Domain Discrepancy Aware Active Learning for Cross-domain LiDAR Point Cloud Segmentation

Anonymous Author(s)

Abstract

The generalization of deep learning-based point cloud semantic segmentation models to target domains is crucial for its practical use. Although many efforts have been made to explore unsupervised domain adaptation (UDA) to transfer knowledge learned from source domains to the target, a significant performance disparity remains between the current UDA model and the one trained on a fully annotated target domain. Recent active domain adaptation (ADA) approaches seek to enhance the model adaptation performance by labelling the most informative subset of target samples. However, current ADA methods for the point cloud segmentation task depend on metrics derived from a model trained on the source domain for sample selection, often overlooking the discrepancy between the source and target domain. Thus, in this paper, we propose a **D**omain **D**iscrepancy-aware **A**ctive **L**earning (DDAL) approach for the cross-domain LiDAR point cloud semantic segmentation, achieving high segmentation accuracy with minimum annotation costs. Firstly, a source-prototype-guided data selection method is proposed to identify and annotate those target points which deviate from the data distribution of the source domain. To further learn domain-invariant representation, we propose a balanced source-target mixing strategy to mix the annotated data with source samples of the same scale. The annotated target data and the randomly selected source data construct an intermediate dataset for knowledge transferring, which gradually narrows the domain gap between source and target datasets. Extensive experiments on synthetic-to-real and real-to-real benchmarks verify that our approach outperforms the state-of-the-arts using only 0.1% active learning budget, improving by 7.8%, 8.2%, 4.4%, 5.1%, respectively.

CCS Concepts

• **Computing methodologies** → *Scene understanding*; **Transfer learning**; Active learning settings.

Keywords

Point cloud Semantic segmentation; Domain adaptation; Active learning

1 Introduction

Point cloud semantic segmentation plays a key role in 3D environment perception and has a wide range of applications in autonomous driving [22], navigation [43], and robotics [47]. Recent works have made significant breakthroughs in the segmentation performance with the help of dense annotations in a fully supervised way [32, 54, 66, 71]. However, the acquirement of large amounts of point-level annotations for 3D point clouds is extremely time-consuming and labour-intensive. Moreover, in real-world situations, it is often impractical to annotate all points to train an accurate semantic segmentation model. A straightforward solution is to adapt a pre-trained model to target scenes. However, each dataset is unique, with variations in distributions influenced by factors such as different LiDAR configurations and environmental conditions. These differences lead to domain shifts that hinder models from effectively generalizing to new scenes.

To address this issue, many recent unsupervised domain adaptation (UDA) methods are proposed to improve the knowledge transfer from source to target domains. Yuan et al. [67] proposed a density-guided translator to translate point density between domains and integrate it into a two-stage self-training pipeline. Li et al. [20] proposed Cooperative Context Learning (CCL), which leverages local context embeddings and global prototype-based attention to bridge the context gap across domains, enhancing domain-invariant feature learning for improved adaptation. Although these UDA methods avoid data annotation, a substantial performance gap still exists compared to fully supervised approaches.

To enhance performance in the target domain, active learning (AL) emerged as an optimal approach. Recent adaptive domain adaptation (ADA) methods [50, 61, 68] aim to greatly improve adaptation performance by selecting and annotating a few informative target samples in an active learning manner. Annotator [58] proposes a voxel-centric query method to select salient and exemplar voxels and then annotate all points within these voxels. Traditional active learning methods employ random, entropy and margin based selection strategies to identify the most representative samples for annotation. However, these methods mainly rely on the predictions of the segmentation model to identify the most important target points without considering the domain discrepancy between the source and target. As the segmentation model is trained on the source, the data selection based on the predictions tends to result in the selected points being concentrated in local areas, which harms the representativeness of the selected data in terms of bridging the domain gap.

Thus, in this paper, we propose a **D**omain **D**iscrepancy-aware **A**ctive **L**earning (DDAL) approach for the cross-domain LiDAR point cloud semantic segmentation. It quantifies domain discrepancy via source prototypes by computing feature differences between target points and the source domain. Specifically, to describe the source feature structures, we first construct prototypes in the feature space for the source domain. As the semantic information for the source data is available, for each category, the feature centre is calculated as the prototype. Guided by the source prototype, we then use it to represent the structure of source domain and calculate the difference in feature space from each target point to the prototypes using a best-to-second method. This difference serves as an indicator of domain discrepancy between the target and the source domain. Subsequently, we weight this domain discrepancy indicator with the model's prediction uncertainty, selecting target points that are both representative and exhibit domain discrepancy. The classification of a point depends on its own value and the context of its regional neighborhood. Therefore, if the model can

access a point cloud scan containing both source and target domain information during training, it can learn domain-invariant feature representations. This fundamental consideration motivates our initial exploration of mixing-based methodologies. Current mixing techniques primarily operate by preserving spatial consistency during point selection. LaserMix [15] selects laser beams from different LiDAR scans to mix, PolarMix [56] swaps points within the same range of azimuth angles in scene level or rotating-pasting selects points of certain classes from one LiDAR scan and rotates for multiple copies before pasting into another scan in the instance-level. These methods share a fundamental characteristic of requiring structurally complete or continuous point clouds. However, in ADA scenarios where AL selects extremely sparse but high-value target points, if the quantity of source data is not constrained as in existing mix technologies, then a large amount of source information could overshadow the very limited target information. Thus, we propose a balanced source-target mixing strategy to mix the source and annotated target points. Specifically, we randomly sample an equal number of points from the source to match the annotated target points, then concatenate them to form a new intermediate-domain dataset. This mixed dataset contains an equal proportion of source and target semantic information, which helps further reducing the domain gap.

The main contributions of this research are as follows:

- We propose a domain discrepancy-aware active learning method, i.e. DDAL, for cross-domain point cloud semantic segmentation. Extensive experiments demonstrate that DDAL outperforms state-of-the-art approaches by a clear margin in synthetic-to-real and real-to-real benchmark. Moreover, leveraging knowledge from the source, our method surpasses the fully-supervised performance in these experiments by 5.4%, 2.2% and 0.8%, respectively.
- We propose a source-prototype-guided query strategy to measure the domain discrepancy based on the distance between target points and the centers of source categories in feature space. The most representative target points are identified based on the measurement of domain discrepancy and the uncertainty predicted with the segmentation model.
- A balanced mixing technique, which adapts point cloud mixing to the ADA framework, is proposed and incorporated into the active learning process. This technique combines newly annotated target data with randomly selected source data at matching scales through mixed sampling, enabling more reliable domain-invariant feature learning.

2 Relate Work

2.1 Point Cloud Semantic Segmentation

The goal of point cloud semantic segmentation is to assign a predefined semantic label to each point. With the development of deep neural networks, CNN and Transformer architectures have been proposed for feature extraction, significantly advancing semantic segmentation tasks. Currently, point cloud semantic segmentation methods can be broadly categorized into three categories. **Projection-based** [1, 3, 7, 14, 16, 23, 51, 52]. These methods map 3D point clouds to 2D images using techniques such as multi-view or bird's-eye view etc. Although these methods benefit from mature 2D segmentation networks, they also suffer from the loss of geometric prior information. **Point-based** [2, 8, 18, 24, 29, 34, 64]. These methods directly extract features from the individual points in the point cloud. PointNet [33] is the first proposed using max-pooling to flexibly handle the unordered issue of point clouds. However, these methods require significant computational resources and are more suitable for small-scale indoor scenarios. **Voxel-based** [17, 42, 53, 62, 63, 70]. These methods convert point clouds into voxel grids, making the structure more regular. By utilizing sparse convolutions, computational cost is reduced and information loss is minimized, making this approach more suitable for large-scale outdoor scenes. Therefore, in this work, we adopt the voxel-based MinkNet [6] network as our backbone.

2.2 Unsupervised 3D Point Cloud Semantic Segmentation with Domain Adaptation

In 3D point cloud semantic segmentation, most domain adaptation research has concentrated on unsupervised approaches [9, 11, 20, 21, 37, 39, 55, 56, 67, 69]. The goal is to transfer a model trained on a labeled source dataset to an unlabeled target dataset. Roughly, UDA methods can be divided into two groups: self-training and adversarial training. For self-training, PolarMix [56] cuts the same angle points from two point cloud scans along the azimuth axis then edit and blend them for data augmentation. CosMix [39] proposes a semantic mixing strategy that creates two new intermediate domains of composite point clouds at input level to narrow the domain gap. DGT-ST [67] proposes a density-guided translator, which translates point density between domains, and integrates it into a two-stage self-training pipeline. Li [20] proposes Cooperative Context Learning (CCL) to mitigate the context gap through effective context modeling and modulations in a cooperative manner to reduce the severer shifts. UniMix [69] blends samples from the source and target domains, considering spatial, intensity and semantic distribution, to learn domain invariant features, enabling the adaptation and generalization of models from clear to adverse weather scenes. For adversarial training, the Complete & Label [65] leverages geometric priors and transformss the domain adaptation problem into a 3D surface completion task, so as to narrow the domain gap at the input level. LiDARNet [12] enhances the model's ability to perceive boundary information in point clouds by extracting private and shared features. This enables effective boundary-aware domain adaptation, thereby improving the generalization performance of point cloud semantic segmentation across different domains. PCT [57] proposes a LiDAR point cloud transformer that transforms synthetic point clouds into representations with similar features and distributions to real point clouds to mitigate the domain gap. ASM [21] reduces the domain gap caused by target noise via learning to mask the source points during the adaptation procedure. However, the performance of these methods still exists a significant gap compared to fully supervised results.

2.3 Active Domain Adaptation

Active domain adaptation (ADA) [25, 27, 31, 41, 50, 58, 60, 61, 68]employs an iterative querying strategy to select the most informative target samples, achieving an optimal balance between

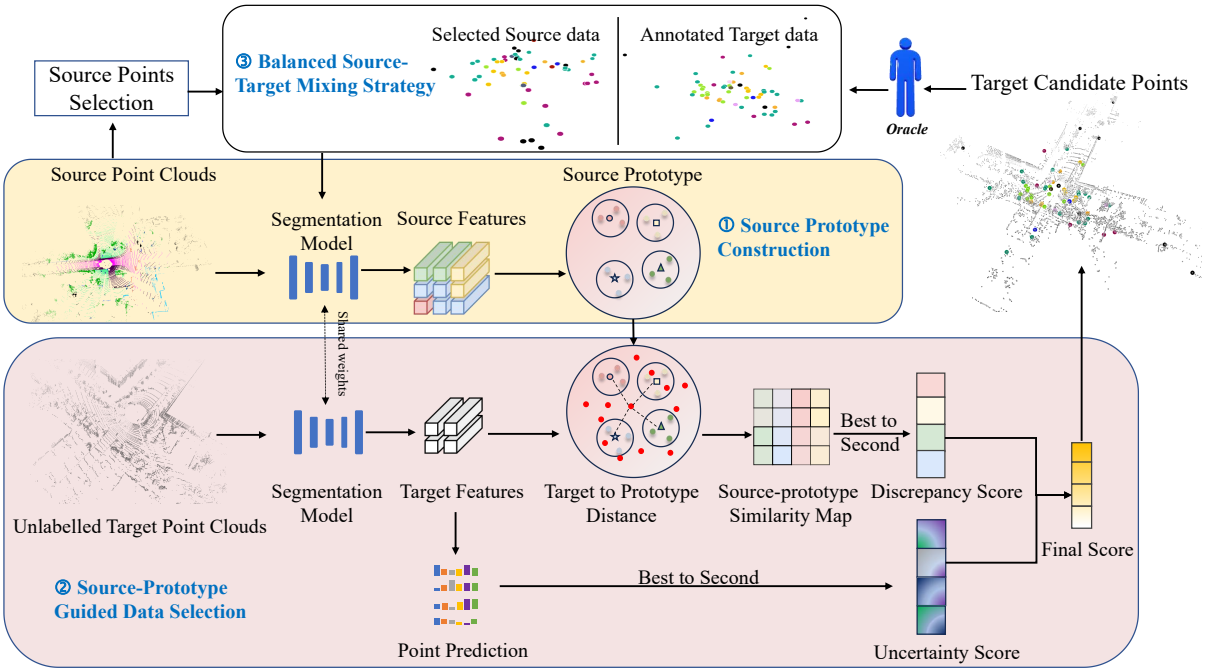

Figure 1: Overview of the proposed domain discrepancy-aware active learning framework for cross-domain LiDAR point cloud segmentation. ① Source Prototype Construction. A segmentation model extracts features from the source point clouds then using them to construct source prototypes as a semantic representation of source domain. ② Source-Prototype Guided Data Selection (SPGDS). A target-to-prototype distance and a source-prototype similarity map are calculated, generating a discrepancy score. This discrepancy score is combined with an uncertainty score from the prediction to compute a final score, guiding the selection of target candidate points for annotation. ③ Balanced Source-Target Mixing Strategy (BSTM). We randomly select an equal number of points from the source point cloud to match the annotated target data, then blend them to generate intermediate data samples for fine-tuning the segmentation model.

model performance and annotation cost. However, due to domain shifts, conventional query strategies are often ineffective in domain adaptation. EADA [61] proposes an energy-based approach, using an Energy-Based Models (EBMs) to help realize the potential of active learning under domain shift and select the most informative points from target. Peng et al. [31] proposes a prototype-guided pseudo-label generating approach that leverages the relationships between source prototypes and target features. By generating target pseudo-labels based on diverse source prototypes, mitigating reliance on potentially biased classifier predictions. LFTL [25] proposes a contrastive active sampling model that generates global intransferability by comparing and applying a weighted summation to the results of the previous and current model iterations, and then local uncertainty is then added to determine the final candidate points. However, these ADA methods are predominantly applied in the field of 2D image segmentation, with limited exploration in LiDAR point cloud semantic segmentation.

3 Method

3.1 Preliminaries

In the active domain adaptation for point cloud semantic segmentation task, we have a labelled source dataset $\mathcal{S} = \{(\mathbf{X}_i^s, \mathbf{Y}_i^s)\}_{i=1}^{N^s}$, and an unlabelled target dataset $\mathcal{T} = \{(\mathbf{X}_i^t, \mathbf{Y}_i^t)\}_{i=1}^{N^t}$, where $\mathbf{X}_i^s \in \mathbb{R}^{n_i \times 4}$ and $\mathbf{X}_i^t \in \mathbb{R}^{n_i \times 4}$ denote the i-th source and target point cloud scan respectively; $\mathbf{Y}_i^s$ and $\mathbf{Y}_i^t$ represent the semantic labels of the i-th scan; n_i is the number of points in the i-th scan and each point containing (x, y, z) coordinates and intensity values. Here, N^s and N^t denote the total number of scans in the source and target datasets. Considering a model $\mathcal{G}$ that is trained on the labelled source domain, our goal is to adapt it to the target domain to achieve accurate point cloud semantic segmentation with the help of active learning strategies.

3.2 Source Prototype Construction

Current ADA approaches for point cloud semantic segmentation mainly utilize the prediction of the pre-trained model on the source domain to identify uncertainty data on the target domain that the current source model cannot discriminate accurately. However, this process ignores the distribution discrepancy between source and target domains and tends to locate those data concentrating on certain long-tailed categories, making selected data lack diversity and representativeness. Also, when using the source-trained segmentation model to predict the target data, the prediction is inaccurate, resulting in unreliable query strategies.

Thus, in this research, the prediction and the discrepancy between the selected data and the source data distribution are considered to ensure reliable and well-distributed data selection. To

describe the structure of the source data distribution, we first construct source prototypes based on the feature centers of each category. For a class of points in the source, their feature vectors are extracted using the segmentation model $\mathcal{G}$, and their average feature vector is regarded as the prototype for the class. Specifically, as shown in Figure 1, a labeled source dataset $\mathcal{S}$ is input into the current network $\mathcal{G}$ to extract features $\mathbf{F} \in \mathbb{R}^{N^s \times d_f}$, d_f represents the dimension of the feature. Based on the category information of the source data, we sum the features of points in the same class and compute their mean to obtain the class prototype $\mathbf{p}_{b,c}$ with Equation (1):

$$\mathbf{p}_{b,c} = \frac{\sum_{i=1}^{N_c} \mathbf{f}_{i,c}}{N_c}, \tag{1}$$

where $c \in [1, C]$ denotes the class c, and C is the total number of classes in the source dataset; $\mathbf{f}_i^c$ represents the feature of the $i-th$ point belonging to class c; N_c is the number of points belonging to the class c and it is calculated using Equation (2) below:

$$N_c = \sum_{i=1}^{N_P^s} \mathbb{I}(y_i = c)\mathbf{f}_i, \tag{2}$$

where y_i represents the label corresponding to $\mathbf{f}_i$, N_P^s represents the all labeled points in source, $\mathbb{I}(y_i = c)$ is an indicator function that equals 1 if y_i belongs to the class c, and 0 otherwise.

Due to the large size of point cloud datasets, loading all data into memory is impractical, making it infeasible to compute prototypes directly from the entire source dataset. Therefore, we use an Incremental Mean Algorithm [49] to compute prototypes progressively, as shown in Equation (3):

$$\mathbf{p}_{b+1,c} = \mathbf{p}_{b,c} + \frac{\mathbf{p}_{b+1}^c - \mathbf{p}_{b,c}}{x_{b+1}}, \tag{3}$$

where x_{b+1} represents the cumulative count of class c points up to batch $b+1$. This approach allows us to obtain the final source dataset prototypes $\mathbf{P} \in \mathbb{R}^{C \times d_f} = \{\mathbf{p}_i\}_{i=1}^C$.

The constructed prototypes serve as centroids in the feature space for each class in the source data, capturing distinct class semantics. These prototypes are updated and utilized at each active learning stage.

3.3 Source-Prototype Guided Data Selection

Once the source prototypes are constructed, we can use them to guide the selection of target data. As shown in Figure 1, we input the unlabeled target point clouds into the segmentation network to obtain their features $\mathbf{F}^t$, then calculate their distances to the source prototypes and the formula is as follows:

$$\mathbf{d}_{i,c} = \|\mathbf{f}_i^t - \mathbf{p}_c\|, \tag{4}$$

where $\mathbf{f}_i^t \in \mathbf{F}^t$ represents the feature vector of the i-th point in the target dataset. $d_{i,c}$ is the Euclidean distance between the i-th unlabeled target point feature and the prototype for class c.

By calculating $d_{i,c}$ for all $c \in [1, C]$, we obtain a $1 \times C$ distance vector $\mathbf{D}_i = [d_{i,1}, d_{i,2}, \ldots, d_{i,C}]$, where each element $d_{i,c}$ represents the distance between the target point i and each class prototype.

Subsequently, the similarity map $\mathbf{D}_i'$ can be obtained by taking the reciprocal of the target-to-prototype distance $\mathbf{D}_i$, and then we derive a discrepancy score $M_{i,ds}$ of target point i by best to second, as shown in Equation (5):

$$M_{i,ds} = S_{R1}(f(\mathbf{D}_i')) - S_{R2}(f(\mathbf{D}_i')), \tag{5}$$

where $f(\cdot)$ is a normalization and here we use $softmax$ function, S_{Ri} represents the $i-th$ value in the descending order, $S_{R1}(f(\mathbf{D}_i'))$ and $S_{R2}(f(\mathbf{D}_i'))$ are the highest and the second highest value.

Next, we compute the final score $M_{i,final}$ for each unlabeled target point i. This is achieved by combining the discrepancy score with the uncertainty score $M_{i,us}$ through weighted fusion. The details are provided in Equation (6) and Equation (7), respectively:

$$M_{i,us} = S_{R1}(\mathbf{p}_i^t) - S_{R2}(\mathbf{p}_i^t), \tag{6}$$

$$M_{i,final} = \alpha \times M_{i,ds} + (1-\alpha) \times M_{i,us}, \tag{7}$$

where $\mathbf{p}_i^t \in \mathbb{R}^{1 \times C}$ is computed as $f(h(\mathbf{f}_i^t))$ is a segmentation head, α is a weight parameter ranging from 0 to 1. We compute the final score by the weighted sum of the domain difference score and the uncertainty score, in order to select target points that balance both domain diversity and representativeness.

We rank the target points based on $M_{i,final}$ and select the k candidate points with smallest final score values. These selected points are then queried for annotation by an Oracle (e.g., an expert or specialist).

3.4 Balanced Source-Target Mixing Strategy

Using only isolated target data is insufficient to effectively adapt the network to the target domain. Therefore, we introduce a mixing strategy to construct intermediate-domain data, bridging the domain gap by learning domain-invariant features. We found that mixing active annotated target points with source data performs better than without mixing. Additionally, we observed that mixing them in equal proportions yields the best results. Specifically, during the data loading stage, we first apply data augmentation to both datasets, compute the number of labeled points in each frame of the target data, and then retrieve an equivalent number of points $\mathcal{Q}_i$ from the source point cloud scans following Equation (8), where $\mathcal{Q}_i$ is source point set that was selected, $\mathcal{T}_i^a$ denote target point set with annotation cloud of the i-th scan. Then, we mixed the selected source points with the target points to a new intermediate dataset $\mathbf{I} = \{\hat{\mathbf{X}}, \hat{\mathbf{Y}}\}$, where $\hat{\mathbf{X}}$ present the mixing point data set and $\hat{\mathbf{Y}}$ denote the corresponding label set:

$$\begin{aligned} \mathcal{Q}_i \subset \mathcal{S}_i \quad & \|\mathcal{Q}_i\| = \|\mathcal{T}_i^a\|, \\ \hat{\mathbf{X}} &= concat\{X_{\mathcal{Q}_i}, X_{\mathcal{T}_i^a}\}, \\ \hat{\mathbf{Y}} &= concat\{Y_{\mathcal{Q}_i}, Y_{\mathcal{T}_i^a}\}, \end{aligned} \tag{8}$$

The newly constructed intermediate data is then fed to the network for fine-tuning. Since only labeled points in the constructed data are used for back propagation, we adopt cross-entropy as our loss function following Equation(9), where $\hat{x}_i \in \hat{\mathbf{X}}$ and $\hat{y}_i \in \hat{\mathbf{Y}}$:

$$\mathcal{L}_{\text{CE}} = -\frac{1}{|I|} \sum_{i \in I} \sum_{c=1}^{C} \hat{y}_{i,c} \log h(\hat{x}_{i,c}), \tag{9}$$

	Methods	car	bi.cle	mt.cle	truck	oth-v.	pers.	bi.clst	mt.clst	road	parki.	sidew.	other-g.	build.	fence	veget.	trunk	terr.	pole	traf.	mIoU
	Source-Only	50.9	4.8	6.2	0.5	2.6	20.4	53.1	4.4	35.7	4.1	36.5	0.0	16.5	26.1	60.5	29.3	49.7	27.5	5.0	22.8
DA	ADDA [44]	52.5	4.5	11.9	0.3	3.9	9.4	27.9	0.5	52.8	4.9	27.4	0.0	61.0	17.0	57.4	34.5	42.9	23.2	4.5	23.0
DA	AdvEnt [45]	58.3	5.1	14.3	0.3	1.8	14.3	44.5	0.5	50.4	4.3	34.8	0.0	48.3	19.7	67.5	34.8	52.0	33.0	6.1	25.8
DA	CRST [72]	62.0	5.0	12.4	1.3	9.2	16.7	44.2	0.4	53.0	2.5	28.4	0.0	57.1	18.7	69.8	35.0	48.7	32.5	6.9	26.5
DA	ST-PCT [57]	70.8	7.3	13.1	1.9	8.4	12.6	44.0	0.6	56.4	4.5	31.8	0.0	66.7	23.7	73.3	34.6	48.4	39.4	11.7	28.9
DA	CoSMix [39]	75.1	6.8	29.4	27.1	11.1	22.1	25.0	**24.7**	79.3	14.9	46.7	0.1	53.4	13.0	67.7	31.4	32.1	37.9	13.4	32.2
DA	PolarMix [56]	76.3	8.4	17.8	3.9	6.0	26.6	40.8	15.9	70.3	0.0	44.4	0.0	68.4	14.7	69.6	38.1	37.1	40.6	10.6	31.0
DA	DGT-ST [67]	92.9	17.3	43.4	15.0	6.1	49.2	54.2	4.2	86.4	19.1	62.3	0.0	78.2	9.2	83.3	56.0	59.1	51.2	32.3	43.1
SSDA	MME [36]	51.0	5.6	13.1	1.3	7.3	15.1	54.4	4.4	43.1	0.2	28.3	0.0	60.7	13.3	66.1	30.1	39.9	24.8	6.6	24.5
SSDA	APE [13]	58.6	6.2	16.6	3.1	11.3	14.2	35.8	3.7	61.5	1.7	30.3	0.0	54.7	15.4	64.6	20.0	45.5	23.9	9.1	25.1
SSDA	APE-PCT [57]	58.1	7.3	17.8	2.6	13.9	24.7	46.5	5.1	60.5	1.9	31.3	0.0	56.8	14.6	67.9	23.7	44.3	26.1	9.3	27.0
SSDA	CoSMix-SSDA [38]	76.9	10.4	27.1	23.1	13.4	24.0	21.7	27.9	75.8	17.9	49.7	0.1	60.3	14.7	69.8	36.8	40.9	45.6	16.2	34.3
SSDA	DDAL (Ours)(0.04%)	95.1	0.0	27.7	51.8	45.0	39.6	0.0	0.0	91.6	40.0	78.0	2.6	89.4	58.5	88.2	65.9	76.1	60.5	50.2	50.5
ADA	Annotator [58]	95.2	22.0	59.7	69.0	49.4	63.4	82.1	3.6	84.1	28.9	71.4	**1.7**	85.4	58.8	85.6	60.1	73.2	60.3	41.6	57.7
ADA	**DDAL (Ours)**(0.1%)	**96.9**	**51.0**	**78.1**	**77.5**	**64.6**	**76.1**	**88.5**	0.6	**93.7**	**47.7**	**81.0**	1.1	**89.9**	**58.9**	**87.8**	**69.4**	**73.5**	**63.0**	**44.7**	**65.5**
	Target-Only	95.5	17.6	61.7	56.0	54.7	69.9	83.0	0.1	92.6	43.0	79.3	0.9	89.9	61.4	88.0	65.0	74.7	62.7	45.9	60.1

Table 1: Per-class results on task of SynLiDAR [57] → SemanticKITTI [4]. Domain adaptation (DA) results are reported from [39, 44, 45, 56–58, 67, 72].

	Methods	car	bike	pers.	rider	grou.	buil.	fence	plants	trunk	pole	traf.	garb.	cone	mIoU
	Source-Only	51.5	1.9	43.8	42.7	79.7	58.2	36.3	66.4	23.1	30.2	0.8	11.8	2.9	34.6
DA	CRST [72]	22.0	6.8	23.5	31.8	60.3	58.2	9.1	63.2	18.9	41.6	1.9	13.5	1.0	27.1
DA	ST-PCT [57]	27.8	6.6	28.9	34.8	63.9	64.1	12.1	63.7	18.6	41.0	4.9	16.6	1.6	29.6
DA	CoSMix [39]	36.2	10.6	55.8	51.4	78.7	66.2	24.9	71.3	23.5	34.2	22.5	28.9	20.4	40.4
DA	PolarMix [56]	25.0	10.7	32.6	39.1	79.0	44.8	23.8	64.2	11.9	29.6	5.8	15.3	13.3	30.4
DA	DGT-ST [67]	55.1	**70.7**	46.1	**74.2**	30.1	36.3	44.1	**81.0**	4.3	**62.8**	10.3	**78.5**	**67.2**	50.8
SSDA	MMD [44]	28.9	13.9	25.5	35.7	70.4	53.3	30.2	64.3	6.7	23.2	1.7	5.6	3.3	27.9
SSDA	MME [36]	25.0	6.7	33.2	40.2	71.2	56.1	37.1	61.9	11.0	31.2	0.4	7.3	5.7	33.2
SSDA	APE [13]	21.5	9.6	34.3	40.1	71.6	55.9	34.3	62.6	16.3	31.3	0.9	2.3	13.3	30.3
SSDA	APE-PCT [57]	27.2	9.3	34.7	36.3	70.9	62.3	35.1	62.9	15.8	31.6	0.8	8.7	9.8	31.2
SSDA	CoSMix-SSDA [38]	33.4	27.4	54.9	50.6	79.0	68.5	29.6	73.0	22.5	38.4	13.6	26.4	16.2	41.0
SSDA	DDAL (Ours) (0.01%)	73.3	59.6	70.4	57.6	80.2	78.0	61.1	77.6	36.3	40.8	40.4	32.2	40.1	57.5
ADA	Annotator [58]	67.4	18.0	64.0	52.0	78.5	61.5	1.5	68.6	**48.5**	32.7	37.9	50.8	43.8	52.0
ADA	**DDAL (Ours) (0.1%)**	**75.5**	57.1	**69.7**	62.6	**80.0**	**78.7**	**58.4**	78.1	47.0	39.5	**48.0**	38.7	48.6	**60.2**
	Target-Only	73.2	60.3	68.9	62.3	82.2	77.3	63.7	78.1	40.1	29.1	45.8	27.0	45.7	58.0

Table 2: Per-class results on task of SynLiDAR [57] → SemanticPOSS [20].

4 Experiments

4.1 Setup

Dataset. To demonstrate the effectiveness of the proposed active learning method for domain adaptation task, we perform experiments on four different datasets: SynLiDAR [57], SemanticPOSS [28], SemanticKITTI [4] and nuScenes [5]. Experiments are conducted on SynLiDAR → SemanticKITTI with 19 shared classes, SynLiDAR → SemanticPOSS with 13 shared classes, nuScenes → SemanticKITTI and reverse adaptation task SemanticKITTI → nuScenes with 7 shared classes.

SynLiDAR [57] is a synthetic point cloud dataset that includes scenes such as urban areas, suburban towns, residential communities, and ports. It consists of 13 LiDAR point cloud sequences, 32 semantic categories, 198, 396 scans, and 19, 482 million points. We use the officially designated subset for training.

SemanticKITTI [4] is the most prevalent large-scale point cloud dataset for point cloud segmentation. This dataset is collected from the real world by Velodyne HDL-64E LiDAR. It contains 22 sequences with 41000 frames and 25 semantic categories. Following [20, 58], we choose sequences 00-10 for training and sequence 08 for validation.

SemanticPOSS [28] is a real-world dataset collected at Peking University using a Pandora 40-line LiDAR sensor, featuring a total of 2, 988 scans. This dataset includes 13 semantic categories, with each scan containing approximately 80, 000 to 100, 000 points, capturing typical urban outdoor scenes. Following prior works [20, 39, 56, 58], sequence 03, comprising 500 scans, is allocated for validation, while the remaining 2, 488 scans are utilized for training.

nuScenes-lidarseg [5] is a real-world LiDAR dataset captured using a 32-beam sensor with a field of view ranging from [−30 °, 10 °]. It comprises a total of 40, 000 frames collected across two major cities: Singapore and Boston. 850 scenes are designated for training, while the remaining 150 scenes are reserved for validation, providing diverse urban driving conditions for segmentation task.

Methods	vehicle	person	road	sidewalk	terrain	manmade	vegetation	mIoU
Source-Only	24.5	13.8	63.2	33.9	20.2	52.2	19.7	32.5
Mix3D [26]	37.9	6.7	42.0	5.7	27.6	41.2	65.4	32.4
CoSMix [10]	44.6	13.9	36.1	10.2	29.3	54.4	69.1	36.8
SN [48]	25.7	5.5	19.6	2.2	23.5	27.7	61.1	23.6
RayCast [19]	28.3	16.1	45.8	9.4	20.6	38.6	61.8	31.5
LiDOG [40]	60.1	9.0	47.4	16.4	32.6	54.2	68.8	41.2
Annotator [59]	95.8	66.1	88.5	74.9	75.9	84.2	86.9	81.8
DDAL (Ours)	**98.1**	**75.4**	**91.6**	**81.1**	**76.0**	**91.7**	**89.6**	**86.2**
Target-Only	97.8	74.0	90.6	79.5	76.5	89.8	89.5	85.4

Table 3: Per-class results on task of nuScenes[5] → SemanticKITTI[28] using only 0.1% budgets. DA results are reported from [59]

Methods	vehicle	person	road	sidewalk	terrain	manmade	vegetation	mIoU
Source-Only	51.1	8.9	55.4	25.1	3.7	40.1	51.7	33.7
Mix3D [26]	33.7	11.2	58.5	12.9	5.3	50.4	48.6	31.5
CoSMix [10]	35.9	0.0	58.1	11.6	9.0	45.2	49.1	29.8
SN [48]	21.4	0.0	60.5	15.1	6.2	31.9	45.7	25.8
RayCast [19]	28.8	0.0	59.3	16.1	12.5	49.7	49.8	30.9
LiDOG [40]	24.0	14.9	70.6	24.6	14.0	45.3	50.9	34.9
Annotator [59]	**88.8**	58.6	93.1	62.6	68.4	82.4	77.3	75.9
DDAL (Ours)	86.5	**68.8**	**95.5**	**71.2**	**75.1**	**87.2**	**83.2**	**81.0**
Target-Only	88.4	75.8	95.5	71.2	75.4	87.5	84.9	82.7

Table 4: Per-class results on task of SemanticKITTI[4] → nuScenes[5].

Implementation. In our experiments, we use the PyTorch [30] implementation of MinkNet [6] in Annotator [58] as the backbone network. All experiments are conducted on a single NVIDIA RTX A6000 GPU. The model is optimized using SGD with a momentum of 0.9 and weight decay of 0.0001. The warmup is applied for the first epoch to linearly increase the learning rate to the base learning rate of 0.01. A cosine learning rate decay schedule with an initial learning rate of 0.01 is adopted to adjust the learning rate dynamically. For the synthetic-to-real, the batch size is set to 16 for the source-only/target-only stages with 10 epochs for SynLiDAR → SemanticKITTI and 20 epochs for SynLiDAR → SemanticPOSS, and 14 for the adaptation stage with 50 epochs. The weight parameter α is set to 0.4 for SynLiDAR → SemanticKITTI and 0.6 for SynLiDAR → SemanticPOSS, respectively. In the experiments on real-to-real benchmarks, the batch size is set to 16 for the source-only/target-only stages with 10 epochs, and 10 for the adaptation stage with 50 epochs. The weight parameter α is set to 0.6. All experiments were performed at a 0.1% labeling budget unless otherwise noted.

4.2 Comparisons with SOTA Methods

We compare our method with state-of-the-art domain adaptation approaches on point cloud semantic segmentation tasks. The segmentation results in the setting of SynLiDAR→SemanticKITTI is shown in Table 1. Our proposed DDAL obtains 65.5% mIoU and significantly outperforms the other methods only using 0.1% labels. Compared with the source-only method which trains the segmentation model on the source data and directly evaluates its performance on the target, we gain 42.7% mIoU improvements. For a fair comparison with the semi-supervised domain adaptation (SSDA) methods, we reduce the annotation budget to the same level, i.e. 0.04% annotations. It is clear to see that our results still surpass the best semi-supervised method by 6.2%. In the case of ADA, the proposed DDAL outperforms the state-of-the-art Annotator [58] by 7.8%. Thus, compared with the query strategies used in Annotator, the proposed source-prototype-guided target data selection can identify those data that is more aware of the domain disparity between source and target, and allows for more representative target point selection. Moreover, our DDAL method can surpass the full-supervised performance (Target-Only) by 5.4%. This is because we gradually incorporate the most representative target points in the active learning process while simultaneously preserving knowledge from the source domain by mixing source and target data during training. This can effectively enhance the segmentation performance on the target domain.

We also present the semantic segmentation results of SynLiDAR → SemanticPOSS in Table 2. Our method, with 0.1% annotation budget, achieves the best performance of 60.2% mIoU, outperforming the state-of-the-art approaches. Compared to the source-only baseline, our method improves by 25.8%. Similarly, when we reduced the annotation budget to 0.01% DDAL (0.01%), its performance still significantly outperforms other semi-supervised methods. The Annotator [58] reaches the segmentation accuracy of 52.0% with

8.2% lower than ours. The DDAL method maintains a clear edge, particularly in categories like car, person, ground, building, fence and traffic sign. Additionally, our method outperforms the fully supervised target-only results in the two experiments by 5.3% and 2.2%, respectively.

To further validate the efficacy of our method, we conduct experiments on real-to-real domain adaptation scenarios. On the nuScenes→SemanticKITTI task in Table 3, our approach achieves 86.2% mIoU, surpassing Annotatorby 4.4% and outperforming the Target-Only by 0.8%. For the reverse adaptation task (SemanticKITTI→ nuScenes), as shown in Table 4, our method obtains a result of 81.0%, exceeding Annotator by 5.1% while reducing the performance gap to the Target-Only benchmark to 1.7%. These results demonstrate the robustness and generalizability of our framework across diverse real-world domain shifts.

Our performance surpasses that of fully supervised learning on the target domain in the above three experiments because the source data, to some extent, serves as a form of data augmentation and enrichment, broadening the model's knowledge and thereby aiding in the improvement of target data performance.

4.3 Ablation Studies

In this section, we conduct ablation studies on the synthetic-to-real cases to demonstrate the effectiveness of each module of DDAL.

Methods	SynLiDAR → SemanticKITTI	SynLiDAR → SemanticPOSS
Source-/Target-Only	22.0 / 61.1	34.5 / 57.9
Random	62.2	58.3
Entropy	61.8	55.6
Margin	64.0	60.0
VCD [58]	56.8	45.5
SPGDS	**65.5**	**60.2**

Table 5: Comparison of different active query strategies for domain adaptation with a mixing rate of 1x. Our query strategy consistently achieves superior performance on both SynLiDAR → SemanticKITTI and SynLiDAR → SemanticPOSS.

Effect of prototype-guide active query strategy. As shown in Table 5, we provide comparative results with various active query strategies under the 0.1% budget and the mixing rate of 1x. Our proposed method SPGDS achieves notable performance improvements over other query strategies in both settings of SynLiDAR → SemanticKITTI and SynLiDAR → SemanticPOSS. In the SynLiDAR → SemanticKITTI task, SPGDS attains an accuracy of 65.5%, surpassing the Random by 3.3%, the Entropy by 3.7%, and the Margin by 1.5%. For the SynLiDAR → SemanticPOSS task, SPGDS also outperforms other methods, achieving a mIoU of 60.2%, outperforming Random by 1.9%, Entropy by 4.6%, and Margin by 0.2%, respectively. These results demonstrate the effectiveness of our proposed data selection strategy, i.e. SPGDS, in identifying the most representative points on the target domain.

Effect of active learning budget. As shown in Figure 2, we evaluate the semantic segmentation accuracy under different active learning budgets with balanced source-target mixing strategy (BSTM). Please note that the semantic segmentation accuracy is improved with the active learning budget increasing. When the budget reaches 1% and 3%, our method can achieve the highest accuracy of 67.1%. However, for a fair comparison with the current domain adaptation methods which only present the accuracy under the budget of 0.1%, we adopt the same budget in our experiments.

We also analyze the cost rate under different active learning budgets. The cost rate is calculated as the ratio of segmentation accuracy over the labelling cost. As shown in Figure 2, the 0.1% budget achieves the highest cost rate which is 654. As the budget increases, the cost rate declines abruptly. This indicates that the 0.1% budget provides the best balance between performance and efficiency, maximizing impact with minimal investment in domain adaptation.

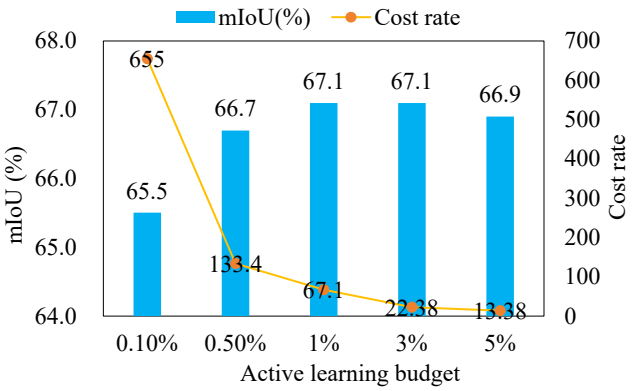

Figure 2: Comparisons with different active learning budgets on the task of SynLiDAR → SemanticKITTI using our proposed SPGDS query strategy.

Effect of mixing rate. As shown in Figure 3, the impact of mixing rates (0x, 1x, 10x, 20x and 30x) on the semantic segmentation accuracy across mainstream query strategies and our SPGDS query method, with target-only as the baseline. It is obvious that the 1x mixing rate consistently delivers the best performance for all query strategies. With the mixing rate increasing, more source data participates in the training, hindering the knowledge transfer to the target and decreasing the segmentation accuracy. These results demonstrate that the proposed balanced source-target mixing (1x rate) yields the most significant gains in domain adaptation.

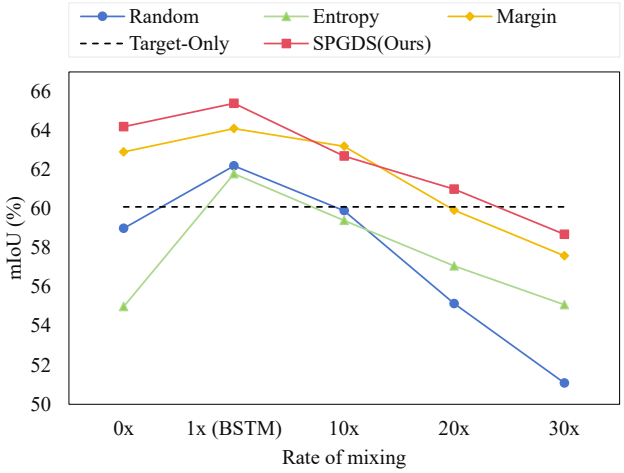

Figure 3: Comparisons with different mixing rates on SynLiDAR → SemanticKITTI using different active learning query strategies.

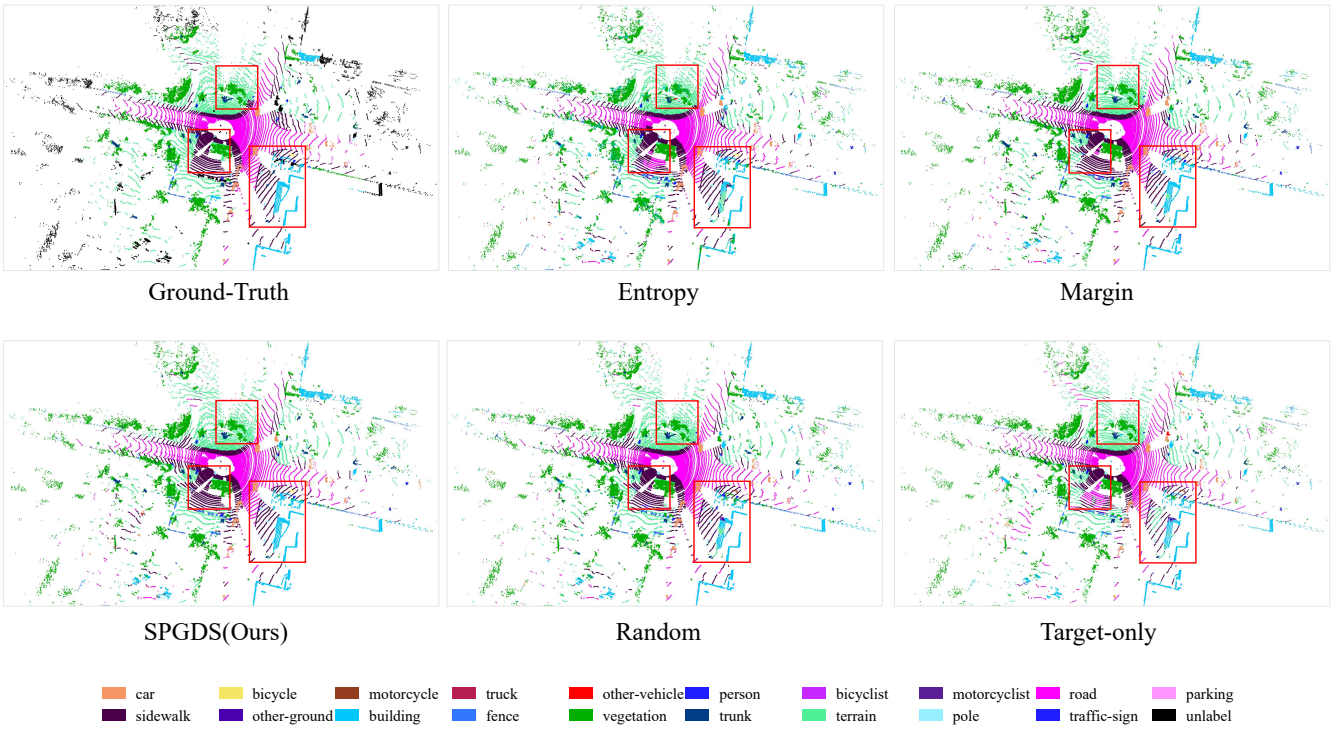

Figure 4: Visualization of Segmentation Results (SynLiDAR [57] → SemanticKITTI [4]). We compare our method with ground truth, target-only, Margin [35], Entropy [46] and Random. Best viewed in colour.

Impact of AL and Mixing Modules. As shown in Table 6, we present the ablation study results of each component. The baseline model that only uses the source data to train the segmentation model and test on the target achieves a mIoU of 22.8%. Adding the active learning module using the proposed query strategy SPGDS and training the model only with the selected target points raise the mIoU to 62.2%, demonstrating a substantial improvement. Similarly, when we do not use the SPGDS method, but instead select the same number of points as in the SPGDS method and perform BSTM, the result is 64.2%. When both the SPGDS and BSTM are combined, mIoU further increases to 65.4%, which proves that the integration of these modules maximizes performance. These results confirm that our proposed AL and Mixing modules significantly enhance the model's effectiveness in domain adaptation.

Method	SPGDS	BSTM	mIoU (%)
MinkNet [6]			22.8
	✓		62.2
		✓	64.2
	✓	✓	**65.5**

Table 6: The ablation study of each component on SynLiDAR→SemanticKITTI.

Visualizations. As show in Figure 4, we present the visualization of segmentation results on SynLiDAR → SemanticKITTI. We compare our method with the ground truth, segmentation with target-only, margin, entropy and random. It can be observed that the segmentation results of our approach are more accurate and some tiny and challenging parts can be segmented correctly, as highlighted in the red bounding boxes. *More qualitative and quantitative evaluation of real-to-real benchmarks can be found in the supplementary materials.*

5 Conclusion

In this work, we propose a Domain Discrepancy-aware Active Learning (DDAL) approach for cross-domain LiDAR point cloud semantic segmentation. By introducing a source-prototype-guided active query strategy (SPGDS), we effectively identify and annotate the highly representative target points that deviate significantly from the source domain. A balanced source-target mixing strategy (BSTM) is presented to construct an intermediate dataset to reduce the domain gap and enhance model adaptation. Both qualitative and quantitative experiments demonstrate the superior segmentation performance on SynLiDAR → SemanticKITTI, SynLiDAR → SemanticPOSS, nuScenes → SemanticKITTI and SemanticKITTI → nuScenes benchmarks.

References

[1] Eren Erdal Aksoy, Saimir Baci, and Selcuk Cavdar. 2020. Salsanet: Fast road and vehicle segmentation in lidar point clouds for autonomous driving. In *2020 IEEE intelligent vehicles symposium (IV)*. IEEE, 926–932.

[2] Zhaochong An, Guolei Sun, Yun Liu, Fayao Liu, Zongwei Wu, Dan Wang, Luc Van Gool, and Serge Belongie. 2024. Rethinking Few-shot 3D Point Cloud Semantic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3996–4006.

[3] Angelika Ando, Spyros Gidaris, Andrei Bursuc, Gilles Puy, Alexandre Boulch, and Renaud Marlet. 2023. Rangevit: Towards vision transformers for 3d semantic segmentation in autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5240–5250.

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

[5] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. 2020. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 11621–11631.

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

[7] Tiago Cortinhal, George Tzelepis, and Eren Erdal Aksoy. 2020. Salsanext: Fast, uncertainty-aware semantic segmentation of lidar point clouds. In *Advances in Visual Computing: 15th International Symposium, ISVC 2020, San Diego, CA, USA, October 5–7, 2020, Proceedings, Part II 15*. Springer, 207–222.

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

[9] Maximilian Jaritz, Tuan-Hung Vu, Raoul de Charette, Emilie Wirbel, and Patrick Pérez. 2020. xmuda: Cross-modal unsupervised domain adaptation for 3d semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 12605–12614.

[10] Maximilian Jaritz, Tuan-Hung Vu, Raoul de Charette, Emilie Wirbel, and Patrick Pérez. 2020. xmuda: Cross-modal unsupervised domain adaptation for 3d semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 12605–12614.

[11] Maximilian Jaritz, Tuan-Hung Vu, Raoul De Charette, Émilie Wirbel, and Patrick Pérez. 2022. Cross-modal learning for domain adaptation in 3d semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 2 (2022), 1533–1544.

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

[13] Taekyung Kim and Changick Kim. 2020. Attract, perturb, and explore: Learning a feature alignment network for semi-supervised domain adaptation. In *European conference on computer vision*. Springer, 591–607.

[14] Lingdong Kong, Youquan Liu, Runnan Chen, Yuexin Ma, Xinge Zhu, Yikang Li, Yuenan Hou, Yu Qiao, and Ziwei Liu. 2023. Rethinking range view representation for lidar segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 228–240.

[15] Lingdong Kong, Jiawei Ren, Liang Pan, and Ziwei Liu. 2023. Lasermix for semi-supervised lidar semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 21705–21715.

[16] Abhijit Kundu, Xiaoqi Yin, Alireza Fathi, David Ross, Brian Brewington, Thomas Funkhouser, and Caroline Pantofaru. 2020. Virtual multi-view fusion for 3d semantic segmentation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIV 16*. Springer, 518–535.

[17] Xin Lai, Yukang Chen, Fanbin Lu, Jianhui Liu, and Jiaya Jia. 2023. Spherical transformer for lidar-based 3d recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 17545–17555.

[18] Xin Lai, Jianhui Liu, Li Jiang, Liwei Wang, Hengshuang Zhao, Shu Liu, Xiaojuan Qi, and Jiaya Jia. 2022. Stratified transformer for 3d point cloud segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 8500–8509.

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

[20] Guangrui Li. 2024. Construct to Associate: Cooperative Context Learning for Domain Adaptive Point Cloud Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 27917–27926.

[21] Guangrui Li, Guoliang Kang, Xiaohan Wang, Yunchao Wei, and Yi Yang. 2023. Adversarially masking synthetic to mimic real: Adaptive noise injection for point cloud segmentation adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 20464–20474.

[22] Jiale Li, Hang Dai, Hao Han, and Yong Ding. 2023. Mseg3d: Multi-modal 3d semantic segmentation for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 21694–21704.

[23] Rong Li, Shijie Li, Xieyuanli Chen, Teli Ma, Juergen Gall, and Junwei Liang. 2024. Tfnet: Exploiting temporal cues for fast and accurate lidar semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4547–4556.

[24] Shoutong Luo, Zhengxing Sun, Yi Wang, Yunhan Sun, and Chendi Zhu. 2024. LDCNet: Long-Distance Context Modeling for Large-Scale 3D Point Cloud Scene Semantic Segmentation. In *Proceedings of the 32nd ACM International Conference on Multimedia*. 1321–1330.

[25] Mengyao Lyu, Tianxiang Hao, Xinhao Xu, Hui Chen, Zijia Lin, Jungong Han, and Guiguang Ding. 2025. Learn from the learnt: Source-free active domain adaptation via contrastive sampling and visual persistence. In *European Conference on Computer Vision*. Springer, 228–246.

[26] Alexey Nekrasov, Jonas Schult, Or Litany, Bastian Leibe, and Francis Engelmann. 2021. Mix3d: Out-of-context data augmentation for 3d scenes. In *2021 international conference on 3d vision (3dv)*. IEEE, 116–125.

[27] Munan Ning, Donghuan Lu, Dong Wei, Cheng Bian, Chenglang Yuan, Shuang Yu, Kai Ma, and Yefeng Zheng. 2021. Multi-anchor active domain adaptation for semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 9112–9122.

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

[29] Jinyoung Park, Sanghyeok Lee, Sihyeon Kim, Yunyang Xiong, and Hyunwoo J Kim. 2023. Self-positioning point-based transformer for point cloud understanding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 21814–21823.

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

[31] Junkun Peng, Mingjie Sun, Eng Gee Lim, Qiufeng Wang, and Jimin Xiao. 2024. Prototype Guided Pseudo Labeling and Perturbation-based Active Learning for domain adaptive semantic segmentation. *Pattern Recognition* 148 (2024), 110203.

[32] Gilles Puy, Alexandre Boulch, and Renaud Marlet. 2023. Using a waffle iron for automotive point cloud semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 3379–3389.

[33] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. 2017. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 652–660.

[34] Guocheng Qian, Yuchen Li, Houwen Peng, Jinjie Mai, Hasan Hammoud, Mohamed Elhoseiny, and Bernard Ghanem. 2022. Pointnext: Revisiting pointnet++ with improved training and scaling strategies. *Advances in neural information processing systems* 35 (2022), 23192–23204.

[35] Dan Roth and Kevin Small. 2006. Margin-based active learning for structured output spaces. In *Machine Learning: ECML 2006: 17th European Conference on Machine Learning Berlin, Germany, September 18-22, 2006 Proceedings 17*. Springer, 413–424.

[36] Kuniaki Saito, Donghyun Kim, Stan Sclaroff, Trevor Darrell, and Kate Saenko. 2019. Semi-supervised domain adaptation via minimax entropy. In *Proceedings of the IEEE/CVF international conference on computer vision*. 8050–8058.

[37] Cristiano Saltori, Fabio Galasso, Giuseppe Fiameni, Nicu Sebe, Fabio Poiesi, and Elisa Ricci. 2023. Compositional semantic mix for domain adaptation in point cloud segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023).

[38] Cristiano Saltori, Fabio Galasso, Giuseppe Fiameni, Nicu Sebe, Fabio Poiesi, and Elisa Ricci. 2023. Compositional semantic mix for domain adaptation in point cloud segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 12 (2023), 14234–14247.

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

[40] Cristiano Saltori, Aljosa Osep, Elisa Ricci, and Laura Leal-Taixé. 2023. Walking your lidog: A journey through multiple domains for lidar semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 196–206.

[41] Jong-Chyi Su, Yi-Hsuan Tsai, Kihyuk Sohn, Buyu Liu, Subhransu Maji, and Manmohan Chandraker. 2020. Active adversarial domain adaptation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 739–748.

[42] Haotian Tang, Zhijian Liu, Shengyu Zhao, Yujun Lin, Ji Lin, Hanrui Wang, and Song Han. 2020. Searching efficient 3d architectures with sparse point-voxel convolution. In *European conference on computer vision*. Springer, 685–702.

[43] Hugues Thomas, Ben Agro, Mona Gridseth, Jian Zhang, and Timothy D Barfoot. 2021. Self-supervised learning of lidar segmentation for autonomous indoor navigation. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 14047–14053.

[44] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. 2017. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 7167–7176.

[45] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Pérez. 2019. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2517–2526.

[46] Dan Wang and Yi Shang. 2014. A new active labeling method for deep learning. In *2014 International joint conference on neural networks (IJCNN)*. IEEE, 112–119.

[47] Gang Wang, Wenlong Li, Cheng Jiang, Dahu Zhu, Zhongwei Li, Wei Xu, Huan Zhao, and Han Ding. 2021. Trajectory planning and optimization for robotic machining based on measured point cloud. *IEEE transactions on robotics* 38, 3 (2021), 1621–1637.

[48] Yan Wang, Xiangyu Chen, Yurong You, Li Erran Li, Bharath Hariharan, Mark Campbell, Kilian Q Weinberger, and Wei-Lun Chao. 2020. Train in germany, test in the usa: Making 3d object detectors generalize. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11713–11723.

[49] Barry Payne Welford. 1962. Note on a method for calculating corrected sums of squares and products. *Technometrics* 4, 3 (1962), 419–420.

[50] Lu Wen, Yuanyuan Xu, Zhenghao Feng, Jiliu Zhou, Luping Zhou, and Yan Wang. 2024. Semi-supervised Domain Adaptation for Semantic Segmentation via Active Learning with Feature-and Semantic-Level Alignments. *IEEE Transactions on Intelligent Vehicles* (2024).

[51] Bichen Wu, Alvin Wan, Xiangyu Yue, and Kurt Keutzer. 2018. 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)*. IEEE, 1887–1893.

[52] Bichen Wu, Xuanyu Zhou, Sicheng Zhao, Xiangyu Yue, and Kurt Keutzer. 2019. 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)*. IEEE, 4376–4382.

[53] Xiaopei Wu, Yuenan Hou, Xiaoshui Huang, Binbin Lin, Tong He, Xinge Zhu, Yuexin Ma, Boxi Wu, Haifeng Liu, Deng Cai, et al. 2024. TASeg: Temporal Aggregation Network for LiDAR Semantic Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 15311–15320.

[54] Peng Xiang, Xin Wen, Yu-Shen Liu, Hui Zhang, Yi Fang, and Zhizhong Han. 2023. Retro-fpn: Retrospective feature pyramid network for point cloud semantic segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*. 17826–17838.

[55] Aoran Xiao, Dayan Guan, Xiaoqin Zhang, and Shijian Lu. 2023. Domain Adaptive LiDAR Point Cloud Segmentation with 3D Spatial Consistency. *IEEE Transactions on Multimedia* (2023).

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

[57] Aoran Xiao, Jiaxing Huang, Dayan Guan, Fangneng Zhan, and Shijian Lu. 2022. Transfer learning from synthetic to real lidar point cloud for semantic segmentation. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 36. 2795–2803.

[58] Binhui Xie, Shuang Li, Qingju Guo, Chi Liu, and Xinjing Cheng. 2023. Annotator: A generic active learning baseline for lidar semantic segmentation. *Advances in Neural Information Processing Systems* 36 (2023).

[59] Binhui Xie, Shuang Li, Qingju Guo, Chi Liu, and Xinjing Cheng. 2023. Annotator: A generic active learning baseline for lidar semantic segmentation. *Advances in Neural Information Processing Systems* 36 (2023).

[60] Binhui Xie, Longhui Yuan, Shuang Li, Chi Harold Liu, and Xinjing Cheng. 2022. Towards fewer annotations: Active learning via region impurity and prediction uncertainty for domain adaptive semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 8068–8078.

[61] Binhui Xie, Longhui Yuan, Shuang Li, Chi Harold Liu, Xinjing Cheng, and Guoren Wang. 2022. Active learning for domain adaptation: An energy-based approach. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 36. 8708–8716.

[62] Jianyun Xu, Ruixiang Zhang, Jian Dou, Yushi Zhu, Jie Sun, and Shiliang Pu. 2021. Rpvnet: A deep and efficient range-point-voxel fusion network for lidar point cloud segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*. 16024–16033.

[63] Xu Yan, Jiantao Gao, Chaoda Zheng, Chao Zheng, Ruimao Zhang, Shuguang Cui, and Zhen Li. 2022. 2dpass: 2d priors assisted semantic segmentation on lidar point clouds. In *European Conference on Computer Vision*. Springer, 677–695.

[64] Xu Yan, Chaoda Zheng, Zhen Li, Sheng Wang, and Shuguang Cui. 2020. Pointasnl: Robust point clouds processing using nonlocal neural networks with adaptive sampling. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5589–5598.

[65] Li Yi, Boqing Gong, and Thomas Funkhouser. 2021. 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*. 15363–15373.

[66] Fukun Yin, Zilong Huang, Tao Chen, Guozhong Luo, Gang Yu, and Bin Fu. 2023. Dcnet: Large-scale point cloud semantic segmentation with discriminative and efficient feature aggregation. *IEEE Transactions on Circuits and Systems for Video Technology* 33, 8 (2023), 4083–4095.

[67] Zhimin Yuan, Wankang Zeng, Yanfei Su, Weiquan Liu, Ming Cheng, Yulan Guo, and Cheng Wang. 2024. Density-guided Translator Boosts Synthetic-to-Real Unsupervised Domain Adaptive Segmentation of 3D Point Clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 23303–23312.

[68] Hao Zhang and Ruimao Zhang. 2022. Active domain adaptation with multi-level contrastive units for semantic segmentation. In *Proceedings of the Asian Conference on Computer Vision*. 1640–1657.

[69] Haimei Zhao, Jing Zhang, Zhuo Chen, Shanshan Zhao, and Dacheng Tao. 2024. Unimix: Towards domain adaptive and generalizable lidar semantic segmentation in adverse weather. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14781–14791.

[70] Xinge Zhu, Hui Zhou, Tai Wang, Fangzhou Hong, Yuexin Ma, Wei Li, Hongsheng Li, and Dahua Lin. 2021. Cylindrical and asymmetrical 3d convolution networks for lidar segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 9939–9948.

[71] Zhuangwei Zhuang, Rong Li, Kui Jia, Qicheng Wang, Yuanqing Li, and Mingkui Tan. 2021. Perception-aware multi-sensor fusion for 3d lidar semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 16280–16290.

[72] Yang Zou, Zhiding Yu, Xiaofeng Liu, BVK Kumar, and Jinsong Wang. 2019. Confidence regularized self-training. In *Proceedings of the IEEE/CVF international conference on computer vision*. 5982–5991.