Supplemental Material for Perturbed Self-Distillation

1. Overview

In this document, we start with more details of the training setup. Then we analyze the role of the GCN in the framework. Moreover, we give the per-class scores of Semantic3D[5] and ScanNet-v2 [4]. Finally, we present the visualization results evaluated on Semantic3D [5] and ScanNet-v2 [4] datasets.

1.1. Training Setup

Weakly setting. Inspired by the previous work [13] on weakly supervised point cloud setting. We create weakly supervised dataset by randomly annotating a tiny fraction points in a category for each original point cloud. Specifically, we set up two weakly supervised training methods: 1pt and 1%. At 1pt setting, we annotate one point for each category for each point cloud sample. For example, there are only 3 categories in a point cloud, and only 3 points are annotated with ground truth. Intuitively, our 1% setting means that 1% of the points are labeled for each category. These labeled points will not change during the training.

You may define this learning style as semi-supervised learning. From the perspective of point classification, this problem can be regarded as semi-supervised learning. At the semantic level, we only annotate some points for each semantic category which is a form of weak supervision (incomplete supervision). In [15], incomplete supervision is defined as a kind of weak supervision. In addition, Xu [13] also defines incomplete supervision as a weakly-supervised task. Therefore, we follow the definition in our paper.

Training config. Here we have supplemented the experimental details of the main paper. Our network training is conducted on a single RTX Titan GPU with 24 GB memory. The batch size is kept fixed to 4 in all datasets. The neighborhood is set to K = 16 for backbone and GCNs. Our network for all datasets takes input point clouds of size 40960 points.

1.2. Experiment Results

The importance of the topological relationship. In order to verify the importance of the topological relationship, we replace the GCN layers in the context-aware module with MLP layers, and presente the comparison results under different settings in the Figure 1. It can be seen that using GCN can greatly improve the performance of segmentation. And the less the labeled points, the more obvious the improvement is. The results also demonstrate that the topological relationship is very important for weakly supervised semantic segmentation tasks.

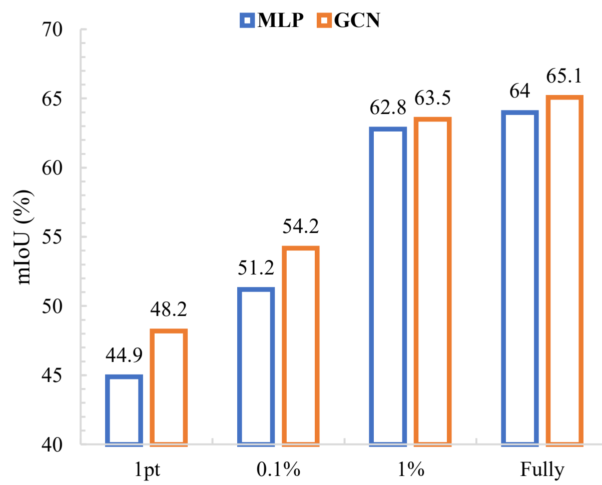

Figure 1. The comparison of GCN and MLP on Area-5 of S3DIS [1] at different settings.

Evaluation on Semantic3D. We conduct the quantitative evaluations on Semantic3D (reduced-8) [5] and list the per-class scores in Table 1. Mean Intersection-over-Union (mIoU) and Overall Accuracy (OA) of all classes are used as the standard metrics. We compared some full supervised methods published in recent years such as SnapNet [3], SEGCloud [10], ShellNet [14], KPConv [11], RandLA-Net [6], and PointGCR [8]. At 1% setting, PSD achieves 75.8% and 94.3% in terms of both mIoU and OA, which are the comparable performance to the fully-supervised methods. Compared with the fully supervised RandLA-Net, our PSD is 1.6% and 0.5% lower than RandLA-Net in mIoU and OA, respectively. But, we achieve the best performance in the category of "man-made terrain" (man-made.) and "high vegetation" (high veg.). Overall, the results show that PSD has a more reliable performance on the outdoor dataset.

Evaluation on ScanNet-v2 We present the segmentation performance of per category on the ScanNet-v2 dataset [4] and choose the weakly supervised setting of 1% for com-

		mIoU	OA	man-made.	natural.	high veg.	low veg.	buildings	hard scape	scanning art.	cars
Fully	SnapNet ('17)[3]	59.1	88.6	82.0	77.3	79.7	22.9	91.1	18.4	37.3	64.4
	SEGCloud ('17)[10]	61.3	88.1	83.9	66.0	86.0	40.5	91.1	30.9	27.5	64.3
	ShellNet ('19)[14]	69.3	93.2	96.3	90.4	83.9	41.0	94.2	34.7	43.9	70.2
	KPConv ('19)[11]	74.6	92.9	90.9	82.2	84.2	47.9	94.9	40.0	**77.3**	79.7
	RandLA-Net ('20)[6]	77.4	94.8	95.6	**91.4**	86.6	51.5	**95.7**	**51.5**	69.8	76.8
	PointGCR ('20)[8]	69.5	92.1	93.8	80.0	64.4	**66.4**	93.2	39.2	34.3	**85.3**
Weakly	ours (1%)	75.8	94.3	**97.1**	91.0	**86.7**	48.1	95.1	46.5	63.2	79.0

Table 1. Quantitative results of per class on Semantic3D (reduced-8) [5]. (mIoU %)

		mIoU	bath-tub	bed	book-shelf	cabinet	chair	counter	curtain	desk	door	floor
Fully	PointNet++ ('17)[9]	33.9	58.4	47.8	45.8	25.6	36.0	25.0	24.7	27.8	26.1	67.7
	PCNN ('18)[2]	49.8	55.9	64.4	56.0	42.0	71.1	22.9	41.4	43.6	35.2	94.1
	SegGCN ('20)[7]	58.9	83.3	73.1	53.9	51.4	78.9	44.8	46.7	57.3	48.4	93.6
	PointConv ('19)[12]	66.6	78.1	75.9	69.9	64.4	82.2	47.5	77.9	56.4	50.4	95.3
	KP-FCNN ('19)[11]	68.4	84.7	75.8	78.4	64.7	81.4	47.3	77.2	60.5	59.4	93.5
Weakly	baseline (1%)	51.3	52.6	61.8	63.0	33.7	78.8	38.8	41.9	47.9	30.7	91.2
	ours (1%)	54.7	57.1	67.8	65.9	$46.5^{12.8\uparrow}$	77.8	38.8	$52.8^{10.9\uparrow}$	49.2	30.4	93.3

		other-furniture	picture	refrig-erator	shower-curtain	sink	sofa	table	toilet	wall	window
Fully	PointNet++ ('17)[9]	18.3	11.7	21.2	14.5	36.4	34.6	23.2	54.8	52.3	25.2
	PCNN ('18)[2]	32.4	15.5	23.8	38.7	49.3	52.9	50.9	81.3	75.1	50.4
	SegGCN ('20)[7]	39.6	6.1	50.1	50.7	59.4	70.0	56.3	87.4	77.1	49.3
	PointConv ('19)[12]	42.8	20.3	58.6	75.4	66.1	75.3	58.8	90.2	81.3	64.2
	KP-FCNN ('19)[11]	45.0	18.1	58.7	80.5	69.0	78.5	61.4	88.2	81.9	63.2
Weakly	baseline (1%)	40.3	26.4	45.2	29.9	42.9	64.7	53.7	72.2	61.7	51.4
	ours (1%)	38.7	30.7	43.1	$38.2^{8.3\uparrow}$	$52.6^{9.7\uparrow}$	66.9	57.2	71.6	60.9	50.6

Table 2. Quantitative results of different approaches on ScanNet-v2 semantic label prediction [4]. (mIoU %)

parison. From Table 2. It can be seen that our PSD achieves 54.7% mIoU and 3.4% improvements against baseline. We also achieve the performance close to the fully supervised method SegGCN [7]. This shows that our method is effective for weakly supervised point cloud semantic segmentation. Moreover, we highlight the improvement of PSD relative to baseline in some categories with red superscripts ($^{x\uparrow}$). For the four categories: "cabinet", "curtain", "shower-curtain", and "sink", PSD achieves the gains of 12.8%, 10.9%, 8.3%, and 9.7% against baseline, respectively. These categories are easily confused each other or confused with other categories. While PSD can greatly improve the performance of these categories. The results show that our method has a better generalization.

1.3. Visualization of Results

Qualitative results on Semantic3D. Figure 2 shows the visualization results on the test set of Semantic3D. Since there is no public ground truth, we show the original point cloud at the left column and our segmentation results at right column. In general, it can be seen that PSD achieves good qualitative segmentation results at 1% setting. Our method can also make more accurate predictions for some categories (*e.g.*, "hard scape", "high vegetation" and "car") with a small number of points.

Qualitative results on ScanNet. In order to further show the effectiveness of PSD with fewer labels, we add experiments at the 0.1% setting and give more qualitative results. In Figure 3, we show the original point clouds, the segmentation results at the 1% and 0.1% settings from left to right on ScanNet-v2 [4], respectively. Whether it is at 1% or 0.1% setting, PSD can achieve good segmentation results for most categories. At the 1% setting, the segmentation precision of small corners and boundaries, is further improved.

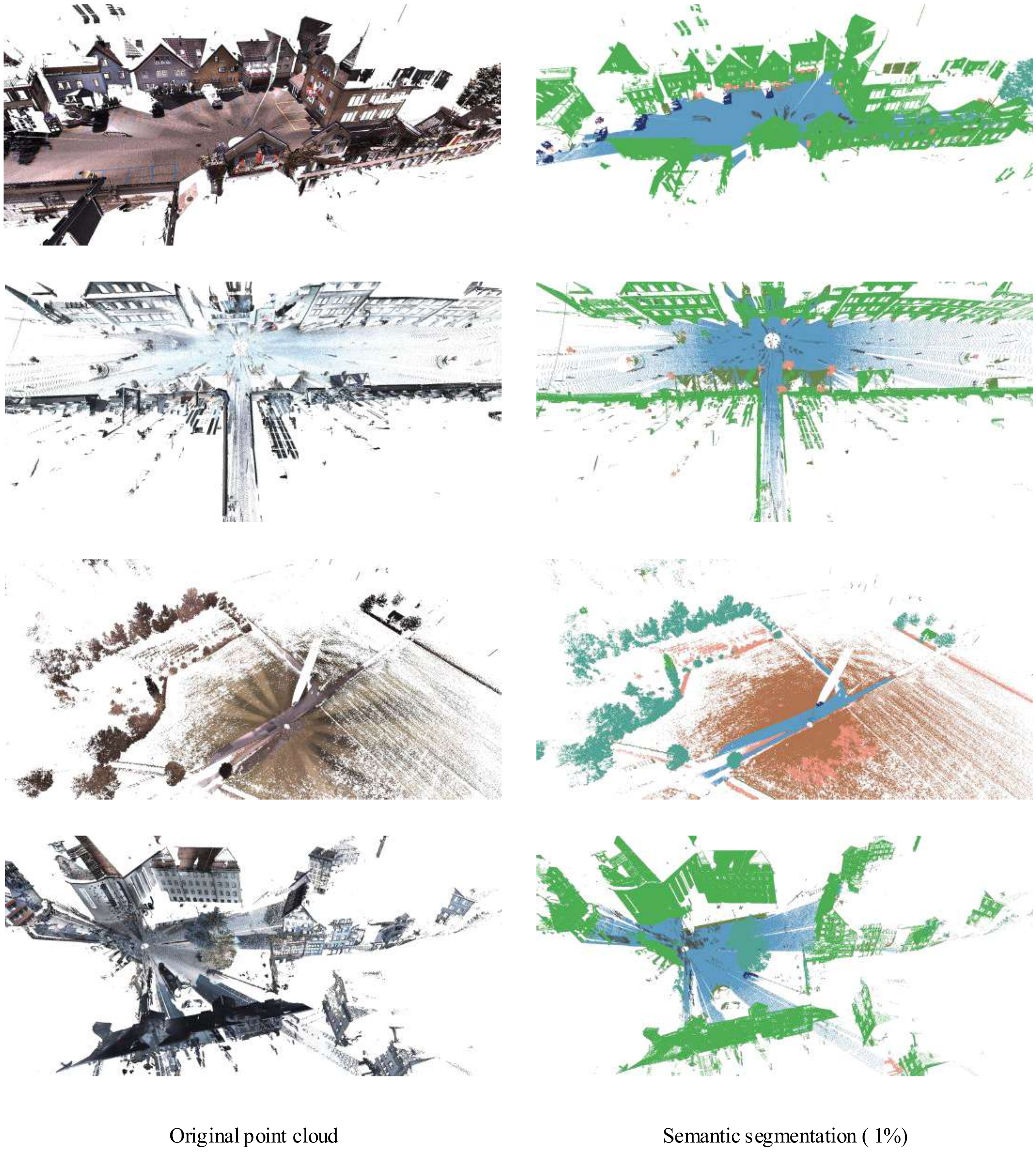

Figure 2. The visualization results on Semantic3D at the 1% setting.

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 *CVPR*, pages 1534–1543, 2016.

Figure 3. The visualization results on ScanNet-v2 at the 1% and 0.1% settings.

[2] Matan Atzmon, Haggai Maron, and Yaron Lipman. Point convolutional neural networks by extension operators. *ACM SIGGRAPH*, 2018.

[3] Alexandre Boulch, Bertrand Le Saux, and Nicolas Audebert. Unstructured point cloud semantic labeling using deep segmentation networks. *3DOR*, 2:7, 2017.

[4] Angela Dai, Angel X. Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *CVPR*, pages 5828–5839, 2017.

[5] Timo Hackel, N. Savinov, L. Ladicky, Jan D. Wegner, K. Schindler, and M. 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.

[6] 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 *CVPR*, pages 11108–11117, 2020.

[7] Huan Lei, Naveed Akhtar, and Ajmal Mian. Seggcn: Efficient 3d point cloud segmentation with fuzzy spherical kernel. In *CVPR*, June 2020.

[8] Yanni Ma, Yulan Guo, Hao Liu, Yinjie Lei, and Gongjian Wen. Global context reasoning for semantic segmentation of 3d point clouds. In *CVPR*, pages 2931–2940, 2020.

[9] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *NeurIPS*, pages 5099–5108, 2017.

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

[11] Hugues Thomas, Charles R Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas J Guibas. Kpconv: Flexible and deformable convolution for point clouds. In *ICCV*, pages 6411–6420, 2019.

[12] Wenxuan Wu, Zhongang Qi, and Li Fuxin. Pointconv: Deep convolutional networks on 3d point clouds. In *CVPR*, pages 9621–9630, 2019.

[13] Xun Xu and Gim Hee Lee. Weakly supervised semantic point cloud segmentation: Towards 10x fewer labels. In *CVPR*, pages 13706–13715, 2020.

[14] Zhiyuan Zhang, Binh-Son Hua, and Sai-Kit Yeung. Shellnet: Efficient point cloud convolutional neural networks using concentric shells statistics. In *ICCV*, pages 1607–1616, 2019.

[15] Zhi-Hua Zhou. A brief introduction to weakly supervised learning. *National Science Review*, 5(1):44–53, 2018.