Supplementary Materials: LaserMix for Semi-Supervised LiDAR Semantic Segmentation

Lingdong Kong^{1,2,3,*} Jiawei Ren^{1,*} Liang Pan¹ Ziwei Liu^{1, \boxtimes}

¹S-Lab, Nanyang Technological University ²National University of Singapore ³CNRS@CREATE

{lingdong001,jiawei011}@e.ntu.edu.sg {liang.pan,ziwei.liu}@ntu.edu.sg

In this file, we supplement more content from the following aspects to support the findings and experimental results in the main body of this paper:

- Sec. A provides more technical details of the LiDAR range view and voxel representations.
- Sec. B gives a concrete case study on the strong spatial prior in the outdoor LiDAR data.
- Sec. C elaborates on additional implementation details for different SSL algorithms in our experiments.
- Sec. D provides additional experimental results, including class-wise IoU scores (quantitative results) and visual comparisons (qualitative results).
- Sec. E acknowledges the public resources used during the course of this work.

A. LiDAR Representation

The LiDAR data has a unique and structural format. Various representations have been proposed to better capture the internal information in LiDAR data, including raw points [9, 15, 19], range view (RV) [6, 13, 21, 23], bird's eye view [3, 22], and voxel [17, 24] representations. This section reviews the technical details for RV projection and cylindrical voxel partition, which are currently the most efficient and the best-performing LiDAR representations, respectively.

A.1. Range View Projection

Given a LiDAR sensor with a fixed number (typically 32, 64, and 128) of laser beams and T times measurement in one scan cycle, we project LiDAR point (p^x, p^y, p^z) within this scan into a matrix $x_{rv}(u, v)$ (*i.e.*, range image) of size $h \times w$ via a mapping $\Pi : \mathbb{R}^3 \mapsto \mathbb{R}^2$, where h and w are the height and width, respectively. More concretely, this can be formulated as follows:

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \left[1 - \arctan(p^y, p^x) \pi^{-1} \right] w \\ \left[1 - (\arcsin(p^z, r^{-1}) + \phi_{\text{down}}) \xi^{-1} \right] h \end{pmatrix},$$
(1)

where (u, v) denotes the matrix grid coordinates of x_{rv} ; $r = \sqrt{(p^x)^2 + (p^y)^2 + (p^z)^2}$ is the range between the point and the LiDAR sensor; $\xi = |\phi_{up}| + |\phi_{down}|$ denotes the inclination range (also known as field-of-view or FOV) of the sensor; ϕ_{up} and ϕ_{down} are the inclinations at the upward direction and the downward direction, respectively.

Note that *h* is set based on the number of laser beams of the LiDAR sensor, and *w* is determined by its horizontal angular resolution. The projected range image $x_{rv}(u, v)$ serves as the input for RV-based LiDAR segmentation networks [6, 13, 23]. The semantic labels are projected in the same way as $x_{rv}(u, v)$.

For range view representation, training losses are calculated on the range view predictions of size [k, h, w], where k denotes the number of semantic classes.

A.2. Cylindrical Partition

The cylinder voxels used in [24] exhibit better segmentation performance than the conventional cubic voxels on the LiDAR data. This is because the outdoor LiDAR point clouds have varying density, which decreases as the range increases. More formally, the cylindrical partition transforms points in the Cartesian coordinate (p^x, p^y, p^z) into cylinder coordinate (ρ, α, p^z) , where ρ is the distance to the origin in X-Y plane and α is the azimuth in the sensor horizontal direction. The transformation can be formulated as follows:

$$\rho = \sqrt{(p^x)^2 + (p^y)^2}, \qquad \alpha = \arctan(\frac{p^y}{p^x}).$$
(2)

Given a predefined voxel resolution $[n_{\rho}, n_{\alpha}, n_z]$, points in the cylinder coordinate can be partitioned into the corresponding voxel cells. The semantic labels are split into partitioned cylinder voxels, where all points within the same voxel are assigned a unified label via majority voting.

For cylindrical representation, training losses are calculated on the voxel predictions of size $[k, n_{\rho}, n_{\alpha}, n_z]$, where k denotes the number of semantic classes.

B. Case Study: Spatial Prior in LiDAR Data

As mentioned in the main body of this paper, the Li-DAR point clouds collected by the LiDAR sensor on top of the autonomous vehicle contain inherent spatial cues, which lead to strong patterns in laser beam partition. In this section, we conduct a case study on SemanticKITTI [1] to verify our findings (see Tab. A).

B.1. Laser Partition

The LiDAR scans in the SemanticKITTI [1] dataset are collected by the Velodyne-HDLE64 sensor, which contains 64 laser beams emitted isotropically around the ego-vehicle with predefined inclination angles. In this study, we split each LiDAR point cloud into eight non-overlapping areas, *i.e.*, $A = \{a_1, a_2, ..., a_8\}$. Each area a_i contains points captured from the consecutive 8 laser beams.

B.2. Spatial Prior

As can be seen from the fourth column in Tab. A, different semantic classes have their own behaviors in these predefined areas. Specifically, the *road* class occupies mostly the first four areas (close to the ego-vehicle) while hardly appearing in the last two areas (far from the egovehicle). The *vegetation* class and the *building* class behavior conversely to *road* and appear at the long-distance areas (*e.g.*, a_6 , a_7 , a_8). The dynamic classes, including *car*, *bicyclist*, *motorcyclist*, and *person*, tend to appear in the middle-distance areas (*e.g.*, a_4 , a_5 , a_6). Similarly, from the heatmaps shown in the fifth column in Tab. A, we can see that these semantic classes tend to appear (lighter colors) in only certain areas. For example, the *traffic-sign* class has a high likelihood to appear in the long-distance regions from the ego-vehicle (upper areas in the corresponding heatmap).

These unique distributions reflect the spatial layout of street scenes in the real world. In this work, we propose to leverage these strong spatial cues to construct our SSL framework. The experimental results verify that the spatial prior can better encourage consistency regularization in Li-DAR segmentation under annotation scarcity.

C. Additional Implementation Detail

In this section, we first compare the configuration details for the three LiDAR segmentation datasets (nuScenes [7], SemanticKITTI [1], and ScribbleKITTI [20]) used in this work (see Tab. B). We then provide more detailed information on different SSL algorithms implemented in our semisupervised LiDAR segmentation benchmark.

C.1. Dataset

nuScenes. As a comprehensive autonomous driving dataset, nuScenes¹ [7] provides 1000 driving scenes of 20s duration each collected by a 32-beam LiDAR sensor from Boston and Singapore. We follow the official *train* and *val* sample splittings. The total number of LiDAR scans is 40000. The training and validation sets contain 28130 and 6019 scans, respectively. The semantic labels are annotated within the ranges: $p^x \in [50m, -50m], p^y \in [50m, -50m]$, and $p^z \in [3m, -5m]$. Points outside the range are labeled as *ignored*. The inclination range is $[10^\circ, -30^\circ]$. We use the official label mapping which contains 16 semantic classes.

SemanticKITTI. Derived from the famous KITTI Vision Odometry Benchmark, SemanticKITTI [1] is another large-scale LiDAR segmentation dataset widely adopted in academia. It consists of 22 driving sequences, which are split into a *train* set (Seq. 00 to 10, where 08 is used for validation) and a *test* set (Seq. 11 to 21). The LiDAR point clouds are captured from Karlsruhe, Germany, by a 64-beam LiDAR sensor. The inclination range is $[3^{\circ}, -25^{\circ}]$. We follow the official label mapping and use 19 semantic classes in our experiments.

ScribbleKITTI. Efficiently annotating LiDAR point clouds is a viable solution for scaling up LiDAR segmentation. ScribbleKITTI [20] adopts scribbles to annotate SemanticKITTI [1], resulting in around 8.06% semantic labels compared to the dense annotations. The other configurations are the same as SemanticKITTI [1]. We use the densely annotated set (Seq. 08 in SemanticKITTI [1]) as the validation set.

In summary, we choose datasets with different numbers of laser beams (*i.e.*, 32 for nuScenes [7] and 64 for SemanticKITTI [1] and ScribbleKITTI [20]), different inclination ranges (*i.e.*, $[10^{\circ}, -30^{\circ}]$ for nuScenes [7] and $[3^{\circ}, -25^{\circ}]$ for SemanticKITTI [1] and ScribbleKITTI [20]), and different annotation proportions (*i.e.*, 100% for nuScenes [7] and SemanticKITTI [1] and 8.06% for ScribbleKITTI [20]). Our proposed SSL framework exhibit constant and evident improvements on all three datasets, which further verifies the scalability of our approaches.

C.2. Model Configuration

FIDNet. We use the *ResNet34-point* variant in FIDNet [23] as our *range view* segmentation backbone. It contains fewer parameters (6.05M) than the one used in the original paper (19.64M) while still maintaining good segmentation performance: 58.8% mIoU (compared to 59.5% mIoU) on the *val* set of SemanticKITTI [1], and 71.6\% mIoU (compared to 72.3\% mIoU) on the *val* set of nuScenes [7]. We refer to the FIDNet [23] paper for more details on the model archi-

¹Refer to the *lidarseg* set in nuScenes, details at https://www.nuscenes.org/lidar-segmentation.

Table A. A case study on the **strong spatial prior** in the LiDAR data (statistics calculated from the SemanticKITTI [1] dataset in this example). For each semantic class, we show its type (static or dynamic), occupation (valid # of points in percentage), distribution among eight areas ($A = \{a_1, a_2, ..., a_8\}$, *i.e.*, eight laser beam groups), and the heatmap in range view (lighter colors correspond to areas that have a higher likelihood to appear and vice versa).

Class	Туре	Proportion	Distribution	Heatmap
vegetation	static	24.825%		
road	static	22.545%		
sidewalk	static	16.353%		Name and
building	static	12.118%		
terrain	static	8.122%		And a second
fence	static	7.827%		
car	dynamic	4.657%		
parking	static	1.681%	I I I I I I I I I I I I I I I I I I I	
trunk	static	0.580%		
other-ground	static	0.396%		
pole	static	0.296%		No. of a contract of the object of the objec
other-vehicle	dynamic	0.229%		
truck	dynamic	0.193%		
traffic-sign	static	0.061%		
motorcycle	dynamic	0.045%		
person	dynamic	0.036%		
bicycle	dynamic	0.018%		
bicyclist	dynamic	0.014%		
motorcyclist	dynamic	0.004%		nengang terlek in kenyangkang bertakan kenyangkan kenyan kenyangkan kenyangkan kenyangkan kenyangkan kenyangkan

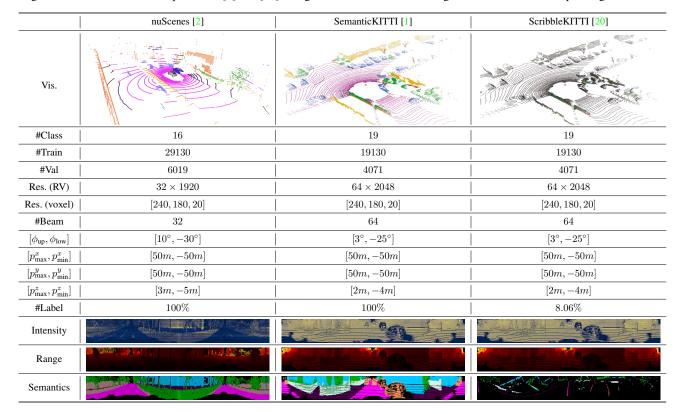
tecture and other related configurations.

Cylinder3D. We use a more compact version of Cylinder3D [24] as the *voxel* segmentation backbone in our experiments, which has 28.13M parameters (compared to 56.26M for the one used in the original paper). We also use a smaller voxel resolution ([240, 180, 20]) compared to the original configuration ([480, 360, 32]). This saves around

 $4\times$ memory consumption and further helps to speed up training. We found that with the smaller resolution (larger voxel size), the performance drops from 76.1% mIoU to 74.1% mIoU on the *val* set of nuScenes [7]. We refer to the Cylinder3D [24] paper for more details on the model architecture and other related configurations.

Training Configurations. All SSL algorithms imple-

Table B. Configuration details for the three LiDAR segmentation datasets (nuScenes [2], SemanticKITTI [1], and ScribbleKITTI [20]) used in this work. Rows from top to bottom: visualization examples, number of semantic classes, number of training scans, number of validation scans, resolution for range view inputs, resolution for voxel inputs, number of laser beams, inclination angle range, x-axis range, y-axis range, z-axis range, the proportion of semantic labels, sensor intensity examples, range examples, and semantic label examples. Images in the second row are adopted from [7] and [20]. Images in the last three rows are generated from the corresponding datasets.



mented in this work share the same LiDAR segmentation backbones, *i.e.*, FIDNet [23] for the LiDAR *range view* representation and Cylinder3D [24] for the LiDAR *voxel* representation. For both FIDNet and Cylinder3D, we adopt AdamW [12] as the optimizer and use the OneCycle learning rate scheduler [16]. The maximum learning rate is 0.0025 for FIDNet and 0.001 for Cylinder3D. The batch size for the LiDAR *range view* representation is 10 for nuScenes and 4 for SemanticKITTI and ScribbleKITTI. The batch size for the LiDAR *voxel* representation is 8 for nuScenes and 4 for SemanticKITTI and ScribbleKITTI.

Data Augmentation. The data augmentations used for the *range view* inputs for all SSL algorithms include random jittering, scaling, flipping (for nuScenes), and shifting (for SemanticKITTI and ScribbleKITTI). The data augmentations used for the *voxel* inputs for all SSL algorithms include random rotation and flipping (for nuScenes, SemanticKITTI, and ScribbleKITTI).

Other Configurations. For LaserMix, the number of spatial areas is uniformly sampled from 1 to 6. The weight λ_{mix} is set as 1 for all three datasets. The weight λ_{mt} is set as 1e3 for nuScenes and 2e3 for SemanticKITTI and Scrib-

bleKITTI. For CPS [4], the weight λ_{cps} is set as 1 for all three datasets. We tried 2 and 6 and found 1 yielded the best results. For MeanTeacher [18], the weight λ_{mt} is set as 1e3 for nuScenes and 2e3 for SemanticKITTI and ScribbleKITTI. For CutMix-Seg [8], the weight λ_{cons} is set as 1 which is the same as the original paper. For CBST [25], we use the *sup.-only* checkpoints to generate the pseudo-labels and then train the segmentation network from scratch with the pseudo-labels. We refer to the original papers for the aforementioned algorithms [4, 8, 18, 25] for additional technical or implementation details.

GPC Split. In the main body, we compared our approach with GPC [10], a 3D SSL method using contrastive learning on point clouds. Since this model is not open-sourced, we directly use the scores reported in their paper for comparison, which might involve factors that are not aligned, *e.g.*, different backbones and data splits. To better align the benchmark settings, we form a sequential track in our codebase² taking into account the LiDAR data collection nature. Kindly refer to our benchmark for more details on this track.

²https://github.com/ldkong1205/LaserMix.

Method	1/16	1/8	1/4	1/2
MeanTeacher [18]	66.1	71.2	74.4	76.3
w/ Ours	68.7	72.3	75.7	76.8
$\Delta \uparrow$	+2.6	+1.1	+1.3	+0.5
CCT [14]	66.4	72.5	75.7	76.8
GCT [11]	65.8	71.3	75.3	77.1
CPS [4]	69.8	74.4	76.9	78.6
CPS-CutMix [4]	74.5	76.6	77.8	78.8
w/ Ours	75.5	77.1	78.3	79.1
$\Delta \uparrow$	+1.0	+0.5	+0.5	+0.3

Table C. Benchmarking results on the val set of Cityscapes [5].

D. Additional Experimental Result

In this section, we provide the class-wise IoU results for our comparative studies and ablation studies in the main body of this paper. Since our proposed SSL framework is a generic design, we also include the benchmarking results on Cityscapes [5] to further verify our generalizability on structural RGB data. To provide more qualitative comparisons, we attach a video demo containing visualizations from the val set of SemanticKITTI [1].

Comparative Study. Tab. D, Tab. E, and Tab. F provide the class-wise IoU scores for different SSL algorithms on the val set of nuScenes [7], SemanticKITTI [1], and ScribbleKITTI [20], respectively. For almost all semantic classes, we observe overt improvements from LaserMix. This can be credited to the strong consistency regularization encouraged by our SSL framework.

Ablation Study. Tab. G and Tab. H provide the class-wise IoU scores for the granularity studies of the LiDAR range view and voxel representations, respectively. Among different LaserMix strategies, we find that increasing the granularity along inclination tends to yield better segmentation performance. In our benchmarking experiments, we combine different strategies together by uniformly sampling the number of spatial areas. This simple ensembling further increases diversity and provides higher segmentation scores.

Extension to RGB Data. To further attest to the scalability of our proposed spatial-prior SSL framework, we conduct experiments on Cityscapes [5], which contains structural RGB images collected from street scenes. We follow the data split from recent work [4] and show the results in Tab. C. Since the images from this dataset also contain strong spatial cues, the mixing strategy used here is similar to that for the LiDAR range view representation, i.e., partitioning areas along the image vertical direction. We combine our proposed \mathcal{L}_{mix} with MeanTeacher [18] (\mathcal{L}_{mt}) and CPS [4] (\mathcal{L}_{cDS}). The results verify that our SSL framework can also encourage consistency for image data. For all four splits, our approaches constantly improve the segmentation performance on top of the SoTA methods [4, 18].

Video Demos. We have attached three demos to show more qualitative results of our approach (see our project page). Specifically, we show the *error maps*, *i.e.*, the *differences* between the model predictions and the ground-truth, on the val sets of SemanticKITTI [1]. The models are trained with 1% labeled data, as discussed in our experiment section. We compare the *sup.-only* model and MeanTeacher [18]. As usual, the error maps are visualized from the LiDAR bird's eve view and range view. Each sub-figure in the video frame shows a LiDAR point cloud of a street scene of size 50m $(|p_{\max}^{x}|)$ by 50m $(|p_{\max}^{y}|)$ by 6m $(|p_{\max}^{z}| - |p_{\min}^{z}|)$. Additionally, we have included several examples from the demos in this file, *i.e.*, Fig. A, Fig. B, Fig. C, Fig. D, and Fig. E.

E. Public Resources Used

We acknowledge the use of the following public resources, during the course of this work:

• nuScenes ³ CC BY-NC-SA 4.0
• nuScenes-devkit ⁴ Apache License 2.0
• SemanticKITTI ⁵ CC BY-NC-SA 4.0
SemanticKITTI-API ⁶ MIT License
• ScribbleKITTI ⁷ Unknown
• FIDNet ⁸ Unknown
• Cylinder3D ⁹ Apache License 2.0
• TorchSemiSeg ¹⁰ MIT License
• Mix3D ¹¹ Unknown
• MixUp ¹² Attribution-NonCommercial 4.0
• CutMix ¹³ MIT License
• CutMix-Seg ¹⁴ MIT License
• CBST ¹⁵ Attribution-NonCommercial 4.0
• MeanTeacher ¹⁶ Attribution-NonCommercial 4.0
<pre>³https://www.nuscenes.org/nuscenes. ⁴https://github.com/nutonomy/nuscenes-devkit. ⁵http://semantic-kitti.org. ⁶https://github.com/PRBonn/semantic-kitti-api. ⁷https://github.com/ouenal/scribblekitti. ⁸https://github.com/placeforyiming/IROS21- FIDNet-SemanticKITTI. ⁹https://github.com/xinge008/Cylinder3D.</pre>
<pre>10https://github.com/charlesCXK/TorchSemiSeg. 11https://github.com/kumuji/mix3d</pre>

F

¹¹https://github.com/kumuji/mix3d.

¹²https://github.com/facebookresearch/mixupcifar10.

¹³https://github.com/clovaai/CutMix-PyTorch.

¹⁴https://github.com/Britefury/cutmix-semisup-

¹⁵https://github.com/yzou2/CBST.

¹⁶ https://github.com/CuriousAI/mean-teacher.

Split	Repr.	Method	mIoU	barr	bicy	bus	car	const	moto	ped	cone	trail	truck	driv	othe	walk	terr	manm	veg
		Suponly	38.3	23.7	0.5	34.2	68.5	0.9	2.6	25.1	26.6	13.1	28.6	89.8	41.3	49.7	61.2	73.4	74.2
	Range View	MeanTeacher [18]	42.1	30.5	0.9	35.6	71.9	0.5	4.2	39.6	33.2	15.2	26.3	92.2	51.3	55.2	63.3	77.4	75.9
	> e	CBST [25]	40.9	28.1	1.7	39.3	71.1	2.0	2.9	27.6	32.6	14.6	32.8	90.8	44.1	51.9	63.6	75.3	75.9
	ngu	CutMix-Seg [8]	43.8	44.6	0.7	30.8	75.7	0.4	2.7	36.9	38.1	18.1	29.8	92.5	45.9	56.9	67.6	80.1	80.0
	R	CPS [4]	40.7	28.4	0.4	38.3	73.5	0.3	0.5	36.1	25.4	13.6	22.4	91.5	44.1	54.4	66.1	77.8	78.8
1%		LaserMix	49.5	50.7	1.8	39.6	80.7	0.6	17.9	53.4	47.6	23.2	41.9	93.5	45.5	60.9	69.4	82.1	82.4
		Suponly	50.9	41.1	1.9	60.0	77.2	7.4	33.7	47.6	39.6	21.3	51.1	93.4	51.9	60.2	65.8	82.1	80.8
	el	MeanTeacher [18]	51.6	48.9	0.8	70.4	79.2	1.5	33.8	50.6	13.9	26.9	58.0	93.8	54.5	62.1	66.5	82.9	82.6
	Voxel	CBST [25]	53.0	59.3	2.6	68.2	77.8	14.2	8.3	54.6	42.2	24.9	51.3	93.1	57.5	60.5	67.0	83.5	82.6
		CPS [4]	52.9	44.1	1.8	63.3	79.1	4.4	28.1	54.2	42.7	22.7	57.6	92.9	56.4	63.5	68.4	83.8	84.0
		LaserMix	55.3	53.6	1.9	67.2	79.2	21.1	29.7	57.3	46.8	28.0	55.5	93.6	54.8	62.1	66.5	83.9	83.3
		Suponly	57.5	65.4	14.7	58.9	82.0	20.7	17.1	60.4	54.0	35.1	54.2	94.6	60.5	65.3	70.0	84.1	82.9
	ew	MeanTeacher [18]	60.4	69.0	12.5	67.0	83.6	27.2	22.0	63.7	55.0	40.4	58.8	95.0	63.8	67.2	71.3	85.6	84.6
	i v	CBST [25]	60.5	68.0	16.8	61.6	83.2	28.4	40.6	62.1	56.4	34.5	54.1	94.9	62.7	66.3	71.2	84.5	83.3
	Range View	CutMix-Seg [8]	63.9	70.5	22.7	69.2	83.4	26.6	69.5	65.2	54.7	39.4	59.2	95.2	59.2	68.3	72.3	84.3	82.8
		CPS [4]	60.8	66.6	9.8	66.6	85.7	22.4	19.7	63.5	57.2	43.4	62.2	95.4	65.3	69.8	73.8	85.9	85.3
10%		LaserMix	68.2	73.9	27.0	74.8	86.9	34.2	69.1	68.7	60.9	48.0	65.5	95.8	68.1	71.0	74.3	87.3	86.0
1070	er	Suponly	65.9	69.5	14.0	87.2	83.5	30.8	61.9	63.5	55.7	47.6	75.2	95.1	63.5	68.1	69.6	85.8	84.1
		MeanTeacher [18]	66.0	71.1	19.7	85.1	83.3	42.0	43.5	64.0	54.9	45.6	73.7	95.3	66.8	69.8	69.6	86.7	84.9
	Voxel	CBST [25]	66.5	70.1	13.6	85.9	82.2	35.1	59.1	61.9	52.1	57.5	74.0	94.5	65.0	70.1	71.8	86.6	85.0
	-	CPS [4]	66.3	72.3	16.4	84.5	81.8	38.5	60.3	62.7	53.4	47.1	70.1	94.7	65.4	70.1	71.7	87.1	85.5
		LaserMix	69.9	72.1	23.3	87.7	84.6	41.3	72.4	67.9	57.2	56.7	77.2	95.5	67.4	70.8	71.2	87.0	85.6
		Suponly	62.7	69.4	19.1	69.7	84.9	29.1	36.4	64.0	58.5	44.6	61.0	95.2	63.6	67.2	70.9	85.7	84.6
	ew	MeanTeacher [18]	65.4	70.7	18.9	75.3	85.6	32.4	48.5	72.2	59.0	46.1	64.0	95.2	65.3	68.3	72.8	86.9	85.7
	5	CBST [25]	64.3	71.6	19.0	70.4	84.4	29.9	49.7	66.2	60.8	46.3	61.3	95.4	62.3	68.4	71.7	86.1	84.6
	Range View	CutMix-Seg [8]	64.8	72.7	23.2	71.8	86.3	34.3	38.2	69.4	59.1	46.9	63.2	95.5	62.0	69.1	72.7	86.9	85.6
	Ra	CPS [4]	64.9	69.6	7.0	75.1	86.6	23.5	50.8	68.5	59.0	50.2	66.5	95.8	68.8	71.1	73.9	85.9	85.6
20%		LaserMix	70.6	74.1	26.1	80.3	89.2	36.2	74.6	73.1	62.8	55.0	73.4	96.0	68.6	71.3	74.3	88.1	86.7
-070		Suponly	66.6	71.5	27.1	82.1	82.7	37.2	68.6	63.6	53.4	42.2	70.5	94.8	65.9	67.8	69.4	85.1	83.7
	-	MeanTeacher [18]	67.1	72.1	26.0	89.1	84.4	39.5	18.4	71.3	57.6	59.3	77.5	95.6	66.9	71.3	71.9	87.6	85.8
	Voxel	CBST [25]	69.6	73.4	29.5	86.1	83.7	37.0	75.7	66.7	56.6	53.0	73.7	95.5	68.5	71.5	70.8	87.3	85.6
		CPS [4]	70.0	73.1	29.3	88.0	83.4	37.2	76.0	66.6	57.8	54.5	75.7	95.5	67.8	71.2	70.5	87.4	85.9
		LaserMix	71.8	73.6	32.1	89.6	84.1	41.4	77.0	69.0	60.0	60.9	78.7	95.8	69.6	72.2	72.9	87.9	84.5
		Suponly	67.6	72.5	32.6	78.5	87.4	32.8	43.6	70.6	62.3	54.0	68.3	95.7	66.4	69.8	72.7	87.7	86.4
	ew	MeanTeacher [18]	69.4	73.4	33.0	81.2	87.6	35.2	61.0	71.9	62.3	55.1	69.4	95.8	66.5	71.1	73.1	87.5	86.1
	Z	CBST [25]	69.3	72.7	35.2	80.8	88.0	35.7	53.7	68.2	62.9	60.2	72.0	95.5	67.4	70.3	73.0	87.3	85.8
	Range View	CutMix-Seg [8]	69.8	74.4	33.5	79.9	88.7	37.3	60.8	70.9	62.0	57.8	70.6	95.8	67.3	70.9	73.3	87.5	85.8
	Ra	CPS [4]	68.0	71.2	31.8	71.9	87.1	29.0	57.4	67.4	62.3	58.6	69.0	95.6	68.7	71.1	74.1	86.7	85.4
50%		LaserMix	73.0	76.0	35.6	85.0	89.9	43.3	76.6	72.5	63.9	61.5	75.1	96.1	69.6	72.3	74.8	88.2	86.9
5070		Suponly	71.2	73.1	35.6	89.0	85.2	41.2	73.3	67.9	59.2	50.9	78.4	95.6	71.5	72.0	73.0	87.3	85.9
	5	MeanTeacher [18]	71.7	73.7	36.2	90.6	85.0	42.3	76.5	68.3	54.9	61.4	74.3	95.7	69.9	72.2	72.6	87.2	86.0
	Voxel	CBST [25]	71.6	73.3	36.1	90.2	84.8	42.2	75.7	67.8	56.6	61.5	74.3	95.7	69.1	72.2	72.7	87.1	85.9
		CPS [4]	72.5	73.9	35.6	91.0	84.9	42.9	79.0	68.6	60.3	60.1	78.3	95.8	71.2	72.3	73.2	87.6	85.2
		LaserMix	73.2	74.5	36.3	91.1	84.9	48.2	78.5	70.5	59.6	59.8	78.9	95.1	70.7	73.5	74.1	88.6	86.9

Table D. **Class-wise IoU scores** of different SSL algorithms on the *val* set of **nuScenes** [2]. All IoU scores are given in percentage (%). The *sup.-only* and the **best** scores for each semantic class within each split are highlighted in **red** and **blue**, respectively.

Split	Repr.	Method	mIoU	car	bicy	moto	truck	bus	ped	b.cyc	m.cyc	road	park	walk	o.gro	build	fence	veg	trunk	terr	pole	sign
		Suponly	36.2	86.8	0.6	0.0	13.0	5.7	12.1	6.6	0.0	87.9	13.4	71.3	0.1	80.4	42.3	78.7	38.1	62.8	52.5	35.7
	View	MeanTeacher [18]	37.5	88.0	0.1	0.1	12.4	3.6	13.0	12.6	0.0	89.2	19.6	73.0	0.0	81.6	44.8	80.2	41.8	64.4	54.0	33.3
	e V	CBST [25]	39.9	89.4	1.9	0.0	4.6	5.8	27.3	3.4	0.0	91.3	25.9	76.5	0.0	83.9	49.1	82.7	56.4	68.1	57.5	33.6
	Range	CutMix-Seg [8]	37.4	86.6	0.2	0.0	3.2	1.5	18.6	6.4	0.0	90.8	24.2	74.9	0.0	81.5	45.5	81.3	50.0	65.7	52.9	34.6
	Rå	CPS [4]	36.5	88.9	0.0	0.0	3.1	0.4	5.7	2.7	0.0	90.8	13.7	76.7	0.0	83.4	52.2	79.9	40.8	63.8	55.9	32.3
1%		LaserMix	43.4	88.8	37.1	0.2	2.1	4.1	10.7	40.7	0.2	91.9	32.3	77.0	0.0	83.9	48.8	81.4	55.9	69.4	59.0	41.7
		Suponly	45.4	90.9	24.5	2.8	35.1	20.4	31.7	49.5	0.0	85.5	23.4	67.5	1.3	85.0	46.0	84.1	49.1	70.3	55.0	40.6
	Voxel	MeanTeacher [18]	45.4	91.2	13.2	5.4	47.3	14.5	29.0	37.3	0.0	86.8	22.6	70.3	1.2	86.7	45.4	84.7	59.4	70.9	55.8	40.8
	Vo	CBST [25] CPS [4]	48.8 46.7	92.4 92.0	$16.3 \\ 13.5$	$6.4 \\ 7.1$	61.9 37.8	27.0 12.7	$35.7 \\ 33.0$	49.4 54.5	0.0 0.0	88.9 89.8	29.4 25.0	73.2 73.8	0.7 0.0	89.1 88.8	49.5 50.1	83.9 83.6	51.4 57.4	$68.1 \\ 67.8$	59.8 58.2	44.0 42.1
		LaserMix	50.6	91.8	35.7	19.8	37.5	25.6	53.6	45.7	2.5	87.8	33.5	71.3	0.7	87.3	43.8	84.6	62.7	69.3	59.8	47.6
		Suponly	52.2	90.4	34.2	22.6	48.2	24.5	59.7	60.9	0.0	92.2	31.8	78.1	0.5	85.7	47.9	83.9	59.3	69.3	59.0	44.2
	3	MeanTeacher [18]	53.1	90.4	30.8	22.0	40.2 58.9	24.5	60.1	57.9	0.0	92.2	34.7	78.7	0.5	87.3	53.5	83.3	59.5 59.6	66.9	57.0	44.2
	View	CBST [25]	53.4	91.7	33.7	28.9	62.0	27.5 29.7	57.9	55.2	0.0	92.9 92.9	32.5	78.7	0.9	87.1	53.7	83.5	59.0 59.4	68.1	57.0 56.7	44.1
	ıge	CutMix-Seg [8]	54.3	90.9	34.9	37.2	57.4	31.7	56.1	63.9	0.0	92.9	34.5	78.6	0.5	87.0	52.3	83.6	58.8	68.8	55.2	44.7
	Range	CPS [4]	52.3	90.2	32.8	19.7	54.0	23.8	56.8	50.5	0.0	92.7	36.3	79.5	0.4	87.6	52.0	85.7	59.4	69.2	58.6	45.1
10%		LaserMix	58.8	92.0	43.5	50.4	76.1	37.1	69.9	74.3	0.0	93.4	38.8	80.1	0.6	87.1	53.3	84.2	63.2	68.3	58.8	45.3
1070		Suponly	56.1	93.4	38.4	47.7	65.7	31.0	61.9	64.9	0.0	90.7	37.7	75.3	0.9	89.2	50.5	86.4	56.0	73.9	56.2	46.0
	5	MeanTeacher [18]	57.1	94.1	40.5	58.4	56.0	38.0	66.5	75.6	0.0	88.4	22.7	72.0	1.5	87.9	49.3	86.7	66.1	74.2	58.0	49.2
	Voxel	CBST [25]	58.3	93.6	40.3	43.5	80.4	33.8	57.6	78.1	0.0	91.6	36.3	76.6	5.1	89.2	51.1	86.3	61.9	71.2	61.3	49.7
	-	CPS [4]	58.7	94.0	38.7	51.0	60.3	39.8	65.7	80.0	0.0	91.4	33.2	76.4	2.9	89.8	53.8	87.2	65.7	74.6	61.5	50.0
		LaserMix	60.0	93.8	44.9	58.4	65.6	39.4	65.8	80.9	0.2	92.0	44.2	77.1	3.9	89.1	49.0	86.2	66.8	72.3	58.4	51.2
		Suponly	55.9	92.2	38.4	34.9	68.8	35.1	63.1	69.4	0.0	93.1	33.8	79.0	1.1	86.6	50.4	84.1	60.9	69.2	56.9	45.3
	View	MeanTeacher [18]	56.1	93.2	33.1	36.3	67.3	39.1	64.9	66.8	0.0	93.3	36.7	79.8	1.0	87.6	54.0	83.9	60.7	67.7	56.5	43.7
	e V	CBST [25]	56.1	92.8	33.2	33.9	64.9	38.9	66.6	69.1	0.0	93.2	36.9	79.7	1.7	87.3	53.6	84.5	60.7	69.1	55.1	44.9
	Range	CutMix-Seg [8]	56.6 56.3	91.5 90.8	42.8 44.0	$39.8 \\ 40.7$	$60.6 \\ 67.9$	32.9 30.7	64.3 65.5	71.6 58.0	$0.0 \\ 0.0$	93.1 93.3	39.8 39.1	79.3 79.5	$0.6 \\ 1.1$	87.1 87.5	53.8 55.6	$85.0 \\ 83.8$	$61.6 \\ 60.4$	71.0 67.9	$56.1 \\ 56.8$	45.4
	Я	CPS [4]								74.7											58.0	46.7
20%		LaserMix	59.4	92.5	43.3	51.5	73.1	45.8	69.4		0.0	94.0	40.4	80.4	5.0	87.3	53.7	83.8	64.1	66.7		44.6
		Suponly	57.8	94.0	31.6	47.3	89.5	38.3	57.9	79.1	0.0	91.6	29.6	76.1	0.9	87.8	43.6	86.6	63.7	72.5	61.8	47.5
	Voxel	MeanTeacher [18]	59.2	94.4	38.7	52.5	81.2	45.8	64.2	78.0	0.0	90.9	35.2	75.7	1.8	89.2	49.8	86.3	65.6	72.6	56.0	47.6
	Vo	CBST [25] CPS [4]	59.4 59.6	94.2 94.2	41.8 41.8	51.4 52.9	77.7 78.2	$39.8 \\ 39.6$	$65.4 \\ 66.1$	$79.8 \\ 80.6$	$0.0 \\ 0.0$	91.7 91.9	29.8 30.2	76.3 76.4	3.5 3.7	89.2 89.2	$49.7 \\ 50.0$	87.1 87.0	$66.1 \\ 66.6$	74.2 73.7	$60.1 \\ 60.0$	51.3 51.1
		LaserMix	61.9	94.4	46.0	68.0	74.3	47.6	68.1	83.7	0.2	92.6	42.7	78.0	1.9	89.7	52.9	86.0	69.3	70.6	59.2	51.7
		Suponly	57.2	91.3	41.1	47.7	70.2	41.2	66.0	74.4	0.0	93.0	39.2	79.2	2.0	86.0	44.2	83.4	59.3	68.6	55.5	45.1
	3		57.4	91.5	38.6	47.7	61.0	41.2	65.7	74.4	0.0	93.1	39.2	79.2	2.0	87.5	44.2 53.8	85.0	60.3	71.5	53.6	47.2
	View	MeanTeacher [18] CBST [25]	56.9	95.1	40.0	42.4 42.9	66.1	45.0 41.7	64.8	73.9	0.0	93.1 93.0	34.9	79.4 79.2	1.2	87.0	$\frac{55.8}{48.7}$	83.7	59.6	68.9	55.0	47.2
	ae Be	CutMix-Seg [8]	57.6	92.0	43.3	48.9	44.6	40.7	67.4	78.5	0.0	93.3	39.1	79.7	3.0	87.2	54.2	86.0	61.6	74.8	55.1	44.8
	Range	CPS [4]	57.4	92.1	38.5	44.3	69.6	45.2	66.5	71.0	0.0	93.5	36.6	80.1	1.7	87.0	48.0	83.9	62.3	68.0	58.0	43.7
F 007		LaserMix	61.4	92.5	45.6	58.8	73.0	53.2	71.2	82.4	0.0	93.7	43.2	80.7	5.5	87.5	52.6	85.4	64.0	71.9	57.9	47.9
50%	i	Suponly	58.7	93.9	40.4	48.0	81.4	33.7	65.7	79.7	0.0	91.9	32.6	76.7	1.3	89.0	51.8	87.2	61.4	72.5	58.7	48.7
		MeanTeacher [18]	60.0	94.1	41.3	57.7	64.6	39.5	65.3	86.8	0.0	91.3	32.8	75.2	3.5	89.7	48.6	85.4	65.9	70.6	58.7	49.1
	Voxel	CBST [25]	59.7	94.9	40.9	54.4	75.3	43.8	67.3	86.8	0.0	91.5	33.3	75.7	2.6	89.3	50.7	86.7	63.9	72.4	56.4	48.8
	>	CPS [4]	60.5	94.6	43.3	55.3	80.5	42.5	67.9	84.6	0.0	92.0	34.3	76.9	2.2	89.8	52.3	86.0	67.4	71.1	59.5	49.4
		LaserMix	62.3	94.7	48.4	64.7	65.2	44.5	71.0	88.3	2.1	92.7	43.0	78.4	2.0	90.3	54.9	88.1	68.1	75.3	66.6	51.7

Table E. **Class-wise IoU scores** of different SSL algorithms on the *val* set of **SemanticKITTI** [1]. All IoU scores are given in percentage (%). The *sup.-only* and the **best** scores for each semantic class within each split are highlighted in **red** and **blue**, respectively.

Split	Repr.	Method	mIoU	car	bicy	moto	truck	bus	ped	b.cyc	m.cyc	road	park	walk	o.gro	build	fence	veg	trunk	terr	pole	sign
		Suponly	33.1	81.3	2.6	0.4	11.7	8.3	11.5	8.7	0.0	76.7	10.9	61.8	0.1	75.8	26.3	73.8	40.7	56.1	48.9	32.5
	View	MeanTeacher [18]	34.2	82.3	1.7	0.1	10.4	6.7	6.1	4.7	0.0	78.6	13.4	67.8	0.1	80.7	31.3	76.1	43.1	60.0	53.3	32.8
	e <	CBST [25]	35.7	84.8	1.6	0.4	11.7	10.6	14.9	8.0	0.0	83.6	13.4	68.1	0.1	79.5	32.4	77.1	44.6	60.5	53.0	34.4
	Range	CutMix-Seg [8] CPS [4]	36.7 33.7	84.7 82.7	$0.9 \\ 0.1$	$0.0 \\ 0.0$	$5.5 \\ 0.9$	$0.9 \\ 0.1$	$18.7 \\ 2.9$	$1.9 \\ 4.1$	$0.0 \\ 0.0$	$89.3 \\ 85.9$	25.1 8.9	74.6 70.8	0.1 0.0	82.6 81.2	27.0 47.3	77.7 78.1	52.1 36.0	65.0 61.2	54.7 51.9	$35.8 \\ 27.5$
	~ 	LaserMix	38.3	86.5	1.9	0.0	12.8	2.9	25.9	2.6	0.0	90.8	25.0	75.8	1.0	83.9	26.4	77.8	55.5	63.9	56.7	38.2
1%		Suponly	39.2	83.2	13.8	3.4	26.3	11.8	28.0	25.2	0.0	72.5	13.0	59.5	0.2	86.6	33.7	78.7	55.7	58.4	54.0	40.3
	' _	MeanTeacher [18]	41.0	82.3	15.8	7.1	32.0	15.4	23.7	36.3	0.0	75.0	12.6	61.4	0.9	85.3	30.0	80.1	57.0	67.0	56.1	41.3
	Voxel	CBST [25]	41.5	83.7	22.1	5.9	28.3	13.4	27.1	34.7	0.0	74.0	14.4	61.7	0.2	88.1	36.6	80.3	58.7	60.4	57.1	41.4
	>	CPS [4]	41.4	82.8	18.2	11.4	20.9	15.1	22.5	35.5	0.0	74.7	15.7	61.6	0.4	86.0	34.2	82.2	58.4	69.9	56.7	40.0
		LaserMix	44.2	82.6	25.5	18.8	29.0	19.8	41.1	47.2	0.6	71.5	10.5	64.2	2.2	85.1	33.5	82.0	59.9	65.8	54.5	45.2
		Suponly	47.7	85.1	30.2	20.4	40.4	20.9	54.4	55.9	0.0	82.8	21.6	68.4	0.5	84.2	40.5	80.3	58.6	61.9	56.0	45.2
	View	MeanTeacher [18]	49.8	83.5	30.0	22.7	62.2	31.1	59.1	52.4	0.0	77.9	17.6	70.5	2.0	86.8	42.6	82.0	61.2	62.6	58.4	46.5
	e V	CBST [25]	50.7	90.3	27.2	18.1	53.1	24.6	60.3	56.5	0.0	90.3	32.4	76.0	0.7	86.1	46.0	81.1	58.7	64.5	51.7	45.0
	Range	CutMix-Seg [8]	50.7	87.4	28.1	25.9	60.5	24.5	58.4	57.7	0.0	85.5	27.5	72.1	1.3	84.7	39.4	82.4	58.8	68.3	56.4	44.4
	22 	CPS [4]	50.0 54.4	85.8	26.7 35.4	17.4 44.4	54.5 62.5	20.5 36.4	54.4 66.9	53.7 72.6	0.0	88.9 80.8	29.2 27.8	74.4 73.7	0.6	86.5 85.2	48.6 35.2	82.4 83.9	58.6 60.6	65.2 70.0	57.3 59.3	45.1 51.6
10%	 	Suponly	48.0	85.7	25.6	21.3	52.8	29.9	46.5	47.2	0.0	79.5	15.4	63.8	0.0	85.4	39.6	84.8	59.7	70.0	59.5 57.7	45.8
	1	1 2				-																
	Voxel	MeanTeacher [18] CBST [25]	50.1 50.6	83.7 85.8	32.6 31.4	$45.1 \\ 30.5$	41.0 58.5	34.7 24.4	$56.0 \\ 55.1$	$59.2 \\ 58.8$	$0.0 \\ 0.0$	75.9 82.6	$14.0 \\ 15.3$	$64.0 \\ 67.8$	$0.7 \\ 0.5$	$85.6 \\ 87.7$	37.9 40.0	83.3 82.8	$62.6 \\ 62.5$	$68.2 \\ 65.0$	59.7 62.0	47.0 50.8
		CPS [4]	51.8	84.6	34.9	47.1	37.5	29.5	60.1	69.1	0.0	79.8	16.5	67.3	2.7	88.0	39.2	84.5	64.5	71.0	60.4	47.9
		LaserMix	53.7	85.8	34.7	45.6	54.9	35.8	63.2	73.6	1.3	79.8	25.0	68.2	1.8	87.7	35.4	84.0	65.8	70.8	59.4	48.2
		Suponly	49.9	86.3	32.2	23.8	49.5	30.3	60.5	58.4	0.0	83.6	22.4	69.5	1.1	85.1	40.6	80.9	59.9	62.3	55.9	46.4
	No.	MeanTeacher [18]	51.6	82.9	27.7	43.1	59.5	32.8	59.5	60.7	0.0	80.8	25.7	70.3	0.7	85.3	41.6	82.1	60.5	66.0	55.6	45.7
	View	CBST [25]	52.7	90.0	33.1	30.2	53.6	33.8	60.0	60.4	0.0	89.3	30.3	75.8	0.6	85.6	44.8	83.5	58.6	70.3	54.7	47.1
	Range	CutMix-Seg [8]	52.9	86.9	30.0	35.6	64.8	35.7	60.9	63.6	0.0	88.3	29.0	74.7	0.9	85.2	40.3	82.0	59.4	65.2	56.5	45.5
	Ra	CPS [4]	52.8	86.3	35.4	28.1	67.1	27.7	59.5	59.2	0.0	89.0	28.0	75.0	0.8	86.7	47.3	83.1	61.0	66.9	58.1	44.6
20%		LaserMix	55.6	87.3	36.0	34.3	69.5	40.6	66.3	70.6	0.0	84.2	27.2	72.3	2.4	86.4	44.6	84.1	62.8	69.8	59.4	57.7
		Suponly	52.1	86.9	38.0	39.5	67.3	29.7	56.5	69.9	0.0	79.0	16.0	66.0	0.3	87.0	38.6	84.3	60.6	66.2	58.8	45.2
	e l	MeanTeacher [18]	52.8	85.9	27.9	41.5	55.5	33.0	64.1	72.0	1.2	81.0	22.5	67.8	1.2	89.1	39.9	82.9	63.7	66.9	60.5	46.7
	Voxel	CBST [25]	53.3	86.6	36.8	40.9	72.9	28.3	58.0	69.5	0.0	81.1	18.3	68.2	0.7	88.7	44.3	83.6	63.3	64.4	60.3	47.5
		CPS [4]	53.9	85.4	37.2	44.7	58.9	32.9	63.5	71.0	0.0	81.6	23.1	69.2	1.9	88.4	38.2	83.8	65.7	69.2	60.2	48.9
		LaserMix	55.1	88.0	38.8	51.3	54.8	36.6	60.2	73.9	0.0	78.8	22.7	71.9	1.5	90.3	43.3	85.3	66.5	70.9	60.3	51.6
		Suponly	52.5	86.7	35.9	40.2	55.9	30.1	63.2	62.9	0.1	83.6	25.3	70.8	1.1	85.1	40.0	82.9	60.4	69.0	56.3	48.3
	View	MeanTeacher [18]	53.3	86.9	$31.9 \\ 36.0$	$37.5 \\ 36.6$	58.6	36.3	63.3	62.0	0.0	87.6	29.5	74.1	1.0 3.8	86.4	40.7	82.6	61.3	68.9	58.0	47.0
	Se	CBST [25] CutMix-Seg [8]	54.6 54.3	90.1 88.1	35.3	30.0 40.0	64.7 68.8	41.6 39.3	$61.2 \\ 62.4$	$66.7 \\ 69.8$	$0.0 \\ 0.0$	90.4 88.0	33.8 32.0	76.8 74.3	3.8 0.9	84.5 85.1	44.3 38.4	83.7 82.4	57.6 59.3	$70.2 \\ 67.4$	48.4 56.3	$47.8 \\ 44.5$
	Range	Cutwix-Seg [8] CPS [4]	54.6	87.1	35.0	40.0	66.1	39.3 40.8	63.2	65.5	0.0	87.9	30.0	74.5 74.6	1.4	86.1	38.4 42.4	82.4 82.7	59.5 60.9	67.9	50.5 57.5	44.5 48.2
		LaserMix	58.7	88.2	37.1	56.0	80.9	51.8	70.8	75.0	0.0	87.0	31.8	74.7	0.8	86.6	41.3	84.6	62.1	72.9	59.8	53.7
50%	' 	Suponly	53.8	87.5	37.2	41.3	71.4	29.6	58.8	80.4	0.0	81.1	16.7	67.5	0.4	88.4	39.4	83.1	64.4	65.5	61.8	47.5
	I	MeanTeacher [18]	53.9	86.9	33.6	46.2	48.9	33.2	62.8	77.7	0.0	82.7	22.8	68.6	3.2	89.2	38.6	83.8	66.4	68.0	62.3	48.5
	Voxel	CBST [25]	54.5	87.6	39.5	$\frac{40.2}{36.7}$	46.9 65.9	35.2 35.7	62.8	78.1	0.0	82.4	22.8	69.6	0.1	89.2 88.8	42.3	84.2	64.0	67.4	60.1	$\frac{48.5}{50.1}$
	×	CPS [4]	54.8	85.1	35.2	45.2	68.6	32.0	65.7	77.9	0.2	81.2	21.7	69.0	1.6	89.2	40.2	84.5	65.1	70.1	60.9	48.5
		LaserMix	56.8	88.0	40.8	51.6	63.1	38.4	61.7	79.9	2.0	83.1	26.1	71.2	2.8	90.1	41.7	85.9	69.5	70.5	63.0	51.6

Table F. **Class-wise IoU scores** of different SSL algorithms on the *val* set of **ScribbleKITTI** [20] (the same as SemanticKITTI [1]). All IoU scores are given in percentage (%). The *sup.-only* and the **best** scores for each semantic class within each split are highlighted in red and **blue**, respectively.

U		U	. ,								U	U						
Method	Illustr.	mIoU	barr	bicy	bus	car	const	moto	ped	cone	trail	truck	driv	othe	walk	terr	manm	veg
Baseline		60.4	69.0	12.5	67.0	83.6	27.2	22.0	63.7	55.0	40.4	58.8	95.0	63.8	67.2	71.3	85.6	84.6
$(1\alpha, 2\phi)$	\bigcirc	63.5	70.8	17.8	65.3	84.9	26.9	44.7	65.8	59.2	46.6	62.2	95.5	64.3	69.2	72.5	86.1	84.9
$(1\alpha, 3\phi)$	\bigcirc	65.2	72.3	21.5	67.1	85.1	26.2	57.1	70.4	59.3	45.8	60.7	95.6	65.4	69.3	73.7	87.0	85.9
$(1\alpha, 4\phi)$	\bigcirc	66.5	73.7	22.4	72.9	87.0	26.3	59.4	70.2	60.3	44.7	64.7	95.8	67.8	70.9	74.2	87.0	85.9
$(1\alpha, 5\phi)$		66.2	72.8	24.1	74.0	85.7	36.3	47.8	71.5	60.8	45.8	64.5	95.7	64.8	69.9	73.3	87.1	85.9
$(1\alpha, 6\phi)$	6	65.4	72.6	25.2	69.8	84.6	33.8	48.3	70.1	60.5	44.8	61.9	95.4	65.3	68.9	73.3	86.8	85.5
$(2\alpha, 1\phi)$	×	61.5	68.4	19.1	67.1	83.5	28.1	26.1	64.8	57.6	41.5	59.0	95.1	64.4	68.1	72.0	85.4	84.3
$(2\alpha, 2\phi)$	$\left \bigotimes \right $	63.3	70.8	11.9	64.8	84.4	27.3	51.4	69.2	58.3	41.8	59.5	95.6	63.8	69.8	73.2	86.2	85.3
$(2\alpha, 3\phi)$	Ø	65.9	71.9	24.2	69.1	86.3	28.2	58.5	71.5	60.1	44.7	63.4	95.7	65.3	70.1	73.4	86.9	85.9
$(2\alpha, 4\phi)$	Ø	66.1	72.9	27.8	70.4	86.4	34.2	54.1	71.4	61.5	42.9	61.3	95.4	64.6	69.0	73.3	86.9	85.8
$(2\alpha, 5\phi)$		66.7	73.1	20.5	69.6	87.5	27.1	67.8	71.1	61.0	43.4	64.6	95.6	69.1	70.1	74.1	87.0	85.7
$(2\alpha, 6\phi)$		65.3	71.9	23.5	68.6	85.3	32.2	51.0	69.6	60.2	45.2	63.6	95.4	63.6	69.2	73.3	86.4	85.1
$(3\alpha, 1\phi)$		60.9	67.4	14.7	65.3	82.9	25.2	39.5	63.6	57.0	34.7	55.2	95.0	64.6	67.1	71.4	85.6	84.7
$(3\alpha, 2\phi)$		64.2	71.4	15.5	66.3	86.2	25.8	54.4	68.4	60.4	44.3	63.2	95.5	63.3	69.3	72.6	86.3	85.1
$(3\alpha, 3\phi)$		65.9	73.1	15.7	74.4	85.3	28.0	59.7	69.0	60.6	45.3	63.4	95.6	68.8	70.5	74.3	86.5	85.2
$(3\alpha, 4\phi)$		66.3	72.1	16.7	73.1	86.2	34.2	66.6	67.3	58.4	45.5	62.4	95.7	65.6	70.5	74.0	86.7	85.6
$(3\alpha, 5\phi)$		66.0	72.1	18.4	72.7	86.0	31.9	68.8	66.0	58.3	42.9	61.9	95.6	65.7	70.0	73.9	86.5	85.4
$(3\alpha, 6\phi)$		65.2	71.0	26.4	68.6	85.6	29.3	54.7	69.1	59.6	43.5	62.1	95.3	63.9	68.7	73.4	86.3	85.1
$(4\alpha, 1\phi)$		60.9	69.1	16.7	66.2	83.7	23.5	26.8	64.4	56.2	40.6	60.4	95.0	63.7	67.0	71.5	85.6	84.6
$(4\alpha, 2\phi)$		64.7	71.0	16.7	66.4	84.7	28.4	59.4	70.6	60.7	42.1	60.3	95.4	65.4	69.6	73.2	86.4	85.5
$(4\alpha, 3\phi)$		65.3	72.0	24.9	68.5	85.6	25.1	52.7	71.3	59.3	45.2	62.3	95.5	66.9	69.6	73.1	86.8	85.8
$(4\alpha, 4\phi)$		65.6	71.9	21.3	71.3	86.4	32.8	44.0	69.3	61.8	45.5	64.0	95.7	67.7	70.2	73.9	87.1	86.2
$(4\alpha, 5\phi)$		65.7	71.9	24.8	72.1	84.5	31.6	52.2	71.6	60.6	43.6	60.4	95.5	65.1	70.1	73.8	87.1	86.1
$(4\alpha, 6\phi)$		65.2	72.1	18.8	69.9	83.7	32.5	57.2	69.0	60.2	43.2	59.1	95.5	66.1	69.8	73.5	87.0	86.1

Table G. Class-wise IoU scores for **ranularity studies** on the *range view* representation (under 10% split on the *val* set of nuScenes [2]). All scores are given in percentage (%). The best score for each semantic class is highlighted in **bold**.

										-								
Method	Illustr.	mIoU	barr	bicy	bus	car	const	moto	ped	cone	trail	truck	driv	othe	walk	terr	manm	veg
Baseline		66.0	71.1	19.7	85.1	83.3	42.0	43.5	64.0	54.9	45.6	73.7	95.3	66.8	69.8	69.6	86.7	84.9
$(1\alpha, 2\phi)$	\bigcirc	68.7	72.4	20.7	87.0	83.1	38.3	66.7	65.8	57.5	54.9	76.6	95.4	67.1	70.3	71.6	86.9	84.6
$(1\alpha, 3\phi)$	\bigcirc	69.0	71.5	21.5	87.0	83.7	39.5	68.1	66.1	57.5	56.6	77.2	95.5	66.5	70.7	71.9	86.9	84.6
$(1\alpha, 4\phi)$	\bigcirc	69.4	71.8	24.1	87.2	84.5	40.8	69.6	66.7	57.1	55.4	76.1	95.5	67.2	70.7	70.9	86.7	85.4
$(1\alpha, 5\phi)$	\bigcirc	69.6	71.1	24.4	88.1	83.0	42.1	72.2	66.4	57.4	57.7	75.2	95.4	67.2	70.5	70.7	86.9	85.5
$(1\alpha, 6\phi)$		69.3	70.3	23.1	87.3	83.5	38.6	74.1	65.8	56.5	57.2	76.7	95.5	65.7	70.8	71.0	86.9	85.7
$(2\alpha, 1\phi)$	×	67.2	70.0	19.7	84.3	86.3	39.6	65.7	62.6	52.1	50.7	73.4	95.2	64.4	69.2	71.4	86.6	84.6
$(2\alpha, 2\phi)$		67.7	70.7	25.9	84.7	84.7	37.4	65.3	63.5	52.6	53.6	71.9	95.3	65.8	69.8	71.6	86.6	84.7
$(2\alpha, 3\phi)$		68.5	70.3	19.8	86.2	86.2	38.1	70.9	64.0	55.7	55.7	74.7	95.3	65.9	69.8	72.0	87.0	84.7
$(2\alpha, 4\phi)$		69.6	72.3	24.2	86.3	85.0	41.8	72.1	66.2	56.6	56.5	76.4	95.4	66.7	70.6	71.4	86.9	85.5
$(2\alpha, 5\phi)$		69.3	72.4	21.7	86.3	84.8	40.5	68.7	66.9	56.6	58.0	76.8	95.3	67.0	70.3	71.5	86.9	85.5
$(2\alpha, 6\phi)$		69.1	71.9	20.5	87.7	84.2	41.6	69.3	66.5	56.9	55.0	76.5	95.4	66.4	70.7	71.1	86.9	85.5
$(3\alpha, 1\phi)$		67.3	70.1	20.7	82.8	86.3	34.4	66.5	62.5	53.6	55.2	74.5	95.1	64.4	69.1	71.4	86.6	84.4
$(3\alpha, 2\phi)$		68.3	71.9	16.6	85.9	83.0	39.5	66.3	66.0	57.0	56.6	76.3	95.4	64.3	70.5	71.7	86.7	84.6
$(3\alpha, 3\phi)$		68.6	70.9	20.2	86.7	86.4	39.1	68.0	66.2	56.7	53.3	74.5	95.3	66.2	70.0	71.8	87.1	85.0
(3α, 4φ)		68.5	70.6	19.9	86.7	86.1	39.3	67.6	66.2	56.8	54.1	74.2	95.3	65.6	70.1	71.9	87.1	84.9
$(3\alpha, 5\phi)$		69.8	72.2	24.3	87.3	84.8	41.2	73.6	67.0	57.1	55.9	76.9	95.4	67.1	70.4	71.2	86.9	85.6
$(3\alpha, 6\phi)$		69.1	71.9	20.5	87.7	84.2	41.6	69.3	66.5	56.9	55.0	76.5	95.4	66.4	70.7	71.1	86.9	85.5
$(4\alpha, 1\phi)$		66.7	70.5	15.4	85.1	86.5	35.2	67.5	62.5	51.3	51.0	72.9	95.1	63.9	68.5	71.6	86.2	84.5
$(4\alpha, 2\phi)$	\otimes	68.4	70.9	18.6	86.0	85.3	39.5	69.9	65.6	56.4	54.2	73.6	95.4	65.5	70.0	71.7	87.0	84.8
$(4\alpha, 3\phi)$		68.7	70.3	22.1	86.5	86.6	39.6	67.8	66.0	57.1	53.2	75.4	95.3	66.0	69.9	71.9	87.2	85.0
$(4\alpha, 4\phi)$		68.9	71.5	23.8	86.9	83.1	38.6	74.5	65.8	56.5	50.8	75.1	95.6	65.8	70.8	70.4	87.1	85.6
$(4\alpha, 5\phi)$		69.0	71.0	23.3	87.7	83.5	40.3	73.7	66.1	56.9	51.6	75.1	95.6	65.3	70.7	70.4	87.1	85.6
$(4\alpha, 6\phi)$		69.4	70.9	25.2	87.9	83.5	40.8	73.0	66.5	57.2	54.0	76.4	95.6	65.1	70.9	70.6	87.1	85.8

Table H. Class-wise IoU scores for **granularity studies** on the *voxel* **representation** (under 10% split on the *val* set of nuScenes [2]). All scores are given in percentage (%). The best score for each semantic class is highlighted in **bold**.

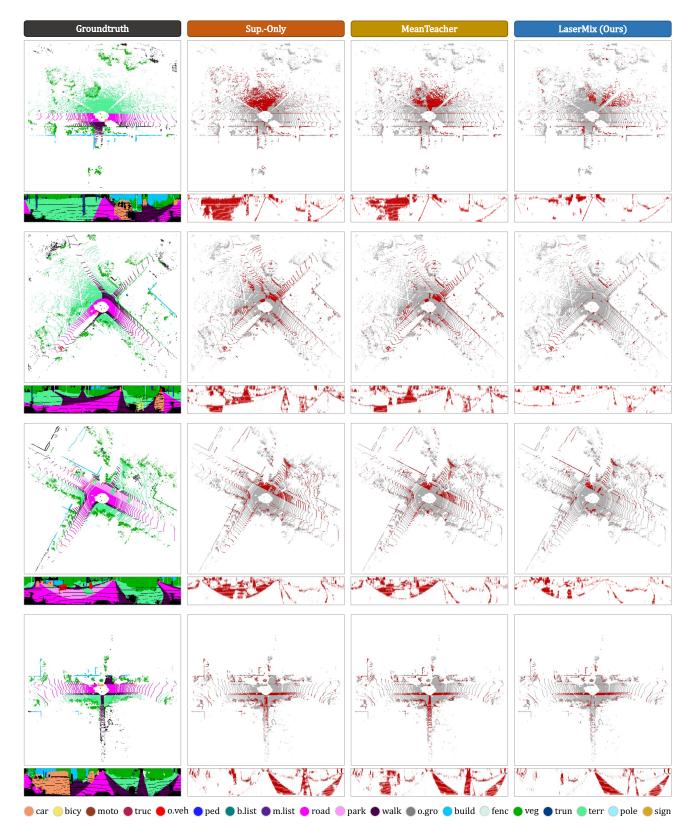


Figure A. Additional qualitative results (error maps). Models are trained with 1% labeled data on SemanticKITTI [1]. To highlight the differences, the correct and incorrect predictions are painted in gray and red, respectively. Each scene is visualized from the bird's eye view (top) and range view (bottom) and covers a region of size 50m by 50m, centered around the ego-vehicle. Best viewed in colors.

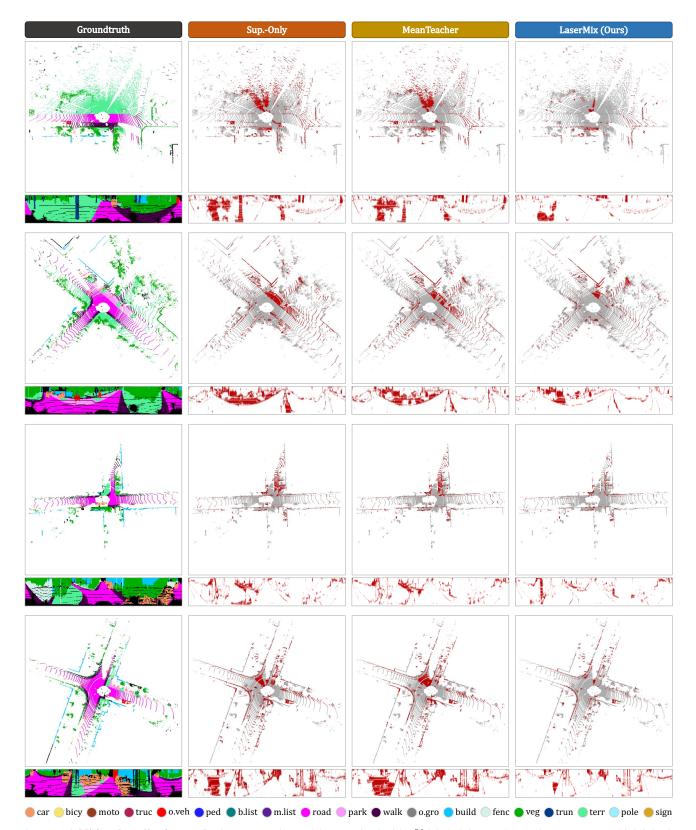


Figure B. Additional qualitative results (error maps). Models are trained with 1% labeled data on SemanticKITTI [1]. To highlight the differences, the correct and incorrect predictions are painted in gray and red, respectively. Each scene is visualized from the bird's eye view (top) and range view (bottom) and covers a region of size 50m by 50m, centered around the ego-vehicle. Best viewed in colors.

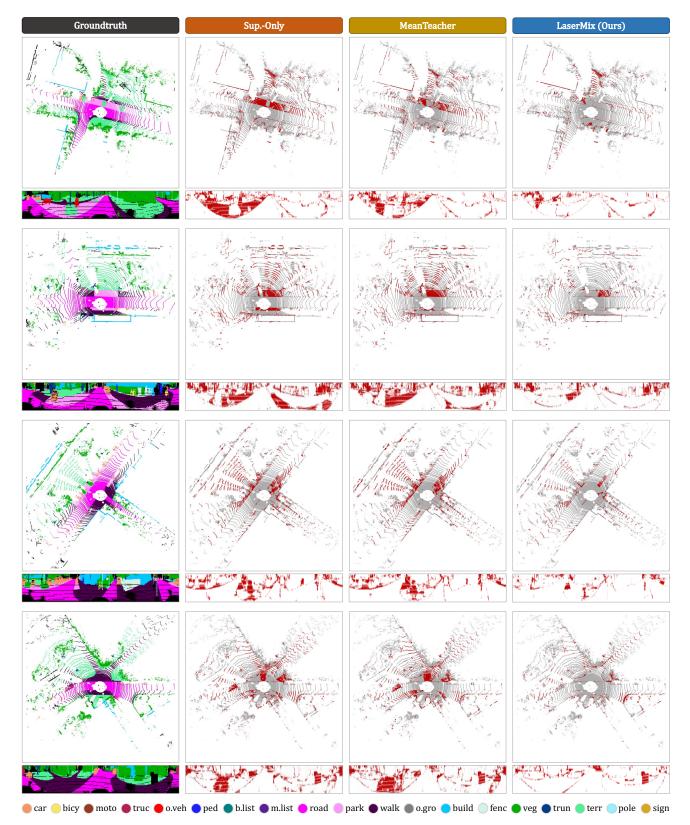


Figure C. Additional qualitative results (error maps). Models are trained with 1% labeled data on SemanticKITTI [1]. To highlight the differences, the correct and incorrect predictions are painted in gray and red, respectively. Each scene is visualized from the bird's eye view (top) and range view (bottom) and covers a region of size 50m by 50m, centered around the ego-vehicle. Best viewed in colors.

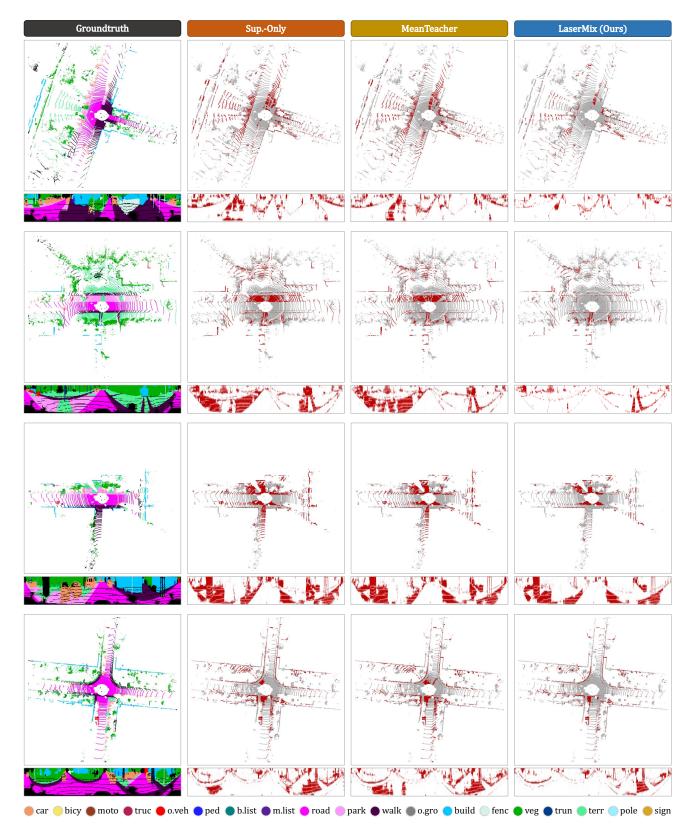


Figure D. Additional qualitative results (error maps). Models are trained with 1% labeled data on SemanticKITTI [1]. To highlight the differences, the correct and incorrect predictions are painted in gray and red, respectively. Each scene is visualized from the bird's eye view (top) and range view (bottom) and covers a region of size 50m by 50m, centered around the ego-vehicle. Best viewed in colors.

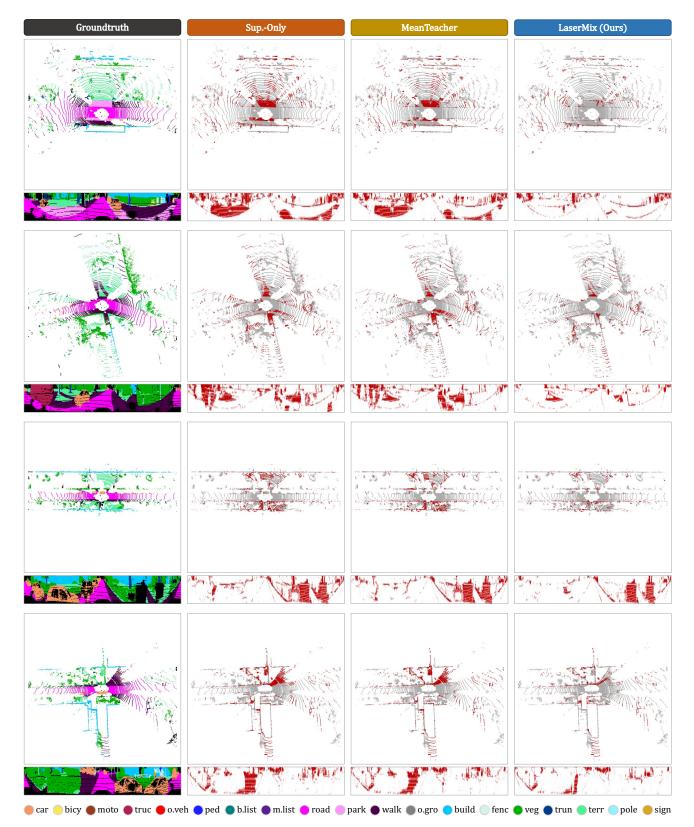


Figure E. Additional qualitative results (error maps). Models are trained with 1% labeled data on SemanticKITTI [1]. To highlight the differences, the correct and incorrect predictions are painted in gray and red, respectively. Each scene is visualized from the bird's eye view (top) and range view (bottom) and covers a region of size 50m by 50m, centered around the ego-vehicle. Best viewed in colors.

References

- Jens Behley, Martin Garbade, Andres Milioto, Jan Quenzel, Sven Behnke, Cyrill Stachniss, and Juergen Gall. Semantickitti: A dataset for semantic scene understanding of lidar sequences. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9297–9307, 2019. 2, 3, 4, 5, 7, 8, 11, 12, 13, 14, 15
- [2] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11621–11631, 2020. 4, 6, 9, 10
- [3] Qi Chen, Sourabh Vora, and Oscar Beijbom. Polarstream: Streaming lidar object detection and segmentation with polar pillars. In Advances in Neural Information Processing Systems (NeurIPS), 2021. 1
- [4] Xiaokang Chen, Yuhui Yuan, Gang Zeng, and Jingdong Wang. Semi-supervised semantic segmentation with cross pseudo supervision. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2613–2622, 2021. 4, 5, 6, 7, 8
- [5] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3213–3223, 2016. 5
- [6] Tiago Cortinhal, George Tzelepis, and Eren Erdal Aksoy. Salsanext: Fast, uncertainty-aware semantic segmentation of lidar point clouds for autonomous driving. arXiv preprint arXiv:2003.03653, 2020. 1
- [7] Whye Kit Fong, Rohit Mohan, Juana Valeria Hurtado, Lubing Zhou, Holger Caesar, Oscar Beijbom, and Abhinav Valada. Panoptic nuscenes: A large-scale benchmark for lidar panoptic segmentation and tracking. *IEEE Robotics and Automation Letters*, pages 3795–3802, 2022. 2, 3, 4, 5
- [8] Geoff French, Timo Aila, Samuli Laine, Michal Mackiewicz, and Graham Finlayson. Semi-supervised semantic segmentation needs strong, high-dimensional perturbations. In *British Machine Vision Conference (BMVC)*, 2020. 4, 6, 7, 8
- [9] Qingyong Hu, Bo Yang, Linhai Xie, Stefano Rosa, Yulan Guo, Niki Trigoni Zhihua Wang, and Andrew Markham. Randla-net: Efficient semantic segmentation of large-scale point clouds. In *IEEE/CVF Conference on Computer Vision* and Pattern Recognition (CVPR), pages 11108–11117, 2020.
- [10] Li Jiang, Shaoshuai Shi, Zhuotao Tian, Xin Lai, Shu Liu, Chi-Wing Fu, and Jiaya Jia. Guided point contrastive learning for semi-supervised point cloud semantic segmentation. In *IEEE/CVF International Conference on Computer Vision* (*ICCV*), pages 6423–6432, 2021. 4
- [11] Zhanghan Ke, Di Qiu, Kaican Li, Qiong Yan, and Rynson WH Lau. Guided collaborative training for pixel-wise semi-supervised learning. In *European Conference on Computer Vision (ECCV)*, pages 429–445, 2020. 5

- [12] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations (ICLR)*, 2018. 4
- [13] Andres Milioto, Ignacio Vizzo, Jens Behley, and Cyrill Stachniss. Rangenet++: Fast and accurate lidar semantic segmentation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4213–4220, 2019. 1
- [14] Yassine Ouali, Céline Hudelot, and Myriam Tami. Semisupervised semantic segmentation with cross-consistency training. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12674–12684, 2020. 5
- [15] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 652–660, 2017. 1
- [16] Leslie N. Smith and Nicholay Topin. Super-convergence: Very fast training of neural networks using large learning rates. arXiv preprint arXiv:1708.07120, 2017. 4
- [17] Haotian Tang, Zhijian Liu, Shengyu Zhao, Yujun Lin, Ji Lin, Hanrui Wang, and Song Han. Searching efficient 3d architectures with sparse point-voxel convolution. In *European Conference on Computer Vision (ECCV)*, pages 685–702, 2020.
- [18] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In Advances in Neural Information Processing Systems (NeurIPS), volume 30, 2017. 4, 5, 6, 7, 8
- [19] 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 *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6411–6420, 2019. 1
- [20] Ozan Unal, Dengxin Dai, and Luc Van Gool. Scribblesupervised lidar semantic segmentation. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022. 2, 4, 5, 8
- [21] Chenfeng Xu, Bichen Wu, Zining Wang, Wei Zhan, Peter Vajda, Kurt Keutzer, and Masayoshi Tomizuka. Squeezesegv3: Spatially-adaptive convolution for efficient pointcloud segmentation. In *European Conference on Computer Vision (ECCV)*, pages 1–19, 2020. 1
- [22] Yang Zhang, Zixiang Zhou, Philip David, Xiangyu Yue, Zerong Xi, Boqing Gong, and Hassan Foroosh. Polarnet: An improved grid representation for online lidar point clouds semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9601–9610, 2020. 1
- [23] Yiming Zhao, Lin Bai, and Xinming Huang. Fidnet: Lidar point cloud semantic segmentation with fully interpolation decoding. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4453–4458, 2021. 1, 2, 4
- [24] Xinge Zhu, Hui Zhou, Tai Wang, Fangzhou Hong, Yuexin Ma, Wei Li, Hongsheng Li, and Dahua Lin. Cylindrical and

asymmetrical 3d convolution networks for lidar segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9939–9948, 2021. 1, 3, 4

[25] Yang Zou, Zhiding Yu, B V K Vijaya Kumar, and Jinsong Wang. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *European Conference on Computer Vision (ECCV)*, pages 289–305, 2018. 4, 6, 7, 8