Computers, Materials & Continua

Tech Science Press

https://doi.org/10.32604/cmc.2026.086285

ARTICLE

Mamba-Transformer: A LiDAR Semantic Segmentation Algorithm for Adverse Weather

Lili Zhang, Jinming Cheng* and Yingyou Wen

School of Computer Science and Engineering, Northeastern University, Shenyang, China

*Corresponding Author: Jinming Cheng. Email: chengjm@mails.neu.edu.cn

Received: 27 May 2026; Accepted: 20 August 2026; Published: 22 October 2026

ABSTRACT: Adverse weather conditions substantially degrade the performance of light detection and ranging (LiDAR) semantic segmentation because weather-induced interference often intensifies point-cloud noise, weakens the representation of long-range objects, and causes the loss of local structural details. These degradations make it difficult for segmentation models to maintain reliable perception across regions with different levels of uncertainty and complexity. To address these challenges, this paper proposes a reliability-guided Mamba–Transformer hybrid framework for adverse-weather LiDAR semantic segmentation. The proposed method first evaluates the reliability and difficulty of adverse-weather point clouds and constructs region-level priors to guide subsequent feature modeling and semantic inference. Based on these priors, the Mamba architecture is employed to capture global contextual dependencies, enabling effective modeling of long-range semantic relationships in degraded point clouds. Meanwhile, a Transformer module is introduced to perform local refinement on critical and difficult regions, thereby enhancing the discrimination of complex structures and improving segmentation robustness. Finally, an adaptive gating mechanism fuses global and local multi-source features so that complementary information from different modeling branches can be integrated for final semantic prediction. Experimental results on the SemanticSTF dataset demonstrate that the proposed method achieves strong overall performance in adverse-weather LiDAR semantic segmentation and shows advantages in overall segmentation accuracy and the recognition of complex regions.

KEYWORDS: LiDAR semantic segmentation; adverse weather; reliability guidance; Mamba; Transformer; RMT-Seg

1 Introduction

Light detection and ranging (LiDAR) has become an indispensable sensing modality for three-dimensional environmental perception in autonomous driving and intelligent robotics, owing to its accurate depth perception capability. As a fundamental component of scene understanding, LiDAR point-cloud semantic segmentation requires not only accurate object recognition and boundary delineation, but also efficient inference for practical deployment. In recent years, LiDAR semantic segmentation has developed along multiple technical directions, including range-view projection, polar or bird's-eye-view (BEV) grid representation, and enhanced contextual encoding. Representative range-view and projection-based methods such as RangeNet++, SalsaNext, PolarNet, and CENet have advanced this task from the perspectives of representation design, network architecture, and real-time inference [1–4]. Point-based networks such as PointNet, PointNet++, KPConv, and RandLA-Net have also established direct or efficient processing schemes for large-scale point clouds [5–8]. Sparse convolutional and point-voxel frameworks, including MinkowskiNet, SPVNAS, Cylinder3D, and Point Transformer, further improve three-dimensional feature encoding for

Copyright © 2026 The Authors. Published by Tech Science Press. This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

outdoor LiDAR scenes [9–12]. In addition, public autonomous-driving benchmarks such as SemanticPOSS, nuScenes, Waymo Open Dataset, CADC, and KITTI have provided important data foundations for LiDAR perception research [13–17]. However, most existing mainstream models are trained and evaluated on standardized datasets collected under favorable weather conditions, and their perceptual robustness and stability in complex degraded scenarios, such as rain, snow, and fog, remain insufficiently explored.

Unlike regular scenes, LiDAR point clouds captured under adverse weather suffer from more severe information degradation and quality imbalance. Due to the absorption and scattering of infrared laser pulses by suspended particles and precipitation in the atmosphere, the received point clouds are often affected by numerous outlier pseudo-points, severe echo attenuation, and structural incompleteness of long-range objects. Studies such as SemanticSTF have shown that three-dimensional semantic segmentation models that perform well under normal weather conditions generally experience a sharp performance decline when facing such degraded point clouds [18]. To alleviate this problem, recent studies have begun to investigate robust semantic segmentation under adverse weather. For example, UniMix improves model stability under complex weather conditions from the perspectives of data mixing and environmental robustness enhancement [19], while other works address the long-tailed distribution of data through data augmentation and annotation-efficiency-oriented strategies [20,21]. Related LiDAR denoising and weather-simulation studies have further shown that fog, snowfall, and other adverse conditions can introduce non-negligible perturbations into point clouds [22–24]. Although these strategies improve generalization to some extent, they mainly focus on data-level processing, leaving an important technical gap in explicitly modeling the reliability of degraded regions according to their underlying physical characteristics.

At the feature modeling level, the extreme sparsity and strong noise interference of adverse-weather point clouds make it particularly important for networks to capture both long-range spatial dependencies and local structural patterns. Traditional convolutional neural networks are limited by local receptive fields and therefore struggle to maintain global semantic consistency under strong noise. Transformer-based methods represented by SphereFormer provide strong global relationship modeling capability, but their full-scale self-attention mechanisms usually introduce substantial computational and memory overhead, making them difficult to deploy in lightweight settings [25,26]. Window-based attention represented by Swin Transformer reduces the computational burden of vision Transformers, but local attention alone is still insufficient for long-range degraded point clouds [27]. Recently, state space models (SSMs), represented by Mamba and VMamba, have offered a new paradigm for efficient global feature extraction because of their linear-time complexity and strong long-sequence modeling capability [28,29]. This development is closely related to structured state-space sequence modeling and recent bidirectional visual SSM designs [30,31]. SSMs have also been successfully introduced into three-dimensional point-cloud analysis by works such as Point Cloud Mamba and Mamba3D [32,33]. Nevertheless, pure state space models remain limited when handling complex local geometric details. More importantly, existing methods generally lack explicit characterization of information reliability and segmentation difficulty in degraded regions [18–21]. As a result, noisy regions and real physical structures are often incorporated into the same unified modeling process. This not only wastes computational resources, but also easily contaminates global context with noise and weakens the effectiveness of local detail discrimination.

Based on the above observations, this paper argues that the key to adverse-weather LiDAR semantic segmentation is not simply to expand the range of global modeling or continuously increase the computational cost of local self-attention. Instead, computational resources should be dynamically and selectively allocated according to the information quality and physical reliability of different regions. The core challenge in adverse-weather scenarios lies in the highly uneven distribution of information degradation; therefore, a more appropriate strategy is to assess first and then allocate. To this end, this paper proposes a

reliability-guided Mamba–Transformer hybrid segmentation framework, termed RMT-Seg. The framework first employs a dual-head evaluation module to explicitly predict the reliability and segmentation difficulty of feature maps based on physical and geometric priors. It then uses a reliability-modulated Mamba backbone to perform efficient global contextual scanning, thereby suppressing noise propagation. Finally, for a small number of difficult key regions selected by the difficulty map, a lightweight local Transformer is adopted for targeted refinement. The algorithmic contribution is this assessment-driven allocation: reliability modulates global state propagation, whereas difficulty activates local attention only where needed.

The main contributions of this paper are summarized as follows:

- A reliability-guided hybrid framework, RMT-Seg, is proposed for adverse-weather LiDAR semantic segmentation. It combines the efficient global modeling capability of state space models with the local fine-grained modeling capability of Transformers, improving segmentation robustness in degraded scenarios without introducing excessive computational overhead.
- A reliability- and difficulty-aware mechanism based on physical and geometric priors is designed. This mechanism explicitly identifies low-confidence regions affected by weather degradation and uses them as guidance signals to dynamically modulate subsequent global feature propagation paths and the selection of difficult key windows.
- A lightweight local Transformer refinement module and an adaptive gated fusion strategy are constructed. Through targeted local enhancement and consistency constraints, the proposed method improves the delineation of long-range small objects and semantic boundaries in complex degraded regions, and demonstrates strong overall performance and potential for lightweight deployment on the real-world adverse-weather SemanticSTF benchmark.

2 Method

2.1 Overall Framework

The overall workflow of the RMT-Seg framework consists of three stages. First, because raw three-dimensional point clouds are unordered and sparse, the data representation and feature initialization stage projects the input data into a regularly arranged two-dimensional range-view representation and extracts initial features containing physical and geometric attributes through a lightweight encoder, thereby establishing a unified computational carrier for subsequent processing. Second, in the reliability-guided global modeling stage, the model employs a dedicated assessment module to identify the quality of local features and generate reliability and difficulty guidance maps. On this basis, a multi-directional scanning Mamba backbone is adopted to efficiently capture global long-range dependencies under the modulation of reliability signals. Finally, in the key-region refinement and fusion decision stage, the model selects difficult local windows according to the difficulty map, invokes a lightweight Transformer to enhance local feature discrimination within the 2D range-view representation, and fuses global priors with locally refined features through an adaptive gating mechanism to output point-wise semantic labels.

To facilitate an overall understanding of the above workflow, Fig. 1 presents the system architecture of RMT-Seg. Starting from the range-view projection of the input point cloud, the figure sequentially illustrates the key components, including initial feature encoding, reliability and difficulty dual-head assessment, reliability-modulated Mamba global modeling, difficulty-map-based hard-window selection, local Transformer refinement, and gated fusion classification. The contribution lies not in the individual backbone components, but in using learned reliability and difficulty signals to coordinate global propagation and selective local attention asymmetrically.

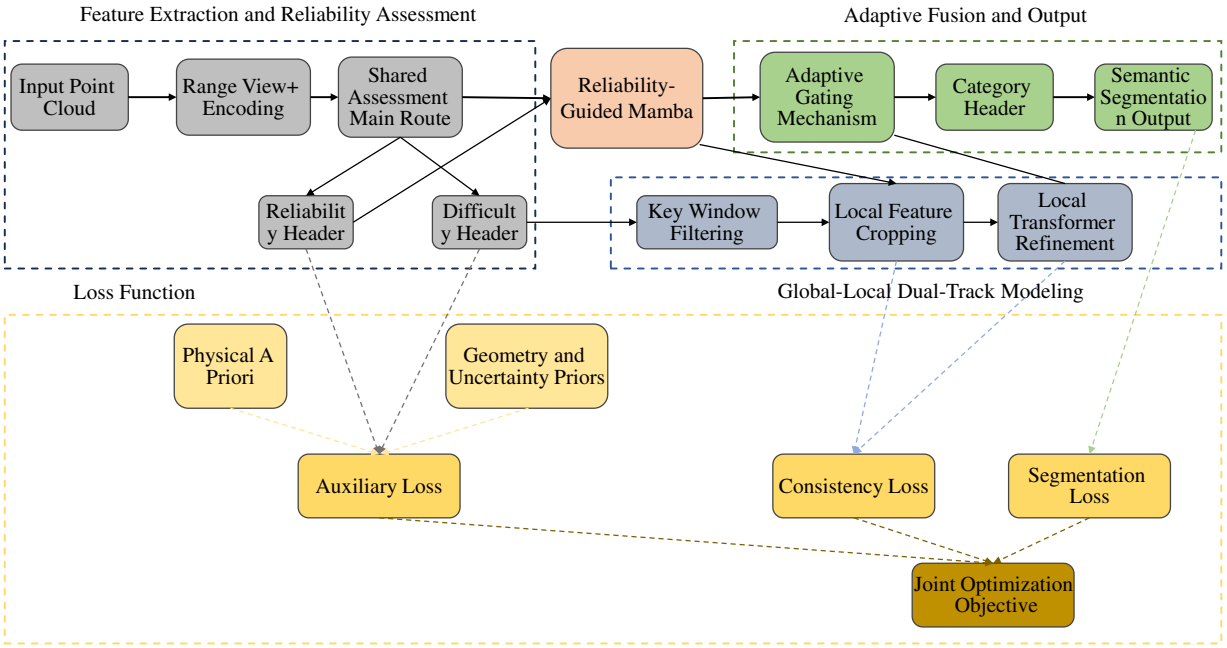


Figure 1: Overall system architecture of RMT-Seg.

2.2 Reliability-Guided Global Modeling

Under adverse weather conditions, different regions are often affected by weather interference to different degrees, resulting in significant variations in point-cloud quality. To more accurately evaluate the data quality of local regions, this paper introduces a reliability assessment module before backbone feature extraction. This module takes the initial feature tensor $\boldsymbol{F}_0 \in \mathbb{R}^{H\times W\times C}$ as input, where H, W, and C denote the height, width, and number of channels of the range-view feature map, respectively. It first extracts degradation-aware features through a shared lightweight network, and then uses two branches to predict a single-channel reliability score map $\boldsymbol{R} \in \mathbb{R}^{H\times W\times 1}$ and a single-channel difficulty score map $\boldsymbol{D} \in \mathbb{R}^{H\times W\times 1}$. The computation can be formulated as

$$\boldsymbol{F}_{\mathrm{a}} = \phi_{\mathrm{assess}}(\boldsymbol{F}_0), \quad \boldsymbol{R} = \phi_{\mathrm{rel}}(\boldsymbol{F}_{\mathrm{a}}), \quad \boldsymbol{D} = \phi_{\mathrm{diff}}(\boldsymbol{F}_{\mathrm{a}}). \tag{1}$$

The reliability map reflects the confidence of information in the current region and helps identify regions severely contaminated by noise, whereas the difficulty map marks difficult regions that require focused refinement according to the complexity of local structures. This shared-front-end and dual-head-output design enables the model to suppress low-quality information while allocating computational resources based on unified features.

After obtaining the above assessment results, the model further extracts global features. For large two-dimensional range-view maps, conventional Transformer architectures usually introduce excessive computational cost, whereas the Mamba architecture can capture long-range contextual information with lower linear complexity. The two-dimensional selective scanning structure adopted in this paper is illustrated in Fig. 2. Inspired by the cross-directional scanning strategy of VMamba, this structure unfolds range-view features along horizontal, vertical, and reverse paths, enabling the state space model to cover long-range dependencies in the two-dimensional space while maintaining linear complexity. Although range-view projection does not preserve exact 3D neighborhoods, each pixel retains its original coordinates, range,

and intensity; the scans model context along the LiDAR azimuth–elevation lattice, while local refinement enhances feature representation in structurally complex regions.

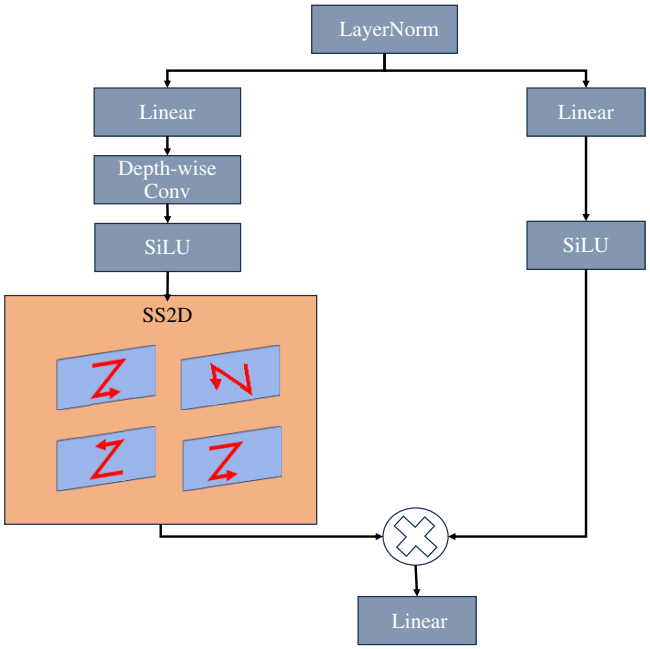


Figure 2: VMamba architecture and two-dimensional multi-directional scanning.

To preserve the two-dimensional structure of the range view, this paper adopts a multi-directional two-dimensional scanning method that scans features along the forward and reverse directions of both the horizontal and vertical axes. Meanwhile, to prevent low-quality noise from spreading in global features, the reliability map $\boldsymbol{R}$ is used as a weight to pre-modulate the input features:

$$\boldsymbol{F}_0^{\mathrm{r}} = \boldsymbol{R} \odot \boldsymbol{F}_0 + (1 - \boldsymbol{R}) \odot \psi(\boldsymbol{F}_0), \tag{2}$$

where $\odot$ denotes element-wise multiplication and $\psi(\cdot)$ denotes a lightweight feature transformation operation. The adjusted features $\boldsymbol{F}_0^{\mathrm{r}}$ are then fed into the two-dimensional Mamba encoder to generate the global feature map $\boldsymbol{F}_{\mathrm{g}} \in \mathbb{R}^{H \times W \times C}$. This design dynamically adjusts the contribution of different regions during feature propagation, helping weaken the influence of contaminated regions. At the same time, it allows reliable regions to play a more important role in global feature extraction.

2.3 Key-Region Refinement and Gated Fusion

Although the global feature map $\boldsymbol{F}_{\mathrm{g}}$ provides overall scene context, it may still be insufficient for handling local details such as blurred boundaries and incomplete small objects. Therefore, this paper proposes a key-region refinement strategy based on the difficulty guidance map $\boldsymbol{D}$. Specifically, the range-view map with spatial size $H \times W$ is first uniformly divided into local windows of size $P \times P$. The windows are then ranked according to the accumulated responses of difficulty scores within each window, and the top K difficult windows are selected to form the set $\Omega = \{\omega_k\}_{k=1}^{K}$. For these difficult regions, the model crops the corresponding sub-blocks $\boldsymbol{F}_{\mathrm{g}}^{(k)} \in \mathbb{R}^{P \times P \times C}$ from the global feature map and uses a lightweight Transformer module for local enhancement. The self-attention mechanism of the Transformer helps more accurately capture spatial structures within each window and outputs locally refined features $\boldsymbol{F}_{\mathrm{l}}^{(k)}$. This

sparse computation design, which applies Transformer refinement only to difficult windows, improves local accuracy while avoiding the high cost of full-image computation.

Subsequently, the scatter operator $\mathcal{S}(\cdot)$ restores these K local features to their original spatial positions, while non-critical regions are filled with zeros, producing a dense feature map $\tilde{\boldsymbol{F}}_{\mathrm{l}} \in \mathbb{R}^{H\times W\times C}$ aligned with the original map. To integrate the global features $\boldsymbol{F}_{\mathrm{g}}$ and the locally refined features $\tilde{\boldsymbol{F}}_{\mathrm{l}}$, this paper designs an adaptive gated fusion module. The module first concatenates multi-source features along the channel dimension, and then dynamically generates a spatial weight map $\boldsymbol{G} \in [0,1]^{H\times W\times C}$ through a lightweight network and a sigmoid function:

$$\boldsymbol{G} = \sigma\big(\phi_{\mathrm{gate}}([\boldsymbol{F}_{\mathrm{g}}, \tilde{\boldsymbol{F}}_{\mathrm{l}}, \boldsymbol{R}, \boldsymbol{D}])\big). \tag{3}$$

The global and local features are then weighted and fused using this map, and the reliability and difficulty maps are further incorporated as auxiliary features to obtain the final fused representation $\boldsymbol{F}_{\mathrm{fuse}}$:

$$\boldsymbol{F}_{\mathrm{fuse}} = \boldsymbol{G} \odot \tilde{\boldsymbol{F}}_{\mathrm{l}} + (1-\boldsymbol{G}) \odot \boldsymbol{F}_{\mathrm{g}} + \phi_{\mathrm{aux}}(\boldsymbol{R}, \boldsymbol{D}). \tag{4}$$

Through this gating mechanism, the model can adaptively balance feature weights. In regular clear regions, the model emphasizes global features; in disturbed difficult regions, it assigns greater importance to locally refined features. Finally, the fused features are fed into the classification head ϕ_{cls} to output point-wise semantic predictions $\hat{\boldsymbol{Y}}$.

2.4 Joint Optimization Objective

To balance overall segmentation accuracy, local detail refinement capability, and robustness under adverse weather, this paper designs a joint optimization objective from three aspects. First, the main semantic segmentation task is supervised by a combination of the standard cross-entropy loss and the Lovasz-Softmax loss, aiming to improve global category prediction accuracy and alleviate the long-tailed distribution problem [34]. This loss is directly applied to the final classification output:

$$\mathcal{L}_{\mathrm{seg}} = \mathcal{L}_{\mathrm{ce}}(\hat{\boldsymbol{Y}}, \boldsymbol{Y}) + \lambda_1 \mathcal{L}_{\mathrm{lovasz}}(\hat{\boldsymbol{Y}}, \boldsymbol{Y}), \tag{5}$$

where $\boldsymbol{Y}$ denotes the ground-truth semantic labels and λ_1 is a balancing coefficient. Second, because direct ground-truth labels are unavailable for the reliability and difficulty assessment module, this paper introduces pseudo-supervision signals based on physical and geometric priors. For the reliability pseudo-label $\boldsymbol{R}^*$, the normalized radial distance $\tilde{\boldsymbol{Z}}$, normalized reflection intensity $\tilde{\boldsymbol{I}}$, and local depth gradient $\tilde{\boldsymbol{E}}$ are jointly considered, and the pseudo-label is computed as

$$\boldsymbol{R}^* = \mathrm{clip}\left(1 - \eta_1 \tilde{\boldsymbol{Z}} - \eta_2 (1-\tilde{\boldsymbol{I}}) - \eta_3 \tilde{\boldsymbol{E}},\ 0,\ 1\right). \tag{6}$$

This constraint reflects the physical rule that point-cloud information becomes less reliable when the distance is longer, the echo intensity is lower, or the local depth variation is stronger. Similarly, the difficulty pseudo-label $\boldsymbol{D}^*$ is constructed from the semantic boundary response $\boldsymbol{B}$ extracted from the ground truth, the normalized local sparsity $\tilde{\boldsymbol{V}}$, and the prediction uncertainty $\tilde{\boldsymbol{U}}$ of the main branch, which is consistent with uncertainty modeling in Bayesian deep learning for visual perception [35]:

$$\boldsymbol{D}^* = \mathrm{clip}\left(\gamma_1 \boldsymbol{B} + \gamma_2 \tilde{\boldsymbol{V}} + \gamma_3 \tilde{\boldsymbol{U}},\ 0,\ 1\right). \tag{7}$$

Based on these pseudo-labels, the auxiliary loss of the reliability assessment module $\mathcal{L}_{\text{rel}}$ is defined as the error between the predicted maps and the pseudo-labels:

$$\mathcal{L}_{\text{rel}} = \mathcal{L}_{\text{r}}(\boldsymbol{R}, \boldsymbol{R}^*) + \lambda_2 \mathcal{L}_{\text{d}}(\boldsymbol{D}, \boldsymbol{D}^*). \tag{8}$$

The physical and geometric priors act as soft, domain-informed supervision rather than deterministic guidance, fixed masks, or decision rules. The learnable dual-head network predicts input-dependent reliability and difficulty maps; at inference, pseudo-label construction is removed and segmentation uses the predicted maps and learned features. Thus, the priors provide domain-informed auxiliary signals that regularize data-driven learning. However, the current experiments do not establish reduced training-data requirements because all models are trained using the full training set.

Finally, to ensure the effectiveness of the local Transformer refinement module and prevent gradient drift or overfitting within difficult windows, this paper introduces a consistency constraint. Let $\hat{\boldsymbol{Y}}^{(k)}_{\text{pre}}$ and $\hat{\boldsymbol{Y}}^{(k)}_{\text{post}}$ denote the class-distribution predictions before and after local refinement in the k-th key window, respectively. The refinement loss $\mathcal{L}_{\text{ref}}$ is then defined by computing the Kullback–Leibler divergence between them:

$$\mathcal{L}_{\text{ref}} = \frac{1}{|\Omega|} \sum_{\omega_k \in \Omega} \text{KL}\left(\hat{\boldsymbol{Y}}^{(k)}_{\text{post}} \,\|\, \hat{\boldsymbol{Y}}^{(k)}_{\text{pre}}\right). \tag{9}$$

Combining the above terms, the total loss function of the RMT-Seg framework is expressed as

$$\mathcal{L} = \mathcal{L}_{\text{seg}} + \alpha \mathcal{L}_{\text{rel}} + \beta \mathcal{L}_{\text{ref}}, \tag{10}$$

where α and β are hyperparameters controlling the weights of the corresponding loss terms. Through this multi-objective joint optimization strategy, the network can achieve collaborative convergence among global context modeling, local detail refinement, and quality assessment modules.

3 Experiments and Analysis

3.1 Dataset

The core validation experiments are conducted on the SemanticSTF dataset. SemanticSTF is a large-scale real-world physical benchmark specifically constructed for adverse-weather three-dimensional point-cloud semantic segmentation, and its underlying road-test data are derived from the public multi-modal perception dataset Seeing Through Fog (STF) [36]. SemanticSTF was collected using a 64-beam LiDAR sensor on real open roads, recording the absorption and scattering effects of suspended particles and precipitation on infrared laser pulses. The dataset contains 2076 representative adverse-weather point-cloud scans, including 694 snowy scenes, 637 dense-fog scenes, 631 light-fog scenes, and 114 rainy scenes. It also provides high-quality dense point-wise semantic annotations, covering 21 semantic categories aligned with the SemanticKITTI standard [37]. Since all data are collected in real rain, snow, and fog environments, the scenes contain numerous local distortions, long-range object sparsification, and outlier pseudo-points that follow realistic optical attenuation patterns. All experiments strictly follow the official training, validation, and test splits of the dataset to ensure objective and fair quantitative comparison with other state-of-the-art baseline methods.

3.2 Evaluation Metrics

To comprehensively quantify the proposed framework in terms of semantic segmentation accuracy, perceptual robustness, and computational efficiency, the following metrics are adopted. For segmentation accuracy and robustness, mean intersection over union (mIoU) is used as the primary evaluation metric. Considering the long-range information attenuation characteristics under adverse weather, distance-stratified mIoU metrics are further introduced to report performance within the ranges of 0–20 m (mIoU-20), 20–40 m (mIoU-40), and beyond 40 m (mIoU-40+). By comparing the variation of mIoU across different distance ranges, the perceptual stability of the model under complex degraded conditions can be evaluated more objectively. In addition, to further evaluate performance robustness across different weather subsets, this paper reports the per-subset mIoU under four types of weather conditions: snow, light fog, dense fog, and rain.

For model complexity and runtime efficiency, the number of parameters (Params), floating-point operations (FLOPs), and inference frame rate (FPS) are jointly evaluated. Params reflects the storage resource consumption of the model, FLOPs measures the computational cost of a single forward pass, and FPS reflects network-level runtime efficiency under the specified hardware and inference protocol.

3.3 Experimental Setup

To feed sparse three-dimensional point clouds into the network, the experiments first project them onto regular two-dimensional range-view maps through spherical projection. Each projected pixel retains the physical attributes of the original point, including three-dimensional spatial coordinates (X, Y, Z), radial distance, and reflection intensity, and all attributes are normalized to zero mean before being read by the network.

The proposed algorithm is implemented and evaluated using the PyTorch deep learning framework [38]. All training and testing procedures are performed on an NVIDIA GeForce RTX 4060 GPU. To improve the robustness of the model to spatial variations in adverse environments, data augmentation strategies, including random rotation around the Z-axis, spatial scaling, and horizontal flipping, are introduced during the training input stage.

During network optimization, the AdamW optimizer is adopted to effectively suppress weight overfitting [39]. The initial learning rate is set to 0.001, and the weight decay coefficient is set to 0.01. The entire training process lasts for 150 epochs, with a batch size of 16 for each iteration. The learning rate is scheduled using cosine annealing [40], and linear warmup is applied during the first 5 epochs to ensure stable gradient descent at the beginning of training.

For the weighting coefficients in the reliability and difficulty pseudo-labels, this paper determines them by combining empirical settings with validation-set fine-tuning. Considering that adverse-weather point-cloud degradation is mainly affected by long-range sparsification and low reflection intensity, while local depth discontinuities may also introduce noise interference, the coefficients in the reliability pseudo-label are set to $\eta_1 = 0.4$, $\eta_2 = 0.4$, and $\eta_3 = 0.2$. In the construction of the difficulty pseudo-label, semantic boundaries are usually the regions where local mis-segmentation is most concentrated; therefore, they are assigned a higher weight. Local sparsity and prediction uncertainty are treated as auxiliary difficulty factors with relatively lower weights, and the corresponding coefficients are set to $\gamma_1 = 0.5$, $\gamma_2 = 0.3$, and $\gamma_3 = 0.2$. In the total loss function, the weights of the auxiliary reliability constraint and the local refinement consistency constraint are set to $\alpha = 0.2$ and $\beta = 0.1$, respectively, to avoid excessive interference from auxiliary tasks on the main segmentation loss during the early training stage. These coefficients are first initialized according to the normalized numerical ranges of the corresponding physical quantities, and then slightly adjusted

around candidate values on the validation set. The final configuration is selected based on stable mIoU and long-range mIoU-40+ performance.

All methods are evaluated under supervised target-domain training using the official SemanticSTF splits. The comparisons include RangeNet++, SalsaNext, CENet, PolarNet, a Transformer baseline, UniMix implemented on the CENet range-view baseline, and a Pure-Mamba control model.

To facilitate reproducibility, the complete PyTorch implementation of the RMT-Seg framework, including the pseudo-label construction, difficulty-based window selection, Rel-Mamba scanning block, local Transformer, and gated fusion mechanism, along with the detailed configurations, is publicly available at https://github.com/Missss-ll/RTSEG.

3.4 Experimental Results and Analysis

Table 1 reports the comprehensive performance and runtime efficiency of RMT-Seg and the baseline methods on the SemanticSTF test set under the supervised target-domain training setting. In addition to quantitative metrics, Fig. 3 further presents qualitative semantic segmentation comparisons under different adverse weather scenarios, linking the numerical differences in the table with actual segmentation effects. The figure selects one to two representative range-view or point-cloud projection samples from snowy, light-fog, dense-fog, and rainy scenes, and simultaneously shows the input, ground truth, Transformer baseline prediction, Pure-Mamba prediction, and RMT-Seg prediction. Long-range objects, small-scale traffic participants, traffic signs, semantic boundaries, and noise-contaminated regions are highlighted to intuitively illustrate the differences among models in terms of mis-segmentation, missed detection, and boundary delineation under degraded scenes.

RMT-Seg achieves 54.8% mIoU, exceeding RangeNet++, SalsaNext, CENet, and UniMix by 7.6, 5.8, 4.4, and 2.0 points, respectively. Its mIoU-40+ reaches 41.2%, compared with 38.0% for UniMix and 39.6% for the Transformer baseline, while remaining smaller than the Transformer baseline.

Considering the model design together with the overall performance changes, the advantage of RMT-Seg is more evident in the recognition and segmentation of small-scale objects, such as pedestrians and traffic signs, as well as complex boundary structures. In real rain, snow, and fog environments, scattering and absorption effects tend to degrade the observable local structure of these fine-grained targets. By extracting difficult windows and applying local Transformer refinement, RMT-Seg helps refine locally ambiguous features while avoiding the high computational cost of full-image self-attention.

Table 1: Comprehensive performance and runtime efficiency comparison between RMT-Seg and baseline methods on the SemanticSTF test set.

Method	mIoU (%)	mIoU-20 (%)	mIoU-40 (%)	mIoU-40+ (%)	Params (M)	FLOPs (G)	FPS
RangeNet++	47.2	60.5	48.1	31.0	1.8	34.2	42.3
SalsaNext	49.0	61.8	50.1	33.5	6.7	49.5	29.8
CENet	50.4	62.6	51.4	34.9	7.4	53.1	25.6
PolarNet	46.5	59.1	46.8	29.9	14.0	22.8	37.6
UniMix (CENet-based)	52.8	64.0	53.5	38.0	–	–	–
Transformer Baseline	53.4	65.9	54.7	39.6	12.9	87.4	14.2
Pure-Mamba	51.6	64.4	52.5	36.6	8.9	39.6	29.4
RMT-Seg (Ours)	54.8	66.7	56.0	41.2	10.1	45.8	24.7

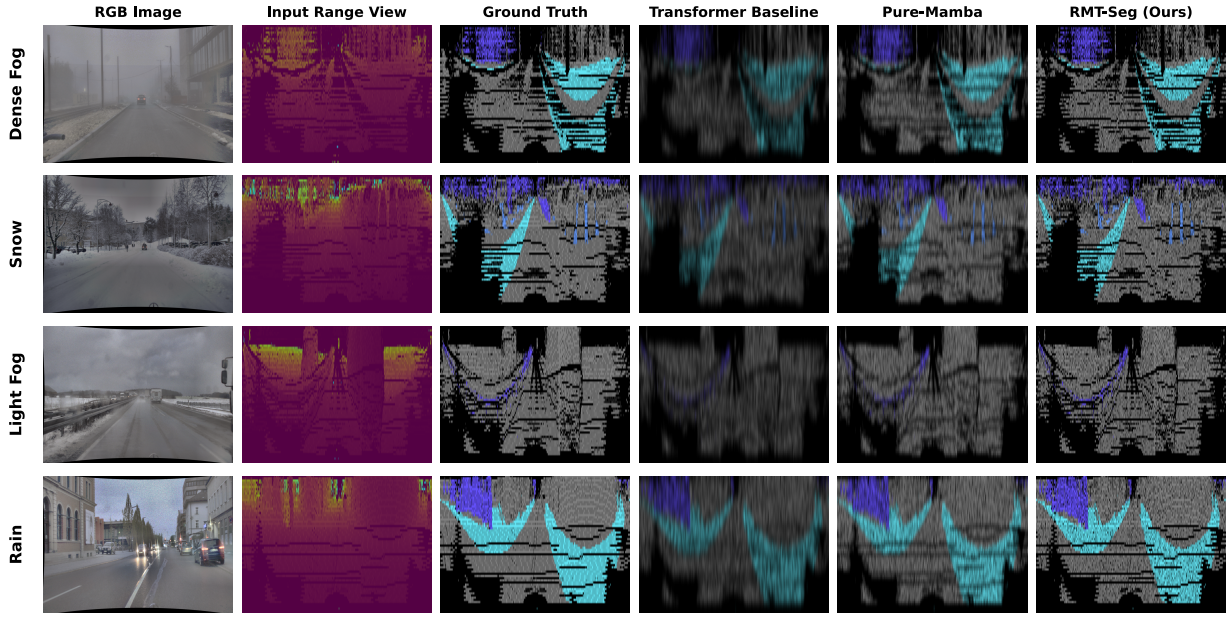


Figure 3: Qualitative semantic segmentation comparison under different adverse weather scenarios.

Regarding the accuracy–efficiency trade-off, Table 1 compares RMT-Seg with the Transformer and Pure-Mamba baselines. Relative to the Transformer, RMT-Seg improves mIoU by 1.4 points while reducing parameters by 21.7% (12.9M to 10.1M) and FLOPs by 47.6% (87.4G to 45.8G), and increasing speed by 73.9% (14.2 to 24.7 FPS). Relative to Pure-Mamba, selective local refinement adds only 1.2M parameters and 6.2G FLOPs while improving mIoU and mIoU-40+ by 3.2 and 4.6 points, respectively.

To evaluate practical deployment potential, inference was conducted on an NVIDIA Jetson Orin Nano configured in the 15 W power mode, using TensorRT 8.5 and CUDA 11.4. The evaluation used an input range-view resolution of 64×2048, a batch size of 1, and FP16 precision. Following 100 warm-up iterations, performance was measured over 1000 consecutive timed runs. Under this protocol, the model achieved an average network-forward latency of 46.7 ms per scan, with a latency variability of ±1.2 ms, corresponding to 21.4 FPS. The peak memory consumption was approximately 1.6 GB. The reported latency covers only TensorRT engine execution and excludes range-view projection, input preprocessing, output postprocessing, and point-cloud unprojection. Although this network-level throughput exceeds the input rate of a typical 10-Hz LiDAR sensor under the tested configuration, it demonstrates the edge-deployment potential of the proposed framework rather than establishing end-to-end real-time performance for a complete autonomous-driving system.

In the near-range clear interval of 0–20 m, the differences among mainstream models are usually not substantial. However, when the distance increases beyond 40 m, the point-cloud distribution becomes extremely sparse and is accompanied by stronger outlier noise, leading to a significant accuracy drop in conventional lightweight algorithms. In contrast, RMT-Seg shows a milder decline in the mIoU-40+ metric. This indicates that reliability-modulated global scanning and local on-demand refinement more effectively suppress the influence of long-range degraded regions on the overall segmentation results.

As shown in Table 2, RMT-Seg achieves the highest mIoU across all four weather subsets. It exceeds Pure-Mamba by 3.5 points in dense fog and 3.1 points in rain, and exceeds UniMix by 2.4 points in rain. These measured gains indicate more consistent performance across the evaluated weather conditions.

Table 2: Segmentation performance comparison under different weather conditions.

Method	Snow (%)	Light Fog (%)	Dense Fog (%)	Rain (%)
RangeNet++	48.0	47.2	45.8	43.5
SalsaNext	49.8	49.2	47.5	45.3
PolarNet	47.3	46.5	45.1	42.8
CENet	51.0	50.4	49.1	46.8
UniMix (CENet-based)	53.5	53.0	52.0	48.5
Transformer Baseline	54.0	53.4	52.1	49.0
Pure-Mamba	52.4	51.8	50.3	47.8
RMT-Seg (Ours)	55.4	55.0	53.8	50.9

To further explain the source of the above performance improvement, Fig. 4 visualizes the reliability map, difficulty map, and key-window selection results. The figure presents the original range view, the predicted reliability heatmap $\boldsymbol{R}$, the difficulty heatmap $\boldsymbol{D}$, and the selected local Transformer refinement windows side by side. Generally, the reliability map exhibits lower responses in long-range sparse regions, low-echo-intensity regions, and areas near noisy pseudo-points, whereas the difficulty map is more concentrated around semantic boundaries, small objects, and locally fragmented structures. This visualization more intuitively shows that RMT-Seg does not uniformly allocate computational resources to the entire range view, but instead implements targeted modeling according to regional quality and segmentation difficulty through an "assess first, allocate later" strategy.

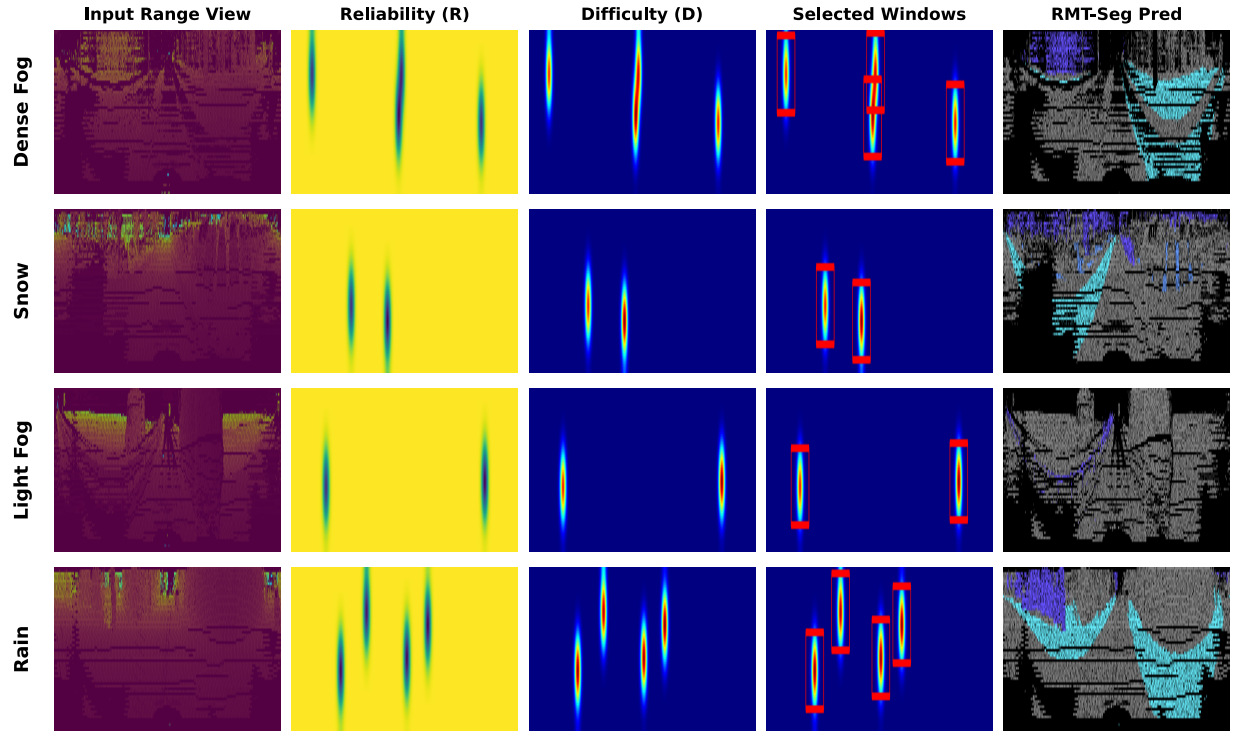


Figure 4: Visualization of the reliability map, difficulty map, and difficult-window selection results.

3.5 Ablation Study

To verify the actual contribution of each key component and the joint optimization objective in RMT-Seg, this paper conducts systematic ablation experiments under the same supervised target-domain training configuration. The experimental baseline is set as a basic network containing only the initial feature encoder and the conventional segmentation loss. Table 3 shows the performance changes after progressively introducing each module and optimization objective. First, after introducing the reliability and difficulty dual-head assessment module on top of the baseline model, the overall mIoU increases from 48.7% to 49.8%. This demonstrates that explicitly assessing the degradation degree of local regions can provide effective priors for subsequent feature extraction. Next, after incorporating the Mamba global scanning backbone modulated by reliability signals, the mIoU further increases to 51.9%, and the mIoU-40+ increases from 34.8% to 38.1%. This result verifies the advantage of state space models in long-range dependency modeling and the effectiveness of the noise isolation mechanism. Selective local refinement then increases mIoU from 51.9% to 53.5% and mIoU-40+ from 38.1% to 39.9%, while adding 4.5G FLOPs (39.6G to 44.1G), an 11.4% increase. Finally, with adaptive gated fusion, the model reaches an overall accuracy of 54.4%, showing that dynamic coordination between global and local features is more effective than simple addition.

Table 3: Ablation study on key modules and loss terms of RMT-Seg; the final three configurations report mean ± standard deviation over three random seeds.

Configuration	Dual-Head	Rel-Mamba	Local Trans.	Gated Fusion	$\mathcal{L}_{\text{rel}}$	$\mathcal{L}_{\text{ref}}$	mIoU (%)	mIoU-40+ (%)
Baseline	–	–	–	–	–	–	48.7	33.6
+ Assessment	✓	–	–	–	–	–	49.8	34.8
Rel-Mamba	✓	✓	–	–	–	–	51.9	38.1
+ Local Refinement	✓	✓	✓	–	–	–	53.5	39.9
+ Gated Fusion	✓	✓	✓	✓	–	–	54.4 ± 0.14	40.7 ± 0.15
+ $\mathcal{L}_{\text{rel}}$	✓	✓	✓	✓	✓	–	54.6 ± 0.11	41.0 ± 0.13
+ $\mathcal{L}_{\text{ref}}$ (Full Model)	✓	✓	✓	✓	✓	✓	54.8 ± 0.10	41.2 ± 0.12

Across the gated-fusion and auxiliary-loss configurations, three-seed standard deviations remain at or below 0.15 points. The full model reaches 54.8 ± 0.10% mIoU and 41.2 ± 0.12% mIoU-40+, indicating stable run-to-run behavior. The modest auxiliary-loss gains are therefore interpreted as consistent regularization effects rather than large accuracy increases.

Finally, Table 4 evaluates both pseudo-label weight sets while keeping all other settings fixed. Alternative configurations retain 54.4%–54.6% mIoU and 40.4%–40.9% mIoU-40+, compared with 54.8% and 41.2% for the defaults; the maximum decreases are only 0.4 and 0.8 percentage points. RMT-Seg is therefore relatively insensitive to moderate coefficient changes. This is consistent with their role as auxiliary supervisory weights rather than direct segmentation rules.

Table 4: Sensitivity to reliability and difficulty pseudo-label weights.

Pseudo-Label	Configuration	Weights	mIoU (%)	mIoU-40+ (%)
Reliability η	Distance-focused	(0.60, 0.20, 0.20)	54.6	40.8
	Intensity-focused	(0.20, 0.60, 0.20)	54.5	40.5
	Equal weights	(0.33, 0.33, 0.33)	54.4	40.4
	Default (Ours)	(0.40, 0.40, 0.20)	54.8	41.2

(Continued)

Table 4 (continued)

Pseudo-Label	Configuration	Weights	mIoU (%)	mIoU-40+ (%)
Difficulty γ	Boundary-focused	(0.70, 0.15, 0.15)	54.5	40.6
	Sparsity-focused	(0.33, 0.50, 0.17)	54.6	40.9
	Equal weights	(0.33, 0.33, 0.33)	54.4	40.5
	Default (Ours)	(0.50, 0.30, 0.20)	54.8	41.2

Note: For η, weights follow (distance, intensity, depth gradient); for γ, they follow (boundary, sparsity, uncertainty).

4 Conclusions

To address the strong noise interference and object degradation challenges faced by LiDAR point-cloud semantic segmentation under adverse weather conditions, this paper proposes a reliability-guided hybrid framework, RMT-Seg. The proposed method follows an asymmetric "assess first, allocate later" computational strategy. Specifically, the model first estimates regional degradation levels through an assessment module guided by physical and geometric priors. It then employs a Mamba backbone to perform efficient long-range contextual denoising scans, and finally invokes a local Transformer only for a small number of difficult windows to conduct fine-grained local feature refinement within the 2D range-view domain. Experiments on the real-world adverse-weather SemanticSTF benchmark demonstrate that the proposed method has certain advantages in long-range object segmentation and the delineation of ambiguous and structurally incomplete boundaries, while maintaining a favorable balance between parameter scale and computational cost. Future research will focus on two directions. First, the stability and scalability of the model will be explored under more extreme weather types, larger scene scales, and higher-resolution point-cloud inputs. Second, hardware-level acceleration and model compression strategies for state space models will be investigated to further promote lightweight deployment on resource-constrained devices.

Acknowledgement: Not applicable.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm contribution to the paper as follows: conceptualization, Lili Zhang and Jinming Cheng; methodology, Lili Zhang; software, Jinming Cheng; validation, Lili Zhang, Jinming Cheng and Yingyou Wen; formal analysis, Yingyou Wen; investigation, Lili Zhang; resources, Lili Zhang; data curation, Jinming Cheng; writing—original draft preparation, Jinming Cheng; writing—review and editing, Jinming Cheng; visualization, Lili Zhang; supervision, Yingyou Wen; project administration, Yingyou Wen; funding acquisition, Lili Zhang. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The SemanticSTF dataset analyzed during the current study is openly available in its public repository. To facilitate reproducibility, the complete PyTorch implementation of the RMT-Seg framework, including the pseudo-label construction, difficulty-based window selection, Rel-Mamba scanning block, local Transformer, and gated fusion mechanism, along with the detailed configurations, is publicly available at https://github.com/Missss-ll/RTSEG.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Milioto A, Vizzo I, Behley J, Stachniss C. RangeNet++: fast and accurate LiDAR semantic segmentation. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 2019 Nov 3–8; Macau, China. p. 4213–20. doi:10.1109/IROS40897.2019.8967762.
2. Cortinhal T, Tzelepis G, Aksoy EE. SalsaNext: fast, uncertainty-aware semantic segmentation of LiDAR point clouds. In: Proceedings of the International Symposium on Visual Computing (ISVC); 2020 Oct 5–7; San Diego, CA, USA. p. 207–22. doi:10.1007/978-3-030-64559-5_16.
3. Zhang Y, Zhou Z, David P, Yue X, Xi Z, Gong B, et al. PolarNet: an improved grid representation for online LiDAR point clouds semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 14–19; Virtual. p. 9601–10. doi:10.1109/CVPR42600.2020.00962.
4. Cheng HX, Han XF, Xiao GQ. CENet: toward concise and efficient LiDAR semantic segmentation for autonomous driving. arXiv:2207.12691. 2022. doi:10.48550/arXiv.2207.12691.
5. Qi CR, Su H, Mo K, Guibas LJ. PointNet: deep learning on point sets for 3D classification and segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017 Jul 21–26; Honolulu, HI, USA. p. 652–60. doi:10.1109/CVPR.2017.16.
6. Qi CR, Yi L, Su H, Guibas LJ. PointNet++: deep hierarchical feature learning on point sets in a metric space. Adv Neural Inf Process Syst. 2017;30:5099–108.
7. Thomas H, Qi CR, Deschaud JE, Marcotegui B, Goulette F, Guibas LJ. KPConv: flexible and deformable convolution for point clouds. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV); 2019 Oct 27–Nov 2; Seoul, Republic of Korea. p. 6411–20. doi:10.1109/ICCV.2019.00651.
8. Hu Q, Yang B, Xie L, Rosa S, Guo Y, Wang Z, et al. RandLA-Net: efficient semantic segmentation of large-scale point clouds. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 14–19; Virtual. p. 11108–17. doi:10.1109/CVPR42600.2020.01112.
9. Choy C, Gwak J, Savarese S. 4D spatio-temporal ConvNets: Minkowski convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 Jun 15–20; Long Beach, CA, USA. p. 3075–84. doi:10.1109/CVPR.2019.00319.
10. Tang H, Liu Z, Zhao S, Lin Y, Lin J, Wang H, et al. Searching efficient 3D architectures with sparse point-voxel convolution. In: Proceedings of the European Conference on Computer Vision (ECCV); 2020 Aug 23–28; Glasgow, UK. p. 685–702. doi:10.1007/978-3-030-58604-1_41.
11. Zhu X, Zhou H, Wang T, Hong F, Ma Y, Li W, et al. Cylindrical and asymmetrical 3D convolution networks for LiDAR segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2021 Jun 19–25; Virtual. p. 9939–48. doi:10.1109/CVPR46437.2021.00981.
12. Zhao H, Jiang L, Jia J, Torr PHS, Koltun V. Point transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV); 2021 Oct 11–17; Virtual. p. 16259–68. doi:10.1109/ICCV48922.2021.01595.
13. Pan Y, Gao B, Mei J, Geng S, Li C, Zhao H. SemanticPOSS: a point cloud dataset with large quantity of dynamic instances. In: Proceedings of the IEEE Intelligent Vehicles Symposium (IV); 2020 Oct 19–Nov 13; Virtual. p. 687–93. doi:10.1109/IV47402.2020.9304596.
14. Caesar H, Bankiti V, Lang AH, Vora S, Liong VE, Xu Q, et al. nuScenes: a multimodal dataset for autonomous driving. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 14–19; Virtual. p. 11621–31. doi:10.1109/CVPR42600.2020.01164.
15. Sun P, Kretzschmar H, Dotiwalla X, Chouard A, Patnaik V, Tsui P, et al. Scalability in perception for autonomous driving: Waymo Open Dataset. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 13–19; Virtual. p. 2446–54. doi:10.1109/CVPR42600.2020.00252.
16. Pitropov M, Garcia DE, Rebello J, Smart M, Wang C, Czarnecki K, et al. Canadian adverse driving conditions dataset. Int J Robot Res. 2021;40(4–5):681–90. doi:10.1177/0278364920979368.
17. Geiger A, Lenz P, Stiller C, Urtasun R. Vision meets robotics: the KITTI dataset. Int J Robot Res. 2013;32(11):1231–7. doi:10.1177/0278364913491297.
18. Xiao A, Huang J, Xuan W, Ren R, Liu K, Guan D, et al. 3D semantic segmentation in the wild: learning generalized models for adverse-condition point clouds. In: Proceedings of the IEEE/CVF Conference on Computer Vision and

Pattern Recognition (CVPR); 2023 Jun 17–24; Vancouver, BC, Canada. p. 9382–91. doi:10.1109/CVPR52729.2023.00905.
19. Zhao H, Zhang J, Chen Z, Zhao S, Tao D. UniMix: towards domain adaptive and generalizable LiDAR semantic segmentation in adverse weather. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2024 Jun 16–22; Seattle, WA, USA. p. 14781–91. doi:10.1109/CVPR52733.2024.01400.
20. Park J, Kim K, Shim H. Rethinking data augmentation for robust LiDAR semantic segmentation in adverse weather. In: Proceedings of the European Conference on Computer Vision (ECCV); 2024 Sep 29–Oct 4; Milan, Italy. p. 320–36. doi:10.1007/978-3-031-72640-8_18.
21. Piroli A, Dallabetta V, Kopp J, Walessa M, Meissner D, Dietmayer K. Label-efficient semantic segmentation of LiDAR point clouds in adverse weather conditions. IEEE Robot Autom Lett. 2024;9(6):5575–82. doi:10.1109/LRA.2024.3396099.
22. Heinzler R, Piewak F, Schindler P, Stork W. CNN-based LiDAR point cloud de-noising in adverse weather. IEEE Robot Autom Lett. 2020;5(2):2514–21. doi:10.1109/LRA.2020.2972865.
23. Hahner M, Sakaridis C, Dai D, Van Gool L. Fog simulation on real LiDAR point clouds for 3D object detection in adverse weather. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV); 2021 Oct 11–17; Virtual. p. 15263–72. doi:10.1109/ICCV48922.2021.01500.
24. Hahner M, Sakaridis C, Bijelic M, Heide F, Yu F, Dai D, et al. LiDAR snowfall simulation for robust 3D object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2022 Jun 18–24; New Orleans, LA, USA. p. 16364–74. doi:10.1109/CVPR52688.2022.01588.
25. Lai X, Chen Y, Lu F, Liu J, Jia J. Spherical transformer for LiDAR-based 3D recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2023 Jun 18–23; Vancouver, BC, Canada. p. 17545–55. doi:10.1109/CVPR52729.2023.01683.
26. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Adv Neural Inf Process Syst. 2017;30:5998–6008. doi:10.65215/r5bs2d54.
27. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al. Swin Transformer: hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV); 2021 Oct 11–17; Virtual. p. 10012–22. doi:10.1109/ICCV48922.2021.00986.
28. Gu A, Dao T. Mamba: linear-time sequence modeling with selective state spaces. arXiv:2312.00752. 2023. doi:10.48550/arXiv.2312.00752.
29. Liu Y, Tian Y, Zhao Y, Yu H, Xie L, Wang Y, et al. VMamba: visual state space model. Adv Neural Inf Process Syst. 2024.
30. Gu A, Goel K, Re C. Efficiently modeling long sequences with structured state spaces. In: Proceedings of the International Conference on Learning Representations (ICLR); 2022 Apr 25–29; Virtual.
31. Zhu L, Liao B, Zhang Q, Wang X, Liu W, Wang X. Vision Mamba: efficient visual representation learning with bidirectional state space model. In: Proceedings of the International Conference on Machine Learning (ICML); 2024 Jul 21–27; Vienna, Austria.
32. Zhang T, Yuan H, Qi L, Zhang J, Zhou Q, Ji S, et al. Point Cloud Mamba: point Cloud learning via state space model. arXiv:2403.00762. 2024. doi:10.48550/arXiv.2403.00762.
33. Han X, Tang Y, Wang Z, Li X. Mamba3D: enhancing local features for 3D point cloud analysis via state space model. In: Proceedings of the ACM International Conference on Multimedia (ACM MM); 2024 Oct 28–Nov 1; Melbourne, Australia. p. 4995–5004. doi:10.1145/3664647.3681173.
34. Berman M, Triki AR, Blaschko MB. The Lovasz-Softmax loss: a tractable surrogate for the optimization of the intersection-over-union measure in neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2018 Jun 18–23; Salt Lake City, UT, USA. p. 4413–21. doi:10.1109/CVPR.2018.00464.
35. Kendall A, Gal Y. What uncertainties do we need in Bayesian deep learning for computer vision? Adv Neural Inf Process Syst. 2017;30:5574–84.

36. Bijelic M, Gruber T, Mannan F, Kraus F, Ritter W, Dietmayer K, et al. Seeing through fog without seeing fog: deep multimodal sensor fusion in unseen adverse weather. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 14–19; Virtual. p. 11682–92. doi:10.1109/CVPR42600.2020.01170.
37. Behley J, Garbade M, Milioto A, Quenzel J, Behnke S, Stachniss C, et al. SemanticKITTI: a dataset for semantic scene understanding of LiDAR sequences. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV); 2019 Oct 27–Nov 2; Seoul, Republic of Korea. p. 9297–306. doi:10.1109/ICCV.2019.00939.
38. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: an imperative style, high-performance deep learning library. Adv Neural Inf Process Syst. 2019;32:8024–35.
39. Loshchilov I, Hutter F. Decoupled weight decay regularization. In: Proceedings of the International Conference on Learning Representations (ICLR); 2019 May 6–9; New Orleans, LA, USA.
40. Loshchilov I, Hutter F. SGDR: stochastic gradient descent with warm restarts. In: Proceedings of the International Conference on Learning Representations (ICLR); 2017 Apr 24–26; Toulon, France.