Bridging LiDAR Gaps: A Multi-LiDARs Domain Adaptation Dataset for 3D Semantic Segmentation

Shaoyang Chen[1], **Bochun Yang**[1], **Yan Xia**[2,3*], **Ming Cheng**[1], **Siqi Shen**[1] and **Cheng Wang**[1*]
[1]Fujian Key Laboratory of Sensing and Computing for Smart Cities, Xiamen University
[2]Technical University of Munich
[3]Munich Center for Machine Learning (MCML)
{sychan, yangbc}@stu.xmu.edu.cn, yan.xia@tum.de, {chm99, siqishen, cwang}@xmu.edu.cn

Abstract

We focus on the domain adaptation problem for 3D semantic segmentation, addressing the challenge of data variability in point clouds collected by different LiDARs. Existing benchmarks often mix different types of datasets, which blurs and complicates segmentation evaluations. Here, we introduce a ***M***ulti-***L***iDARs ***D***omain ***A***daptation ***S***egmentation (MLDAS) dataset, which contains point-wise semantic annotated point clouds captured simultaneously by a 128-beam LiDAR, a 64-beam LiDAR, a 32-beam LiDAR. We select 31,875 scans from 2 representative scenarios: ***campus*** and ***urban street***. Furthermore, we evaluate the current 3D segmentation unsupervised domain adaptation methods on the proposed dataset and propose **H**ierarchical **S**egmentation Network with **S**patial **C**onsistency (**HSSC**) as a novel knowledge transfer method to mitigate the domain gap significantly using spatial-temporal consistency constraints. Extensive experiments show that HSSC greatly improves the state-of-the-art cross-domain semantic segmentation methods. Our project is available at: https://sychen320.github.io/projects/MLDAS.

1 Introduction

3D semantic segmentation is a foundational task in robotics [Grigorescu *et al.*, 2020; Xia *et al.*, 2021a] and autonomous driving [Shen *et al.*, 2023; Li *et al.*, 2023]. Compared with 2D images, point clouds collected by LiDAR offer accurate 3D geometric properties and depth insights [Xia *et al.*, 2024]. Over the past years, a series of learning-based 3D segmentation methods [Hu *et al.*, 2020; Thomas *et al.*, 2019; Huang *et al.*, 2024] achieve excellent performance when training with extensively annotated datasets (e.g., SemanticKITTI [Behley *et al.*, 2019], nuScenes [Caesar *et al.*, 2020], and SemanticPOSS [Pan *et al.*, 2020]). However, these SOTA segmentation methods often rely on a crucial assumption: *the training and testing datasets should have similar data distributions.*

In practice, the quality and characteristics of scanned point clouds differ significantly, influenced by various LiDAR sensor configurations and physical environments. Notably, the differences in the disparities in laser line count, field of view, and detection range especially pose substantial challenges for cross-domain research in 3D point cloud domain adaptation.

Current 3D domain adaptation methods [Yi *et al.*, 2021; Langer *et al.*, 2020] leverage existing datasets [Sun *et al.*, 2020; Pan *et al.*, 2020] as benchmarks to validate their effectiveness in cross-dataset domain adaptation. These datasets are typically collected by different authors using diverse sensors in various scenes and annotated according to distinct rules. For example, SemanticKITTI [Behley *et al.*, 2019] provides point cloud data with 64-beam and annotations for 19 semantic categories, while nuScenes [Caesar *et al.*, 2020] is collected using a 32-beam LiDAR and annotated with 16 categories. In addition, nuScenes cover multiple road scenes in different countries to provide rich samples and add diversity for 3D semantic segmentation tasks but not semantic segmentation domain adaptation. Consequently, such benchmarks amalgamate multiple factors influencing domain gaps, posing challenges for meticulous individual analysis. Recently, some studies [Wu *et al.*, 2018; Xiao *et al.*, 2022b] utilize virtual engines or simulation methods to rapidly acquire annotated point cloud data, concurrently presenting adaptive approaches from virtual to real data. While such methods effectively alleviate the burden of acquiring labeled data, the domain gap between synthetic and real data remains highly challenging. From the scenario above, there is still a lack of datasets specifically designed to address domain adaptation issues concerning ***cross-sensor*** and ***cross-scenario***.

To address this issue, we introduce a **M**ulti-**L**iDAR **D**omain **A**daptation Dataset for 3D semantic **S**egmentation (**MLDAS**). The dataset is equipped with 3 typical types of multi-beam LiDAR: 32-beam, 64-beam, and 128-beam. LiDARs are time-synchronized and calibrated to capture the same scenario, mitigate the impact of dynamic objects on LiDAR scans. Utilizing a vehicle-mounted platform, we collected 10,625 frames of data per LiDAR across two distinct scenes, with 6,395 frames on campus and 4,230 frames on the street. We performed detailed point-wise annotations for 14 categories on the 128-beam data and subsequently transferred the labels to the 64-beam and 32-beam data. Thanks to

*Corresponding author

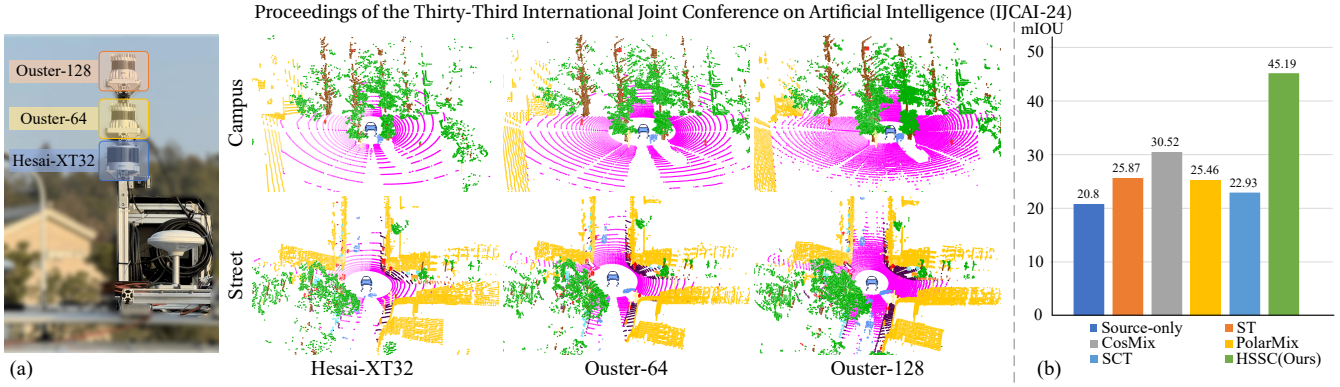

Figure 1: (a) We create MLDAS, a spatiotemporally synchronized Multi-LiDARs domain adaptation dataset for 3D segmentation. MLDAS comprises over 31.8k annotated frames of 14 semantic classes, which are collected by a 128-beam LiDAR (Ouster-OS128), a 64-beam LiDAR (Ouster-OS64), and a 32-beam LiDAR (Hesai-XT32) from campus and urban street scenarios. (b) Segmentation performance on the proposed MLDAS. Our hierarchical segmentation network (HSSC) performs better than previous SOTA methods.

precise sensor calibration and time synchronization, the label transfer process significantly reduced the human effort and time required for data annotation. Based on this foundation, we will elaborate on the dataset's capacity to investigate the impact of cross-sensor and cross-scene factors on semantic segmentation domain adaptation.

Building upon MLDAS, we propose a method based on **H**ierarchical **S**egmentation network with **S**patial **C**onsistency (**HSSC**) to investigate the cross-sensor and cross-scenario challenge in 3D semantic segmentation domain adaptation (DA). Unlike existing methods [Xie *et al.*, 2020; Xiao *et al.*, 2023a] that achieve consistency constraints between different views through data augmentation, our approach utilizes sensor calibration relationships and synchronization to establish corresponding relationships in the same spatial-temporal context across extra LiDAR data. It allows us to focus on the search between different sensors, mitigating the impact of environmental variables. Employing such consistency constraints, we apply the typical mean-teacher architecture to transfer knowledge from the source to the target sensor. In unsupervised domain adaptation (UDA) tasks, experimental results demonstrate the effectiveness of our HSSC in cross-sensor knowledge transfer compared to previous SOTA methods. Furthermore, to assess the generalization capability of our HSSC, we collect extra data from a location distinct from the Campus and Street. This supplementary dataset serves as a transition domain for cross-domain training of the model. In the absence of target domain data, we verify the effectiveness of our HSSC in UDA.

To summary, our main contributions include:

- We introduce a **M**ulti-**L**iDAR **D**omain **A**daptation Dataset for 3D semantic **S**egmentation(**MLDAS**), including 3 types of multi-beam spinning LiDARs scanning 2 scenes with point-wise annotations. MLDAS will establish a robust groundwork for exploring uncharted territories in cross-LiDAR and cross-scenario domain adaptation in LiDAR point cloud segmentation research.
- We investigate the significant factors underlying the domain gap in cross-sensor and cross-scenario, leading to the design of **HSSC**, a framework capable of transferring knowledge from the source sensor to the target sensor, effectively mitigating the domain gap.
- We benchmark various UDA algorithms for LiDAR semantic segmentation, providing comprehensive baselines to support cross-sensor and scenario segmentation research.

2 Related Works

2.1 LiDAR Segmentation Datasets

The recent emergence of outstanding learning-based LiDAR semantic segmentation methods [Zhu *et al.*, 2021; Tang *et al.*, 2022; Unal *et al.*, 2022] has sparked a demand for extensive semantic annotation data, leading to the introduction of numerous datasets. Currently, widely utilized datasets include SemanticKITTI, nuScenes, Waymo [Sun *et al.*, 2020] and semanticPOSS [Pan *et al.*, 2020]. These datasets offer rich varieties of LiDAR types and diverse scene data, significantly advancing the field of semantic segmentation tasks. However, employing these datasets directly for domain adaptation introduces complications related to various influencing factors such as cross-sensor and cross-scenario effects, hindering effective evaluation.

To address the needs of certain semantic segmentation domain adaptation tasks, specific datasets have been proposed. SynLiDAR [Xiao *et al.*, 2022b] uses a simulation platform to simulate a 64-beam LiDAR.While it alleviates the issue of annotating real data, the gap between virtual and real data remains challenging to overcome. Since existing datasets are collected under normal weather conditions, they are not conducive to studying semantic segmentation under adverse-weather conditions. SemanticSTF [Xiao *et al.*, 2023b] conducts an adverse-weather dataset to facilitate research on semantic segmentation adaptation in all-weather conditions. Despite the availability of numerous datasets for 3D semantic segmentation, there is still a lack of datasets in real-world scenarios for semantic segmentation domain adaptation while decoupling the influence of scenes and LiDAR types.

2.2 Domain Adaptation in 3D Semantic Segmentation

Earlier research [Zhao *et al.*, 2021; Wu *et al.*, 2019; Jiang and Saripalli, 2021] applied the 2D DA method to range

Sensor	Type	FOV(°)	Resolution	FPS
XT32	LiDAR	[-16, 15]	32 * 2048	10
OS64	LiDAR	[-22.5, 22.5]	64 * 1024	10
OS128	LiDAR	[-22.5, 22.5]	128 * 2048	10
M39	GNSS/IMU	-	-	200

Table 1: Sensors specifications. FOV means vertical field of view. FPS refers to frames per second.

images projected from 3D point clouds. As the development of point cloud processing networks [Qi *et al.*, 2017; Zhou and Tuzel, 2018] progresses, researchers are increasingly focusing on the characteristics of 3D point clouds. Some researchers employ sparse point cloud completion method [Xia *et al.*, 2021b] to ease sensor density variation. Data augmentation [Yun *et al.*, 2019; Wu *et al.*, 2020] is commonly employed in 2D images to mitigate domain gaps between diverse datasets. Inspired by this, Cosmix [Saltori *et al.*, 2022] and Polarmix [Xiao *et al.*, 2022a] encourage the model to learn cross-domain representations by blending source and target data differently. Similarly, PCT [Xiao *et al.*, 2022b] adapts synthetic point clouds to match the appearances and sparsity of real point clouds. SCT [Xiao *et al.*, 2023a] achieves domain adaptive segmentation by enforcing spatial consistency constraints on the target domain data before and after data augmentation. Distinguishing from it, our approach avoids the necessity for intricate augmentation techniques. Instead, we directly leverage the spatial-temporal consistency constraints among real LiDAR datas to learn cross-sensor&scenario representations.

3 MLDAS Dataset

3.1 Sensors Setup

MLDAS comprises point cloud data from three mechanical spinning LiDARs with various beams, including: (1) Hesai XT32, a 32-beam low resolution LiDAR (denoted as **XT32**). (2) Ouster OS1-64, a 64-beam LiDAR (denoted as **OS64**). (3) Ouster OS1-128, a 128-beam high resolution LiDAR (denoted as **OS128**). These LiDARs cover the common application and research. In addition, we install a GNSS/IMU device (M39) to obtain the ego-motion pose. Detailed specifications for these sensors are described in Table. 1.

Figure 2 shows a rigid support system for securing the LiDARs to ensure that the sensor suites operate reliably. Different LiDAR brands have varying capabilities and accuracy (shown in the Supplementary Material), and each uses its own coordinate system, as illustrated in Figure 2 (b). Theses disparities facilitate our research on cross-sensor domain adaptation. The LiDAR units are mounted vertically to minimize the interference or occlusion between them during operation and to maximize coverage of a 360° scene. Additionally, the positions of the LiDAR centers are not aligned on a single axis to diversify collection positions.

3.2 Sensors Synchronization and Calibration

Precision Time Protocol (PTP) [Eidson and Kang, 2002] is used to synchronize time amongst three LiDAR units, reliant on GNSS clocks provided by M39 for time reference.

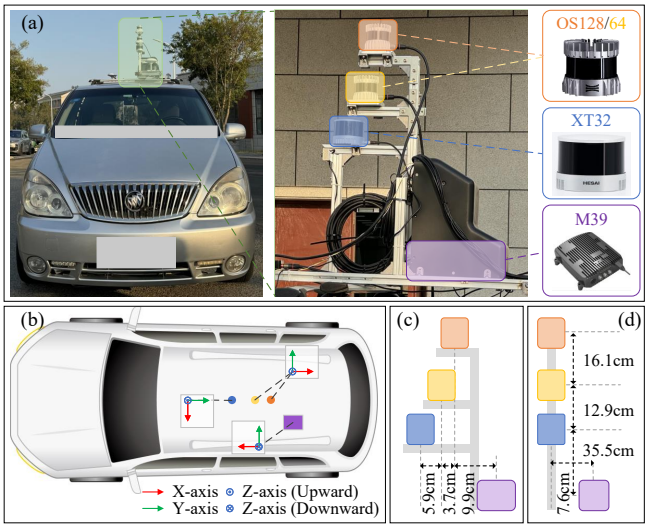

Figure 2: The platform and sensor location diagram. (a) a multi-purpose vehicle, our data collection platform, equipped with multiple sensors. (b-d) Coordinate frames illustrate the direction and distance of each sensor on the vehicle with the convention: X-axis (red), Y-axis (green), and Z-axis (blue). The measurements listed are approximated to the nearest centimeter.

The synchronization mechanism ensures that the cross-sensor data is aligned properly. In addition, we also achieve precise extrinsic calibration among sensors. Specifically, MLDAS sensor calibration includes both LiDAR-to-LiDAR and LiDAR-to-GNSS/IMU calibrations. We strategically position rigid objects with regular shapes within the detection range of the LiDARs for precise calibration of LiDAR-to-LiDAR. We subsequently identify corresponding corner points and apply the Generalized-ICP [Segal *et al.*, 2009] algorithm to derive the final external parameters. The calibration method of LiDAR-to-GNSS/IMU refers to KITTI [Geiger *et al.*, 2012].

3.3 Scene Selection

To decouple the effects of cross-sensor and cross-scenario in domain adaption, we collect data in two representative scenarios: **campus** and urban **street**. These scenarios are vastly different in point number and distribution of some categories. For example, on campus, there will be fewer cars on campus, more bicycles, and more vegetation, as well as a relatively dispersed pattern of buildings. The distance between dynamic categories and the collection vehicle is random on campus due to the lack of strict traffic rules. However, in city streets, there is a relatively large flow of traffic and people, and there are densely populated areas on both sides of the street. Traffic rules ensure that people and vehicles are able to move within the prescribed range. Consequently, there is a significant shift in the domain between the two scenes. Compared with existing multi-scene collected dataset [Xiao *et al.*, 2021; Caesar *et al.*, 2020], MLDAS covers a wider area, which guarantees a comprehensive and representative distribution of data. This methodology is intended to provide ample and stable domain gaps for subsequent domain adaptation research.

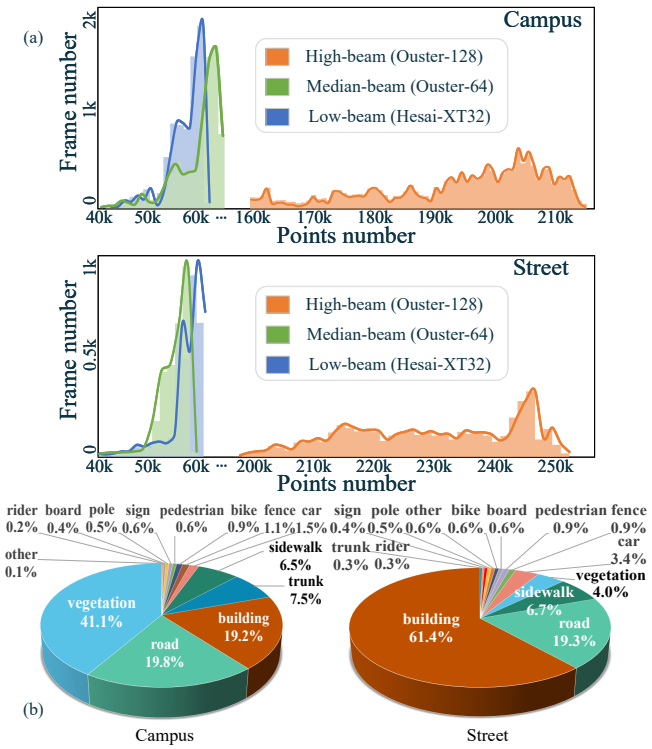

Figure 3: (a) The point number distribution and (b) annotation details of campus and street scenarios, respectively.

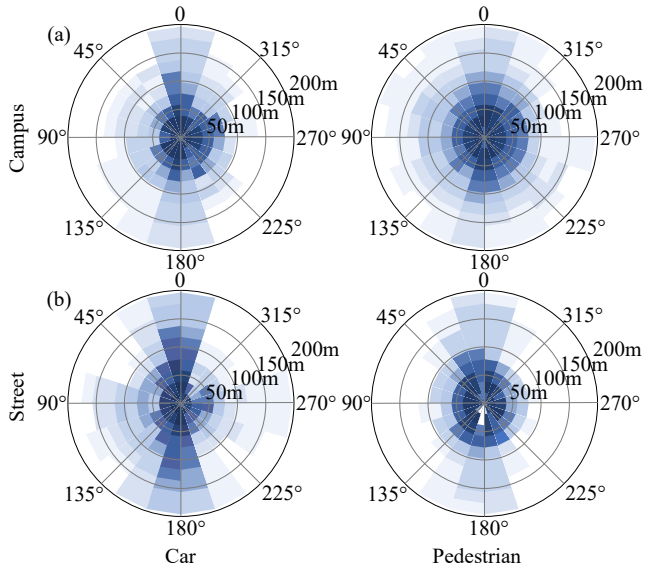

Figure 4: Polar log-scaled density map for point annotations of (a) campus and (b) street scenarios, respectively. The radial and polar axis are the distance and yaw angle from the ego-vehicle, respectively. Darker bins denote more point annotations in that area. Here, we only show the density up to 200m radial distance for all maps.

3.4 Annotation

MLDAS provides point-wise semantic segmentation annotations for three types of LiDAR data. Our methodology commences with the annotation of the densely populated OS128 data to achieve efficient and high-quality annotations. Subsequently, we capitalize on sensor calibration and synchronization relationships to propagate these annotations to the other two LiDAR datasets. We now elaborate on the annotations generation and propagation of MLDAS.

In the annotation generation stage, inspired by SemanticKITTI [Behley *et al.*, 2019], we build the point cloud map from OS128 data for efficient labeling. Specifically, we initialize the map using the poses provided by GNSS/IMU. Then, we employ an offline SLAM algorithm [Zheng *et al.*, 2022] to enhance map quality. Finally, we annotate 14 categories for semantic segmentation evaluation.

In the annotation propagation stage, we employ a temporal-spatial constrained method to transfer the annotations from OS128 to the other two LiDARs data. The first step is to transform the XT32 data into OS128 coordinate system using calibration relationships and position information obtained through SLAM. Our subsequent step is to search for corresponding points in the OS128 map for the XT32 points using the nearest neighbor search algorithm [Bentley, 1975]. This process establishes a spatial correspondence between the two LiDAR data. Due to the presence of dynamic objects in the scene, points within the candidate set may represent different categories at different times but occupy the same location. We eliminate points from the candidate set that do not align with the temporal information of XT32 data by leveraging the temporal synchronization relationships between LiDARs. We assign the label of the nearest neighbor OS128 points for the remaining candidate points to XT32 points. OS64 data annotation follows a similar procedure, which is not repeated here. We ensure the accuracy and consistency of semantic annotations across different LiDAR data sets by calibrating and synchronizing devices, providing a foundation for further research on domain adaptation.

3.5 Dataset Statistics

MLDAS collects 31,875 frames of point cloud data and annotates 14 categories in total. We partition each type of LiDAR data from the two scenes into training, validation, and test sets in a ratio of 8:1:1.

Distribution of Point Numbers. The differences in the quantity and distribution of point clouds from different LiDARs are evident, as illustrated in Figure 3 (a). The data volume of OS128 is significantly higher than that of other LiDARs, posing a challenge for cross-sensor domain adaptation. To better understand the distribution of different LiDAR data, we visualize them in Figure 1 (a). We can observe that the density of three types of LiDAR points decreases with the increasing distance from the sensor. However, due to differences in the field of view and the number of lines, the distribution characteristics of the data are markedly different.

Distribution of Categories. Figure 3 (b) illustrates the distribution of points for 14 semantic categories in MLDAS, showing significant variations in the point count among different categories. Clearly, *Vegetation* and *building* have much higher point counts than others. In addition to the diversity in quantity, the spatial distribution diversity of categories is

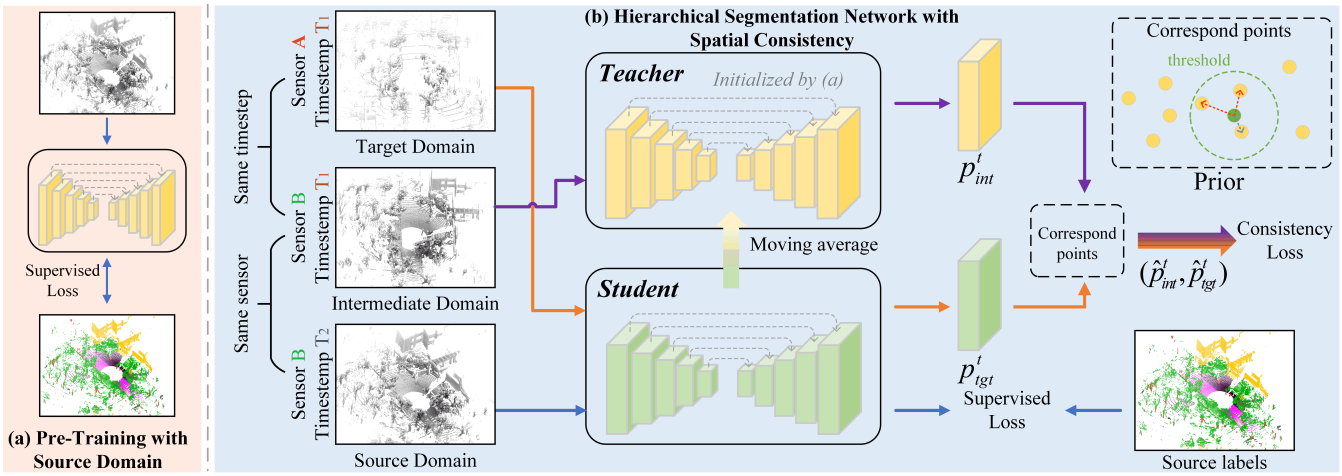

Figure 5: The pipeline of our framework. Hierarchical segmentation contains two steps: (1) Source domain→intermediate domain: Teacher model initialized by(a) predicts intermediate domain, the logits predictions p^t_{int} as spatial consistency supervised signals. (2) Intermediate domain→target domain: Student model generate the logits predictions p^t_{tgt} of target domain, and the adaptation achieved by enforcing $\hat{p}^t_{tgt}$ to be consistent with $\hat{p}^t_{int}$. $\hat{p}^t_{tgt}$ and $\hat{p}^t_{int}$ refer to matching points selected by the nearest neighbor algorithm.

crucial for DA research. Therefore, we analyze the distribution of labels around the sensors for *car*, *pedestrian* categories in both scenes through a polar log-scaled density map, as shown in Figure 4. As mentioned in Section 3.3, traffic rules constrain the distribution of pedestrians and cars in the streets, resulting in higher density along the direction of vehicle movement. In contrast, pedestrians on the campus exhibit a more uniform distribution.

4 HSSC Methods

4.1 Definition

According to UDA setup, we have point cloud from a source domain dataset $D_{src} = \{x^i_{src}, y^i_{src}\}^{N_{src}}_{i=1}$ which contains $N_{src} = |D^{src}|$ point clouds x_{src} with point-wise label y_{src} and an unlabeled target domain $D_{tgt} = \{x^i_{tgt}\}^{N_{tgt}}_{i=1}$. Because MLDAS incorporates multiple LiDARs, for clarity and distinction, we refer to the data, excluding the source and target domains, as the **intermediate domain** $D_{int} = \{x^i_{int}\}^{N_{int}}_{i=1}$. The intermediate domain aims to establish connections between the source and target domains. In this study, D_{int} and D_{src} share the same LiDAR but distinct scenarios. In contrast, D_{int} and D_{tgt} utilize different LiDARs while scanning the data simultaneously. As a result, $N_{int} = N_{tgt}$ and the timestamps of x^i_{int} and x^i_{tgt} are the same.

4.2 Points-wise Matching and Spatial Consistency

Based on MLDAS, we propose a concise and effective method for point-wise correspond matching between two frames of point clouds collected by different sensors. Initially, leveraging the time synchronization within MLDAS, we identify the point clouds x^t_{int} and x^t_{tgt} from D_{int} and D_{tgt} at the time t. Subsequently, employing the calibration relationship T, we transform the point cloud x^t_{tgt} to the coordinate system of x^t_{int}. Finally, We utilize the nearest neighbor algorithm [Bentley, 1975] to obtain the nearest correspond point of x^t_{tgt} in x^t_{int} with a distance below a predefined threshold τ. We record the indexes of the matching pairs in the array as a correspondence mapping M^t. A pair of correspond points reflect the modeling of the same object in spatial by different sensors, possessing consistent contextual information. Building on this spatial consistency, we train the model to adapt to diverse sensor data.

4.3 Hierarchical Segmentation

Utilizing the intermediate domain, we propose the Hierarchical Segmentation Network with Spatial Consistency (HSSC) learning strategy. Our learning paradigm build up on the teacher-student framework with two same backbone of 3D point cloud segmentation, as illustrated in Figure 5. The process of our HSSC can be divided into two hierarchical learning level:

(1) Adapting from source D_{src} to intermediate domain D_{int}. We feed x_{int} into the teacher model. Based on knowledge of D_{src}, the network output the logits predictions p_{int}, which applied as the supervised signals in the next level.

(2) Adapting from intermediate D_{int} to target domain D_{tgt}. We sequentially input x_{tgt} into the student model to generate logits predictions p_{tgt}. The adaptation can be achieved by a spatial-consistency loss that enforce p_{tgt} to be consistent with p_{init}.

4.4 Network Architecture and Training

Both networks share the same 3D semantic segmentation backbone [Choy *et al.*, 2019]. For the source domain, we apply cross entropy loss $\mathcal{L}_{seg}$ to optimize the student network and train model with labeled data, $\mathcal{F}(\cdot)$ refers to the network:

$$\mathcal{L}_{seg} = \frac{1}{N_{src}} \sum_{i=1}^{N_{src}} CE\left(y^i_{src}, \mathcal{F}_{stu}\left(x^i_{src}\right)\right). \quad (1)$$

To narrow the gap between the target and source domains, we construct an intermediate domain based on MLDAS, as

Method	mIOU	car	bicycle	pede.	rider	road	sidewalk	buil.	fence	vege.	trunk	pole	sign	board	oth-obj
Source-only	20.80	35.09	8.54	46.16	17.23	3.08	16.52	87.89	2.01	16.72	11.84	8.71	20.84	**14.76**	1.77
ST	25.87	26.11	7.77	55.66	29.87	0.11	34.22	**94.88**	11.79	17.02	11.78	15.49	43.60	5.48	8.35
CosMix	30.52	40.55	11.71	**58.69**	50.07	32.41	23.76	91.32	2.81	25.07	13.88	13.49	27.18	31.60	4.70
PolarMix	25.46	24.98	13.21	47.38	**52.03**	34.08	2.43	87.36	0.69	20.84	25.04	1.93	42.02	0.16	4.36
SCT	22.93	34.85	10.86	55.32	19.21	14.42	20.11	93.02	8.58	17.20	11.56	7.47	13.19	13.08	2.18
Ours	**45.19**	**88.26**	**21.66**	51.41	48.00	**75.22**	**51.18**	92.46	**27.94**	**48.65**	**27.30**	**18.56**	**49.22**	8.93	**23.90**

Table 2: Adaptation results on Campus(XT32)→Street(OS128). Street(XT32) as intermediate domain.

Method	mIOU	car	bicycle	pede.	rider	road	sidewalk	buil.	fence	vege.	trunk	pole	sign	board	oth-obj
Source-only	37.89	63.27	8.27	43.46	45.4	74.68	46.55	88.57	0.96	53.75	35.14	15.98	42.57	7.60	4.30
ST	39.21	85.43	10.11	43.37	43.18	63.66	40.85	86.84	1.85	37.76	**38.91**	**23.16**	51.12	14.80	**7.86**
CosMix	41.69	61.11	41.06	**55.08**	**65.95**	72.46	46.08	86.91	**5.40**	54.92	30.49	18.52	37.04	2.18	6.42
PolarMix	39.83	64.93	**26.54**	36.33	41.90	72.92	37.14	88.21	5.00	**58.48**	33.82	19.45	54.34	14.44	4.20
SCT	41.50	71.31	12.98	43.94	46.56	78.90	47.29	**89.36**	0.61	53.86	35.05	19.85	**65.58**	11.21	4.52
Ours	**43.52**	**84.45**	18.63	48.78	54.96	**84.83**	**55.44**	88.85	0.99	53.66	33.98	19.41	17.85	**17.85**	5.56

Table 3: Adaptation results on Campus(OS128)→Street(XT32). Street(OS128) as intermediate domain.

described in Section 4.1. We employ a self-supervised training approach for the intermediate and target domain data. Initially, we select a pair of point cloud data $\left\{x_{int}^t, x_{tgt}^t\right\}$ at time t. We feed x_{int}^t data to the teacher network, obtaining prediction logits $p_{int}^t = \mathcal{F}_{tea}(x_{int}^t)$. Similarly, we obtain the logits prediction $p_{tgt}^t = \mathcal{F}_{stu}(x_{tgt}^t)$ through the student network. Subsequently, based on the mapping relationship obtained M^t in Section 2.2, we derive the correspond prediction logits $\hat{p}_{int}^t$ and $\hat{p}_{tgt}^t$. This operation is denoted as $\Psi(\cdot)$:

$$\hat{p}_{int}^t, \hat{p}_{tgt}^t = \Psi\left(p_{int}^t, p_{tgt}^t, M^t\right), \tag{2}$$

where $|\hat{p}_{int}^t| = |\hat{p}_{tgt}^t|$. We then calculate the cosine similarity loss between $\hat{p}_{int}^t$ and $\hat{p}_{tgt}^t$:

$$\mathcal{L}_{cons} = \frac{1}{T}\sum_{t=i}^{T}\frac{1}{|\hat{p}_{int}^t|}\sum_{i=1}^{|\hat{p}_{int}^t|}(1-\cos(\hat{p}_{int}^{t,i}, \hat{p}_{tgt}^{t,i})), \tag{3}$$

where $\hat{p}_{int}^{t,i}$ and $\hat{p}_{tgt}^{t,i}$ is the prediction logits of the i-th correspond points between x_{int}^t and x_{tgt}^t. The total objective loss $\mathcal{L}$ defined as:

$$\mathcal{L} = \mathcal{L}_{seg} + \mathcal{L}_{cons}. \tag{4}$$

In the end, We deploy the Expoential Moving Average (EMA) algorithm to update the weights of the teacher model:

$$\theta_{tea} = \alpha\theta_{tea} + (1-\alpha) + \theta_{stu}, \tag{5}$$

where α is a smoothing coefficient hyperparameter.

5 Experiments

5.1 Experiments Setups and Metrics

Setup. We design three experiment setups for 3D semantic segmentation UDA: (1) **cross-sensor&scenario**, (2) **cross-sensor**, (3) **cross-scenario**. All experiments are implemented in Pytorch [Paszke *et al.*, 2019] and MinkowskiEngine [Choy *et al.*, 2019] on a single NVIDIA RTX 3090 GPU. For a fair comparison, all methods for experiments select the MinkowskiNet [Choy *et al.*, 2019] segmentation architecture as the backbone. The voxel size of the network is set to 0.05m. We train the model on the source domain data using the SGD optimizer and cross-entropy loss function for 30 epochs. Subsequently, we select the best pre-trained model based on evaluating the source validation dataset and re-use the same model for all methods. We select five baseline methods, including source-only, ST [Zou *et al.*, 2018], CosMix [Saltori *et al.*, 2022], PolarMix [Xiao *et al.*, 2022a], and SCT [Xiao *et al.*, 2023a]. Source-only means applying the pre-trained model to the target data without any extra fine-tuning. ST is a domain adaptation method for 2D semantic segmentation, and others are used in 3D point clouds. As SCT has yet to release its source code, we replicated the method based on the information provided in the paper. We utilize target and intermediate domain data for the baselines above to ensure a fair comparison.

Evaluation Metrics. Following previous research [Behley *et al.*, 2019], we employ per-class Intersection-over-Union (IoU) and mean IoU (mIoU) metrics to evaluate the model on 14 categories. In the following sections, we will describe the experimental settings under different conditions.

5.2 Experiments on Cross-Sensor&Scenario

The simultaneous influence of two factors on DA is the most common scenario. Therefore, this experimental setup most directly reflects the effectiveness of UDA methods. We conduct two UDA experiments, considering two LiDAR types (XT32 and OS128) and two scenes (Campus and Street): (1) Campus(XT32)→Street(OS128), (2) Campus(OS128)→Street(XT32). We adapt the model from the source to the target domain. In addition, we apply Street(XT32), Street(O128) as intermediate domain individually.

Comparisons with SOTA. Table 2, 3 present the detailed experimental results. Our method consistently outperforms all state-of-the-art UDA methods on mIoU across different experimental settings, significantly achieving an improvement of +14.67 mIoU over the suboptimal result. These findings demonstrate the ability of our approach to robustly learn knowledge from different sensors and scenes. Our approach maintains commendable performance when migrating from a high-beam LiDAR to a low-beam LiDAR.

Method	mIOU	car	bicycle	pede.	rider	road	sidewalk	buil.	fence	vege.	trunk	pole	sign	board	oth-obj
Source only	57.10	38.59	59.07	60.02	53.48	72.19	68.58	86.78	33.81	81.89	53.85	62.02	54.87	44.6	29.59
ST	59.06	51.57	61.66	64.97	49.04	76.19	72.5	90.19	38.53	82.31	53.16	58.73	41.61	55.80	30.56
CosMix	54.03	38.56	54.35	63.39	52.48	61.94	62.46	82.36	**40.82**	75.1	51.07	53.03	41.07	**60.09**	19.64
Polarmix	58.00	38.83	60.79	63.01	54.58	74.64	70.23	87.93	36.52	82.61	53.99	61.62	49.13	44.68	33.51
SCT	57.36	46.54	59.03	60.08	**58.13**	75.45	71.22	88.87	31.50	82.40	52.82	57.67	47.40	44.32	27.65
Ours	**63.47**	**63.23**	**64.29**	**68.26**	57.40	**81.31**	**77.50**	**92.01**	34.63	**84.64**	**55.55**	**63.00**	**55.31**	54.61	**36.78**

Table 4: Adaptation results on Campus(OS128)→Campus(XT32).

Method	mIOU	car	bicycle	pede.	rider	road	sidewalk	buil.	fence	vege.	trunk	pole	sign	board	oth-obj
Source-only	48.20	92.25	15.14	58.74	63.48	86.11	53.19	91.65	7.01	73.13	31.27	21.36	62.58	8.36	10.47
ST	**53.96**	**92.69**	**42.75**	**65.11**	**67.73**	83.05	58.90	**94.26**	13.32	**75.84**	33.52	23.68	**68.33**	11.06	**25.15**
CosMix	52.17	87.09	31.70	63.25	65.84	86.29	58.62	88.11	**33.76**	68.25	33.55	23.17	60.14	11.68	18.96
PolarMix	49.29	85.22	15.86	56.13	47.37	87.25	53.40	91.74	29.44	78.26	**42.70**	21.20	65.81	5.36	10.27
SCT	48.79	92.15	19.03	62.23	66.71	87.92	**59.79**	92.42	10.88	65.55	26.09	20.33	60.8	7.76	11.28
Ours	52.01	89.82	23.29	58.95	56.39	**88.86**	54.72	92.28	23.34	75.63	35.78	**27.31**	67.59	**12.07**	22.09

Table 5: Adaptation results on Campus(XT32)→Street(XT32). Street(OS128) as intermediate domain.

Training data	mIOU
Source-only	36.66
Source & Place C (OS128) & Place C (XT32)	42.31
Target & intermediate	43.96
Source & target & intermediate	**45.24**

Table 6: Adaptation results on Street(OS128)→Campus(XT32). Performances of our approach trained on different datasets. "intermediate" refers to Campus(OS128). "Place C" shares the same LiDAR type with the source and target domain, but the data is scanned from other places.

5.3 Experiments on Cross-Sensor

In this experiment, we select the Campus to conduct sensor adaptation from OS128 to XT32 data, denoted as Campus(OS128)→Campus(XT32), without introducing an intermediate domain.

Comparisons with SOTA. Analyzing the experimental results from Table 4, we can draw several conclusions: (1) The pre-trained model achieves satisfactory metrics in the target domain. (2) However, the large gap between sensors makes the improvement of existing methods insignificant and even performance degradation. (3) Our proposed method, designed to adapt to sensor variations, successfully learns the target data's characteristics and outperforms other methods.

5.4 Experiments on Cross-Scenario

In this experiment, we select XT32 data and design the experiments: Campus(XT32)→Street(XT32), and apply Street(OS128) as intermediate domain. As we exclude sensor factors, introducing additional intermediate domain data may adversely affect other methods. Hence, we conduct experiments with and without intermediate domain data for other methods and select the best results.

Comparisons with SOTA. From the results of Table 5, when adapting from Campus to Street, ST achieves the best results for its class-balance strategy. Due to the introduction of intermediate domain data (considering the massive data volume of OS128), our method inevitably incurs negative impacts. Our method still achieves results comparable to it.

5.5 Discussion

The above experiments assume that the intermediate and target domains share similar scenes and scan simultaneously. This assumption may not hold since we only have one data type for the target domain (LiDAR type is known). Sometimes, we may have no access to the source domain. Here, we will briefly examine the generalization of our approach to the above cases. First, we construct an intermediate data domain using the MLDAS multi-LiDAR system to collect data, where the LiDAR type corresponds to the source domain and the target domain. Due to the lack of intersection between Campus and Street and semantic annotation, we name the scene **Place C**. Under the setting of Street(OS128)→Campus(XT32), different datasets were used for training and verification on the target domain, and the experimental results are presented in Table 6. The results indicate that our method can be generalized without relying on assumptions derived from previous experiments. More ablation studies are presented in the Supplementary Material.

6 Conclusion

We introduced a new dataset, MLDAS, for 3D segmentation domain adaptation, incorporating three precisely synchronized and calibrated LiDARs and two diverse scenes. Moreover, we proposed a novel hierarchical segmentation network with spatial consistency (HSSC) to address cross-sensor and cross-scenario problems. Extensive experiments demonstrate that the proposed HSSC significantly improves the 3D segmentation performance over the state-of-the-art methods. We hope our work will inspire more investigation into the 3D segmentation domain adaptation problem.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (No.20720230033), and PDL (2022-PDL-12).

Contribution Statement

Shaoyang Chen and Bochun Yang have made equal contributions to this work.

References

[Behley *et al.*, 2019] Jens Behley, Martin Garbade, Andres Milioto, Jan Quenzel, Sven Behnke, Cyrill Stachniss, and Jurgen Gall. Semantickitti: A dataset for semantic scene understanding of lidar sequences. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9297–9307, 2019.

[Bentley, 1975] Jon Louis Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517, 1975.

[Caesar *et al.*, 2020] 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11621–11631, 2020.

[Choy *et al.*, 2019] Christopher Choy, JunYoung Gwak, and Silvio Savarese. 4d spatio-temporal convnets: Minkowski convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3075–3084, 2019.

[Eidson and Kang, 2002] John Eidson and L Kang. Ieee standard for a precision clock synchronization protocol for networked measurement and control systems. *IEEE Std*, 1588, 2002.

[Geiger *et al.*, 2012] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE conference on computer vision and pattern recognition*, pages 3354–3361. IEEE, 2012.

[Grigorescu *et al.*, 2020] Sorin Grigorescu, Bogdan Trasnea, Tiberiu Cocias, and Gigel Macesanu. A survey of deep learning techniques for autonomous driving. *Journal of Field Robotics*, 37(3):362–386, 2020.

[Hu *et al.*, 2020] Qingyong Hu, Bo Yang, Linhai Xie, Stefano Rosa, Yulan Guo, Zhihua Wang, Niki Trigoni, and Andrew Markham. Randla-net: Efficient semantic segmentation of large-scale point clouds. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11108–11117, 2020.

[Huang *et al.*, 2024] Weijie Huang, Pufan Zou, Yan Xia, Chenglu Wen, Yu Zang, Cheng Wang, and Guoqing Zhou. Opoca: One point one class annotation for lidar point cloud semantic segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[Jiang and Saripalli, 2021] Peng Jiang and Srikanth Saripalli. Lidarnet: A boundary-aware domain adaptation model for point cloud semantic segmentation. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2457–2464. IEEE, 2021.

[Langer *et al.*, 2020] Ferdinand Langer, Andres Milioto, Alexandre Haag, Jens Behley, and Cyrill Stachniss. Domain transfer for semantic segmentation of lidar data using deep neural networks. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8263–8270. IEEE, 2020.

[Li *et al.*, 2023] Wen Li, Shangshu Yu, Cheng Wang, Guosheng Hu, Siqi Shen, and Chenglu Wen. Sgloc: Scene geometry encoding for outdoor lidar localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9286–9295, 2023.

[Pan *et al.*, 2020] Yancheng Pan, Biao Gao, Jilin Mei, Sibo Geng, Chengkun Li, and Huijing Zhao. Semanticposs: A point cloud dataset with large quantity of dynamic instances. In *2020 IEEE Intelligent Vehicles Symposium (IV)*, pages 687–693. IEEE, 2020.

[Paszke *et al.*, 2019] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[Qi *et al.*, 2017] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017.

[Saltori *et al.*, 2022] Cristiano Saltori, Fabio Galasso, Giuseppe Fiameni, Nicu Sebe, Elisa Ricci, and Fabio Poiesi. Cosmix: Compositional semantic mix for domain adaptation in 3d lidar segmentation. In *European Conference on Computer Vision*, pages 586–602. Springer, 2022.

[Segal *et al.*, 2009] Aleksandr Segal, Dirk Haehnel, and Sebastian Thrun. Generalized-icp. In *Robotics: science and systems*, volume 2, page 435. Seattle, WA, 2009.

[Shen *et al.*, 2023] Shuo Shen, Yan Xia, Andreas Eich, Yusheng Xu, Bisheng Yang, and Uwe Stilla. Segtrans: Semantic segmentation with transfer learning for mls point clouds. *IEEE Geoscience and Remote Sensing Letters*, 2023.

[Sun *et al.*, 2020] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2446–2454, 2020.

[Tang *et al.*, 2022] Haotian Tang, Zhijian Liu, Xiuyu Li, Yujun Lin, and Song Han. Torchsparse: Efficient point cloud inference engine. *Proceedings of Machine Learning and Systems*, 4:302–315, 2022.

[Thomas *et al.*, 2019] 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 *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6411–6420, 2019.

[Unal *et al.*, 2022] Ozan Unal, Dengxin Dai, and Luc Van Gool. Scribble-supervised lidar semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2697–2707, 2022.

[Wu *et al.*, 2018] Bichen Wu, Alvin Wan, Xiangyu Yue, and Kurt Keutzer. Squeezeseg: Convolutional neural nets with recurrent crf for real-time road-object segmentation from 3d lidar point cloud. In *2018 IEEE international conference on robotics and automation (ICRA)*, pages 1887–1893. IEEE, 2018.

[Wu *et al.*, 2019] Bichen Wu, Xuanyu Zhou, Sicheng Zhao, Xiangyu Yue, and Kurt Keutzer. Squeezesegv2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a lidar point cloud. In *2019 international conference on robotics and automation (ICRA)*, pages 4376–4382. IEEE, 2019.

[Wu *et al.*, 2020] Yuan Wu, Diana Inkpen, and Ahmed El-Roby. Dual mixup regularized learning for adversarial domain adaptation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIX 16*, pages 540–555. Springer, 2020.

[Xia *et al.*, 2021a] Yan Xia, Yusheng Xu, Shuang Li, Rui Wang, Juan Du, Daniel Cremers, and Uwe Stilla. Soe-net: A self-attention and orientation encoding network for point cloud based place recognition. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pages 11348–11357, 2021.

[Xia *et al.*, 2021b] Yan Xia, Yusheng Xu, Cheng Wang, and Uwe Stilla. Vpc-net: Completion of 3d vehicles from mls point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 174:166–181, 2021.

[Xia *et al.*, 2024] Yan Xia, Letian Shi, Zifeng Ding, João F Henriques, and Daniel Cremers. Text2loc: 3d point cloud localization from natural language. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

[Xiao *et al.*, 2021] Pengchuan Xiao, Zhenlei Shao, Steven Hao, Zishuo Zhang, Xiaolin Chai, Judy Jiao, Zesong Li, Jian Wu, Kai Sun, Kun Jiang, et al. Pandaset: Advanced sensor suite dataset for autonomous driving. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pages 3095–3101. IEEE, 2021.

[Xiao *et al.*, 2022a] Aoran Xiao, Jiaxing Huang, Dayan Guan, Kaiwen Cui, Shijian Lu, and Ling Shao. Polarmix: A general data augmentation technique for lidar point clouds. *Advances in Neural Information Processing Systems*, 35:11035–11048, 2022.

[Xiao *et al.*, 2022b] Aoran Xiao, Jiaxing Huang, Dayan Guan, Fangneng Zhan, and Shijian Lu. Transfer learning from synthetic to real lidar point cloud for semantic segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 2795–2803, 2022.

[Xiao *et al.*, 2023a] Aoran Xiao, Dayan Guan, Xiaoqin Zhang, and Shijian Lu. Domain adaptive lidar point cloud segmentation with 3d spatial consistency. *IEEE Transactions on Multimedia*, 2023.

[Xiao *et al.*, 2023b] Aoran Xiao, Jiaxing Huang, Weihao Xuan, Ruijie Ren, Kangcheng Liu, Dayan Guan, Abdulmotaleb El Saddik, Shijian Lu, and Eric P Xing. 3d semantic segmentation in the wild: Learning generalized models for adverse-condition point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9382–9392, 2023.

[Xie *et al.*, 2020] Saining Xie, Jiatao Gu, Demi Guo, Charles R Qi, Leonidas Guibas, and Or Litany. Pointcontrast: Unsupervised pre-training for 3d point cloud understanding. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pages 574–591. Springer, 2020.

[Yi *et al.*, 2021] Li Yi, Boqing Gong, and Thomas Funkhouser. Complete & label: A domain adaptation approach to semantic segmentation of lidar point clouds. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 15363–15373, 2021.

[Yun *et al.*, 2019] 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 *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6023–6032, 2019.

[Zhao *et al.*, 2021] Sicheng Zhao, Yezhen Wang, Bo Li, Bichen Wu, Yang Gao, Pengfei Xu, Trevor Darrell, and Kurt Keutzer. epointda: An end-to-end simulation-to-real domain adaptation framework for lidar point cloud segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 3500–3509, 2021.

[Zheng *et al.*, 2022] Chunran Zheng, Qingyan Zhu, Wei Xu, Xiyuan Liu, Qizhi Guo, and Fu Zhang. Fast-livo: Fast and tightly-coupled sparse-direct lidar-inertial-visual odometry. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4003–4009. IEEE, 2022.

[Zhou and Tuzel, 2018] Yin Zhou and Oncel Tuzel. Voxelnet: End-to-end learning for point cloud based 3d object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4490–4499, 2018.

[Zhu *et al.*, 2021] 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9939–9948, 2021.

[Zou *et al.*, 2018] Yang Zou, Zhiding Yu, BVK Kumar, and Jinsong Wang. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *Proceedings of the European conference on computer vision (ECCV)*, pages 289–305, 2018.