Supplementary Material for JSENet: Joint Semantic Segmentation and Edge Detection Network for 3D Point Clouds

Abstract. This supplementary document is organized as follows:

- Section A explains in more detail about the dataset selection and preparation.
- Section B compares the model sizes and speeds of our network with others.
- Section C provides some qualitative comparison examples and more visualization results on the ScanNet [1] dataset.
- Section D enumerates detailed semantic segmentation results with class scores.

A Dataset selection and preparation.

In the main paper, we conduct all the experiments on two indoor-scene datasets: S3DIS [2] and ScanNet [1]. The main reason for choosing no outdoor-scene dataset is that we find semantic edges are not well defined in existing outdoor-scene datasets. As shown in Fig. 1, compared to the indoor-scene datasets, existing outdoor-scene datasets suffer more from incompletion. Objects in an outdoor-scene are often not densely connected due to the missing parts in the point cloud. Therefore, it is hard to define meaningful semantic edges on these point clouds for our evaluation.

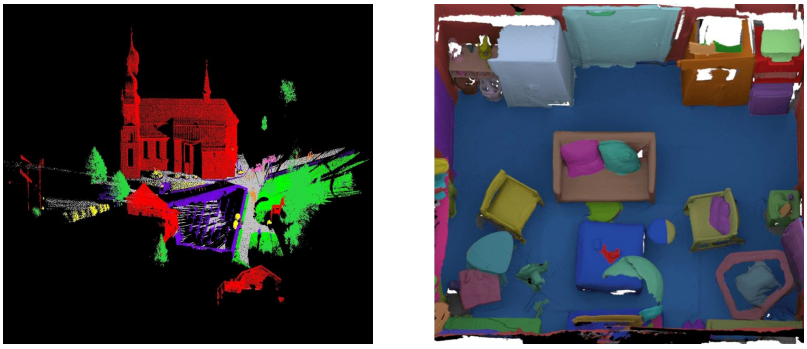

Fig. 1: (Left) An outdoor scene from Semantic3D [3]; (Right) An indoor scene from ScanNet [1]

We generate 3D semantic edges following the idea from 2D works [4, 5] with slight differences. In 2D, thin semantic edges of one or two pixels width are generated. In contrast, we generate thick semantic edges in 3D since points in a point cloud are much sparser than pixels in an image. Moreover, boundaries between an object and the background are considered as semantic edges in 2D images. However, these boundaries are meaningless in the 3D case. Thus, in 3D, we only consider semantic edges between different objects. In general, all semantic edge points will have two or more than two class labels. Since there are unconsidered classes in the ScanNet dataset, semantic edges between a considered class and an unconsidered class might have only one class label.

B Complexity of the network, in comparison with other works.

In this section, we present the comparison on the complexity of our network against state-of-the-art methods. All the experiments have been conducted on a PC with 8 Intel(R) i7-7700 CPUs and a single GeForce GTX 1080Ti GPU.

Training. We train KPConv and JSENet on the ScanNet dataset. Using the setting presented in their paper, KPConv takes about 0.7s for one training iteration and converges in 160K iterations, taking about 31h in total. Using the setting presented in our paper, in the first step, JSENet takes about 0.9s for one training iteration and converges in 170K iterations. In the second step, JSENet takes about 0.6s for one training iteration and converges in 40K iterations. The whole training takes about 49h.

Table 1: Comparison on runtime complexity of JSENet against state-of-the-art methods.

Method	Average time (s)	Parameters (m)
KPConv [6]	0.044	14.1
PointConv [7]	0.307	21.7
MinkowskiNet [8]	0.185	37.9
JSENet	0.097	16.2

Inference. We compare KPConv, PointConv, MinkowskiNet, and JSENet for their runtime complexity given the same sets of points extracted from the ScanNet dataset (13000 points each). Results are shown in Table. 1. It can be seen that for both the inference time and the parameter size, JSENet is largely comparable to KPConv and is both more efficient and compact than PointConv (another recent point-based method) and MinkowskiNet (the SOTA voxel-based method).

C Qualitative Visualization.

In this section, we present more visualization results. More visualization results of our method on the ScanNet dataset are shown in Fig. 2. Qualitative comparison on the effects of joint refinement are shown in Fig. 3 and Fig. 4. Black points in the GT SSP masks are unlabeled points or points of unconsidered classes. All semantic edges are thickened for visualization.

Point Cloud GT-SSP Mask Pred-SSP Mask GT-SEP Map Pred-SEP Map

Fig. 2: Qualitative results on ScanNet val set.

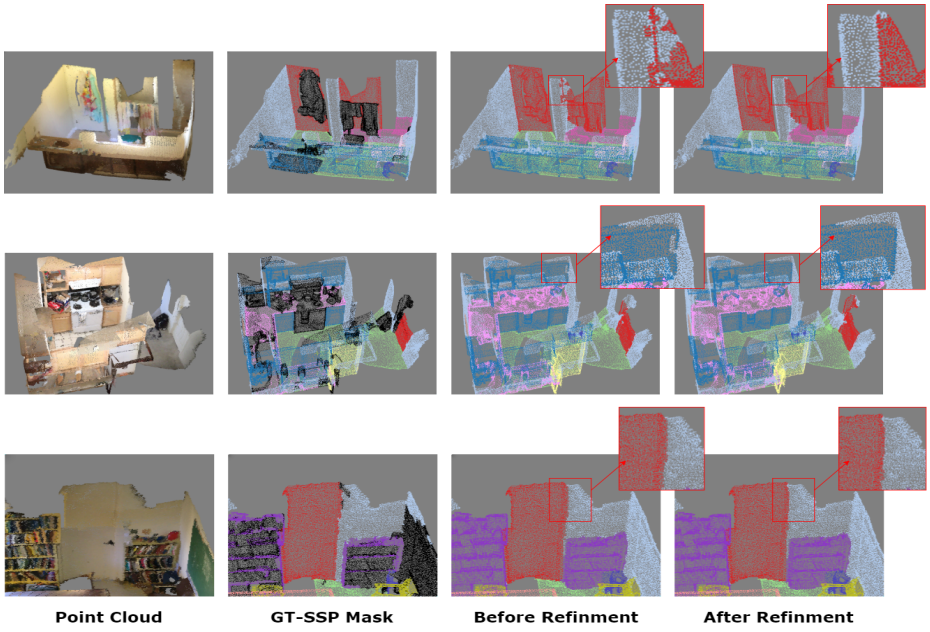

Fig. 3: Some visualization comparison examples for semantic segmentation before and after joint refinement (best viewed in color).

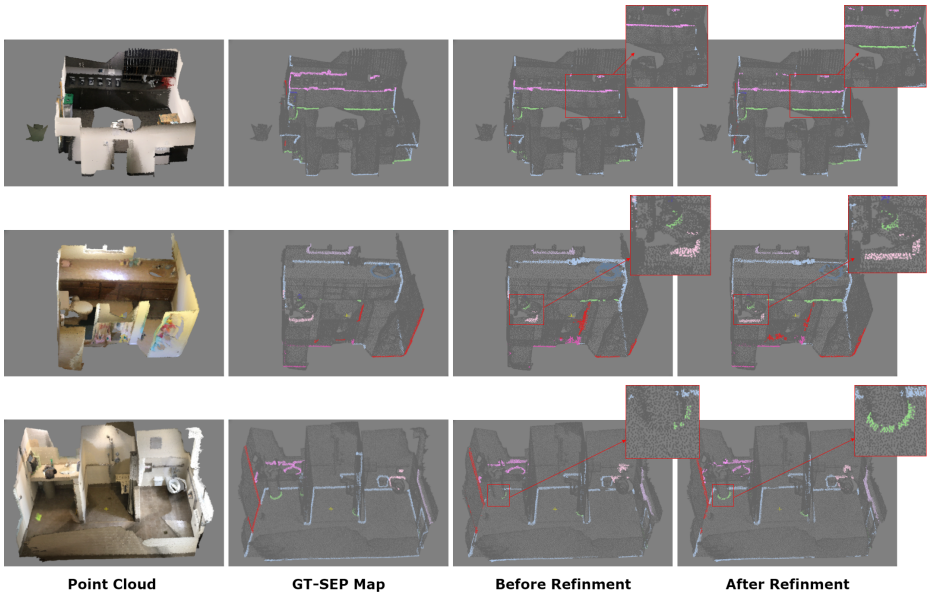

Fig. 4: Some visualization comparison examples for semantic edge detection before and after joint refinement (best viewed in color). For better visualization, we thickened all the semantic edges.

D Detailed semantic segmentation results.

In this section, we provide more details on our semantic segmentation experiments, for benchmarking purpose with future works. Detailed class scores for the S3DIS dataset and the ScanNet dataset are presented in Table 2 and Table 3, respectively.

Table 2: Detailed mIoU scores (%) of semantic segmentation on S3DIS Area-5.

Method	mIoU	ceil.	floor	wall	beam	col.	wind.	door	chair	table	book.	sofa	board	clut.
Pointnet [9]	41.1	88.8	97.3	69.8	0.1	3.9	46.3	10.8	52.6	58.9	40.3	5.9	26.4	33.2
SegCloud [10]	48.9	90.1	96.1	69.9	0.0	18.4	38.4	23.1	75.9	70.4	58.4	40.9	13.0	41.6
Eff 3D Conv [11]	51.8	79.8	93.9	69.0	0.2	28.3	38.5	48.3	71.1	73.6	48.7	59.2	29.3	33.1
TangentConv [12]	52.6	90.5	97.7	74.0	0.0	20.7	39.0	31.3	69.4	77.5	38.5	57.3	48.8	39.8
RNN Fusion [13]	53.4	**95.2**	98.6	77.4	**0.8**	9.8	52.7	27.9	78.3	76.8	27.4	58.6	39.1	51.0
PointCNN [14]	57.3	92.3	98.2	79.4	0.0	17.6	22.8	62.1	74.4	80.6	31.7	66.7	62.1	56.7
SPGraph [15]	58.0	89.4	96.9	78.1	0.0	**42.8**	48.9	61.6	84.7	75.4	69.8	52.6	2.1	52.2
ParamConv [16]	58.3	92.3	96.2	75.9	0.3	6.0	**69.5**	63.5	66.9	65.6	47.3	68.9	59.1	46.2
SPH3D-GCN [17]	59.5	93.3	97.1	81.1	0.0	33.2	45.8	43.8	79.7	86.9	33.2	71.5	54.1	53.7
HPEIN [18]	61.9	91.5	98.2	81.4	0.0	23.3	65.3	40.0	75.5	**87.7**	58.5	67.8	65.6	49.4
MinkowskiNet [8]	65.4	91.8	**98.7**	**86.2**	0.0	34.1	48.9	62.4	89.8	81.6	74.9	47.2	**74.4**	58.6
KPConv rigid [6]	65.4	92.6	97.3	81.4	0.0	16.5	54.5	69.5	**90.1**	80.2	74.6	66.4	63.7	58.1
JSENet (ours)	**67.7**	93.8	97.0	83.0	0.0	23.2	61.3	**71.6**	89.9	79.8	**75.6**	**72.3**	72.7	**60.4**

Table 3: Detailed mIoU scores (%) of semantic segmentation on ScanNet test set.

Method	mIoU	bath	bed	bksf	cab	chair	cntr	curt	desk	door	floor	othr	pic	ref	show	sink	sofa	tab	toil	wall	wind
ScanNet [1]	30.6	20.3	36.6	50.1	31.1	52.4	21.1	0.2	34.2	18.9	78.6	14.5	10.2	24.5	15.2	31.8	34.8	30.0	46.0	43.7	18.2
PointNet++ [19]	33.9	58.4	47.8	45.8	25.6	36.0	25.0	24.7	27.8	26.1	67.7	18.3	11.7	21.2	14.5	36.4	34.6	23.2	54.8	52.3	25.2
SPLATNet [20]	39.3	47.2	51.1	60.6	31.1	65.6	24.5	40.5	32.8	19.7	92.7	22.7	0 0	0.1	24.9	27.1	51.0	38.3	59.3	69.9	26.7
TangentConv [12]	43.8	43.7	64.6	47.4	36.9	64.5	35.3	25.8	28.2	27.9	91.8	29.8	14.7	28.3	29.4	48.7	56.2	42.7	61.9	63.3	35.2
PointCNN [14]	45.8	57.7	61.1	35.6	32.1	71.5	29.9	37.6	32.8	31.9	94.4	28.5	16.4	21.6	22.9	48.4	54.5	45.6	75.5	70.9	47.5
PanopticFusion [21]	52.9	49.1	68.8	60.4	38.6	63.2	22.5	70.5	43.4	29.3	81.5	34.8	24.1	49.9	66.9	50.7	64.9	44.2	79.6	60.2	56.1
TextureNet [22]	56.6	67.2	66.4	67.1	49.4	71.9	44.5	67.8	41.1	39.6	93.5	35.6	22.5	41.2	53.5	56.5	63.6	46.4	79.4	68.0	56.8
SPH3D-GCN [17]	61.0	85.8	77.2	48.9	53.2	79.2	40.4	64.3	57.0	50.7	93.5	41.4	4.6	51.0	70.2	60.2	70.5	54.9	85.9	77.3	53.4
HPEIN [18]	61.8	72.9	66.8	64.7	59.7	76.6	41.4	68.0	52.0	52.5	94.6	43.2	21.5	49.3	59.9	63.8	61.7	57.0	89.7	80.6	60.5
KP-FCNN [6]	68.4	84.7	75.8	78.4	64.7	81.4	47.3	77.2	60.5	59.4	93.5	45.0	18.1	58.7	80.5	69.0	78.5	61.4	88.2	81.9	63.2
SparseCOnvNet [23]	72.5	64.7	82.1	84.6	72.1	86.9	53.3	75.4	60.3	61.4	95.5	57.2	32.5	71.0	87.0	72.4	82.3	62.8	93.4	86.5	68.3
MinkowskiNet [8]	73.6	85.9	81.8	83.2	70.9	84.0	52.1	85.3	66.0	64.3	95.1	54.4	28.6	73.1	89.3	67.5	77.2	68.3	87.4	85.2	72.7
JSENet (ours)	69.9	88.1	76.2	82.1	66.7	80.0	52.2	79.2	61.3	60.7	93.5	49.2	20.5	57.6	85.3	69.1	75.8	65.2	87.2	82.8	64.9

References

1. Dai, A., Chang, A.X., Savva, M., Halber, M., Funkhouser, T., Nießner, M.: Scannet: Richly-annotated 3d reconstructions of indoor scenes. In: Proc. Computer Vision and Pattern Recognition (CVPR), IEEE. (2017)
2. Armeni, I., Sener, O., Zamir, A.R., Jiang, H., Brilakis, I., Fischer, M., Savarese, S.: 3d semantic parsing of large-scale indoor spaces. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. (2016)
3. Hackel, T., Savinov, N., Ladicky, L., Wegner, J.D., Schindler, K., Pollefeys, M.: SEMANTIC3D.NET: A new large-scale point cloud classification benchmark. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Volume IV-1-W1. (2017) 91–98
4. Yu, Z., Feng, C., Liu, M.Y., Ramalingam, S.: Casenet: Deep category-aware semantic edge detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2017) 5964–5973
5. Prasad, M., Zisserman, A., Fitzgibbon, A., Kumar, M.P., Torr, P.H.: Learning class-specific edges for object detection and segmentation. In: Computer Vision, Graphics and Image Processing. Springer (2006) 94–105
6. Thomas, H., Qi, C.R., Deschaud, J.E., Marcotegui, B., Goulette, F., Guibas, L.J.: Kpconv: Flexible and deformable convolution for point clouds. In: Proceedings of the IEEE International Conference on Computer Vision. (2019) 6411–6420
7. Wu, W., Qi, Z., Fuxin, L.: Pointconv: Deep convolutional networks on 3d point clouds. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2019) 9621–9630
8. Choy, C., Gwak, J., Savarese, S.: 4d spatio-temporal convnets: Minkowski convolutional neural networks. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (Jun 2019)
9. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2017) 652–660
10. Tchapmi, L., Choy, C., Armeni, I., Gwak, J., Savarese, S.: Segcloud: Semantic segmentation of 3d point clouds. In: 2017 international conference on 3D vision (3DV), IEEE (2017) 537–547
11. Zhang, C., Luo, W., Urtasun, R.: Efficient convolutions for real-time semantic segmentation of 3d point clouds. In: 2018 International Conference on 3D Vision (3DV), IEEE (2018) 399–408
12. Tatarchenko, M., Park, J., Koltun, V., Zhou, Q.Y.: Tangent convolutions for dense prediction in 3d. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2018) 3887–3896
13. Ye, X., Li, J., Huang, H., Du, L., Zhang, X.: 3d recurrent neural networks with context fusion for point cloud semantic segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV). (2018) 403–417
14. Li, Y., Bu, R., Sun, M., Wu, W., Di, X., Chen, B.: Pointcnn: Convolution on x-transformed points. In: Advances in neural information processing systems. (2018) 820–830
15. Landrieu, L., Simonovsky, M.: Large-scale point cloud semantic segmentation with superpoint graphs. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2018) 4558–4567
16. Wang, S., Suo, S., Ma, W.C., Pokrovsky, A., Urtasun, R.: Deep parametric continuous convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2018) 2589–2597

17. Lei, H., Akhtar, N., Mian, A.: Spherical kernel for efficient graph convolution on 3d point clouds. arXiv preprint arXiv:1909.09287 (2019)
18. Jiang, L., Zhao, H., Liu, S., Shen, X., Fu, C.W., Jia, J.: Hierarchical point-edge interaction network for point cloud semantic segmentation. In: Proceedings of the IEEE International Conference on Computer Vision. (2019) 10433–10441
19. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In: Advances in neural information processing systems. (2017) 5099–5108
20. Su, H., Jampani, V., Sun, D., Maji, S., Kalogerakis, E., Yang, M.H., Kautz, J.: Splatnet: Sparse lattice networks for point cloud processing. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (Jun 2018)
21. Narita, G., Seno, T., Ishikawa, T., Kaji, Y.: Panopticfusion: Online volumetric semantic mapping at the level of stuff and things. arXiv preprint arXiv:1903.01177 (2019)
22. Huang, J., Zhang, H., Yi, L., Funkhouser, T., Nießner, M., Guibas, L.J.: Texturenet: Consistent local parametrizations for learning from high-resolution signals on meshes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2019) 4440–4449
23. Graham, B., Engelcke, M., van der Maaten, L.: 3d semantic segmentation with submanifold sparse convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2018) 9224–9232