# EEP-Net: Enhancing Local Neighborhood Features and Efficient Semantic Segmentation of Scale Point Clouds

Yicheng Liu[1], Fuxiang Wu[1,2,3], Qieshi Zhang[1,2,3(✉)], Ziliang Ren[4], and Jun Chen[1,2,3]

[1] Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
qs.zhang@siat.ac.cn

[2] Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Beijing, China

[3] The Chinese University of Hong Kong, Hong Kong, China

[4] DongGuan University of Technology, Dongguan, China

**Abstract.** Semantic segmentation of point clouds at the scene level is a challenging task. Most existing work relies on expensive sampling techniques and tedious pre- and post-processing steps, which are often time-consuming and laborious. To solve this problem, we propose a new module for extracting contextual features from local regions of point clouds, called EEP module in this paper, which converts point clouds from Cartesian coordinates to polar coordinates of local regions, thereby Fade out the geometric representation with rotation invariance in the three directions of XYZ, and the new geometric representation is connected with the position code to form a new spatial representation. It can preserve geometric details and learn local features to a greater extent while improving computational and storage efficiency. This is beneficial for the segmentation task of point clouds. To validate the performance of our method, we conducted experiments on the publicly available standard dataset S3DIS, and the experimental results show that our method achieves competitive results compared to existing methods.

**Keywords:** Semantic segmentation · Scale point clouds · Rotational invariance

## 1 Introduction

In recent years, image recognition technology has developed rapidly relying on deep learning. In addition to 2D image recognition, people are increasingly interested in 3D vision, because directly learning 3D tasks from acquired 2D images always has certain limitations. With the rapid development of 3D acquisition technology, the availability and value of 3D sensors are increasing, including various types of 3D scanners, lidar, and RGB-D cameras. The advent of large-scale

high-resolution 3D datasets with scale information has also brought about the context of using deep neural networks to reason about 3D data. As a common 3D data format, the point cloud retains the original geometric information in three-dimensional space, so the point cloud is called the preferred data form for scene understanding. Efficient semantic segmentation of scene-level 3D point clouds has important applications in areas such as autonomous driving and robotics. In recent years, the pioneering work PointNet has become the most popular method for directly processing 3D point clouds. The PointNet [1] architecture directly processes point clouds through a shared MltiLayer Perceptron (MLP), using the MLP layer to learn the features of each point independently, using maximum pooling to obtain global features. On the other hand, since PointNet learns the features of each point individually, it ignores the local structure between points. Therefore, to improve this, the team introduced PointNet++ [2], as the core of this network hierarchy, with an ensemble abstraction layer consisting of three layers: a sampling layer, a grouping layer, and a PointNet-based learning layer. By overlaying several ensemble abstraction layers, this network learns features from local geometric structures and abstracts local features layer by layer. PointWeb [3] connects all local point pairs using local geographical context, and finally forms a local fully connected network, and then adjusts point features by learning point-to-point features. This strategy can enrich the point features of the local region and form aggregated features, which can better describe the local region and perform 3D recognition.

The effect of these methods in processing small-scale point clouds is gratifying, but it will bring some limitations to processing scene-level point clouds, mainly because: 1) High computational volume and low storage efficiency caused by the sampling method. 2) Most existing local feature learners usually rely on computationally expensive kernelization or graph construction, and thus cannot handle large numbers of points. 3) Existing learners have limited acceptance and size to effectively capture complex structures, and do not capture enough local area features for large-scale point clouds, RandLA-Net [4] provides us with a solution to these problems. RandLA-Net [4] based on the principle of simple random sampling and effective local feature aggregator, can increase the sampling rate while gradually increasing the receptive field of each neural layer through the feature aggregation module to help effectively learn complex local structures. However, after research, it was found that RandLA-Net [4] did not pay attention to the relationship between each point and point in a neighborhood when learning the local structure.

Our main contributions are as follows:

– As the input point cloud is direction-sensitive, we propose a new local space representation that is rotationally invariant in the $X$-$Y$-$Z$ axis, which can effectively improve the performance of point cloud segmentation.
– We propose a new local feature aggregation module, Local Representation of Rotation Invariance (LRRI) , which connects the spatial representation with rotation invariance in  $X$-$Y$-$Z$ axis to the local relative point position

representation to form a new local geometric representation that effectively preserves local geometric details.

– We perform experimental validation on a representative S3DIS dataset, and our method is compared with state-of-the-art methods and achieves good performance.

## 2   Related Work

The goal of point cloud semantic segmentation is to give a point cloud and divide it into subsets according to the semantics of the points. There are three paradigms for semantic segmentation: projection-based, discretization-based and point-based.

### 2.1   Projection-Based Methods

To leverage the 2D segmentation methods, many existing works aim to project 3D point clouds into 2D images and then process 2D semantic segmentation. By which conventional convolution of 2D images can be used to process point cloud data, to solve target detection and semantic segmentation tasks. There are two main categories of such methods: (1) multi-viewpoint representation [5–8]. (2) spherical representation [9–12]. In general, the performance of multi-viewpoint segmentation methods is sensitive to viewpoint selection and occlusion. In addition, these methods do not fully utilize the underlying geometric and structural information because the projection step inevitably introduces information loss; the spherical projection representation retains more information than the single-view projection and is suitable for LiDAR point cloud labeling, however, this intermediate representation also inevitably introduces problems such as dispersion errors and occlusion.

### 2.2   Discretization-Based Methods

Discretization-based methods, which voxelized point clouds into 3D meshes and then apply the powerful 3D CNN in [13–17]. But the performance of these methods is sensitive to the granularity of voxels, and voxelization itself introduces discretization artifacts. On the other hand, the main limitation of such methods is their large computational size when dealing with large-scale point clouds. This method is very important in practical applications when choosing a suitable grid resolution.

### 2.3   Point-Based Methods

Different with the first two methods, point-based networks act directly on irregular point clouds. However, point clouds are disordered and unstructured, so standard CNNs cannot be used directly. For this reason, the paper [1] proposes the pioneering network PointNet. The irregular format and envelope invariance

of the point set are discussed, and a network that uses point clouds directly is proposed. The method uses a shared MLP as the basic unit of its network. however, the point-like features extracted by the shared MLP cannot capture the local geometric structure and interactions between points in the point cloud. So PointNet++ [2] not only considers global information, but also extends Point-Net [1] with local details of the farthest sampling and grouping layers. Although PointNet++ [2] makes use of the local environment, using only maximum pooling may not aggregate information from local regions well. For better access to contextual features and geometric structures, some works try to use graph networks [18–20] and Recurrent Neural Networks (RNN) [21–23] to implement segmentation. The article [4] proposes an efficient lightweight network Rand-LA for large-scale point cloud segmentation, which utilizes random sampling and achieves very high efficiency in terms of memory and computation, and proposes a local feature aggregation module to capture and preserve geometric features.

## 3    EEP-Net

In this section, we discuss the EEP module for large-scale point cloud segmentation, which mainly consists of two blocks: Local Representation of Rotation Invariance (LRRI), Attentive Pooling (AP). Then we introduce EEP-Net, which is an encoder-decoder network structure with EEP modules.
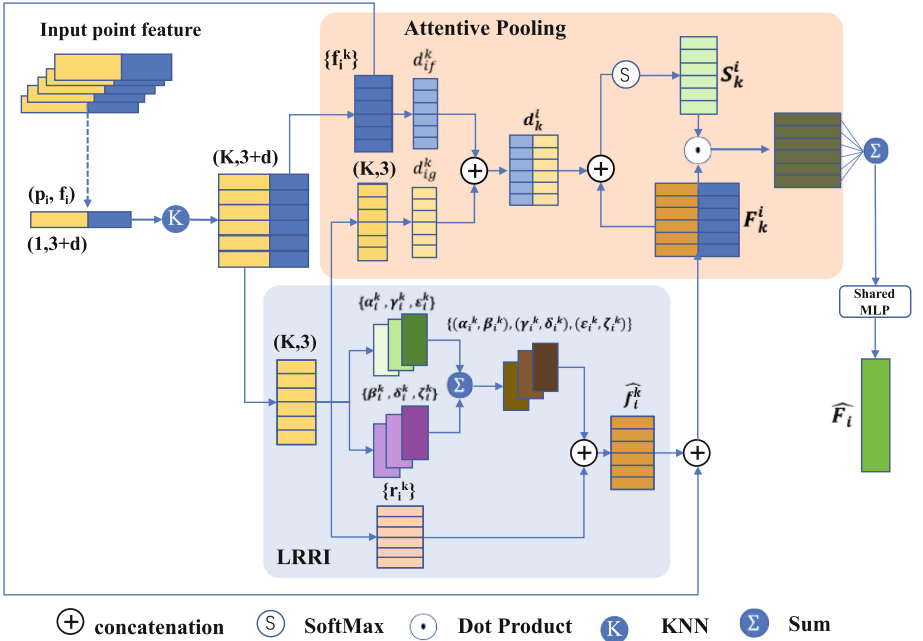


**Fig. 1.** Architecture of the EEP module.

## 3.1   Architecture of EEP Module

The architecture of the EEP module is shown in Fig. 1. Given a point cloud $P$ and the features of each point (including spatial information and intermediate learned features), the local features of each point can be learned efficiently using two blocks, LPPR and AP, as shown in Fig. 1. It shows how the local features of a point are learned and applied to each point in parallel. The local space representation constructed by LPPR with $XYZ$ axis rotation invariance is automatically integrated by AP, and we perform LPPR/AP operation twice for the same point to obtain the information of $K$-squared neighboring points, which can significantly increase the perceptual field of each point and obtain more information. The final output of this module learns the local features with $XYZ$ axis rotation invariance.

**Local Representation of Rotation Invariance (LRRI).** As a geometric object, the learned representation of a point set should be invariant to rotation transformations. Points rotated together should not modify the global point cloud category, nor the segmentation of points. In many real scenes, such as the common chairs in indoor scenes as shown in the figure below, the orientations of objects belonging to the same category are usually different. Further, it can be clearly understood that the same object is not only represented by the rotation invariance of the $Z$-axis (Figs. 2(d)(e)), the $X$-axis and the $Y$-axis also have certain rotation invariance. To address this issue, we propose to learn a rotation-invariant local representation, which utilizes polar coordinates to locally represent individual points, and the overall structure of LRRI is shown in the figure. As shown in Figure, local spatial information is input into the LRRI block and the output is a local representation with rotationally invariant features in the $X$, $Y$, and $Z$ axes. LRRI includes the following steps: Finding neighboring points:  For the point $P_i$, the neighboring points are collected by the $K$-Nearest Neighbors (KNN) algorithm based on the point-by-point Euclidean distance to improve the efficiency of local feature extraction. Representation of local geometric features in two coordinate systems:

(a) Local geometric representation based on polar coordinates: for the nearest $K$ points $P_1, P_2, P_3, \ldots, P_k$ of the center point $P_i$, we use the $X, Y, Z$ of each point (based on the Cartesian coordinate system) to convert to the polar representation of each point, and then subtract the polar representation of the neighboring
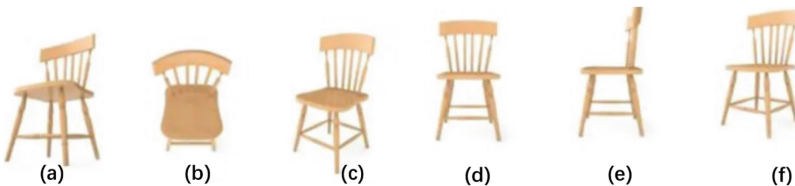


**Fig. 2.** Pictures of the same chair from different angles.

points and the center point to obtain the local geometric representation based on polar coordinates, the specific operation is as follows:

1) Local representation is constant for $Z$-axis rotation:

$$\alpha_i^k = \arctan\left(\frac{y_i^k}{x_i^k}\right) - \arctan\left(\frac{y_{im}}{x_{im}}\right), \tag{1}$$

$$\beta_i^k = \arctan\left(\frac{z_i^k}{\sqrt{x_i^{k^2} + y_i^{k^2}}}\right) - \arctan\left(\frac{z_{im}}{\sqrt{x_{im}^2 + y_{im}^2}}\right). \tag{2}$$

2) Local representation is constant for $X$-axis rotation:

$$\gamma_i^k = \arctan\left(\frac{z_i^k}{y_i^k}\right) - \arctan\left(\frac{z_{im}}{y_{im}}\right), \tag{3}$$

$$\delta_i^k = \arctan\left(\frac{x_i^k}{\sqrt{z_i^{k^2} + y_i^{k^2}}}\right) - \arctan\left(\frac{x_{im}}{\sqrt{z_{im}^2 + y_{im}^2}}\right). \tag{4}$$

3) Local representation is constant for $Y$-axis rotation:

$$\epsilon_i^k = \arctan\left(\frac{x_i^k}{z_i^k}\right) - \arctan\left(\frac{x_{im}}{z_{im}}\right), \tag{5}$$

$$\zeta_i^k = \arctan\left(\frac{y_i^k}{\sqrt{x_i^{k^2} + z_i^{k^2}}}\right) - \arctan\left(\frac{y_{im}}{\sqrt{x_{im}^2 + z_{im}^2}}\right). \tag{6}$$

(b) Relative point position encoding: For each of the nearest $K$ points of the centroid $P_i$, we encode the location of the points as follows:

$$r_i^k = MLP\left(P_i \oplus P_i^k \oplus \left(P_i - P_i^k\right) \oplus \|P_i - P_i^k\|\right). \tag{7}$$

(c) Point Feature Augmentation: The enhanced local geometric representation of a point can be obtained by concatenating the relative position codes of adjacent points and the representation of local geometric features in their corresponding two coordinate systems.

$$\widehat{f_i^k} = MLP\left(\alpha \oplus \beta \oplus \gamma \oplus \delta \oplus \zeta \oplus r_i^k\right). \tag{8}$$

**Attentive Pooling (AP).** In the previous section, we have given the point cloud local geometric feature representation, most of the existing work for aggregating neighboring features uses max/mean pooing, but this approach leads to most of the information loss, we are inspired by SCF-Net [29] network, our attention pooling consists of the following steps:

(a) Calculate the distance: point features and local geometric features generated by the LRRI block, and neighboring point geometric distances are input

to the AP module to learn the contextual features of the local region. We want to express the correlation between points by distance, the closer the distance, the stronger the correlation. Two distances are calculated: geometric distance between points and feature distance between point features:

$$d_{if}^k = \mathrm{mean}(|v(i) - v(k)|), \tag{9}$$

$$\mathbf{d_i^k} = \exp\left(-d_{ig}^k\right) \oplus \lambda \exp\left(-d_{if}^k\right). \tag{10}$$

(b) Calculate the attention score: use a shared MLP to learn the attention score of each feature:

$$\mathbf{A_i^k} = \mathrm{softmax}\left(MLP\left(\mathbf{d_i^k} \oplus \mathbf{f_i^k}\right)\right). \tag{11}$$

(c) Weighted sum: use the learned attention scores to calculate the weighted sum of neighboring point features to learn important local features:

$$F_i^k = MLP\left(f_i^k \oplus \hat{f}_i^k\right), \tag{12}$$

$$\hat{F}_i = \sum_{k=1}^K \left(F_i^k \cdot A_i^k\right). \tag{13}$$

To summarize: given the input point cloud, for the $i$-th point $P_i$, we learn to aggregate the local features of its $K$ nearest points through two blocks LPPI and AP, and generate a feature vector.

## 3.2   Global Feature (GF)

To improve the reliability of segmentation, in addition to learning locally relevant features, we borrowed the GF module from SCF-Net to complement the global features. The relationship between position and volume ratio is used.

$$B_i = \frac{V_i}{V_g} \tag{14}$$

where $B_i$ is the volume of the neighborhood's bounding sphere corresponding to $P_i$, and is the volume of the bounding sphere of the point cloud.

$$\mathbf{f_{iG}} = MLP\left((x_i, y_i, z_i) \oplus B_i\right) \tag{15}$$

The $x$-$y$-$z$ coordinate of $P_i$ is used to represent the location of the local neighborhood. Therefore, the global contextual features are defined as $f_{iG}$.

## 3.3   Architecture of EEP-Net

In this section, we embed the proposed EEP module into the widely used encoder-decoder architecture, resulting in a new network we named EEP-Net, as shown in Fig. 3. The input of the network is a point cloud of size $N \times d$, where $N$ is the number of points and $d$ is the input feature dimension. The point cloud is first fed to a shared MLP layer to extract the features of each point, and the feature dimension is uniformly set to 8. We use five encoder-decoder layers to learn the features of each point, and finally three consecutive fully connected layers and an exit layer are used to predict the semantic label of each point.
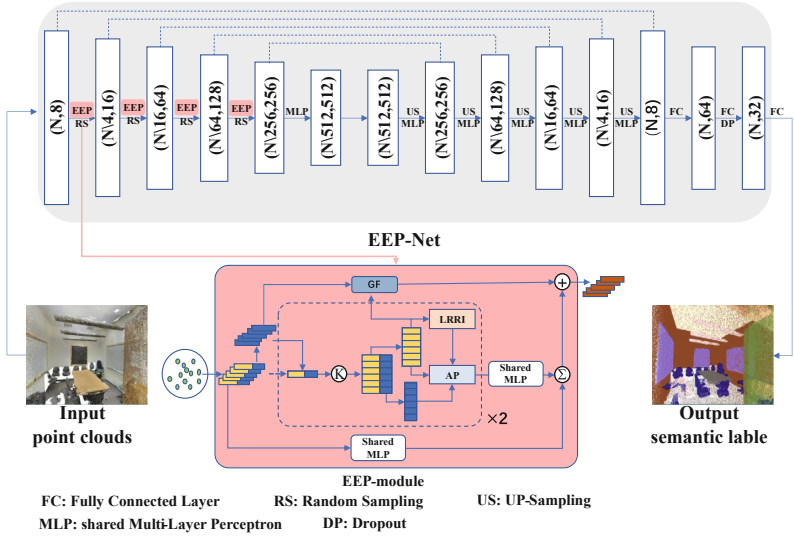
**Fig. 3.** Architecture of the EEP-Net and the EEP module
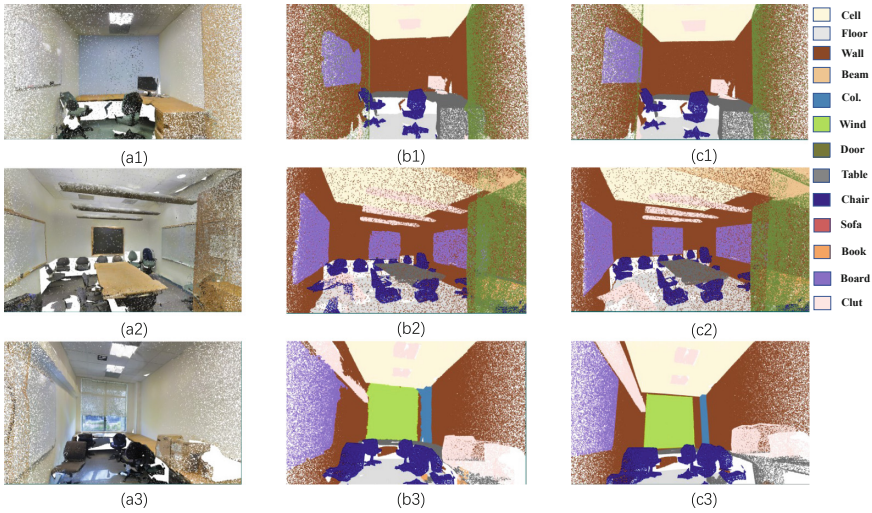
## 4    Experiments

In this section, we evaluate our EEP-Net on a typical indoor field attraction cloud benchmark dataset S3DIS. S3DIS is a large-scale indoor point cloud dataset, which consists of point clouds of 6 areas including 271 rooms. Each point cloud is a medium-sized room, and each point is annotated with one of the semantic labels from 13 classes. Our experiments are performed on Tensorflow (2.1.0). We also report the corresponding results of 8 methods on S3DIS. In addition, after verifying the effectiveness of each module, we focus on ablation experiments on Area 5 of S3DIS.

### 4.1    Evalution on S3DIS Dataset

We performed six cross-validations to evaluate our method, using mIoU as the criterion, the quantitative results of all reference methods are shown in Table 1, our method mIoU outperforms all other methods on this metric, and achieves the best performance on 2 categories, including clut and sofa. Also near the best performance in other categories. Figure 4 shows the visualization results of a typical indoor scene, including an office and a conference room. In generally, the semantic segmentation of indoor scenes is difficult, and the whiteboard on the white wall is easily confused with the white wall itself, but our network can still identify it more accurately.

**Table 1.** Quantitative results of different methods on S3DIS and the classwise metric is IoU (%).

| Methods | mIoU(%) | Ceil | Floor | Wall | Beam | Col | Wind | Door | Table | Chair | Sofa | Book | Board | Clut |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PointNet [1] | 47.6 | 88.0 | 88.7 | 69.3 | 42.4 | 23.1 | 47.5 | 51.6 | 54.1 | 42.0 | 9.6 | 38.2 | 29.4 | 35.2 |
| RSNet [24] | 56.5 | 92.5 | 92.8 | 78.6 | 32.8 | 34.4 | 51.6 | 68.1 | 59.7 | 60.1 | 16.4 | 50.2 | 44.9 | 52.0 |
| SPG [25] | 62.1 | 89.9 | 95.1 | 76.4 | 62.8 | 47.1 | 55.3 | 68.4 | 73.5 | 69.2 | 63.2 | 45.9 | 8.7 | 52.9 |
| PointCNN [26] | 65.4 | 94.8 | 97.3 | 75.8 | 63.3 | 51.7 | 58.4 | 57.2 | 71.6 | 69.1 | 39.1 | 61.2 | 52.2 | 58.6 |
| Pointweb [3] | 66.7 | 93.5 | 94.2 | 80.8 | 52.4 | 41.3 | 64.9 | 68.1 | 71.4 | 67.1 | 50.3 | 62.7 | 62.2 | 58.5 |
| ShellNet [27] | 66.8 | 90.2 | 93.6 | 79.9 | 60.4 | 44.1 | 64.9 | 52.9 | 71.6 | 84.7 | 53.8 | 64.6 | 48.6 | 59.4 |
| RandLA-Net [4] | 70.0 | 93.1 | 96.1 | 80.6 | 62.4 | 48.0 | 64.4 | 69.4 | 67.4 | 76.4 | 60.0 | 64.2 | 65.9 | 60.1 |
| MuGNet [28] | 69.8 | 92.0 | 95.7 | 82.5 | 64.4 | 60.1 | 60.7 | 69.7 | 82.6 | 70.3 | 64.4 | 52.1 | 52.8 | 60.6 |
| Ours | **70.9** | 93.3 | 96.0 | 80.6 | 61.8 | 46.4 | 64.4 | 68.1 | 69.9 | 82.0 | 68.5 | 63.4 | 65.4 | 61.5 |



**Fig. 4.** Visualization examples of three typical indoor scenes on S3DIS. (a) RGB colored input point clouds, (b) Predictions obtained via the proposed EEP-Net, (c) Ground truths.

## 4.2   Ablation Study

The experimental results on the S3DIS dataset validate the effectiveness of our proposed method, and in order to better understand the network, we evaluate it and conduct the following experiments, which will be performed on Area 5, the location of the S3DIS dataset, for this set of experiments. As shown in Fig. 5, it is easy to see that the segmentation result of (d) is obviously closer to the ground truth than (c), which proves that the performance of EEP-Net is better than that of the network containing only *Z*-axis rotation invariance Table 2.

**Table 2.** The mean IoU scores of all ablated networks based on our full EEP-Net.

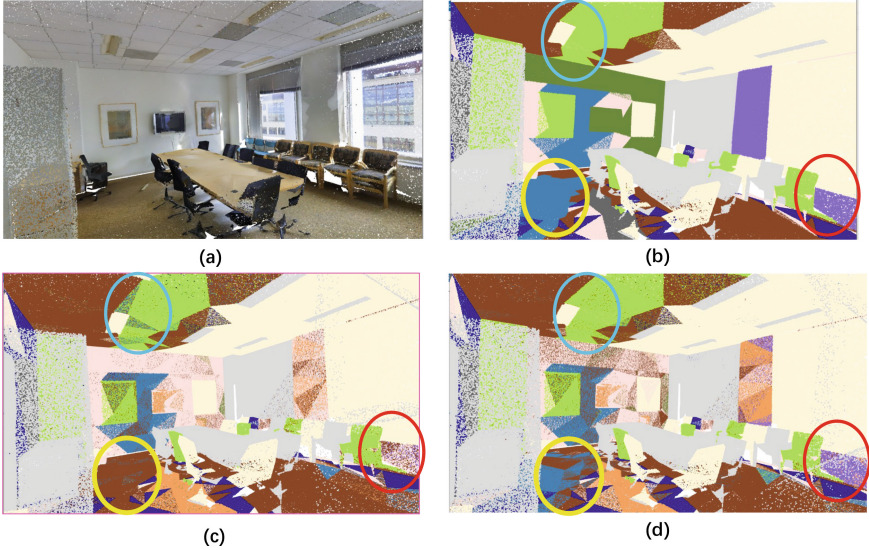|  | mIoU(%) |
|---|---|
| (1) Rotational invariance for $Z$-axis only | 63.04 |
| (2) Local relative feature $+$ $Z$-axis rotation invariance | 63.38 |
| (3) Rotational invariance of $X - Y - Z$ axes | 63.84 |
| (4) Local relative feature $+$ $X - Y - Z$ axis rotation invariance | **64.00** |



**Fig. 5.** Visualization example on S3DIS Area 5, (a) RGB color input point cloud, (b) ground truth, (c) network segmentation result with only $Z$-axis rotation invariance, (d) EEP-Net segmentation result. The circles of the same color are the comparison of segmentation results of the same region.

## 5   Conclusion

In this paper, to better learn local contextual features of point clouds, we propose a new local feature aggregation module EEP, which works by representing point clouds from Cartesian coordinates to polar coordinates of local regions. To verify the effectiveness of the method, we conduct experiments on the representative dataset S3DIS, and compare with eight methods to verify the advanced nature of our method. And we conduct ablation experiments on Area 5 of S3DIS to verify the effectiveness of EEP-Net.

# References

1. Qi, C., Su, H., Mo, K., Guibas, L.J.: PointNet: Deep learning on point sets for 3D classification and segmentation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 77–85 (2017)

2. Qi, C., Yi, L., Su, H., Guibas, L.J.: PointNet++: Deep hierarchical feature learning on point sets in a metric space. In: Conference and Workshop on Neural Information Processing Systems (NIPS) (2017)

3. Zhao, H., Jiang, L., Fu, C.W., Jia, J.: PointWeb: Enhancing local neighborhood features for point cloud processing. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5560–5568 (2019)

4. Hu, Q., et al.: RandLA-Net: Efficient semantic segmentation of large-scale point clouds. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11105–11114 (2020)

5. Lawin, F.J., Danelljan, M., Tosteberg, P., Bhat, G., Khan, F.S., Felsberg, M.: Deep projective 3D semantic segmentation. In: International Conference on Computer Analysis of Images and Patterns (CAIP) (2017)

6. Boulch, A., Saux, B.L., Audebert, N.: Unstructured point cloud semantic labeling using deep segmentation networks. In: Eurographics Workshop on 3D Object Retrieval (3DOR)@Eurographics (2017)

7. Audebert, N., Le Saux, B., Lefèvre, S.: Semantic segmentation of earth observation data using multimodal and multi-scale deep networks. In: Lai, S.-H., Lepetit, V., Nishino, K., Sato, Y. (eds.) ACCV 2016. LNCS, vol. 10111, pp. 180–196. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-54181-5_12

8. Tatarchenko, M., Park, J., Koltun, V., Zhou, Q.Y.: Tangent convolutions for dense prediction in 3D. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3887–3896 (2018)

9. Wu, B., Wan, A., Yue, X., Keutzer, K.: SqueezeSeg: Convolutional neural nets with recurrent CRF for real-time road-object segmentation from 3D Lidar point cloud. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 1887–1893 (2018)

10. Wu, B., Zhou, X., Zhao, S., Yue, X., Keutzer, K.: Squeezesegv 2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a Lidar point cloud. In: International Conference on Robotics and Automation (ICRA), pp. 4376–4382 (2019)

11. Landola, F.N., Moskewicz, M.W., Ashraf, K., Han, S., Dally, W.J., Keutzer, K.: SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1mb model size. ArXiv abs/1602.07360 (2016)

12. Milioto, A., Vizzo, I., Behley, J., Stachniss, C.: RangeNet ++: Fast and accurate Lidar semantic segmentation. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4213–4220 (2019)

13. Meng, H.Y., Gao, L., Lai, Y.K., Manocha, D.: VV-Net: Voxel VAE-Net with group convolutions for point cloud segmentation. In: IEEE/CVF International Conference on Computer Vision (ICCV), pp. 8499–8507 (2019)

14. Rethage, D., Wald, J., Sturm, J., Navab, N., Tombari, F.: Fully-convolutional point networks for large-scale point clouds. ArXiv abs/1808.06840 (2018)

15. Huang, J., You, S.: Point cloud labeling using 3D convolutional neural network. In: 23rd International Conference on Pattern Recognition (ICPR), pp. 2670–2675 (2016)

16. Tchapmi, L.P., Choy, C.B., Armeni, I., Gwak, J., Savarese, S.: SEGCloud: semantic segmentation of 3D point clouds. In: International Conference on 3D Vision (3DV), pp. 537–547 (2017)

17. Shelhamer, E., Long, J., Darrell, T.: Fully convolutional networks for semantic segmentation. In: IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) vol. 39(4), pp. 640–651 (2017)

18. Landrieu, L., Simonovsky, M.: Large-scale point cloud semantic segmentation with superpoint graphs. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4558–4567 (2018)

19. Landrieu, L., Boussaha, M.: Point Cloud oversegmentation with graph structured deep metric learning. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7432–7441 (2019)

20. Wang, L., Huang, Y., Hou, Y., Zhang, S., Shan, J.: Graph attention convolution for point cloud semantic segmentation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10288–10297 (2019)

21. Ye, X., Li, J., Huang, H., Du, L., Zhang, X.: 3D recurrent neural networks with context fusion for point cloud semantic segmentation. In: European Conference on Computer Vision (ECCV), pp. 403–417 (2018)

22. Huang, Q., Wang, W., Neumann, U.: Recurrent slice networks for 3D segmentation of point clouds. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2626–2635 (2018)

23. Liu, F., et al.: 3DCNN-DQN-RNN: A deep reinforcement learning framework for semantic parsing of large-scale 3D point clouds. In: IEEE International Conference on Computer Vision (ICCV), pp. 5679–5688 (2017)

24. Huang, Q., Wang, W., Neumann, U.: Recurrent slice networks for 3D segmentation of point clouds. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2626–2635 (2018)

25. Landrieu, L., Simonovsky, M.: Large-scale point cloud semantic segmentation with superpoint graphs. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4558–4567 (2018)

26. Li, Y., Bu, R., Sun, M., Wu, W., Di, X., Chen, B.: PointCNN: Convolution on x-transformed points. In: Conference and Workshop on Neural Information Processing Systems (NIPS), pp. 820–830 (2018)

27. Zhang, Z., Hua, B.S., Yeung, S.K.: ShellNet: Efficient point cloud convolutional neural networks using concentric shells statistics. In: IEEE/CVF International Conference on Computer Vision (ICCV), pp. 1607–1616 (2019)

28. Xie L., Furuhata T., Shimada K.: Multi-resolution graph neural network for large-scale point cloud segmentation. arXiv preprint arXiv:2009.08924 (2020)

29. Fan, S., Dong, Q., Zhu, F., Lv, Y., Ye, P., Wang, F.: SCF-Net: Learning spatial contextual features for large-scale point cloud segmentation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 14499–14508 (2020)