Single Domain Generalization for LiDAR Semantic Segmentation

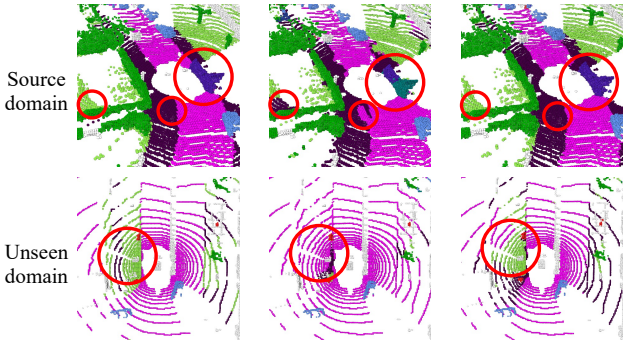
Hyeonseong Kim*, Yoonsu Kang*, Changgyoon Oh, and Kuk-Jin Yoon
Visual Intelligence Lab., KAIST, Korea
{brian617,gzgzys9887,changgyoon,kjyoon}@kaist.ac.kr

Abstract

With the success of the 3D deep learning models, various perception technologies for autonomous driving have been developed in the LiDAR domain. While these models perform well in the trained source domain, they struggle in unseen domains with a domain gap. In this paper, we propose a single domain generalization method for LiDAR semantic segmentation (DGLSS) that aims to ensure good performance not only in the source domain but also in the unseen domain by learning only on the source domain. We mainly focus on generalizing from a dense source domain and target the domain shift from different LiDAR sensor configurations and scene distributions. To this end, we augment the domain to simulate the unseen domains by randomly subsampling the LiDAR scans. With the augmented domain, we introduce two constraints for generalizable representation learning: sparsity invariant feature consistency (SIFC) and semantic correlation consistency (SCC). The SIFC aligns sparse internal features of the source domain with the augmented domain based on the feature affinity. For SCC, we constrain the correlation between class prototypes to be similar for every LiDAR scan. We also establish a standardized training and evaluation setting for DGLSS. With the proposed evaluation setting, our method showed improved performance in the unseen domains compared to other baselines. Even without access to the target domain, our method performed better than the domain adaptation method. The code is available at https://github.com/gzgzys9887/DGLSS.

1. Introduction

Understanding the surrounding scene is essential for autonomous driving systems. Using LiDAR sensors for perception has recently gained popularity from their ability to provide accurate distance information. Among such tasks, LiDAR semantic segmentation (LSS) predicts point-wise semantic labels from a single sweep of LiDAR data. Together with the recent development of 3D point cloud deep-learning models and the release of several real-world 3D annotated datasets [4, 7, 53, 64], numerous research on LiDAR semantic segmentation have emerged lately. However, since these models do not consider the data distribution differences between the train and test domains during the learning process, severe performance deterioration occurs when applied to real-world applications where domain gaps exist.



(a) Ground Truth (b) Domain Adaptation (c) Ours

Figure 1. Segmentation results on the source (SemanticKITTI [4]) and unseen (nuScenes-lidarseg [7]) domains. The results are from (a) ground truth, (b) domain adaptation method [60] adapted to Waymo [64] dataset, and (c) ours. Our method predicts successfully in both source and unseen domains, while the domain adaptation method struggles in both domains.

Two main domain gaps arise for real-scene LiDAR datasets: 1) differences in sparsity due to different sensor configurations and 2) differences in scene distribution. Depending on the type of LiDAR sensor, the total number of beams, vertical field of view (FOV), and vertical and horizontal resolutions differ, which leads to different sampling patterns. SemanticKITTI [4] dataset used a 64-beam LiDAR sensor, but a 32-beam sensor was used in nuScenes-lidarseg [7], and for SemanticPOSS [53], a 40-beam sensor was used. Waymo [64] dataset also used a 64-beam LiDAR, but there is a sparsity difference because the vertical resolution differs from SemanticKITTI. Table 1 summarizes configuration details of LiDAR sensors used in SemanticKITTI, nuScenes-lidarseg, Waymo, and SemanticPOSS. In addition, each dataset was acquired from different

*The first two authors contributed equally. In alphabetical order.

locations, so the scene structure as well as category distribution can vary enormously. For example, SemanticKITTI mainly contains suburban areas and therefore has a high number of cars, roads, and vegetation. Whereas nuScenes-lidarseg and Waymo datasets were acquired from dense urban environments and contain crowded pedestrian scenes, downtown, and also residential areas. Fig. 1 shows the differences in sparsity and scene between different datasets.

Recently, unsupervised domain adaptation (UDA) methods for LiDAR point clouds [1, 60, 72, 76, 81] have been proposed to mitigate performance degradation in a specific target domain. While UDA methods perform well in the target domain, they cannot guarantee high performance in unseen domains, which is critical for safe driving systems. As shown in Fig. 1(b), the UDA method [60] trained on the source domain (SemanticKITTI) and adapted to the target domain (Waymo) has low performance in the unseen domain (nuScenes-lidarseg). Also, the cumbersome process of acquiring new data and retraining is required to apply UDA to a new target domain. These problems can be tackled by building domain generalizable LSS models that guarantee performance on unseen domains and robustness against domain gaps. Among previous works, [76] did test its UDA methodology in a domain generalization (DG) setting. However, the method was not specifically designed for DG and was deemed impractical for real-world applications due to its limited evaluation with 2 classes. All of these circumstances further highlight the importance of developing methodologies that primarily aim at domain generalization for LiDAR semantic segmentation.

In this paper, we propose a domain generalization approach for LiDAR semantic segmentation (DGLSS), which aims to achieve high performance in unseen domains while training the segmentation model only on source domains. In particular, we propose a representation learning approach for DGLSS by focusing on differences in sparsity and scene distribution. Since obtaining multiple fully-labeled LiDAR datasets as source domains for training is cost-expensive, we choose to perform generalization using a single source domain. Also, as using sparser LiDAR data in an application is more efficient and a more likely scenario for autonomous driving, we focus on learning from a denser source domain. To this end, we first simulate the unseen domain during the learning process while considering the characteristics of the actual LiDAR sensor. We augment the domain by randomly subsampling LiDAR beams from the LiDAR scan of the source domain at every iteration of the learning process. With the augmented sparse domain, we introduce two constraints for generalizable representation learning: *sparsity invariant feature consistency (SIFC)* and *semantic correlation consistency (SCC)*. The purpose of SIFC is to align the internal sparse features of the source domain with the augmented domain based on the feature affinity. For SCC, we build scene-wise class prototypes and constrain the correlation between class prototypes to be similar for every LiDAR scene regardless of domain. Learning with the proposed constraints, the model can generalize well on unseen domains that have different sparsity and scene distribution compared to the source domain.

Table 1. Configuration of sensors used to acquire LiDAR datasets.

Dataset	LiDAR beams	vertical FOV(°)	vertical res(°)	horizontal res(°)	range (m)
SemanticKITTI [4]	64	[-23.6, 3.2]	0.4	0.08	120
nuScenes-lidarseg [7]	32	[-30, 10]	1.33	0.1–0.4	70
Waymo [64]	64	[-17.6, 2.4]	0.31	0.16	75
SemanticPOSS [53]	40	[-16, 7]	0.33/1	0.2	200

In addition, we build standardized training and evaluation settings for DGLSS, as such setup for domain generalization has been absent. For this, we employ 4 real-world LiDAR datasets [4, 7, 53, 64] and carefully select 10 common classes existing in datasets. Then, we remap the labels of each dataset to the common labels. Furthermore, we implement baseline methods using general-purpose DG methods applicable to LiDAR point cloud [38, 51] and evaluate them in the proposed evaluation setting. With the proposed evaluation setting, our method showed good performance in the unseen domains compared to the baselines. Even without access to the target domain, our method showed comparable performance to the domain adaptation method as shown in Fig. 1(c).

In summary, our contributions are as follows:

- To the best of our knowledge, we propose the first approach that primarily aims at domain generalization for LiDAR semantic segmentation (DGLSS).
- We build standardized training and evaluation settings for DGLSS and implement several DG baseline methods applicable to the point cloud domain.
- We propose *sparsity invariant feature consistency* and *semantic correlation consistency* for generalizable representation learning for DGLSS, which can effectively deal with domain gaps of LiDAR point clouds.
- Extensive experiments show that our approach outperforms both UDA and DG baselines.

2. Related Work

2.1. LiDAR Semantic Segmentation

LiDAR semantic segmentation (LSS) aims to assign semantic labels to each point in the LiDAR point cloud. Unlike 2D images, LiDAR point clouds are irregular, unordered, and have non-uniform sparsity, making it difficult to apply traditional 2D deep learning approaches. LSS approaches can be categorized into point-based, projection-based, and voxel-based methods, depending on how they represent point clouds. Point-based methods [25, 43, 57, 66]

directly operate on points by using MLP following the pioneer PointNet [56]. Although they have a small number of parameters and low information loss, they require heavy computation and memory to apply to large-scale LiDAR data. For efficiency and the use of advanced 2D CNN architectures, projection-based methods have been proposed, projecting points to range-view image using spherical projection [2, 11, 15, 19, 48, 71, 72] or bird-eye-view [79]. However, these methods might lose some geometrical information during projection. Recent voxel-based methods [12, 65, 82], instead of using dense 3D CNN that requires heavy computation and memory, represent point cloud as sparse 3D voxels and use sparse convolution [14, 23] with high performance and less computation. We also use MinkowskiNet [14] as our baseline to take advantage of the high performance and efficiency of sparse convolution.

2.2. Unsupervised Domain Adaptation for LiDAR Semantic Segmentation

In an attempt to reduce the domain discrepancy between the source and target domains, unsupervised domain adaptation (UDA) approaches for LSS use both labeled source data and unlabeled target data. [32, 72, 75] learn domain-invariant features by reducing the feature difference between the source and target domain and creating a shared feature space. Also, domain-mapping methods [1, 36, 59, 73, 81] aim at transforming the source data to resemble the appearance of target data and reduce the physical difference. Others construct intermediate domains with common representations by mixing portions of source and target domains [35, 60], or by utilizing completion to restore the underlying 3D surface invariant to sensor differences [76]. Still, UDA methods cannot ensure high performance for domains other than the specified target domain, highlighting the necessity of DG approaches. [76] did test their UDA method in a DG setting but with only 2 classes (pedestrian and car), whereas we directly target DG by proposing an experimental setup more suitable for DGLSS using 10 classes.

2.3. Domain Generalization

Domain generalization (DG) [5, 50] aims to achieve high performance in unseen domains that are not used in the learning process by learning only on source domains. For 2D computer vision applications, DG has been widely studied in digit and object recognition [21, 38–41, 44, 61, 84], semantic segmentation [13, 51, 58, 68, 69, 77], and medical imaging [42, 45, 46, 78]. For DG using multiple sources, aligning source domain distributions [18, 20, 41, 47, 49, 80], disentangling domain-invariant and -specific features [9, 29, 33, 39, 55], domain augmentation [8, 61, 83–85], and meta-learning-based methods [3, 16, 17, 38, 40, 44, 45] are proposed. As using only a single source to learn domain invariant representations is difficult, many single-source DG methods augment additional domains by employing domain-specifically designed image transformation [10, 68, 78], task adversarial gradient [58, 62, 69], random augmentation network [74], and style transfer network [6, 63, 77]. Recently, in semantic segmentation, normalization and whitening methods have received attention and achieved satisfactory performance [13, 28, 51, 52, 54]. However, domain generalization on 3D point cloud [26, 27], especially for LiDAR point cloud [37], is still an open research problem. This is mainly because the domain gap of 3D point clouds is different from that of 2D images, e.g. domain discrepancies of images come from different viewpoints, appearance, color, or brightness, and domain gaps in point clouds stem from different sparsity, object shape, or occlusion. In this paper, we propose a domain generalization method for LSS, and implement baselines adopted from generally applicable DG approaches [38, 51] to the point cloud domain.

3. Proposed Method

3.1. Overview

Our goal is to learn a LiDAR semantic segmentation model that performs well on unseen domains when trained only on a single source domain. We have a single source domain $D^s = \{(P_i^s, y_i^s)\}_{i=1}^N$ for training, which consists of LiDAR point cloud P_i^s and its corresponding point-wise labels y_i^s where N is the number of scans in the source domain. We first augment the source domain by manipulating the sparsity of the source domain and obtain the augmented domain $D^a = \{(P_i^a, y_i^a)\}_{i=1}^N$ (Sec. 3.2). For given source and augmented domains, both point clouds and labels are voxelized into $\{V_i^s, \tilde{y}_i^s\}_{i=1}^N$ and $\{V_i^a, \tilde{y}_i^a\}_{i=1}^N$, where V_i^* and $\tilde{y}_i^*$ denote voxels and corresponding labels. Then, the model consisting of encoder Φ_{enc}, decoder Φ_{dec}, and classifier $\mathcal{C}$, receives voxels and predicts a semantic class for each voxel, $\hat{y}_i^s$ and $\hat{y}_i^a$. We propose two constraints SIFC (Sec. 3.3) and SCC (Sec. 3.4) to encourage generalizable representation learning. The semantic segmentation performance is evaluated on unseen target domains as well as the source domain. Fig. 2 shows the proposed DGLSS framework, and each part will be explained in detail in the following sections.

3.2. Augmentation

Learning domain-invariant features with a single source domain is difficult, which is why additional augmented domains are needed. To achieve high performance even in unseen domains with unknown sparsity, we focus on one of the main domain gaps in LiDAR data, sparsity difference, and augment domains with diverse sparsity. To this end, scan lines in the source domain are randomly dropped instead of individual points to simulate the local structures of

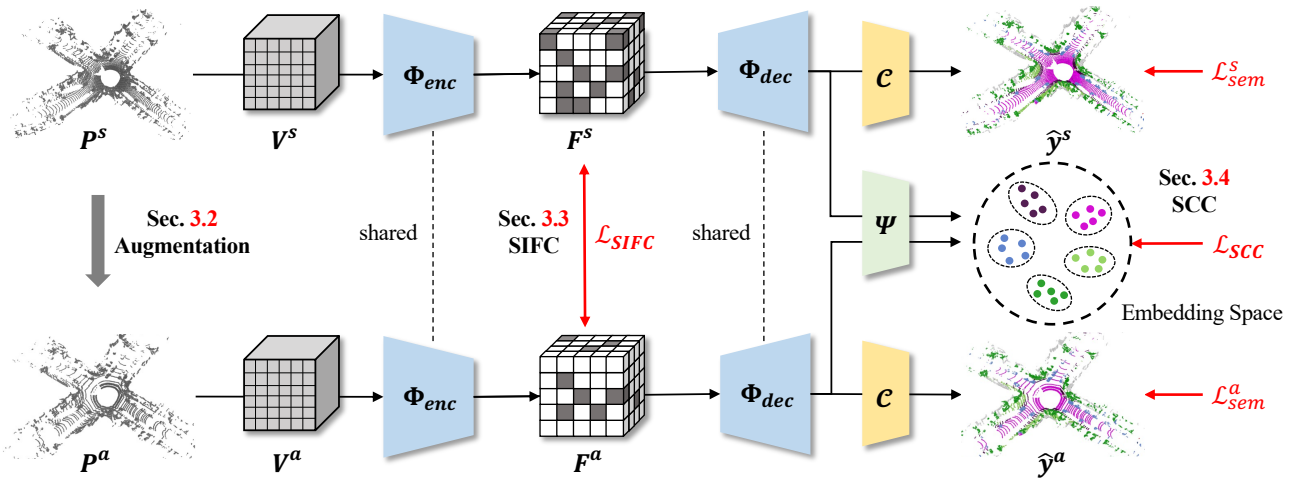


Figure 2. Overall DGLSS framework. From the single source domain P^s, we augment a new domain with different sparsity P^a to learn sparsity-invariant representations. The two domains are encoded by Φ_{enc}, and the encoded internal features F^s and F^a are constrained with the proposed *sparsity invariant feature consistency* (SIFC). The decoded features from Φ_{dec} are fed to the metric learner Ψ to construct a feature embedding space. The embedding space is constrained by the proposed *semantic correlation consistency* (SCC). The semantic class predictions $\hat{y}^s$ and $\hat{y}^a$ are supervised by the semantic segmentation ground truth for both source and augmented domains.

real LiDAR data [22, 24, 31, 70, 76]. For given P^s_i and y^s_i, we first obtain the point-wise distance from the sensor r_i, azimuth θ_i, and altitude φ_i. Then, we project P^s_i into a range-view image as in [48] with Eq. 1 whose height corresponds to the number of LiDAR sensor beams:

$$\begin{pmatrix} r^x_i \\ r^y_i \end{pmatrix} = \begin{pmatrix} \frac{1}{2}[1-\theta_i \pi^{-1}]W \\ [1-(\varphi_i + f_{max})f^{-1}]H \end{pmatrix}, \tag{1}$$

where r^x_i, r^y_i denote the pixels of the range-view image, f_{max}, f denote the upper bound FOV and FOV itself, and H, W are the height and width of the image. After randomly selecting a ratio p between $[p_{min}, p_{max}]$ where p_{min} and p_{max} are the minimum and maximum probability for dropping a certain beam, a total of $p \times H$ beams are randomly selected for dropping. The points and labels corresponding to the selected beams are then removed. From this process, we can obtain a new augmented domain $D^a = \{(P^a_i, y^a_i)\}^N_{i=1}$ with diverse sparsity levels. This augmentation process is performed on-the-fly for every training iteration, and the augmented data are only used for representation learning and are not kept.

3.3. Sparsity Invariant Feature Consistency

To better generalize across multiple domains with different sparsity, we let the model learn the sparsity-invariant features. In particular, we propose sparsity invariant feature consistency (SIFC) to maintain consistency between encoded sparse features of source and augmented domains with the same scene but different sparsity. Given P^s_i and P^a_i from source and augmented domains, we voxelize them to obtain sparse voxel representations V^s_i and V^a_i. Then, sparse voxels are fed into the encoder Φ_{enc} to obtain sparse voxel features $F^s_i = \Phi_{enc}(V^s_i) \in \mathbb{R}^{N^s_i \times d}$ and $F^a_i = \Phi_{enc}(V^a_i) \in \mathbb{R}^{N^a_i \times d}$, where N^s_i and N^a_i denote the number of occupied voxels of F^s_i and F^a_i, and d is the dimension of voxel features. Our goal is to impose SIFC to enforce the F^s_i and F^a_i to be consistent. However, because the sparsity of the source and augmented domains are different, $N^s_i \geq N^a_i$, not all voxels in the source domain may have corresponding voxel pairs in the same spatial locations within the augmented domain. Therefore, we deal with F^s_i in a different manner depending on whether they have a corresponding pair in F^a_i.

We first find voxel features in F^s_i that have corresponding paired voxel features of the same spatial locations in F^a_i, and denote them as $F^s_{i,p}$. The voxels that do not have corresponding pairs are $F^s_{i,n} = F^s_i \backslash F^s_{i,p}$. The paired voxel features $F^s_{i,p}$ and F^a_i are self-supervised with L1 loss, which allows the sparse voxel features in the same spatial locations for both source and augmented domains to be similar. To impose supervision for $F^s_{i,n}$, we aggregate neighboring features from the augmented domain based on the feature affinity in the source domain, as shown in Fig. 3. For voxel feature $f^s \in F^s_{i,n}$ with voxel coordinate x^s, we compute the coordinate distances between f^s and $F^s_{i,p}$. We find the k nearest voxel features $\{f^s_j\}^k_{j=1}$ with coordinates $\{x^s_j\}^k_{j=1}$ in $F^s_{i,p}$, whose corresponding paired voxel features and coordinates in F^a_i are $\{f^a_j\}^k_{j=1}$ and $\{x^a_j\}^k_{j=1}$. Then, we aggregate $\{f^a_j\}^k_{j=1}$ by a weighted sum where the weights are the reciprocal of the coordinate distances. However, using all k neighboring features for aggregation may adversely affect the learning of precise decision boundaries by allow-

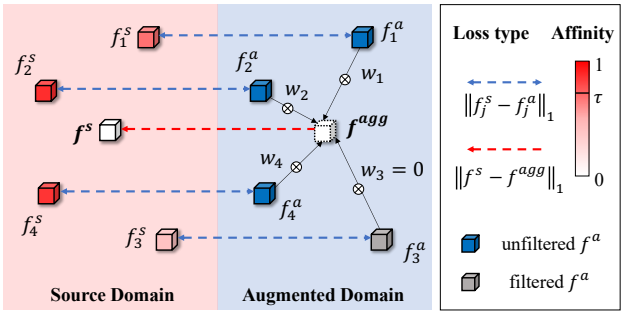


Figure 3. Illustration of SIFC loss. The blue arrow shows the self-supervision of sparse voxel features from the source f_j^s and augmented domain f_j^a with the same coordinates. The red arrow shows the self-supervision between source domain voxel feature f^s and aggregated voxel feature f^{agg} in the augmented domain.

ing the propagation of irrelevant class features. Therefore, we further filter out the neighboring features by zeroing out the features that have affinity lower than the threshold τ. Specifically, the affinity is calculated in source domain between f^s and $\{f_j^s\}_{j=1}^k$ using cosine similarity. The aggregated voxel feature $f^{agg} \in F_i^{agg}$ is calculated as Eq. 2:

$$f^{agg} = \frac{\sum_{j=1}^{k} w_j f_j^a}{\sum_{j=1}^{k} w_j}, \qquad w_j = \begin{cases} 1/\|x^s - x_j^s\|_2, & \text{if } \frac{<f^s, f_j^s>}{\|f^s\|\|f_j^s\|} \geq \tau \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

In summary, the proposed SIFC loss $\mathcal{L}_{SIFC}$ is defined as Eq. 3:

$$\mathcal{L}_{SIFC} = \frac{1}{N} \sum_{i=1}^{N} \|F_{i,p}^s - F_i^a\|_1 + \|F_{i,n}^s - F_i^{agg}\|_1 \tag{3}$$

3.4. Semantic Correlation Consistency

Although SIFC allows the model to extract invariant features from point clouds with different sparsity, it might have difficulty when the differences in scene distribution becomes a domain gap. To provide an additional constraint, we propose a semantic correlation consistency (SCC), which enforces the correlation between class prototypes within each data to be similar. The motivation is that even though scenes may be different, the relationship between class semantic features should be maintained throughout domains. For example, cars may have a strong correlation with trucks as they both belong to automobile categories, and can also be related to road class because cars are located on top of the road in most cases. By learning these relationships through the proposed constraint, the model can generalize across different scenes.

In order to impose SCC, we first obtain class-specific prototypes for each LiDAR scan regardless of the domain. Let us consider a single batch of B scans, and denote the union of the encoded features as $\{F_i\}_{i=1}^{2B} = \{F_i^s\}_{i=1}^{B} \cup \{F_i^a\}_{i=1}^{B}$. Then, the encoded features F_i are fed into the decoder Φ_{dec} followed by the metric learner Ψ to obtain feature embeddings $\Psi(\Phi_{dec}(F_i)) \in \mathbb{R}^{M_i \times l}$, where M_i and l are the total number of features in $\Psi(\Phi_{dec}(F_i))$ and feature dimension. From the features embeddings, we obtain class-specific prototypes $z_{i,c} \in \mathbb{R}^l$ for class c following Eq. 4:

$$z_{i,c} = \frac{\sum_{j=1}^{M_i} \mathbb{1}_{\tilde{y}_{i,j}=c} \Psi(\Phi_{dec}(F_i))_j}{\sum_{j=1}^{M_i} \mathbb{1}_{\tilde{y}_{i,j}=c}}, \tag{4}$$

where $\Psi(\Phi_{dec}(F_i))_j$ and $\tilde{y}_{i,j}$ are the j^{th} feature embedding of $\Psi(\Phi_{dec}(F_i))$ and the corresponding j^{th} label in $\tilde{y}_i$.

From the above equation, we can obtain a prototype matrix $Z_i \in \mathbb{R}^{C \times l}$ where C is the total number of classes. Finally, the SCC loss $\mathcal{L}_{SCC}$ is defined as Eq. 5:

$$\mathcal{L}_{SCC} = \frac{1}{L} \sum_{i} \sum_{j \neq i} (Z_i Z_i^T - Z_j Z_j^T), \tag{5}$$

where L is the total number of valid combinations in $\{F_i\}_{i=1}^{2B}$. Note that we only consider class correlations whose number of voxels belonging to that class is nonzero.

3.5. Overall Loss Function

The final outputs of the model are voxel-wise semantic class predictions for source and augmented domain, $\hat{y}^s$ and $\hat{y}^a$. These are supervised by $\tilde{y}^s$ and $\tilde{y}^a$ with weighted cross-entropy loss [48] $\mathcal{L}_{sem}^s$ and $\mathcal{L}_{sem}^a$. The weights are calculated by the inverse of each class frequency within the source domain. In conclusion, the total loss function is the weighted sum of individual losses.

$$\mathcal{L}_{total} = \mathcal{L}_{sem}^s + \mathcal{L}_{sem}^a + \lambda_1 \mathcal{L}_{SIFC} + \lambda_2 \mathcal{L}_{SCC} \tag{6}$$

where λ_1 and λ_2 are the weights for $\mathcal{L}_{SIFC}$ and $\mathcal{L}_{SCC}$.

4. Dataset Setup

4.1. Datasets

We utilize four real-world LiDAR datasets for autonomous driving to train and evaluate the generalization performance of the LiDAR semantic segmentation model. **SemanticKITTI** [4] dataset adopts a 64-beam LiDAR and has 19 class labels. We follow the official sequence split and use 00-07, 09-10 scenes for training (19,130 frames) and scene 08 for validation (4,071 frames).
nuScenes-lidarseg [7] dataset uses a 32-beam LiDAR, consists of driving scenes from Singapore and Boston, and has 32 class labels. We train our model using 700 scenes for training (28,130 frames) and 150 scenes for validation (6,019 frames).

Waymo [64] dataset was acquired diversely from 3 cities with various weather conditions using a 64-beam LiDAR, and has 23 segmentation labels. Following the official recommendation, 798 training scenes (23,691 frames) and 202 validation scenes (5,976 frames) are used.

SemanticPOSS [53] dataset adopts a 40-beam LiDAR to capture campus scenes that have a large number of cars, riders, and pedestrians. It has 14 class labels, and we only use SemanticPOSS for evaluation (scene 03 with 501 frames) due to the small number of training scenes (2488 frames).

4.2. Label Mapping

As there has been no commonly shared setup for DGLSS until now, we propose standardized training and evaluation settings to further encourage future DGLSS research. We select 10 common classes that overlap among most datasets and include other minor classes: {*car, bicycle, motorcycle, truck, other-vehicle, pedestrian, drivable-surface, sidewalk, walkable, vegetation*}. We add a {*background*} class to map unused classes and ignore them. Then, we map the classes of each dataset into common classes. In SemanticKITTI, nuScenes-lidarseg, and Waymo datasets, 10 common classes were finally used. But in SemanticPOSS, we integrate {vegetation} and {walkable} according to the observation. Also, since there is no {motorcycle, truck, other-vehicle, sidewalk}, 5 classes {car, bicycle, pedestrian, drivable-surfaces, walkable} are used. Label mapping for each dataset is described in the *Supplementary material*.

5. Experiments

5.1. Experimental Setup

Metrics We evaluate the segmentation performance using the Intersection over the Union (IoU) for every class, and the mean IoU (mIoU) for each dataset. To evaluate the generalization performance over the source and unseen domains, we compute the arithmetic mean (AM) of mIoU over datasets, as well as the harmonic mean (HM).

Implementation Details We use MinkowskiNet [14], specifically MinkUNet34, as the encoder Φ_{enc} and decoder Φ_{dec}. The classifier $\mathcal{C}$ consists of a single linear layer. For the metric learner Ψ, we use 2-layer MLP with ReLU activation functions. We use Adam optimizer [34] with lr=1e-3, β_1=0.9, and β_2=0.999, and train with a batch size of 8. We also employ classical 3D data augmentation (*e.g.* arbitrary scaling, random rotation, flipping, translation) on the source domain data during training. Please refer to the *Supplementary material* for more implementation details.

5.2. Baselines

Base method uses the source domain for training and is directly evaluated on the target domains. We use the same model architecture as our proposed method except for the metric learner. Only $\mathcal{L}^s_{sem}$ is used for training.

Augment method additionally uses an augmented domain (Sec. 3.2) compared to the baseline and optimized using $\mathcal{L}^s_{sem} + \mathcal{L}^a_{sem}$.

IBN-Net [51] combines Instance Normalization (IN) [67] and Batch Normalization [30] to learn appearance invariant feature while maintaining content information. Following [51], we implement the IBN-b block by adding IN right after the second conv layer (conv2) and the second, third convolution group (conv3_x and conv4_x) of MinkUNet34.

MLDG [38] is a meta-learning method for DG that simulates the domain shift during training. Since we only have a single source domain, we randomly split each batch into meta-train and meta-test sets. We also try to split a batch by maximizing mean feature distances between meta-train and meta-test sets using a pre-trained baseline model. We denote the MLDG using random splitting and maximizing feature distances as MLDG(A) and MLDG(B), respectively.

CoSMIX [60] is a syn-to-real DA method, but the result of SemanticPOSS→ SemanticKITTI results in the paper indicates that the model is capable of real-to-real DA. We follow the overall setting of [60] for training. We refer CoSMIX(K), CoSMIX(N), and CoSMIX(W) as the method adapted to SemanticKITTI, nuScenes-lidarseg, and Waymo datasets, respectively.

For fair comparison, all the baselines use MinkUNet34 as the backbone. Please refer to the *Supplementary material* for the detailed implementations of baseline methods.

Table 2. Per-dataset mIoU(%) and AM(%) and HM(%) over all datasets compared with other DG and DA methods. The results are reported using a model trained on SemanticKITTI.

Method	K	N	W	P	AM	HM
Base	57.31	37.42	35.24	40.92	42.72	41.24
Augment	58.25	40.27	38.16	45.68	45.59	44.40
IBN-Net	57.74	38.74	36.99	43.11	44.15	42.85
MLDG(A)	56.26	36.77	35.39	37.41	41.46	40.02
MLDG(B)	54.71	40.26	36.39	41.30	43.16	42.19
Ours	**59.62**	**44.83**	**40.67**	45.09	**47.55**	**46.60**
CoSMIX(N)	49.98	43.25	38.05	**46.42**	44.42	43.98
CoSMIX(W)	49.35	38.94	39.46	43.89	42.91	42.52

Table 3. Per-dataset mIoU(%) and AM(%) and HM(%) over all datasets compared with other DG and DA methods. The results are reported using a model trained on Waymo.

Method	W	K	N	P	AM	HM
Base	75.37	49.40	47.83	51.13	55.93	54.07
Augment	**75.66**	50.66	**50.55**	52.30	57.29	55.66
IBN-Net	75.47	51.13	44.72	49.58	55.22	53.09
MLDG(A)	72.47	48.94	48.64	49.35	54.85	53.29
MLDG(B)	68.25	44.60	45.77	44.96	50.90	49.28
Ours	75.28	**51.23**	49.61	**54.28**	**57.60**	**56.04**
CoSMIX(N)	65.68	40.99	47.98	52.69	51.83	50.35
CoSMIX(K)	66.68	44.71	49.96	52.34	53.42	52.29

Table 4. Ablation study of SIFC and SCC with per-class IoU(%). The experiment is conducted using SemanticKITTI dataset as the source domain. Aug. denotes the use of sparsity augmentation and T denotes the evaluated target domain. The best value is **bolded** and the second best value is underlined. We omit the non-existent common classes in SemanticPOSS and mark them with '-'.

Aug.	SIFC	SCC	T	car	bicycle	motorcycle	truck	other vehicle	pedestrian	drivable surface	sidewalk	walkable	vegetation	mIoU
				91.23	10.04	35.69	52.89	37.95	40.99	83.86	62.78	66.34	91.33	57.31
✓				90.42	5.32	32.92	61.56	39.52	39.69	85.00	65.59	**70.28**	**92.24**	58.25
✓	✓		K	**92.83**	9.06	30.04	71.62	**47.17**	40.83	84.89	65.07	66.61	91.00	**59.91**
✓		✓		92.19	2.69	**36.15**	68.97	38.86	**44.25**	**86.79**	**67.05**	66.20	91.01	59.42
✓	✓	✓		92.65	**11.99**	27.09	**72.50**	45.95	36.39	84.76	65.64	67.98	91.28	59.62
				68.91	**2.51**	12.18	11.30	20.35	29.47	80.17	31.91	40.19	77.18	37.42
✓				77.07	1.03	16.93	23.79	19.44	30.68	76.75	29.92	**45.21**	81.83	40.27
✓	✓		N	76.79	1.41	**37.95**	**26.94**	19.24	36.08	81.07	36.97	44.48	**82.29**	44.32
✓		✓		**77.15**	1.09	26.23	19.87	19.04	24.64	**82.91**	34.56	42.29	78.91	40.67
✓	✓	✓		76.36	1.51	35.18	26.47	**25.49**	**37.09**	82.03	**38.12**	44.20	81.79	**44.83**
				72.12	2.52	4.52	7.77	13.36	40.86	64.92	30.12	34.84	81.40	35.24
✓				79.35	1.70	**10.98**	13.75	13.99	42.03	65.77	31.47	38.15	84.42	38.16
✓	✓		W	**83.55**	1.06	7.03	**20.06**	13.24	37.57	64.94	33.28	34.85	**85.86**	38.14
✓		✓		78.24	0.47	8.53	9.75	13.45	19.93	**72.62**	34.71	**38.89**	82.68	35.93
✓	✓	✓		82.26	**4.85**	9.72	16.80	**17.67**	**52.55**	68.20	**35.91**	33.33	85.41	**40.67**
				58.37	0.37	-	-	-	43.40	32.84	-	69.64	-	40.92
✓				58.19	1.04	-	-	-	50.55	**43.58**	-	**75.04**	-	**45.68**
✓	✓		P	62.35	0.36	-	-	-	47.60	27.88	-	73.35	-	42.31
✓		✓		**64.54**	0.23	-	-	-	36.88	42.65	-	73.75	-	43.61
✓	✓	✓		63.05	**1.68**	-	-	-	**52.14**	34.15	-	74.44	-	45.09

Table 5. Per-dataset mIoU(%) and AM(%) and HM(%) over all datasets. The results are from the models trained on **n**uScenes-lidarseg (**N**) and **S**ynLiDAR (**S**).

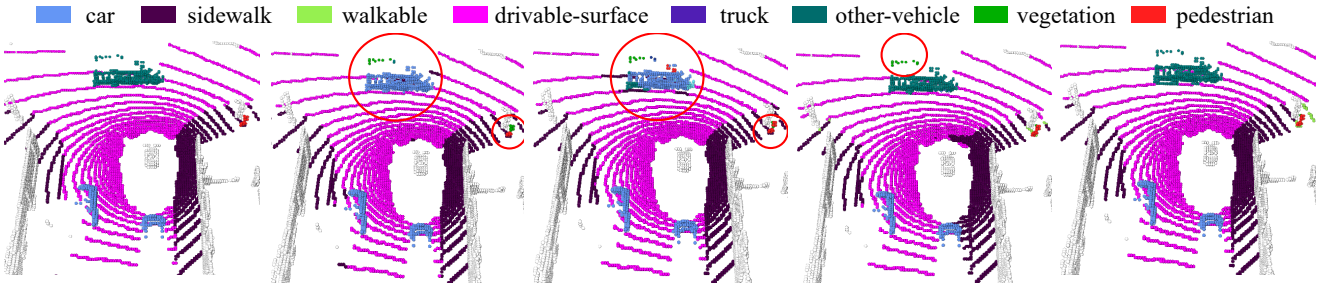
Source	Method	N	K	W	P	AM	HM
N	Base	65.78	36.24	38.65	38.51	44.79	42.26
	Augment	**66.97**	39.34	43.12	44.85	48.57	46.60
	IBN-Net	65.31	36.93	36.53	48.11	46.72	44.17
	MLDG(A)	61.23	32.70	36.33	35.96	41.56	39.12
	Ours	65.32	38.98	40.93	45.32	47.64	45.73
	Ours(w/o SIFC)	65.12	**39.39**	**46.56**	**45.87**	**49.23**	**47.60**
S	Base	29.78	27.37	27.76	37.58	30.62	30.13
	Ours	32.45	28.17	28.25	39.29	32.04	31.45
	CoSMIX(K)	36.68	**31.28**	**31.63**	**43.02**	**35.70**	**35.10**
	CoSMIX(W)	**37.93**	30.72	30.99	42.69	35.58	34.90

5.3. Quantitative Comparisons

In this section, we report the results of our method trained on various source datasets and compare them with the implemented baselines. For the comparison with [76], please refer to the *Supplementary material*. We abbreviate Semantic**K**ITTI, **n**uScenes-lidarseg, **W**aymo, and Semantic**P**OSS as **K**, **N**, **W**, and **P** in all the tables for brevity. As our main objective is to generalize from the denser domain, we focus on learning from a source domain with a higher number of beams, *i.e*. SemanticKITTI and Waymo datasets.

Table 2 and Table 3 show the results of our method and baselines trained on SemanticKITTI and Waymo as source datasets, respectively. Our method successfully shows improvement over the base method, where the gain is +4.84% AM and +5.36% HM using SemanticKITTI as the source domain, and +1.67% AM and +1.97% HM using Waymo as the source domain. The major improvement comes from the unseen domains, which supports the generalization strength of our method. Meanwhile, the Augment method does show better performance compared to other baselines, suggesting the importance of considering sparsity in LSS. Thanks to the proposed constraints, SIFC and SCC, our method gains additional performance improvement.

IBN-Net trained on the SemanticKITTI obtains better results than the base method and is comparable when trained on Waymo, due to the generalization capacity of instance normalization. Nevertheless, since the domain gaps of 2D images are different from that of point clouds and the concept of appearance feature is ambiguous in the LiDAR domain, the performance improvement from learning appearance invariant features may not be sufficient. MLDG(B) has higher mIoU than the base method and MLDG(A) for most datasets when trained on the SemanticKITTI dataset, as the meta-learning framework is able to learn from more distinguished meta-train and meta-test sets. Unfortunately, meta-learning schemes trained on the Waymo dataset perform worse than the baseline, and even MLDG(B) is worse than MLDG(A). It seems that only using a simple meta-learning framework has difficulty learning from the vast and enormously diverse data distribution of the Waymo dataset in a stable manner. We also report the results of DA method [60]. Impressively, our method surpasses the DA method in both training settings. This shows that our method is effective in preserving domain-invariant information that is applicable to other domains.



(a) Ground Truth (b) Base (c) IBN-Net [51] (d) CoSMIX(N) [60] (e) Ours

Figure 4. Qualitative comparison between the proposed method and the baseline methods on nuScenes-lidarseg dataset. The circles shown in the figure indicate the mislabeled parts.

We additionally test our approach in a challenging setting of sparse-to-dense generalization, using the nuScenes-lidarseg as the source domain. As shown in Table 5, the augment method outperforms IBN-Net, MLDG(A), and our method. The result shows the limitation of aligning to severely sparse data using SIFC. Still, using SCC without SIFC achieves the highest performance, indicating that SCC can improve generalizability in sparse-to-dense situations by considering scene distributions.

We also report the results of syn-to-real generalization when trained on SynLiDAR [73], a synthetic dataset with 64 beams. As there is no official training and validation split, we choose scenes 05 and 10 for validation out of the 13 scenes and the others for training. In Table 5, we can see that our method performs better than the base method and has the potential to generalize from a synthetic domain, but is still worse than the DA method. This implies that additional domain gaps need to be addressed such as unrealistic sampling patterns from synthetic environments that do not exist in real LiDAR data.

5.4. Ablation Study

To demonstrate the strength of the proposed constraints, we conduct ablation experiments using the SemanticKITTI dataset as the source domain. In Table 4, we report the per-class IoU and mIoU each time an element is added. Utilizing an additional augmented domain increases the overall segmentation performance for all datasets compared to the base method. When applying SIFC with the augmented domain, the performance enhancement is best seen in large objects (*i.e.* car and truck) whose appearance is largely affected by sparsity, and also when the sparsity gap with the target domain is large (*i.e.* nuScenes-lidarseg). Moreover, adding SCC on top of SIFC further increased the mIoU in unseen target domains. Particularly, we observed improvements in bicycle, other-vehicle, pedestrian, drivable-surface, and sidewalk classes for every unseen domain. However, in the case of applying SCC without SIFC, maintaining the inter-class correlation does enhance the classes that are less affected by sparsity difference (drivable-surface and sidewalk), but has an adverse effect on vehicles. This implies that SIFC plays a key role in leveraging the potential of SCC and boosting its generalization ability in addition to sparsity invariant feature learning. For the hyperparameter analysis, please refer to the *Supplementary material*.

5.5. Qualitative Comparisons

In Fig. 4, we visualize the qualitative results of our method and baselines on the nuScenes-lidarseg dataset when using SemanticKITTI as the source domain. The red circles show the wrongly segmented parts from the baseline results. The base method and IBN-Net confuse the other-vehicle with the car and a pedestrian with vegetation. CoSMIX adapted to the nuScenes-lidarseg mostly gives correct predictions with slight errors due to the adaptation process. Thanks to the proposed SIFC and SCC, our method correctly segments both classes. Please refer to the *Supplementary material* for more comparative results.

6. Conclusion

Existing LiDAR semantic segmentation methods for autonomous driving usually suffer performance degradation in the presence of domain gaps where LiDAR sensor configuration or driving scene changes. In this paper, we propose a novel domain generalization approach for LiDAR semantic segmentation (DGLSS) to ensure performance in unseen domains as well as a source domain. To generalize well in presence of diverse sparsity and scenes with the single source domain, we constrain the model with sparsity invariant feature consistency (SIFC) and semantic correlation consistency (SCC). For training and evaluating the model, we also introduce a standardized setting and implement baselines for DGLSS using four real-world LiDAR datasets. As a result, our method achieved improved performance in the unseen domains compared to other baselines including the domain adaptation method.

Acknowledgements This work was supported by Institute of Information & Communications Technology Planning & Evaluation(IITP) grant funded by Korea government(MSIT) (No.2020-0-00440) and the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (NRF-2022R1A2B5B03002636).

References

[1] Inigo Alonso, Luis Riazuelo Montesano, Ana C Murillo, et al. Domain adaptation in lidar semantic segmentation by aligning class distributions. *arXiv preprint arXiv:2010.12239*, 2020. 2, 3

[2] Inigo Alonso, Luis Riazuelo, Luis Montesano, and Ana C Murillo. 3d-mininet: Learning a 2d representation from point clouds for fast and efficient 3d lidar semantic segmentation. *IEEE Robotics and Automation Letters*, 5(4):5432–5439, 2020. 3

[3] Yogesh Balaji, Swami Sankaranarayanan, and Rama Chellappa. Metareg: Towards domain generalization using meta-regularization. *Advances in neural information processing systems*, 31, 2018. 3

[4] 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. 1, 2, 5

[5] Gilles Blanchard, Gyemin Lee, and Clayton Scott. Generalizing from several related classification tasks to a new unlabeled sample. *Advances in neural information processing systems*, 24, 2011. 3

[6] Francesco Cappio Borlino, Antonio D'Innocente, and Tatiana Tommasi. Rethinking domain generalization baselines. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 9227–9233. IEEE, 2021. 3

[7] 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. 1, 2, 5

[8] Fabio Maria Carlucci, Paolo Russo, Tatiana Tommasi, and Barbara Caputo. Hallucinating agnostic images to generalize across domains. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 3227–3234. IEEE, 2019. 3

[9] Prithvijit Chattopadhyay, Yogesh Balaji, and Judy Hoffman. Learning to balance specificity and invariance for in and out of domain generalization. In *European Conference on Computer Vision*, pages 301–318. Springer, 2020. 3

[10] Chen Chen, Wenjia Bai, Rhodri H Davies, Anish N Bhuva, Charlotte H Manisty, Joao B Augusto, James C Moon, Nay Aung, Aaron M Lee, Mihir M Sanghvi, et al. Improving the generalizability of convolutional neural network-based segmentation on cmr images. *Frontiers in cardiovascular medicine*, 7:105, 2020. 3

[11] Ke Chen, Ryan Oldja, Nikolai Smolyanskiy, Stan Birchfield, Alexander Popov, David Wehr, Ibrahim Eden, and Joachim Pehserl. Mvlidarnet: Real-time multi-class scene understanding for autonomous driving using multiple views. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2288–2294. IEEE, 2020. 3

[12] Ran Cheng, Ryan Razani, Ehsan Taghavi, Enxu Li, and Bingbing Liu. 2-s3net: Attentive feature fusion with adaptive feature selection for sparse semantic segmentation network. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12547–12556, 2021. 3

[13] Sungha Choi, Sanghun Jung, Huiwon Yun, Joanne T Kim, Seungryong Kim, and Jaegul Choo. Robustnet: Improving domain generalization in urban-scene segmentation via instance selective whitening. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11580–11590, 2021. 3

[14] 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. 3, 6

[15] Tiago Cortinhal, George Tzelepis, and Eren Erdal Aksoy. Salsanext: Fast, uncertainty-aware semantic segmentation of lidar point clouds. In *International Symposium on Visual Computing*, pages 207–222. Springer, 2020. 3

[16] Qi Dou, Daniel Coelho de Castro, Konstantinos Kamnitsas, and Ben Glocker. Domain generalization via model-agnostic learning of semantic features. *Advances in Neural Information Processing Systems*, 32, 2019. 3

[17] Yingjun Du, Jun Xu, Huan Xiong, Qiang Qiu, Xiantong Zhen, Cees GM Snoek, and Ling Shao. Learning to learn with variational information bottleneck for domain generalization. In *European Conference on Computer Vision*, pages 200–216. Springer, 2020. 3

[18] Sarah Erfani, Mahsa Baktashmotlagh, Masud Moshtaghi, Xuan Nguyen, Christopher Leckie, James Bailey, and Rao Kotagiri. Robust domain generalisation by enforcing distribution invariance. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16)*, pages 1455–1461. AAAI Press, 2016. 3

[19] Martin Gerdzhev, Ryan Razani, Ehsan Taghavi, and Liu Bingbing. Tornado-net: multiview total variation semantic segmentation with diamond inception module. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9543–9549. IEEE, 2021. 3

[20] Muhammad Ghifary, David Balduzzi, W Bastiaan Kleijn, and Mengjie Zhang. Scatter component analysis: A unified framework for domain adaptation and domain generalization. *IEEE transactions on pattern analysis and machine intelligence*, 39(7):1414–1430, 2016. 3

[21] Muhammad Ghifary, W Bastiaan Kleijn, Mengjie Zhang, and David Balduzzi. Domain generalization for object recognition with multi-task autoencoders. In *Proceedings of the IEEE international conference on computer vision*, pages 2551–2559, 2015. 3

[22] Leonardo Gigli, B Ravi Kiran, Thomas Paul, Andres Serna, Nagarjuna Vemuri, Beatriz Marcotegui, and Santiago Velasco-Forero. Road segmentation on low resolution lidar point clouds for autonomous vehicles. *arXiv preprint arXiv:2005.13102*, 2020. 4

[23] Benjamin Graham, Martin Engelcke, and Laurens Van Der Maaten. 3d semantic segmentation with submani-

fold sparse convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9224–9232, 2018. 3

[24] Jordan SK Hu and Steven L Waslander. Pattern-aware data augmentation for lidar 3d object detection. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pages 2703–2710. IEEE, 2021. 4

[25] 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. 2

[26] Chao Huang, Zhangjie Cao, Yunbo Wang, Jianmin Wang, and Mingsheng Long. Metasets: Meta-learning on point sets for generalizable representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8863–8872, 2021. 3

[27] Hao Huang, Cheng Chen, and Yi Fang. Manifold adversarial learning for cross-domain 3d shape representation. In *European Conference on Computer Vision*, pages 272–289. Springer, 2022. 3

[28] Lei Huang, Dawei Yang, Bo Lang, and Jia Deng. Decorrelated batch normalization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 791–800, 2018. 3

[29] Maximilian Ilse, Jakub M Tomczak, Christos Louizos, and Max Welling. Diva: Domain invariant variational autoencoders. In *Medical Imaging with Deep Learning*, pages 322–348. PMLR, 2020. 3

[30] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015. 6

[31] Maximilian Jaritz, Raoul De Charette, Emilie Wirbel, Xavier Perrotton, and Fawzi Nashashibi. Sparse and dense data with cnns: Depth completion and semantic segmentation. In *2018 International Conference on 3D Vision (3DV)*, pages 52–60. IEEE, 2018. 4

[32] 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. 3

[33] Aditya Khosla, Tinghui Zhou, Tomasz Malisiewicz, Alexei A Efros, and Antonio Torralba. Undoing the damage of dataset bias. In *European Conference on Computer Vision*, pages 158–171. Springer, 2012. 3

[34] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6

[35] Lingdong Kong, Niamul Quader, and Venice Erin Liong. Conda: Unsupervised domain adaptation for lidar segmentation via regularized domain concatenation. *arXiv preprint arXiv:2111.15242*, 2021. 3

[36] 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. 3

[37] Alexander Lehner, Stefano Gasperini, Alvaro Marcos-Ramiro, Michael Schmidt, Mohammad-Ali Nikouei Mahani, Nassir Navab, Benjamin Busam, and Federico Tombari. 3d-vfield: Adversarial augmentation of point clouds for domain generalization in 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17295–17304, 2022. 3

[38] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy Hospedales. Learning to generalize: Meta-learning for domain generalization. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018. 2, 3, 6

[39] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M Hospedales. Deeper, broader and artier domain generalization. In *Proceedings of the IEEE international conference on computer vision*, pages 5542–5550, 2017. 3

[40] Da Li, Jianshu Zhang, Yongxin Yang, Cong Liu, Yi-Zhe Song, and Timothy M Hospedales. Episodic training for domain generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1446–1455, 2019. 3

[41] Haoliang Li, Sinno Jialin Pan, Shiqi Wang, and Alex C Kot. Domain generalization with adversarial feature learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5400–5409, 2018. 3

[42] Haoliang Li, YuFei Wang, Renjie Wan, Shiqi Wang, Tie-Qiang Li, and Alex Kot. Domain generalization for medical imaging classification with linear-dependency regularization. *Advances in Neural Information Processing Systems*, 33:3118–3129, 2020. 3

[43] Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, and Baoquan Chen. Pointcnn: Convolution on x-transformed points. *Advances in neural information processing systems*, 31, 2018. 2

[44] Yiying Li, Yongxin Yang, Wei Zhou, and Timothy Hospedales. Feature-critic networks for heterogeneous domain generalization. In *International Conference on Machine Learning*, pages 3915–3924. PMLR, 2019. 3

[45] Quande Liu, Qi Dou, and Pheng-Ann Heng. Shape-aware meta-learning for generalizing prostate mri segmentation to unseen domains. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 475–485. Springer, 2020. 3

[46] Quande Liu, Qi Dou, Lequan Yu, and Pheng Ann Heng. Ms-net: multi-site network for improving prostate segmentation with heterogeneous mri data. *IEEE transactions on medical imaging*, 39(9):2713–2724, 2020. 3

[47] Toshihiko Matsuura and Tatsuya Harada. Domain generalization using a mixture of multiple latent domains. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 11749–11756, 2020. 3

[48] Andres Milioto, Ignacio Vizzo, Jens Behley, and Cyrill Stachniss. Rangenet++: Fast and accurate lidar semantic segmentation. In *2019 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 4213–4220. IEEE, 2019. 3, 4, 5

[49] Saeid Motiian, Marco Piccirilli, Donald A Adjeroh, and Gianfranco Doretto. Unified deep supervised domain adaptation and generalization. In *Proceedings of the IEEE international conference on computer vision*, pages 5715–5725, 2017. 3

[50] Krikamol Muandet, David Balduzzi, and Bernhard Schölkopf. Domain generalization via invariant feature representation. In *International Conference on Machine Learning*, pages 10–18. PMLR, 2013. 3

[51] Xingang Pan, Ping Luo, Jianping Shi, and Xiaoou Tang. Two at once: Enhancing learning and generalization capacities via ibn-net. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 464–479, 2018. 2, 3, 6, 8

[52] Xingang Pan, Xiaohang Zhan, Jianping Shi, Xiaoou Tang, and Ping Luo. Switchable whitening for deep representation learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1863–1871, 2019. 3

[53] 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. 1, 2, 6

[54] Duo Peng, Yinjie Lei, Munawar Hayat, Yulan Guo, and Wen Li. Semantic-aware domain generalized segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2594–2605, 2022. 3

[55] Vihari Piratla, Praneeth Netrapalli, and Sunita Sarawagi. Efficient domain generalization via common-specific low-rank decomposition. In *International Conference on Machine Learning*, pages 7728–7738. PMLR, 2020. 3

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

[57] 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. 2

[58] Fengchun Qiao, Long Zhao, and Xi Peng. Learning to learn single domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12556–12565, 2020. 3

[59] Mrigank Rochan, Shubhra Aich, Eduardo R Corral-Soto, Amir Nabatchian, and Bingbing Liu. Unsupervised domain adaptation in lidar semantic segmentation with self-supervision and gated adapters. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 2649–2655. IEEE, 2022. 3

[60] 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. 1, 2, 3, 6, 7, 8

[61] Shiv Shankar, Vihari Piratla, Soumen Chakrabarti, Siddhartha Chaudhuri, Preethi Jyothi, and Sunita Sarawagi. Generalizing across domains via cross-gradient training. In *International Conference on Learning Representations*, 2018. 3

[62] Aman Sinha, Hongseok Namkoong, and John Duchi. Certifiable distributional robustness with principled adversarial training. In *International Conference on Learning Representations*, 2018. 3

[63] Nathan Somavarapu, Chih-Yao Ma, and Zsolt Kira. Frustratingly simple domain generalization via image stylization. *arXiv preprint arXiv:2006.11207*, 2020. 3

[64] 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. 1, 2, 6

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

[66] 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. 2

[67] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6924–6932, 2017. 6

[68] Riccardo Volpi and Vittorio Murino. Addressing model vulnerability to distributional shifts over image transformation sets. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7980–7989, 2019. 3

[69] Riccardo Volpi, Hongseok Namkoong, Ozan Sener, John C Duchi, Vittorio Murino, and Silvio Savarese. Generalizing to unseen domains via adversarial data augmentation. *Advances in neural information processing systems*, 31, 2018. 3

[70] Yi Wei, Zibu Wei, Yongming Rao, Jiaxin Li, Jie Zhou, and Jiwen Lu. Lidar distillation: bridging the beam-induced domain gap for 3d object detection. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXIX*, pages 179–195. Springer, 2022. 4

[71] 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. 3

[72] 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. 2, 3

[73] 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. 3, 8

[74] Zhenlin Xu, Deyi Liu, Junlin Yang, Colin Raffel, and Marc Niethammer. Robust and generalizable visual representation learning via random convolutions. In *International Conference on Learning Representations*, 2021. 3

[75] Eojindl Yi, JuYoung Yang, and Junmo Kim. Enhanced prototypical learning for unsupervised domain adaptation in lidar semantic segmentation. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 01–07. IEEE, 2022. 3

[76] 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. 2, 3, 4, 7

[77] Xiangyu Yue, Yang Zhang, Sicheng Zhao, Alberto Sangiovanni-Vincentelli, Kurt Keutzer, and Boqing Gong. Domain randomization and pyramid consistency: Simulation-to-real generalization without accessing target domain data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2100–2110, 2019. 3

[78] Ling Zhang, Xiaosong Wang, Dong Yang, Thomas Sanford, Stephanie Harmon, Baris Turkbey, Bradford J Wood, Holger Roth, Andriy Myronenko, Daguang Xu, et al. Generalizing deep learning for medical image segmentation to unseen domains via deep stacked transformation. *IEEE transactions on medical imaging*, 39(7):2531–2540, 2020. 3

[79] Yang Zhang, Zixiang Zhou, Philip David, Xiangyu Yue, Zerong Xi, Boqing Gong, and Hassan Foroosh. Polarnet: An improved grid representation for online lidar point clouds semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9601–9610, 2020. 3

[80] Shanshan Zhao, Mingming Gong, Tongliang Liu, Huan Fu, and Dacheng Tao. Domain generalization via entropy regularization. *Advances in Neural Information Processing Systems*, 33:16096–16107, 2020. 3

[81] 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. 2, 3

[82] Hui Zhou, Xinge Zhu, Xiao Song, Yuexin Ma, Zhe Wang, Hongsheng Li, and Dahua Lin. Cylinder3d: An effective 3d framework for driving-scene lidar semantic segmentation. *arXiv preprint arXiv:2008.01550*, 2020. 3

[83] Kaiyang Zhou, Yongxin Yang, Timothy Hospedales, and Tao Xiang. Deep domain-adversarial image generation for domain generalisation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 13025–13032, 2020. 3

[84] Kaiyang Zhou, Yongxin Yang, Timothy Hospedales, and Tao Xiang. Learning to generate novel domains for domain generalization. In *European conference on computer vision*, pages 561–578. Springer, 2020. 3

[85] Kaiyang Zhou, Yongxin Yang, Yu Qiao, and Tao Xiang. Domain generalization with mixstyle. In *International Conference on Learning Representations*, 2021. 3