Supplementary Material for ShellNet: Efficient Point Cloud Convolutional Neural Networks using Concentric Shells Statistics

Zhiyuan Zhang[1] Binh-Son Hua[2] Sai-Kit Yeung[3]

[1]Singapore University of Technology and Design [2]The University of Tokyo
[3]Hong Kong University of Science and Technology

Abstract

In this supplemental document, we demonstrate more detailed experiments to evaluate the design choice of ShellConv and ShellNet [17]. Particularly, we demonstrate the use of additional statistics when learning features in ShellConv, which we found that maxpooling alone is giving the best performance while other statistics provide very good results. We also further discuss the neighbor sampling used in ShellConv. Next, we present the per-class comparison for object classification on ModelNet40 dataset [14], object part segmentation on ShapeNet [3], semantic segmentation for indoor scenes on S3DIS dataset [1], and semantic segmentation for outdoor scenes on Semantic3D dataset [4]. Finally, we show more qualitative results for part and semantic segmentations on ShapeNet dataset [15], S3DIS dataset [1], Semantic3D dataset [4], respectively.

1. Additional Statistics

We show that it is possible to use other types of statistics for feature learning and still achieve competitive performance. For example, we can use average pooling to aggregate the mean feature vectors instead of maxpool features. The results are shown in Table 1.

We can see that using average pooling leads to over 92% classification accuracy, which is competitive but not as good as maxpooling (as in our main paper). Therefore, we simply use a maxpooling to aggregate statistics in ShellConv.

statistics	max	mean	max+mean
Accuracy	93.1	92.4	92.5

Table 1: Classification accuracy (%) with different statistics. Empirically, all statistics works very well, but maxpooling gives a slightly better results.

2. Further Discussion on Neighbor Sampling

In ShellConv, we define the shell size (ss) as the number of points contained in each shell which is fixed in this work. For each representative point, we construct the shells within the local neighborhood where the neighborhood size increases arithmetically (e.g. 16 for 1 shell, 32 for 2 shells, 64 for 4 shells, $ss = 16$). Intuitively, such design may be sensitive to the point distribution. When the points are severe non-uniform, the shells will deviate from equidistant. Thus, in the paper, we assumed a more or less uniform distribution of input points. However, in implementation we set a larger ss to cover larger local area such that small varying density will not much affect the performance. An alternative sampling scheme is by fixed shell radius that the shells are equidistant and the number of points contained in each shell will become different when the points are non-uniform. We also test this scheme but find no improvement. This can be explained in two aspects. First, the points of the testing datasets are approximately uniform distributed. Second, non-uniform distribution also affects the statistic in each shell. Thus, we choose fixed ss throughout this paper for the sake of easy implementation.

3. Per-class Accuracy on ModelNet40

We further evaluate the accuracy of the classification task per object class. The mean accuracy over the classes and the per-class accuracies are shown in Table 2 and Table 3. Comparing to the results by PointNet [11], PointNet++ [12], PointCNN [10], and Pointwise [6], we see that our method obtains a superior performance on most classes. In total, our method achieves 1st (in bold) in 25/40 classes, while PointNet [11], PointNet++ [12], PointCNN [10], and Pointwise [6] has 10, 13, 14, and 3 classes achieving 1st, respectively.

Method	mA
PointNet [11]	86.2
PointNet++ [12]	88.1
PointCNN [10]	88.1
Pointwise [6]	81.4
Ours	**89.2**

Table 2: Comparisons of mean per-class accuracy (mA, %) on ModelNet40.

4. Per-class Accuracy on ShapeNet

Please see Table 4.

5. Per-class Accuracy on S3DIS

Please see Table 5.

6. Per-class Accuracy on Semantic3D

In this section, we show the per-class accuracy on Semantic3D benchmark in Table 6. ShellNet ranks 2nd and achieves competitive performance on many classes. Unlike existing methods that employ postprocessing such as CRF, our method uses pure coordinates as input.

7. Visualizations of Part Segmentation

The part segmentation experiment is conducted on the ShapeNet dataset [15] which comprises 16 different categories. For each category, we randomly choose an object and show the qualitative segmentation results in Figure 1, Figure 2, Figure 3 and Figure 4.

8. Visualizations of Indoor Semantic Segmentation

We plot more qualitative segmentation results in Figure 5 and Figure 6

9. Visualizations of Outdoor Semantic Segmentation

We evaluated the performance on the reduced set of 4 subsampled scans in the Semantic3D dataset [4]. In the paper, we showed the qualitative results on two scans. Here, we visualize the segmentation results on the other two scenes in Figure 7.

References

[1] Iro Armeni, Ozan Sener, Amir R Zamir, Helen Jiang, Ioannis Brilakis, Martin Fischer, and Silvio Savarese. 3d semantic parsing of large-scale indoor spaces. In *Computer Vision and Pattern Recognition*, pages 1534–1543, 2016.

[2] Alexandre Boulch, Bertrand Le Saux, and Nicolas Audebert. Unstructured point cloud semantic labeling using deep segmentation networks. In *Eurographics Workshop on 3D Object Retrieval*, volume 2, page 7, 2017.

[3] Angel X. Chang, Thomas A. Funkhouser, Leonidas J. Guibas, Pat Hanrahan, Qi-Xing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. Shapenet: An information-rich 3d model repository. *CoRR*, abs/1512.03012, 2015.

[4] Timo Hackel, Nikolay Savinov, Lubor Ladicky, Jan D Wegner, Konrad Schindler, and Marc Pollefeys. Semantic3d.net: A new large-scale point cloud classification benchmark. In *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 91–98, 2017.

[5] Timo Hackel, Jan D Wegner, and Konrad Schindler. Fast semantic segmentation of 3d point clouds with strongly varying density. *ISPRS annals of the photogrammetry, remote sensing and spatial information sciences*, 3(3):177–184, 2016.

[6] Binh-Son Hua, Minh-Khoi Tran, and Sai-Kit Yeung. Pointwise convolutional neural network. In *Computer Vision and Pattern Recognition*, pages 984–993, 2018.

[7] Qiangui Huang, Weiyue Wang, and Ulrich Neumann. Recurrent slice networks for 3d segmentation on point clouds. In *Computer Vision and Pattern Recognition*, pages 2626–2635, 2018.

[8] Loic Landrieu and Martin Simonovsky. Large-scale point cloud semantic segmentation with superpoint graphs. In *Computer Vision and Pattern Recognition*, pages 4558–4567, 2018.

[9] Felix Järemo Lawin, Martin Danelljan, Patrik Tosteberg, Goutam Bhat, Fahad Shahbaz Khan, and Michael Felsberg. Deep projective 3d semantic segmentation. In *International Conference on Computer Analysis of Images and Patterns*, pages 95–107, 2017.

[10] 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.

[11] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Computer Vision and Pattern Recognition*, pages 652–660, 2017.

[12] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in Neural Information Processing Systems*, pages 5105–5114, 2017.

[13] Lyne Tchapmi, Christopher Choy, Iro Armeni, JunYoung Gwak, and Silvio Savarese. Segcloud: Semantic segmentation of 3d point clouds. In *International Conference on 3D Vision*, pages 537–547, 2017.

[14] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Computer Vision and Pattern Recognition*, pages 1912–1920, 2015.

Network	airplane	bathtub	bed	bench	bookshelf	bottle	bowl	car	chair	cone
PointNet [11]	**100.0**	80.0	94.0	75.0	93.0	94.0	**100.0**	97.9	96.0	**100.0**
PointNet++ [12]	**100.0**	92.0	96.0	75.0	94.0	96.0	90.0	**99.0**	96.0	95.0
PointCNN [10]	**100.0**	92.0	99.0	**80.0**	**100.0**	96.0	90.0	**99.0**	97.0	95.0
Pointwise [6]	**100.0**	82.0	93.0	68.4	91.8	93.9	95.0	95.6	96.0	80.0
Ours	**100.0**	**96.0**	**100.0**	65.0	97.0	**99.0**	95.0	**99.0**	**100.0**	90.0
	cup	curtain	desk	door	dresser	flower pot	glass box	guitar	keyboard	lamp
PointNet[11]	70.0	85.0	79.0	**95.0**	65.1	30.0	94.0	**100.0**	**100.0**	**90.0**
PointNet++ [12]	70.0	90.0	88.4	**95.0**	70.9	25.0	95.0	**100.0**	**100.0**	85.0
PointCNN [10]	60.0	**95.0**	86.0	90.0	**87.2**	**35.0**	92.0	99.0	95.0	85.0
Pointwise [6]	60.0	80.0	76.7	75.0	67.4	10.0	80.8	98.0	**100.0**	83.3
Ours	**75.0**	**95.0**	**90.7**	80.0	**87.2**	5.0	**96.0**	**100.0**	95.0	85.0
	laptop	mantel	monitor	night stand	person	piano	plant	radio	range hood	sink
PointNet [11]	**100.0**	96.0	95.0	82.6	85.0	88.8	73.0	70.0	91.0	80.0
PointNet++ [12]	**100.0**	97.0	99.0	72.1	90.0	**96.0**	79.0	75.0	**97.0**	**90.0**
PointCNN [10]	**100.0**	94.0	**100.0**	81.4	**95.0**	94.0	83.0	**85.0**	94.0	70.0
Pointwise [6]	95.0	93.9	92.9	70.2	89.5	84.5	78.8	65.0	88.9	65.0
Ours	**100.0**	**99.0**	**100.0**	**83.7**	**95.0**	**96.0**	**96.0**	80.0	93.0	80.0
	sofa	stairs	stool	table	tent	toilet	tv stand	vase	wardrobe	xbox
PointNet [11]	96.0	85.0	**90.0**	88.0	**95.0**	99.0	87.0	78.8	60.0	70.0
PointNet++ [12]	98.0	**95.0**	80.0	83.0	**95.0**	**100.0**	89.0	80.0	**80.0**	75.0
PointCNN [10]	98.0	90.0	70.0	87.0	**95.0**	**100.0**	90.0	**82.0**	65.0	80.0
Pointwise [6]	96.0	80.0	83.3	**90.9**	90.0	94.9	84.5	81.3	30.0	75.0
Ours	**100.0**	**95.0**	85.0	83.0	**95.0**	**100.0**	**91.0**	80.0	75.0	**90.0**

Table 3: Per-class accuracy of object classification on the ModelNet40 dataset.

Network	aero	bag	cap	car	chair	earphone	guitar	knife	lamp	laptop	motor	mug	pistol	rocket	skate	table
PointNet [11]	83.4	78.7	82.5	74.9	89.6	73.0	91.5	85.9	80.8	95.3	65.2	93.0	81.2	57.9	72.8	80.6
PointNet++ [12]	82.4	79.0	87.7	77.3	**90.8**	71.8	91.0	85.9	83.7	95.3	71.6	94.1	81.3	58.7	76.4	82.6
ResNet [7]	82.7	86.4	84.1	78.2	90.4	69.3	91.4	87.0	83.5	95.4	66.0	92.6	81.8	56.1	75.8	82.2
SynNet [16]	81.6	81.7	81.9	75.2	90.2	74.9	**93.0**	86.1	84.7	95.6	66.7	92.7	81.6	60.6	**82.9**	82.1
PointCNN [10]	84.1	**86.5**	86.0	**80.8**	90.6	**79.7**	92.3	**88.4**	**85.3**	**96.1**	**77.2**	**95.3**	**84.2**	64.2	80.0	**83.0**
Ours	**84.3**	79.6	**88.9**	79.1	90.0	79.4	91.3	85.9	**82.3**	95.4	68.6	94.9	82.7	61.5	79.7	81.7

Table 4: Per-class accuracy of object part segmentation on the ShapeNet dataset.

[15] Li Yi, Vladimir G. Kim, Duygu Ceylan, I-Chao Shen, Mengyan Yan, Hao Su, Cewu Lu, Qixing Huang, Alla Sheffer, and Leonidas Guibas. A scalable active framework for region annotation in 3d shape collections. *ACM Transactions on Graphics*, 2016.

[16] Li Yi, Hao Su, Xingwen Guo, and Leonidas J Guibas. Syncspeccnn: Synchronized spectral cnn for 3d shape segmentation. In *Computer Vision and Pattern Recognition*, pages 2282–2290, 2017.

[17] Zhiyuan Zhang, Binh-Son Hua, and Sai-Kit Yeung. Shellnet: Efficient point cloud convolutional neural networks using concentric shells statistics. In *International Conference on Computer Vision (ICCV)*, 2019.

Network	OA	mIoU	ceiling	floor	wall	beam	column	window	door	table	chair	sofa	bookcase	board	clutter
PointNet [11]	78.5	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
SPG [8]	85.5	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
ResNet [7]	-	56.5	92.5	92.8	78.6	32.8	34.4	51.6	68.1	60.1	59.7	50.2	16.4	44.9	52.0
PointCNN [10]	**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
Ours	87.1	**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**

Table 5: Segmentation results on the S3DIS [1] dataset in overall accuracy (OA, %), micro-averaged IoU (mIoU, %) and per-class IoU.

Method	OA	mIoU	man-made terrain	natural terrain	high vegetation	low vegetation	buildings	hard-scape	scanning artefact	cars
TMLC-MSR [5]	86.2	54.2	89.8	74.5	53.7	26.8	88.8	18.9	36.4	44.7
DeePr3SS [9]	88.9	58.5	85.6	83.2	74.2	32.4	89.7	18.5	25.1	59.2
SnapNet [2]	88.6	59.1	82.0	77.3	79.7	22.9	91.1	18.4	37.3	64.4
SegCloud [13]	88.1	61.3	83.9	66.0	86.0	40.5	91.1	30.9	27.5	64.3
SPG [8]	94.0	73.2	97.4	92.6	87.9	44.0	93.2	31.0	63.5	76.2
ShellNet (Ours)	93.7	69.4	96.5	92.2	84.1	40.4	94.5	35.5	44.9	67.4

Table 6: Segmentation results on the Semantic3D [4] dataset in overall accuracy (OA, %), micro-averaged IoU (mIoU, %) and per-class IoU.

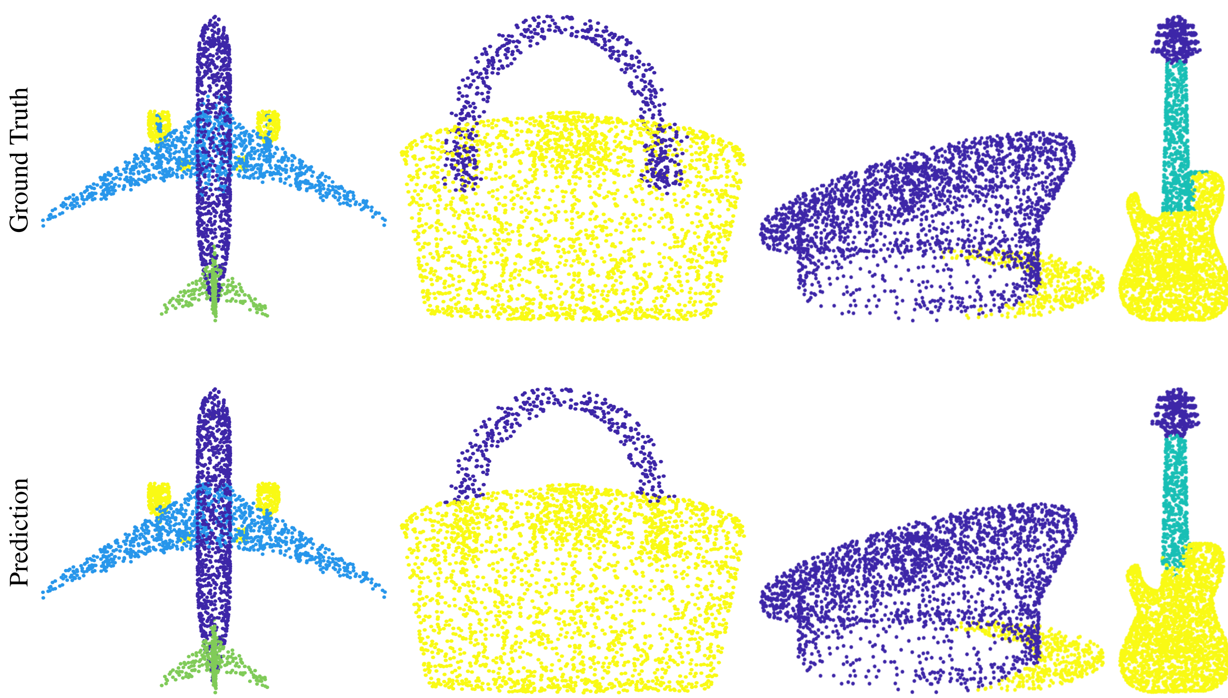

Figure 1: Part segmentation plane, bag, cap, and guitar objects.

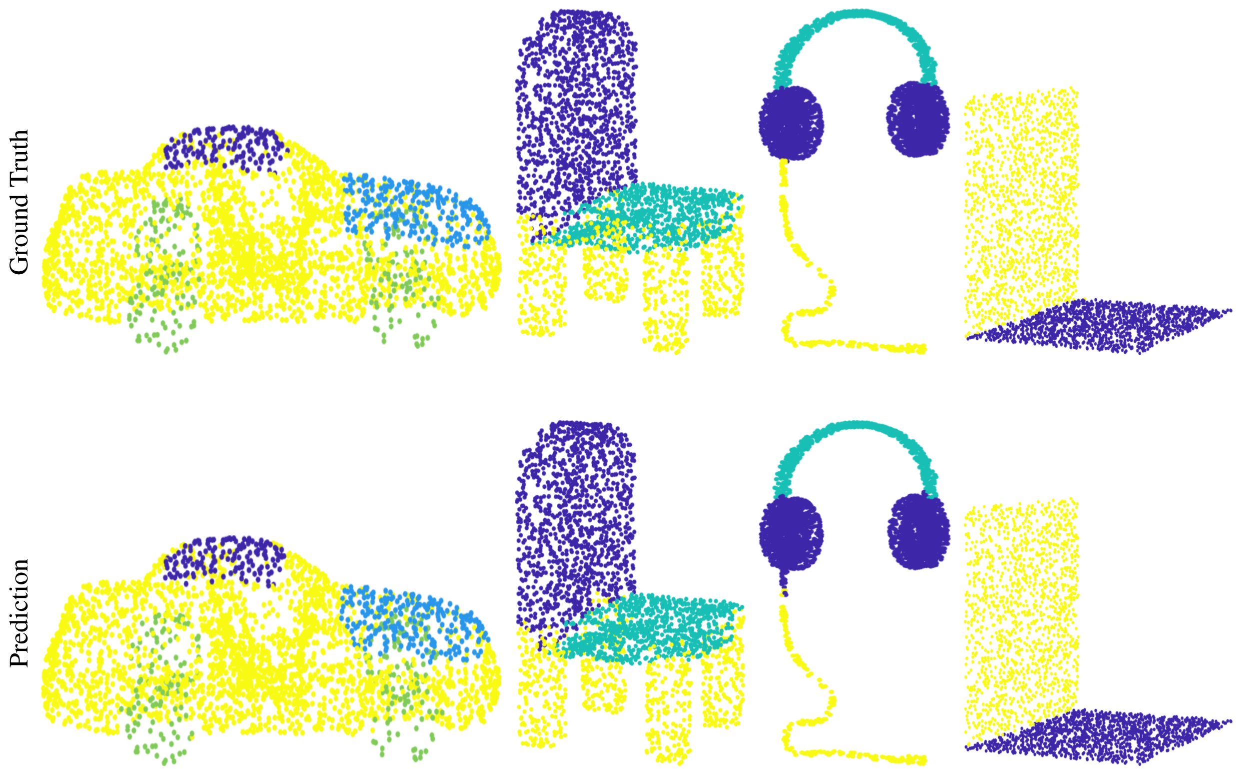

Figure 2: Part segmentation car, chair, earphone, and laptop objects.

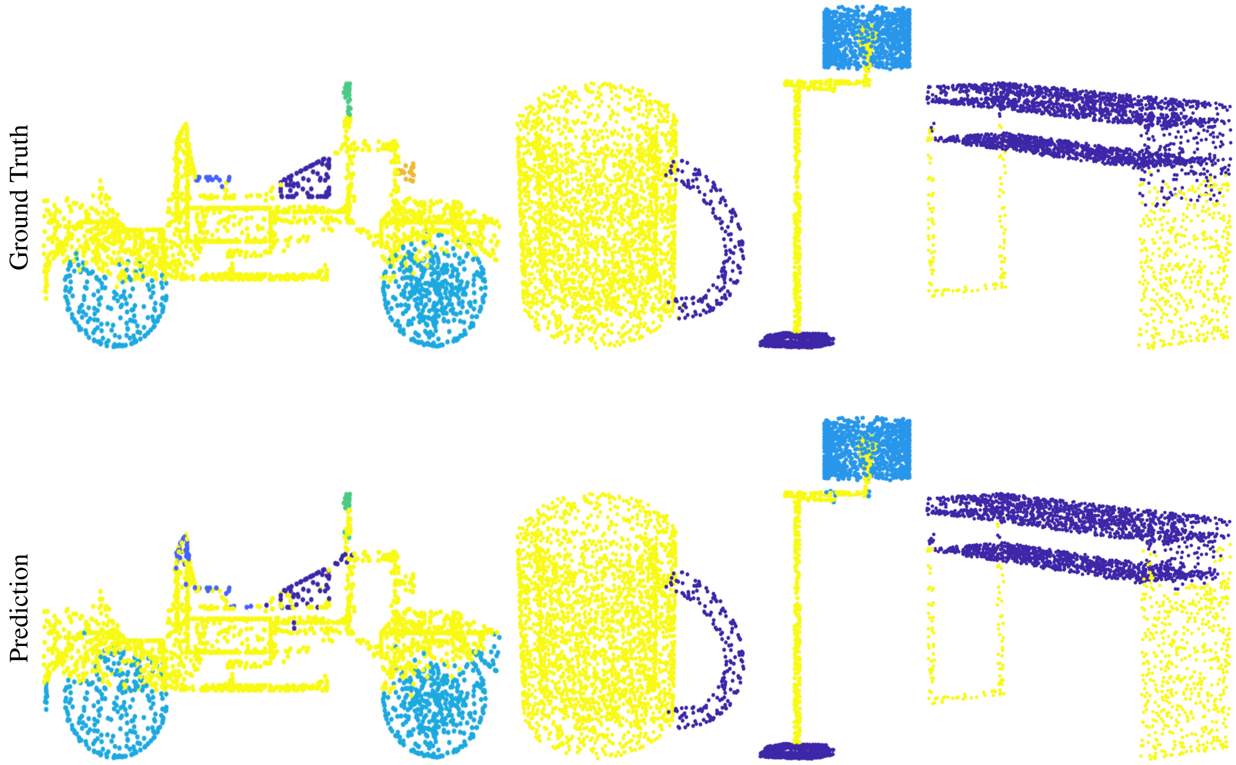

Figure 3: Part segmentation on motobike, mug, lamp, and table objects.

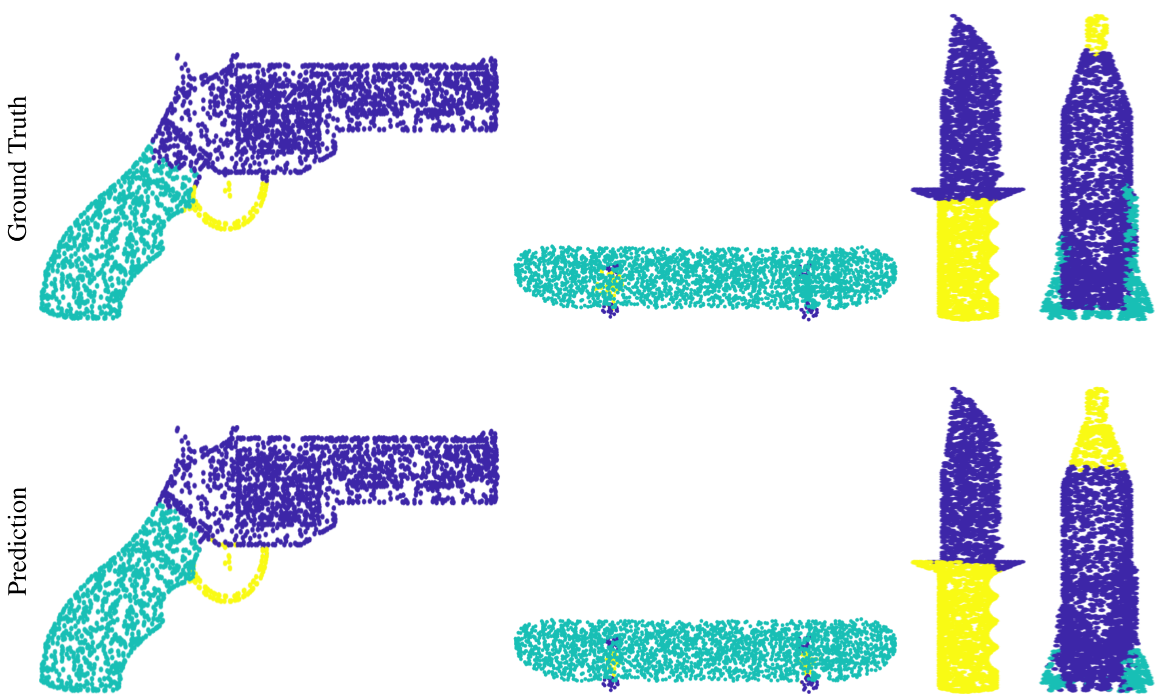

Figure 4: Part segmentation on pistol, skateboard, knife, and rocket objects.

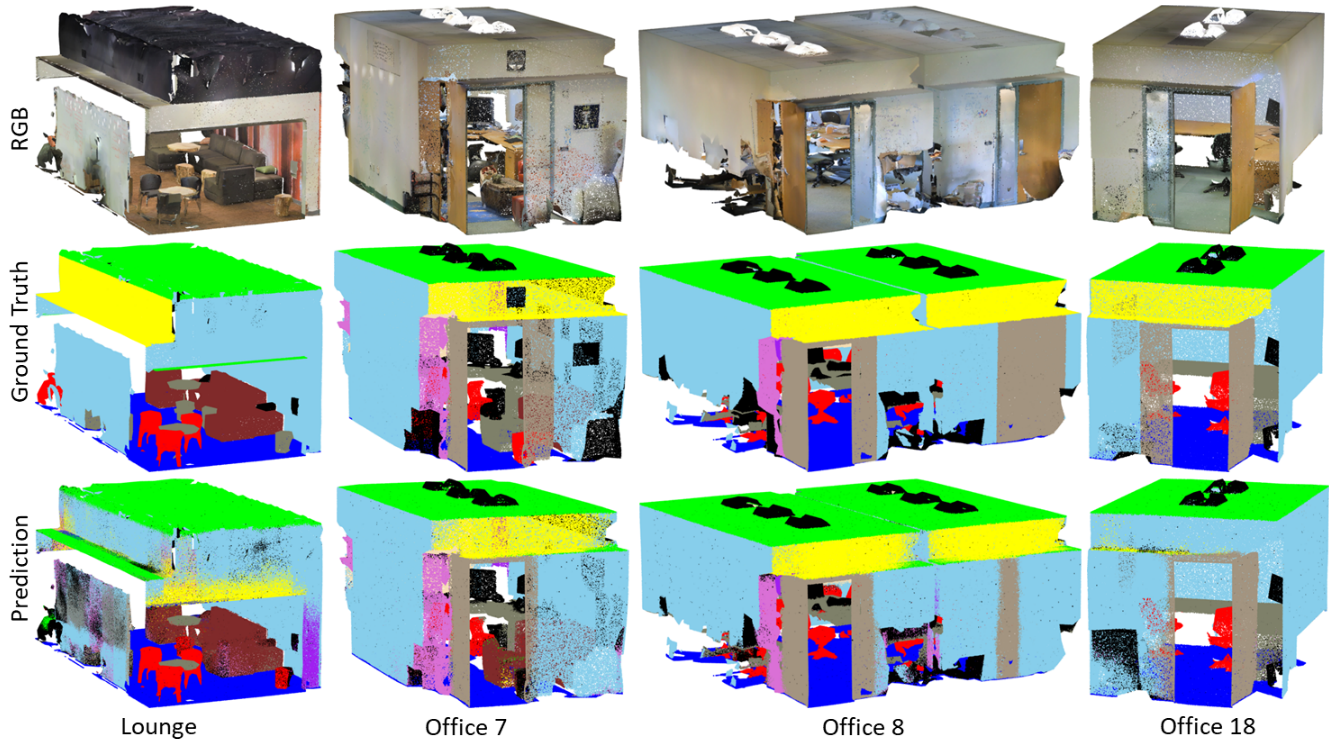

Figure 5: Semantic segmentation with S3DIS dataset.

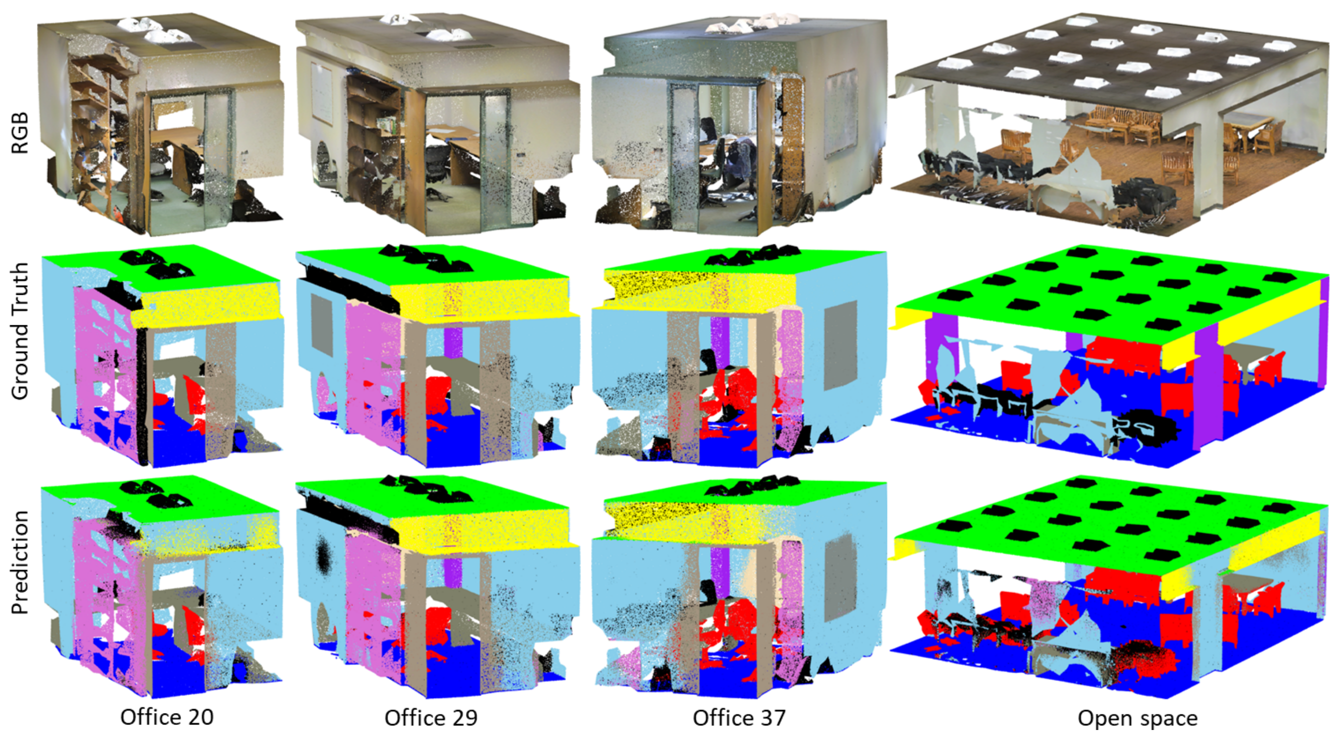

Figure 6: Semantic segmentation with S3DIS dataset.

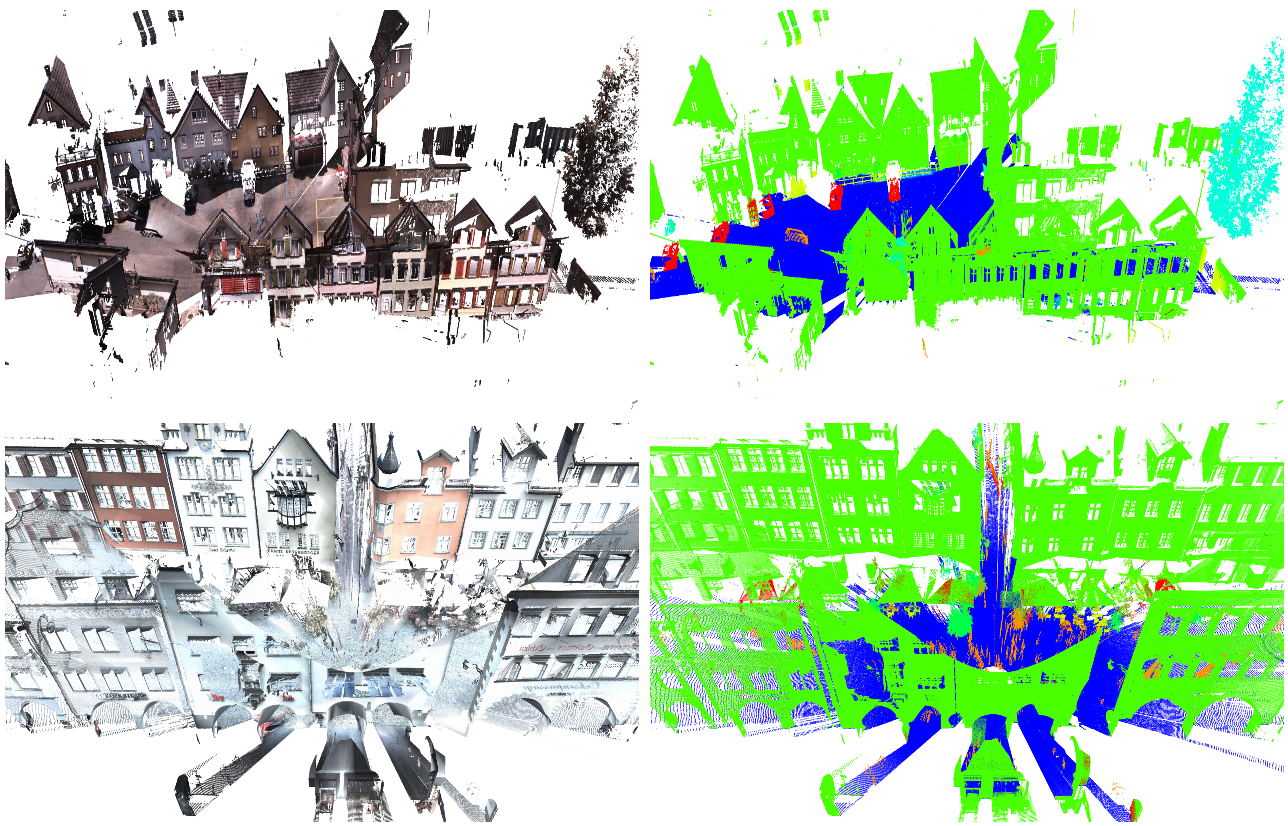

Figure 7: Semantic segmentation for outdoor scenes in the Semantic3D dataset [4]. Left: colored point clouds (for visualization only). Right: our segmentation. Note that the ground truth of the test set is not publicly available.