This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Towards Explicit Geometry-Reflectance Collaboration for Generalized LiDAR Segmentation in Adverse Weather

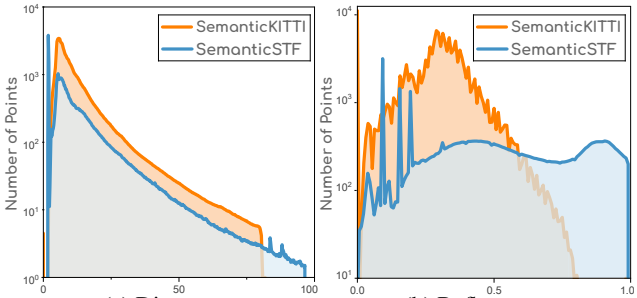
Longyu Yang[1], Ping Hu[1*], Shangbo Yuan[1], Lu Zhang[2], Jun Liu[3], Hengtao Shen[1,4], Xiaofeng Zhu[1]
[1]UESTC [2]DLUT [3]Lancaster University [4]Tongji University

Abstract

Existing LiDAR semantic segmentation models often suffer from decreased accuracy when exposed to adverse weather conditions. Recent methods addressing this issue focus on enhancing training data through weather simulation or universal augmentation techniques. However, few works have studied the negative impacts caused by the heterogeneous domain shifts in the geometric structure and reflectance intensity of point clouds. In this paper, we delve into this challenge and address it with a novel Geometry-Reflectance Collaboration (GRC) framework that explicitly separates feature extraction for geometry and reflectance. Specifically, GRC employs a dual-branch architecture designed to independently process geometric and reflectance features initially, thereby capitalizing on their distinct characteristic. Then, GRC adopts a robust multi-level feature collaboration module to suppress redundant and unreliable information from both branches. Consequently, without complex simulation or augmentation, our method effectively extracts intrinsic information about the scene while suppressing interference, thus achieving better robustness and generalization in adverse weather conditions. We demonstrate the effectiveness of GRC through comprehensive experiments on challenging benchmarks, showing that our method outperforms previous approaches and establishes new state-of-the-art results.

1. Introduction

LiDAR semantic segmentation is a fundamental task in 3D scene understanding and plays a crucial role in applications like robotics and autonomous driving [12, 54]. It works by predicting a semantic label for each point in a LiDAR scan, thus enabling the recognition of objects and environmental features essential for downstream tasks. With the rapid advancement of deep learning techniques, significant progress has been made in this area [7, 19, 33, 48, 49, 58, 62]. However, most traditional research is conducted under relatively idealized conditions, utilizing standard datasets for training and testing that often exclude various interference factors encountered in real-world applications, such as adverse weather conditions like fog, rain, snow, etc [6, 13, 14, 53]. Consequently, although current methods perform well in standard environments, they tend to experience significant performance degradation when exposed to more challenging scenarios [20, 56]. This highlights a critical challenge in the task: the lack of robustness and adaptability of LiDAR semantic segmentation models across different scenarios, especially under adverse weather.



(a) Distance (b) Reflectance

Figure 1. Histogram of distance and reflectance intensity for the averaged point cloud in SemanticKITTI [5] and SemanticSTF [53]. Reflectance intensity exhibits a much more pronounced domain shift compared to the geometric layout.

In addressing the challenge, recent efforts focus on employing weather-simulators [13, 14, 57, 61] and universal augmentation techniques [17, 18, 20, 24, 30, 53]. Although previous approaches have shown promising results, several limitations may still exist. On the one hand, weather simulation methods synthesize intrinsic characteristics of specific weather conditions but struggle to capture the full spectrum of weather types and severities, leading to limited adaptability when models encounter real-world conditions [61]. On the other hand, universal augmentation aims to learn general representations with reduced overfitting to the training data but often produces redundant and futile information, resulting in increased complexity and difficulty in model training [44, 63]. Unlike previous approaches that focus on enhancing the diversity of training data, we present a perspective to achieve robust LiDAR segmentation under

*Corresponding author

Input	Dense-fog	Light-fog	Rain	Snow	All
MinkNet w/ R	29.5	26.0	28.4	21.4	24.4
MinkNet w/o R	32.3	37.7	35.4	32.1	37.3
Ours	38.1	40.1	38.1	34.1	42.5

Table 1. Generalized segmentation accuracy of MinkNet [7] and the proposed method on the SemanticKITTI→SemanticSTF task. Excluding reflectance from the input improves MinkNet's performance across various adverse weather conditions.

adverse weather without relying on complex augmentation or simulation. As plotted in Fig.1, we observe that adverse weather conditions affect the geometry and reflectance of point clouds in distinct ways. Notably, geometric layout experiences much slighter shifts between normal and adverse conditions, suggesting that reflectance may be a critical factor for performance degradation. We verify the impact of reflectance in Tab.1. As shown, excluding reflectance from input improves generalization performance across four types of adverse weather conditions, demonstrating its negative effects.

Although discarding reflectance intensity significantly improves generalization capability, this simple strategy can be suboptimal, as corrupted reflectance still contains valuable information beneficial for semantic segmentation. To address this, we propose a novel Geometry-Reflectance Collaboration (GRC) framework designed to effectively extract intrinsic and generalizable features from point clouds. In GRC, we utilize a dual-branch architecture to explicitly separate the processing of geometric and reflectance information, leveraging their distinct and domain-invariant characteristics. Then, a robust multi-level feature collaboration module combines features from both the geometric and reflectance branches, while further suppressing noise and corruption caused by adverse weather. Consequently, our GRC framework effectively reduces mutual interference between geometry and reflectance, allowing for more precise feature extraction and thus achieving robust generalization across adverse conditions without the need for complex data augmentation or simulation, as demonstrated in Tab. 1. Experimental results on challenging benchmark datasets validate the effectiveness and efficiency of our approach. In summary, our contributions are as follows:

- We delve into the challenge in generalized LiDAR segmentation, under heterogeneous domain shifts in geometry and reflectance caused by adverse weather, and propose an effective Geometry-Reflectance Collaboration (GRC) framework to address this issue.
- We introduce a Robust Multi-level Feature Collaboration mechanism that effectively harnesses useful information from both geometry and reflectance while mitigating mutual interference.
- We conduct extensive experiments on challenging benchmark datasets, demonstrating that the proposed method achieves new state-of-the-art results.

2. Related Work

Point cloud semantic segmentation is an active research area that has witnessed significant advancement with the success of deep learning. Existing methods can be roughly divided into point-based, projection-based, and voxel-based approaches. Point-based methods process 3D points directly using multi-layer perceptrons (MLPs) [33–35], kernel point convolutions [28, 41, 47], graph neural networks (GNNs) [16, 23, 45], or transformer architectures [21, 48, 49, 60] to extract features and capture geometric structures. Projection-based methods transform LiDAR points into 2D images, with rangeview-based techniques [4, 8, 26, 29, 38] using spherical projection to create compact, dense, and computationally efficient representations that facilitate semantic segmentation via established 2D convolutional methods and pretrained 2D models, albeit at the cost of potential geometric information loss. Voxel-based methods [7, 11, 22, 32, 50, 64] divide the 3D space into voxel grids [7], cylindrical partitions [64], and radial windows [22]. The sparsity and irregularity of point clouds can lead to redundant computations. It can be addressed by 3D sparse convolutions [7, 11, 39], which focus computations on non-empty voxels, thereby enhancing efficiency and reducing computational load and memory usage.

Despite these advancements, traditional LiDAR segmentation models still face significant challenges in real-world applications, especially when exposed to adverse weather conditions [6, 13, 14, 20, 56]. To mitigate performance degradation, recent efforts have focused on simulating weather-related corruptions in point clouds during training [13, 14, 57, 61]. Although these physically-based strategies are grounded in realistic principles, they are often limited by the specificity of physic model based simulations and may lack adaptability in dynamic and complex environments. To develop more generalizable segmentation models, reducing overfitting through weather-agnostic data augmentation has become a popular approach [15, 17, 18, 20, 24, 30, 53]. For instance, Xiao *et al*. [53] apply geometry style randomization to point clouds, while He *et al*. [15] introduce augmentation in the feature space. Park *et al*. [30] first evaluate the impact of various universal augmentations and then employ reinforcement learning to predict optimal augmentation strategies. Although such universal augmentation based techniques can effectively generate more generalizable features, they often come with increased training costs and complexity [44, 63]. In contrast to these data-centric approaches, we thoroughly examine adverse weather's heterogeneous impacts on different aspects of point clouds, and accordingly introduce a robust and generalizable segmentation framework without relying on complex data augmentation or simulations.

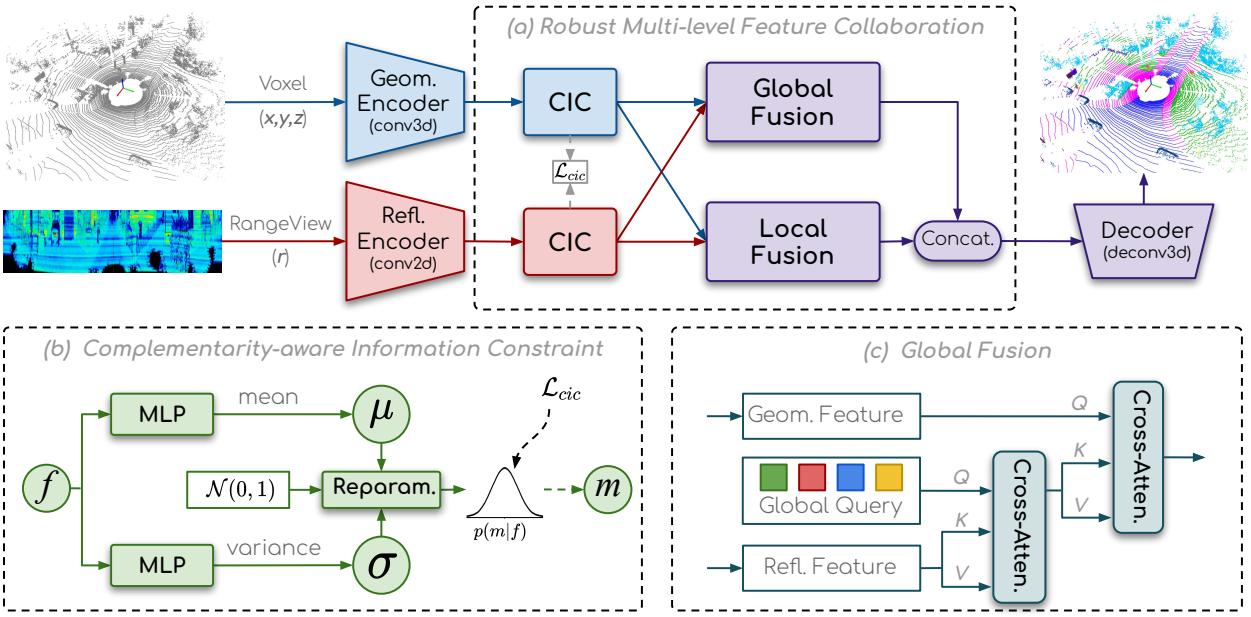


Figure 2. Overview of the Proposed Geometry-Reflectance Collaboration Network (GRCNet). GRCNet begins by independently processing geometric structure and reflectance intensity through separate feature encoders. These features are then integrated using a Robust Multi-level Feature Collaboration (RMFC) module. Within the RMFC, both geometric and reflectance features are augmented by Complementarity-aware Information Constraint (CIC) before being fused at global and local levels. By explicitly separating the feature encoding and then fusing, GRCNet mitigates mutual interference between geometric and reflectance features, enabling the extraction of intrinsic scene features that are robust across various weather conditions.

3. Methodology

This section presents the Geometry-Reflectance Collaboration Network (GRCNet), designed for generalized LiDAR semantic segmentation under adverse weather conditions. As illustrated in Fig.2, GRCNet employs a dual-branch network in the encoding phase to separately process geometric structure and reflectance intensity information, which are then fused to predict semantic labels. We begin by outlining the problem formulation in Sec.3.1. Next, in Sec.3.2, we detail the dual-branch design for separate feature encoding of geometric structure and reflectance intensity. In Sec.3.3, we describe the feature fusion using a Robust Multi-level Feature Collaboration (RMFC) module. Finally, we present the decoder and discuss training details in Sec. 3.4.

3.1. Problem Formulation

We formulate the task as a domain-generalized LiDAR segmentation problem. The objective is to achieve robust performance on target domains with adverse weather conditions by training exclusively on source domain data under standard conditions. Formally, during training, the model has access to source domain data $\mathcal{S} = \{(P_i^s, L_i^s)\}_{i=1}^{N^s}$, where N^s is the number of scans in the source domain dataset. For the i-th source sample (P_i^s, L_i^s) containing n LiDAR points, $L_i^s \in \mathbb{R}^n$ denotes the point-wise label annotations, and $P_i^s = \{(x_{ij}^s, y_{ij}^s, z_{ij}^s, r_{ij}^s)\}_{j=1}^n$ represents the i-th point cloud with (x, y, z) and r representing coordinates and reflectance intensity of each point respectively. Then, the goal of the task is to learn a semantic segmentation model $f : P \rightarrow L$ with only the source domain data $\mathcal{S}$ to perform well on the target domain $\mathcal{T} = \{P_i^t\}_{i=1}^{N^t}$, which remains unavailable during training.

3.2. Separated Feature Encoding

As discussed in Sec. 1, geometric structure and reflectance intensity capture fundamentally different aspects of LiDAR data: geometric structure represents spatial layout while reflectance intensity relates to surface characteristics and material properties. These two features undergo distinct degradation when transitioning from standard conditions to adverse weather. Ignoring these inherent differences and treating geometric structure and reflectance intensity as unified inputs can lead to suboptimal feature extraction when generalizing to challenging weather scenarios. Corruptions or distortions in reflectance can negatively impact the extraction of meaningful geometric features, and vice versa. To address this, we explicitly separate the feature extraction and propose a dual-branch model to independently and specializedly encode geometry and reflectance information.

3.2.1 Encoding Geometric Information

The voxel-based approach demonstrates robustness in dealing with variations in geometric information. The voxelization process discretizes continuous 3D spatial space into a

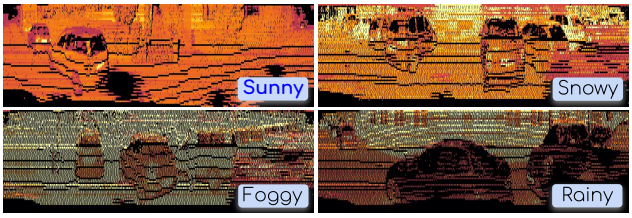


Figure 3. Visualization of reflectance intensity under different weather conditions, with intensity values represented in the same color scale.

regular grid of voxels, which reduces the sensitivity to noise and distortion. Moreover, operations like Conv3D on voxels can further enhance robustness due to the feature smoothing effect and multi-scale context aggregation. Inspired by this, we exploit voxel-based deep networks for generalizable geometric information encoding. Formally, we exclude the reflectance intensity r from a input point cloud P, resulting in $P_{geo} = [\mathbf{x}, \mathbf{y}, \mathbf{z}] \in \mathbb{R}^{n\times 3}$. Then, P_{geo} is voxelized and further processed by a geometric encoder,

$$F_{geo} = \mathbf{E}_{geo}(P_{geo}) \tag{1}$$

where $F_{geo} \in \mathbb{R}^{n\times c}$ is the resulting geometric feature in the form of sparse representation, and E_{geo} is the encoder network composed of a series of sparse Conv3D blocks [7] that gradually aggregates spatial resolutions and geometric context.

3.2.2 Encoding Reflectance Information

Compared to geometric structure, reflectance intensity is typically more affected by adverse weather, as adverse weather can significantly alter or attenuate intensity values through scattering or absorption of the LiDAR signal. This sensitivity to environmental factors causes drastic changes in the distribution of reflectance intensity, hindering the model's ability to generalize effectively. Nevertheless, despite these substantial variations, reflectance intensity still conveys valuable appearance information that aids in semantic segmentation. As shown in Fig. 3, even with intensity fluctuations and noise, in the dense 2D projection, appearance contours remain distinguishable due to relative intensity differences between objects.

Range view is a natural way to remove 3D layout and better retain 2D appearance in reflectance. To capture weather-invariant semantic information from reflectance intensity, we apply spherical projection [46] to transform 3D point cloud into 2D range-view image $P_{ref} \in \mathbb{R}^{H\times W}$, where each point's reflectance intensity is retained as input feature for the corresponding pixel. Reflectance features are then extracted as follows,

$$F_{ref} = E_{ref}(P_{ref}) \tag{2}$$

where $F_{ref} \in \mathbb{R}^{h\times w\times c}$ represents the extracted features. Since the input P_{ref} is in 2D, we design the reflectance encoder E_{ref} using 2D depthwise separable convolution blocks [36] equipped with Instance Normalization [42] . Similar to the geometric encoder, spatial size and appearance context are progressively aggregated through a series of convolution blocks. Notably, the spherical projection is invertible, enabling the extracted reflectance features to be mapped back to their corresponding locations in 3D space.

We recognize that while recent works [3, 27, 55] adopt multi-branch frameworks for point cloud segmentation, our approach is fundamentally different in three critical ways. First, we tackle the challenging domain generalization problem, which is ignored by [3, 27, 55]. Second, we delve into the distinct domain shifts in geometry and reflectance under adverse weather, which are often treated as unified inputs in previous works. This insight leads to a specialized dual-branch encoding architecture to achieve better generalization ability. Moreover, we introduce a novel fusion mechanism specifically designed to handle corruption and distortion in point clouds due to adverse weather, an issue unaddressed in [3, 27, 55].

3.3. Robust Multi-level Feature Collaboration

After obtaining the geometric and reflectance features, the next step is to fuse them. However, simple fusion strategies, such as concatenation or addition, are often less effective for domain generalized LiDAR semantic segmentation, as the extracted features can still be impacted by noise and distortions from adverse weather conditions. To effectively harness useful information from both features while minimizing mutual interference, in this part,we propose a robust multi-level feature collaboration mechanism, which further augment features with complementary task-relevant information and then perform fusion at local and global levels.

3.3.1 Complementarity-aware Information Constraint

To encourage the encoders to extract discriminative and complementary information from the scene, we impose information constraints on the extracted geometric and reflectance features. To achieve this, we design a Complementarity-aware Information Constraint inspired by the recent success of variational information bottlenecks [1, 2, 31]. Specifically, for a feature vector $f \in \mathcal{R}^c$ from a given feature map, we employ two fully connected networks to estimate the corresponding mean and variance vectors for each element,

$$\mu = \mathcal{W}_\mu(f), \quad \sigma = \mathcal{W}_\sigma(f) \tag{3}$$

where $\mathcal{W}_\mu(\cdot)$ and $\mathcal{W}_\sigma(\cdot)$ are composed fully connected layers, and $\{\mu \in \mathcal{R}^c, \sigma \in \mathcal{R}^c_+\}$ are the mean and standard deviation vectors. To ensure that σ remains positive, $\mathcal{W}_\sigma(\cdot)$

is followed by a softplus activation function [9]. We then define the enhanced feature vector m using reparameterization, which conforms to a multivariate Gaussian distribution characterized by the mean μ and standard deviation σ:

$$m = \mu + \epsilon \cdot \sigma, \quad \epsilon \sim \mathcal{N}(0, \mathbf{I}) \tag{4}$$

where $\mathbf{I}$ is an identity matrix, and ϵ represents noise sampled from a standard Gaussian distribution. Then, we separately apply such distributional projection for feature vectors in the geometric feature map F_{geo} and the reflectance feature map F_{ref}, as illustrated in Fig. 2 (b).

For each feature vector f_{ref} in F_{ref}, we map its location to 3D volume and locate the corresponding feature vector f_{geo} in F_{geo}. Then we have paired feature distributions,

$$\begin{aligned} p(m_{geo} \mid f_{geo}) &\sim \mathcal{N}(\mu_{geo},\ \sigma^2_{geo}\mathbf{I}) \\ p(m_{ref} \mid f_{ref}) &\sim \mathcal{N}(\mu_{ref},\ \sigma^2_{ref}\mathbf{I}) \end{aligned} \tag{5}$$

To explicitly enhance the robustness and complementarity of the geometric and reflectance features, we adopt the following objective:

$$\begin{aligned} \mathcal{L}_{cic} = &\mathcal{KL}\left(p(m_{geo} \mid f_{geo}) \parallel \mathbf{r}(m_{geo})\right) \\ &+ \mathcal{KL}\left(p(m_{ref} \mid f_{ref}) \parallel \mathbf{r}(m_{ref})\right) \\ &- \mathcal{KL}\left(p(m_{geo} \mid f_{geo}) \parallel p(m_{ref} \mid f_{ref})\right) \\ &- \mathcal{KL}\left(p(m_{ref} \mid f_{ref}) \parallel p(m_{geo} \mid f_{geo})\right) \end{aligned} \tag{6}$$

where $\mathcal{KL}(\cdot||\cdot)$ denotes the Kullback-Leibler divergence, and $\mathbf{r}(\cdot)$ represents a prior marginal distribution set as a standard Gaussian. Minimizing $\mathcal{L}_{cic}$ serves two primary purposes. First, it reduces the dependence between m_{geo} and f_{geo}, as well as between m_{ref} and f_{ref}, allowing m_{geo} and m_{ref} to discard domain-specific noise and capture more robust, generalizable, and task-relevant information. Second, it decreases the correlation between m_{geo} and m_{ref}, thereby reducing redundancy and encouraging complementarity between the geometric and reflectance features. This dual effect ensures that the features are both robust and complementary, thereby enhancing the subsequent feature fusion and final prediction.

3.3.2 Local-level Fusion

So far, we have established distributional feature representations for each vector in the voxel-based 3D geometric feature map F_{geo} and the range-view-based 2D reflectance feature map F_{ref}. To facilitate local fusion, we retrieve reflectance features for geometric feature vectors via spherical projection. Then, we perform local fusion by combining geometric feature with reflectance features. Given the a geometric feature $\{\mu_{geo}, \sigma_{geo}\}$ and the corresponding reflectance feature $\{\mu_{ref}, \sigma_{ref}\}$, we apply a dynamic fusion approach, defined as:

$$f_{local} = \alpha \cdot \mu_{geo} + (1 - \alpha) \cdot \mu_{ref} \tag{7}$$

where $\alpha = \frac{e^{\frac{1}{\bar{\sigma}_{geo}}}}{e^{\frac{1}{\bar{\sigma}_{geo}}} + e^{\frac{1}{\bar{\sigma}_{ref}}}}$ with $\bar{\sigma}_{geo}$ and $\bar{\sigma}_{ref}$ as the standard deviation averaged along the channel dimmension. By incorporating the standard deviations into the fusion weights, the model dynamically assesses the reliability and contribution of each feature, enhancing the robustness of the fused features. Eq. 7 is applied to all the geometric feature vectors, thus resulting in locally fused feature map F_{local}.

3.3.3 Global-level Fusion

The reflectance maps under adverse weather contain substantial noise at the local scale, yet they convey clear object semantics at the global level. Therefore, we adopt a global-level fusion module to further exploit this valuable information. An intuitive approach is to use Cross-Attention [43] to directly aggregate global context from reflectance features into the geometric space. However, this straightforward solution can face challenges in computational efficiency and may remain vulnerable to overall distortions in the reflectance features caused by adverse weather. To address these limitations, we introduce a two-stage cross-attention mechanism, as illustrated in Fig. 2 (c). Given the reflectance feature map $M_{ref} \in \mathbb{R}^{h \times w \times c}$, we introduce a set of learnable global-query tokens $Q \in \mathbb{R}^{m \times c}$, with $m \ll hw$. The reflectance feature map M_{ref} is first converted into Key and Value features, which are aggregated by the query tokens Q to produce intermediate global features. These intermediate global features are then further aggregated by the geometric features M_{geo}:

$$F_{global} = \mathcal{CA}(M_{geo},\ \mathcal{CA}(Q, M_{ref})) \tag{8}$$

where $\mathcal{CA}(\cdot, \cdot)$ represents the cross-attention operation, $M_{geo}$and M_{ref} are the processed geometric features and relectance features, respectively. The learnable query Q serves as an intermediary, integrating global information from the range-view features across different perspectives and effectively passing it to the geometric information in the voxel domain. It avoids the direct computation of attention between two large-scale feature sets in a single step, optimizing computational efficiency while ensuring meaningful cross-domain interaction. During the cross-attention process, the learnable Q not only facilitates the fusion but also compresses and refines the range-view's global features, making the information more compact and relevant.

3.4. Decoding and Training

Since both the local-level and global-level fusion modules are designed to transfer information from the reflectance feature to the geometric feature, the resulting features F_{local} and F_{global} are both represented in the voxelized 3D space. We concatenate these features at each location and pass

the combined representation to a decoder network to generate the final predictions. During training, we optimize the model using the following loss function:

$$\mathcal{L}_{total} = \mathcal{L}_{CE} + \beta \mathcal{L}_{cic} \tag{9}$$

where $\mathcal{L}_{CE}$ is the standard cross-entropy loss, and $\mathcal{L}_{cic}$ is the information constraint loss defined in Eq. 6.

4. Experiments

We evaluate the generalization ability of our method under adverse conditions with three settings: SemanticKITTI $\rightarrow$ SemanticSTF, SynLiDAR $\rightarrow$ SemanticSTF, and SemanticKITTI $\rightarrow$ SemanticKITTI-C. In these settings, the clear-weather real dataset SemanticKITTI [5] and synthetic dataset SynLiDAR [52] serve as the source domains, while the adverse-condition target domains are represented by SemanticSTF [53] and SemanticKITTI-C [20].

4.1. Experimental details

Datasets. *SemanticKITTI*[5] is a large-scale semantic segmentation dataset based on the KITTI Vision Benchmark[10]. The data is collected with a 64-beam LiDAR and annotated with over 19 semantic categories. Following the official protocol, we use sequences 00-07 and 09-10 as the training set, and sequence 08 as the validation set. *SynLiDAR*[52] is a large-scale synthetic LiDAR dataset derived from various virtual environments, consisting of 13 sequences of LiDAR point clouds with approximately 20,000 scans. We split the training and validation sets in line with previous works[30, 52, 53, 61]. *SemanticSTF*[53] is an adverse-weather LiDAR segmentation dataset that extends the realistic STF Detection Benchmark[6] by providing point-wise annotations for 21 semantic categories. It includes four common adverse weather conditions: dense fog, light fog, snow, and rain. We follow the official protocol and utilize its validation set, with 19 semantic categories, for testing. *SemanticKITTI-C*[20] is a corrupted LiDAR segmentation dataset created by applying eight types of corruption at three severity levels to the validation set of SemanticKITTI[5]. We use mean Intersection over Union (mIoU) as the evaluation metric in our experiments.

Implementation details. Unless otherwise specified, we use the encoder and decoder of MiniNet-18/16 [7] as our geometric encoder and task decoder, respectively. For the reflectance input, we employ an encoder built with the efficient Inverted Residual Blocks [36] and Instance Normalization [42]. Fusion is performed on the outputs of the two encoders. To train the network, we apply standard data augmentations, including random dropping, rotation, flipping, scaling, etc. We use the momentum SGD optimizer with a momentum of 0.9, a weight decay of 0.0001, and a batch size of 6 for both SemanticKITTI and SynLiDAR. The learning rate is initially set to 0.24 and is adjusted with the OneCycleLR policy [37] for 50 epochs in total. All experiments are conducted on a single Nvidia RTX4090 GPU.

4.2. Main Results

SemanticKITTI to SemanticSTF/SemanticKITTI-C. As shown in the upper part of Tab. 2, our proposed method demonstrates significant improvements of the generalization performance on SemanticSTF dataset. Specifically, compared to the widely adopted baseline MinkNet, our approach achieves an impressive +18.1 increase in mIoU. Additionally, our method surpasses the state-of-the-art method RDA [30], by +3.0 in mIoU. This substantial performance gain is attributed to our strategy of decoupling geometric and appearance information, which effectively reduces negative impact of distortions and corruptions. Our method demonstrates the best results across all tested adverse weather conditions. Specifically, we observe mIoU improvements of +2.1 in dense fog, +2.6 in light fog, +0.5 in rain, and +1.0 in snow over the state-of-the-art [30]. Compared with the MinkNet baseline, our method achieves remarkable increases of +8.6 mIoU in dense fog, +14.1 mIoU in light fog, +9.7 mIoU in rain, and +12.7 mIoU in snow. Beyond performing well across adverse weather conditions, our method also yields superior results across most of the semantic categories. Compared to the baseline, we achieve more than +10 mIoU improvement in 13 categories. Notably, our approach shows gains of over +5 mIoU compared to the state-of-the-art method [30] in categories such as trucks, other vehicles, motorcyclists, roads, parking, and terrain. It is noteworthy that the previous state-of-the-art method [30] relies on reinforcement learning for data augmentation, requiring four A6000 GPUs for training. In contrast, our approach achieves superior performance through a simpler framework, without the need for such complex augmentation techniques, and can be trained efficiently on a single RTX 4090 GPU.

To further evaluate our method's effectiveness, we conduct experiments with more baseline network and datasets. As shown in Tab. 3, with SPVCNN [39] or MinkNet18/32 [7], our method outperforms the previous state-of-the-art on both SemanticSTF and SemanticKITTI-C datasets with significant gaps. This again demonstrates the superioity of our proposed framework.

SynLiDAR to SemanticSTF. The transition from SynLiDAR to SemanticSTF poses an even greater challenge than the previous setting, due to the substantial disparity between synthetic clear weather environments and real-world adverse weather conditions. As shown in the lower part of Tab. 2, our method achieves the best overall performance in this challenging context, slightly outperforming the state-of-the-art UniMix method [61]. While the improvement may appear modest, it is noteworthy that UniMix relies on

Method	car	bi.cle	mt.cle	truck	oth-v.	pers.	bi.clst	mt.clst	road	parki.	sidew.	oth-g.	build.	fence	veget.	trunk	terra.	pole	traf.	D-fog	L-fog	Rain	Snow	mIoU
Oracle	89.4	42.1	0.0	59.9	61.2	69.6	39.0	0.0	82.2	21.5	58.2	45.6	86.1	63.6	80.2	52.0	77.6	50.1	61.7	51.9	54.6	57.9	53.7	54.7
SemanticKITTI→SemanticSTF																								
MinkNet[7]	55.9	0.0	0.2	1.9	10.9	10.3	6.0	0.0	61.2	10.9	32.0	0.0	67.9	41.6	49.8	27.9	40.8	29.6	17.5	29.5	26.0	28.4	21.4	24.4
PolarMix[51]	57.8	1.8	3.8	16.7	3.7	26.5	0.0	2.0	65.7	2.9	32.5	0.3	71.0	48.7	53.8	20.5	45.4	25.9	15.8	29.7	25.0	28.6	25.6	26.0
MMD[25]	63.6	0.0	2.6	0.1	11.4	28.1	0.0	0.0	67.0	14.1	37.9	0.3	67.3	41.2	57.1	27.4	47.9	28.2	16.2	30.4	28.1	32.8	25.2	26.9
PCL[59]	65.9	0.0	0.0	17.7	0.4	8.4	0.0	0.0	59.6	12.0	35.0	1.6	74.0	47.5	60.7	15.8	48.9	26.1	27.5	28.9	27.6	30.1	24.6	26.4
PointDR[53]	67.3	0.0	4.5	19.6	9.0	18.8	2.7	0.0	62.6	12.9	38.1	0.6	73.3	43.8	56.4	32.2	45.7	28.7	27.4	31.3	29.7	31.9	26.2	28.6
UniMix[61]	82.7	**6.6**	8.6	4.5	15.1	35.5	**15.5**	37.7	55.8	10.2	36.2	1.3	72.8	40.1	49.1	33.4	34.9	23.5	33.5	34.8	30.2	34.9	30.9	31.4
RDA[30]	**86.1**	4.8	13.8	39.7	26.6	**55.4**	8.5	50.4	63.7	14.9	37.9	5.5	75.2	**52.7**	60.4	**39.7**	44.9	30.1	40.8	36.0	37.5	37.6	33.1	39.5
Ours	85.6	2.5	**16.7**	**47.6**	**33.0**	52.9	4.2	**60.0**	**70.3**	**20.7**	**42.2**	**12.4**	**78.2**	51.4	**63.2**	38.3	**53.0**	**32.1**	**43.3**	**38.1**	**40.1**	**38.1**	**34.1**	**42.5**
SynLiDAR→SemanticSTF																								
MinkNet[7]	27.1	**3.0**	0.6	15.8	0.1	25.2	1.8	5.6	23.9	0.3	14.6	0.6	36.3	19.9	37.9	17.9	41.8	9.5	2.3	19.9	17.2	17.2	11.9	15.0
PolarMix[51]	39.2	1.1	1.2	8.3	1.5	17.8	0.8	0.7	23.3	1.3	17.5	0.4	45.2	24.8	46.2	20.1	38.7	7.6	1.9	16.1	15.5	19.2	15.6	15.7
MMD[25]	25.5	2.3	2.1	13.2	0.7	22.1	1.4	7.5	30.8	0.4	17.6	0.2	30.9	19.7	37.6	19.3	**43.5**	9.9	2.6	17.3	16.3	20.0	12.7	15.1
PCL[59]	30.9	0.8	1.4	10.0	0.4	23.3	**4.0**	7.9	28.5	1.3	17.7	1.2	39.4	18.5	40.0	16.0	38.6	12.1	2.3	17.8	16.7	19.3	14.1	15.5
PointDR[53]	37.8	2.5	2.4	23.6	0.1	26.3	2.2	3.3	27.9	7.7	17.5	0.5	47.6	25.3	45.7	21.0	37.5	17.9	5.5	19.5	19.91	21.1	16.9	18.5
UniMix[61]	**65.4**	0.1	**3.9**	16.9	**5.3**	**32.3**	2.0	**19.3**	52.1	5.0	**27.3**	**3.0**	49.4	20.3	**58.5**	22.7	23.2	**26.9**	**10.4**	**24.3**	22.9	**26.1**	**20.9**	23.4
RDA[30]	39.3	2.9	0.9	19.4	0.8	27.7	2.2	3.8	42.5	**9.4**	21.6	0.3	51.9	33.5	47.3	23.1	33.3	23.2	6.8	19.0	21.2	23.1	17.3	20.5
Ours	49.5	2.0	2.3	**24.4**	1.8	31.8	2.2	9.0	**62.3**	4.6	25.1	0.2	**59.1**	**35.0**	55.7	**24.1**	32.8	18.8	6.1	21.7	**23.5**	24.9	20.5	**23.5**

Table 2. Generalized segmentation performance with SemanticKITTI and SynLiDAR as the source and SemanticSTF as the target.

Method	SemanticSTF	SemanticKITTI-C
SPVCNN	28.1	52.5
SPVCNN+RDA [30]	38.4	52.9
SPVCNN+Ours	**40.8**	**54.2**
MinkNet18/32	31.4	53.0
MinkNet18/32+RDA [30]	39.5	58.6
MinkNet18/32+Ours	**43.7**	**60.3**

Table 3. Generalized segmentation performance on SemanticSTF and SemanticKITTI-C.

GB	RB	CIC	LF	GF	D-fog	L-fog	Rain	Snow	All
					29.5	26.0	28.4	21.4	24.4
✓					32.3	37.7	35.4	32.1	37.3
✓	✓				31.6	35.7	35.1	32.8	36.5
✓	✓	✓			33.1	40.5	36.8	33.4	38.6
✓	✓	✓	✓		36.6	39.3	37.8	33.4	41.2
✓	✓	✓	✓	✓	38.1	40.1	38.1	34.1	42.5

Table 4. Generalized segmentation performance on SemanticSTF with different components. "GB" is the geometric encoder. "RB" is the reflectance encoder. "CIC" , "LF" and "GF"denotes the complementarity-aware information constraint, local fusion, and global fusion, respectively.

complex weather simulations and exponential moving average (EMA) ensembling [40], which result in increased complexity and heavy resource consumption. In contrast, our method achieves competitive performance without requiring these intricate techniques, highlighting its efficiency and effectiveness in adapting to challenging environments.

4.3. Method Analysis

Ablation on Method Desgin. We first analyze the effectiveness of the different designs of our method in Tab. 4. The first row represents MinkNet, which is the baseline model and achieves the lowest accuracy. The second row indicate our geometric-encoder-only model, which is equivalent to the baseline without reflectance as input. By simply excluding reflection intensity, we observed a significant improvement of baseline across all adverse weather conditions, achieving a +12.9 mIoU overall. This shows the severe negative impact of drastic changes in reflectance intensity distribution on robustness. In the third row, we introduce a reflectance encoder and simply merge the features by addition. However, we observe that the overall accuracy drops by 0.8 mIoU. This shows that the features are suffering from noise and distortion. Yet it is worth noting that even with minor degradation, the accuracy is still much higher than the baseline, which also demonstrate the rationality of separating the feature extraction. In the forth row, we apply the proposed information constraint loss during training and simply adding the processed features. We find that the performance gets improved and outperforms the baseline model without reflectance, showing the effectiveness of the propose information constraint. In the fifth row, when introducing the Complementarity-aware Information Constraint with local fusion, the model achieves improvements in dense fog(+7.1 mIoU), light fog(+13.3 mIoU), rain(+9.4 mIoU), snow(+12.0 mIoU), and an overall performence increase of +16.8 mIoU. This shows the importance of the proposed information constraint, which alleviates noise interference and weather-relate distortion , and the stochastic modeling it introduces makes the model more robust to domain shifts. In the last row, further applying the global fusion results in improvements in dense fog(+8.6 mIoU), light

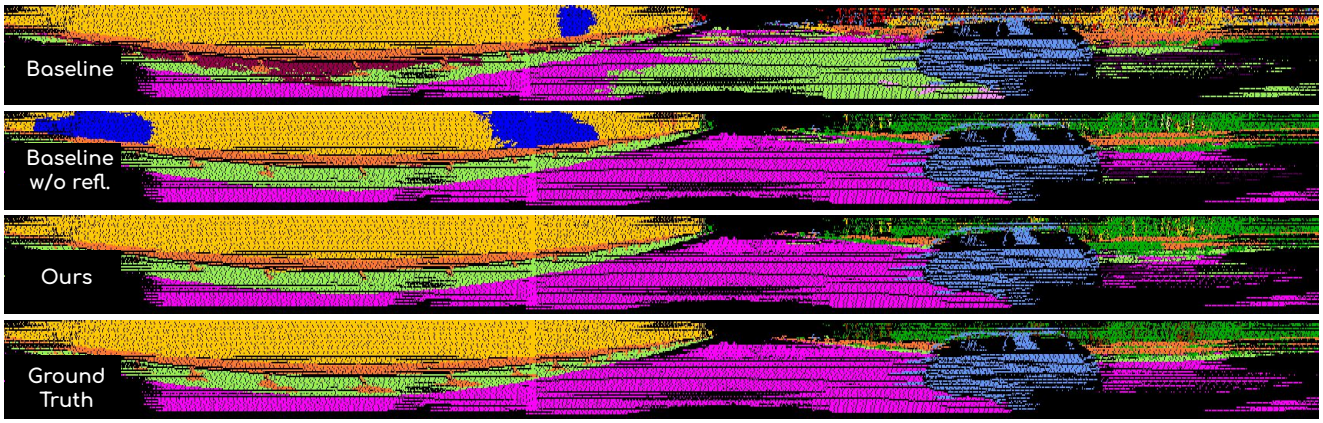


Figure 4. Qualitative results of our method on SemanticSTF. Without reflectance, the baseline achieves better overall segmentation accuracy but introduces new mispredictions, such as the blue region on the left half of the second image. Our method successfully corrects these mispredictions while maintaining overall accuracy.

Addition [3]	Gated Fusion [55]	Cross-Atten. [43]	Ours
36.5	37.7	37.4	42.5

Table 5. Impact of different fusion strategies.

fog(+14.1 mIoU), rain(+9.7 mIoU), snow(+12.7 mIoU), indicating that the success extraction of global information. In addition, removing reflectance intensity reduces MinkNet's performance from 63.8 mIoU to 62.6 mIoU in the source domain, while GRC raises it to 63.1 mIoU, showing its effectiveness in utilizing both geometric and reflectance cues. In Fig. 5, we also show generalized segmentation performance across different distance intervals. As shown, our method achieve significant performance gains compared to the baseline model. We visualize some qualitative results in Fig. 4

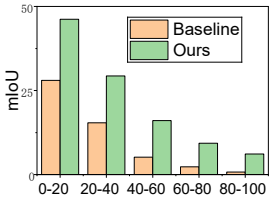


Figure 5. Generalized segmentation performance across different distance on SemanticSTF.

Fusion Strategy. In Tab. 5, we compare the impact of different fusion strategies that can be applied in our framework. As shown, fusing the geometric feature and reflectance features by simply addition leads to the worst performance, while applying gate fusion [55] or cross-attention [43] improves the accuracy by about +1 in mIoU. We also observe that directly applying cross-attention between the two types of features incurs out-of-memory when trained with a RTX4090 GPU. In contrast, our proposed method achieves the best performance without exhausting the memory, demonstrating it efficacy. We also experiment by progressively adding extra RMFC modules to shallower blocks in the encoders. However, the results in the target domain consistently showed a decline in performance. This may be caused by the fact that shallower features convey more heavily noise and distortion.

	MACs (G)	Param (M)	FPS	mIoU
MinkNet18/32	141.7	21.7	19	31.4
MinkNet18/16	37.5	5.4	25	24.4
MinkNet18/16+Ours	70.8	8.4	19	42.5

Table 6. Analsyis of model size and inference complexity.

Efficiency Analysis. Tab. 6 presents the efficiency analysis of our proposed method alongside the baseline model. Compared to the baseline MinkNet18/16, our approach modestly increases model size and complexity, yet achieves a significant improvement in performance, with a +18% boost in mIoU. In contrast, MinkNet18/32 yields only a +7% mIoU increase despite its much larger model size and higher computational cost. This discrepancy highlights that the efficiency and effectiveness of our model stem from the inherent strengths of the proposed architectural design.

5. Conclusion

In this paper, we introduced a novel geometry-reflectance collaboration framework for generalized LiDAR semantic segmentation under adverse weather conditions. We delved into the challenge posed by heterogeneous domain shifts in the geometric structure and reflectance intensity of point clouds and proposed a solution that explicitly separates feature extraction for geometry and reflectance. Additionally, we developed a robust multi-level feature collaboration module to suppress redundant and unreliable information from both branches, achieving more robust and generalizable segmentation. Our results demonstrate that the propose approach outperforms prior methods, establishing new state-of-the-art performance.

Acknowledgement

This work was supported by National Key Research and Development Program of China under Grant 2022YFA1004100 and National Natural Science Foundation of China under Grant 62476048.

References

[1] Kartik Ahuja, Ethan Caballero, Dinghuai Zhang, Jean-Christophe Gagnon-Audet, Yoshua Bengio, Ioannis Mitliagkas, and Irina Rish. Invariance principle meets information bottleneck for out-of-distribution generalization. *Adv. Neural Inform. Process. Syst.*, 34:3438–3450, 2021. 4

[2] Alexander A Alemi, Ian Fischer, Joshua V Dillon, and Kevin Murphy. Deep variational information bottleneck. *arXiv preprint arXiv:1612.00410*, 2016. 4

[3] Yara Ali Alnaggar, Mohamed Afifi, Karim Amer, and Mohamed ElHelw. Multi projection fusion for real-time semantic segmentation of 3d lidar point clouds. In *WACV*, 2021. 4, 8

[4] Angelika Ando, Spyros Gidaris, Andrei Bursuc, Gilles Puy, Alexandre Boulch, and Renaud Marlet. Rangevit: Towards vision transformers for 3d semantic segmentation in autonomous driving. In *CVPR*, 2023. 2

[5] 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 *ICCV*, 2019. 1, 6

[6] Mario Bijelic, Tobias Gruber, Fahim Mannan, Florian Kraus, Werner Ritter, Klaus Dietmayer, and Felix Heide. Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. In *CVPR*, 2020. 1, 2, 6

[7] Christopher Choy, JunYoung Gwak, and Silvio Savarese. 4d spatio-temporal convnets: Minkowski convolutional neural networks. In *CVPR*, 2019. 1, 2, 4, 6, 7

[8] Tiago Cortinhal, George Tzelepis, and Eren Erdal Aksoy. Salsanext: Fast, uncertainty-aware semantic segmentation of lidar point clouds. In *ISVC*, 2020. 2

[9] Charles Dugas, Yoshua Bengio, François Bélisle, Claude Nadeau, and René Garcia. Incorporating second-order functional knowledge for better option pricing. *Adv. Neural Inform. Process. Syst.*, 13, 2000. 5

[10] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *CVPR*, 2012. 6

[11] Benjamin Graham, Martin Engelcke, and Laurens Van Der Maaten. 3d semantic segmentation with submanifold sparse convolutional networks. In *CVPR*, 2018. 2

[12] Yulan Guo, Hanyun Wang, Qingyong Hu, Hao Liu, Li Liu, and Mohammed Bennamoun. Deep learning for 3d point clouds: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43(12):4338–4364, 2020. 1

[13] Martin Hahner, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Fog simulation on real lidar point clouds for 3d object detection in adverse weather. In *ICCV*, 2021. 1, 2

[14] Martin Hahner, Christos Sakaridis, Mario Bijelic, Felix Heide, Fisher Yu, Dengxin Dai, and Luc Van Gool. Lidar snowfall simulation for robust 3d object detection. In *CVPR*, 2022. 1, 2

[15] Pei He, Licheng Jiao, Lingling Li, Xu Liu, Fang Liu, Wenping Ma, Shuyuan Yang, and Ronghua Shang. Domain generalization-aware uncertainty introspective learning for 3d point clouds segmentation. In *ACMMM*, 2024. 2

[16] Chang-Qin Huang, Fan Jiang, Qiong-Hao Huang, Xi-Zhe Wang, Zhong-Mei Han, and Wei-Yu Huang. Dual-graph attention convolution network for 3-d point cloud classification. *IEEE Trans. Neural Networks Learn. Syst.*, 35(4): 4813–4825, 2022. 2

[17] Hyeonseong Kim, Yoonsu Kang, Changgyoon Oh, and Kuk-Jin Yoon. Single domain generalization for lidar semantic segmentation. In *CVPR*, 2023. 1, 2

[18] Jaeyeul Kim, Jungwan Woo, Jeonghoon Kim, and Sunghoon Im. Domain generalization in lidar semantic segmentation leveraged by density discriminative feature embedding. *arXiv preprint arXiv:2312.12098*, 2023. 1, 2

[19] Lingdong Kong, Youquan Liu, Runnan Chen, Yuexin Ma, Xinge Zhu, Yikang Li, Yuenan Hou, Yu Qiao, and Ziwei Liu. Rethinking range view representation for lidar segmentation. In *ICCV*, 2023. 1

[20] Lingdong Kong, Youquan Liu, Xin Li, Runnan Chen, Wenwei Zhang, Jiawei Ren, Liang Pan, Kai Chen, and Ziwei Liu. Robo3d: Towards robust and reliable 3d perception against corruptions. In *ICCV*, 2023. 1, 2, 6

[21] Xin Lai, Jianhui Liu, Li Jiang, Liwei Wang, Hengshuang Zhao, Shu Liu, Xiaojuan Qi, and Jiaya Jia. Stratified transformer for 3d point cloud segmentation. In *CVPR*, 2022. 2

[22] Xin Lai, Yukang Chen, Fanbin Lu, Jianhui Liu, and Jiaya Jia. Spherical transformer for lidar-based 3d recognition. In *CVPR*, 2023. 2

[23] Loic Landrieu and Martin Simonovsky. Large-scale point cloud semantic segmentation with superpoint graphs. In *CVPR*, 2018. 2

[24] 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 *CVPR*, 2022. 1, 2

[25] Haoliang Li, Sinno Jialin Pan, Shiqi Wang, and Alex C Kot. Domain generalization with adversarial feature learning. In *CVPR*, 2018. 7

[26] Li Li, Hubert PH Shum, and Toby P Breckon. Rapid-seg: Range-aware pointwise distance distribution networks for 3d lidar segmentation. In *ECCV*, 2025. 2

[27] Xiaoyan Li, Gang Zhang, Hongyu Pan, and Zhenhua Wang. Cpgnet: Cascade point-grid fusion network for real-time lidar semantic segmentation. In *ICRA*, 2022. 4

[28] Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, and Baoquan Chen. Pointcnn: Convolution on x-transformed points. *Adv. Neural Inform. Process. Syst.*, 31, 2018. 2

[29] Andres Milioto, Ignacio Vizzo, Jens Behley, and Cyrill Stachniss. Rangenet++: Fast and accurate lidar semantic segmentation. In *IROS*, 2019. 2

[30] Junsung Park, Kyungmin Kim, and Hyunjung Shim. Rethinking data augmentation for robust lidar semantic segmentation in adverse weather. *arXiv preprint arXiv:2407.02286*, 2024. 1, 2, 6, 7

[31] Xue Bin Peng, Angjoo Kanazawa, Sam Toyer, Pieter Abbeel, and Sergey Levine. Variational discriminator bottleneck: Im-

proving imitation learning, inverse rl, and gans by constraining information flow. *arXiv preprint arXiv:1810.00821*, 2018. 4

[32] Charles R Qi, Hao Su, Matthias Nießner, Angela Dai, Mengyuan Yan, and Leonidas J Guibas. Volumetric and multi-view cnns for object classification on 3d data. In *CVPR*, 2016. 2

[33] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *CVPR*, 2017. 1, 2

[34] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Adv. Neural Inform. Process. Syst.*, 30, 2017.

[35] Guocheng Qian, Yuchen Li, Houwen Peng, Jinjie Mai, Hasan Hammoud, Mohamed Elhoseiny, and Bernard Ghanem. Pointnext: Revisiting pointnet++ with improved training and scaling strategies. *Adv. Neural Inform. Process. Syst.*, 35:23192–23204, 2022. 2

[36] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *CVPR*, 2018. 4, 6

[37] Leslie N Smith and Nicholay Topin. Super-convergence: Very fast training of neural networks using large learning rates. In *Artificial intelligence and machine learning for multi-domain operations applications*, pages 369–386, 2019. 6

[38] Xiaorui Sun, Jun Liu, Heng Tao Shen, Xiaofeng Zhu, and Ping Hu. On efficient variants of segment anything model: A survey. *arXiv preprint arXiv:2410.04960*, 2024. 2

[39] 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 *ECCV*, 2020. 2, 6

[40] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *NeurIPS*, 2017. 7

[41] 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 *ICCV*, 2019. 2

[42] D Ulyanov. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016. 4, 6

[43] A Vaswani. Attention is all you need. *Adv. Neural Inform. Process. Syst.*, 2017. 5, 8

[44] Jindong Wang, Cuiling Lan, Chang Liu, Yidong Ouyang, Tao Qin, Wang Lu, Yiqiang Chen, Wenjun Zeng, and S Yu Philip. Generalizing to unseen domains: A survey on domain generalization. *IEEE Trans. Knowl. Data Eng.*, 35(8): 8052–8072, 2022. 1, 2

[45] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *ACM Trans. Graph.*, 38(5):1–12, 2019. 2

[46] 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 *ICRA*, 2018. 4

[47] Wenxuan Wu, Zhongang Qi, and Li Fuxin. Pointconv: Deep convolutional networks on 3d point clouds. In *CVPR*, 2019. 2

[48] Xiaoyang Wu, Yixing Lao, Li Jiang, Xihui Liu, and Hengshuang Zhao. Point transformer v2: Grouped vector attention and partition-based pooling. *Adv. Neural Inform. Process. Syst.*, 35:33330–33342, 2022. 1, 2

[49] Xiaoyang Wu, Li Jiang, Peng-Shuai Wang, Zhijian Liu, Xihui Liu, Yu Qiao, Wanli Ouyang, Tong He, and Hengshuang Zhao. Point transformer v3: Simpler faster stronger. In *CVPR*, 2024. 1, 2

[50] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *CVPR*, 2015. 2

[51] Aoran Xiao, Jiaxing Huang, Dayan Guan, Kaiwen Cui, Shijian Lu, and Ling Shao. Polarmix: A general data augmentation technique for lidar point clouds. *Adv. Neural Inform. Process. Syst.*, 35:11035–11048, 2022. 7

[52] Aoran Xiao, Jiaxing Huang, Dayan Guan, Fangneng Zhan, and Shijian Lu. Transfer learning from synthetic to real lidar point cloud for semantic segmentation. In *AAAI*, 2022. 6

[53] Aoran Xiao, Jiaxing Huang, Weihao Xuan, Ruijie Ren, Kangcheng Liu, Dayan Guan, Abdulmotaleb El Saddik, Shijian Lu, and Eric P Xing. 3d semantic segmentation in the wild: Learning generalized models for adverse-condition point clouds. In *CVPR*, 2023. 1, 2, 6, 7

[54] Aoran Xiao, Xiaoqin Zhang, Ling Shao, and Shijian Lu. A survey of label-efficient deep learning for 3d point clouds. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2024. 1

[55] Jianyun Xu, Ruixiang Zhang, Jian Dou, Yushi Zhu, Jie Sun, and Shiliang Pu. Rpvnet: A deep and efficient range-point-voxel fusion network for lidar point cloud segmentation. In *CVPR*, 2021. 4, 8

[56] Xu Yan, Chaoda Zheng, Ying Xue, Zhen Li, Shuguang Cui, and Dengxin Dai. Benchmarking the robustness of lidar semantic segmentation models. *Int. J. Comput. Vis.*, pages 1–24, 2024. 1, 2

[57] Donglin Yang, Zhenfeng Liu, Wentao Jiang, Guohang Yan, Xing Gao, Botian Shi, Si Liu, and Xinyu Cai. Realistic rainy weather simulation for lidars in carla simulator. *arXiv preprint arXiv:2312.12772*, 2023. 1, 2

[58] Yu-Qi Yang, Yu-Xiao Guo, Jian-Yu Xiong, Yang Liu, Hao Pan, Peng-Shuai Wang, Xin Tong, and Baining Guo. Swin3d: A pretrained transformer backbone for 3d indoor scene understanding. *arXiv preprint arXiv:2304.06906*, 2023. 1

[59] Xufeng Yao, Yang Bai, Xinyun Zhang, Yuechen Zhang, Qi Sun, Ran Chen, Ruiyu Li, and Bei Yu. Pcl: Proxy-based contrastive learning for domain generalization. In *CVPR*, 2022. 7

[60] Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip HS Torr, and Vladlen Koltun. Point transformer. In *ICCV*, 2021. 2

[61] Haimei Zhao, Jing Zhang, Zhuo Chen, Shanshan Zhao, and Dacheng Tao. Unimix: Towards domain adaptive and gener-

alizable lidar semantic segmentation in adverse weather. In *CVPR*, 2024. 1, 2, 6, 7

[62] 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. In *ICCV*, 2021. 1

[63] Kaiyang Zhou, Ziwei Liu, Yu Qiao, Tao Xiang, and Chen Change Loy. Domain generalization: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(4):4396–4415, 2022. 1, 2

[64] Xinge Zhu, Hui Zhou, Tai Wang, Fangzhou Hong, Yuexin Ma, Wei Li, Hongsheng Li, and Dahua Lin. Cylindrical and asymmetrical 3d convolution networks for lidar segmentation. In *CVPR*, 2021. 2