Supplementary: Recurrent Slice Networks for 3D Segmentation on Point Clouds

Qiangui Huang Weiyue Wang Ulrich Neumann
University of Southern California
Los Angeles, California
{qianguih,weiyuewa,uneumann}@usc.edu

Here, we present more details of experimental settings and more experimental results and discussions.

1. More Results and Discussions on the S3DIS dataset

In the S3DIS dataset, there are 272 indoor scenes captured from 6 areas in 3 buildings. The points are annotated in 13 categories. To process this dataset, our RSNet takes points with 9 dimensional features as inputs as in [5]. The first three, middle three, and last three dimensions represent the xyz coordinates, RGB intensities, and normalized xyz coordinates, respectively.

In the main text, we used the training/testing split in [7] is used to avoid dividing areas from same building to both training and testing sets. However, in [5], the authors reported their performances using 6-fold validation. In order to comprehensively compare with [5], we also present the 6-fold validation performances of RSNet in Table.1. The results show that our RSNet outperforms PointNet by a large margin while requiring less memories and reasonable extra inference times.

Both Table.1 and the Table.1 in the main text show that while all the methods work well on some categories like ceiling, floor and wall, they all fail to achieve the same level of performances on the categories like beam, column, and bookcase. This is because the S3DIS dataset is a highly unbalanced dataset. From the data portion statistics in Table.2 we notice that ceiling, floor and wall are the dominant classes which have $7 \sim 50$ times more training data than the rare classes. This makes the segmentation algorithms fail to generalize well on the rare classes. In order to alleviate this problem, we adopt the median frequency balancing strategy [3] in our RSNet training. The results are compared with the baseline in Table.5. It shows that using median frequency balancing improves performances in terms of the mean accuracy. However, there is a slight decrease in mean IOU.

2. More Results and Discussions on the ScanNet dataset

The ScanNet dataset contain 1,513 scenes captured by the Matterport 3D sensor. We follow the official training/testing split [2] in this paper. The points are annotated in 20 categories and one background class. As shown in Table.3, the ScanNet dataset is also highly unbalanced. Thus, we use the mean IOU and mean accuracy as evaluation metrics in the main text to better measure the performances for this dataset. To process the ScanNet dataset, out RSNet takes points with 3 dimensional features (xyz coordinates) as inputs as in [6].

In order to further improve the performances on the ScanNet dataset, we train a RSNet taking not only xyz coordinates but also RGB intensities as inputs. The results are reported in Table.4. It shows that RGB information can slightly improve the performances of our baseline model. The mean IOU and mean accuracy are improved by 1.81 and 1.97. Moreover, detailed per-class IOUs show that the RGB information is particularly helpful for categories like door, window, and picture. These classes can be easily confused with walls when only geometric information (xyz coordinate) is present. However, RGB information helps the network distinguish them from each other.

References

[1] I. Armeni, O. Sener, A. R. Zamir, H. Jiang, I. Brilakis, M. Fischer, and S. Savarese. 3d semantic parsing of large-scale indoor spaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1534–1543, 2016.

[2] A. Dai, A. X. Chang, M. Savva, M. Halber, T. Funkhouser, and M. Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. *arXiv preprint arXiv:1702.04405*, 2017. 1

[3] D. Eigen and R. Fergus. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2650–2658, 2015. 1

Method	mIOU	mAcc	ceiling	floor	wall	beam	column	window	door	chair	table	bookcase	sofa	board	clutter
PointNet [5]	47.71	-	-	-	-	-	-	-	-	-	-				
Ours	**56.47**	66.45	92.48	92.83	78.56	32.75	34.37	51.62	68.11	59.72	60.13	16.42	50.22	44.85	52.03

Table 1: 6-fold validation results on the Large-Scale 3D Indoor Spaces Dataset (S3DIS). IOU of each category is also reported.

	ceiling	floor	wall	beam	column	window	door	chair	table	bookcase	sofa	board	clutter
Data Per-centage (%)	25.3	23.3	17.3	2.42	1.6	1.1	4.6	3.4	5.3	0.5	3.3	0.7	11.2

Table 2: Data portion of each category in the training set of the S3DIS dataset.

	wall	floor	chair	table	desk	bed	bookshelf	sofa	sink	bathtub
Data Per-centage (%)	36.8	24.9	4.6	2.5	1.7	2.6	2.0	2.6	0.3	0.3
	toilet	curtain	counter	door	window	shower-curtain	refridgerator	picture	cabinet	other furniture
Data Per-centage (%)	0.3	1.5	0.6	2.3	0.9	0.2	0.4	0.4	2.6	2.5

Table 3: Data portion of each category in the training set of the ScanNet dataset.

Method	mIOU	mAcc	wall	floor	chair	table	desk	bed	book-shelf	sofa	sink
RSNet	39.35	48.37	79.23	94.10	**64.99**	**51.04**	34.53	**55.95**	53.02	**55.41**	34.84
RSNet with RGB	**41.16**	**50.34**	**79.38**	**94.21**	63.65	48.67	**35.27**	53.09	**53.67**	51.06	**41.00**
Method	bathtub	toilet	curtain	counter	door	window	shower curtain	refrid-gerator	picture	cabinet	other furniture
RSNet	49.38	54.16	6.78	**22.72**	3.00	8.75	**29.92**	37.90	0.95	**31.29**	18.98
RSNet with RGB	**60.37**	**63.20**	**8.30**	20.90	**15.32**	**15.67**	**24.36**	**39.76**	**4.30**	30.06	**20.98**

Table 4: Results on the ScanNet dataset. IOU of each category is also reported here.

Method	mIOU	mAcc
RSNet	**51.93**	59.42
RSNet-*median*	48.68	**62.09**

Table 5: Results of different training strategies on the S3DIS dataset.

[4] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár. Focal loss for dense object detection. *arXiv preprint arXiv:1708.02002*, 2017.

[5] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. *arXiv preprint arXiv:1612.00593*, 2016. 1, 2

[6] C. R. Qi, L. Yi, H. Su, and L. J. Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *arXiv preprint arXiv:1706.02413*, 2017. 1

[7] L. P. Tchapmi, C. B. Choy, I. Armeni, J. Gwak, and S. Savarese. Segcloud: Semantic segmentation of 3d point clouds. *arXiv preprint arXiv:1710.07563*, 2017. 1

[8] L. Yi, V. G. Kim, D. Ceylan, I. Shen, M. Yan, H. Su, A. Lu, Q. Huang, A. Sheffer, L. Guibas, et al. A scalable active framework for region annotation in 3d shape collections. *ACM Transactions on Graphics (TOG)*, 35(6):210, 2016.