2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Towards Efficient Semi-Supervised Semantic Segmentation for Solid-State LiDAR Point Clouds

Mardanjan Abla†, Eksan Firkat†✉, Bangquan Xie, Eliyas Suleyman,
Jiazhan Gao, Bin Zhu, and Askar Hamdulla✉

***Abstract*— LiDAR-based 3D semantic segmentation is a critical task in autonomous driving, but its scalability is limited by the reliance on large-scale labeled datasets. Semi-supervised learning (SSL) offers a potential solution by leveraging unlabeled data. However, most existing SSL segmentation methods are designed for mechanical spinning LiDAR (MSLR) and fail to generalize well to solid-state LiDAR (SSLR) due to different scanning patterns and point cloud distributions. To address this challenge, we propose *SSLiMix*, a novel semi-supervised segmentation method with checkerboard *mixing* for *solid*-state LiDAR. Unlike prior MSLR-oriented methods, SSLiMix employs 2D grid partitioning with checkerboard mixing to adapt to SSLR's dense and uniform point clouds, thereby preserving spatial consistency even when beam-based augmentations fail. Additionally, we introduce a hierarchical confidence-aware pseudo-labeling mechanism (HCAP), which classifies pseudo-labels by confidence and applies targeted processing to enhance pseudo-label reliability. Experiments on the PandaSet dataset show that SSLiMix improves mIoU by 11.3% over the fully-supervised baseline using only 1% labeled data, demonstrating its effectiveness in low-label regimes and providing a strong benchmark for semi-supervised SSLR segmentation.**

I. INTRODUCTION

LiDAR (Light Detection and Ranging) point cloud semantic segmentation is a fundamental task in autonomous driving, aiming to assign a semantic label to every point in a 3D scene [1], [3], [8], [16], [31]. In recent years, deep learning has made significant progress in this field. However, most existing methods still rely heavily on large-scale, high-quality labeled datasets. The fine-grained annotation of LiDAR point clouds is both time-consuming and labor-intensive task, which makes the deployment of the model in practical applications a great challenge [7], [9], [17]. Semi-supervised learning (SSL) offers an effective alternative by leveraging a small set of labeled data alongside abundant unlabeled samples, thereby enhancing model performance while significantly reducing annotation costs.

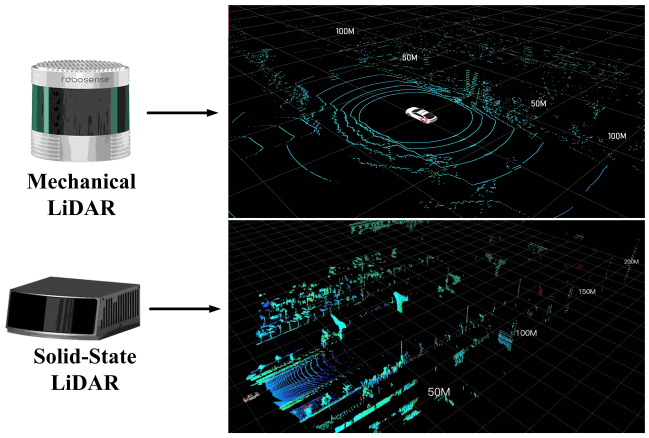

Fig. 1: Comparison of Mechanical and Solid-State LiDAR point clouds. The top image shows sparse, rotationally scanned point clouds from Mechanical LiDAR, while the bottom image illustrates denser, more uniform point clouds from Solid-State LiDAR with extended range.

Recent years have seen the development of various SSL-based methods for 3D semantic segmentation [4], [10], [13], [15], [20]. These methods mainly focus on utilizing the prior information of the point cloud, such as spatial structure, reflectance features, or even 2D semantic cues, to provide supervised signals for unlabeled data, thereby efficiently transferring semantic knowledge from labeled to unlabeled data. While these methods have achieved significant results in Mechanical Spinning LiDAR (MSLR) scenarios, their effectiveness largely depends on the inherent characteristics of MSLR, including the regular rotational scanning patterns and radially sparse point cloud distribution over distance [30], [34]. In contrast, Solid-State LiDAR (SSLR) presents fundamentally different data characteristics. As shown in Fig. 1, SSLR uses Micro-Electro-Mechanical Systems (MEMS) or Flash scanning technologies to produce denser and more uniformly distributed point clouds, lacking the structured laser beams of MSLR [11], [24]. These differences limit the applicability of beam-based augmentation methods (e.g., LaserMix) to SSLR data, often resulting in poor or unstable performance. In addition, the quality of

This work was supported by the Tianshan Talents Cultivation Program - Leadings Talents for Scientifc and Technological Innovation under Grant 2024TSYCLJ0002, in part by the Guangxi Science and Technology Major Project under Grant GUIKEAA23062031 and Grant GUIKEAA23062073.

[1]Mardanjan Abla, Jiazhan Gao and Askar Hamdulla are with the School of Computer Science and Technology, Xinjiang University, Urumqi 830046, China.

[2]Eksan Firkat is with Dongguan Key Laboratory of Intelligent Equipment and Smart Industry, School of Advanced Engineering, Greater Bay University, Dongguan 523000, China, and Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen 518057, China.

[3]Bangquan Xie is with Dongguan Key Laboratory of Intelligent Equipment and Smart Industry, School of Advanced Engineering, Great Bay University, Dongguan 523000, China.

[4]Eliyas Suleyman is with the School of Computing Science, University of Glasgow, UK.

[5]Bin Zhu is with the Department of Automation, Tsinghua University, Beijing 100084, China.

† These authors contributed equally.

✉ Corresponding author.

pseudo-labels plays a crucial role in SSL [10], [12], [18], [19], [43]. Due to the dense and uniform distribution of SSLR point clouds, generating high-confidence pseudo-labels becomes more challenging. Directly applying MSLR-designed pseudo-labeling strategies, which use a fixed threshold, may introduce significant noise, degrading learning efficiency.

To address these challenges, We propose SSLiMix, a novel semi-supervised segmentation method for SSLR point clouds. Unlike traditional laser beam-based methods [4], [13], [43], SSLiMix employs a 2D grid-based partitioning combined with checkerboard-style mixing, which is better suited for the dense and uniform point clouds of SSLR while preserving spatial consistency. We further introduce a Hierarchical Confidence-Aware Pseudo-Labeling (HCAP) mechanism, which classifies pseudo-labels into high, ambiguous, and low-confidence regions and dynamically adjusts thresholds based on category characteristics. In contrast to fixed-threshold pseudo-labeling methods [4], [10], [15], [20], HCAP effectively leverages low-confidence labels with latent semantic value, improving supervision and overall performance. We validate our approach on the large-scale PandaSet dataset [5]. Extensive experiments demonstrate that SSLiMix consistently outperforms state-of-the-art semi-supervised LiDAR segmentation methods [4], [13], [20] under varying annotation-scarcity settings, confirming its effectiveness in SSLR scenarios. Our main contributions are as follows:

- We propose SSLiMix, a novel augmentation strategy based on 2D grid partitioning and checkerboard mixing, explicitly designed for SSLR's dense and uniform point clouds.
- We introduce HCAP, a hierarchical pseudo-labeling mechanism that dynamically adapts thresholds and leverages low-confidence labels for richer supervision.
- Through extensive experiments, we validate the effectiveness of our method in low-data scenarios and establish a PandaSet benchmark to facilitate future research.

II. Related Work

LiDAR Semantic Segmentation. Various methods have been proposed for LiDAR semantic segmentation, categorized into point-based, projection-based, and voxel-based approaches [1], [8], [9], [16], [21]. Point-based methods [2], [3] exploit spatial relationships among unordered points to learn point-wise features from raw point clouds, but they incur high computational cost, limiting scalability in large-scale scenes. Projection-based methods [23], [25], [27], [29], [32] map point clouds onto 2D planes using perspective or spherical projections, enabling efficient segmentation with 2D CNNs, though this process may sacrifice geometric fidelity. Voxel-based methods, such as Cylinder3D [31], partition point clouds into 3D grids, preserving spatial structure while balancing accuracy and efficiency. Despite these advances, all these approaches rely on fully supervised training, requiring large labeled datasets and performing poorly when annotations are scarce, which motivates the exploration of semi-supervised solutions.

Semi-supervised LiDAR Segmentation. Semi-supervised 3D semantic segmentation trains models using a small amount of labeled data together with a large set of unlabeled point clouds, effectively reducing annotation cost [4], [6], [9], [10], [17]. Early work in this direction was primarily conducted on indoor RGB-D point clouds. HPAL [6], [33] achieved strong results on indoor datasets, but the limited scale and diversity of indoor scenes, together with their differences from LiDAR spatial structures and sampling priors, restrict the applicability of such methods to outdoor LiDAR data. Recent studies have therefore focused on semi-supervised LiDAR segmentation. LaserMix [4] exploits strong spatial priors by mixing laser beams from different scans to enhance prediction consistency under data augmentation. ItTakesTwo [13] combines range and voxel representations to improve both consistency and contrastive learning, while PLE [20] leverages spatio-temporal consistency to generate reliable pseudo-labels and mitigate label imbalance through a dual-branch model. Nevertheless, these methods are primarily designed for mechanical LiDAR, and due to fundamental differences in scanning patterns and point cloud distributions, they often struggle to adapt effectively to SSLR scenarios.

Solid-State LiDAR Segmentation. Compared to traditional mechanical LiDAR, solid-state LiDAR offers faster scanning speeds, higher resolution, and improved stability, making it an appealing alternative for autonomous driving applications [11], [24], [34]. However, its unique scanning patterns and point cloud distributions pose new challenges for semantic segmentation. Recent fully supervised methods have attempted to address these issues through tailored architectures. SFPNet [35] introduces a sparse focus mechanism to adapt to varying point cloud densities and distributions, while SSL-VoxPart [36] proposes an adaptive voxel segmentation strategy designed specifically for SSL data. Despite their effectiveness, these approaches rely heavily on large-scale annotations, limiting scalability. To mitigate this, CL3D [37] explores unsupervised domain adaptation to transfer knowledge from mechanical to solid-state LiDAR, improving cross-sensor generalization. Nevertheless, semi-supervised segmentation for SSL remains largely unexplored. This work addresses this gap by proposing a semi-supervised framework that leverages SSLR's unique features, while establishing comprehensive SSL benchmarks on popular autonomous driving datasets.

III. Method

Fig. 2 illustrates the overall pipeline of the proposed method. We first define the semi-supervised point cloud segmentation problem, followed by an introduction to our framework's architecture and detailed descriptions of its components. Finally, we present the loss functions used for training.

Problem Definition. In the semi-supervised setting, we are provided with two sets of data: a small set of manually annotated point clouds $\mathcal{D}_l = \{(\mathbf{X}_i, \mathbf{Y}_i)\}_{i=1}^{N_l}$, and a large set of unlabeled point clouds $\mathcal{D}_u = \{\mathbf{X}_i\}_{i=1}^{N_u}$. Each point

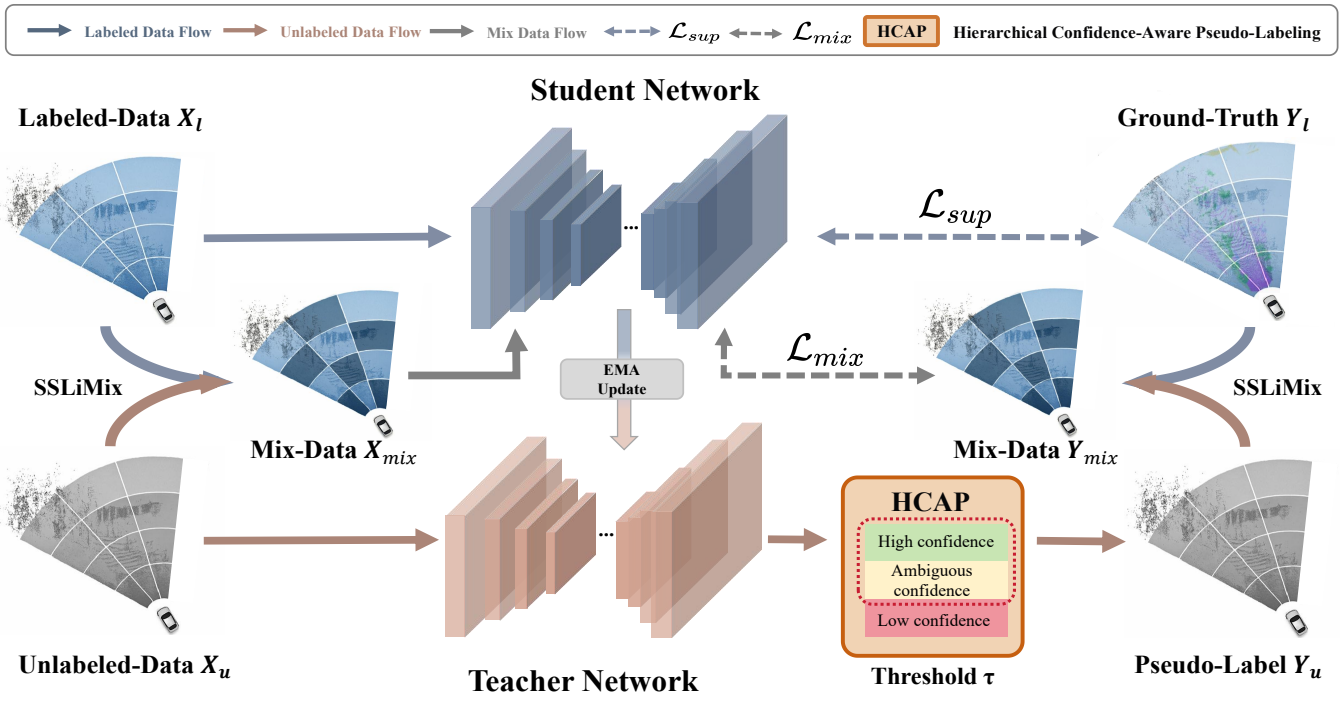

Fig. 2: The labeled data X_l is fed into the Student network to compute the supervised loss L_{sup} with ground-truth Y_l. The unlabeled data X_u and generated pseudo-labels Y_u are mixed with (X_l, Y_l) using SSLiMix (Sec III-B) to form the mixed sample $(X_{\text{mix}}, Y_{\text{mix}})$, which is then fed into the Student network to compute the mixing loss L_{mix}. The Teacher network is updated using EMA, and the mean teacher loss L_{mt} is computed over the Student and Teacher predictions.

cloud sample $\mathbf{X}_i \in \mathbb{R}^{m\times(3+r)}$ consists of m points, where each point is represented by its 3D Cartesian coordinates and r-dimensional additional features (e.g., intensity or color). The corresponding label $\mathbf{Y}_i \in \{1,\ldots,c\}^m$ denotes the per-point semantic annotations over c predefined categories. Typically, $N_u \gg N_l$. The goal of semi-supervised point cloud segmentation is to leverage both limited labeled data and large-scale unlabeled data to improve the performance of an segmentation model.

A. Overall Architecture

We adopt the widely used teacher-student framework [38], consisting of two segmentation networks sharing the same architecture. In our implementation, both the teacher and student networks utilize the same segmentation head f_θ, which learns a point-wise segmentation function $f_\theta : \mathbb{R}^{m\times(3+r)} \to \mathbb{R}^{m\times c}$, mapping an input point cloud $\mathbf{X}_i$ to per-point probability distributions over c semantic classes.

During training, each batch is divided into labeled and unlabeled point clouds. The student network G_{θ_s} is supervised with ground-truth labels for the labeled data and pseudo labels generated by the teacher network G_{θ_t} for the unlabeled data. The student network is updated via standard backpropagation. The teacher network is updated using an Exponential Moving Average (EMA) of the student's parameters:

$$\theta_t^{(k)} \leftarrow \alpha\theta_t^{(k-1)} + (1-\alpha)\theta_s^{(k)}, \tag{1}$$

where α is the EMA decay rate and k denotes the training step.

To enhance sample diversity and better exploit unlabeled data, we propose SSLiMix, a grid-based data augmentation strategy (Sec. III-B). SSLiMix partitions both labeled and unlabeled point clouds into 2D grids and applies a checkerboard-style mixing strategy to combine the data before feeding it into the student network. Furthermore, to address the unreliability of pseudo labels, we introduce a Hierarchical Confidence-Aware Pseudo-Labeling (HCAP) mechanism (Sec. III-C). By categorizing pseudo labels into high-confidence, ambiguous, and low-confidence regions and applying targeted refinement strategies, HCAP improves pseudo-label quality and provides more reliable supervisory signals to the student network.

B. SSLiMix Data Augmentation

Traditional data mixing strategies, such as LaserMix [4], rely on a beam-based partitioning scheme that divides the point cloud into concentric annular regions based on elevation angles. An alternating even-odd mixing strategy is then applied along the radial direction within each ring. While effective for mechanical LiDAR with full 360° scanning coverage, this method relies heavily on the sparse radial distribution and consistent beam arrangement, assumptions that do not hold for solid-state LiDAR due to differing scanning principles.

In contrast, solid-state LiDAR, particularly those employing MEMS-based scanning, produce point clouds that are denser and more uniformly distributed within a constrained field of view [24]. Additionally, solid-state LiDAR point

clouds exhibit strong semantic locality, where neighboring points typically belong to the same object or surface. These characteristics suggest that a different partitioning approach is required. Therefore, we propose a 2D angular grid-based partitioning strategy, more suited to the spatial structure of solid-state LiDAR point clouds. Given a point cloud $\mathbf{X} = \{(x_i, y_i, z_i)\}_{i=1}^{N}$, the azimuth θ) and pitch ϕ for each point are computed as follows:

$$\theta_i = \arctan\left(\frac{y_i}{x_i}\right), \quad \phi_i = \arctan\left(\frac{z_i}{\sqrt{x_i^2 + y_i^2}}\right). \quad (2)$$

The field of view is then discretized into a grid of $M \times N$ angular cells, with the angular bounds of each cell defined by:

$$\theta_m = \theta_{\min} + m \times \frac{\theta_{\max} - \theta_{\min}}{M}, \quad m \in \{0, ..., M-1\} \quad (3)$$

$$\phi_n = \phi_{\min} + n \times \frac{\phi_{\max} - \phi_{\min}}{N}, \quad n \in \{0, ..., N-1\} \quad (4)$$

Each grid cell (m, n) represents a localized region within the point cloud, reflecting the fine spatial structure inherent in solid-state LiDAR data.

Building upon this grid partitioning, we introduce SSLiMix, a checkerboard-style point cloud mixing strategy designed specifically for solid-state LiDAR. Given a labeled point cloud P_{label} and one unlabeled P_{unlabel}, we construct a mixed point cloud P_{mix} by selecting data sources based on the parity of the grid indices:

$$P_{\text{mix}}(m,n) = \begin{cases} P_{\text{label}}(m,n), & \text{if } (m+n) \text{ is even} \\ P_{\text{unlabel}}(m,n), & \text{if } (m+n) \text{ is odd} \end{cases} \quad (5)$$

The SSLiMix process involves two main stages: (1) partitioning the point cloud into 2D grids, and (2) performing a checkerboard-style mixing of labeled and unlabeled data sources based on grid cell index. This mixing strategy maintains spatial consistency within localized regions, enhancing both training diversity and the semantic continuity of the scene.

We hypothesize that SSLiMix improves spatial consistency and robustness compared to traditional beam-based mixing methods, particularly for dense and semantically localized point clouds generated by solid-state LiDAR. This hypothesis will be validated in Section IV.

C. Hierarchical Confidence-Aware Pseudo-Labeling

The quality of pseudo-labels is critical for semi-supervised learning, especially for solid-state LiDAR point clouds whose predictions often exhibit a narrower confidence range than mechanical LiDAR. As a result, a single global confidence threshold tends either to retain excessive noisy supervision or to discard informative but uncertain labels, limiting training effectiveness. To address this issue, we propose Hierarchical Confidence-Aware Pseudo-Labeling (HCAP), which integrates confidence stratification with class-wise adaptive thresholding, as illustrated in Fig. 3.

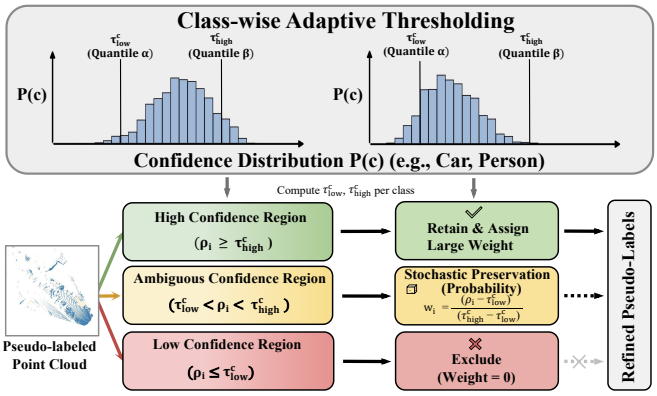

Fig. 3: HCAP mechanism: Pseudo-label classification and processing strategies for three confidence regions.

Given a pseudo-label confidence ρ_i for point i, HCAP partitions pseudo-labels into three regions using two thresholds τ_{low} and τ_{high}: *high-confidence* ($\rho_i \geq \tau_{\text{high}}$), *ambiguous-confidence* ($\tau_{\text{low}} \leq \rho_i < \tau_{\text{high}}$), and *low-confidence* ($\rho_i < \tau_{\text{low}}$). High-confidence pseudo-labels are considered reliable and are fully retained with an increased training weight to amplify their supervisory contribution. Low-confidence pseudo-labels are deemed unreliable, so they are excluded from the loss computation to prevent misleading supervision. For pseudo-labels in the ambiguous-confidence region, we adopt a stochastic retention strategy to partially exploit informative yet uncertain supervision while mitigating systematic bias. Specifically, each ambiguous pseudo-label is preserved via Bernoulli sampling with probability:

$$w_i = \frac{\rho_i - \tau_{\text{low}}}{\tau_{\text{high}} - \tau_{\text{low}}}, \quad \tau_{\text{low}} \leq \rho_i < \tau_{\text{high}} \quad (6)$$

so that higher-confidence ambiguous labels are more likely to be retained, while noisier ones are suppressed.

A fixed global thresholding scheme also fails to accommodate the confidence imbalance across semantic categories, where rare or difficult classes are more prone to underrepresentation during pseudo-label propagation. To improve adaptability, HCAP introduces class-wise adaptive thresholds computed from the accumulated confidence statistics of each class. Let P_c denote the confidence distribution of pseudo-label predictions for class c. We define class-specific thresholds using quantiles:

$$\tau_{\text{high}}^c = \text{Quantile}_\beta(P_c), \quad \tau_{\text{low}}^c = \text{Quantile}_\alpha(P_c). \quad (7)$$

where α and β are quantile levels. This enforces stricter thresholds for frequent classes while relaxing them for rare or difficult ones, improving pseudo-label reliability. Class-wise thresholds are activated only after a minimum number of pseudo-label samples are accumulated for class c; otherwise, HCAP falls back to the global thresholds τ_{low} and τ_{high}.

By combining hierarchical stratification and class-adaptive thresholding, HCAP effectively exploits reliable supervision while reducing noise and class imbalance, leading to more robust semi-supervised learning on solid-state LiDAR point clouds.

D. Loss Functions

Supervised Loss. For labeled data X_l with corresponding ground-truth labels Y_l, the student network is optimized using the standard cross-entropy loss:

$$\mathcal{L}_{\text{sup}} = \frac{1}{m}\sum_{i=1}^{m} \text{CE}\big(Y_l^{(i)}, G_{\theta_s}(X_l)^{(i)}\big). \tag{8}$$

where m denotes the number of points in the input point cloud, and $\text{CE}(\cdot)$ represents the cross-entropy function.

Mixed-Data Loss. To leverage unlabeled data, SSLiMix combines a labeled sample (X_l, Y_l) and an unlabeled sample X_u with its pseudo labels Y_u to construct a mixed input X_{mix} and mixed label Y_{mix}. The student network is supervised on the mixed data using:

$$\mathcal{L}_{\text{mix}} = \frac{1}{m}\sum_{i=1}^{m} \text{CE}\big(Y_{\text{mix}}^{(i)}, G_{\theta_s}(X_{\text{mix}})^{(i)}\big). \tag{9}$$

Consistency Loss. To enforce prediction consistency between the teacher and student networks on the same unlabeled input X_u, we adopt a mean squared error loss:

$$\mathcal{L}_{\text{mt}} = \frac{1}{m}\sum_{i=1}^{m} \Big\| G_{\theta_s}(X_u)^{(i)} - G_{\theta_t}(X_u)^{(i)} \Big\|_2^2. \tag{10}$$

Total Loss. The overall training objective is defined as:

$$\mathcal{L} = \mathcal{L}_{\text{sup}} + \lambda_{\text{mix}}\mathcal{L}_{\text{mix}} + \lambda_{\text{mt}}\mathcal{L}_{\text{mt}}. \tag{11}$$

where λ_{mix} and λ_{mt} are balance weights for the mixed-data and consistency losses, respectively.

During inference, only the teacher network G_{θ_t} is used to generate predictions, as it typically provides more stable outputs.

IV. Experiments

A. Experiment Settings

Dataset. Given the scarcity of publicly available solid-state LiDAR datasets, we conduct our experiments on the **PandaSet** [5] dataset, one of the most comprehensive multi-sensor datasets for semantic segmentation tasks. PandaSet includes LiDAR data collected using two distinct sensors: the rotating Pandar64 and the solid-state PandarGT, both operating at a frequency of 10 Hz. For this study, we focus on the data captured by the PandarGT sensor. Specifically, we extracted labeled scans from PandarGT and merged the original label categorization into 13 semantic categories for evaluation. The dataset is then split into training and validation sets with a 4:1 ratio [4], [40], [41]. Following the settings of previous semi-supervised segmentation studies [4], [13], we evaluate the effectiveness of our method under varying levels of supervision. Specifically, we uniformly sample 1%, 10%, 20%, and 50% of the labeled training scans and treat the remaining scans as unlabeled. This protocol allows us to investigate the robustness of SSL algorithms under different supervision ratios.

Evaluation Metrics. We employ Intersection over Union (**IoU**) and Mean Intersection over Union (**mIoU**) as key metrics for evaluating our model's performance [15]. IoU, defined as the ratio of the intersection to the union of the predicted and ground truth point sets for a specific class, quantitatively assesses the precision of model predictions for individual class labels. The mIoU is the average IoU across all classes, providing an overall performance measure.

Implementation Details. Our experiments are based on the MMDetection3D [28] codebase, configured with reference to LaserMix and optimized for the characteristics of solid-state LiDAR data to better fit the SSLiMix framework.

TABLE I: Segmentation results (mIoU) on the **PandaSet** [5] under varying labeled ratios. Best results are in **bold**.

Repr.	Method	Venue	Backbone	PandaSet			
				1%	10%	20%	50%
Range View	*Sup.-only*	CVPR 2021	FRNet	32.3	36.1	37.5	38.6
	MeanTeacher [38]	NeurIPS 2017	FRNet	28.0	35.3	35.7	36.5
	CBST [22]	ECCV 2018		27.7	33.0	34.5	35.1
	LaserMix [4]	CVPR 2023		31.8	36.5	37.2	37.9
	FrustumMix [32]	CVPR 2024		32.1	37.5	38.4	38.9
	Ours		FRNet	**32.8**	**37.7**	**38.6**	**38.9**
	$\Delta\uparrow$			**+0.5**	**+1.6**	**+1.1**	**+0.5**
Voxel	*Sup.-only*	CVPR 2021	MinkUNet	37.0	46.8	47.2	49.5
	MeanTeacher [38]	NeurIPS 2017	MinkUNet	37.3	42.4	45.8	46.6
	CBST [22]	ECCV 2018		32.7	41.3	43.2	43.8
	LaserMix [4]	CVPR 2023		43.6	47.2	48.3	50.8
	HiLoTs [43]	CVPR 2025		44.0	48.4	49.2	51.6
	Ours		MinkUNet	**46.4**	**54.3**	**54.9**	**56.9**
	$\Delta\uparrow$			**+9.4**	**+7.5**	**+7.7**	**+7.4**
	Sup.-only	CVPR 2021	Cylinder3D	36.9	46.7	48.9	50.8
	LaserMix [4]	CVPR 2023	Cylinder3D	39.1	45.9	46.3	47.7
	PLE [20]	IROS 2024		39.4	43.8	45.1	45.5
	HiLoTs [43]	CVPR 2025		43.1	45.6	48.3	49.7
	Ours		Cylinder3D	**48.2**	**53.6**	**54.4**	**56.3**
	$\Delta\uparrow$			**+11.3**	**+6.9**	**+5.5**	**+5.5**

We conducted experiments using two distinct representations: FRNet [32] for the range view, and MinkUNet [44] and Cylinder3D [31] for the voxel-based representation. These backbones are well-known for their competitive performance in LiDAR semantic segmentation and have demonstrated strong results across multiple benchmarks. For pseudo-labeling, we set confidence thresholds at $\tau_{\text{high}} = 0.9$ and $\tau_{\text{low}} = 0.7$. The model is optimized using the AdamW [26] optimizer, with an initial learning rate of 0.004 and a weight decay of 0.01. A OneCycle scheduler [14] is employed to dynamically adjust the learning rate during training. Data preprocessing included random flipping, rotation, and scaling operations. All experiments are implemented in PyTorch and conducted on NVIDIA RTX 3090 GPUs with a batch size of 4.

B. Comparative Evaluation

To evaluate the generalization and effectiveness of SSLiMix, we conduct extensive experiments on the PandaSet dataset with three representative LiDAR segmentation backbones: FRNet [32] for range-view representation, and MinkUNet [44] and Cylinder3D [31] for voxel-based representation. Comparisons are made against classical 2D SSL methods such as MeanTeacher [38] and CBST [22], as well as recent 3D-specific approaches including LaserMix, FrustumMix, PLE, and HiLoTs. Quantitative results are summarized in Table I.

Results on Range-View Backbone. On the FRNet [32] range-view backbone, SSLiMix demonstrates a steady performance improvement across different annotation ratios, consistently outperforming LaserMix and FrustumMix. These results indicate that the method is applicable under range-view representations; however, its overall mIoU remains lower than that of voxel-based models, reflecting the inherent challenges of range-view representations in capturing the characteristics of solid-state LiDAR data.

Results on Voxel-based Backbones. On voxel-based backbones, SSLiMix demonstrates more pronounced advantages. With 1% annotated data, it achieves 46.4% mIoU with the MinkUNet [44] backbone, surpassing the supervised baseline by 9.4 points, and exhibits a similar improvement on Cylinder3D [31], exceeding the baseline by 11.3 points. Across both backbones, SSLiMix consistently outperforms classical SSL methods and recent 3D-specific approaches. In contrast, 2D SSL methods such as MeanTeacher and CBST perform markedly worse, highlighting the limitations of directly transferring dense image-based consistency strategies to 3D point clouds. Notably, MinkUNet achieves performance on par with Cylinder3D while requiring substantially less memory during training.

Qualitative Analysis. Visualization results in the Fig. 4 show that SSLiMix achieves more coherent segmentation than Sup.-only and LaserMix on road, building, vegetation, and cars. It produces sharper boundaries and stronger seman-

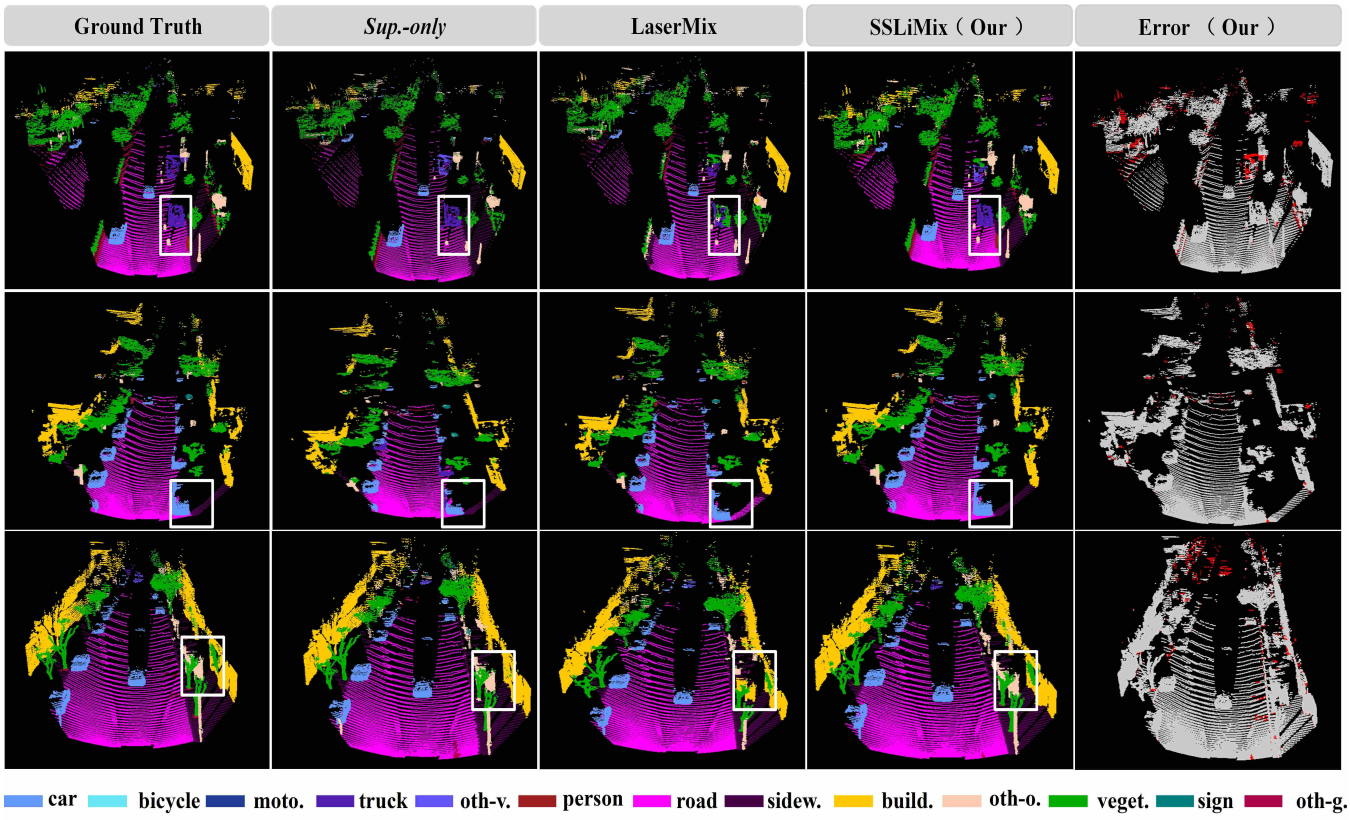

Fig. 4: Qualitative comparison of SSLiMix with the *Sup.-only* and LaserMix models on the PandaSet dataset under the 20% labeling ratio. Best viewed in color.

tic consistency, with clear advantages in sparse and distant regions. The error visualization indicates fewer misclassifications in critical areas, confirming its effectiveness under limited supervision.

C. Ablation Study

To evaluate the contributions of key modules to semi-supervised point cloud segmentation, we conduct system-level and component-level ablation experiments on Cylinder3D. Unless otherwise noted, all experiments share the same data split and training settings, and results are reported under a 1% labeled ratio.

a) System-Level Ablation: Table II reports the system-level ablation results. The first row corresponds to the supervised-only baseline. Introducing the mean teacher (MT) framework alone slightly decreases performance (36.9% $\rightarrow$ 35.1%), indicating that under extremely limited annotations, unreliable pseudo-labels may introduce noise and hinder optimization when no additional regularization or data augmentation is applied. When SSLiMix is incorporated on top of MT, performance improves substantially to 44.5% mIoU, demonstrating that the proposed grid-based mixing strategy effectively stabilizes semi-supervised training and enables better utilization of unlabeled data. Finally, integrating the HCAP module further boosts performance to 48.2% mIoU, confirming that confidence-aware and class-adaptive pseudo-label refinement is complementary to strong spatial augmentation, leading to more reliable supervision.

TABLE II: System-level ablation of key modules.

	Baseline	MT	SSLiMix	HCAP	mIoU (%)
#1	✓				36.9
#2	✓	✓			35.1
#3	✓	✓	✓		44.5
#4	✓	✓	✓	✓	**48.2**

b) Pseudo-Label Filtering Strategies: Table III compares pseudo-label filtering strategies using SSLiMix as the default augmentation. Using a global confidence threshold results in 44.5% mIoU, whereas hierarchical thresholds increase performance to 47.1%. The proposed HCAP method further improves mIoU to 48.2%, indicating that integrating hierarchical thresholds with class-adaptive confidence selection better accounts for confidence variations across classes and enhances pseudo-label reliability.

TABLE III: Ablation of pseudo-label filtering strategies.

Method	mIoU (%)
Global threshold ($\tau = 0.9$)	44.5
Hierarchical threshold	47.1
HCAP (hierarchical + class-adaptive)	**48.2**

c) EMA Ratio Analysis: Table IV reports the effect of different EMA momentum coefficients under 1% and 10% label ratios. Performance steadily improves as the EMA increases from 0.5 to 0.99, reaching the best results at 0.99. However, further increasing the momentum to 0.999 leads to slight performance degradation, likely due to overly smoothed model updates.

TABLE IV: Effect of EMA momentum on segmentation performance.

EMA	1% Labeled	10% Labeled
0.5	47.3	52.8
0.9	47.6	53.1
0.99	**48.2**	**53.6**
0.999	46.8	52.5

d) Class-wise Performance Analysis: Fig. 5 presents the class-wise mIoU for representative static and dynamic categories. SSLiMix consistently outperforms both the supervised baseline and LaserMix across most categories. Notably, low-frequency dynamic classes such as *motorcycle* and *truck* benefit significantly, while static categories like *building* and *vegetation* also improve, indicating more consistent performance across both object-level and scene-level categories.

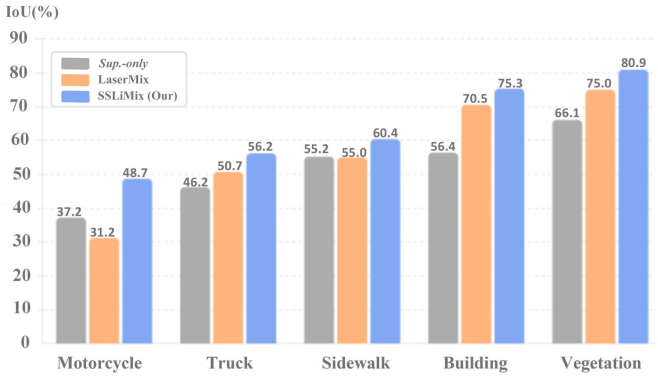

Fig. 5: Class-wise mIoU for representative static and dynamic categories.

V. Conclusion

We present SSLiMix, a semi-supervised framework for semantic segmentation of solid-state LiDAR (SSLR). To address the limited adaptability of MSLR-oriented methods to SSLR's dense and uniform point clouds, SSLiMix introduces a 2D grid-based checkerboard mixing strategy that preserves spatial continuity without relying on beam ordering. We also propose a class-adaptive HCAP module to improve pseudo-label quality and suppress noise. Experiments on PandaSet demonstrate that SSLiMix achieves consistent improvements under limited supervision. To the best of our knowledge, this is the first work to systematically investigate semi-supervised semantic segmentation for SSLR point clouds, and we establish the first benchmark to facilitate future research. Due to the scarcity of public SSLR datasets, our study is currently limited to PandaSet. Future work will extend SSLiMix to broader SSLR scenarios, evaluate its generalization across diverse LiDAR types and backbones, and explore multi-modal sensor fusion for practical deployment in autonomous driving.

REFERENCES

[1] Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu, and M. Bennamoun, "Deep learning for 3D point clouds: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 12, pp. 4338–4364, 2021.

[2] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," in *Proc. NeurIPS*, 2017.

[3] Q. Hu, B. Yang, L. Xie, S. Rosa, Y. Guo, Z. Wang, N. Trigoni, and A. Markham, "RandLA-Net: Efficient semantic segmentation of large-scale point clouds," in *Proc. CVPR*, 2020, pp. 11108–11117.

[4] L. Kong, J. Ren, L. Pan, and Z. Liu, "LaserMix for semi-supervised LiDAR semantic segmentation," in *Proc. CVPR*, 2023, pp. 21705–21715.

[5] P. Xiao, Z. Shao, S. Hao, Z. Zhang, X. Chai, J. Jiao, Z. Li, J. Wu, K. Sun, K. Jiang, Y. Wang, and D. Yang, "PandaSet: Advanced sensor suite dataset for autonomous driving," in *Proc. IEEE ITSC*, 2021, pp. 3095–3101.

[6] Z. Xu, B. Yuan, S. Zhao, Q. Zhang, and X. Gao, "Hierarchical point-based active learning for semi-supervised point cloud semantic segmentation," in *Proc. ICCV*, 2023, pp. 18098–18108.

[7] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall, "SemanticKITTI: A dataset for semantic scene understanding of LiDAR sequences," in *Proc. ICCV*, 2019, pp. 9297–9307.

[8] Y. Li, L. Ma, Z. Zhong, *et al.*, "Deep learning for LiDAR point clouds in autonomous driving: A review," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 8, pp. 3412–3432, 2021.

[9] B. Gao, Y. Pan, C. Li, S. Geng, and H. Zhao, "Are we hungry for 3D LiDAR data for semantic segmentation? A survey of datasets and methods," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 6063–6081, 2022.

[10] L. Li, H. P. H. Shum, and T. P. Breckon, "Less is more: Reducing task and model complexity for 3D point cloud semantic segmentation," in *Proc. CVPR*, 2023, pp. 9361–9371.

[11] J. Chen and Y. Shi, "Research progress in solid-state LiDAR," *Opto-Electronic Engineering*, vol. 46, no. 7, p. 190218, 2019.

[12] L. Ran, Y. Li, G. Liang, and Y. Zhang, "Semi-supervised semantic segmentation based on pseudo-labels: A survey," arXiv:2403.01909, 2024.

[13] Y. Liu, Y. Chen, H. Wang, V. Belagiannis, I. Reid, and G. Carneiro, "It takes two: Leveraging peer representations for semi-supervised LiDAR semantic segmentation," in *Proc. ECCV*, 2024, pp. 81–99.

[14] L. N. Smith and N. Topin, "Super-convergence: Very fast training of neural networks using large learning rates," arXiv:1708.07120, 2017.

[15] W. Xuan, H. Qi, and A. Xiao, "TSG-Seg: Temporal-selective guidance for semi-supervised semantic segmentation of 3D LiDAR point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 216, pp. 217–228, 2024.

[16] R. Zhang, Y. Wu, W. Jin, and X. Meng, "Deep-learning-based point cloud semantic segmentation: A survey," *Electronics*, vol. 12, no. 17, p. 3642, 2023.

[17] O. Unal, D. Dai, and L. Van Gool, "Scribble-supervised LiDAR semantic segmentation," in *Proc. CVPR*, 2022, pp. 2697–2707.

[18] C. Liu, C. Gao, F. Liu, P. Li, D. Meng, and X. Gao, "Hierarchical supervision and shuffle data augmentation for 3D semi-supervised object detection," in *Proc. CVPR*, 2023, pp. 23819–23828.

[19] Y. Liu, Y. Tian, Y. Chen, F. Liu, V. Belagiannis, and G. Carneiro, "Perturbed and strict mean teachers for semi-supervised semantic segmentation," in *Proc. CVPR*, 2022, pp. 4258–4267.

[20] S. Lee, H. Lee, and H. Shim, "Learning from spatio-temporal correlation for semi-supervised LiDAR semantic segmentation," in *Proc. IROS*, 2024, pp. 14095–14102.

[21] X. Wu, Y. Lao, L. Jiang, X. Liu, and H. Zhao, "Point transformer v2: Grouped vector attention and partition-based pooling," in *Proc. NeurIPS*, 2022, pp. 33330–33342.

[22] Y. Zou, Z. Yu, B. V. K. Vijaya Kumar, and J. Wang, "Unsupervised domain adaptation for semantic segmentation via class-balanced self-training," in *Proc. ECCV*, 2018, pp. 289–305.

[23] T. Cortinhal, G. Tzelepis, and E. E. Aksoy, "SalsaNext: Fast, uncertainty-aware semantic segmentation of LiDAR point clouds," in *Proc. ISVC*, 2020, pp. 207–222.

[24] D. Wang, C. Watkins, and H. Xie, "MEMS mirrors for LiDAR: A review," *Micromachines*, vol. 11, no. 5, p. 456, 2020.

[25] E. E. Aksoy, S. Baci, and S. Cavdar, "SalsaNet: Fast road and vehicle segmentation in LiDAR point clouds for autonomous driving," in *Proc. IEEE IV*, 2020, pp. 926–932.

[26] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proc. ICLR*, 2018.

[27] C. Xu, B. Wu, Z. Wang, W. Zhan, P. Vajda, K. Keutzer, and M. Tomizuka, "SqueezeSegV3: Spatially-adaptive convolution for efficient point cloud segmentation," in *Proc. ECCV*, 2020, pp. 1–19.

[28] MMDetection3D Contributors, "MMDetection3D: OpenMMLab next-generation platform for general 3D object detection," 2020. [Online]. Available: https://github.com/open-mmlab/mmdetection3d

[29] A. Milioto, I. Vizzo, J. Behley, and C. Stachniss, "RangeNet++: Fast and accurate LiDAR semantic segmentation," in *Proc. IROS*, 2019, pp. 4213–4220.

[30] Y. Li and J. Ibanez-Guzman, "LiDAR for autonomous driving: The principles, challenges, and trends for automotive LiDAR and perception systems," *IEEE Signal Process. Mag.*, vol. 37, no. 4, pp. 50–61, 2020.

[31] X. Zhu, H. Zhou, T. Wang, F. Hong, Y. Ma, W. Li, H. Li, and D. Lin, "Cylindrical and asymmetrical 3D convolution networks for LiDAR segmentation," in *Proc. CVPR*, 2021, pp. 9939–9948.

[32] X. Xu, L. Kong, H. Shuai, and Q. Liu, "FRNet: Frustum-range networks for scalable LiDAR segmentation," *IEEE Trans. Image Process.*, 2025.

[33] L. Jiang, S. Shi, Z. Tian, X. Lai, S. Liu, C.-W. Fu, and J. Jia, "Guided point contrastive learning for semi-supervised point cloud semantic segmentation," in *Proc. ICCV*, 2021, pp. 6423–6432.

[34] H. Holzhüter, J. Bödewadt, S. Bayesteh, A. Aschinger, and H. Blume, "Technical concepts of automotive LiDAR sensors: A review," *Optical Engineering*, vol. 62, no. 3, p. 031213, 2023.

[35] Y. Wang, W. Zhao, C. Cao, T. Deng, J. Wang, and W. Chen, "SFPNet: Sparse focal point network for semantic segmentation on general LiDAR point clouds," arXiv:2407.11569, 2024.

[36] N. Leuze, H. Schaub, M. Hoh, and A. Schoettl, "SSL-VoxPart: A novel solid-state LiDAR-tailored voxel partition approach for 3D perception," in *Proc. IEEE SENSORS*, 2023, pp. 1–4.

[37] X. Peng, X. Zhu, and Y. Ma, "CL3D: Unsupervised domain adaptation for cross-LiDAR 3D detection," in *Proc. AAAI*, vol. 37, no. 2, 2023, pp. 2047–2055.

[38] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Proc. NeurIPS*, 2017, vol. 30.

[39] K. Sohn, D. Berthelot, C.-L. Li, Z. Zhang, N. Carlini, E. D. Cubuk, A. Kurakin, H. Zhang, and C. Raffe, "FixMatch: Simplifying semi-supervised learning with consistency and confidence," in *Proc. NeurIPS*, 2020, vol. 33, pp. 596–608.

[40] Y. Ouali, C. Hudelot, and M. Tami, "Semi-supervised semantic segmentation with cross-consistency training," in *Proc. CVPR*, 2020, pp. 12674–12684.

[41] X. Chen, Y. Yuan, G. Zeng, and J. Wang, "Semi-supervised semantic segmentation with cross pseudo supervision," in *Proc. CVPR*, 2021, pp. 2613–2622.

[42] Y. Wang, H. Wang, Y. Shen, J. Fei, W. Li, G. Jin, L. Wu, R. Zhao, and X. Le, "Semi-supervised semantic segmentation using unreliable pseudo-labels," in *Proc. CVPR*, 2022, pp. 4248–4257.

[43] R.-D. Lin, P. Weng, Y. Wang, H. Ding, J. Han, and F. Wang, "HiLoTS: High-low temporal sensitive representation learning for semi-supervised LiDAR segmentation in autonomous driving," in *Proc. CVPR*, 2025, pp. 1429–1438.

[44] C. Choy, J. Gwak, and S. Savarese, "4D spatio-temporal ConvNets: Minkowski convolutional neural networks," in *Proc. CVPR*, 2019, pp. 3075–3084.