ISPRS Journal of Photogrammetry and Remote Sensing 218 (2024) 204–219

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Weather-aware autopilot: Domain generalization for point cloud semantic segmentation in diverse weather scenarios ☆

Jing Du [a], John Zelek [a,*], Jonathan Li [a,b,*]

[a] Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[b] Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada

ARTICLE INFO

Keywords:
Point cloud
Semantic segmentation
Deep learning
Domain generalization
Diverse weather scenarios

ABSTRACT

3D point cloud semantic segmentation, a pivotal task in fields such as autonomous driving and urban planning, confronts the challenge of performance degradation under adverse weather conditions. Current methodologies primarily focus on optimal weather scenarios, leaving a significant gap in handling various environmental adversities like fog, rain, and snow. To bridge this gap, we propose a comprehensive deep learning framework featuring unique components — an Adaptive Feature Normalization Module (AFNM) for effective normalization and calibration of features, a Dual-Attention Fusion Module (DAFM) for integrating cross-domain features, and a Proxy Label Generation Module (PLGM) for generating reliable proxy labels within the domain. Utilizing the SemanticKITTI and SynLiDAR datasets as source domains and the SemanticSTF dataset as the target domain, our model has been rigorously evaluated under varying weather conditions. When trained on the SemanticKITTI dataset as the source domain with the SemanticSTF dataset as the target, our approach surpasses the current state-of-the-art models by a margin of 6.2% in terms of overall mean Intersection over Union (mIoU) scores. Similarly, with the SynLiDAR dataset as the source and SemanticSTF as the target, our performance exceeds the best existing models by 3.4% in mIoU. These results substantiate the efficacy of our model in advancing the field of 3D semantic segmentation under diverse weather conditions, showcasing its notable robustness and superiority. The code is available at https://github.com/J2DU/WADG-PointSeg.

1. Introduction

In the domain of computer vision, semantic segmentation of 3D point clouds is a fundamental field. This intricate task involves the precise segmentation of each point in the 3D scan into specific semantic classes, providing a comprehensive understanding of the environment (Guo et al., 2021). Its applications are extensive and diverse, ranging from assisting autonomous vehicles in making precise, real-time decisions to advancing urban planning with sophisticated remote sensing technologies. The integration of semantic segmentation with these varied applications highlights its significance in the advancement of smart technologies, autonomous systems, and sustainable urban development. The 3D semantic segmentation field has undergone a transformative evolution, driven by rapid strides in deep learning and the emergence of sophisticated 3D sensing technologies.

However, a review of commonly used semantic segmentation datasets such as S3DIS (Armeni et al., 2016), ScanNet (Dai et al., 2017), Semantic3D (Hackel et al., 2017), Toronto-3D (Tan et al., 2020), and SemanticKITTI (Behley et al., 2019) reveals a significant gap: these datasets do not account for different weather conditions such as fog, rain, and snow, which are crucial for developing robust, weather-aware segmentation models. This limitation is critical, as real-world environments are ever-changing, and autonomous systems must be reliable under all weather scenarios. Recognizing this need, the research community has made efforts to address this gap. The emergence of the SemanticSTF dataset (Xiao et al., 2023b) in 2023 represented a significant milestone, providing the first extensive adverse-weather benchmark for 3D semantic segmentation with high-quality, point-wise annotations across various weather conditions including dense fog, light fog, snow, and rain.

Despite this progress, several challenges persist. Most current deep learning models (Cortinhal et al., 2020; Zhu et al., 2021) are optimized in ideal weather scenarios, leading to notable performance degradation when confronted with real-world environmental variabilities such as fog, rain, or snow (Xiao et al., 2023b). These conditions change the fundamental characteristics of data, impacting visibility and density, and

☆ This work was supported in part by the National Natural Science Foundation of China, under Grant 41871380, the China Scholarship Council under Ph.D. Scholarship 202208350003, and the National Science and Research Council of Canada.

* Corresponding authors.
E-mail addresses: jzelek@uwaterloo.ca (J. Zelek), junli@uwaterloo.ca (J. Li).

https://doi.org/10.1016/j.isprsjprs.2024.09.006
Received 21 September 2023; Received in revised form 5 September 2024; Accepted 5 September 2024
Available online 17 September 2024

thus posing a significant challenge to the domain adaptability of these models. Consequently, the challenge is not just about achieving high accuracy in the source domain but ensuring consistent performance across these challenging conditions. Furthermore, while datasets like SemanticSTF have started to address this issue by providing benchmarks for adverse-weather conditions, the overall scarcity of diverse datasets that cover various weather scenarios still exacerbates the problem, as creating labeled datasets for training deep learning networks is both labor-intensive and time-consuming (Xiao et al., 2023a).

To address these challenges, our research is motivated by several key objectives. We aim to introduce a robust deep learning framework capable of withstanding the challenges posed by environmental variabilities in diverse weather scenarios, ultimately enhancing the practical applicability and reliability of 3D point cloud segmentation technologies. Additionally, we seek to design models that can be trained on a variety of detailed outdoor point cloud datasets, both real and synthetic, to enhance generalization capabilities. Finally, we intend to utilize existing datasets that encompass a broad range of semantic categories and environmental scenarios to mitigate the annotation bottleneck.

In order to achieve these objectives, we propose a domain generalization approach to realize these objectives. Domain generalization aims to train models on multiple source domains with the goal of generalizing well to unseen target domains. The proposed framework seeks to bridge the gap between controlled training environments and the varied complexities encountered in real-world settings. Through extensive experiments on benchmark datasets, the proposed model demonstrates the ability to maintain consistent performance across varying environmental conditions. By addressing these critical challenges, our research strives to advance the field of 3D semantic segmentation, contributing to the development of more robust and adaptable autonomous systems that can operate safely and reliably in dynamic and unpredictable environments.

The main contributions of our work can be summarized as follows:

- We propose a comprehensive deep learning framework that incorporates unique components such as the Adaptive Feature Normalization Module (AFNM), Dual-Attention Fusion Module (DAFM), and Proxy Label Generation Module (PLGM). These modules work in concert to enhance the model's stability and performance under diverse conditions, including various weather scenarios.
- We present a sophisticated, multi-faceted loss function that effectively directs the learning process. This innovative approach allows the model to efficiently leverage two types of inputs and proxy label learning, enhancing the model's generalization capabilities across unseen data.
- Through extensive experiments on benchmark datasets, our framework consistently demonstrates superior performance across a multitude of weather conditions. Specifically, in dense fog scenarios, utilizing the SemanticKITTI dataset as the source domain and the SemanticSTF dataset as the target, our approach outstrips the state-of-the-art model (Xiao et al., 2023b), achieving a significant 8.2% lead in performance. Similarly, under rainy conditions, when employing the SynLiDAR dataset as the source and the SemanticSTF as the target, our model surpasses the best existing method (Xiao et al., 2023b) with a remarkable 6.1% edge in mIoU.

2. Related work

2.1. 3D point cloud semantic segmentation

In the rapidly advancing field of 3D point cloud semantic segmentation, significant strides have been made (Wang and Yao, 2022; Liu et al., 2022; Tang and Cham, 2022). Innovations in deep learning (Wang and Yao, 2022; Liu et al., 2022) have markedly reduced the necessity for manual labeling while simultaneously improving accuracy. Further advancements in feature extraction (Wang et al., 2022a), annotation efficiency (Hu et al., 2022), and cross-domain segmentation (Yuan et al., 2023; Shuai and Liu, 2023) reflect the extensive progress within this area. The adoption of semi-supervised techniques (Li et al., 2023b) and the integration of multi-modal data (Li et al., 2023a) are significantly propelling the discipline forward, enhancing the utilization of resources. Moreover, the emerging developments in novel class discovery (Riz et al., 2023) and few-shot segmentation (He et al., 2023) hold substantial promise, potentially transforming applications in sectors such as autonomous driving and robotics.

The collection of large-scale, precisely-annotated training data for point clouds is notably time-consuming and costly, limiting dataset scalability and creating obstacles for the efficient utilization of point cloud data across diverse tasks and applications. Label-efficient learning emerges as a viable solution, facilitating efficient deep network training with significantly less annotation effort (Xiao et al., 2023c). In response to the challenge of labeling large-scale point clouds, a novel deep-learning framework (Wang and Yao, 2022) has been proposed, significantly reducing labeling efforts through the utilization of unlabeled data and the application of innovative prediction constraints, complemented by entropy regularization. Concurrently, recognizing the necessity for enhanced labeling efficiency in outdoor scenes, the LESS framework (Liu et al., 2022) adopts a comprehensive approach, incorporating geometry-aware sampling, prototype learning, multi-scale feature aggregation, and multi-scan distillation. This approach markedly diminishes manual labeling while simultaneously improving semantic segmentation and the utilization of point cloud data in outdoor settings. Annotator (Xie et al., 2024) introduces a groundbreaking active learning framework, addressing the high cost and labor intensity of annotating vast point clouds. This method innovates with a voxel-centric online selection strategy, utilizing a novel metric, voxel confusion degree, to efficiently identify and annotate key voxel grids based on their local topology and structure.

To address the increasing complexity and scale of point cloud datasets, several innovative solutions have been proposed. FPS-Net (Xiao et al., 2021) introduces an innovative convolutional fusion network designed for large-scale environments. Distinctively, it separates point attributes into modalities, processes them through a modality-specific encoder, and then fuses them using a high-dimensional feature space. Furthermore, addressing the demand for a scalable and robust solution to manage intricate, large-scale datasets, the Mask Point Transformer (Tang and Cham, 2022) has been developed. This innovation utilizes mask tokens and a transformer-based decoder to achieve enhanced, context-rich feature extraction and segmentation in point cloud data. Additionally, confronting the inadequacies of traditional 3D convolution kernels in handling the irregularities and sparsity of point clouds, Wang et al. (2022a) introduces a feature extraction method that bolsters the effectiveness of the spherical interpolated convolution network. This method integrates a spherical interpolated convolution operator with a self-learned distance-feature density metric, thereby capturing local geometric features more effectively. Hu et al. (2022) has innovated in large-scale point cloud segmentation with the Semantic Query Network (SQN), which combines query-based sampling with self-supervised learning. Additionally, addressing the challenges in capturing global structural knowledge and infusing semantic information into mid-level features, Wang et al. (2022b) unveils SemAffiNet. This approach, utilizing semantic-affine transformation and attention mechanisms, significantly improves accuracy and category comprehension in point cloud segmentation.

Confronting the challenges of domain gaps and data scarcity in outdoor point cloud cross-domain segmentation, a novel category-level adversarial framework (Yuan et al., 2023) has been introduced. This comprehensive framework includes a backbone network for feature extraction, a multi-scale domain conditioned block, and a category-level adversarial training module, collectively enhancing segmentation

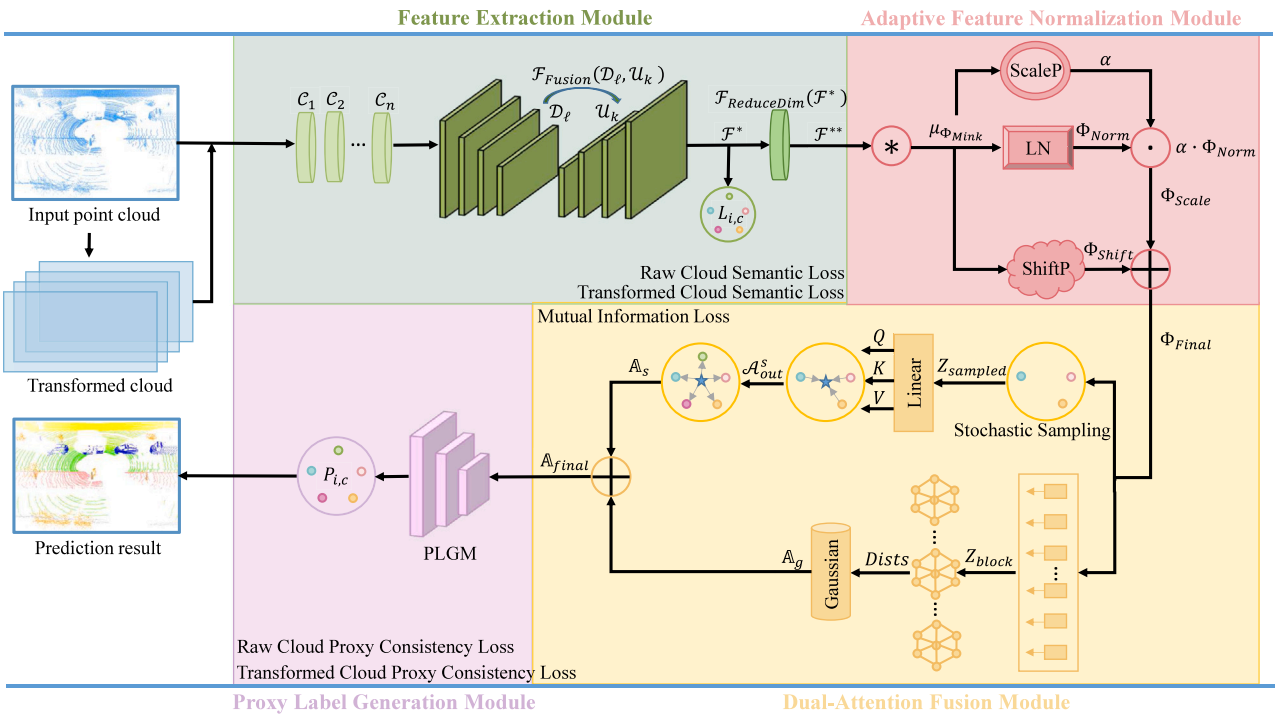

Fig. 1. Framework of the proposed approach. The symbol ⊛ represents the average operation. The symbol ⊙ indicates the multiplication operation. The symbol ⊕ signifies the summation operation. Additionally, within the framework, 'ScaleP' stands for Scale Predictor, 'LN' refers to Layer Normalization, and 'ShiftP' denotes Shift Predictor.

performance by aligning features across domains. DGLSS (Kim et al., 2023) innovatively employs a domain augmentation technique by subsampling LiDAR scans to simulate sparse unseen domains. It introduces two novel constraints: Sparsity Invariant Feature Consistency and Semantic Correlation Consistency, aiming to align features and maintain semantic correlations across domains. Lastly, 3DLabelProp (Sanchez et al., 2023) innovates in domain generalization for 3D semantic segmentation by leveraging the geometry and sequentiality of the LiDAR data. This method enhances model generalization by utilizing partially accumulated point clouds, allowing for effective label propagation and domain alignment across diverse environments.

2.2. Diverse weather scene understanding

Exploring diverse weather scene understanding, several studies have proposed innovative solutions to enhance autonomous vehicle perception and adaptation in challenging conditions. Zhang et al. (2023) comprehensively examines the impact of adverse weather conditions on autonomous vehicle perception and sensing, highlighting the challenges and advancements in sensor technology and machine learning. It discusses state-of-the-art solutions for weather-related perception issues, including sensor fusion, weather classification, and enhancement techniques.

In a related vein, Liao et al. (2022) introduces the Target-Domain driven pseudo label Diffusion (TDo-Dif) framework, an unsupervised domain adaptation (UDA) method for foggy scene understanding, which leverages spatial–temporal knowledge to diffuse sparse, confident pseudo labels across similar and temporally related pixels in sequential driving scene images. Utilizing superpixels and optical flows, the method enhances both the density and accuracy of pseudo labels through spatial similarity and temporal correspondence. Further expanding on domain adaptation strategies, Liao et al. (2023) proposes a Candidate labels disambiguation-based Domain Adaptation (CanDA) model for domain-adaptive semantic segmentation in foggy conditions, replacing traditional one-hot pseudo labels with a Candidate Label Set (CLS) to address labeling uncertainty. By employing a prototypical contrastive learning framework, CanDA refines pixel embeddings towards real semantics, progressively ranking ground truth labels higher within the CLS.

Additionally, Hoyer et al. (2023) proposes Masked Image Consistency (MIC), a novel module enhancing UDA by leveraging spatial context relations in target domains without annotations. Applicable across various tasks such as image classification, semantic segmentation, and object detection, MIC significantly boosts performance in challenging domain shifts like synthetic-to-real and clear-to-adverse weather conditions. PanopticRoad (Sakaino, 2023) introduces an innovative approach to panoptic road condition segmentation under adversarial conditions. This system enhances segmentation accuracy by refining failures in local road condition recognition through the synergistic application of weather and physical constraints. By combining techniques such as depth estimation, obstacle recognition, and condition classification, PanopticRoad significantly enhances the accuracy, stability, and robustness of panoptic segmentation across various adverse scenarios. Moreover, Kalwar et al. (2023) presents a Gated Differentiable Image Processing (GDIP) block designed to enhance object detection in autonomous vehicles under adverse weather and lighting conditions. Integrated with existing detection frameworks, GDIP leverages a unique gating mechanism to weight outputs of various image processing techniques, directly informed by object detection loss. Demonstrated on synthetic and real-world datasets, GDIP significantly advances object detection under challenging conditions like fog and low light.

3. Method

3.1. Method overview

The proposed methodology integrates four key modules designed to process and segment 3D point cloud data efficiently. This integration is comprehensively depicted in Fig. 1. Input data comprises original point clouds and their transformed counterparts, enhanced through operations like rotation, scaling, and flipping, ensuring the model's resilience and accuracy across varied environmental settings. At the core of this approach is the MinkowskiNet-based feature extraction module, adept

at handling sparse 3D data with remarkable precision. Building on this, the Adaptive Feature Normalization Module (AFNM) skillfully regulates the scale and variance of features, promoting model stability and consistent performance. The model's adaptability is further fortified by the incorporation of a Dual-Attention Fusion Module (DAFM), which integrates self-attention mechanisms with both stochastic sampling and Gaussian block attention to achieve a robust refinement of features. To effectively tackle the complexities of domain generalization, the Proxy Label Generation Module (PLGM) is utilized, adeptly generating high-confidence proxy labels. This module plays a pivotal role in enabling the model to adapt and perform accurately across diverse domain shifts. The method culminates in a multifaceted loss function that harmonizes various learning paradigms, offering an intricate balance of robustness and versatility in interpreting 3D point cloud data, thereby enhancing the overall performance in semantic segmentation tasks.

3.2. Feature extraction module

In the field of 3D point cloud semantic segmentation, especially under diverse weather conditions like snowy, dense-foggy, light-foggy, and rainy scenarios, the challenges are manifold. These include the inherent sparsity and uneven distribution of point cloud data, compounded by visibility and clarity issues in adverse weather. To effectively confront these challenges, we employ MinkowskiNet (Choy et al., 2019) as the core feature extraction module. It begins with an array of initial convolutional layers designed to make the first critical contact with the input data, denoted as $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$. These layers play a foundational role in generating the initial feature maps and spotting elementary patterns from point clouds, a task that becomes significantly more challenging in adverse weather conditions due to issues like reduced visibility and increased data sparsity. As the data moves through the network, it encounters a series of transformational blocks that incrementally increase the complexity of feature representations. This progression is essential for adapting to the varied and unpredictable nature of environmental changes, ensuring that key details in the point cloud are preserved and accurately represented. In the latter stages, the integration of up-sampling and feature fusion processes enhances the accuracy of segmentation by synergizing high-level abstract features with detailed low-level data. This structured approach to feature extraction is critical for the network to perform reliable and precise semantic segmentation across different weather scenarios.

The operation of these initial convolutional layers can be denoted as:

$$C_n = \mathcal{F}_{\text{Conv}_n}(\mathcal{X}), \tag{1}$$

where $\mathcal{F}_{\text{Conv}_n}(\cdot)$ is the operation performed by the nth convolutional layer. The network then leads into a series of transformational blocks. Each block is capable of systematically augmenting the complexity of the feature representations and simultaneously reducing the data's spatial resolution. These blocks are structured with residual connections alongside initial layers, effectively mitigating the vanishing gradient problem and enabling the network to learn complex features as the depth increases. The operation of the transformation block can be denoted as:

$$\mathcal{B}_m = \mathcal{F}_{\text{Block}_m}(C_n), \tag{2}$$

where $\mathcal{F}_{\text{Block}_m}(\cdot)$ denotes the operation performed by the mth block.

The output of each block undergoes a downsampling layer, which can be represented as:

$$\mathcal{D}_l = \mathcal{F}_{\text{DS}_l}(\mathcal{B}_m), \tag{3}$$

where $\mathcal{F}_{\text{DS}_l}(\cdot)$ represents the downsampling operation performed by the lth layer. Upon transformation, the network incorporates a set of up-sampling modules that aim to recuperate the spatial resolution that was reduced during the downsampling process. These modules serve a dual purpose: they not only up-sample the feature maps but also conduct feature fusion by integrating high-level abstract features with low-level detailed features from previous layers. This fusion process results in a richer feature representation, thereby enhancing the accuracy of the segmentation process. The up-sampling operation can be represented as:

$$\mathcal{U}_k = \mathcal{F}_{\text{US}_k}(\mathcal{D}_l), \tag{4}$$

where $\mathcal{F}_{\text{US}_k}(\cdot)$ represents the up-sampling operation performed by the kth module. The feature fusion operation can be represented as:

$$\mathcal{F}^* = \mathcal{F}_{\text{Fusion}}(\mathcal{D}_l, \mathcal{U}_k). \tag{5}$$

The feature maps output from the up-sampling modules are then directed to a decision-making layer. This layer plays a pivotal role in assigning class labels to each point in the point cloud data, translating the complex high-level features into discrete class predictions necessary for the semantic segmentation tasks. This process can be represented as:

$$L_i = \mathcal{F}_{\text{Classify}}(\mathcal{F}^*[i]), \tag{6}$$

where $\mathcal{F}_{\text{Classify}}(\cdot)$ represents the classification operation and $\mathcal{F}^*[i]$ is the feature vector corresponding to the ith point.

Lastly, the MinkowskiNet integrates a dimensionality reduction module. This module is vital for tasks such as feature visualization and clustering. It maps high-dimensional features into a lower-dimensional space, thereby easing the interpretation and understanding of the learned feature representations. This can be represented as:

$$\mathcal{F}^{**} = \mathcal{F}_{\text{ReduceDim}}(\mathcal{F}^*), \tag{7}$$

where $\mathcal{F}_{\text{ReduceDim}}(\cdot)$ represents the dimensionality reduction operation and $\mathcal{F}^{**}$ is the reduced-dimension feature map.

3.3. Adaptive feature normalization module

In addressing the complex challenges of 3D point cloud semantic segmentation under varying weather conditions, the Adaptive Feature Normalization Module (AFNM) plays a pivotal role. This module is specifically designed to refine the feature representations extracted by MinkowskiNet, denoted as Φ_{Mink}, ensuring they are robust and generalizable. The AFNM effectively tackles the internal covariate shift, a prevalent issue in deep learning architectures, particularly critical when handling the complex and variable nature of point cloud data influenced by adverse weather conditions. By normalizing the feature data on a per-channel basis and adaptively scaling it, the AFNM significantly enhances the stability and performance of our model. This process is crucial for maintaining the integrity of feature representations despite the fluctuating visibility and sparsity induced by diverse weather conditions. The ability to dynamically adjust the scale and shift of features allows the model to more effectively interpret and segment point clouds across various environmental contexts. Therefore, the AFNM serves as a targeted solution to the specific challenges posed by weather-influenced 3D point cloud segmentation.

The AFNM operates on a per-channel basis and starts by computing the mean of the incoming features, thereby getting a measure of the central tendency of the features. The mean of the features is computed as:

$$\mu_{\Phi_{\text{Mink}}} = \frac{1}{N}\sum_{i=1}^{N} \Phi_{\text{Mink}}[i]. \tag{8}$$

Subsequently, Layer Normalization is applied to the features, normalizing the statistics of the features within each sample. This normalization step ensures that the learned features are not overly dependent on the scale and variance of the input data, thereby enhancing the

model's stability and performance. The normalized features, denoted as Φ_{Norm}, are obtained as:

$$\Phi_{\text{Norm}}[i] = \frac{\Phi_{\text{Mink}}[i] - \mu_{\Phi_{\text{Mink}}}}{\sqrt{Var(\Phi_{\text{Mink}})}}. \tag{9}$$

The module then predicts scale and shift factors to rescale the normalized features adaptively. The scale factor, denoted as α, is predicted by the Scale Predictor, a sub-module within the AFNM that consists of two linear layers with a Layer Normalization and a ReLU activation in between. This predictor takes the mean of the input features, applies a sigmoid activation to ensure the scale factor remains positive, and multiplies it with a learnable parameter gamma, denoted as γ. The operation of the Scale Predictor can be represented as:

$$\alpha = \gamma \cdot Sigmoid(ScalePredictor(\mu_{\Phi_{\text{Mink}}})). \tag{10}$$

The scaled features, denoted as Φ_{Scale}, are obtained by multiplying Φ_{Norm} with the scale factor α:

$$\Phi_{\text{Scale}} = \alpha \cdot \Phi_{\text{Norm}}. \tag{11}$$

The shift factor, denoted as β, is predicted by the Shift Predictor, another sub-module within the AFNM. This predictor follows a similar architecture to the Scale Predictor, with the addition of an extra linear layer to reduce the feature dimension by half. The operation of the Shift Predictor can be represented as:

$$\Phi_{\text{Shift}} = \beta \cdot ShiftPredictor(\mu_{\Phi_{\text{Mink}}}). \tag{12}$$

The final output of the AFNM, denoted as Φ_{Final}, is obtained by adding the shift features Φ_{Shift} to the scaled features Φ_{Scale}:

$$\Phi_{\text{Final}} = \Phi_{\text{Scale}} + \Phi_{\text{Shift}}. \tag{13}$$

In summary, the AFNM enhances the network's ability to extract meaningful features by adaptively normalizing and scaling the output from the MinkowskiNet. By reducing the internal covariate shift and allowing the network to restore the original feature statistics adaptively, the AFNM contributes significantly to the robustness and generalization capabilities of the proposed network.

3.4. Dual-attention fusion module

The Dual-Attention Fusion Module (DAFM) merges stochastic sampling-based self-attention and Gaussian block self-attention to effectively address the challenges of analyzing point cloud data within a domain generalization framework. The stochastic sampling technique targets the computational intensity typically associated with processing vast, sparse datasets characteristic of diverse weather scenarios. This innovative approach achieves a critical balance between computational efficiency and data fidelity, ensuring high-quality segmentation despite environmental variability. Furthermore, the design of the stochastic sampling-based self-attention mechanism is particularly suitable for the domain generalization paradigm. It allows for a comprehensive understanding of the global context of the data, facilitating the recognition of broad patterns and structures crucial in scenarios where training and testing environments differ significantly.

Complementing this, the Gaussian block self-attention zeroes in on local contextual nuances within segmented blocks of point cloud data. This localized approach is particularly effective in adverse weather, where diminished visibility and clarity can obscure critical data features. By applying Gaussian filters, the DAFM enhances the model's capacity to discern fine details, critical for precise segmentation in challenging conditions. The combination of these two attention strategies within the DAFM fosters a robust and versatile module. It is instrumental for nuanced and accurate interpretation of point clouds across varying weather conditions. Consequently, the DAFM emerges as a key component in addressing the challenges of domain generalization for semantic segmentation of 3D point clouds in varied weather conditions.

3.4.1. Self-attention with stochastic sampling

Given an input matrix $Z \in \mathbf{R}^{N \times d}$, where N is the number of points and d is the dimensionality of the feature space, we propose a stochastic sampling strategy to address the high computational demands of traditional self-attention mechanisms. By randomly selecting a subset of size n:

$$Z_{\text{sampled}} = Z[\text{RandomIndices}(N, n)], \tag{14}$$

where n is the size of the sampled subset, we achieve significant computational efficiency without substantial loss of information.

The Query, Key, and Value transformations are obtained using learnable linear transformations:

$$\begin{aligned} Q &= W_q \cdot Z_{\text{sampled}}, \\ K &= W_k \cdot Z_{\text{sampled}}, \\ V &= W_v \cdot Z_{\text{sampled}}, \end{aligned} \tag{15}$$

where W_q, W_k, and W_v are the transformation matrices. The attention weights are then computed as:

$$\mathcal{A}_w^s = \text{softmax}\left(\frac{Q \cdot K^T}{\sqrt{d}}\right). \tag{16}$$

The softmax ensures that the weights are normalized. The attention output is given by:

$$\mathcal{A}_{out}^s = \mathcal{A}_w^s \cdot V. \tag{17}$$

The final output for each point is obtained by averaging the attention output across the sampled dimension and replicating it across all points:

$$\mathbb{A}_s = \text{mean}(\mathcal{A}_{out}^s, \text{dim} = 0) \uparrow N, \tag{18}$$

where $\mathbb{A}_s$ is the final output matrix with the same averaged attention output repeated N times, and $\text{mean}(\mathcal{A}_{out}^s, \text{dim} = 0)$ represents the mean of the attention output across the specified dimension, resulting in a 2D tensor with one row. This tensor is then broadcast across the N points via the $\uparrow$ operation, defined here as replicating the mean vector to form an N-row matrix where each row is identical to the mean vector, thereby assuring that each point in the final output matrix $\mathbb{A}_s$ is assigned the same global attention value. This operation effectively aligns the attention mechanism's focus uniformly across the point cloud.

This approach significantly reduces the complexity from $O(N^2)$ to $O(n^2)$, where $n \ll N$, achieving computational efficiency. Despite the stochastic sampling, the average operation ensures that the original relationships between the features are preserved. This method skillfully balances between efficiency and integrity, rendering it suitable for various applications where both speed and fidelity are required.

3.4.2. Self-attention with Gaussian block attention

For the second strategy, we segment the input matrix into blocks, and apply Gaussian attention within each block. This approach enhances the local context modeling within the data. The input matrix is segmented into blocks of size b:

$$Z_{\text{block}} = Z_{\text{start\_idx:end\_idx}}, \tag{19}$$

where start_idx and end_idx define the block boundaries. The squared Euclidean distances within a block are computed:

$$\text{Dists}_{i,j} = \left\| Z_{\text{block},i} - Z_{\text{block},j} \right\|_2^2, \tag{20}$$

where $\text{Dists}_{i,j}$ denotes the squared Euclidean distance between the ith and jth data points within the block, and $Z_{\text{block},i}$ and $Z_{\text{block},j}$ represent the corresponding data points.

Next, the Gaussian attention weights are computed based on these distances:

$$\mathcal{A}_w^g = \exp\left(-\frac{\text{Dists}}{2 \cdot d}\right), \tag{21}$$

where $\mathcal{A}_w^g$ denotes the attention weights within the block, and d is the dimensionality of the data, controlling the scale of the Gaussian function. The final attention output within the block is:

$$\mathcal{A}_{out}^g = \mathcal{A}_w^g \cdot Z_{\text{block}}. \tag{22}$$

The overall attention output across all blocks is:

$$\mathbb{A}_g = \text{Aggregate}\left(\left\{ \frac{\mathcal{A}_{\text{out},i}^g}{b} \right\}_{i=1}^{m} \right), \tag{23}$$

where m is the total number of blocks, the aggregate function denotes the process of combining these normalized block-wise attention outputs into a single comprehensive attention representation.

This approach inherently takes into account the similarity between points, ensuring that closer points have higher attention weights. The block-based processing allows handling of larger data efficiently, as the computational requirements are determined by the block size rather than the entire data size.

3.4.3. Attention fusion module

We introduce an attention fusion module that combines the two aforementioned attention strategies. The final output integrates the stochastic sampling-based and Gaussian block attention outputs:

$$\mathbb{A}_{final} = w_1 \cdot \mathbb{A}_s + w_2 \cdot \mathbb{A}_g, \tag{24}$$

where w_1 and w_2 are learnable parameters controlling the fusion of the two self-attention strategies.

The fusion of both self-attention mechanisms offers a synergistic blend of advantages, maximizing computational efficiency without compromising the essential data dependencies. By employing both the random sampling and the block-based processing with Gaussian attention, this comprehensive methodology provides substantial applicability across various research domains.

3.5. Proxy label generation module

The Proxy Label Generation Module (PLGM) utilizes the nuanced, context-rich feature representations processed by the DAFM, denoted as Φ_{DAFM}. Unlike traditional methods in transfer learning, This module operates as a transformative bridge, converting the intricate feature space into a semantic label space, achieved through a fully-connected layer adept at mapping the interplay between features and their corresponding semantic classes.

$$\mathbf{L} = \mathbf{F}(\Phi_{\text{DAFM}}), \tag{25}$$

where $\mathbf{L}$ denotes the generated proxy labels, and $\mathbf{F}$ represents the fully-connected layer.

In response to the intrinsic complexities of domain generalization, the PLGM plays a crucial role in bridging the gap between domain-specific variations and a unified segmentation model. By generating high-confidence proxy labels, the PLGM enables a self-supervised learning paradigm, integral to training models in scenarios where direct domain correlation is challenging. This approach not only enhances the model's ability to generalize across various domains but also establishes a feedback loop within the model. This loop facilitates continuous improvement and refinement of the model's understanding of data, thereby addressing the limitations commonly faced in domain generalization, such as the variance in data distribution and environmental conditions.

In summary, the PLGM effectively integrates feature representations from the DAFM with a robust mechanism for proxy label generation. This capability allows the model to learn from the inherent structures and patterns of the input data, an essential factor for adapting to the fluctuating visibility and clarity nature of point cloud data in adverse weather conditions.

3.6. Loss function

In our proposed model, we incorporate a comprehensive and nuanced loss function structure that utilizes various aspects of learning paradigms. This holistic approach ensures the robustness and versatility of the model when interpreting 3D point cloud data. The multifaceted loss structure is composed of five harmonized components: Raw Cloud Semantic Loss (RCSL), Transformed Cloud Semantic Loss (TCSL), Raw Cloud Proxy Consistency Loss (RCPCL), Transformed Cloud Proxy Consistency Loss (TCPCL), and Mutual Information Loss (MIL). The locations where each element operates within the overall learning process are illustrated in Fig. 1.

Raw Cloud Semantic Loss: The RCSL is instrumental in processing the raw 3D point cloud data. Represented by $\mathcal{L}_{RCSL}$, this loss function employs the conventional cross-entropy metric to quantitatively assess the congruence between the predicted semantic labels $L_{i,c}$ and their corresponding ground truths $G_{i,c}$. RCSL is integral to the system's analytical accuracy, ensuring precise and reliable interpretations of the original point cloud. This loss function contributes significantly to the model's robustness in segmenting complex raw point cloud structures and enhances overall performance in semantic analysis. Its mathematical expression is as follows:

$$\mathcal{L}_{RCSL} = -\sum_i \sum_c G_{i,c} \log(L_{i,c}). \tag{26}$$

Transformed Cloud Semantic Loss: The TCSL, denoted as $\mathcal{L}_{TCSL}$, is also derived from the cross-entropy loss function. It quantifies discrepancies between predictions and ground truth for the transformed point cloud input, denoted as $L'_{i,c}$ and $G'_{i,c}$ respectively. By processing the transformed point cloud data, the TCSL enables the model to assimilate and learn from varied data representations. This approach significantly enhances the model's environmental perception, allowing for a more comprehensive understanding of the spatial information within point clouds. The mathematical formulation of the TCSL is as follows:

$$\mathcal{L}_{TCSL} = -\sum_i \sum_c G'_{i,c} \log(L'_{i,c}). \tag{27}$$

Raw Cloud Proxy Consistency Loss: The RCPCL, represented by $\mathcal{L}_{RCPCL}$, evaluates the discrepancy between the proxy labels, generated from the raw 3D point cloud data, denoted as $P_{i,c}$, and the corresponding ground truth labels $G_{i,c}$. The optimization of RCPCL aims to refine the accuracy of proxy labels, aligning them more closely with the actual labels, thus enhancing the functionality of the PLGM. This loss is crucial for ensuring the fidelity of the model's predictions in relation to the original point cloud structures. The mathematical expression for the RCPCL is:

$$\mathcal{L}_{RCPCL} = -\sum_i \sum_c G_{i,c} \log(P_{i,c}). \tag{28}$$

Transformed Cloud Proxy Consistency Loss: The TC-PCL, indicated as $\mathcal{L}_{TCPCL}$, evaluates the discrepancy between the proxy labels generated from the transformed point cloud data, denoted as $P'_{i,c}$, and the corresponding ground truth labels $G'_{i,c}$. This loss function plays a pivotal role in ensuring that the proxy labels derived from the transformed point cloud input align accurately with the ground truth, thereby providing a crucial dimension of feedback for the learning process. The TCPCL, enhancing the model's adaptability to diverse data transformations, is mathematically expressed as:

$$\mathcal{L}_{TCPCL} = -\sum_i \sum_c G'_{i,c} \log(P'_{i,c}). \tag{29}$$

Mutual Information Loss: To ensure the extracted features from both raw point cloud and transformed point cloud inputs encapsulate shared semantic information, a Mutual Information Loss, indicated as $\mathcal{L}_{MIL}$, is utilized. This loss assesses the commonality in the feature representations derived from the raw point cloud, $\mathbf{F}$, and the transformed point cloud, $\mathbf{F}'$. Minimizing this loss fosters a learning environment

Table 1
3D semantic segmentation results (%) with SemanticKITTI as source and SemanticSTF as target.

Method	car	bicycle	motorcycle	truck	other-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic-sign	D-fog	L-fog	Rain	Snow	mIoU
ADDA (Tzeng et al., 2017)	65.6	0.0	0.0	21.0	1.3	2.8	1.3	16.7	64.7	1.2	35.4	0.0	66.5	41.8	57.2	32.6	42.2	23.3	26.4	31.5	27.9	27.4	23.4	26.3
Ent-Min (Vu et al., 2019)	69.2	0.0	10.1	31.0	5.3	2.8	2.6	0.0	65.9	2.6	35.7	0.0	72.5	42.8	52.4	32.5	44.7	24.7	21.1	31.4	28.6	30.3	24.9	27.2
Self-training (Zou et al., 2019)	71.5	0.0	10.3	33.1	7.4	5.9	1.3	0.0	65.1	6.5	36.6	0.0	67.8	41.3	51.7	32.9	42.9	25.1	25.0	31.8	29.3	27.9	25.1	27.6
CoSMix (Saltori et al., 2022)	65.0	1.7	22.1	25.2	7.7	33.2	0.0	0.0	64.7	11.5	31.1	0.9	62.5	37.8	44.6	30.5	41.1	**30.9**	**28.6**	31.6	30.3	**33.1**	**32.9**	28.4
Baseline	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
Dropout (Srivastava et al., 2014)	62.1	0.0	15.5	3.0	**11.5**	5.4	2.0	0.0	58.4	12.8	26.7	1.1	72.1	43.6	52.9	**34.2**	43.5	28.4	15.5	29.3	25.6	29.4	24.8	25.7
Perturbation	**74.4**	0.0	0.0	23.3	0.6	19.7	0.0	0.0	60.3	10.8	33.9	0.7	72.0	45.2	58.7	17.5	42.4	22.1	9.7	26.3	27.8	30.0	24.5	25.9
PolarMix (Xiao et al., 2022a)	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 (Li et al., 2018)	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 (Yao et al., 2022)	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 (Xiao et al., 2023b)	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
Ours	72.0	0.0	**32.9**	**37.0**	1.9	**37.7**	**6.8**	**52.9**	59.9	10.7	31.8	**2.2**	**76.0**	**48.8**	**62.7**	34.0	**49.3**	23.6	20.4	**39.5**	**32.5**	31.7	29.4	**34.8**

where the model is encouraged to capture consistent and shared information across these different input modalities, thus bolstering its ability to generalize. The MIL is mathematically expressed as:

$$\mathcal{L}_{MIL} = -\sum_{f\in\mathbf{F}}\sum_{f'\in\mathbf{F}'} p(f,f')\log\left(\frac{p(f,f')}{p(f)p(f')}\right). \tag{30}$$

The final loss function is a weighted sum of the five loss components, denoted as $\mathcal{L}_{total}$. The flexibility in determining these weights, denoted as α_i, provides control over the importance of each learning objective. It can be represented as:

$$\begin{aligned}\mathcal{L}_{total} =& \alpha_1\mathcal{L}_{RCSL} + \alpha_2\mathcal{L}_{TCSL} + \alpha_3\mathcal{L}_{RCPCL} + \\ & \alpha_4\mathcal{L}_{TCPCL} + \alpha_5\mathcal{L}_{MIL}.\end{aligned} \tag{31}$$

This multifaceted loss structure not only ensures the model's robust performance on the semantic segmentation task but also leverages transformed inputs and proxy label learning, while maintaining the harmony between different input perspectives. Therefore, this comprehensive loss structure significantly contributes to our model's superior performance on 3D point cloud semantic segmentation tasks.

4. Experiments

4.1. Dataset

The SemanticKITTI dataset (Behley et al., 2019), grounded in the KITTI Vision Benchmark's odometry set, includes diverse scenes from inner-city traffic to countryside roads in Karlsruhe, Germany. It provides 23,201 full 3D training scans and 20,351 testing scans, making it the largest publicly accessible dataset of its kind. The dataset features point cloud data from a commonly used automotive LiDAR, the Velodyne HDL64E, allowing research into the impact of aggregating sequential scans on semantic segmentation performance and moving object recognition. The 28 annotated classes overlap with Mapillary Vistas (Neuhold et al., 2017) and scapes datasets (Cordts et al., 2016) to suit the sparsity and vertical field-of-view inherent to point cloud data. Around 1700 man-hours were invested to ensure accurate and high-quality annotations.

The SynLiDAR dataset (Xiao et al., 2022b), created with Unreal Engine 4, offers a detailed collection of LiDAR point cloud sequences from diverse virtual outdoor scenarios, such as urban areas, small towns, and harbors. The dataset comprises 198,396 scans with a total of 19 billion points, averaging 98,000 points per scan. It provides accurate coordinates and precise point-wise annotations across 32 semantic classes, contributing to an in-depth understanding of each scene. With its substantial scale and variety of semantic classes, SynLiDAR is an invaluable resource for research in synthetic-to-real point cloud transfer learning. To overcome the challenge of simulating intensity in LiDAR point clouds, SynLiDAR incorporates a model trained on real-world LiDAR data, thereby ensuring a comprehensive, realistic dataset.

SemanticSTF (Xiao et al., 2023b) is an expansive adverse-weather 3D semantic segmentation benchmark, meticulously created from the STF dataset (Bijelic et al., 2020). It comprises 2,076 LiDAR scans, carefully selected from diverse geographical locations such as Germany, Sweden, Denmark, and Finland to minimize data redundancy. The scans include diverse weather conditions with 694 snowy, 637 dense-foggy, 631 light-foggy, and 114 rainy scenarios. Manual, highly detailed point-wise annotations were painstakingly applied, with a laborious over 6,600 man-hours dedicated to quality control, verification, and correction. The dataset, divided into 1,326 training, 250 validating, and 500 testing scans, is a pivotal resource for the development of robust 3D semantic segmentation models designed to perform reliably across all weather conditions.

4.2. Evaluation and visualization results

4.2.1. Evaluation results: from SemanticKITTI to SemanticSTF

Table 1 presents the performance of our proposed network in 3D semantic segmentation tasks, using SemanticKITTI as the source and SemanticSTF as the target. The results are compared with several other state-of-the-art methods. Our model clearly outperforms other methods in terms of mIoU, scoring a remarkable 34.8%. This suggests that, on average, our model has a superior capability to correctly segment points across all categories. Drilling down to individual categories, our model demonstrates particularly strong performance in "motorcycle", "truck", "person", "motorcyclist", and "terrain" categories, where it outperforms all other models. For instance, in the "truck" category, our model achieves an impressive accuracy of 37.0%, significantly higher than the nearest network, while in the "terrain" category, our model reaches a high score of 49.3%, exceeding all other models. The high precision in these categories can be attributed to the distinctive features of our proposed network, which include specialized convolutional structures, sophisticated attention mechanisms, and advanced fusion modules. These features enhance the network's ability to discern and correctly segment these larger objects, which tend to have consistent shape and size characteristics. However, there are categories where our model's performance could be improved, such as "pole" and "traffic-sign". These categories often include smaller and more variably shaped objects, which present a unique set of challenges for 3D semantic segmentation. Nevertheless, the overall high performance of our model suggests its effectiveness across diverse segmentation tasks. Future research could focus on improving performance in these more challenging categories, potentially through implementing additional attention mechanisms, or introducing novel methods for handling complex object categories.

Leveraging SemanticKITTI dataset as the source and SemanticSTF dataset as the target, we exhibit the model's performance under varying environmental adversities. These conditions, including dense fog, light fog, rain, and snow, present significant challenges to model accuracy due to reduced visibility and increased noise in the data. The proposed model demonstrates notable performance across a range of adverse weather conditions, surpassing other leading semantic segmentation models in mIoU scores. Beginning with dense fog, a condition that typically plummets model performance, our model achieves a remarkable mIoU of 39.5% as shown in Table 2. It should be noted that the

Table 2
3D semantic segmentation results (%) with SemanticKITTI as source and SemanticSTF as target (dense fog).

Method	car	bicycle	motorcycle	truck	other-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic-sign	mIoU
Baseline	74.7	–	–	7.8	0.0	6.4	8.9	0.0	72.2	0.6	**33.8**	0.0	59.6	48.7	56.9	27.4	56.4	27.2	21.1	29.5
Dropout (Srivastava et al., 2014)	67.5	–	–	1.9	0.0	8.9	2.8	0.0	70.9	5.6	29.0	0.8	64.6	44.0	60.0	31.6	60.6	**28.1**	21.3	29.3
Perturbation	68.6	–	–	8.8	0.0	6.0	0.0	0.0	66.6	**14.8**	24.3	0.1	52.2	43.5	60.1	19.4	54.1	16.3	11.5	26.3
PolarMix (Xiao et al., 2022a)	52.3	–	–	17.2	0.0	3.6	0.0	19.3	75.2	0.0	28.7	0.6	62.4	49.5	60.5	29.0	55.4	20.8	30.7	29.7
MMD (Li et al., 2018)	75.5	–	–	0.3	0.0	4.2	0.0	0.0	**75.4**	11.2	33.6	0.5	64.8	**51.7**	64.7	26.1	62.3	23.0	23.0	30.4
PCL (Yao et al., 2022)	64.3	–	–	11.7	0.0	0.6	0.0	0.0	72.4	3.8	31.3	0.8	63.1	46.5	65.7	19.4	**64.3**	18.5	28.9	28.9
PointDR (Xiao et al., 2023b)	69.2	–	–	7.1	0.0	2.4	6.7	0.0	73.5	8.5	33.6	0.2	**65.6**	47.6	63.6	31.0	60.7	24.4	**38.8**	31.3
Ours	**78.8**	–	–	**40.6**	0.0	**7.0**	**47.7**	**67.3**	71.6	0.3	29.8	**1.4**	65.0	48.0	**66.5**	**33.0**	60.8	21.3	33.0	**39.5**

Table 3
3D semantic segmentation results (%) with SemanticKITTI as source and SemanticSTF as target (light fog).

Method	car	bicycle	motorcycle	truck	other-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic-sign	mIoU
Baseline	60.0	0.0	00	1.3	10.9	12.3	0.0	0.0	68.6	4.5	36.0	0.0	61.5	53.1	55.6	38.0	44.7	**29.2**	18.2	26.0
Dropout (Srivastava et al., 2014)	63.2	0.0	0.0	3.2	10.2	5.5	0.0	0.0	63.8	4.9	29.4	0.1	62.5	53.1	58.6	42.5	46.6	27.8	14.3	25.6
Perturbation	**76.6**	0.0	0.0	38.2	0.0	21.9	0.0	0.0	66.6	**8.8**	34.6	0.1	62.4	56.1	63.2	25.3	46.2	22.4	6.5	27.8
PolarMix (Xiao et al., 2022a)	42.6	**0.2**	0.0	29.4	3.3	17.0	0.0	0.2	69.8	0.7	33.1	0.1	56.2	56.3	54.9	24.7	44.8	24.1	16.6	25.0
MMD (Li et al., 2018)	63.6	0.0	0.0	0.1	**13.3**	25.9	0.0	0.0	**73.9**	5.6	42.8	0.1	64.1	55.3	61.9	36.6	50.7	**29.2**	9.9	28.1
PCL (Yao et al., 2022)	66.3	0.0	0.0	26.7	0.2	8.7	0.0	0.0	67.8	5.0	36.7	0.4	64.3	58.0	66.1	21.2	53.1	25.5	**24.6**	27.6
PointDR (Xiao et al., 2023b)	65.9	0.0	0.0	29.7	4.4	11.4	**0.9**	0.0	70.9	**8.8**	**43.3**	0.0	66.5	55.1	61.3	43.0	49.1	29.1	24.3	29.7
Ours	68.5	0.0	0.0	**54.0**	4.8	**43.8**	0.0	0.0	69.1	4.3	34.3	**0.5**	**68.4**	**58.6**	**66.6**	**43.8**	**55.3**	24.2	20.7	**32.5**

Table 4
3D semantic segmentation results (%) with SemanticKITTI as source and SemanticSTF as target (rain).

Method	car	bicycle	motorcycle	truck	other-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic-sign	mIoU
Baseline	72.4	0.0	–	0.0	16.3	6.9	0.0	–	71.6	12.7	58.1	0.0	70.0	33.0	51.8	9.9	24.2	33.3	22.9	28.4
Dropout (Srivastava et al., 2014)	81.3	0.0	–	0.0	**21.2**	5.6	0.0	–	62.2	11.8	44.8	0.6	76.8	44.7	56.0	16.3	23.3	32.8	22.2	29.4
Perturbation	83.9	0.0	–	2.4	0.0	20.9	0.0	–	73.2	12.6	54.7	**7.0**	71.7	43.2	58.3	5.9	29.4	29.4	16.9	30.0
PolarMix (Xiao et al., 2022a)	56.7	**4.0**	–	**9.1**	1.5	29.8	0.0	–	68.2	10.9	50.2	0.5	73.2	**47.2**	48.3	17.8	22.3	32.3	14.1	28.6
MMD (Li et al., 2018)	83.9	0.0	–	0.0	8.9	31.6	0.0	–	**77.9**	17.9	**60.2**	0.3	69.6	39.3	58.4	14.1	**32.5**	34.0	30.0	**32.8**
PCL (Yao et al., 2022)	84.2	0.0	–	0.0	0.1	4.3	0.0	–	68.1	10.9	55.5	4.6	74.7	43.9	59.6	5.8	27.3	34.2	**38.8**	30.1
PointDR (Xiao et al., 2023b)	78.0	0.0	–	0.0	13.8	20.0	0.0	–	72.1	14.7	60.0	1.2	76.1	36.9	58.0	**18.3**	24.7	**36.1**	32.5	31.9
Ours	**84.3**	0.0	–	0.4	0.0	**40.6**	0.0	–	65.1	**19.2**	51.3	4.2	**77.3**	42.7	**61.3**	15.7	22.8	30.9	23.5	31.7

symbol '–' indicates the absence of instances captured under dense fog conditions. This is a clear indication of the model's robustness in managing challenging low-visibility situations. Despite the success, there are areas where performance can be improved, specifically in the "other-vehicle" and "other-ground" categories, which presented lower mIoUs.

As the conditions shift to light fog, our model maintains its exceptional performance, achieving an mIoU of 32.5%, as detailed in Table 3. This consistent performance in light fog conditions exemplifies the model's robustness and adaptability, highlighting its superior resilience in adverse conditions. Under rainy conditions, a weather scenario known for inducing substantial noise and visibility challenges, our model exhibits remarkable performance with an mIoU of 31.7% as shown in Table 4. This accomplishment underlines the model's ability to handle complex and noisy environments, while maintaining a high degree of accuracy.

Under snowy conditions, our model continues to set new standards, achieving an mIoU of 29.4%, as indicated in Table 5. Despite the challenges of reduced visibility in snowy environments, the model's performance affirms its exceptional adaptability and resilience. Across a spectrum of challenging weather conditions, the model demonstrates commendable precision, especially notable in categories such as "motorcycle", "truck", "person", and "vegetation". This consistent accuracy underscores the model's robustness and versatility. Nonetheless, there is potential for further refinement, particularly in enhancing its performance in more complex categories like "other-vehicle" and "other-ground".

4.2.2. Visualization results: from SemanticKITTI to SemanticSTF

Figs. 2, 3, 4, and 5 present the 3D semantic segmentation outcomes, utilizing SemanticKITTI as the source and SemanticSTF as the target, under varying weather conditions such as dense fog, light fog, rain, and snow, respectively. In these figures, (a) shows the ground truth; (b) and (c) present results from PointDR and our method, respectively, where different colors represent distinct semantic categories; (d) and (e) illustrate the right (R) and wrong (W) predictions in bright blue and bright yellow for PointDR and our method, respectively, to highlight differences. Although both the PointDR and our proposed method deliver satisfactory segmentation performance in these challenging visibility conditions, the latter demonstrates superior ability in discerning and segmenting intricate details.

Table 5
3D semantic segmentation results (%) with SemanticKITTI as source and SemanticSTF as target (snow).

Method	car	bicycle	motorcycle	truck	other-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic-sign	mIoU
Baseline	49.5	0.0	0.3	0.5	11.6	10.8	–	–	42.1	14.9	23.9	0.0	71.5	26.7	29.3	24.0	17.8	**30.8**	10.1	21.4
Dropout (Srivastava et al., 2014)	58.5	0.0	30.5	5.4	13.2	5.2	–	–	41.9	**18.0**	20.4	2.5	76.4	30.5	31.8	**32.7**	19.8	28.2	7.0	24.8
Perturbation	**73.6**	0.0	0.0	5.5	1.1	19.8	–	–	45.7	10.9	**34.4**	0.1	**80.6**	32.8	45.2	12.8	20.0	24.4	9.5	24.5
PolarMix (Xiao et al., 2022a)	66.5	**3.4**	9.3	3.5	5.8	32.4	–	–	**55.3**	3.6	30.1	0.1	77.8	36.1	34.2	12.6	25.1	29.8	10.1	25.6
MMD (Li et al., 2018)	59.4	0.0	4.7	0.0	14.7	30.5	–	–	50.8	16.9	32.8	0.2	68.4	24.4	36.6	24.1	24.1	30.0	11.4	25.2
PCL (Yao et al., 2022)	64.0	0.0	0.0	8.2	0.7	9.2	–	–	38.9	15.2	31.6	2.3	79.6	35.1	41.3	11.2	23.1	30.1	**26.8**	24.6
PointDR (Xiao et al., 2023b)	66.2	0.0	10.4	0.0	**16.7**	21.3	–	–	43.0	15.2	33.0	1.7	76.8	30.3	36.1	27.6	22.2	30.0	14.1	26.2
Ours	70.6	0.1	**39.8**	**8.7**	0.9	**39.4**	–	–	39.8	13.8	27.4	**3.4**	**80.6**	**36.6**	**46.8**	30.6	**26.0**	23.5	12.0	**29.4**

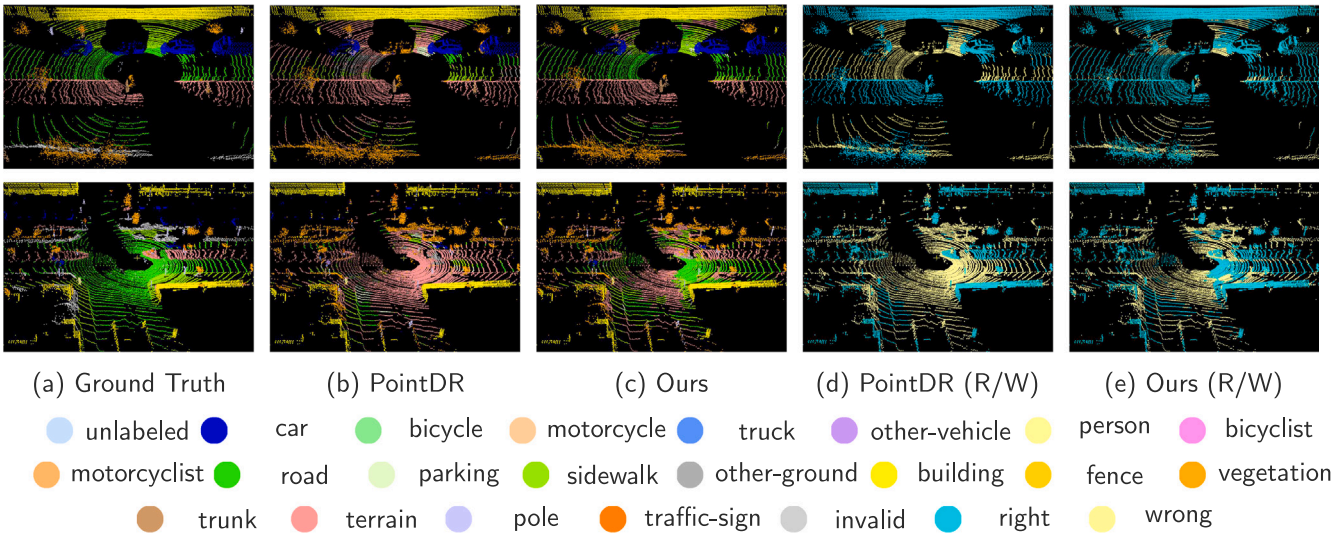

Fig. 2. Qualitative results with SemanticKITTI as source and SemanticSTF as target (dense fog). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

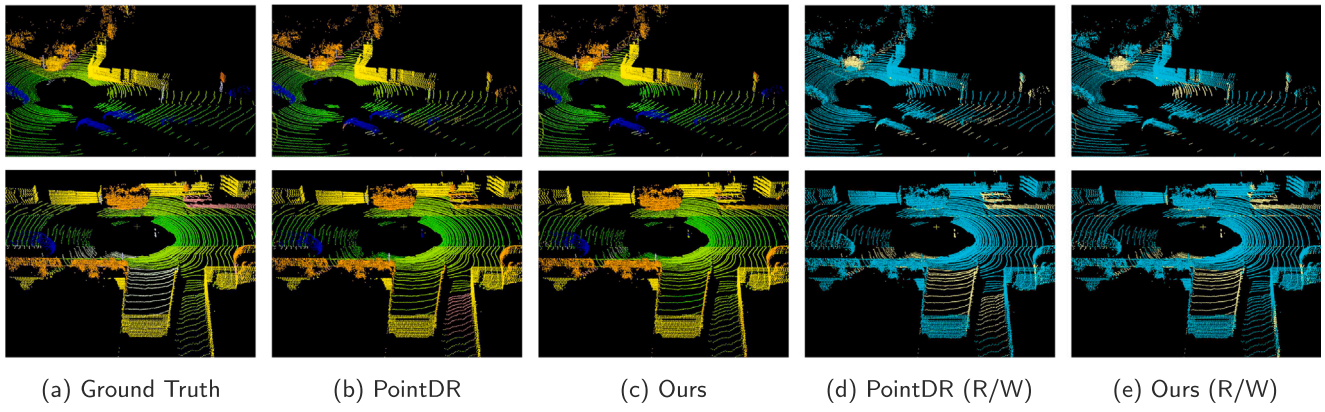

Fig. 3. Qualitative results with SemanticKITTI as source and SemanticSTF as target (light fog). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2.3. Evaluation results: from SynLiDAR to SemanticSTF

In Table 6, we assess the 3D semantic segmentation results utilizing SynLiDAR as the source and SemanticSTF as the target. Our method, showcased in the final row, performs remarkably well across multiple environmental conditions with the highest mIoU score of 21.9%. This showcases a higher average precision across categories in comparison to other methods, with PointDR (Xiao et al., 2023b) trailing behind at an mIoU of 18.5%. Examining category-wise performances, our model secures exemplary results, specifically in categories such as "road" (53.9%), "building" (64.3%), "vegetation" (53.8%), and "pole" (28.6%). The notable performance in the "road" and "building" categories, which are imperative for driving scenarios, underscore our model's adeptness. The robustness of our model extends to adverse weather conditions as well - dense fog (D-fog), light fog (L-fog), rain, and snow. Our method's resilience in these conditions further solidifies its superiority over other models. The underpinning factor for this superior performance can be attributed to our network's robust architecture, capable of encapsulating complex data interrelationships and yielding

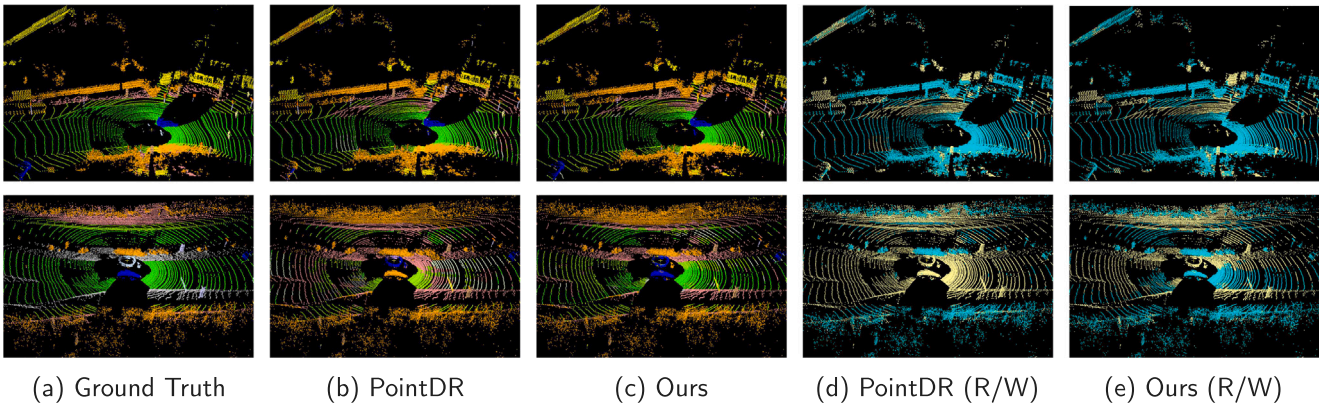

Fig. 4. Qualitative results with SemanticKITTI as source and SemanticSTF as target (rain). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

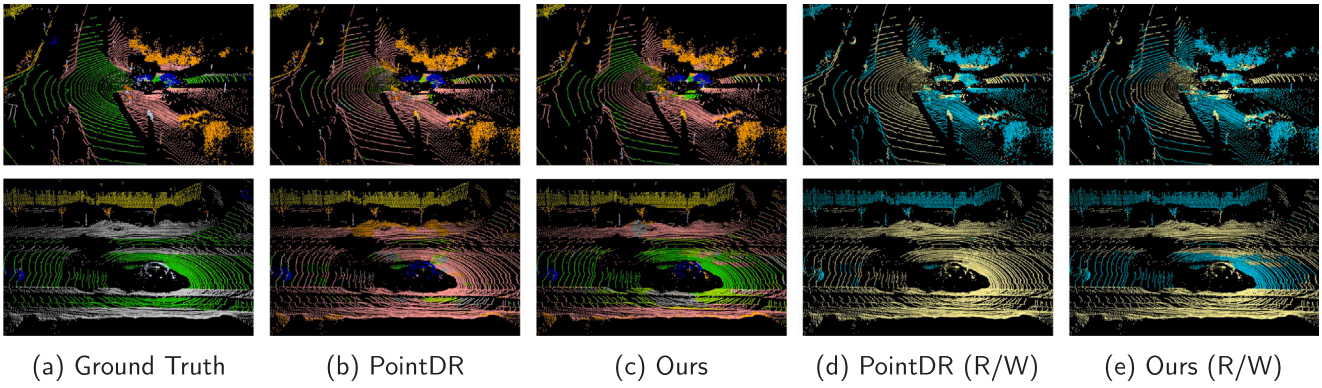

Fig. 5. Qualitative results with SemanticKITTI as source and SemanticSTF as target (snow). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 6
3D semantic segmentation results (%) with SynLiDAR as source and SemanticSTF as target.

Method	car	bicycle	motorcycle	truck	other-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic-sign	D-fog	L-fog	Rain	Snow	mIoU
Baseline	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	16.9	17.2	17.2	11.9	15.0
Dropout (Srivastava et al., 2014)	28.0	**3.0**	1.4	9.6	0.0	17.1	0.8	0.7	34.2	6.8	19.1	0.1	35.5	19.1	42.3	17.6	36.0	14.0	2.8	15.3	16.6	20.4	14.0	15.2
Perturbation	27.1	2.3	2.3	16.0	0.1	23.7	1.2	4.0	27.0	3.6	16.2	0.8	29.2	16.7	35.3	22.7	38.3	17.9	5.1	16.3	16.7	19.3	13.4	15.2
PolarMix (Xiao et al., 2022a)	**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	36.2	20.1	38.7	7.6	1.9	16.1	15.5	19.2	15.6	15.7
MMD (Li et al., 2018)	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	77.6	19.3	**43.5**	9.9	2.6	17.3	16.3	20.0	12.7	15.1
PCL (Yao et al., 2022)	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 (Xiao et al., 2023b)	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	19.9	21.1	16.9	18.5
Ours	33.8	1.1	**2.9**	17.0	0.2	**26.8**	1.0	4.3	**53.9**	5.0	**20.6**	**2.2**	**64.3**	**27.1**	**53.8**	**27.0**	37.0	**28.6**	**8.6**	**21.6**	**23.4**	**27.2**	**21.4**	**21.9**

nuanced segmentation results. Analogous scrutiny into the performance of other methods offers some insight into their limitations. For instance, while PointDR (Xiao et al., 2023b) model showcases commendable results in the "truck" (23.6%) and "parking" (7.7%) categories, it lacks consistent performance across the board. PolarMix (Xiao et al., 2022a), albeit performing well in the "car" category (39.2%), is not as proficient in handling "person", "road", and "trunk" categories.

These four tables illustrate the performance comparison of various methods for 3D semantic segmentation under different environmental conditions, namely dense fog (Table 7), light fog (Table 8), rain (Table 9), and snow (Table 10). These conditions, as represented by the SynLiDAR as source and SemanticSTF as target datasets, showcase a wide range of complexity and provide an extensive testbed for evaluating the robustness of our proposed method compared to other methods.

In the Table 7, our model outperforms the best model, PointDR, by 2.1% mIoU overall. Significant gains can be seen in categories like "road" (13.7% increase), "building" (11.1% increase), "fence" (2.8% increase), "vegetation" (1.9% increase), and "pole" (5.2% increase) when compared with the second-best accuracy achieved in each category. For the light fog condition in Table 8, our model achieves an mIoU of 23.4%, which is a 3.5% improvement over the next best, PointDR. Remarkable improvements are observed in "person" (3.8% increase), "road" (19.4% increase), "sidewalk" (3.2% increase), and "building" (8.3% increase), demonstrating the model's enhanced performance relative to the next best results in these categories. In the Table 9, despite the adverse rainy conditions, the proposed network maintains strong performance, particularly in the categories of "car" (2.2% increase), "road" (17.8% increase), "building" (13.2% increase), "fence" (6.1% increase), "vegetation" (9.7% increase), "pole" (8.7% increase), and "traffic-sign" (5.4% increase). This resilience to adverse weather indicates the robustness of the proposed model in handling weather-related challenges in feature extraction.

In the snowy conditions in Table 10, our model excels with an mIoU of 21.4%, surpassing the second-best model, PointDR, by 4.5%. Compared to the next best performances in each category, our model displays substantial improvements, with notable advances in "road"

Table 7
3D semantic segmentation results (%) with SynLiDAR as source and SemanticSTF as target (dense fog).

Method	car	bicycle	motorcycle	truck	other-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic-sign	mIoU
Baseline	21.6	–	–	6.4	0.0	3.7	2.9	18.9	25.7	0.0	7.7	1.0	41.2	22.5	52.3	15.4	**55.5**	9.3	2.4	16.9
Dropout (Srivastava et al., 2014)	12.7	–	–	7.7	0.0	1.9	0.4	2.5	38.3	0.1	10.2	0.3	37.3	21.8	57.4	13.1	44.5	10.1	1.0	15.3
Perturbation	13.3	–	–	10.4	0.0	**4.3**	2.8	19.1	30.0	**0.7**	8.8	1.2	30.5	17.5	48.9	18.4	50.3	16.3	5.2	16.3
PolarMix (Xiao et al., 2022a)	15.8	–	–	10.6	0.0	1.5	1.7	3.5	27.7	0.0	9.9	0.3	46.2	28.9	59.2	13.5	49.5	4.4	1.7	16.1
MMD (Li et al., 2018)	26.5	–	–	12.7	0.0	2.7	4.0	22.3	30.6	0.0	9.4	0.0	31.6	21.7	52.6	13.9	54.3	8.9	2.5	17.3
PCL (Yao et al., 2022)	22.9	–	–	**20.1**	0.0	2.2	**6.2**	**28.3**	29.0	0.0	9.2	**2.6**	37.9	22.9	54.5	11.4	45.9	8.5	1.1	17.8
PointDR (Xiao et al., 2023b)	**42.5**	–	–	16.6	0.0	2.4	3.2	12.2	31.9	0.2	9.0	0.8	42.8	27.1	59.8	18.3	44.0	15.4	5.7	19.5
Ours	24.4	–	–	13.1	0.0	3.5	0.0	19.9	**52.0**	0.1	**10.5**	0.7	**57.3**	**31.7**	**61.7**	**20.6**	42.0	**21.5**	**7.8**	**21.6**

Table 8
3D semantic segmentation results (%) with SynLiDAR as source and SemanticSTF as target (light fog).

Method	car	bicycle	motorcycle	truck	other-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic-sign	mIoU
Baseline	32.0	**4.2**	0.5	27.3	0.2	14.0	6.2	0.0	31.0	0.0	12.6	0.9	38.7	24.8	51.5	26.7	46.4	8.5	1.3	17.2
Dropout (Srivastava et al., 2014)	22.5	3.0	0.9	16.0	0.1	10.0	5.2	0.2	40.3	1.3	18.1	0.0	38.9	22.1	57.6	23.5	38.5	13.8	3.7	16.6
Perturbation	31.1	1.9	**1.6**	21.5	0.0	12.5	2.6	0.0	33.2	1.6	14.3	1.1	34.3	20.1	48.7	29.8	42.0	16.7	4.5	16.7
PolarMix (Xiao et al., 2022a)	27.3	0.3	0.4	8.9	**1.4**	8.2	1.2	0.0	29.0	0.2	15.5	0.7	39.9	27.4	57.3	28.8	40.9	5.8	1.5	15.5
MMD (Li et al., 2018)	31.0	2.1	0.5	16.0	0.0	10.5	1.7	0.0	37.7	0.3	16.3	0.6	29.2	24.9	51.8	29.6	**47.8**	8.3	1.8	16.3
PCL (Yao et al., 2022)	31.7	0.7	0.8	10.1	0.1	10.2	**21.6**	0.0	33.9	0.6	16.1	0.1	37.8	22.2	52.5	23.8	42.6	11.3	2.2	16.7
PointDR (Xiao et al., 2023b)	**44.7**	1.7	1.0	**33.9**	0.3	12.9	4.7	0.0	36.0	0.9	15.8	0.7	44.4	30.3	60.0	28.3	42.4	15.1	5.7	19.9
Ours	39.8	0.4	1.2	33.0	0.1	**17.8**	5.3	0.0	**59.7**	**3.8**	**21.3**	**5.2**	**52.7**	**31.2**	**63.2**	**33.4**	40.1	**27.1**	**8.6**	**23.4**

Table 9
3D semantic segmentation results (%) with SynLiDAR as source and SemanticSTF as target (rain).

Method	car	bicycle	motorcycle	truck	other-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic-sign	mIoU
Baseline	45.8	4.5	–	6.8	0.4	**38.9**	0.0	–	32.0	0.0	24.3	0.0	43.0	8.0	33.8	11.3	23.9	11.5	7.7	17.2
Dropout (Srivastava et al., 2014)	47.0	**7.6**	–	7.7	0.0	34.0	0.0	–	47.3	**6.9**	**34.6**	0.0	39.8	11.5	37.5	13.8	29.6	21.6	8.6	20.4
Perturbation	57.5	5.3	–	18.2	0.0	36.3	0.1	–	37.1	1.5	26.9	0.3	34.9	10.4	32.6	12.2	20.5	23.2	10.4	19.3
PolarMix (Xiao et al., 2022a)	59.6	1.5	–	6.0	**5.2**	24.6	1.0	–	31.4	0.1	30.4	0.0	55.5	12.2	44.6	13.1	25.0	11.0	4.7	19.2
MMD (Li et al., 2018)	49.5	4.8	–	20.0	4.7	37.6	0.0	–	43.7	0.0	32.4	0.0	42.1	11.3	34.4	12.3	25.1	13.4	8.1	20.0
PCL (Yao et al., 2022)	51.3	0.9	–	4.3	2.1	35.6	0.0	–	41.4	0.0	32.0	0.0	54.8	9.7	37.1	11.4	24.2	16.6	6.3	19.3
PointDR (Xiao et al., 2023b)	42.2	3.3	–	21.9	0.0	30.4	**1.7**	–	35.8	3.2	31.9	0.0	54.0	14.4	40.7	12.5	**31.9**	23.6	11.8	21.1
Ours	**61.8**	1.5	–	**22.0**	0.0	38.4	0.0	–	**65.1**	1.2	31.2	**4.1**	**68.7**	**20.5**	**54.3**	**14.5**	29.7	**32.3**	**17.2**	**27.2**

Table 10
3D semantic segmentation results (%) with SynLiDAR as source and SemanticSTF as target (snow).

Method	car	bicycle	motorcycle	truck	other-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic-sign	mIoU
Baseline	24.6	2.7	1.5	2.4	0.0	32.2	–	–	12.9	0.4	18.3	0.0	33.3	13.8	15.7	14.9	18.1	10.1	1.9	11.9
Dropout (Srivastava et al., 2014)	35.9	2.8	3.7	3.0	0.0	21.9	–	–	20.9	10.0	22.8	0.0	33.2	14.8	17.1	16.8	16.5	15.7	2.6	14.0
Perturbation	27.1	2.4	6.8	6.8	0.2	31.0	–	–	15.4	4.8	19.7	0.0	26.3	12.4	14.0	22.0	16.4	19.0	4.1	13.4
PolarMix (Xiao et al., 2022a)	**53.4**	2.3	4.1	6.0	**1.2**	27.9	–	–	11.7	1.9	21.5	0.3	45.2	**20.8**	21.7	18.8	16.5	10.5	1.7	15.6
MMD (Li et al., 2018)	20.8	2.7	6.0	4.8	0.2	31.3	–	–	20.1	0.5	21.0	0.1	29.6	12.2	15.0	16.6	**21.8**	11.3	2.4	12.7
PCL (Yao et al., 2022)	30.7	1.1	4.4	6.2	0.3	34.6	–	–	19.1	1.7	22.0	0.3	37.8	12.6	16.4	14.2	19.9	14.7	3.0	14.1
PointDR (Xiao et al., 2023b)	34.2	**4.0**	7.4	**7.5**	0.1	**36.2**	–	–	13.8	**12.0**	22.7	0.0	48.8	19.9	19.9	18.9	17.0	20.7	3.4	16.9
Ours	30.7	1.6	**7.9**	2.6	0.4	32.2	–	–	**46.5**	6.1	**25.1**	**0.9**	**69.9**	19.1	**30.2**	**28.0**	21.6	**33.9**	**7.5**	**21.4**

(25.6% increase), "building" (21.1% increase), "vegetation" (8.5% increase), and "pole" (13.2% increase). It is crucial to note that these performance improvements are not evenly distributed across all classes. Some categories see minor improvements or even slight regressions, which suggests room for further model optimization. However, the significant improvements in most categories validate the effectiveness

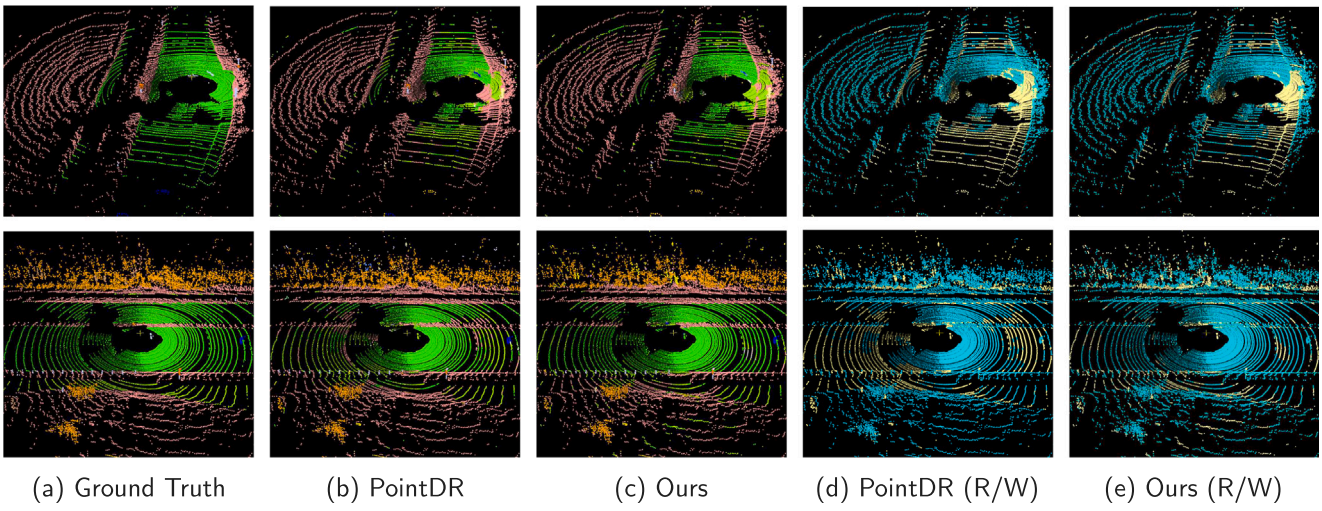

Fig. 6. Qualitative results with SynLiDAR as source and SemanticSTF as target (dense fog). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

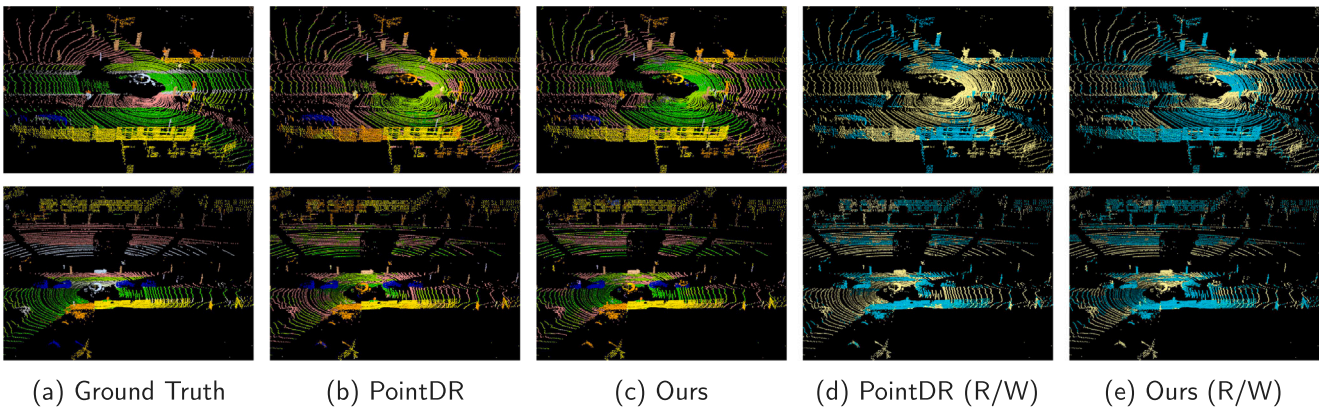

Fig. 7. Qualitative results with SynLiDAR as source and SemanticSTF as target (light fog). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of our proposed model. Specifically, the advanced feature processing, robustness to varying weather conditions, and efficient handling of class imbalance in our model contribute to these performance boosts.

4.2.4. Visualization results: from SynLiDAR to SemanticSTF

In this study, we present the visualization results of 3D semantic segmentation employing SynLiDAR as the source and SemanticSTF as the target domain. Figs. 6, 7, 8, and 9 illustrate these outcomes across a spectrum of challenging weather conditions, specifically dense fog, light fog, rain, and snow, respectively. Comparative analysis reveals that while the PointDR technique demonstrates notable proficiency in segmentation under adverse visibility conditions, our proposed methodology excels in its ability to discern fine details. This capability leads to a segmentation that is both more precise and nuanced, highlighting the advanced interpretative potential of our approach.

4.3. Ablation study

4.3.1. Ablation study on network components

Table 11 illustrates an ablation study that investigates the influence of various network components — AFNM, DAFM, and PLGM, on the semantic segmentation performance. The data sources for the study are SemanticKITTI and SemanticSTF, with SemanticKITTI serving as the source and SemanticSTF as the target. The data is categorized based on different weather conditions including D-fog, L-fog, Rain, Snow, and an aggregate condition labeled as "All". Different categories such as "car", "bicycle", "truck", "person", "road", "building", "vegetation", "pole", and "traffic-sign", etc., are taken into account. Table 11 presents the mIoU values for diverse object categories in each weather condition, given the various combinations of network components. A holistic evaluation of the table reveals that the full combination of AFNM, DAFM, and PLGM provides the highest mIoU values, particularly under the "All" weather conditions with a value of 34.8%. This signifies the effectiveness of integrating all modules in improving the overall model's performance. Upon examining the impacts of the absence of one or more components, a decrease in mIoU is noticeable. This suggests that each component plays a vital role in enhancing the model's performance. For instance, when each module is successively removed from the full combination, the mIoU value under "All" weather conditions reduces to 27.2%, 30.1%, and 30.0%, respectively. This reveals the contributions of individual components and how their absence can degrade the model's overall performance. Assessing category-specific IoU, the model exhibits varying performances. Certain categories, such as "car", "building", and "vegetation", consistently maintain relatively high mIoU values across all weather conditions and module combinations. This could be attributed to these categories being more common and less affected by weather conditions compared to others. On the other hand, categories like "bicycle", "other-ground", and "other-vehicle", exhibit much lower and more inconsistent mIoU values, perhaps due to their more complex shapes and lower occurrence rates in the data. A noteworthy observation is

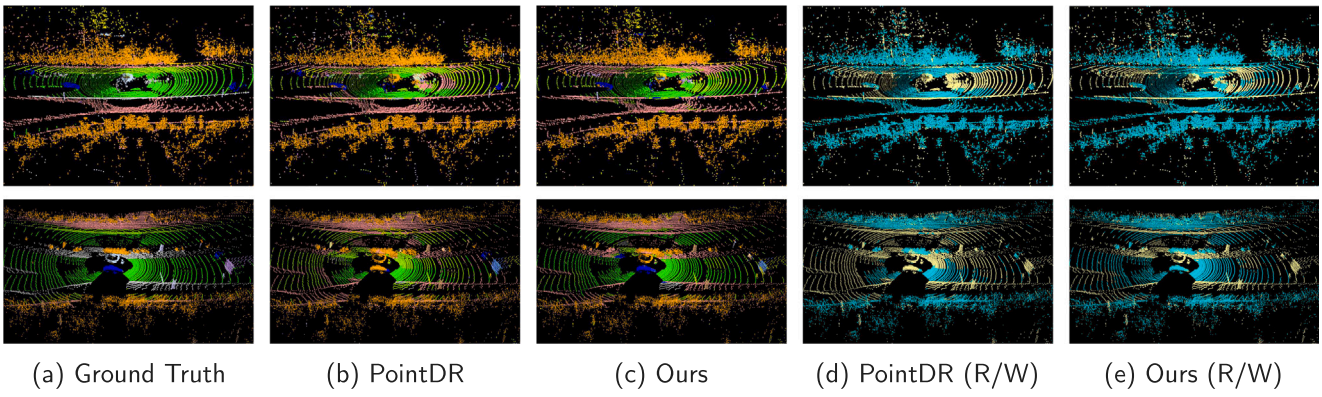

Fig. 8. Qualitative results with SynLiDAR as source and SemanticSTF as target (rain). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

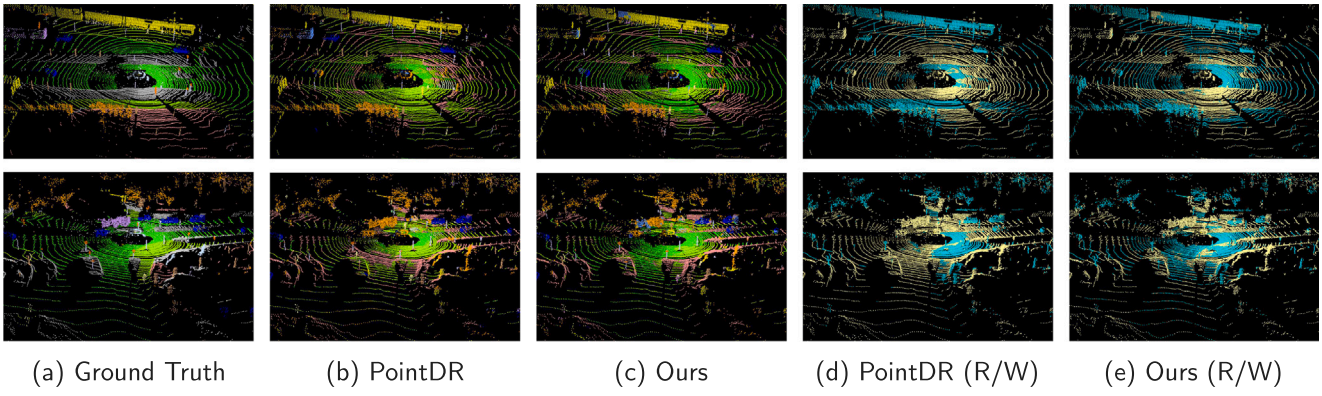

Fig. 9. Qualitative results with SynLiDAR as source and SemanticSTF as target (snow). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 11
Ablation study on the effects of network components on semantic segmentation performance with SemanticKITTI as the source and SemanticSTF as the target.

AFNM	DAFM	PLGM	Weather	car	bicycle	motorcycle	truck	other-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic-sign	D-fog	L-fog	Rain	Snow	mIoU
			D-fog	72.9	–	–	20.8	0.0	3.2	9.3	4.2	73.4	8.4	31.1	0.9	68.2	43.5	63.0	26.8	62.4	24.1	29.6	31.9				
			L-fog	58.6	0.0	0.0	38.0	12.9	2.5	0.0	0.0	70.3	7.8	37.6	0.2	69.4	56.0	60.5	36.3	51.8	26.8	19.6		28.9			
	√	√	Rain	80.7	0.0	–	2.0	1.7	4.2	0.0	–	71.9	15.0	55.4	0.0	75.1	35.2	53.6	11.7	27.2	35.3	32.0			29.5		
			Snow	48.8	0.0	34.3	8.2	7.6	6.7	–	–	37.8	18.7	25.8	0.7	70.9	29.0	28.1	25.9	20.5	27.0	13.7				23.8	
			All	55.6	0.0	16.1	23.7	7.3	5.8	6.7	3.3	60.8	15.4	32.4	0.6	70.6	43.1	52.6	28.1	46.5	26.7	22.1					27.2
			D-fog	75.3	–	–	28.0	0.0	7.1	16.7	0.0	74.7	0.3	30.4	3.6	64.6	50.5	62.3	32.6	60.1	23.4	25.7	32.7				
			L-fog	73.0	0.0	0.0	53.3	5.6	34.5	0.0	0.0	69.4	7.5	39.8	1.9	60.3	54.0	61.7	39.9	51.2	27.2	14.7		31.3			
√		√	Rain	83.7	0.0	–	0.0	0.0	45.2	0.0	–	56.1	17.9	53.7	12.6	73.5	39.6	57.0	19.7	19.4	32.9	13.5			30.9		
			Snow	77.8	0.0	0.0	14.1	0.6	47.1	–	–	47.4	11.0	31.4	1.8	77.0	34.0	48.2	28.5	26.7	30.8	7.5				28.5	
			All	76.9	0.0	0.0	34.6	2.0	42.8	5.9	0.0	62.2	10.0	35.4	3.0	71.9	46.5	59.6	32.3	48.6	28.0	13.0					30.1
			D-fog	60.1	–	–	15.5	0.0	2.3	28.4	0.0	63.8	1.0	20.8	0.8	67.4	43.4	60.2	32.9	54.1	23.5	36.1	30.0				
			L-fog	58.1	0.0	0.0	29.4	5.9	32.5	0.0	0.0	57.7	11.3	29.6	0.1	62.0	42.6	58.8	33.2	48.0	24.3	19.8		27.0			
√	√		Rain	68.9	0.4	–	0.2	16.4	28.5	0.0	–	42.5	8.6	38.0	2.3	76.6	20.8	55.3	20.8	22.5	36.6	13.1			26.5		
			Snow	73.0	8.6	41.2	14.8	14.0	40.3	–	–	34.7	18.8	30.9	0.5	80.8	30.3	47.4	23.7	26.4	24.7	9.1				**30.5**	
			All	67.8	5.7	33.5	23.0	9.5	36.3	13.0	0.0	50.5	15.9	29.0	0.7	75.0	37.8	57.4	28.5	45.2	25.0	17.0					30.0
			D-fog	78.8	–	–	40.6	0.0	7.0	47.7	67.3	71.6	0.3	29.8	1.4	65.0	48.0	66.5	33.0	60.8	21.3	33.0	**39.5**				
			L-fog	68.5	0.0	0.0	54.0	4.8	43.8	0.0	0.0	69.1	4.3	34.3	0.5	68.4	58.6	66.6	43.8	55.3	24.2	20.7		**32.5**			
√	√	√	Rain	84.3	0.0	–	0.4	0.0	40.6	0.0	–	65.1	19.2	51.3	4.2	77.3	42.7	61.3	15.7	22.8	30.9	23.5			**31.7**		
			Snow	70.6	0.1	39.8	8.7	0.9	39.4	–	–	39.8	13.8	27.4	3.4	80.6	36.6	46.8	30.6	26.0	23.5	12.0				29.4	
			All	72.0	0.0	32.9	37.0	1.9	37.7	6.8	52.9	59.9	10.7	31.8	2.2	76.0	48.8	62.7	34.0	49.3	23.6	20.4					**34.8**

the significant improvement in mIoU values for the "motorcyclist" category under D-fog weather condition when all components are active (67.3%), compared to when any of them is inactive (with highest being 4.2%). This indicates that some categories significantly benefit from the combination of all components. In conclusion, this ablation study underlines the importance of incorporating all three components — AFNM, DAFM, and PLGM, to achieve the optimal performance in semantic segmentation tasks across a range of weather conditions and object categories. It emphasizes the individual significance of each module and the synergy achieved when they work together.

Table 12 extends the ablation study to assess the impact of AFNM, DAFM, and PLGM on semantic segmentation, using SynLiDAR as the source and SemanticSTF as the target. The mIoU values for various object categories under different weather conditions are presented, highlighting how the model performs with different combinations of network components. Similar to the findings from Tables 11, 12 demonstrates that the highest mIoU values are achieved when all modules — AFNM, DAFM, and PLGM — are integrated, particularly under the "All" weather conditions. This consistent pattern reinforces the notion that the combined action of these modules significantly elevates the model's performance in semantic segmentation tasks. The results from

Table 12
Ablation study on the effects of network components on semantic segmentation performance with SynLiDAR as the source and SemanticSTF as the target.

AFNM	DAFM	PLGM	Weather	car	bicycle	motorcycle	truck	other-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic-sign	D-fog	L-fog	Rain	Snow	mIoU
			D-fog	35.8	–	–	12.1	0.0	3.0	6.6	24.1	40.7	0.0	9.4	0.1	55.2	25.8	60.4	23.1	44.7	19.3	8.7	21.7				
			L-fog	41.3	1.5	0.6	29.8	0.6	16.9	6.1	0.0	46.4	2.3	16.2	0.0	52.0	25.5	58.8	35.8	41.9	22.2	9.9		21.5			
	√	√	Rain	43.1	1.7	–	14.4	0.0	43.2	0.0	–	48.4	3.3	33.5	0.0	72.1	15.1	45.3	18.6	31.4	31.0	14.8			24.5		
			Snow	30.8	1.2	0.6	5.6	0.2	43.8	–	–	18.9	6.5	23.5	0.0	54.4	19.1	21.5	27.7	17.5	31.5	7.3				18.2	
			All	34.3	1.0	0.4	18.5	0.3	32.1	2.9	4.9	36.4	4.9	18.3	0.0	55.7	23.0	46.9	28.4	38.0	25.6	9.1					20.0
			D-fog	34.0	–	–	14.8	0.0	2.4	2.8	58.0	42.9	0.1	11.1	4.5	54.4	25.6	58.9	14.8	40.5	16.2	5.0	**22.7**				
			L-fog	45.9	2.8	0.3	36.6	1.6	13.3	2.3	0.0	50.6	4.4	19.5	1.9	53.1	27.1	59.7	25.7	40.4	20.7	7.4		21.8			
√		√	Rain	67.7	6.0	–	20.3	0.1	33.7	1.6	–	57.2	5.7	37.8	0.6	63.7	11.0	44.4	13.2	33.0	27.5	11.9			25.6		
			Snow	44.6	8.5	2.2	6.2	1.0	37.9	–	–	24.6	14.1	26.1	1.7	51.3	18.7	20.9	16.3	20.0	23.0	5.2				19.0	
			All	45.0	4.7	0.8	23.2	1.0	26.6	1.4	13.3	41.0	10.0	20.7	2.8	53.2	23.2	46.5	18.4	36.2	20.9	6.3					20.8
			D-fog	33.9	–	–	11.6	0.0	2.7	3.3	18.2	49.7	0.0	10.9	2.3	50.9	25.8	60.6	19.4	40.8	20.4	9.7	21.2				
			L-fog	49.3	1.6	0.8	19.2	0.1	13.0	3.4	0.0	54.1	6.0	21.0	0.1	43.2	26.3	59.4	28.8	39.2	24.0	11.2		21.1			
√	√		Rain	69.3	1.9	–	11.9	0.0	37.1	0.0	–	59.1	1.9	36.5	0.6	58.4	11.9	44.8	14.9	31.7	27.1	15.6			24.9		
			Snow	37.4	2.6	8.8	6.2	0.1	38.7	–	–	32.9	12.7	24.5	0.0	51.6	15.1	20.7	19.9	20.1	30.2	9.0				19.4	
			All	41.6	1.9	2.3	14.2	0.0	27.8	1.7	3.6	46.8	9.9	20.9	1.1	50.4	22.0	46.7	21.9	35.9	25.8	10.4					20.3
			D-fog	24.4	–	–	13.1	0.0	3.5	0.0	19.9	52.0	0.1	10.5	0.7	57.3	31.7	61.7	20.6	42.0	21.5	7.8	21.6				
			L-fog	39.8	0.4	1.2	33.0	0.1	17.8	5.3	0.0	59.7	3.8	21.3	5.2	52.7	31.2	63.2	33.4	40.1	27.1	8.6		**23.4**			
√	√	√	Rain	61.8	1.5	–	22.0	0.0	38.4	0.0	–	65.1	1.2	31.2	4.1	68.7	20.5	54.3	14.5	29.7	32.3	17.2			**27.2**		
			Snow	30.7	1.6	7.9	2.6	0.4	32.2	–	–	46.5	6.1	25.1	0.9	69.9	19.1	30.2	28.0	21.6	33.9	7.5				**21.4**	
			All	33.8	1.1	2.9	17.0	0.2	26.8	1.0	4.3	53.9	5.0	20.6	2.2	64.3	27.1	53.8	27.0	37.0	28.6	8.6					**21.9**

Table 13
Ablation study on the effects of loss components on semantic segmentation performance with SynLiDAR as the source and SemanticSTF as the target.

RCSL	TCSL	RCPCL	TCPCL	MIL	Weather	car	bicycle	motorcycle	truck	other-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic-sign	D-fog	L-fog	Rain	Snow	mIoU
					D-fog	46.8	–	–	14.3	0.0	1.8	3.6	25.8	42.7	0.3	10.3	0.4	43.2	32.7	61.8	17.9	39.0	16.8	5.2	21.3				
					L-fog	56.6	3.2	1.6	37.0	7.9	16.4	4.3	0.0	46.1	4.4	17.6	0.2	43.8	33.1	63.1	33.0	37.2	20.5	6.7		22.8			
√		√	√	√	Rain	68.7	5.6	–	20.3	0.4	40.3	1.8	–	47.4	5.5	34.2	0.0	54.1	14.9	43.9	15.7	30.4	33.7	15.6			25.4		
					Snow	48.9	7.1	8.1	5.7	2.7	41.2	–	–	21.2	11.7	24.7	0.0	47.3	22.9	22.6	22.8	19.3	26.2	5.9				19.9	
					All	51.6	4.3	3.1	23.7	3.8	30.4	2.3	5.6	37.4	8.5	19.2	0.2	46.5	28.5	49.1	24.1	34.4	22.5	6.6					21.2
					D-fog	9.2	–	–	15.3	0.0	2.7	6.2	49.7	55.8	0.0	12.7	0.0	49.1	24.9	62.5	12.1	29.7	18.6	3.5	20.7				
					L-fog	17.8	0.9	0.9	35.5	0.0	13.9	13.1	0.6	60.5	0.1	24.1	0.0	45.0	27.8	59.4	21.1	30.2	20.3	7.2		19.9			
√	√			√	Rain	48.4	2.7	–	16.9	0.0	39.5	0.0	–	66.7	0.0	36.7	0.0	66.5	20.9	50.0	12.2	20.8	29.3	7.5			24.6		
					Snow	37.5	2.8	4.7	14.2	0.5	34.0	–	–	47.5	0.1	28.1	0.0	56.7	18.1	23.3	10.1	15.9	27.2	5.1				19.2	
					All	24.5	1.4	1.5	26.6	0.3	25.3	4.2	24.2	55.6	0.1	23.8	0.0	54.1	23.8	49.1	13.8	27.4	23.1	5.4					20.2
					D-fog	16.4	–	–	8.6	0.0	2.3	10.3	43.3	46.7	0.1	11.8	2.9	42.7	22.0	59.9	12.3	40.1	19.4	7.1	20.3				
					L-fog	40.3	2.7	0.0	13.5	0.4	15.7	5.8	0.0	51.5	0.1	20.6	2.6	35.8	24.1	59.2	21.9	38.8	21.9	5.6		19.0			
√	√	√		√	Rain	64.0	5.2	–	13.2	0.1	38.9	1.5	–	61.6	0.0	38.7	3.5	61.2	16.8	49.7	12.0	28.1	30.4	12.6			25.7		
					Snow	38.0	4.2	6.1	5.5	0.2	36.6	–	–	39.7	0.1	25.7	0.8	58.3	15.1	24.6	15.8	21.9	29.8	5.9				19.3	
					All	36.8	3.5	1.6	10.1	0.2	27.7	3.2	8.4	47.2	0.1	21.8	2.4	50.7	20.6	49.1	16.5	35.7	24.8	6.7					19.3
					D-fog	35.8	–	–	9.2	0.0	3.0	3.0	13.0	54.2	0.8	10.5	0.3	50.7	17.5	59.5	18.1	35.5	19.9	8.2	19.9				
					L-fog	45.6	2.1	0.3	28.5	1.5	10.9	2.8	0.1	58.4	3.4	20.5	0.0	44.9	18.3	58.2	29.2	32.5	20.9	10.4		20.4			
√	√	√	√		Rain	63.5	4.7	–	15.0	5.0	40.6	2.4	–	60.3	2.5	32.3	0.0	61.8	6.6	45.9	13.6	23.2	31.6	16.5			25.0		
					Snow	40.5	3.2	9.7	3.5	0.3	35.1	–	–	43.5	8.1	26.1	0.0	50.6	11.8	21.7	22.4	17.6	29.6	9.3				19.6	
					All	42.8	2.3	3.0	17.4	1.0	25.2	1.6	3.2	52.7	6.2	20.6	0.1	50.5	15.2	46.7	22.6	30.8	24.6	9.8					19.8
					D-fog	24.4	–	–	13.1	0.0	3.5	0.0	19.9	52.0	0.1	10.5	0.7	57.3	31.7	61.7	20.6	42.0	21.5	7.8	**21.6**				
					L-fog	39.8	0.4	1.2	33.0	0.1	17.8	5.3	0.0	59.7	3.8	21.3	5.2	52.7	31.2	63.2	33.4	40.1	27.1	8.6		**23.4**			
√	√	√	√	√	Rain	61.8	1.5	–	22.0	0.0	38.4	0.0	–	65.1	1.2	31.2	4.1	68.7	20.5	54.3	14.5	29.7	32.3	17.2			**27.2**		
					Snow	30.7	1.6	7.9	2.6	0.4	32.2	–	–	46.5	6.1	25.1	0.9	69.9	19.1	30.2	28.0	21.6	33.9	7.5				**21.4**	
					All	33.8	1.1	2.9	17.0	0.2	26.8	1.0	4.3	53.9	5.0	20.6	2.2	64.3	27.1	53.8	27.0	37.0	28.6	8.6					**21.9**

Table 12 further validate the integral roles of AFNM, DAFM, and PLGM, especially when dealing with the varying complexities of different weather conditions.

4.3.2. Ablation study on loss function

Table 13 presented here is an ablation study on the impacts of various loss components on semantic segmentation performance. This study utilizes SynLiDAR as the source and SemanticSTF as the target. The study examines several different environmental conditions: D-fog, L-fog, Rain, Snow, and a combination of all these conditions (All). Each row of the table represents a different combination of five loss components: RCSL, TCSL, RCPCL, TCPCL, and MIL. Each combination is then evaluated under the various environmental conditions and the mIoU metric is reported. In the rainy condition, when all five loss components (RCSL, TCSL, RCPCL, TCPCL, and MIL) are activated, the model achieves the highest mIoU of 27.2%. For individual categories, the IoU varies greatly. For example, under rainy conditions with all loss components, "vegetation" segmentation reaches a score of 54.3%. From the perspective of the impact of different loss modules, the addition of each component generally contributes to the improvement of the mIoU metric. However, the impact differs. For example, including RCSL, RCPCL, TCPCL, and MIL results in an mIoU of 25.4% in rainy conditions, while adding TCSL to these components increases the mIoU to 27.2% in the same conditions. Therefore, the impact of the TCSL component seems particularly significant in this case. Looking at all the loss components combined, the best performance is observed under the light fog condition with an mIoU of 23.4%. This indicates that using all components provides the best performance, suggesting that each component offers unique benefits that collectively contribute to improved performance. Regarding weather conditions, it is clear that the segmentation performance varies greatly depending on the weather. The best performance is achieved under rainy conditions, which could suggest that the model is more capable of handling such weather conditions compared to others. Finally, when looking at the "All" rows, representing the results under a combination of all weather conditions, one can notice that the best mIoU (21.9%) is achieved when using the RCSL, TCSL, RCPCL, TCPCL, and MIL components. This might suggest that this combination of loss components is the most versatile in handling different weather conditions.

Table 14 delineates a comprehensive ablation study focusing on the impact of various loss components on semantic segmentation performance, with SemanticKITTI as the source and SemanticSTF as the target. An in-depth analysis of the table reveals that the combination of all loss components yields the highest mIoU values, particularly under the comprehensive "All" conditions, where it reaches 34.8%. This outcome underscores the collective efficacy of the integrated loss components in optimizing the overall model performance under varying weather conditions. The study further explores the influence of individual loss components on the model's segmentation accuracy.

Table 14
Ablation study on the effects of loss components on semantic segmentation performance with SemanticKITTI as the source and SemanticSTF as the target.

RCSL	TCSL	RCPCL	TCPCL	MIL	Weather	car	bicycle	motorcycle	truck	other-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic-sign	D-fog	L-fog	Rain	Snow	mIoU
					D-fog	77.1	–	–	30.4	0.0	5.8	7.8	0.0	71.1	2.0	31.8	0.0	66.2	50.5	61.0	33.8	61.0	25.9	39.9	33.2					
					L-fog	66.4	0.0	0.0	57.9	18.3	29.4	0.0	0.0	63.1	7.0	34.7	0.0	71.7	60.8	62.4	41.9	48.7	25.6	32.3		**32.6**			
√		√	√	√	Rain	85.5	0.0	–	0.1	3.8	40.6	0.0	–	66.9	11.7	49.5	0.6	77.7	44.2	57.3	19.8	23.3	33.5	40.9			32.7		
					Snow	64.2	0.0	35.9	12.9	8.6	32.1	–	–	35.9	21.1	24.2	0.0	79.9	33.9	37.1	31.4	18.8	27.2	27.6				28.9	
					All	67.5	0.0	20.6	36.3	9.7	31.0	1.9	0.0	56.6	15.5	30.7	0.1	76.3	49.3	56.5	34.2	43.7	26.8	34.2					31.1
					D-fog	71.7	–	–	36.8	0.0	4.7	0.0	23.4	72.3	0.1	30.8	0.4	64.9	49.5	57.1	35.9	55.2	23.5	30.6	32.8					
					L-fog	61.1	0.0	0.0	53.2	1.1	26.5	0.0	0.0	69.8	6.5	37.3	0.0	70.9	59.1	56.6	46.6	46.1	25.5	19.7		30.5			
√	√			√	Rain	62.4	0.0	–	0.1	0.1	40.7	0.0	–	74.7	18.2	55.7	0.2	73.7	48.2	56.7	24.3	24.2	31.6	37.1			32.2		
					Snow	64.4	0.0	16.0	12.7	1.2	44.9	–	–	35.8	19.7	25.0	1.0	78.8	31.5	33.6	34.4	18.6	24.4	21.0				27.2	
					All	64.3	0.0	5.8	35.8	0.9	39.3	0.0	18.0	59.8	15.6	31.8	0.5	75.0	47.8	51.9	37.6	41.3	25.0	24.9					30.3
					D-fog	70.2	–	–	17.9	0.0	6.2	0.0	1.6	70.8	0.0	33.2	0.1	63.1	45.9	58.8	36.0	50.0	21.1	31.4	29.8					
					L-fog	67.3	0.0	0.0	41.9	12.6	31.7	0.0	0.0	74.5	1.9	45.9	0.0	63.7	53.2	58.0	48.4	48.1	23.7	23.4		31.3			
√	√	√		√	Rain	81.5	0.0	–	2.7	14.5	47.0	0.0	–	80.8	13.6	60.8	3.5	69.4	39.7	55.9	28.0	31.8	33.8	29.5			**34.9**		
					Snow	63.4	4.9	13.7	10.8	11.9	47.2	–	–	58.8	12.1	34.0	0.0	70.7	25.6	31.6	34.9	22.1	24.7	15.2				28.3	
					All	66.3	3.1	4.0	30.1	11.7	41.1	0.0	1.3	68.9	9.3	39.2	0.3	68.3	41.1	52.0	38.3	42.7	24.0	24.4					29.8
					D-fog	70.3	–	–	9.0	0.0	3.5	0.0	0.5	66.0	3.9	23.1	0.6	64.3	43.6	66.8	33.4	63.7	22.7	47.6	30.5					
					L-fog	74.9	0.0	0.0	5.4	0.4	35.5	0.0	0.0	60.1	13.1	28.5	0.3	70.2	60.8	63.9	44.5	54.8	27.2	33.8		30.2			
√	√	√	√		Rain	85.9	0.1	–	0.0	1.8	39.6	0.0	–	61.3	16.4	48.2	3.7	75.2	44.8	59.8	24.2	30.2	27.2	38.5			32.8		
					Snow	76.4	0.0	19.1	1.2	9.2	44.2	–	–	31.0	18.9	27.3	1.1	76.0	29.3	39.2	32.1	25.6	24.0	27.6				28.4	
					All	76.1	0.1	15.9	4.8	4.8	39.7	0.0	0.4	52.3	16.8	28.6	0.9	73.3	45.2	59.3	35.3	50.4	24.7	37.0					29.8
					D-fog	78.8	–	–	40.6	0.0	7.0	47.7	67.3	71.6	0.3	29.8	1.4	65.0	48.0	66.5	33.0	60.8	21.3	33.0	**39.5**				
					L-fog	68.5	0.0	0.0	54.0	4.8	43.8	0.0	0.0	69.1	4.3	34.3	0.5	68.4	58.6	66.6	43.8	55.3	24.2	20.7		32.5			
√	√	√	√	√	Rain	84.3	0.0	–	0.4	0.0	40.6	0.0	–	65.1	19.2	51.3	4.2	77.3	42.7	61.3	15.7	22.8	30.9	23.5			31.7		
					Snow	70.6	0.1	39.8	8.7	0.9	39.4	–	–	39.8	13.8	27.4	3.4	80.6	36.6	46.8	30.6	26.0	23.5	12.0				**29.4**	
					All	72.0	0.0	32.9	37.0	1.9	37.7	6.8	52.9	59.9	10.7	31.8	2.2	76.0	48.8	62.7	34.0	49.3	23.6	20.4					**34.8**

The removal of any single component results in a discernible decrease in mIoU, indicating the integral role of each component in the segmentation process.

5. Conclusion

In the context of 3D point cloud semantic segmentation, weather variability presents a significant challenge that can substantially affect the performance and reliability of autonomous systems. In this work, we have proposed and validated a novel deep learning framework aimed at enhancing the robustness of semantic segmentation under diverse weather conditions, thereby contributing to the safe and efficient operation of systems such as autonomous vehicles. Our research introduces the incorporation of unique modules, including the Adaptive Feature Normalization Module (AFNM), Dual-Attention Fusion Module (DAFM), and Proxy Label Generation Module (PLGM). Together, these components effectively normalize and calibrate features, fuse cross-domain features, and generate reliable proxy labels. Furthermore, we have presented an innovative multi-faceted loss function that guides the learning process efficiently. Experimental evaluations conducted on standard benchmarks have demonstrated the effectiveness and superiority of our framework over state-of-the-art methods under a variety of weather conditions. The results attest to the robustness and resilience of our proposed method in the face of environmental variability. However, despite these promising findings, there is room for further investigation and development. Our future work will focus on refining the framework to improve segmentation performance in more challenging and less common categories. Additionally, exploring and integrating other auxiliary data sources may offer avenues to further enhance the model's performance under extreme weather conditions. We hope that our work will inspire further research in this important and challenging area.

CRediT authorship contribution statement

Jing Du: Funding acquisition, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **John Zelek:** Funding acquisition, Methodology, Supervision, Validation, Writing – review & editing. **Jonathan Li:** Funding acquisition, Methodology, Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Armeni, I., Sener, O., Zamir, A.R., Jiang, H., Brilakis, I.K., Fischer, M., Savarese, S., 2016. 3D semantic parsing of large-scale indoor spaces. In: Proc. CVPR. IEEE, pp. 1534–1543. http://dx.doi.org/10.1109/CVPR.2016.170.

Behley, J., Garbade, M., Milioto, A., Quenzel, J., Behnke, S., Stachniss, C., Gall, J., 2019. SemanticKITTI: A dataset for semantic scene understanding of LiDAR sequences. In: Proc.ICCV. IEEE, pp. 9296–9306. http://dx.doi.org/10.1109/ICCV.2019.00939.

Bijelic, M., Gruber, T., Mannan, F., Kraus, F., Ritter, W., Dietmayer, K., Heide, F., 2020. Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. In: Proc. CVPR. IEEE, pp. 11679–11689. http://dx.doi.org/10.1109/CVPR42600.2020.01170.

Choy, C.B., Gwak, J., Savarese, S., 2019. 4D spatio-temporal ConvNets: Minkowski convolutional neural networks. In: Proc. CVPR. IEEE, pp. 3075–3084. http://dx.doi.org/10.1109/CVPR.2019.00319.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B., 2016. The cityscapes dataset for semantic urban scene understanding. In: Proc.CVPR. IEEE Computer Society, pp. 3213–3223. http://dx.doi.org/10.1109/CVPR.2016.350.

Cortinhal, T., Tzelepis, G., Aksoy, E.E., 2020. SalsaNext: Fast, uncertainty-aware semantic segmentation of LiDAR point clouds. In: Proc. ISVC. Vol. 12510, Springer, pp. 207–222. http://dx.doi.org/10.1007/978-3-030-64559-5_16.

Dai, A., Chang, A.X., Savva, M., Halber, M., Funkhouser, T.A., Nießner, M., 2017. ScanNet: Richly-annotated 3D reconstructions of indoor scenes. In: Proc. CVPR. IEEE, pp. 2432–2443. http://dx.doi.org/10.1109/CVPR.2017.261.

Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., Bennamoun, M., 2021. Deep learning for 3D point clouds: A survey. IEEE Trans. Pattern Anal. Mach. Intell. 43 (12), 4338–4364. http://dx.doi.org/10.1109/TPAMI.2020.3005434.

Hackel, T., Savinov, N., Ladicky, L., Wegner, J.D., Schindler, K., Pollefeys, M., 2017. Semantic3D.net: A new large-scale point cloud classification benchmark. CoRR URL: http://arxiv.org/abs/1704.03847.

He, S., Jiang, X., Jiang, W., Ding, H., 2023. Prototype adaption and projection for few- and zero-shot 3D point cloud semantic segmentation. IEEE Trans. Image Process. 32, 3199–3211. http://dx.doi.org/10.1109/TIP.2023.3279660.

Hoyer, L., Dai, D., Wang, H., Gool, L.V., 2023. MIC: Masked image consistency for context-enhanced domain adaptation. In: Proc. CVPR. IEEE, pp. 11721–11732. http://dx.doi.org/10.1109/CVPR52729.2023.01128.

Hu, Q., Yang, B., Fang, G., Guo, Y., Leonardis, A., Trigoni, N., Markham, A., 2022. SQN: Weakly-supervised semantic segmentation of large-scale 3D point clouds. In: Proc. ECCV. In: Lecture Notes in Computer Science, Vol. 13687, Springer, pp. 600–619. http://dx.doi.org/10.1007/978-3-031-19812-0_35.

Kalwar, S., Patel, D., Aanegola, A., Konda, K.R., Garg, S., Krishna, K.M., 2023. GDIP: Gated differentiable image processing for object detection in adverse conditions. In: Proc. ICRA. IEEE, pp. 7083–7089. http://dx.doi.org/10.1109/ICRA48891.2023.10160356.

Kim, H., Kang, Y., Oh, C., Yoon, K., 2023. Single domain generalization for LiDAR semantic segmentation. In: Proc. CVPR. IEEE, pp. 17587–17598. http://dx.doi.org/10.1109/CVPR52729.2023.01687.

Li, J., Dai, H., Han, H., Ding, Y., 2023a. MSeg3D: Multi-modal 3D semantic segmentation for autonomous driving. In: Proc. CVPR. IEEE, pp. 21694–21704. http://dx.doi.org/10.48550/arXiv.2303.08600.

Li, H., Pan, S.J., Wang, S., Kot, A.C., 2018. Domain generalization with adversarial feature learning. In: Proc. CVPR. IEEE, pp. 5400–5409. http://dx.doi.org/10.1109/CVPR.2018.00566.

Li, L., Shum, H.P.H., Breckon, T.P., 2023b. Less is more: Reducing task and model complexity for 3d point cloud semantic segmentation. In: Proc. CVPR. pp. 9361–9371. http://dx.doi.org/10.48550/arXiv.2303.11203.

Liao, L., Chen, W., Xiao, J., Wang, Z., Lin, C., Satoh, S., 2022. Unsupervised foggy scene understanding via self spatial-temporal label diffusion. IEEE Trans. Image Process. 31, 3525–3540. http://dx.doi.org/10.1109/TIP.2022.3172208.

Liao, L., Chen, W., Zhang, Z., Xiao, J., Yang, Y., Lin, C., Satoh, S., 2023. Only a few classes confusing: Pixel-wise candidate labels disambiguation for foggy scene understanding. In: Proc.AAAI. AAAI Press, pp. 1558–1567. http://dx.doi.org/10.1609/AAAI.V37I2.25242.

Liu, M., Zhou, Y., Qi, C.R., Gong, B., Su, H., Anguelov, D., 2022. LESS: Label-efficient semantic segmentation for LiDAR point clouds. In: Proc. ECCV. In: Lecture Notes in Computer Science, Vol. 13699, Springer, pp. 70–89. http://dx.doi.org/10.1007/978-3-031-19842-7_5.

Neuhold, G., Ollmann, T., Bulò, S.R., Kontschieder, P., 2017. The Mapillary vistas dataset for semantic understanding of street scenes. In: Proc.ICCV. IEEE Computer Society, pp. 5000–5009. http://dx.doi.org/10.1109/ICCV.2017.534.

Riz, L., Saltori, C., Ricci, E., Poiesi, F., 2023. Novel class discovery for 3D point cloud semantic segmentation. In: Proc. CVPR. IEEE, pp. 9393–9402. http://dx.doi.org/10.48550/arXiv.2303.11610.

Sakaino, H., 2023. PanopticRoad: Integrated panoptic road segmentation under adversarial conditions. In: Proc. CVPR. IEEE, pp. 3591–3603. http://dx.doi.org/10.1109/CVPRW59228.2023.00367.

Saltori, C., Galasso, F., Fiameni, G., Sebe, N., Ricci, E., Poiesi, F., 2022. CoSMix: Compositional semantic mix for domain adaptation in 3D LiDAR segmentation. In: Avidan, S., Brostow, G.J., Cissé, M., Farinella, G.M., Hassner, T. (Eds.), Proc. ECCV. Vol. 13693, Springer, pp. 586–602. http://dx.doi.org/10.1007/978-3-031-19827-4_34.

Sanchez, J., Deschaud, J., Goulette, F., 2023. Domain generalization of 3D semantic segmentation in autonomous driving. In: Proc.ICCV. IEEE, pp. 18031–18041. http://dx.doi.org/10.1109/ICCV51070.2023.01657.

Shuai, H., Liu, Q., 2023. Geometry-injected image-based point cloud semantic segmentation. IEEE Trans. Geosci. Remote Sens. 61, 1–10. http://dx.doi.org/10.1109/TGRS.2023.3264292.

Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (1), 1929–1958. http://dx.doi.org/10.5555/2627435.2670313.

Tan, W., Qin, N., Ma, L., Li, Y., Du, J., Cai, G., Yang, K., Li, J., 2020. Toronto-3D: A large-scale mobile LiDAR dataset for semantic segmentation of urban roadways. In: Proc. CVPR. IEEE, pp. 797–806. http://dx.doi.org/10.1109/CVPRW50498.2020.00109.

Tang, Z.J., Cham, T., 2022. MPT-Net: Mask point transformer network for large scale point cloud semantic segmentation. In: Proc. IROS. IEEE, pp. 10611–10618. http://dx.doi.org/10.1109/IROS47612.2022.9981809.

Tzeng, E., Hoffman, J., Saenko, K., Darrell, T., 2017. Adversarial discriminative domain adaptation. In: Proc. CVPR. IEEE, pp. 2962–2971. http://dx.doi.org/10.1109/CVPR.2017.316.

Vu, T., Jain, H., Bucher, M., Cord, M., Pérez, P., 2019. ADVENT: Adversarial entropy minimization for domain adaptation in semantic segmentation. In: Proc. CVPR. IEEE, pp. 2517–2526. http://dx.doi.org/10.1109/CVPR.2019.00262.

Wang, Z., Rao, Y., Yu, X., Zhou, J., Lu, J., 2022b. SemAffiNet: Semantic-affine transformation for point cloud segmentation. In: Proc. CVPR. IEEE, pp. 11809–11819. http://dx.doi.org/10.1109/CVPR52688.2022.01152.

Wang, G., Yang, Y., Zhang, H., Liu, Z., Wang, H., 2022a. Spherical interpolated convolutional network with distance-feature density for 3-D semantic segmentation of point clouds. IEEE Trans. Cybern. 52 (12), 13546–13556. http://dx.doi.org/10.1109/TCYB.2021.3124954.

Wang, P., Yao, W., 2022. A new weakly supervised approach for ALS point cloud semantic segmentation. ISPRS J. Photogramm. Remote Sens. (ISSN: 0924-2716) 188, 237–254. http://dx.doi.org/10.1016/j.isprsjprs.2022.04.016.

Xiao, A., Huang, J., Guan, D., Cui, K., Lu, S., Shao, L., 2022a. PolarMix: A general data augmentation technique for LiDAR point clouds. In: NeurIPS. URL: https://arxiv.org/abs/2208.00223.

Xiao, A., Huang, J., Guan, D., Zhan, F., Lu, S., 2022b. Transfer learning from synthetic to real LiDAR point cloud for semantic segmentation. In: Proc.AAAI. AAAI Press, pp. 2795–2803, URL: https://ojs.aaai.org/index.php/AAAI/article/view/20183.

Xiao, A., Huang, J., Guan, D., Zhang, X., Lu, S., Shao, L., 2023a. Unsupervised point cloud representation learning with deep neural networks: A survey. IEEE Trans. Pattern Anal. Mach. Intell. 45 (9), 11321–11339. http://dx.doi.org/10.1109/TPAMI.2023.3262786.

Xiao, A., Huang, J., Xuan, W., Ren, R., Liu, K., Guan, D., El-Saddik, A., Lu, S., Xing, E.P., 2023b. 3D semantic segmentation in the Wild: Learning generalized models for adverse-condition point clouds. In: Proc. CVPR. http://dx.doi.org/10.48550/arXiv.2304.00690.

Xiao, A., Yang, X., Lu, S., Guan, D., Huang, J., 2021. FPS-Net: A convolutional fusion network for large-scale LiDAR point cloud segmentation. ISPRS J. Photogramm. Remote Sens. 176, 237–249. http://dx.doi.org/10.1016/j.isprsjprs.2021.04.011.

Xiao, A., Zhang, X., Shao, L., Lu, S., 2023c. A survey of label-efficient deep learning for 3D point clouds. http://dx.doi.org/10.48550/ARXIV.2305.19812, CoRR abs/2305.19812.

Xie, B., Li, S., Guo, Q., Liu, C.H., Cheng, X., 2024. Annotator: A generic active learning baseline for LiDAR semantic segmentation. In: NeurIPS.

Yao, X., Bai, Y., Zhang, X., Zhang, Y., Sun, Q., Chen, R., Li, R., Yu, B., 2022. PCL: proxy-based contrastive learning for domain generalization. In: Proc. CVPR. IEEE, pp. 7087–7097. http://dx.doi.org/10.1109/CVPR52688.2022.00696.

Yuan, Z., Wen, C., Cheng, M., Su, Y., Liu, W., Yu, S., Wang, C., 2023. Category-level adversaries for outdoor LiDAR point clouds cross-domain semantic segmentation. IEEE Trans. Intell. Transp. Syst. 24 (2), 1982–1993. http://dx.doi.org/10.1109/TITS.2022.3219853.

Zhang, Y., Carballo, A., Yang, H., Takeda, K., 2023. Perception and sensing for autonomous vehicles under adverse weather conditions: A survey. ISPRS J. Photogramm. Remote Sens. 196, 146–177. http://dx.doi.org/10.1016/j.isprsjprs.2022.12.021.

Zhu, X., Zhou, H., Wang, T., Hong, F., Ma, Y., Li, W., Li, H., Lin, D., 2021. Cylindrical and asymmetrical 3D convolution networks for LiDAR segmentation. In: Proc. CVPR. IEEE, pp. 9939–9948.

Zou, Y., Yu, Z., Liu, X., Kumar, B.V.K.V., Wang, J., 2019. Confidence regularized self-training. In: Proc. ICCV. IEEE, pp. 5981–5990. http://dx.doi.org/10.1109/ICCV.2019.00608.