Supplementary Material of SCPNet: Semantic Scene Completion on Point Cloud

Zhaoyang Xia[1], **Youquan Liu**[1], **Xin Li**[2], **Xinge Zhu**[3], **Yuexin Ma**[4],
Yikang Li[1], **Yuenan Hou**[1]†, **and Yu Qiao**[1]
[1]Shanghai AI Laboratory [2]East China Normal University
[3]The Chinese University of Hong Kong [4]ShanghaiTech University
[1]{houyuenan, liyikang, qiaoyu}@pjlab.org.cn

1. Ablation studies

Loss coefficient β. We investigate the effect of the loss coefficient of the DSKD loss on the final performance. As shown in Table 1, when we change the loss coefficient from 1, 000 to 4, 000, the completion performance of SCPNet first improves and then declines. Therefore, we set the loss coefficient of the DSKD loss as 3, 000 to obtain the best performance.

Detailed performance comparison on DSKD loss. The detailed performance comparison of SCPNet with and without DSKD is shown in Table 2. On motorcycle, truck, person and bicyclist, the proposed DSKD loss can bring more than 3 IoU improvement.

Detailed performance comparison on the downsampling operation. We examine the effect of adding the downsampling operation to the completion sub-network of SCPNet. The detailed performance comparison of SCPNet with and without the downsampling operation is shown in Table 3. It is apparent that the completion performance drops significantly, especially for truck, other-vehicle, other-ground and traffic-sign. The severe performance degradation strongly shows the necessity of removing the lossy downsampling operation for the completion sub-network.

2. Elaborated implementation details

Range mismatch. On SemanticKITTI, the point cloud range used by our segmentation sub-network, *i.e.*, Cylinder3D, is [-36.2, 36.2] m, [-36.2, 36.2] m and [-4, 2] m for x, y, z, respectively. For semantic scene completion, the range of the completion labels is [0, 51.2] m, [-25.6, 25.6] m and [-2, 4.4] m for x, y, z, respectively. The range mismatch problem will cause the existence of many empty voxels, which will significantly hamper the completion performance. To address this problem, we directly use the point cloud range of the completion labels.

†: Corresponding author.

Why conv bias and BN layers breaks the sparsity of voxel features. The voxel features, which are treated as the input of the completion sub-network, are sparse, *i.e.*, only a part of the whole voxel space is occupied. The completion sub-network uses the vanilla dense convolution for dilation. However, the bias of 3D convolution weight, the mean and variance of the Batch Normalization (BN) layers will result in non-zero values of all empty voxel positions. This will cause all empty voxel features to become occupied, which breaks the sparsity of the original voxel features and significantly increases the computation burden of the segmentation sub-network.

How does changing the random seed influence the mIoU values. We conduct experiments on SemanticKITTI using three different random seeds, *i.e.*, 100, 240 and 666. Experiments on SemanticKITTI show that the performance variance of SCPNet is within 0.3 mIoU.

Apply the proposed distillation loss to other architectures. We apply the DSKD loss to JS3CNet. It improves the performance of JS3CNet from 24.0 mIoU to 26.2 mIoU on SemanticKITTI val set.

Apply label rectification to other models. We apply label rectification to JS3CNet. Experimental results show that on SemanticKITTI val set, the proposed label rectification can bring considerable gains to JS3CNet, improving the performance from 24.0 mIoU to 26.8 mIoU.

Computational impact of the proposed adjustments. We calculate the computational overhead of the completion sub-network and finds that it only introduces around 24.4 ms overhead.

Error bands for the results. We run the experiments on SemanticKITTI and SemanticPOSS datasets for three times. The performance variance of SCPNet on these benchmarks is within 0.3 mIoU.

Why panoptic labels are useful in label rectification. The panoptic labels provide instance-level annotations for those thing classes (*e.g.*, cars and persons) and these instance-level annotations are helpful to remove the long traces of moving objects in completion labels which only provide

Table 1. Impact of the loss coefficient β on the performance.

β	mIoU	completion	car	bicycle	motorcycle	truck	other-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic-sign
4000	35.1	49.9	49.5	25.4	28.9	47.0	40.2	15.3	16.1	**5.1**	70.2	58.5	51.4	11.5	33.0	30.1	40.3	31.5	49.3	37.0	27.5
3000	**37.2**	49.9	**50.5**	**28.5**	**31.7**	**58.4**	41.4	**19.4**	**19.9**	0.2	**70.5**	**60.9**	**52.0**	**20.2**	**34.1**	**33.0**	35.3	**33.7**	**51.9**	**38.3**	27.5
2000	35.3	**50.4**	50.0	25.8	31.3	56.4	**41.6**	17.2	9.8	0.0	70.0	58.0	51.3	8.3	31.7	28.7	**40.7**	32.7	51.6	38.1	**27.9**
1000	35.0	48.8	49.2	27.1	29.6	56.2	36.4	16.4	14.2	0.0	69.6	57.7	50.7	7.7	29.7	30.4	34.7	30.3	48.0	37.0	27.4

Table 2. Impact of DSKD loss on the performance.

Methods	mIoU	completion	car	bicycle	motorcycle	truck	other-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic-sign
SCPNet w/ DSKD	**37.2**	**49.9**	**50.5**	**28.5**	**31.7**	**58.4**	41.4	**19.4**	**19.9**	**0.2**	**70.5**	**60.9**	**52.0**	**20.2**	**34.1**	**33.0**	35.3	**33.7**	**51.9**	**38.3**	**27.5**
SCPNet w/o DSKD	34.4	48.5	48.5	26.4	28.1	54.6	**41.7**	14.5	13.1	0.0	70.2	58.3	51.3	2.9	31.7	30.4	**37.9**	31.6	49.2	36.7	25.7

Table 3. Impact of downsampling on the performance.

Methods	mIoU	completion	car	bicycle	motorcycle	truck	other-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic-sign
w/o downsampling	**37.2**	49.9	**50.5**	**28.5**	**31.7**	**58.4**	**41.4**	**19.4**	**19.9**	**0.2**	70.5	**60.9**	**52.0**	**20.2**	**34.1**	**33.0**	35.3	**33.7**	**51.9**	**38.3**	**27.5**
w/ downsampling	33.1	**51.0**	48.7	21.3	28.9	40.3	30.3	17.6	16.0	0.0	**70.7**	58.8	51.3	11.5	33.6	29.4	**41.2**	32.3	51.5	35.9	8.9

Table 4. Training and inference time using A100.

Methods	Train (h)	Inference (ms)	mIoU
SCPNet	34	143.2	37.2
JS3CNet	28	120.6	24.0

semantic segmentation annotations and do not differentiate each single instance.

Training and inference time and a comparison with SOTA. We summarized the training and inference time between JS3CNet and our SCPNet in Table 4. Our SCPNet has comparable training and inference time but exhibits much better completion performance than JS3CNet.

3. Qualitative results

We provide visual comparison of SCPNet with and without DSKD in Fig. 1. Compared with SCPNet without DSKD, the single-frame SCPNet with DSKD achieves better completion and segmentation performance by distilling dense and relation-based information from the multi-frame teacher model. SCPNet without DSKD performs badly on the parking areas and small objects while SCPNet with DSKD exhibits much better completion performance owing to the proposed distillation objective.

And we also provide visual comparison between original SCPNet and SCPNet with downsampling and upsampling operations. As can be seen from Fig. 2, the downsampling and upsampling operations will cause over dilation and shape distortion for these objects marked by the red ellipses.

References

[1] Jens Behley, Martin Garbade, Andres Milioto, Jan Quenzel, Sven Behnke, Cyrill Stachniss, and Jurgen Gall. SemanticKITTI: A Dataset for Semantic Scene Understanding of Lidar Sequences. In *IEEE International Conference on Computer Vision*, pages 9297–9307, 2019. 3

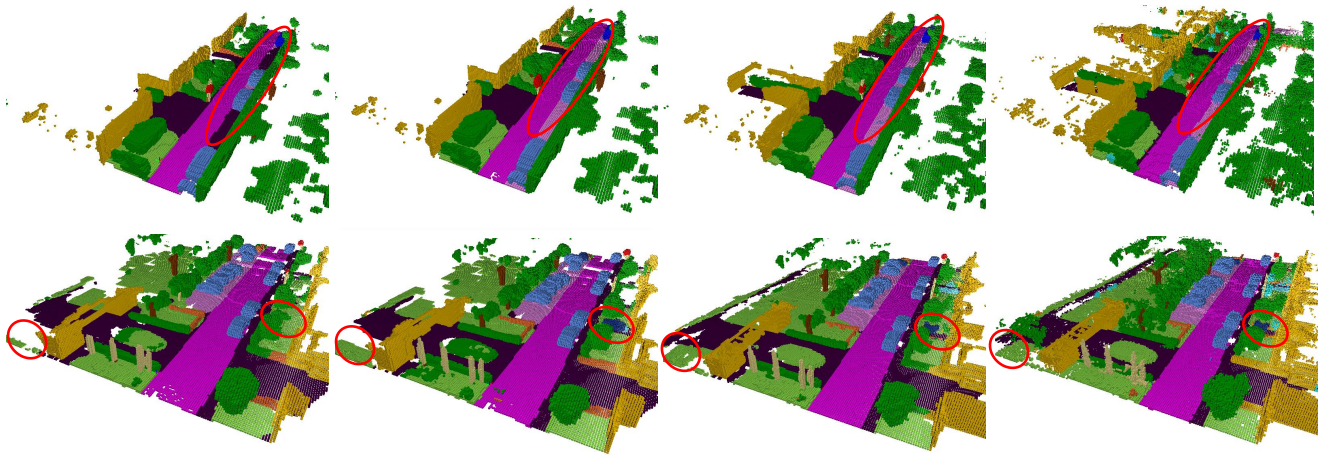

Figure 1. Visual comparison of SCPNet with and without DSKD on the SemanticKITTI [1] validation set. From left to right: SCPNet without the DSKD loss, SCPNet with DSKD, SCPNet-4Frames and ground-truth. Different color represents different class.

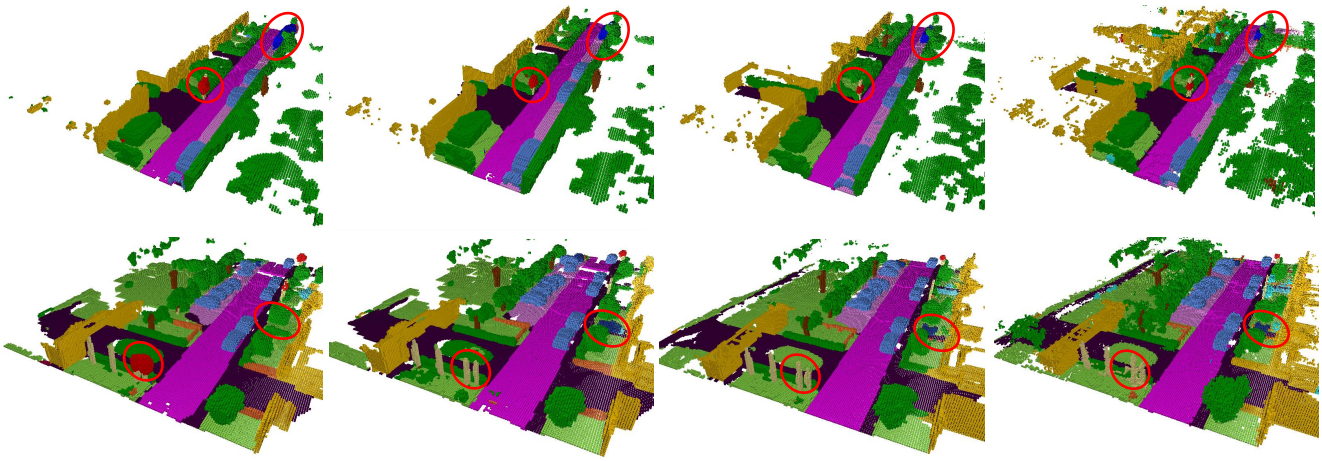

Figure 2. Visual comparison of SCPNet with and without the downsampling operation on the SemanticKITTI [1] validation set. From left to right: SCPNet with downsampling operation, SCPNet without downsampling operation, SCPNet-4Frames and ground-truth. Different color represents different class.