1. Details of Architectures

In this section, we present the design of RPNet-W (Tab. 1) and RPNet-D (Tab. 2).

2. Detailed Segmentation Results

We show more detailed segmentation results on S3DIS in Tab. 3. Our RPNet-D27 outperforms other methods on the whole accuracy and most of the detailed accuracy. We argue that our model are less precise on some categories since such objects are similar to other shapes and may confuse the model.

3. Visualization

In this section, we present more visualization results. We show some attention maps from our groupwise self-attention on ModelNet40 in Sec. 3.1. To further verify the effectiveness of our RPNet-D, we visualize segmentation labeling with more examples on S3DIS and ScanNet in Sec. 3.2.

3.1. Attention Maps

Shown in Fig. 1, we show more examples of attention maps with different scales of grouping. The edge points are more likely to be important in a relatively simple group (i.e., the desk), while for a complex surface, the important points can be anywhere (i.e., the toilet). This observation is reasonable in the real world. To distinguish a shape, we first focus on its outline. But we will consider its internal structure on a complex object.

3.2. Segmentation Labeling

We demonstrate the precision of our RPNet-D with more visual examples. Shown in Fig. 2 and Fig. 3, we present the labeling results on ScanNet and S3DIS, respectively.

References

[1] Qingyong Hu, Bo Yang, Linhai Xie, Stefano Rosa, Yulan Guo, Zhihua Wang, Niki Trigoni, and Andrew Markham. Randla-net: Efficient semantic segmentation of large-scale point clouds. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 11108–11117, 2020.

[2] Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, and Baoquan Chen. Pointcnn: Convolution on x-transformed points. In Advances in neural information processing systems, pages 820–830, 2018.
| Method       | OA  | mIoU | ceiling | floor | wall | beam | column | window | door | table | chair | sofa | bookcase | board | clutter |
|--------------|-----|------|---------|-------|------|------|--------|--------|------|-------|-------|------|----------|-------|----------|
| PointNet [3] | 78.5| 47.6 | 88.0    | 88.7  | 69.3 | 42.4 | 23.1   | 47.5   | 51.6 | 42.0  | 54.1  | 38.2 | 9.6      | 29.4  | 35.2     |
| PointCNN [2] | 88.1| 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 [5] | 87.3| 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     |
| PointASNL [4]| 88.8| 68.7 | 95.3    | 97.9  | 81.9 | 47.0 | 48.0   | 67.3   | 70.5 | 71.3  | 77.8  | 50.7 | 60.4     | 63.0  | 62.8     |
| RandLA-Net [11]| 88.0| 70.0 | 93.1    | 96.1  | 80.6 | 62.4 | 48.0   | 64.4   | 69.4 | 69.4  | 76.4  | 60.0 | 64.2     | 65.9  | 60.1     |
| RPNet-D27    | **90.1** | **70.8** | **96.1** | **98.5** | **83.4** | **47.2** | **48.7** | **69.9** | **73.0** | **74.7** | **79.1** | 54.6 | **64.3** | **66.8** | **64.1** |

Table 3. Semantic segmentation results on the S3DIS dataset with 6-fold cross validation.

Figure 1. More examples of self-attention maps with the group size of 128 (above of each object) and 32 (below of each object). The blue balls are the center or query points.
Figure 2. More examples of ScanNet dataset.

Figure 3. More examples of S3DIS dataset.
[3] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 652–660, 2017.

[4] Xu Yan, Chaoda Zheng, Zhen Li, Sheng Wang, and Shuguang Cui. Pointasnnl: Robust point clouds processing using nonlocal neural networks with adaptive sampling. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 5589–5598, 2020.

[5] Hengshuang Zhao, Li Jiang, Chi-Wing Fu, and Jiaya Jia. Pointweb: Enhancing local neighborhood features for point cloud processing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5565–5573, 2019.