LaserMix for Semi-Supervised LiDAR Semantic Segmentation

Lingdong Kong[1,2,3,*] Jiawei Ren[1,*] Liang Pan[1] Ziwei Liu[1,✉]
[1]S-Lab, Nanyang Technological University [2]National University of Singapore [3]CNRS@CREATE
{lingdong001,jiawei011}@e.ntu.edu.sg {liang.pan,ziwei.liu}@ntu.edu.sg

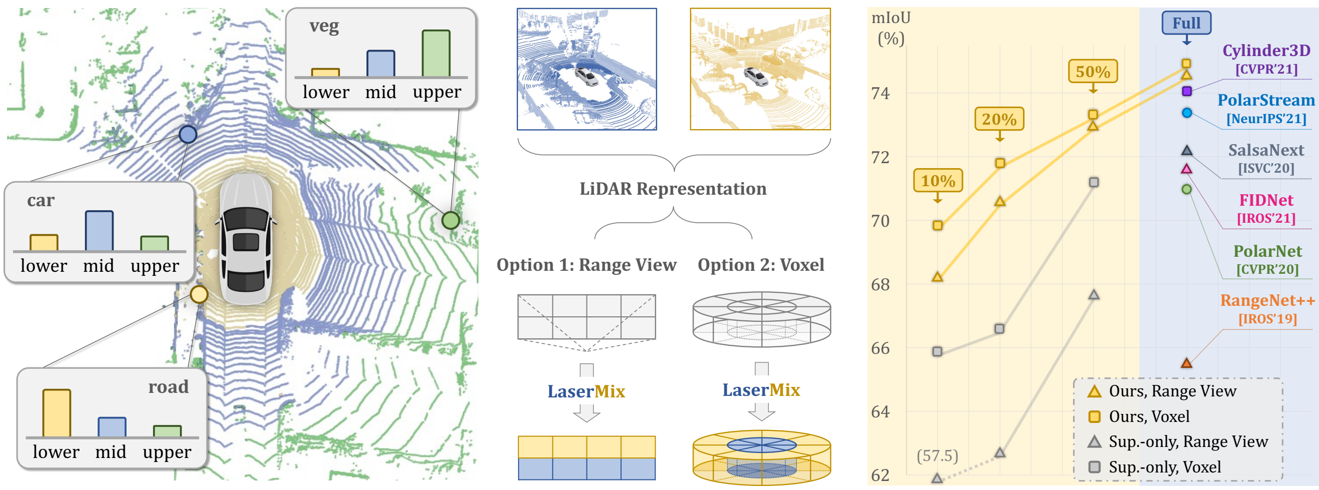


Figure 1. **Left:** The LiDAR point cloud contains strong spatial prior. Objects and backgrounds around the ego-vehicle have a patterned distribution on different (lower, middle, upper) laser beams. **Middle:** Following the scene structure, the proposed LaserMix blends beams from different LiDAR scans, which is compatible with various popular LiDAR representations. **Right:** We achieve superior results over SoTA methods in both low-data (10%, 20%, and 50% semantic labels) and high-data (full semantic labels) regimes on nuScenes [11].

Abstract

*Densely annotating LiDAR point clouds is costly, which often restrains the scalability of fully-supervised learning methods. In this work, we study the underexplored semi-supervised learning (SSL) in LiDAR semantic segmentation. Our core idea is to leverage the strong spatial cues of LiDAR point clouds to better exploit unlabeled data. We propose **LaserMix** to mix laser beams from different LiDAR scans and then encourage the model to make consistent and confident predictions before and after mixing. Our framework has three appealing properties. **1) Generic:** LaserMix is agnostic to LiDAR representations (e.g., range view and voxel), and hence our SSL framework can be universally applied. **2) Statistically grounded:** We provide a detailed analysis to theoretically explain the applicability of the proposed framework. **3) Effective:** Comprehensive experimental analysis on popular LiDAR segmentation datasets (nuScenes, SemanticKITTI, and ScribbleKITTI) demonstrates our effectiveness and superiority. Notably, we achieve competitive results over fully-supervised counterparts with $2\times$ to $5\times$ fewer labels and improve the supervised-only baseline significantly by relatively 10.8%. We hope this concise yet high-performing framework could facilitate future research in semi-supervised LiDAR segmentation. Code is publicly available[1].*

1. Introduction

LiDAR segmentation is one of the fundamental tasks in autonomous driving perception [41]. It enables autonomous vehicles to semantically perceive the dense 3D structure of the surrounding scenes [15, 34, 39]. However, densely annotating LiDAR point clouds is inevitably expensive and labor-intensive [18, 23, 47], which restrains the scalability of fully-supervised LiDAR segmentation methods. Semi-supervised learning (SSL) that directly leverages the easy-to-acquire unlabeled data is hence a viable and promising

(∗) Lingdong and Jiawei contributed equally to this work. (✉) Ziwei serves as the corresponding author. E-mail: ziwei.liu@ntu.edu.sg.

[1]https://github.com/ldkong1205/LaserMix.

solution to achieve scalable LiDAR segmentation [13, 14].

Yet, semi-supervised LiDAR segmentation is still under-explored. Modern SSL frameworks are mainly designed for image recognition [2, 3, 42] and semantic segmentation [6, 21, 37] tasks, which only yield sub-par performance on LiDAR data due to the large modality gap between 2D and 3D. Recent research [20] proposed to consider semi-supervised point cloud semantic segmentation as a fresh task and proposed a point contrastive learning framework. However, their solutions do not differentiate indoor and outdoor scenes and therefore overlook the intrinsic and important properties that only exist in LiDAR point clouds.

In this work, we explore the use of spatial prior for semi-supervised LiDAR segmentation. Unlike the general 2D/3D segmentation tasks, the spatial cues are especially significant in LiDAR data. In fact, LiDAR point clouds serve as a perfect reflection of real-world distributions, which is highly dependent on the spatial areas in the LiDAR-centered 3D coordinates. As shown in Fig. 1 (left), the top laser beams travel outward long distance and perceive mostly *vegetation*, while the middle and bottom beams tend to detect *car* and *road* from the medium and close distances, respectively. To effectively leverage this strong spatial prior, we propose **LaserMix** to mix laser beams from different LiDAR scans, and then encourage the LiDAR segmentation model to make consistent and confident predictions before and after mixing. Our SSL framework is statistically grounded, which consists of the following components:

1) Partitioning the LiDAR scan into low-variation areas. We observe a strong distribution pattern on laser beams as shown in Fig. 1 (left) and thus propose the laser partition.

2) Efficiently mixing every area in the scan with foreign data and obtaining model predictions. We propose LaserMix to manipulate the laser-grouped areas from two LiDAR scans in an intertwining way as depicted in Fig. 1 (middle) and serves as an efficient LiDAR mixing strategy for SSL.

3) Encouraging models to make confident and consistent predictions on the same area in different mixing. We hence propose a mixing-based teacher-student training pipeline.

Despite the simplicity of our overall pipeline, it achieves competitive results over the fully supervised counterpart using $2\times$ to $5\times$ fewer labels as shown in Fig. 1 (right) and significantly outperforms all prevailing semi-supervised segmentation methods on nuScenes [11] (up to $+5.7\%$ mIoU) and SemanticKITTI [1] (up to $+3.5\%$ mIoU). Moreover, LaserMix directly operates on point clouds so as to be agnostic to different LiDAR representations, *e.g.*, range view [32] and voxel [58]. Therefore, our pipeline is highly compatible with existing state-of-the-art (SoTA) LiDAR segmentation methods under various representations [46, 56, 57]. Besides, our pipeline achieves competitive performance using very limited annotations on weak supervision dataset [47]: it achieves 54.4% mIoU on SemanticKITTI [1] using only 0.8% labels, which is on-par with PolarNet [56] (54.3%), RandLA-Net [19] (53.9%), and RangeNet++ [32] (52.2%) using 100% labels. Spatial prior is proven to play a pivotal role in the success of our framework through comprehensive empirical analysis. To summarize, this work has the following key contributions:

- We present a statistically grounded SSL framework that effectively leverages the spatial cues in LiDAR data to facilitate learning with semi-supervisions.
- We propose LaserMix, a novel and representation-agnostic mixing technique that strives to maximize the "strength" of the spatial cues in our SSL framework.
- Our overall pipeline significantly outperforms previous SoTA methods in both low- and high-data regimes. We hope this work could lay a solid foundation for semi-supervised LiDAR segmentation.

2. Related Work

LiDAR Segmentation. Various approaches from different aspects have been proposed for LiDAR scene segmentation, *i.e.*, range view [8, 32, 49, 57], bird's eye view [56], voxel [44, 58], and multi-view [26, 50] methods. Although appealing results have been achieved, these fully-supervised methods rely on large-scale annotated LiDAR datasets and their performance would degrade severely in the low-data regime [13]. Recent works seek weak [18,55], scribble [47], and box [28] supervisions or activate learning [27, 31] to ease the annotation cost. We tackle this problem from the perspective of semi-supervised learning (SSL), aiming at directly leveraging the easy-to-acquire unlabeled data to boost the LiDAR semantic segmentation performance.

SSL in 2D. Well-known SSL algorithms are prevailing in handling image recognition problems [2, 3, 24, 42, 45]. In the context of semantic segmentation, CutMix-Seg [12] and PseudoSeg [60] apply perturbations on inputs and hope the decision boundary lies in the low-density region. CPS [6] and GCT [21] enforce consistency between two perturbed networks [29]. These perturbations [30, 36, 52], however, are either inapplicable or only yield sub-par results in 3D. Another line of research is entropy minimization. Methods like CBST [59] and ST++ [51] generate pseudo-labels [25] offline per round during self-training. The extra storage needed might become costly for large-scale LiDAR datasets [1, 4, 11]. Our framework encourages both consistency regularization and entropy minimization and does not require extra overhead, which better maintains scalability.

SSL in 3D. Most works focus on developing SSL for object-centric point clouds [22, 43] or indoor scenes [7, 9, 17, 34], whose scale and diversity are much lower than the outdoor LiDAR point clouds [1, 4]. Some other works [38, 40, 48] try to utilize SSL for 3D object detection on LiDAR data. A recent work [20] tackles semi-supervised point cloud semantic segmentation using contrastive learning, but it still mainly focused on indoor scenes and does not distinguish between the uniformly distributed indoor point clouds and the spatially structured LiDAR point clouds. We are one of the first works to explore SSL for LiDAR segmentation. Our work also establishes comprehensive SSL benchmarks upon popular autonomous driving databases [1, 11, 47].

3. Approach

In this section, we first introduce our SSL framework that leverages the spatial prior of LiDAR data by encouraging confidence and consistency in predictions (Sec. 3.1). We then present LaserMix which strives to maximize the "strength" of the spatial prior and mixes LiDAR scans in an efficient manner (Sec. 3.2). Finally, we elaborate on the overall pipeline (Sec. 3.3) and the pseudo-code (Algo. 1).

3.1. Leveraging the Spatial Prior for SSL

Spatial Prior Formulation. The distribution of real-world objects/backgrounds has a strong correlation to their spatial positions in the LiDAR scan. Objects/backgrounds inside a specified spatial area of a LiDAR point cloud follow similar patterns, *e.g.*, the close-range area is most likely *road* while the long-range area consists of *building*, *vegetation*, etc. In another word, there exists a spatial area $a \in A$ where LiDAR points and semantic labels inside the area (denoted as X_{in} and Y_{in}, respectively) will have relatively low variations. Formally, the conditional entropy $H(X_{\text{in}}, Y_{\text{in}}|A)$ is smaller. Therefore, when estimating the parameter θ of the segmentation network $\mathcal{G}_\theta$, we would expect:

$$\mathbb{E}_\theta[H(X_{\text{in}}, Y_{\text{in}}|A)] = c\,, \tag{1}$$

where c is a small constant. Similar to the classic entropy minimization [16], the constraint in Eq. (1) can be converted to a prior on the model parameter θ using the principle of maximum entropy:

$$\begin{aligned} P(\theta) &\propto \exp(-\lambda H(X_{\text{in}}, Y_{\text{in}}|A)) \\ &\propto \exp(-\lambda H(Y_{\text{in}}|X_{\text{in}}, A))\,, \end{aligned} \tag{2}$$

where $\lambda > 0$ is the Lagrange multiplier corresponding to constant c; $H(X_{\text{in}}|A)$ has been ignored for being independent of the model parameter θ. We consider Eq. (2) as the formal formulation of the spatial prior and discuss how to empirically compute it in the following sections.

Marginalization. To utilize the spatial prior defined in Eq. (2), we empirically compute the entropy $H(Y_{\text{in}}|X_{\text{in}}, A)$ of the LiDAR points *inside* area A as follows:

$$\begin{aligned} &\hat{H}(Y_{\text{in}}|X_{\text{in}}, A) = \\ &\hat{\mathbb{E}}_{X_{\text{in}},Y_{\text{in}},A}[P(Y_{\text{in}}|X_{\text{in}}, A)\log P(Y_{\text{in}}|X_{\text{in}}, A)]\,, \end{aligned} \tag{3}$$

where $\hat{\cdot}$ denotes the empirical estimation. The end-to-end LiDAR segmentation model $\mathcal{G}_\theta$ usually takes full-sized data as inputs during inference. Therefore, to compute $P(Y_{\text{in}}|X_{\text{in}}, A)$ in Eq. (3), we first pad the data *outside* the area to obtain the full-sized data. Here we denote the data *outside* the area as X_{out}; we then let the model infer $P(Y_{\text{in}}|X_{\text{in}}, X_{\text{out}}, A)$, and finally marginalize X_{out} as:

$$P(Y_{\text{in}}|X_{\text{in}}, A) = \hat{\mathbb{E}}_{X_{\text{out}}}[P(Y_{\text{in}}|X_{\text{in}}, X_{\text{out}}, A)]\,. \tag{4}$$

The generative distribution of the padding $P(X_{\text{out}})$ can be directly obtained from the dataset.

Training. Finally, we train the segmentation model $\mathcal{G}_\theta$ using the standard maximum-a-posteriori (MAP) estimation. We maximize the posterior that can be computed by Eq. (2), Eq. (3) and Eq. (4), which is formulated as follows:

$$\begin{aligned} &C(\theta) = L(\theta) - \lambda\hat{H}(Y_{\text{in}}|X_{\text{in}}, A) = L(\theta) \\ &- \lambda\hat{\mathbb{E}}_{X_{\text{in}},Y_{\text{in}},A}[P(Y_{\text{in}}|X_{\text{in}}, A)\log P(Y_{\text{in}}|X_{\text{in}}, A)]. \end{aligned} \tag{5}$$

Here, $L(\theta)$ is the likelihood function computed using labeled data, *i.e.*, the conventional supervised learning. Minimizing $\hat{H}(Y_{\text{in}}|X_{\text{in}}, A)$ requires the marginal probability $P(Y_{\text{in}}|X_{\text{in}}, A)$ to be confident, which further requires $P(Y_{\text{in}}|X_{\text{in}}, X_{\text{out}}, A)$ to be both confident and consistent with respect to different outside data X_{out}.

In summary, our proposed SSL framework in Eq. (5) encourages the segmentation model to make confident and consistent predictions at a predefined area, regardless of the data outside the area. The predefined area set A determines the "strength" of the prior. When setting A to the full area (*i.e.*, the whole point cloud), our framework degrades to the classic entropy minimization framework [16].

Implementation. There are three key steps for implementing our framework:

- *Step 1)*: Select a proper partition set A which maintains strong spatial prior;
- *Step 2)*: Efficiently compute the marginal probability, *i.e.*, $P(Y_{\text{in}}|X_{\text{in}}, A)$;
- *Step 3)*: Efficiently minimize the marginal entropy, *i.e.*, $\hat{H}(Y_{\text{in}}|X_{\text{in}}, A)$.

We propose a simple yet effective implementation following these steps in the next section.

3.2. LaserMix

Partition. LiDAR sensors have a fixed number (*e.g.*, 32, 64, and 128) of laser beams which are emitted isotropically around the ego-vehicle with predefined inclination angles as shown in Fig. 2. To obtain a proper set of spatial areas A, we propose to partition the LiDAR point cloud based on laser beams. Specifically, points captured by the same laser beam have a unified inclination angle to the sensor plane. For point i, its inclination ϕ_i is defined as follows:

$$\phi_i = \arctan\left(\frac{p_i^z}{\sqrt{(p_i^x)^2 + (p_i^y)^2}}\right), \tag{6}$$

where (p^x, p^y, p^z) is the Cartesian coordinates of the LiDAR points. Given two LiDAR scans x_1 and x_2, we first group all points from each scan by their inclination angles. Concretely, to form m non-overlapping areas, a set of $m+1$ inclination angles $\Phi = \{\phi_0, \phi_1, \phi_2, ..., \phi_m\}$ will be evenly sampled within the range of the minimum and maximum inclination angles in the dataset (defined by sensor configurations), and the area set $A = \{a_1, a_2, ..., a_m\}$ can be formed by bounding area a_i in the inclination range $[\phi_{i-1}, \phi_i)$.

Role in our framework: Laser partition effectively "excites" a strong spatial prior in the LiDAR point cloud, as described by *Step 1* in our framework. As shown in Fig. 1 (left), we find an overt pattern in semantic classes detected by each laser beam. More concrete evidence on this aspect has been included in the Appendix. Despite being an empirical choice, we will show in later sections that laser partition significantly outperforms other partition choices, including random points (*MixUp*-like partition [54]), random areas (*CutMix*-like partition [53]), and other heuristics like azimuth α (sensor horizontal direction) or radius r (sensor range direction) partitions.

Mixing. To this end, we propose LaserMix, a simple yet effective LiDAR mixing strategy that can better control the "strength" of the spatial prior. LaserMix mixes the aforementioned laser partitioned areas A from two scans in an intertwining way, *i.e.*, one takes from odd-indexed areas $A_1 = \{a_1, a_3, ...\}$ and the other takes from even-indexed areas $A_2 = \{a_2, a_4, ...\}$, so that each area's neighbor will be from the other scan:

$$\begin{aligned}\tilde{x}_1, \tilde{x}_2 &= \text{LaserMix}(x_1, x_2)\,,\\ \tilde{x}_1 &= x_1^{a_1} \cup x_2^{a_2} \cup x_1^{a_3} \cup \cdots\,,\\ \tilde{x}_2 &= x_2^{a_1} \cup x_1^{a_2} \cup x_2^{a_3} \cup \cdots\,,\end{aligned} \tag{7}$$

where $x_i^{a_j}$ is the data crop of x_i in the area a_j. The semantic labels are mixed in the same way. LaserMix is directly applied to the point clouds and is thus agnostic to the various LiDAR representations [19,32,56,58]. We show LaserMix's instantiations with the *range view* and *voxel* representations as in Fig. 1 (middle), since they are currently the most efficient and the best-performing options, respectively.

Role in our framework: LaserMix helps to efficiently compute the marginal probability $P(Y_{\text{in}}|X_{\text{in}}, A)$, as described by *Step 2* in our framework. The cost for directly computing the marginal probability in Eq. (4) on real-world LiDAR data is prohibitive; we need to iterate through all areas in A and all outside data in X_{out}, which requires $|A|\cdot|X_{\text{out}}|$ predictions in total. To reduce the training overhead, we take advantage of the fact that a prediction in an area will be largely affected by its neighboring areas and let X_{out} fill only the neighbors instead of all the remaining areas. LaserMix mixes two scans by *intertwining* the areas so that the neighbors of each area are filled with data from the other scan. As a result, we obtain the prediction on

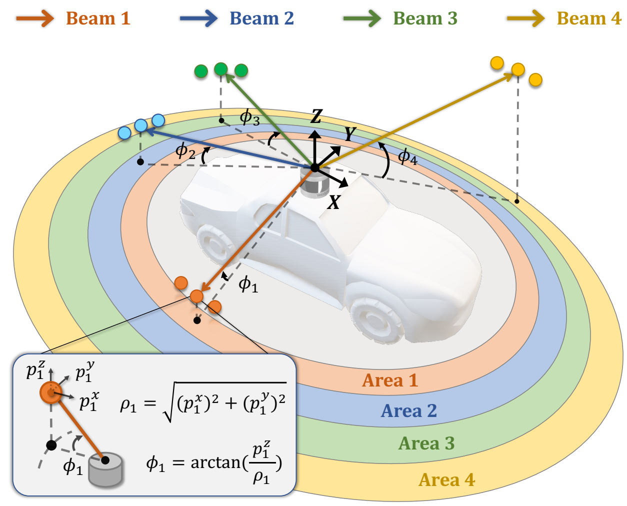


Figure 2. **Laser partition example**. We group LiDAR points (p_i^x, p_i^y, p_i^z) whose inclinations ϕ_i are within the same inclination range into the same area, as depicted in the color regions.

all areas A of two scans from only two predictions, which on average reduces the cost from $|A|$ to 1. The scan before and after mixing counts as two data fillings, therefore $|X_{\text{out}}| = 2$. Overall, the training overhead is reduced from $|A|\cdot|X_{\text{out}}|$ to 2: only one prediction on original data and one additional prediction on mixed data are required for each LiDAR scan. During training, the memory consumption for a batch will be $2\times$ compared to a standard SSL framework, and the training speed will not be affected.

3.3. Overall Pipeline

We show the overall framework in Fig. 3 and the pseudo-code in Algo. 1. There are two branches in our pipeline, one Student net $\mathcal{G}_\theta^s$ and one Teacher net $\mathcal{G}_\theta^t$. During training, a batch is composed of half labeled data and half unlabeled data. We collect the predictions from both $\mathcal{G}_\theta^s$ and $\mathcal{G}_\theta^t$, and produce pseudo-labels from Teacher net's prediction with a predefined confidence threshold T. For labeled data, we compute the cross-entropy loss between the Student net's prediction and the ground-truth as $\mathcal{L}_{\text{sup}}$. For unlabeled data, LaserMix blends every scan with a random labeled scan, together with their pseudo-label or ground-truth. Then, we let $\mathcal{G}_\theta^s$ predict on the mixed data and compute the cross-entropy loss $\mathcal{L}_{\text{mix}}$ (*w/* mixed labels). The point-wise cross-entropy loss for a scan x and its corresponding ground-truth/pseudo-label y on the segmentation net $\mathcal{G}_\theta$ is defined as:

$$\mathcal{L}_{\text{ce}} = \frac{1}{|x|}\sum_{i=1}^{|x|} \text{CrossEntropy}(y^{(i)}, \mathcal{G}_\theta^{(i)}(x))\,, \tag{8}$$

where (i) denotes the i-th point. Moreover, we adopt the mean teacher idea in [45] and use Exponential Moving Average (EMA) to update the weights of $\mathcal{G}_\theta^t$ from $\mathcal{G}_\theta^s$, and com-

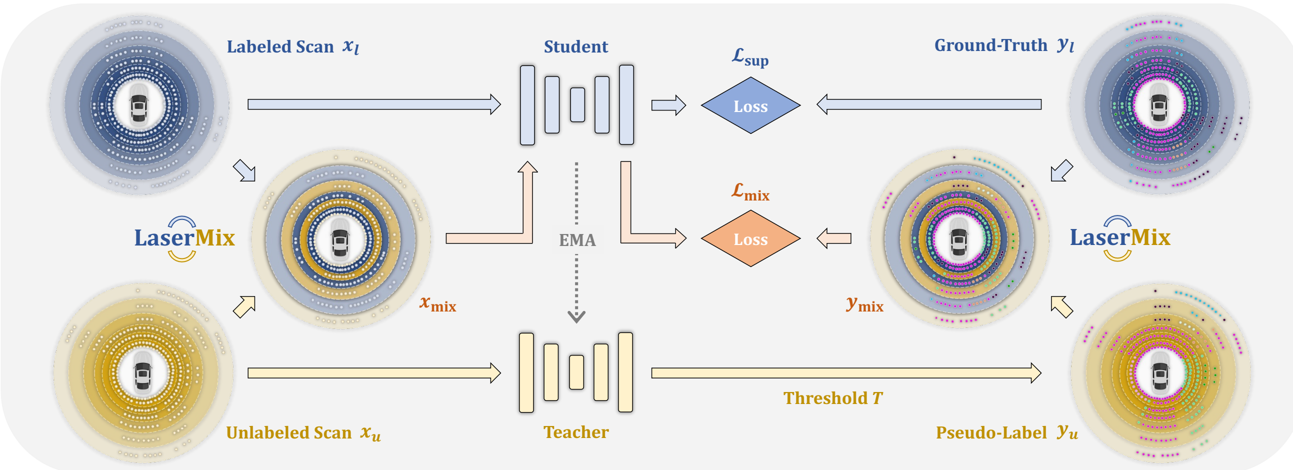


Figure 3. **Framework overview**. The labeled scan x_l is fed into the Student net to compute the supervised loss $\mathcal{L}_{\text{sup}}$ (*w/* ground-truth y_l). The unlabeled scan x_u and the generated pseudo-label y_u are mixed with (x_l, y_l) via LaserMix (Sec. 3.2) to produce mixed data sample $(x_{\text{mix}}, y_{\text{mix}})$, which is then fed into the Student net to compute the mixing loss $\mathcal{L}_{\text{mix}}$. Additionally, we adopt the EMA update in [45] for the Teacher net and compute the mean teacher loss $\mathcal{L}_{\text{mt}}$ over Student net's and Teacher net's predictions.

pute the L2 loss between their predictions as $\mathcal{L}_{\text{mt}}$:

$$\mathcal{L}_{\text{mt}} = ||\mathcal{G}_\theta^s(x) - \mathcal{G}_\theta^t(x)||_2^2 \,, \tag{9}$$

where $||\cdot||_2^2$ is the L2 norm. The overall loss function is $\mathcal{L} = \mathcal{L}_{\text{sup}} + \lambda_{\text{mix}}\mathcal{L}_{\text{mix}} + \lambda_{\text{mt}}\mathcal{L}_{\text{mt}}$, where λ_{mix} and λ_{mt} are loss weights. We use the Teacher net during inference as it empirically gives more stable results. There will be no extra inference overhead in our framework.

Role in our framework: Our overall pipeline minimizes the marginal entropy, as described by *Step 3* in our framework. Since the objective for minimizing the entropy has a hard optimization landscape, pseudo-labeling is a common resort in practice [25]. Unlike conventional pseudo-label optimization in SSL which only aims to encourage the predictions to be confident, minimizing the marginal entropy requires all predictions to be both confident and consistent. Therefore, we use the ground-truth and pseudo-label as an anchor and encourage the model's predictions to be confident and consistent with these supervision signals.

Algorithm 1 Pseudo-code for one training iteration.

```
Input: Shuffled labeled batch (X_l, Y_l) = {(x_l^(b), y_l^(b)); b ∈ (1,...,B)}, shuffled unlabeled batch X_u = {x_u^(b); b ∈ (1,...,B)}, threshold T, loss weights λ_mix and λ_mt, Student net and Teacher net.
for b = 1 to B do
    x_mix^(2b-1), x_mix^(2b) = LaserMix(x_l^(b), x_u^(b))    // LaserMix data
end for
X_mix = {x_mix^(i); i ∈ (1,...,2B)}
S_l, S_u, S_mix = Student(Concat(X_l, X_u, X_mix))    // Student pred
Ŝ_l, Ŝ_u = Teacher(Concat(X_l, X_u))    // Teacher pred
Y_u = PseudoLabel(Ŝ_u, T)    // Pseudo-label generation process
for b = 1 to B do
    y_mix^(2b-1), y_mix^(2b) = LaserMix(y_l^(b), y_u^(b))    // LaserMix label
end for
Y_mix = {y_mix^(i); i ∈ (1,...,2B)}
L_sup = CrossEntropy(S_l, Y_l)    // Supervised loss
L_mix = CrossEntropy(S_mix, Y_mix)    // Mixing loss
L_mt = L2(Concat(S_l, S_u), Concat(Ŝ_l, Ŝ_u))    // MeanTeacher loss
L = L_sup + λ_mix L_mix + λ_mt L_mt    // Overall loss
Backward(L), Update(Student), UpdateEMA(Teacher)
```

4. Experiments

4.1. Settings

Protocol. We follow the Realistic Evaluation Protocol [35] when building the benchmark. Specifically, all experiments share the same backbones and are within the same codebase. All configurations and augmentations are unified to ensure a fair comparison among different SSL algorithms.

Data. We build three SSL benchmarks upon nuScenes [11], SemanticKITTI [1], and ScribbleKITTI [47]. nuScenes [11] and SemanticKITTI [1] are the two most popular LiDAR segmentation datasets, with 29130 and 19130 training scans and 6019 and 4071 validation scans, respectively. ScribbleKITTI [47] is a recent variant of SemanticKITTI [1], which contains the same number of scans but is annotated with line scribbles (approximately 8.06% valid semantic labels) rather than dense annotations. For all three LiDAR segmentation datasets, we uniformly sample 1%, 10%, 20%, and 50% labeled training scans and assume the remaining ones as unlabeled. This is in line with the conventional settings from the semi-supervised image segmentation community [6, 21, 37]. We also conduct experiments with sequential splits and show the results in the Appendix.

Implementation Details. We adopt FIDNet [57] and Cylinder3D [58] as the segmentation backbones for the *range*

Table 1. Benchmarking results among different SSL methods with the LiDAR *range view* (top) and *voxel* (bottom) representations. All mIoU scores are given in percentage (%). The **best** and second best score for each data split is highlighted in **bold** and underline.

Repr.	Method	nuScenes [11]				SemanticKITTI [1]				ScribbleKITTI [47]			
		1%	10%	20%	50%	1%	10%	20%	50%	1%	10%	20%	50%
Range View	*Sup.-only*	38.3	57.5	62.7	67.6	36.2	52.2	55.9	57.2	33.1	47.7	49.9	52.5
	MeanTeacher [45]	42.1	60.4	65.4	69.4	37.5	53.1	56.1	57.4	34.2	49.8	51.6	53.3
	CBST [59]	40.9	60.5	64.3	69.3	39.9	53.4	56.1	56.9	35.7	50.7	52.7	54.6
	CutMix-Seg [12]	43.8	63.9	64.8	69.8	37.4	54.3	56.6	57.6	36.7	50.7	52.9	54.3
	CPS [6]	40.7	60.8	64.9	68.0	36.5	52.3	56.3	57.4	33.7	50.0	52.8	54.6
	LaserMix (Ours)	**49.5**	**68.2**	**70.6**	**73.0**	**43.4**	**58.8**	**59.4**	**61.4**	**38.3**	**54.4**	**55.6**	**58.7**
	$\Delta\uparrow$	+11.2	+10.7	+7.9	+5.4	+7.2	+6.6	+3.5	+4.2	+5.2	+6.7	+5.7	+6.2
Voxel	*Sup.-only*	50.9	65.9	66.6	71.2	45.4	56.1	57.8	58.7	39.2	48.0	52.1	53.8
	MeanTeacher [45]	51.6	66.0	67.1	71.7	45.4	57.1	59.2	60.0	41.0	50.1	52.8	53.9
	CBST [59]	53.0	66.5	69.6	71.6	48.8	58.3	59.4	59.7	41.5	50.6	53.3	54.5
	CPS [6]	52.9	66.3	70.0	72.5	46.7	58.7	59.6	60.5	41.4	51.8	53.9	54.8
	LaserMix (Ours)	**55.3**	**69.9**	**71.8**	**73.2**	**50.6**	**60.0**	**61.9**	**62.3**	**44.2**	**53.7**	**55.1**	**56.8**
	$\Delta\uparrow$	+4.4	+4.0	+5.2	+2.0	+5.2	+3.9	+4.1	+3.6	+5.0	+5.7	+3.0	+3.0

Table 2. Comparison to the state-of-the-art 3D SSL method [20] on the *val* set of SemanticKITTI [1]. All mIoU scores are given in percentage (%).

Method	5%	10%	20%	30%	40%
GPC [20]	41.8	49.9	58.8	59.4	59.9
Ours (RV)	54.6	58.8	59.4	60.1	60.8
$\Delta\uparrow$	+12.8	+8.9	+0.6	+0.7	+0.9
Ours (Voxel)	56.7	60.0	61.9	62.1	62.3
$\Delta\uparrow$	+14.9	+10.1	+3.1	+1.7	+1.4

Table 3. Ablation results on the *val* set of nuScenes [11]. (1) Baseline results [45]; (2) Results with Student net supervision (SS); (3) Results with Teacher net supervision (TS). All mIoU scores are given in percentage (%).

#	$\mathcal{L}_{\text{mt}}$	$\mathcal{L}_{\text{mix}}$	SS	TS	1%	10%	20%	50%
(1)	✓				42.1	60.4	65.4	69.4
(2)		✓	✓		45.6(+3.5)	64.3(+3.9)	67.8(+2.4)	71.6(+2.2)
	✓	✓	✓		47.0(+4.9)	65.5(+5.1)	69.5(+4.1)	72.0(+2.6)
(3)		✓		✓	46.0(+3.9)	64.1(+3.7)	69.5(+4.1)	72.3(+2.9)
	✓	✓		✓	49.5(+7.4)	68.2(+7.8)	70.6(+5.2)	73.0(+3.6)

view and the *voxel* options, respectively. The input resolution of range images is set as 64×2048 for SemanticKITTI [1] and ScribbleKITTI [47], and 32×1920 for nuScenes [11]. The voxel resolution is fixed as $[240, 180, 20]$ for all three sets. The number of spatial areas m in LaserMix is uniformly sampled from 2 to 6 areas. We denote the supervised-only baseline as *sup.-only*. Due to the lack of LiDAR SSL works [20], we also compare SoTA consistency regularization [6, 12, 45] and entropy minimization [59] methods from semi-supervised image segmentation. We report the intersection-over-union (IoU) over each semantic class and the mean IoU (mIoU) scores over all classes in our benchmarks. All experiments are implemented using PyTorch on NVIDIA Tesla V100 GPUs with 32GB RAM.

4.2. Comparative Study

Improvements over Baseline. Tab. 1 benchmarks results on nuScenes [11], SemanticKITTI [1], and ScribbleKITTI [47]. For all three sets under different data splits, we observe significant improvements in our approach over the *sup.-only* baseline. Such gains are especially evident in *range view*, which reach up to 11.2% mIoU. We also observe constant improvements for the *voxel* option, which provide on average 4.1% mIoU gains over all splits across all sets. The results verify the effectiveness of our framework and further highlight the importance of leveraging unlabeled data in LiDAR semantic segmentation.

Compare with SoTA. We compare LaserMix with GPC[2] [20], the SoTA 3D SSL method tested on SemanticKITTI [1]. The results in Tab. 2 show that our approach exhibits much better results than GPC [20], especially in scenarios where very few annotations are available. We also reimplemented popular SSL algorithms from the image segmentation domain and show their results in Tab. 1. We find that these methods, albeit competitive in 2D, only yield sub-par performance in the LiDAR SSL benchmark, highlighting the importance of exploiting the LiDAR data structure.

Compare with Full Labels. As shown in Fig. 1 (right), the comparisons between the prevailing LiDAR segmentation methods attests that our approach can achieve more competitive scores over the fully-supervised counterparts [5, 8, 32, 56] while with $2\times$ to $5\times$ fewer annotations. The strong augmentation and regularization ability of LaserMix have yielded better results in the high-data regime and extreme low-data regime (*i.e.*, 0.8% labels on [47]), which validates the generalizability of our approach.

Qualitative Examination. Fig. 5 visualizes the scene segmentation results for different SSL algorithms on the *val*

[2]Note that GPC uses private backbone / split. See Appendix for details.

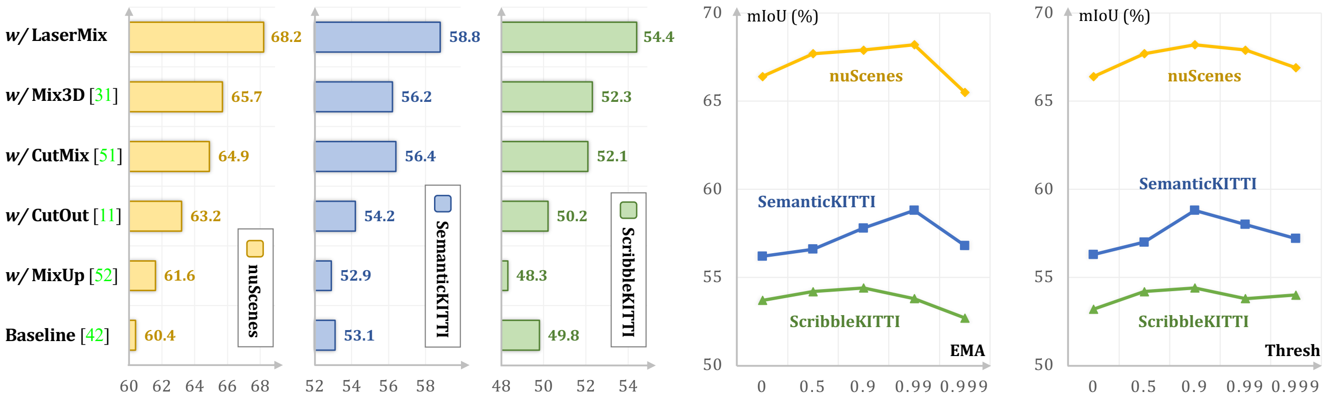


Figure 4. Ablation studies on: **Left:** Different mixing-based techniques used in point partition & mixing; **Middle:** Different EMA decay rates between the Teacher net and the Student net; **Right:** Different confidence thresholds T used in the pseudo-label generation.

set of nuScenes [11], where each example covers a driving scene centered by the ego-vehicle. We find that previous arts [6, 45, 59] can only improve predictions in limited regions, while our approach holistically eliminates false predictions in almost every direction around the ego-vehicle. The consistency enlightened by LaserMix has yielded better segmentation accuracy under annotation scarcity.

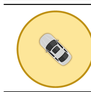
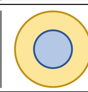
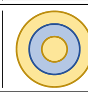
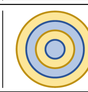
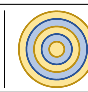
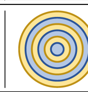
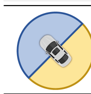
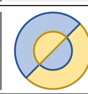
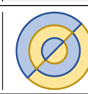
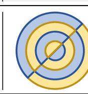
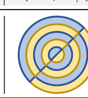
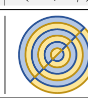
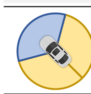
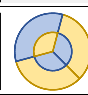
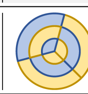
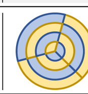
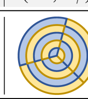
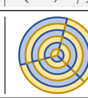
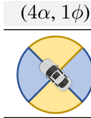
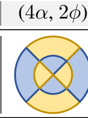
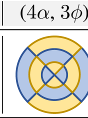
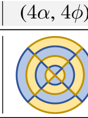
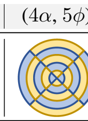
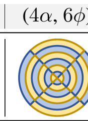
4.3. Ablation Study

Without loss of generalizability, we stick with the 10% budget setting and range view backbones in our ablations.

Framework Setups. The component analysis in Tab. 3 shows that $\mathcal{L}_{\text{mix}}$ contributes significantly to the overall improvement. Meanwhile, using the Teacher net instead of the Student net to generate pseudo-labels leads to better results, as the formal is temporally ensembled and encourages consistency in-between mixed and original data, which is crucial besides enforcing confident predictions. It is worth noting that all of our model configurations have achieved superior results than the baseline MeanTeacher [45], which further emphasizes the effectiveness of our framework in tackling LiDAR segmentation with semi-supervisions.

Mixing Strategies. Fig. 4 (left) compares LaserMix with other mixing methods [10, 33, 53, 54]. MixUp and CutMix can be considered as setting A to random points and random areas, respectively. We observe that MixUp has no improvements over the baseline on average since there is no distribution pattern in random points. CutMix has a considerable improvement over the baseline, as there is always a structure prior in scene segmentation, *i.e.*, the same semantic class points tend to cluster, which reduces the entropy in any continuous area. This prior is often used in image semantic segmentation SSL [12]. However, our spatial prior is much stronger, where not only the area structure but also the area's spatial position has been considered. LaserMix outperforms CutMix by a large margin (up to 3.3% mIoU) on all sets. CutOut can be considered as setting X_{out} to

Table 4. Ablation studies on laser beam partitions (horizontal: inclination direction ϕ; vertical: azimuth direction α). (i-α, j-ϕ) denotes that there are i azimuth and j inclination partitions.

Baseline	$(1\alpha, 2\phi)$	$(1\alpha, 3\phi)$	$(1\alpha, 4\phi)$	$(1\alpha, 5\phi)$	$(1\alpha, 6\phi)$
					
60.4	$63.5_{(+3.1)}$	$65.2_{(+4.8)}$	$66.5_{(+6.1)}$	$66.2_{(+5.8)}$	$65.4_{(+5.0)}$
$(2\alpha, 1\phi)$	$(2\alpha, 2\phi)$	$(2\alpha, 3\phi)$	$(2\alpha, 4\phi)$	$(2\alpha, 5\phi)$	$(2\alpha, 6\phi)$
					
$61.5_{(+1.1)}$	$63.3_{(+2.9)}$	$65.9_{(+5.5)}$	$66.1_{(+5.7)}$	$66.7_{(+6.3)}$	$65.3_{(+4.9)}$
$(3\alpha, 1\phi)$	$(3\alpha, 2\phi)$	$(3\alpha, 3\phi)$	$(3\alpha, 4\phi)$	$(3\alpha, 5\phi)$	$(3\alpha, 6\phi)$
					
$60.9_{(+0.6)}$	$64.2_{(+3.8)}$	$65.9_{(+5.5)}$	$66.3_{(+5.9)}$	$66.0_{(+5.6)}$	$65.2_{(+4.8)}$
$(4\alpha, 1\phi)$	$(4\alpha, 2\phi)$	$(4\alpha, 3\phi)$	$(4\alpha, 4\phi)$	$(4\alpha, 5\phi)$	$(4\alpha, 6\phi)$
					
$60.9_{(+0.6)}$	$64.7_{(+4.3)}$	$65.3_{(+4.9)}$	$65.6_{(+5.2)}$	$65.7_{(+5.3)}$	$65.2_{(+4.8)}$

a dummy filling instead of sampling from datasets, and it leads to a considerable performance drop from CutMix.

Orderless Mix. We revert the area ordering (*i.e.*, put the topmost laser beam at the bottom, and vice versa) in LaserMix, and the performance drops from 68.2% to 64.4% (−3.8% mIoU). When we shuffle the ordering, the performance drops to 63.8% (−4.4% mIoU), which becomes comparable with CutMix. The results once again confirm the superiority of using spatial prior in LiDAR SSL.

Other Heuristics. Besides our proposed inclination partition, the LiDAR scans can also be split based on azimuth (sensor horizontal direction). Results in Tab. 4 reveal that in contrast to laser partitioning, pure azimuth splitting (the

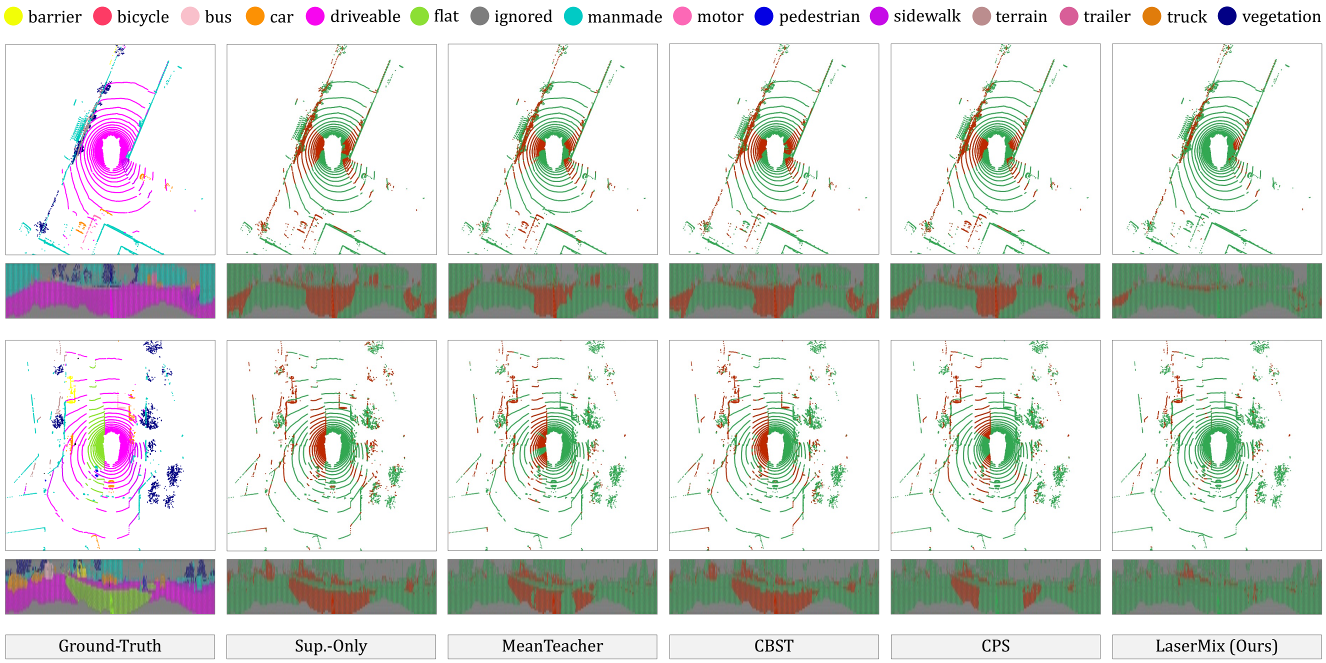


Figure 5. Error maps visualized from LiDAR *bird's eye view* and *range view* on nuScenes [11]. Each example shows a street scene of size 50m by 50m by 8m. The **correct** and **incorrect** predictions are painted in **green** and **red** to highlight the differences. Best viewed in color.

first column) does not improve the performance, which is attributed to the fact that the semantic distribution has a weak correlation in the azimuth direction. We also observe that the results tend to improve as the mixing granularity increases (row direction in Tab. 4). The scores start to drop when the granularity is beyond a certain limit (*e.g.*, the last two columns). We conjecture that over-fine-grained partition and mixing tend to hurt semantic coherence.

Mix Unlabeled Data Only. To verify that our method is more than trivially augmenting seen data and label pairs, we apply LaserMix only on unlabeled data. Instead of mixing an unlabeled scan with a labeled scan as described in Sec. 3.3, we mix two unlabeled scans with their pseudo-labels. The score drops from 68.2% to 66.9% (-1.3% mIoU) but still outperforms all existing methods by large margins. This verifies that LaserMix indeed brings a strong consistency regularization effect during SSL.

EMA. Fig. 4 (middle) provides results with different EMA decay rates, Typically, a rate between 0.9 and 0.99 yields the best possible results. Large rates like 0.999 tend to hurt the consistency between the two networks. The results also verify that the teacher-student pipeline [45] has good synergy with our proposed SSL framework. Thanks to this simplicity, more modern SSL techniques can be easily incorporated into the current framework in future works.

Confidence Threshold. As pseudo-labels play an important role in our framework, we further analyze the impact of the threshold parameter T used in pseudo-label generation and show results in Fig. 4 (right). When T is too low, a forced consistency to low-quality pseudo-labels tends to deteriorate the performance. When T is too high, the benefits from mixing might diminish. Generally, T is a dataset-dependent parameter and we find that a value around 0.9 leads to the best possible results on the three tested LiDAR segmentation datasets [1, 11, 47] in our benchmark.

5. Conclusion

In this work, we exploit the unique spatial prior in LiDAR scenes for semi-supervised LiDAR semantic segmentation. We proposed a statistically-principled and effective SSL pipeline, including LaserMix, a novel LiDAR mixing technique that intertwines laser beams from different scans. Through comprehensive empirical analysis, we show the importance of spatial prior and the superiority of our approach on three popular benchmarks. The effectiveness and simplicity of our framework have shed light on the scalable deployment of the LiDAR semantic mapping system.

Acknowledgments. This study is supported by the Ministry of Education, Singapore, under its MOE AcRF Tier 2 (MOE-T2EP20221-0012), the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG2-PhD-2021-08-018), NTU NAP, and under the RIE2020 Industry Alignment Fund – Industry Collaboration Projects (IAF-ICP) Funding Initiative, as well as cash and in-kind contribution from the industry partner(s). This research is part of the programme DesCartes and is supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme.

References

[1] 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, 5, 6, 8

[2] David Berthelot, Nicholas Carlini, Ekin D Cubuk, Alex Kurakin, Kihyuk Sohn, Han Zhang, and Colin Raffel. Remixmatch: Semi-supervised learning with distribution alignment and augmentation anchoring. *arXiv preprint arXiv:1911.09785*, 2019. 2

[3] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin A Raffel. Mixmatch: A holistic approach to semi-supervised learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019. 2

[4] 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. 2

[5] 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. 6

[6] 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. 2, 5, 6, 7

[7] Mingmei Cheng, Le Hui, Jin Xie, and Jian Yang. Sspc-net: Semi-supervised semantic 3d point cloud segmentation network. In *AAAI Conference on Artificial Intelligence (AAAI)*, pages 1140–1147, 2021. 2

[8] 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. 2, 6

[9] Shuang Deng, Qiulei Dong, Bo Liu, and Zhanyi Hu. Superpoint-guided semi-supervised semantic segmentation of 3d point clouds. *arXiv preprint arXiv:2107.03601*, 2021. 2

[10] Terrance DeVries and Graham W. Taylor. Improved regularization of convolutional neural networks with cutout. *arXiv preprint arXiv:1708.04552*, 2017. 7

[11] 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. 1, 2, 3, 5, 6, 7, 8

[12] 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. 2, 6, 7

[13] Biao Gao, Yancheng Pan, Chengkun Li, Sibo Geng, and Huijing Zhao. Are we hungry for 3d lidar data for semantic segmentation? a survey of datasets and methods. *IEEE Transactions on Intelligent Transportation Systems*, 2021. 2

[14] Awet Haileslassie Gebrehiwot, Patrik Vacek, David Hurych, Karel Zimmermann, Patrick Pérez, and Tomáš Svoboda. Teachers in concordance for pseudo-labeling of 3d sequential data. *IEEE Robotics and Automation Letters*, pages 536–543, 2022. 2

[15] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012. 1

[16] Yves Grandvalet and Yoshua Bengio. Semi-supervised learning by entropy minimization. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 17, 2004. 3

[17] Ji Hou, Benjamin Graham, Matthias Nießner, and Saining Xie. Exploring data-efficient 3d scene understanding with contrastive scene contexts. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15587–15597, 2021. 2

[18] Qingyong Hu, Bo Yang, Guangchi Fang, Yulan Guo, Ales Leonardis, Niki Trigoni, and Andrew Markham. Sqn: Weakly-supervised semantic segmentation of large-scale 3d point clouds with 1000x fewer labels. *arXiv preprint arXiv:2104.04891*, 2021. 1, 2

[19] 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. 2, 4

[20] 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. 2, 6

[21] 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. 2, 5

[22] Amit Pal Singh Kohli, Vincent Sitzmann, and Gordon Wetzstein. Semantic implicit neural scene representations with semi-supervised training. In *IEEE International Conference on 3D Vision (3DV)*, pages 423–433, 2020. 2

[23] Lingdong Kong, Niamul Quader, and Venice Erin Liong. Conda: Unsupervised domain adaptation for lidar segmentation via regularized domain concatenation. *arXiv preprint arXiv:2111.15242*, 2021. 1

[24] Samuli Laine and Timo Aila. Temporal ensembling for semi-supervised learning. In *International Conference on Learning Representations (ICLR)*, 2017. 2

[25] Dong-Hyun Lee. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *International Conference on Machine Learning Workshops (ICMLW)*, volume 3, 2013. 2, 5

[26] Venice Erin Liong, Thi Ngoc Tho Nguyen, Sergi Widjaja, Dhananjai Sharma, and Zhuang Jie Chong. Amvnet: Assertion-based multi-view fusion network for lidar semantic segmentation. *arXiv preprint arXiv:2012.04934*, 2020. 2

[27] Minghua Liu, Yin Zhou, Charles R Qi, Boqing Gong, Hao Su, and Dragomir Anguelov. Less: Label-efficient semantic segmentation for lidar point clouds. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 2

[28] Yan Liu, Qingyong Hu, Yinjie Lei, Kai Xu, Jonathan Li, and Yulan Guo. Box2seg: Learning semantics of 3d point clouds with box-level supervision. *arXiv preprint arXiv:2201.02963*, 2022. 2

[29] Yuyuan Liu, Yu Tian, Yuanhong Chen, Fengbei Liu, Vasileios Belagiannis, and Gustavo Carneiro. Perturbed and strict mean teachers for semi-supervised semantic segmentation. *arXiv preprint arXiv:2111.12903*, 2021. 2

[30] Wenfeng Luo and Meng Yang. Semi-supervised semantic segmentation via strong-weak dual-branch network. In *European Conference on Computer Vision (ECCV)*, pages 784–800, 2020. 2

[31] Jilin Mei, Biao Gao, Donghao Xu, Wen Yao, Xijun Zhao, and Huijing Zhao. Semantic segmentation of 3d lidar data in dynamic scene using semi-supervised learning. *IEEE Transactions on Intelligent Transportation Systems*, 21(6):2496–2509, 2019. 2

[32] 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. 2, 4, 6

[33] Alexey Nekrasov, Jonas Schult, Or Litany, Bastian Leibe, and Francis Engelmann. Mix3d: Out-of-context data augmentation for 3d scenes. In *IEEE International Conference on 3D Vision (3DV)*, pages 116–125, 2021. 7

[34] Lucas Nunes, Rodrigo Marcuzzi, Xieyuanli Chen, Jens Behley, and Cyrill Stachniss. Segcontrast: 3d point cloud feature representation learning through self-supervised segment discrimination. *IEEE Robotics and Automation Letters*, pages 2116–2123, 2022. 1, 2

[35] Avital Oliver, Augustus Odena, Colin A Raffel, Ekin D Cubuk, and Ian Goodfellow. Realistic evaluation of deep semi-supervised learning algorithms. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 31, 2018. 5

[36] Viktor Olsson, Wilhelm Tranheden, Juliano Pinto, and Lennart Svensson. Classmix: Segmentation-based data augmentation for semi-supervised learning. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1369–1378, 2021. 2

[37] Yassine Ouali, Céline Hudelot, and Myriam Tami. Semi-supervised semantic segmentation with cross-consistency training. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12674–12684, 2020. 2, 5

[38] Jinhyung Park, Chenfeng Xu, Yiyang Zhou, Masayoshi Tomizuka, and Wei Zhan. Detmatch: Two teachers are better than one for joint 2d and 3d semi-supervised object detection. *arXiv preprint arXiv:2203.09510*, 2022. 2

[39] Scott Drew Pendleton, Hans Andersen, Xinxin Du, Xiaotong Shen, Malika Meghjani, You Hong Eng, Daniela Rus, and Marcelo H Ang. Perception, planning, control, and coordination for autonomous vehicles. *Machines*, 5, 2017. 1

[40] Charles R Qi, Yin Zhou, Mahyar Najibi, Pei Sun, Khoa Vo, Boyang Deng, and Dragomir Anguelov. Offboard 3d object detection from point cloud sequences. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6134–6144, 2021. 2

[41] Ricardo Roriz, Jorge Cabral, and Tiago Gomes. Automotive lidar technology: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 2021. 1

[42] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin D Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, 2020. 2

[43] Chun-Yu Sun, Yu-Qi Yang, Hao-Xiang Guo, Peng-Shuai Wang, Xin Tong, Yang Liu, and Heung-Yeung Shum. Semi-supervised 3d shape segmentation with multilevel consistency and part substitution. *arXiv preprint arXiv:2204.08824*, 2022. 2

[44] 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. 2

[45] 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. 2, 4, 5, 6, 7, 8

[46] 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. 2

[47] Ozan Unal, Dengxin Dai, and Luc Van Gool. Scribble-supervised lidar semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 1, 2, 3, 5, 6, 8

[48] Jianren Wang, Haiming Gang, Siddarth Ancha, Yi-Ting Chen, and David Held. Semi-supervised 3d object detection via temporal graph neural networks. In *IEEE International Conference on 3D Vision (3DV)*, pages 413–422, 2021. 2

[49] Chenfeng Xu, Bichen Wu, Zining Wang, Wei Zhan, Peter Vajda, Kurt Keutzer, and Masayoshi Tomizuka. Squeezesegv3: Spatially-adaptive convolution for efficient point-cloud segmentation. In *European Conference on Computer Vision (ECCV)*, pages 1–19, 2020. 2

[50] Jianyun Xu, Ruixiang Zhang, Jian Dou, Yushi Zhu, Jie Sun, and Shiliang Pu. Rpvnet: A deep and efficient range-point-voxel fusion network for lidar point cloud segmentation. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 16024–16033, 2021. 2

[51] Lihe Yang, Wei Zhuo, Lei Qi, Yinghuan Shi, and Yang Gao. St++: Make self-training work better for semi-supervised semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 2

[52] Jianlong Yuan, Yifan Liu, Chunhua Shen, Zhibin Wang, and Hao Li. A simple baseline for semi-supervised semantic segmentation with strong data augmentation. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8229–8238, 2021. 2

[53] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6023–6032, 2019. 4, 7

[54] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. In *International Conference on Learning Representations (ICLR)*, 2018. 4, 7

[55] Yachao Zhang, Yanyun Qu, Yuan Xie, Zonghao Li, Shanshan Zheng, and Cuihua Li. Perturbed self-distillation: Weakly supervised large-scale point cloud semantic segmentation. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 15520–15528, 2021. 2

[56] 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. 2, 4, 6

[57] 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. 2, 5

[58] 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. 2, 4, 5

[59] 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. 2, 6, 7

[60] Yuliang Zou, Zizhao Zhang, Han Zhang, Chun-Liang Li, Xiao Bian, Jia-Bin Huang, and Tomas Pfister. Pseudoseg: Designing pseudo labels for semantic segmentation. In *International Conference on Learning Representations (ICLR)*, 2020. 2