fused features. This module is embedded in the last two mlps to capture edge-to-edge interaction features, improve contextual information, and employ the cross-entropy loss between predicted and true values as the loss function of EIC.

3. EXPERIMENTS

3.1 Experiment details

Our experiments were conducted on a single NVIDIA GeForce RTX 3060 TI GPU and an Intel® Core^TMi7-10700K CPU @ 3.80GHz *16 and implemented in TensorFlow on a server configured with CUDA11.3 and CUDNN8.2.1. We use the Adam algorithm with default parameters as the optimizer. The initial learning rate is set as 0.01 and decreases by 10% after each epoch. We trained 100 epochs on our own dataset, with batch size set to 2 and the number of nearest points K set to 16, and sampled a fixed number of points (40960) from each point cloud to feed into the network during training, while feeding the entire point cloud data during testing.

3.2 Dataset

This study relies on the actual historical architectural heritage, using mature UAV tilt photography and 3D laser scanning and other technical means to obtain the complete internal and external mapping data of historical buildings. We reviewed relevant information to master the basic composition of historical building structures, and used a variety of software to process and label the data to establish the historical building point cloud datasets.

In this study, two measurement areas were selected. Experimental Data 1 was built in the Ming Dynasty, it is located in Tucheng outside Deshengmen, Chaoyang District, and is not only a landmark on the northern extension of the central axis of Beijing city, but also a Beijing-level cultural relics protection unit. Experimental Data 2 was built in early 1908 and is located in the southwest of the Qing agricultural test site, a typical Chinese classical building.

We use UAV tilt photography and 3D laser scanning to obtain the interior and exterior point cloud data and roof data to obtain the complete historical building scene, and use CloudCompare software to align the generated roof point cloud with the scanned point cloud to fuse into a complete historical building point cloud data including the interior and exterior structures of historical buildings. To control the amount of point cloud data, we use 0.01m spatial subsampling to streamline the point cloud data by referring to a study related to European architectural heritage, which is sufficient to enable the neural network to learn effective structural features from the point cloud. The final number of points obtained for our dataset is 20182330 points, each with coordinates and colors. Finally, the data of the required experimental range is cropped for subsequent annotation work, and the complete point cloud after fusion is shown in Figure 4.
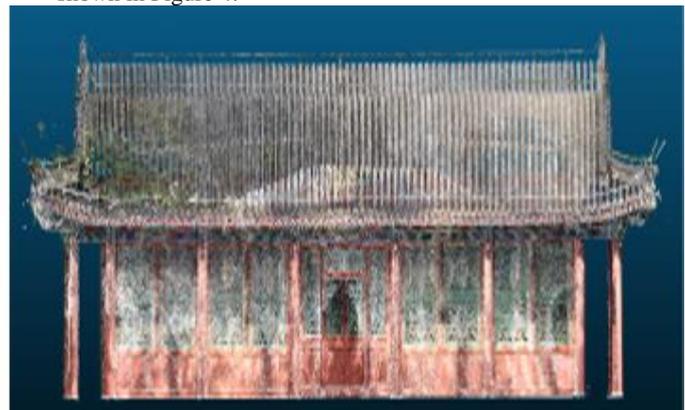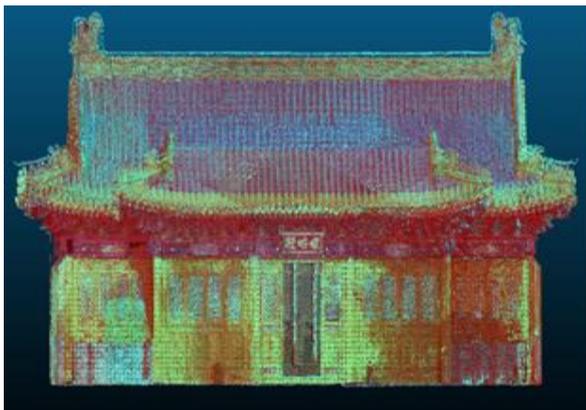
Figure 4. The experimental data 1 (left) and experimental data 2 (right).

The manual annotation of the point cloud of historical buildings is shown in Figure 5 and Figure 6. In order to scientifically evaluate the generalization ability of the segmentation network on historical building data, we cut both historical buildings along the central axis of the scene into a left part point cloud and a right part point cloud, a total of four parts, in order to carry out a four-fold cross-validation experiment. The point cloud scene is a complete point cloud containing both external and internal structures, with six semantic categories labelled: roof, column, square, door, window, and wall.

Figure 5. Experimental data 1 labeling results (left and right).
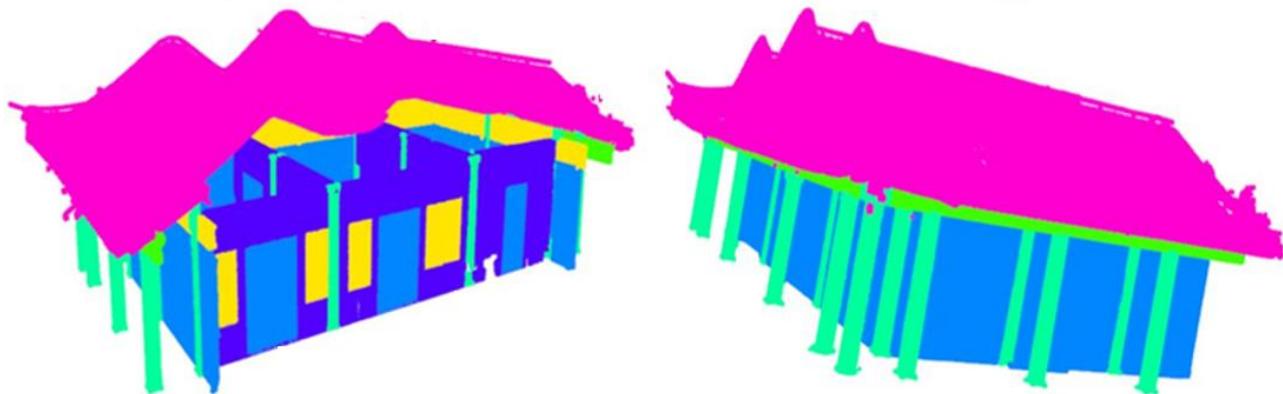
Figure 6. Experimental data 2 labeling results (left and right).

3.3 Semantic Segmentation Results

In this paper, we choose RandLA-Net and our network to validate the present and past point cloud semantic segmentation experiments on the historical architecture dataset using a four-fold crossover. To fully evaluate the semantic segmentation effect of the EEI-Net network model on the historical architecture dataset. We used two models: (a) region one for testing, regions 2 to 4 for training, and again, each region in turn for testing and the other regions for training. (b) k-fold cross-validation (k=4). We used overall accuracy(OA), mean class accuracy(mAcc), the mean intersection over union(mIoU) as standard metrics. As shown below：

$$OA = \sum_{i=0}^{k} \frac{T_P^i}{N_P}$$

(6)

$$mAcc = \frac{1}{k+1} \sum_{i=0}^{k} \frac{T_P^i}{N_P^i}$$

(7)

$$mIoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{T_P}{T_P + F_P + F_N}$$

(8)

There, k is the number of categories, T_P is the number of correctly classified positive samples, F_P is the number of incorrectly classified positive samples, F_N is the number of incorrectly classified negative samples, and N_P is the total number of points.

We adopted a four-fold cross validation method to evaluate our method, and the experimental results are shown in Table 1 and Figure 7. It is obvious from Table 1 that the performance of our proposed EEI-Net network in OA, mAcc and mIoU is better than that of the original RandLA-Net. As shown in Fig 7, the part shown in red box can clearly indicate that our method can segment the scene more smoothly, especially for the object boundaries.

	OA（%）	mACC(%)	mIOU(%)	column	door	ef	wall	roof	window
RandLA-Net	94.5	91.1	84.6	80.0	88.1	83.7	85.8	98.1	71.9
our	94.8	91.3	85.1	81.0	88.8	93.7	86.1	98.1	72.8

Table 1. Comparison of edge contour feature extraction methods.
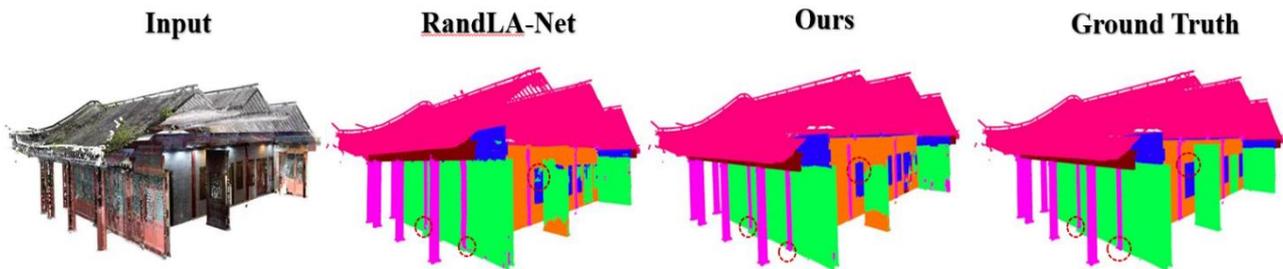
Figure 7. Qualitative results of this method on historical architecture data set.

4. CONCLUSIONS

Nowadays, historical building point clouds have complex geometry, but the point cloud dataset of historical building scenes is very limited, which is a key problem we need to solve. Therefore, a dataset of historical buildings is constructed in this paper. In this study, we propose EEI-Net in order to investigate semantic segmentation for the special nature of point clouds of historical buildings. The network consists of an edge-enhanced interpolation codec and an edge interaction classifier designed to fully explore the local neighborhood and edge features. The edge enhancement interpolation module performs edge enhancement interpolation by fusing multi-layer features between the encoder and decoder. The edge interaction classifier enables the interaction of edge information through information transfer between individual nodes. EEI-Net incorporates contextual features and better preserves and enhances the edge information of point clouds. The whole framework exhibits significant feature representation capability, which makes the features more discriminative and achieves excellent performance on the historical architecture dataset. However it still has some problems to overcome, such as incomplete and imbalanced categories. In addition to this, the work is performed under full tagging, and large point cloud datasets are becoming more and more common, resulting in high costs. Therefore, in future work, we will continue to improve the accuracy while reducing the annotation point cloud share to achieve better performance with reduced annotation cost and effort.

ACKNOWLEDGEMENTS

REFERENCES

Bello, S. A., Yu, S., Wang, C., Adam, J. M., & Li, J. (2020). Review:Deep learning on 3D point clouds. *Remote Sensing,* 12(11), 1729.

Chen, Y., Liu, X., Xiao, Y., Zhao, Q., & Wan, S. (2021a). Three-Dimensional Urban Land Cover Classification by Prior-Level Fusion of LiDAR Point Cloud and Optical Imagery. *Remote Sensing,* 13(23), 4928.

Chen, Y., Liu, G., Xu, Y., Pan, P., & Xing, Y. (2021b). PointNet++ network architecture with individual point level and global features on centroid for ALS point cloud classification. *Remote Sensing,* 13(3), 472.

Cheng, S., Chen, X., He, X., Liu, Z., & Bai, X. (2021). Pra-net: Point relation-aware network for 3d point cloud analysis. *IEEE Transactions on Image Processing,* 30, 4436-4448.

Chun, Q., Van Balen, K., Pan, J., & Sun, L. (2015). Structural performance and repair methodology of the Wenxing lounge bridge in China. *International Journal of Architectural Heritage,* 9(6), 730-743.

Elsner, P., Dornbusch, U., Thomas, I., Amos, D., Bovington, J., & Horn, D. (2018). Coincident beach surveys using UAS, vehicle mounted and airborne laser scanner: Point cloud inter-comparison and effects of surface type heterogeneity on elevation accuracies. *Remote Sensing of Environment,* 208, 15-26.

Feng, M., Zhang, L., Lin, X., Gilani, S. Z., & Mian, A. (2020). Point attention network for semantic segmentation of 3D point clouds. *Pattern Recognition,* 107, 107446.

Guo, M. H., Cai, J. X., Liu, Z. N., Mu, T. J., Martin, R. R., & Hu, S. M. (2021). Pct: Point cloud transformer. *Computational Visual Media,* 7, 187-199.

Hu, H., Zhang, Z., Xie, Z., & Lin, S. (2019). Local relation networks for image recognition. *In Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 3464-3473).

Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., ... & Markham, A. (2020). Randla-net: Efficient semantic segmentation of large-scale point clouds. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11108-11117).

Ji, Y., Dong, Y., Hou, M., Qi, Y., & Li, A. (2021). An extraction method for roof point cloud of historical building using deep learning framework. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* 46, 321-327.

Le, T., & Duan, Y. (2018). Pointgrid: A deep network for 3d shape understanding. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 9204-9214).

Lyu, Y., Huang, X., & Zhang, Z. (2020). Learning to segment 3d point clouds in 2d image space. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 12255-12264).

Milioto, A., Vizzo, I., Behley, J., & Stachniss, C. (2019, November). Rangenet++: Fast and accurate lidar semantic segmentation. *In 2019 IEEE/RSJ international conference on intelligent robots and systems (IROS)* (pp. 4213-4220). IEEE.

Meng, H. Y., Gao, L., Lai, Y. K., & Manocha, D. (2019). Vv-net: Voxel vae net with group convolutions for point cloud segmentation. *In Proceedings of the IEEE/CVF international conference on computer vision* (pp. 8500-8508).

Matrone, F., Grilli, E., Martini, M., Paolanti, M., Pierdicca, R., & Remondino, F. (2020). Comparing machine and deep learning methods for large 3D heritage semantic segmentation. *ISPRS International Journal of Geo-Information,* 9(9), 535.

Poli, D., & Caravaggi, I. (2013). 3D modeling of large urban areas with stereo VHR satellite imagery: Lessons learned. *Natural hazards,* 68, 53-78.

Pierdicca, R., Paolanti, M., Matrone, F., Martini, M., Morbidoni, C., Malinverni, E. S., ... & Lingua, A. M. (2020). Point cloud semantic segmentation using a deep learning framework for cultural heritage. *Remote Sensing,* 12(6), 1005.

Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017a). Pointnet: Deep learning on point sets for 3d classification and segmentation. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 652-660).

Qi, C. R., Yi, L., Su, H., & Guibas, L. J. (2017b). Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30.

Qian, G., Hammoud, H. A. A. K., Li, G., Thabet, A. K., & Ghanem, B. (2021). Anisotropic Separable Set Abstraction for Efficient Point Cloud Representation Learning.

Qian, G., Li, Y., Peng, H., Mai, J., Hammoud, H., Elhoseiny, M., & Ghanem, B. (2022). Pointnext: Revisiting pointnet++ with improved training and scaling strategies. *Advances in Neural Information Processing Systems,* 35, 23192-23204.

Triess, L. T., Peter, D., Rist, C. B., & Zöllner, J. M. (2020, October). Scan-based semantic segmentation of lidar point clouds: An experimental study. *In 2020 IEEE Intelligent Vehicles Symposium (IV)* (pp. 1116-1121). IEEE.

Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., & Solomon, J. M. (2019). Dynamic graph cnn for learning on point clouds. *Acm Transactions On Graphics (tog),* 38(5), 1-12.

Xie, Y., Tian, J., & Zhu, X. X. (2020). Linking points with labels in 3D: A review of point cloud semantic segmentation. *IEEE Geoscience and remote sensing magazine,* 8(4), 38-59.

Yang, X., Grussenmeyer, P., Koehl, M., Macher, H., Murtiyoso, A., & Landes, T. (2020). Review of built heritage modelling: Integration of HBIM and other information techniques. *Journal of Cultural Heritage,* 46, 350-360.

Zhang, Z., Li, J., Guo, Y., Li, X., Lin, Y., Xiao, G., & Wang, C. (2019). Robust procedural model fitting with a new geometric similarity estimator. *Pattern Recognition,* 85, 120-131.

Zhu, J.; Xu, Y.; Ye, Z.; Hoegner, L.; Stilla, U. (2021).Fusion of urban 3D point clouds with thermal attributes using MLS data and TIR image sequences. *Infrared Phys. Technol,* 113, 103622.