Reliability-Aware Temporal Multimodal Fusion of LiDAR and 4D Radar for Adverse-Weather 3D Object Detection

Anonymous authors
Paper under double-blind review

Abstract

Robust 3D object detection under adverse weather conditions is critical for safe autonomous driving. Weather-robust 4D Radar can effectively mitigate challenges such as the loss of foreground points that occur with LiDAR under adverse weather conditions. Nevertheless, 4D Radar suffers from limitations such as low resolution, which makes LiDAR and 4D Radar fusion still struggle to fully observe targets under adverse weather conditions in individual frames. To address the limitations of individual observations, this work presents the first multi-frame temporal information fusion for LiDAR-4D Radar 3D object detection in adverse weather. However, multi-frame temporal fusion of LiDAR and 4D Radar in adverse weather faces two key challenges: the accumulation of temporal noises and unreliable historical observations. Consequently, naive temporal fusion of LiDAR and 4D Radar brings little or no gain and can even underperform the single-frame baseline. To address these challenges, we propose **TMF**, a multi-frame temporal fusion framework for adverse-weather 3D object detection. First, instead of treating all frames as equally valid evidence, the Temporal Radar Denoising (TRD) block performs reliability-aware temporal reweighting, discounting clutter at aggregation time so that noise no longer accumulates across frames. Subsequently, the Radar-Guided Temporal Compensation (RGTC) block performs sparsity-guided cross-modal retrieval, querying historical LiDAR features where 4D Radar reveals missing observations, and the Degradation-aware Gated Prior Fusion (DGPF) block gates the retrieved features against the current frame to focus on the important information. Extensive experiments on K-Radar dataset demonstrate consistent improvements across weather conditions, including an average gain of **6.49%** AP_{3D} in adverse weather.

1 Introduction

LiDAR-based 3D object detection has achieved remarkable progress in autonomous driving and robotic perception by leveraging accurate geometric representations from point clouds (Yan et al., 2018; Lang et al., 2019; Zhang et al., 2024a). However, LiDAR performance significantly deteriorates under adverse weather conditions, where scattering and attenuation effects lead to sparse observations and distorted geometric structures (Hahner et al., 2021; Liu et al., 2022). Existing approaches attempt to mitigate this issue through point cloud restoration or knowledge transfer strategies (Huang et al., 2024; Liu et al., 2025), but their reliance on degraded LiDAR measurements fundamentally limits robustness in challenging environments.

To alleviate LiDAR degradation under adverse weather conditions, recent studies have explored LiDAR-Radar fusion by leveraging Radar's robustness to environmental variations (Qian et al., 2021; Li et al., 2023). While LiDAR provides accurate geometric structures, Radar offers complementary measurements that remain reliable under challenging conditions. With the emergence of 4D Radar, additional elevation and velocity information further enhance multimodal representation learning and improve LiDAR-Radar fusion performance in diverse environments (Wang et al., 2022a; Chae et al., 2024b; Huang et al., 2025b). Recent works have further investigated advanced fusion strategies to improve the robustness of LiDAR-4D Radar-based 3D object detection (Chae et al., 2024a; Tong et al., 2026; Qi et al., 2026).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

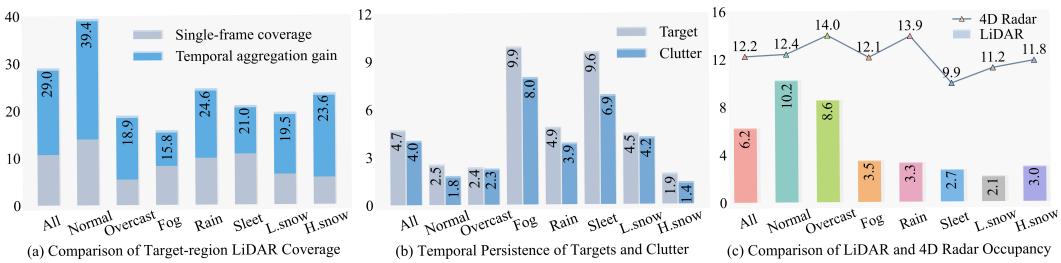


Figure 1: Observation characteristics of LiDAR and 4D Radar under adverse weather conditions. (a) Comparison of target-region LiDAR BEV occupancy between single-frame and multi-frame observations under different weather conditions. (b) Temporal persistence comparison between target and clutter responses in multi-frame 4D Radar. (c) BEV occupancy comparison between LiDAR and 4D Radar.

Despite these advances, LiDAR-4D Radar fusion still faces challenges under adverse weather conditions, where individual frames may fail to capture complete and reliable target information (Huang et al., 2025a). Figure 1(a) presents a comparison of the LiDAR BEV occupancy within the target regions under single-frame and temporal observations with multi-frame point cloud concatenation. We observe that temporal aggregation substantially enhances the coverage of target-region across diverse weather conditions. These observations indicate that temporal context can serve as a complementary source of information to mitigate degraded perception gaps caused by adverse conditions.

However, effective temporal fusion under adverse weather conditions remains challenging due to two critical issues: **(1) Temporal noise accumulation.** 4D Radar observations across different timestamps may contain inconsistent noisy responses caused by weather-induced sensing degradation. Directly aggregating multi-frame observations can amplify unreliable signals and introduce false perception cues. **(2) Unreliable historical observations.** Historical observations cannot always provide reliable temporal references under adverse weather conditions. Sensor degradation may persist across timestamps, causing previous frames to contain distorted or missing target cues. Consequently, suppressing temporal noise accumulation and exploiting reliable historical observations are indispensable for robust LiDAR-4D Radar 3D object detection in adverse weather.

Figure 1(b) compares the temporal persistence of target and clutter responses in multi-frame 4D Radar observations under different weather conditions. It can be observed that target responses preserve more consistent patterns across consecutive frames, while clutter responses show relatively weaker temporal stability. Moreover, we also observe that adverse weather significantly degrades LiDAR occupancy, while 4D Radar maintains more stable observations under various weather conditions, as shown in Figure 1(c). Therefore, developing a weather-robust multimodal temporal fusion requires **suppressing temporally inconsistent clutter** and **utilizing cross-modal cues to identify reliable historical features.**

To this end, we propose **TMF**, the first multi-frame temporal fusion framework for LiDAR-4D Radar 3D object detection under adverse weather conditions. Specifically, TMF introduces a Temporal Radar Denoising (TRD) block to suppress temporally inconsistent 4D Radar responses and mitigate noise accumulation. In addition, a Radar-Guided Temporal Compensation (RGTC) block is proposed to exploit reliable historical observations to compensate for degraded current-frame LiDAR perception gaps. Finally, the Degradation-aware Gated Prior Fusion (DGPF) block adaptively further gates the injection of compensated information into current-frame features to alleviate degraded foreground semantics under adverse weather conditions. Overall, TMF effectively exploits the reliable temporal information from LiDAR and 4D Radar to mitigate the degraded perception gaps in individual frames and improve the robustness of 3D object detection under adverse weather conditions. We conduct extensive experiments on the K-Radar dataset and demonstrate the effectiveness and robustness of the proposed TMF under various weather conditions, including an average gain of **6.49%** AP_{3D} under adverse weather conditions.

The main contributions of our work are as follows:

- We present the first adverse-weather-aware temporal multi-modal fusion approach for LiDAR and 4D Radar robust 3D object detection, termed **TMF**, which applies reliability-aware

temporal fusion to suppress temporal noise accumulation while exploiting reliable historical observations for degraded current perception compensation.

- TMF performs reliability-aware temporal reweighting to attenuate temporally incoherent clutter across consecutive frames. Meanwhile, TMF incorporates a temporal compensation strategy that employs sparsity-guided cross-modal retrieval and degradation-aware gated fusion of reliable historical information.
- Extensive experiments on the K-Radar benchmark demonstrate the effectiveness and robustness of the proposed TMF under diverse adverse weather conditions, achieving consistent improvements under different evaluation settings. In particular, TMF achieves an average improvement of **6.49%** AP_{3D} in various adverse weather scenarios.

2 Related Work

LiDAR-based 3D Object Detection: Existing studies have developed advanced geometric representation learning and detection frameworks for point cloud perception. Voxel-based and point-based methods (Lang et al., 2019; Yang et al., 2020; Deng et al., 2021) improve point cloud encoding and detection by voxelization, sparse convolution, and point-voxel feature abstraction. Recent works further explore sequential point cloud learning and domain adaptation strategies (Liu et al., 2024b; Zhang et al., 2024a; Wozniak et al., 2024) to enhance the generalization capability of LiDAR detectors. However, LiDAR perception remains vulnerable to adverse weather conditions, where point sparsity and noise significantly degrade geometric representations. To improve robustness, existing studies investigate point cloud denoising (Wang et al., 2022b; Huang et al., 2023), generative augmentation (Zhang et al., 2024c; Matteazzi et al., 2025), and knowledge distillation (Huang et al., 2024; Liu et al., 2025). However, these methods rely on single-modal LiDAR and degrade under severe weather conditions.

Radar-based 3D Object Detection: Existing Radar-based studies explore feature learning for 3D perception in challenging environments. However, early approaches based on conventional radar representations suffer from limited elevation information, which hinders accurate 3D location (Meyer et al., 2021; Li et al., 2022). With 4D Radar, elevation and velocity information can provide richer spatial-temporal cues. Existing studies investigate feature learning and spatial representation fusion strategies to enhance 3D detection under diverse conditions (Xu et al., 2021; Paek et al., 2022; 2023; Kong et al., 2023). Despite the robustness of 4D Radar, noise from multipath effects and clutter remains a challenge. Existing studies introduce preprocessing strategies and analyze radar perception (Liu et al., 2024a; Zhang et al., 2024b; Gu et al., 2025), while radar-only detection is still limited by sparsity and measurement uncertainty.

LiDAR-Radar Fusion-based 3D Object Detection: LiDAR-Radar fusion exploits the complementary characteristics of LiDAR and Radar to combine geometric information with robust environmental perception. Early methods (Qian et al., 2021; Li et al., 2023; Wang et al., 2023) focused mainly on cross-modal interaction and feature learning. With 4D Radar, recent studies (Wang et al., 2022a; Chae et al., 2024b; Huang et al., 2025b) further incorporated richer spatial and motion information for LiDAR-4D Radar fusion, achieving promising performance in challenging environments. More recent works (Chae et al., 2024a; 2025; Tong et al., 2026; Qi et al., 2026) explored radar-assisted enhancement, Doppler-aware fusion, and BEV-level multimodal representation learning to improve robustness against sensor degradation. However, existing LiDAR-4D Radar fusion methods focus on spatial feature interaction in individual frames, while the potential of temporal information to compensate for incomplete observations and suppress cross-frame noise remains largely unexplored.

3 Methodology

3.1 Problem Statement and Overall Architecture

Problem Statement: For an outdoor autonomous driving scenario, we formulate the problem of multi-frame LiDAR-4D Radar fusion for robust 3D object detection under adverse weather conditions as follows. Given a sequence of LiDAR and 4D Radar point clouds from multiple timestamps, we denote the LiDAR observations as $\mathcal{P}^L = \{\mathcal{P}^L_{t-k}, ..., \mathcal{P}^L_{t-1}, \mathcal{P}^L_t\}$ and the corresponding 4D Radar observations as $\mathcal{P}^R = \{\mathcal{P}^R_{t-k}, ..., \mathcal{P}^R_{t-1}, \mathcal{P}^R_t\}$, where t denotes the current timestamp and k represents

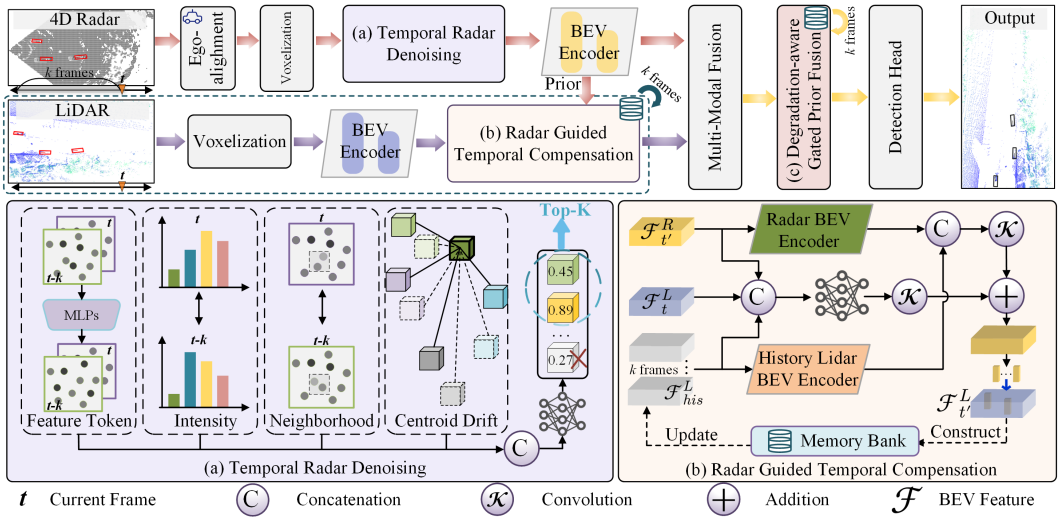


Figure 2: The proposed TMF framework consists of three key modules. First, (a) **T**emporal **R**adar **D**enoising (TRD) performs reliability-aware temporal reweighting of multi-frame 4D Radar features to suppress temporal noise accumulation. Next, (b) **R**adar-**G**uided **T**emporal **C**ompensation (RGTC) leverages 4D Radar guidance to aggregate reliable historical LiDAR cues and compensate for degraded perception gaps. Finally, (c) **D**egradation-aware **G**ated **P**rior **F**usion (DGPF) further gates the injection of compensated cues into current-frame features.

the number of historical frames. Each point p contains spatial information and sensor-specific attributes. Multi-frame multimodal perception seeks to learn a reliable temporal aggregation function $\mathcal{F}^f_{t'} = \mathcal{I}(\mathcal{P}^L_{t-k:t}, \mathcal{P}^R_{t-k:t}; \theta)$, where $\mathcal{I}(\cdot)$ is a temporal multimodal encoder that consolidates complementary information across frames and modalities, and θ represents learnable parameters. The fused temporal feature $\mathcal{F}^f_{t'}$ is then utilized for 3D object detection, with the aim of regressing a set of 3D bounding boxes $\mathcal{B}_t = \{b_i\}_{i=1}^{N_b}$, where $b_i \in \mathbb{R}^7$ indicates the center, dimensions and orientation of the object. The goal of this work is to exploit both temporal consistency and cross-modal complementarity from historical LiDAR and 4D Radar observations, compensating for degraded perception gaps and achieving robust 3D object detection under adverse weather conditions.

Overall Architecture: The overall architecture of the TMF framework is shown in Figure 2. Given multi-frame LiDAR point clouds $\mathcal{P}^L_{t-k:t}$ and 4D Radar point clouds $\mathcal{P}^R_{t-k:t}$, TMF exploits temporal complementary information to achieve robust 3D object detection under adverse weather conditions. However, temporal fusion in adverse weather faces two major challenges: the accumulation of temporal noise and unreliable historical observations. To address these challenges, TMF aims to exploit temporal multimodal observations to derive robust spatial representations from degraded inputs, thereby improving the reliability of temporal information aggregation. Specifically, the input 4D Radar and LiDAR observations are first voxelized. 4D Radar voxels are retained as $\mathcal{X}^R_{t-k:t}$ for subsequent temporal denoising, and LiDAR voxels are encoded in high-dimensional features $\mathcal{F}^L_{t-k:t}$. First, the Temporal Radar Denoising (TRD) block performs reliability-aware temporal reweighting on multi-frame 4D Radar features, producing a temporally refined 4D Radar representation $\mathcal{X}^R_{t'} = \textbf{TRD}(\mathcal{X}^R_{t-k:t})$, where unreliable clutter is suppressed to alleviate temporal noise accumulation across frames. Subsequently, the refined 4D Radar features $\mathcal{X}^R_{t'}$ are mapped to the BEV information $\mathcal{F}^R_{t'}$. Moreover, TMF proposes a temporal compensation strategy consisting of Radar-Guided Temporal Compensation (RGTC) and Degradation-aware Gated Prior Fusion (DGPF) blocks. The RGTC employs sparsity-guided cross-modal retrieval to query informative historical LiDAR features, where 4D Radar reveals missing observations, generating a compensated LiDAR representation $\mathcal{F}^L_{t'} = \textbf{RGTC}(\mathcal{F}^L_{t-k:t}, \mathcal{F}^R_{t'})$. Furthermore, DGPF gates retrieved features against the current frame to focus on important information. It derives object-aware and compensation-aware gates as $\{G_o, G_c\} = \textbf{DGPF}(\mathcal{F}^f_t, \mathcal{F}^f_{his}, \mathcal{F}^f_{his} - \mathcal{F}^f_t)$, where G_o and G_c regulate the spatial locations and contributions of temporal compensation, respectively. Consequently, multi-modal temporal features are adaptively refined to obtain fused information $\mathcal{F}^f_{t'}$, which effectively integrates retrieved historical cues with the current observations to alleviate the impact of degraded sensing conditions. Finally, $\mathcal{F}^f_{t'}$ is fed into the detection head to predict the final 3D bounding boxes $\mathcal{B}_t = \{b_i\}_{i=1}^{N_b}$.

3.2 Temporal Radar Denoising (TRD)

Multi-frame 4D Radar observations provide valuable temporal cues in adverse weather. However, historical 4D Radar responses are not always reliable due to the accumulation of temporal clutter over consecutive frames. Therefore, we propose the TRD block in Figure 2 (a). The block performs reliability-aware temporal reweighting, suppressing clutter and temporal noise accumulation.

Given multi-frame 4D Radar voxels $\mathcal{X}^R_{t-k:t}$, TRD estimates voxel-level reliability based on temporal-spatial consistency. Reliable 4D Radar responses usually exhibit persistent temporal support and consistent spatial structures across consecutive observations, whereas temporal clutter tends to be sparse and unstable. Accordingly, TRD first separates the points within each voxel according to their timestamps and extracts timestamp-aware feature representations to characterize the temporal evolution of 4D Radar responses. To quantify voxel-level 4D Radar reliability in adverse weather, TRD constructs a degradation-aware reliability representation by aggregating complementary temporal-spatial cues. Specifically, TRD performs reliability-aware temporal modeling of 4D Radar responses, where feature tokens and statistical priors jointly characterize response consistency in temporal and spatial dimensions, filtering temporal clutter and reducing cross-frame noise accumulation. For each 4D Radar voxel, the extracted reliability cues are encoded into a reliability-aware representation:

$$\boldsymbol{\mathcal{Z}}_i = \Theta(\boldsymbol{\mathcal{Z}}_i^{token}, \boldsymbol{\mathcal{Z}}_i^{int}, \boldsymbol{\mathcal{Z}}_i^{drift}, \boldsymbol{\mathcal{Z}}_i^{nbr}), \tag{1}$$

where Θ is concatenate and $\boldsymbol{\mathcal{Z}}_i^{token}$, $\boldsymbol{\mathcal{Z}}_i^{int}$, $\boldsymbol{\mathcal{Z}}_i^{drift}$, and $\boldsymbol{\mathcal{Z}}_i^{nbr}$ are feature token, intensity, motion, and neighborhood reliability cues, respectively. Accordingly, a lightweight head with sigmoid predicts voxel-wise reliability gates for adaptive filtering: $\boldsymbol{\mathcal{G}}_i = \sigma(\mathrm{Head}(\boldsymbol{\mathcal{Z}}_i))$, where $\boldsymbol{\mathcal{G}}_i \in [0, 1]$ denotes the reliability score of the i-th Radar voxel. The predicted reliability gates enable Top-K filtering to suppress temporal clutter.

$$\hat{\mathcal{X}}^R_t = \boldsymbol{\Gamma}(\mathcal{X}^R_{t-k:t}, \{\boldsymbol{\mathcal{G}}_i\}_{i=1}^V, K), \tag{2}$$

where $\boldsymbol{\Gamma}$ is a filtering operation and $\mathcal{X}^R_{t'} = \hat{\mathcal{X}}^R_t \cdot \{\boldsymbol{\mathcal{G}}_i\}_{i=1}^{KV}$ is the refined 4D Radar voxel. By reliability reweighting, TRD adaptively discounts unreliable clutter, alleviating temporal noise accumulation.

3.3 Radar-Guided Temporal Compensation (RGTC)

Individual-frame LiDAR observations often suffer from incomplete spatial structures due to severe LiDAR degradation in adverse weather, whereas historical context preserves complementary spatial cues that serve as temporal priors for mitigating individual-frame perception degradation. However, effectively exploiting historical cues remains challenging due to temporal misalignment and redundant variations in historical features. To address this issue, we propose RGTC in Figure 2 (b), which exploits 4D Radar cues as guidance for cross-modal retrieval and adaptively incorporates reliable historical LiDAR features to compensate for the degraded perception of current LiDAR observations.

Starting from the current LiDAR features $\mathcal{F}^L_t$, RGTC maintains a temporal memory bank that stores historical LiDAR features $\mathcal{F}^L_{his}$, along with occupancy masks marking stored BEV locations and validity indicators identifying available historical frames as spatial priors. The temporal cues provide complementary information to alleviate perception degradation caused by adverse weather. Specifically, multi-frame historical LiDAR features are encoded in temporal cues by concatenation and MLP, while TRD-filtered 4D Radar BEV features $\mathcal{F}^R_{t'}$ are transformed into spatial priors. The temporal and spatial cues guide temporal cross-modal aggregation with the current LiDAR feature. RGTC considers current LiDAR information and historical cues to capture the temporal correspondence between them, while 4D Radar embeddings provide reliability-aware spatial guidance for adaptive aggregation. The 4D Radar-guided interaction selectively aggregates informative historical cues and compensates for missing spatial structures in degraded LiDAR observations.

$$\Delta\mathcal{F}^L_t = \mathcal{T}(\mathcal{F}^L_t, \mathcal{F}^L_{his}, \mathcal{F}^R_{t'}), \tag{3}$$

The $\mathcal{T}(\cdot)$ denotes the Radar-guided interaction between the current LiDAR features and the aggregated historical context, which produces complementary features for residual compensation.

$$\mathcal{F}^L_{t'} = \mathcal{A}(\mathcal{F}^L_t, \Delta\mathcal{F}^L_t), \tag{4}$$

where $\mathcal{A}(\cdot)$ is a residual aggregation. This formulation of the residual compensation strategy allows RGTC to adaptively retrieve a reliable historical LiDAR context guided by 4D Radar cues, thereby compensating for degraded perception gaps in current-frame representations.

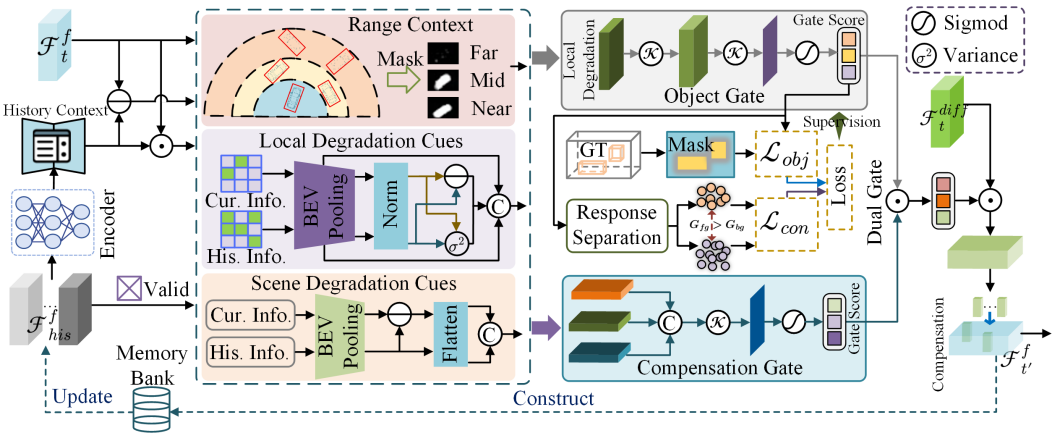


Figure 3: The proposed DGPF block constructs three degradation-aware priors to derive object-aware and compensation-aware gates. Based on degradation-aware gated fusion, multi-modal temporal cues are adaptively aggregated to alleviate degraded foreground semantics under adverse weather conditions.

3.4 Degradation-aware Gated Prior Fusion (DGPF)

Although sensor-specific compensation alleviates modality-specific degradation, the fused LiDAR-4D Radar feature still exhibits compromised foreground semantics and unreliable background activations in adverse weather. Sensing degradation is spatially non-uniform in adverse weather, resulting in location-dependent reliability of historical responses. Therefore, we propose the DGPF in Figure 3, which gates the retrieved features against the current frame to focus on important information.

Starting from fused information $\mathcal{F}_t^f$ and historical features $\mathcal{F}_{his}^f$ stored in the memory bank, DGPF first encodes $\mathcal{F}_{his}^f$ through a lightweight encoder to obtain temporal context $\mathcal{F}_{his}^{f'}$. The temporal discrepancy between current and historical representations is utilized to characterize the reliability of temporal updates. To achieve degradation-aware feature adaptation, DGPF introduces two complementary gates: an object gate G_o and a compensation gate G_c. The DGPF block explicitly considers degradation-related factors to determine where and how historical features should be incorporated. Specifically, DGPF constructs three degradation-aware priors. The range-aware context encodes distance-dependent reliability variations in adverse weather by incorporating normalized range and a spatial range mask, providing degradation-aware guidance for temporal fusion. Local degradation cues infer spatially varying reliability by exploiting current-history discrepancies and temporal uncertainty, enabling adaptive historical feature fusion in locally degraded regions. Scene degradation cues infer global degradation states by aggregating current and historical responses through BEV pooling and feature interaction, providing adaptive guidance for temporal fusion under varying degradation conditions. The $G_o = \sigma(\kappa(\mathcal{D}_{\text{local}}))$ determines spatial regions where temporal compensation is required, while the $G_c = \sigma(\kappa(\Theta(\mathcal{D}_{\text{range}}, \mathcal{D}_{\text{local}}, \mathcal{D}_{\text{scene}})))$ controls the contribution of historical features according to the degradation conditions, where $\mathcal{D}$ and κ are prior cues and 3x3 convolution, respectively. The gates enable reliability-aware temporal aggregation by balancing current and historical observations:

$$\mathcal{F}_{t'}^f = \mathcal{A}(\mathcal{F}_t^f, \alpha\mathcal{M}(\mathcal{F}_{his}^{f'} - \mathcal{F}_t^f)), \tag{5}$$

where $\mathcal{M}(\cdot) = G_o \odot G_c \odot (\cdot)$ is the modulation operator, $\mathcal{A}(\cdot)$ indicates residual aggregation, and α is learnable. The dual-gate design enables DGPF to compensate for missing temporal information, alleviating weakened foreground semantics and unstable background activations in adverse weather.

3.5 Loss Function and Training Strategy

Following L4DR (Huang et al., 2025b), we adopt the standard detection losses $\mathcal{L}_{cls}$ and $\mathcal{L}_{loc}$ with weights β_{cls} and β_{loc}. Additionally, we employ a BCE loss $\mathcal{L}_{obj}$ and a contrastive loss $\mathcal{L}_{con}$ to supervise the object gate G_o in DGPF, weighted by λ_{obj} and λ_{con}, respectively. The detailed loss formulations and settings are provided in Appendix A.2. The overall objective is formulated as:

$$\mathcal{L}_{all} = \beta_{cls}\mathcal{L}_{cls} + \beta_{loc}\mathcal{L}_{loc} + \lambda_{obj}\mathcal{L}_{obj} + \lambda_{con}\mathcal{L}_{con}, \tag{6}$$

Table 1: Quantitative results of different 3D object detection methods on K-Radar dataset V1.0 at IoU=0.5. We present the modality of each method (L: LiDAR, 4DR: 4D radar) and detailed performance for each weather condition. MF denotes multi-frame methods. Best in **bold**, second in underline.

Frame	Methods	Modality	Metric	Total	Normal	Overcast	Fog	Rain	Sleet	Lightsnow	Heavysnow
Single-frame	RTNH (NeurIPS 2022)	4DR	AP_{BEV}	36.0	35.8	41.9	44.8	30.2	34.5	63.9	55.1
			AP_{3D}	14.1	19.7	20.5	15.9	13.0	13.5	21.0	6.36
	PointPillars (CVPR 2019)	L	AP_{BEV}	49.1	48.2	53.0	45.4	44.2	45.9	74.5	53.8
			AP_{3D}	22.4	21.8	28.0	28.2	27.2	22.6	23.2	12.9
	RTNH (NeurIPS 2022)	L	AP_{BEV}	66.3	65.4	87.4	83.8	73.7	48.8	78.5	48.1
			AP_{3D}	37.8	39.8	46.3	59.8	28.2	31.4	50.7	24.6
	InterFusion (IROS 2022)	L+4DR	AP_{BEV}	52.9	50.0	59.0	80.3	50.0	22.7	72.2	53.3
			AP_{3D}	17.5	15.3	20.5	47.6	12.9	9.33	56.8	25.7
	LOD-PDR (ECCV 2024)	L+4DR	AP_{BEV}	71.3	72.1	87.3	83.9	75.2	45.1	79.1	48.1
			AP_{3D}	40.4	40.3	43.3	63.4	31.6	27.9	51.6	24.9
	3D-LRF (CVPR 2024)	L+4DR	AP_{BEV}	73.6	72.3	88.4	86.6	76.6	47.5	79.6	64.1
			AP_{3D}	45.2	45.3	55.8	51.8	38.3	23.4	60.2	36.9
	L4DR (AAAI 2025)	L+4DR	AP_{BEV}	77.5	76.8	88.6	89.7	78.2	59.3	80.9	53.8
			AP_{3D}	53.5	53.0	64.1	73.2	53.8	46.2	52.4	37.0
	DLRFusion (ICCV 2025)	L+4DR	AP_{BEV}	73.2	72.3	86.8	86.7	81.6	43.3	84.0	54.5
			AP_{3D}	45.7	45.0	55.6	45.4	46.1	26.3	61.0	30.6
	REL (AAAI 2026)	L+4DR	AP_{BEV}	75.6	74.0	88.4	90.1	**84.5**	52.6	84.3	59.8
			AP_{3D}	54.5	53.0	**71.5**	70.3	**55.4**	36.6	**67.4**	45.7
Multi-frame	InterFusion-MF	L+4DR	AP_{BEV}	52.7	49.6	56.5	76.6	60.2	43.0	63.2	40.5
			AP_{3D}	16.9	10.3	16.0	22.9	21.1	12.6	18.2	23.1
	L4DR-MF	L+4DR	AP_{BEV}	77.2	76.6	89.3	89.6	78.1	56.7	88.6	58.3
			AP_{3D}	53.3	56.8	63.3	72.2	54.2	48.9	64.5	30.4
	TMF (Ours)	L+4DR	AP_{BEV}	**79.9**	**79.7**	**92.9**	**92.6**	80.1	**68.6**	**88.7**	**65.3**
			AP_{3D}	**57.8**	**60.0**	71.1	**76.0**	53.8	**51.2**	67.2	**46.8**

4 Experiments

4.1 Dataset and Evaluation Metrics

K-Radar dataset: K-Radar (Paek et al., 2022) is a large-scale autonomous driving perception dataset developed for robust 3D perception in challenging environments. In this work, we focus on LiDAR-4D Radar based 3D object detection and utilize only the 64-line LiDAR and 4D Radar data for training and evaluation. Following the official data split, the dataset contains 17458 frames for training and 17536 frames for testing. Since TMF relies on multi-frame observations, we divide K-Radar sequences into non-overlapping temporal clips of h consecutive frames. During training, clip-wise temporal order is preserved, while different clips are randomly shuffled to improve data diversity. The memory bank is initialized at the beginning of each clip and reset after processing. During inference, the memory is updated following the original temporal order to exploit temporal complementary information. We evaluate our approach under two officially released evaluation protocols, i.e., Version 1.0 (released in 2022) and Version 2.0 (updated in 2023). For both evaluation protocols, we follow the official settings and report the standard AP_{BEV} and AP_{3D} metrics under IoU thresholds of 0.5 and 0.3. (Results under IoU=0.3 are provided in the Appendix A.3 and A.5.).

4.2 Implementation Details

We implement TMF based on L4DR (Huang et al., 2025b). To ensure fair comparisons, we maintain the same detection framework and replace the original fusion strategy with our proposed temporal multimodal framework. All experiments are conducted with a batch size of 8 on 2 RTX 3090 GPUs, while other settings are provided in the Appendix A.2.

4.3 Results on K-Radar Dataset

We compare our proposed TMF with several existing 3D object detection methods, including RTNH (Paek et al., 2022), PointPillars (Lang et al., 2019), InterFusion (Wang et al., 2022a), LOD-PDR (Chae et al., 2024a), 3D-LRF (Chae et al., 2024b), L4DR (Huang et al., 2025b), DLRFusion (Chae et al., 2025), and REL (Tong et al., 2026). Table 1 presents the quantitative comparisons in different weather conditions on the K-Radar with the V1.0 and an IoU threshold of 0.5. Compared with existing methods, TMF achieves consistent improvements on both AP_{BEV} and AP_{3D}, showing

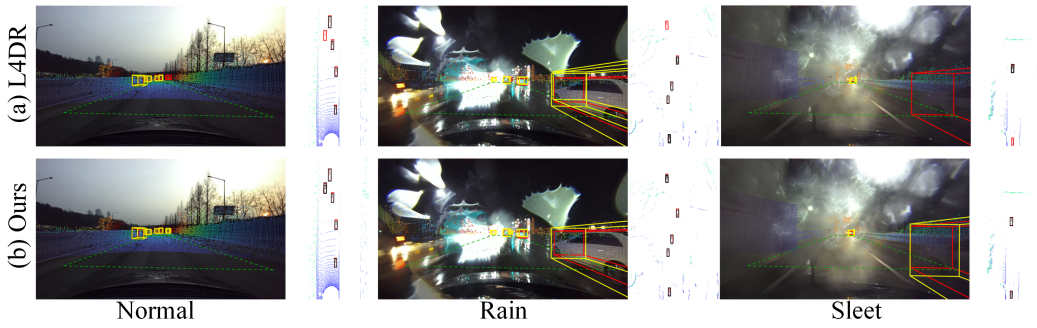


Figure 4: Visualization results of 3D object detection in range view and bird's-eye view. The results in the range view show the image and projected LiDAR with red GT boxes and yellow predicted boxes. In the bird's-eye view, LiDAR points are shown with red GT boxes and black predicted boxes. Each row represents a different 3D object detection method: (a) L4DR and (b) Ours. Each column corresponds to a different weather condition. Detailed IoU results for individual object categories and additional visualizations of 4D Radar heatmap projections on LiDAR views are provided in the Appendix A.4.

Table 2: Quantitative results of different 3D object detection methods on the K-Radar V2.0 at IoU=0.5. SF and MF are single-frame and multi-frame methods, respectively. Best in **bold**, second in underline.

Class	Method	Modality	Frame	Metric	Total	Normal	Overcast	Fog	Rain	Sleet	Li.Snow	He.Snow
Sedan	L4DR (AAAI2025)	L+4DR	SF	AP_{BEV}	74.3	72.7	82.5	90.3	**77.4**	57.1	87.2	**57.9**
				AP_{3D}	51.3	49.4	51.6	77.8	53.5	**42.0**	**61.4**	**43.7**
	L4DR-MF	L+4DR	MF	AP_{BEV}	**75.1**	**74.1**	84.8	90.4	77.2	55.4	86.1	51.6
				AP_{3D}	52.6	51.1	53.9	78.8	52.4	41.2	54.5	40.1
	Ours	L+4DR	MF	AP_{BEV}	74.5	73.6	**85.4**	**93.7**	74.0	**62.4**	**87.5**	**57.9**
				AP_{3D}	**54.1**	**52.2**	**57.3**	**82.9**	**53.8**	38.0	59.4	41.7
Bus or Truck	L4DR (AAAI2025)	L+4DR	SF	AP_{BEV}	**60.6**	51.1	**80.9**	-	4.2	64.7	80.4	**79.5**
				AP_{3D}	40.4	35.9	**63.6**	-	3.6	55.1	65.5	**39.7**
	L4DR-MF	L+4DR	MF	AP_{BEV}	59.8	**56.0**	66.9	-	3.2	**71.3**	81.1	61.4
				AP_{3D}	39.9	**38.3**	58.5	-	1.2	**61.7**	59.6	26.7
	Ours	L+4DR	MF	AP_{BEV}	59.0	52.2	70.0	-	**7.2**	53.9	**88.3**	68.0
				AP_{3D}	**42.0**	37.5	53.1	-	**5.2**	40.2	**74.7**	39.5

the effectiveness of exploiting temporal complementary information to alleviate adverse weather degradation. Specifically, TMF achieves an overall AP_{3D} of 57.8%, obtaining the best performance among all the compared methods. This indicates that historical observations can provide valuable complementary information for enhancing degraded spatial representations under adverse weather conditions. To evaluate the effectiveness of temporal fusion, we compare TMF with InterFusion-MF and L4DR-MF, which directly concatenate multi-frame 4D Radar and LiDAR point clouds with temporal information. However, naive temporal concatenation fails to deliver consistent improvements, with overall performance slightly degrading due to the accumulation of unreliable temporal cues. By better leveraging temporal information, TMF achieves further improvements over these multi-frame baselines. For example, compared to L4DR-MF, TMF improves the overall AP_{3D} from 53.3% to 57.8%. In particular, significant gains are observed under challenging weather conditions, improving from 30.4% to 46.8% on HeavySnow and from 48.9% to 51.2% on Sleet. Furthermore, we provide qualitative comparisons of 3D detection results in both camera and LiDAR views (Figure 4). We also further evaluate TMF on the K-Radar V2.0 and compare it with L4DR and its multi-frame extension in Table 2. The results show that TMF achieves consistent improvements in both the Sedan and Bus or Truck categories in Total AP_{3D}. Moreover, for the Sedan, TMF achieves superior performance across most adverse weather conditions.

4.4 Ablation Study

Temporal Observation Analysis: Table 3 investigates the impact of temporal context length on the proposed TMF in the K-Radar V1.0 benchmark by varying the total number of input frames from 3 to 11. Increasing the temporal context from 3 to 7 frames improves the 3D object detection

performance, achieving the best detection results with 7 input frames ($AP_{\text{3D}} = 57.8\%$ and $AP_{\text{BEV}} = 79.9\%$ at IoU=0.5). This shows that the appropriate temporal context provides complementary information to compensate for degraded spatial representations. However, further increasing the input frames introduces redundant or unreliable temporal responses, resulting in performance degradation. Therefore, TMF adopts 7 input frames as a balance between temporal complementarity and reliability.

Component Contribution Analysis: We further analyze the contribution of each proposed component based on reproduced L4DR. (The implementation details and modifications of the baseline are provided in the Appendix A.5.) Table 4 shows that progressively introducing the proposed blocks improves the robustness of the proposed TMF under diverse weather conditions. Introducing TRD and RGTC enhances the performance of the proposed TMF by discounting unreliable temporal clutter and compensating for degraded spatial observations, achieving notable gains under challenging conditions, especially HeavySnow (40.33%→48.71%). Finally, introducing DGPF further improves overall performance to 57.77%, with consistent gains in multiple weather conditions, especially in Sleet and LightSnow. These results demonstrate that each component contributes to effective temporal fusion, enabling TMF to exploit reliable temporal information under adverse weather conditions.

Computational Cost Analysis: We further analyze the computational overhead introduced by each proposed component in Table 4 (A detailed computational cost analysis is provided in the Appendix A.5.). Table 4 shows that TRD introduces negligible additional cost, increasing only 0.02M parameters while effectively improving the reliability of the 4D radar. RGTC incurs moderate overhead due to historical LiDAR feature aggregation built upon the TRD, introducing an additional 0.30M parameters and 11.37ms inference latency. With all components enabled, TMF requires 93.80ms inference time and introduces 3.85M parameters compared with the baseline. Such computational overhead is acceptable considering the substantial robustness improvement achieved in adverse weather.

Table 3: Ablation study of different temporal observation lengths on the K-Radar V1.0 dataset at IoU=0.5. Best results are highlighted in **bold**, and second-best results are underlined.

Frame	Metric	Total	Normal	Overcast	Fog	Rain	Sleet	Lightsnow	Heavysnow
3	AP_{BEV}	79.6	79.1	92.8	92.3	79.7	67.0	**89.3**	66.1
	AP_{3D}	57.2	56.1	69.3	80.4	**55.7**	47.8	60.3	48.7
5	AP_{BEV}	79.8	79.4	92.3	89.9	**80.6**	67.5	88.2	**66.3**
	AP_{3D}	57.0	56.9	65.4	77.5	55.3	45.4	67.1	**49.8**
7	AP_{BEV}	**79.9**	**79.7**	**92.9**	**92.6**	80.1	68.6	88.7	65.3
	AP_{3D}	**57.8**	**60.0**	**71.1**	76.0	53.8	**51.2**	**67.2**	46.8
9	AP_{BEV}	79.4	79.3	92.8	**92.6**	79.0	67.4	**89.3**	64.8
	AP_{3D}	57.1	57.6	66.6	77.3	55.6	48.2	63.2	39.7
11	AP_{BEV}	79.8	79.2	92.6	92.5	79.4	**70.0**	88.6	65.0
	AP_{3D}	56.8	56.4	66.3	**80.5**	55.4	46.4	59.4	44.9

Table 4: Ablation study of different components and computational cost analysis based on our reproduced L4DR baseline on the K-Radar dataset V1.0 using AP_{3D} at IoU=0.5. Best in **bold**, second in underline.

TRD	RGTC	DGPF	Total	Normal	Overcast	Fog	Rain	Sleet	L.snow	H.snow	Speed	Params
-	-	-	56.64	54.49	68.42	77.33	54.01	50.58	62.82	40.33	66.75ms	52.72M
✓	-	-	57.47	57.08	68.67	76.87	**56.94**	50.19	65.08	42.59	72.52ms (+5.77ms)	52.74M (+0.02M)
✓	✓	-	57.29	57.43	70.05	76.39	53.73	46.13	61.83	**48.71**	78.12ms (+11.37ms)	53.02M (+0.30M)
✓	-	✓	57.49	58.51	65.39	**78.23**	56.65	45.47	64.24	46.12	88.36ms (+21.61ms)	56.28M (+3.56M)
✓	✓	✓	**57.77**	**60.04**	**71.05**	75.99	53.75	**51.22**	**67.24**	46.84	93.80ms (+27.05ms)	56.57M (+3.85M)

5 Conclusion

In this paper, we investigate LiDAR-4D Radar fusion under adverse weather conditions and reveal that sensor degradation causes incomplete individual observations. We find that historical observations provide complementary information to compensate for degraded current representations and propose TMF, the first temporal multimodal learning framework for robust 3D object detection using LiDAR-4D Radar under adverse weather conditions. Extensive experiments on the K-Radar dataset demonstrate that the proposed TMF achieves consistent improvements under diverse weather conditions, validating the effectiveness of temporal complementary cues for robust 3D object detection.

AI Use Statement

In this work, we used generative AI tools for language refinement and improved manuscript readability. We have not used generative AI tools to generate experimental results, design evaluation protocols, or make scientific claims. In addition, we used generative AI tools to assist with figure design and LaTeX formatting. We reviewed all AI-assisted work and verified its accuracy. We take responsibility for the final content of this work, including all text, claims, and artifacts produced with the aid of generative AI.

Reproducibility Statement

For reproducibility, Section 3 describes the TMF architecture, its constituent blocks, and the training objective. Section 4.1 and Appendix A.2 specify dataset splits, data preprocessing, temporal clip construction, and evaluation protocols. Training hyperparameters and hardware configurations are provided in Section 4.2 and Appendix A.2, with additional details on gate supervision in Appendix A.2. Appendix A.5 documents the modifications to the L4DR baseline and presents further ablation studies.

References

Yujeong Chae, Hyeonseong Kim, Changgyoon Oh, Minseok Kim, and Kuk-Jin Yoon. Lidar-based all-weather 3d object detection via prompting and distilling 4d radar. In *European Conference on Computer Vision*, pp. 368–385. Springer, 2024a.

Yujeong Chae, Hyeonseong Kim, and Kuk-Jin Yoon. Towards robust 3d object detection with lidar and 4d radar fusion in various weather conditions. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15162–15172. IEEE, 2024b.

Yujeong Chae, Heejun Park, Hyeonseong Kim, and Kuk-Jin Yoon. Doppler-aware lidar-radar fusion for weather-robust 3d detection. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 27197–27208. IEEE, 2025.

Jiajun Deng, Shaoshuai Shi, Peiwei Li, Wengang Zhou, Yanyong Zhang, and Houqiang Li. Voxel r-cnn: Towards high performance voxel-based 3d object detection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 1201–1209, 2021.

Zijian Gu, Jianwei Ma, Yan Huang, Honghao Wei, Zhanye Chen, Hui Zhang, and Wei Hong. Hgsfusion: Radar-camera fusion with hybrid generation and synchronization for 3d object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 3185–3193, 2025.

Martin Hahner, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Fog simulation on real lidar point clouds for 3d object detection in adverse weather. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 15263–15272. IEEE, 2021.

He Huang, Xinyuan Yan, Junxing Yang, Yuming Cao, and Xin Zhang. Lidsor: A filter for removing rain and snow noise points from lidar point clouds in rainy and snowy weather. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48:733–740, 2023.

Xun Huang, Hai Wu, Xin Li, Xiaoliang Fan, Chenglu Wen, and Cheng Wang. Sunshine to rainstorm: Cross-weather knowledge distillation for robust 3d object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 2409–2416, 2024.

Xun Huang, Jinlong Wang, Qiming Xia, Siheng Chen, Bisheng Yang, Xin Li, Cheng Wang, and Chenglu Wen. V2x-r: Cooperative lidar-4d radar fusion with denoising diffusion for 3d object detection. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 27390–27400. IEEE, 2025a.

Xun Huang, Ziyu Xu, Hai Wu, Jinlong Wang, Qiming Xia, Yan Xia, Jonathan Li, Kyle Gao, Chenglu Wen, and Cheng Wang. L4dr: Lidar-4dradar fusion for weather-robust 3d object detection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 39, pp. 3806–3814, 2025b.

Seung-Hyun Kong, Dong-Hee Paek, and Sangjae Cho. Rtnh+: Enhanced 4d radar object detection network using combined cfar-based two-level preprocessing and vertical encoding. *arXiv preprint arXiv:2310.17659*, 2023.

Alex H Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. Pointpillars: Fast encoders for object detection from point clouds. In *2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 12689–12697. IEEE, 2019.

Peizhao Li, Pu Wang, Karl Berntorp, and Hongfu Liu. Exploiting temporal relations on radar perception for autonomous driving. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17050–17059. IEEE, 2022.

Yu-Jhe Li, Matthew O'Toole, and Kris Kitani. St-mvdnet++: Improve vehicle detection with lidar-radar geometrical augmentation via self-training. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5. IEEE, 2023.

Chen Liu, Florian Kelber, Bernhard Vogginger, and Christian Mayr. Ca-cfar is convolution: Fast target detection with machine learning accelerator. In *2024 13th Mediterranean Conference on Embedded Computing (MECO)*, pp. 1–6. IEEE, 2024a.

Yifan Liu, Yong Zhang, Rukai Lan, Cheng Cheng, and Zhaolong Wu. Awardistill: Adaptive and robust 3d object detection in adverse conditions through knowledge distillation. *Expert Systems with Applications*, 266:126032, 2025.

Yuhang Liu, Yonglin Tian, Boyi Sun, Yiyao Wang, and Fei-Yue Wang. Parallel lidars meet the foggy weather. *IEEE Journal of Radio Frequency Identification*, 6:867–870, 2022.

Zhe Liu, Jinghua Hou, Xinyu Wang, Xiaoqing Ye, Jingdong Wang, Hengshuang Zhao, and Xiang Bai. Lion: Linear group rnn for 3d object detection in point clouds. *Advances in Neural Information Processing Systems*, 37:13601–13626, 2024b.

Andrea Matteazzi, Pascal Colling, Michael Arnold, and Dietmar Tutsch. Adverse weather conditions augmentation of lidar scenes with latent diffusion models. *arXiv preprint arXiv:2501.01761*, 2025.

Michael Meyer, Georg Kuschk, and Sven Tomforde. Graph convolutional networks for 3d object detection on radar data. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pp. 3053–3062. IEEE, 2021.

Dong-Hee Paek, Seung-Hyun Kong, and Kevin Tirta Wijaya. K-radar: 4d radar object detection for autonomous driving in various weather conditions. *Advances in Neural Information Processing Systems*, 35:3819–3829, 2022.

Dong-Hee Paek, Seung-Hyun Kong, and Kevin Tirta Wijaya. Enhanced k-radar: Optimal density reduction to improve detection performance and accessibility of 4d radar tensor-based object detection. In *2023 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1–6. IEEE, 2023.

Yuanfan Qi, Chun Liu, Hangbin Wu, Ruijie Chen, Chenglu Wen, Xun Huang, Shoujun Jia, and Keke Zhang. Fusionbev: Lidar and 4d radar fusion for 3d object detection. *Information Fusion*, pp. 104240, 2026.

Kun Qian, Shilin Zhu, Xinyu Zhang, and Li Erran Li. Robust multimodal vehicle detection in foggy weather using complementary lidar and radar signals. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 444–453. IEEE, 2021.

Tianxu Tong, Xinrun Liu, Hongmin Liu, and Bin Fan. Towards accurate 3d object detection in adverse weather by leveraging 4d radar for lidar geometry enhancement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 9548–9556, 2026.

Li Wang, Xinyu Zhang, Baowei Xv, Jinzhao Zhang, Rong Fu, Xiaoyu Wang, Lei Zhu, Haibing Ren, Pingping Lu, Jun Li, et al. Interfusion: Interaction-based 4d radar and lidar fusion for 3d object detection. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 12247–12253. IEEE, 2022a.

Weiqi Wang, Xiong You, Lingyu Chen, Jiangpeng Tian, Fen Tang, and Lantian Zhang. A scalable and accurate de-snowing algorithm for lidar point clouds in winter. *Remote Sensing*, 14(6):1468, 2022b.

Yingjie Wang, Jiajun Deng, Yao Li, Jinshui Hu, Cong Liu, Yu Zhang, Jianmin Ji, Wanli Ouyang, and Yanyong Zhang. Bi-lrfusion: Bi-directional lidar-radar fusion for 3d dynamic object detection. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13394–13403. IEEE, 2023.

Maciej K Wozniak, Mattias Hansson, Marko Thiel, and Patric Jensfelt. Uada3d: Unsupervised adversarial domain adaptation for 3d object detection with sparse lidar and large domain gaps. *IEEE Robotics and automation letters*, 9(12):11210–11217, 2024.

Baowei Xu, Xinyu Zhang, Li Wang, Xiaomei Hu, Zhiwei Li, Shuyue Pan, Jun Li, and Yongqiang Deng. Rpfa-net: A 4d radar pillar feature attention network for 3d object detection. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pp. 3061–3066. IEEE, 2021.

Yan Yan, Yuxing Mao, and Bo Li. Second: Sparsely embedded convolutional detection. *Sensors*, 18(10):3337, 2018.

Zetong Yang, Yanan Sun, Shu Liu, and Jiaya Jia. 3dssd: Point-based 3d single stage object detector. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11037–11045. IEEE, 2020.

Guowen Zhang, Lue Fan, Chenhang He, Zhen Lei, Zhaoxiang Zhang, and Lei Zhang. Voxel mamba: Group-free state space models for point cloud based 3d object detection. *Advances in Neural Information Processing Systems*, 37:81489–81509, 2024a.

Miao Zhang, Sherif Abdulatif, Benedikt Loesch, Marco Altmann, Marius Schwarz, and Bin Yang. Exploring domain shift on radar-based 3d object detection amidst diverse environmental conditions. In *2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 3513–3518. IEEE, 2024b.

Yuxiao Zhang, Ming Ding, Hanting Yang, Yingjie Niu, Maoning Ge, Kento Ohtani, Chi Zhang, and Kazuya Takeda. Lidar point cloud augmentation for adverse conditions using conditional generative model. *Remote Sensing*, 16(12):2247, 2024c.

A Appendix

A.1 Overview

This supplementary material provides additional implementation details, extended quantitative and qualitative evaluations, and further ablation studies to complement the main paper. Specifically, Section A.2 reports the implementation details, while Section A.3 extends the quantitative evaluation with additional results. Section A.4 provides further qualitative visualizations of 3D object detection, and Section A.5 investigates additional ablation studies to further validate the effectiveness of the proposed framework.

A.2 Implementation Details

For training, we optimize the entire network for 35 epochs using the Adam optimizer with a learning rate of 1×10^{-3}, where $\beta_1 = 0.9$ and $\beta_2 = 0.999$. We adopt the standard detection losses $\mathcal{L}cls$ and $\mathcal{L}loc$ with weights $\beta_{cls} = 1$ and $\beta_{loc} = 2$. The object gate G_o in the DGPF is supervised by a BCE loss $\mathcal{L}obj$ and a foreground-background contrastive loss $\mathcal{L}con$, weighted by $\lambda_{obj} = 0.05$ and $\lambda_{con} = 0.02$, respectively. The denoising ratio K in the TRD block is set to 0.7. All experiments are conducted with a batch size of 8 on 2 RTX 3090 GPUs and the voxel size is set to $(0.4\text{m}, 0.4\text{m}, 0.4\text{m})$.

For the object-aware gate regularization, foreground and background regions are constructed based on the ground-truth 3D bounding boxes. Specifically, BEV locations inside the projected ground-truth boxes are regarded as foreground regions, while the remaining locations are treated as background

regions. A binary target map is generated by assigning labels of 1 and 0 to foreground and background regions, respectively. To alleviate the severe imbalance between foreground and background regions, a foreground-balanced BCE loss is adopted:

$$\mathcal{L}_{obj} = -\frac{1}{\sum_i w_i} \sum_i w_i \left[y_i \log(g_i) + (1 - y_i) \log(1 - g_i) \right], \tag{7}$$

where w_i denotes the foreground-balanced weight assigned to the i-th BEV location, g_i is predicted object gate and y_i represents corresponding foreground/background target. To further enlarge the response gaps between foreground and background regions, we introduce a gate separation constraint:

$$\mathcal{L}_{con} = \max(0, m - (\bar{g}_{fg} - \bar{g}_{bg})), \tag{8}$$

where the separation margin m is set to 0.05. $\bar{g}_{fg}$ and $\bar{g}_{bg}$ are the average gate responses of foreground and background regions, respectively.

For the K-Radar dataset, multi-frame 4D Radar observations are concatenated with an additional temporal coordinate appended to each Radar point, resulting in point features of $(x, y, z, \text{intensity}, t)$. We randomly sample 10,240 points from the sparse 4D Radar tensor, while retaining all LiDAR points as input. Following the official data split, the K-Radar dataset contains 17,458 frames for training and 17,536 frames for testing. After filtering frames with valid object annotations, we retain 11,021 training frames and 10,110 testing frames. Since TMF relies on continuous temporal observations, we divide the K-Radar sequences into non-overlapping temporal clips. We first split each sequence at frames without valid objects to avoid mixing unrelated observations, and then partition the resulting segments into clips with up to $h = 12$ consecutive frames. After temporal partitioning, the training set contains 977 temporal clips. During training, different temporal clips are randomly shuffled to improve data diversity and generalization, while the temporal order within each clip is strictly preserved to maintain effective temporal dependency modeling. We evaluate our approach under two officially released evaluation protocols, i.e., Version 1.0 (released in 2022) and Version 2.0 (updated in 2023). For Version 1.0, the ROI is defined as $[0, 72]\text{m}$ along the X-axis, $[-6.4, 6.4]\text{m}$ along the Y-axis, and $[-2, 6]\text{m}$ along the Z-axis, with the detection target limited to the "Sedan" category. Version 2.0 further extends the evaluation range by expanding the ROI to $[0, 72]\text{m}$, $[-16, 16]\text{m}$, and $[-2, 7.6]\text{m}$ along the X-, Y-, and Z-axis, respectively, and introduces an additional 'Bus or Truck' category. Other settings follow the default OpenPCDet configurations.

A.3 Extended Quantitative Evaluation

Additional quantitative results under an IoU threshold of 0.3 are provided on the K-Radar V1.0 and V2.0 benchmarks in Tables 5 and 6. These results further validate the robustness of TMF under a more relaxed localization criterion. In the K-Radar V1.0 benchmark, TMF achieves the best overall performance among the LiDAR-4D Radar fusion methods, obtaining 79.4% AP_{3D}. Compared with existing single-frame fusion approaches, TMF consistently improves detection performance under challenging weather conditions. In particular, TMF achieves notable advantages in the Sleet and Lightsnow scenarios, reaching 67.0% and 88.5% AP_{3D}, respectively, demonstrating the effectiveness of temporal information in compensating for degraded perception gaps. Moreover, TMF maintains strong performance in Overcast and Fog weather conditions, where temporal complementary cues help alleviate incomplete or unreliable sensor measurements. On the K-Radar V2.0 benchmark, TMF further demonstrates its generalization capability across different object categories. For the Sedan category, TMF achieves competitive performance with significant improvements under challenging weather conditions, including the best results in Overcast (85.0%), Fog (90.1%), Sleet (66.8%), and competitive performance in Li.Snow (87.9%). For the Bus or Truck category, TMF achieves the best HeavySnow performance with 56.6% AP_{3D} and maintains competitive results in other weather conditions. These additional evaluations confirm that TMF effectively exploits temporal complementary information for robust 3D object detection under adverse weather conditions.

A.4 Qualitative Visualization and Analysis

Additional qualitative comparisons in diverse weather conditions are presented in Figure 5. Beyond the visualizations provided in the main paper, these results incorporate detailed detection IoUs and BEV projections of 4D Radar heatmaps, enabling a more comprehensive examination of the temporal multimodal fusion process in TMF. The qualitative results reveal that TMF produces more precise

Table 5: Quantitative results of different 3D object detection methods on K-Radar dataset V1.0 at IoU=0.3. We present the modality of each method (L: LiDAR, 4DR: 4D radar) and detailed performance for each weather condition. MF denotes multi-frame methods. Best in **bold**, second in underline.

Frame	Methods	Modality	Metric	Total	Normal	Overcast	Fog	Rain	Sleet	Lightsnow	Heavysnow
Single-frame	RTNH (NeurIPS 2022)	4DR	AP_{BEV}	41.1	41.0	44.6	45.4	32.9	50.6	81.5	56.3
			AP_{3D}	37.4	37.6	42.0	41.2	29.2	49.1	63.9	43.1
	PointPillars (CVPR 2019)	L	AP_{BEV}	51.9	51.6	53.5	45.4	44.7	54.3	81.2	55.2
			AP_{3D}	47.3	46.7	51.9	44.8	42.4	45.5	59.2	55.2
	RTNH (NeurIPS 2022)	L	AP_{BEV}	76.5	76.5	88.2	86.3	77.3	55.3	81.1	59.5
			AP_{3D}	72.7	73.1	76.5	84.8	64.5	53.4	80.3	52.9
	InterFusion (IROS 2022)	L+4DR	AP_{BEV}	57.5	57.2	60.8	81.2	52.8	27.5	72.6	57.2
			AP_{3D}	53.0	51.1	58.1	80.9	40.4	23.0	71.0	55.2
	LOD-PDR (ECCV 2024)	L+4DR	AP_{BEV}	82.1	83.9	88.3	87.7	84.4	52.4	89.1	59.6
			AP_{3D}	73.2	74.4	76.9	85.5	70.3	45.9	87.8	54.4
	3D-LRF (CVPR 2024)	L+4DR	AP_{BEV}	**84.0**	83.7	89.2	**95.4**	78.3	60.7	88.9	**74.9**
			AP_{3D}	74.8	81.2	87.2	86.1	73.8	49.5	87.9	**67.2**
	L4DR (AAAI 2025)	L+4DR	AP_{BEV}	79.5	**86.0**	89.6	89.9	81.1	62.3	89.1	61.3
			AP_{3D}	78.0	77.7	80.0	88.6	**79.2**	60.1	78.9	51.9
	DLRFusion (ICCV 2025)	L+4DR	AP_{BEV}	82.9	83.7	93.5	88.7	**84.6**	54.6	88.1	62.7
			AP_{3D}	74.8	79.8	84.4	84.5	75.1	44.2	86.8	61.1
Multi-frame	InterFusion-MF	L+4DR	AP_{BEV}	76.7	75.9	88.3	89.8	77.4	61.7	87.4	58.6
			AP_{3D}	52.0	50.3	64.0	63.1	52.9	44.0	59.3	41.3
	L4DR-MF	L+4DR	AP_{BEV}	79.4	84.9	90.1	89.8	80.9	62.3	89.2	59.9
			AP_{3D}	76.9	**83.2**	87.4	88.2	77.9	59.5	79.5	42.7
	TMF (Ours)	L+4DR	AP_{BEV}	83.9	84.4	**93.5**	92.8	83.6	**71.5**	**92.0**	71.8
			AP_{3D}	**79.4**	82.1	**89.0**	**89.1**	77.2	**67.0**	**88.5**	55.6

Table 6: Quantitative results of different 3D object detection methods on the K-Radar V2.0 using AP_{3D} at IoU = 0.3. SF and MF are single-frame and multi-frame methods, respectively. Best in **bold**, second in underline.

Class	Method	Modality	Frame	Total	Normal	Overcast	Fog	Rain	Sleet	Li.Snow	He.Snow
Sedan	PointPillars (CVPR2019)	4DR	SF	42.8	35.0	53.9	77.3	37.4	37.5	53.6	48.3
	RTNH (NeurIPS2022)	4DR	SF	48.2	35.5	58.8	79.3	40.3	48.1	65.6	52.6
	PointPillars (CVPR2019)	L	SF	69.7	68.1	79.0	89.2	77.7	59.1	79.0	51.5
	InterFusion (IROS2022)	L+4DR	SF	69.9	69.0	77.9	89.5	77.1	58.9	79.1	51.7
	L4DR (AAAI2025)	L+4DR	SF	**75.8**	74.6	79.2	89.3	**77.8**	61.4	87.5	**58.4**
	L4DR-MF	L+4DR	MF	75.3	**74.9**	83.4	88.7	77.5	57.9	81.8	51.4
	TMF(Ours)	L+4DR	MF	74.4	74.8	**85.0**	**90.1**	72.7	**66.8**	**87.9**	49.3
Bus or Truck	PointPillars (CVPR2019)	4DR	SF	29.4	25.8	21.5	-	0	18	64.1	34.9
	RTNH (NeurIPS2022)	4DR	SF	34.4	25.3	31.1	-	0	28.5	78.2	46.3
	PointPillars (CVPR2019)	L	SF	53.8	52.9	77.3	-	3.7	61.8	84.1	50.7
	InterFusion (IROS2022)	L+4DR	SF	56.9	56.2	80.5	-	6.4	**70.6**	**85.7**	40.5
	L4DR (AAAI2025)	L+4DR	SF	**59.7**	59.4	**86.4**	-	**8.1**	66.1	84.4	51.9
	L4DR-MF	L+4DR	MF	56.9	**61.5**	72.3	-	4.8	64.9	77.6	36.2
	TMF(Ours)	L+4DR	MF	59.1	58.2	75.9	-	7.6	54.7	84.8	**56.6**

object localization and maintains more consistent predictions across various weather conditions. These observations further demonstrate that leveraging temporally complementary multimodal cues allows TMF to compensate for degraded object representations and improve perception robustness under challenging environmental conditions.

A.5 Comprehensive Ablation Analysis

This section presents five additional analyses, including Baseline Adaptation for Temporal Fusion, Computational Cost Analysis, Quantitative Evaluation of Temporal Observation, Top-K Ratio Analysis of TRD, and Temporal LiDAR Observation Gain.

Baseline Adaptation for Temporal Fusion: To isolate the contribution of temporal fusion, we adapt L4DR as the baseline framework for TMF. Specifically, we introduce two modifications: (1) removing the original FAD block, which performs single-frame Radar denoising, to decouple temporal consistency learning from additional Radar denoising operations, and (2) replacing the original dense coordinate matching operation in the MME block with an efficient coordinate matching strategy, which reduces computational overhead while maintaining identical voxel correspondence.

First, we modify the original L4DR framework by removing the FAD block. The original FAD in L4DR is designed for single-frame 4D Radar denoising, where foreground-aware segmentation is



Figure 5: Visualization results of 3D object detection in range view and bird's-eye view. The range-view results present images with projected LiDAR points, where red and yellow boxes indicate GT and predicted boxes, respectively, along with the corresponding IoU. The bird's-eye-view results show BEV LiDAR and 4D Radar heatmap projections with red GT boxes and black predicted boxes, together with the IoU. Each column represents a different 3D object detection method: (a) InterFusion, (b) L4DR, and (c) Ours. Each row corresponds to a different weather condition.

Table 7: Computational comparison between the original and adapted L4DR baseline for temporal fusion.

Baseline	GFLOPs	Inference Speed	Params
Original Reported L4DR	255.60	96.0ms	61.32M
The Adapted L4DR	170.94 (-84.66)	66.75ms (-29.25ms)	52.72M (-8.6M)

utilized to filter unreliable 4D Radar points before early fusion. However, TMF aims to directly exploit temporal consistency and complementary information from concatenated multi-frame 4D Radar observations. Applying FAD before temporal aggregation may remove 4D Radar responses that are potentially valuable for cross-frame correlation, thereby affecting the learning of temporal 4D Radar characteristics and making it difficult to accurately evaluate the contribution of temporal information itself. Therefore, we remove the FAD when constructing the baseline to preserve the original multi-frame Radar observations and fairly evaluate the effectiveness of temporal fusion.

Table 8: Computational cost analysis of different components based on the adapted L4DR baseline.

TRD	RGTC	DGPF	GFLOPs	Inference Speed	Params
-	-	-	170.94	66.75ms	52.72M
✓	-	-	170.98 (+0.04)	72.52ms (5.77ms)	52.74M (+0.02M)
-	✓	-	192.67 (+21.73)	67.23ms (+0.48ms)	53.01M (+0.29M)
✓	✓	-	192.71 (+21.77)	78.12ms (+11.37ms)	53.02M (+0.30M)
✓	✓	✓	260.81 (+89.87)	93.80ms (+27.05ms)	56.57M (+3.85M)

Table 9: Quantitative comparison with different frame numbers on the K-Radar V1.0 dataset. Best results are highlighted in **bold**, and second-best results are underlined.

Frame	IoU	Metric	Total	Normal	Overcast	Fog	Rain	Sleet	Lightsnow	Heavysnow
3	0.3	AP_{BEV}	**84.1**	**84.6**	93.4	92.4	**83.7**	73.4	89.9	**72.4**
		AP_{3D}	79.3	**82.5**	89.0	**89.6**	**78.4**	**69.8**	84.1	58.1
	0.5	AP_{BEV}	79.6	79.1	92.8	92.3	79.7	67.0	89.3	66.1
		AP_{3D}	57.2	56.1	69.3	80.4	**55.7**	47.8	60.3	48.7
5	0.3	AP_{BEV}	83.4	83.9	93.1	92.2	82.6	71.3	89.3	72.2
		AP_{3D}	**79.4**	81.6	88.6	89.2	77.7	66.5	86.8	**58.4**
	0.5	AP_{BEV}	79.8	79.4	92.3	89.9	**80.6**	67.5	88.2	**66.3**
		AP_{3D}	57.0	56.9	65.4	77.5	55.3	45.4	67.1	**49.8**
7	0.3	AP_{BEV}	83.9	84.4	**93.5**	**92.8**	83.6	71.5	**92.0**	71.8
		AP_{3D}	**79.4**	82.1	**89.0**	89.1	77.2	67.0	**88.5**	55.6
	0.5	AP_{BEV}	**79.9**	**79.7**	**92.9**	**92.6**	80.1	68.6	88.7	65.3
		AP_{3D}	**57.8**	**60.0**	**71.1**	76.0	53.8	**51.2**	**67.2**	46.8
9	0.3	AP_{BEV}	83.7	83.8	93.5	**92.8**	83.5	**74.1**	89.6	72.2
		AP_{3D}	78.8	81.9	88.4	89.3	76.4	67.6	85.8	51.1
	0.5	AP_{BEV}	79.4	79.3	92.8	**92.6**	79.0	67.4	**89.3**	64.8
		AP_{3D}	57.1	57.6	66.6	77.3	55.6	48.2	63.2	39.7
11	0.3	AP_{BEV}	83.7	83.8	93.2	**92.8**	83.5	**74.1**	89.6	72.2
		AP_{3D}	79.3	82.0	88.3	88.9	77.3	69.4	86.0	54.7
	0.5	AP_{BEV}	79.8	79.2	92.6	92.5	79.4	**70.0**	88.6	65.0
		AP_{3D}	56.8	56.4	66.3	**80.5**	55.4	46.4	59.4	44.9

Second, we further optimize the LiDAR-4D Radar voxel matching operation in the MME block. The original implementation searches for corresponding spatial locations by computing the pairwise distance matrix between LiDAR and 4D Radar voxel coordinates:

$$\text{dist}_{ij} = \sum_{d} (L_i^d - R_j^d)^2. \tag{9}$$

However, this pairwise matching operation introduces $O(NM)$ computational complexity and requires an $N \times M$ distance matrix, resulting in considerable computational and memory overhead when the number of voxels increases. Therefore, we replace the original distance-based matching with an efficient coordinate hashing-based matching strategy. Specifically, the voxel coordinates are first encoded into compact integer keys, and the common voxel coordinates between LiDAR and 4D Radar are efficiently retrieved through sorting and binary search. This modification avoids explicitly constructing the pairwise distance matrix while preserving the same voxel correspondence as the original implementation. The computational cost comparison of L4DR before and after adaptation is presented in Table 7. All ablation experiments are conducted based on the adapted L4DR, ensuring that the performance improvements are attributed to the TMF components rather than implementation differences.

Computational Cost Analysis: The computational overhead introduced by different components of TMF is summarized in Table 8 (The first row denotes our reproduced L4DR baseline.), including GFLOPs, inference speed, and parameter number. TRD introduces negligible additional cost with only 0.02M extra parameters, demonstrating its efficiency in enhancing Radar reliability. In contrast, RGTC incurs moderate computational overhead due to historical LiDAR feature aggregation, introducing 21.73 GFLOPs and 0.29M parameters when applied independently. With all components integrated, TMF requires 93.80ms inference time and 260.81 GFLOPs, while introducing only 3.85M additional parameters compared to the baseline. These results indicate that TMF achieves an effective balance between temporal robustness and computational efficiency.

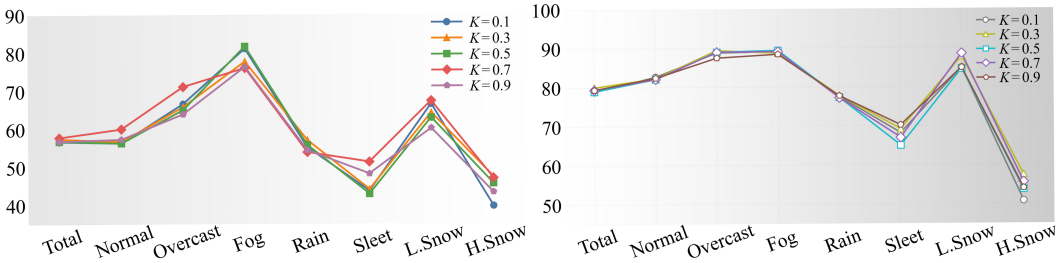


Figure 6: Performance comparison of different Top-K selection ratios in the TRD block on the K-Radar V1.0 dataset. The line charts report AP_{3D} under IoU=0.5 (left) and IoU=0.3 (right) across various weather conditions.

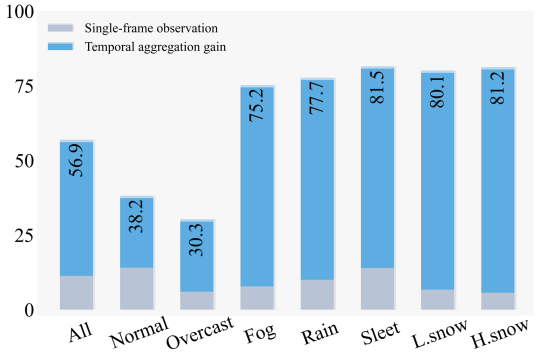


Figure 7: Object-region LiDAR observation coverage before and after temporal aggregation of the proposed TMF under adverse weather conditions.

Quantitative Evaluation of Temporal Observation: Additional results under an IoU threshold of 0.3 on the K-Radar V1.0 benchmark are presented in Table 9. Compared with the IoU=0.5 evaluation reported in the main paper, the performance variation with different temporal context lengths shows a consistent trend, where 7 input frames achieve the best overall detection performance. This indicates that the benefit of temporal information remains stable under different localization criteria and is not sensitive to the IoU threshold. Furthermore, increasing the temporal context from a limited number of frames improves detection performance by providing additional complementary observations, while excessively long temporal contexts may introduce redundant or unreliable responses. These results further demonstrate that TMF can effectively exploit temporal complementary cues to compensate for degraded observations and achieve robust 3D object detection under adverse weather conditions.

Top-K Ratio Analysis of TRD. We examine the effect of the Top-K selection ratio on temporal Radar refinement within the TRD block. As shown in Figure 6, different Top-K configurations are systematically evaluated on the K-Radar V1.0 under IoU thresholds of 0.5 and 0.3, covering a wide range of weather conditions. The results reveal that the Top-K selection ratio has a significant impact on detection performance. An excessively small K may remove useful 4D Radar responses, while an overly large K can introduce more unreliable observations during temporal aggregation, leading to suboptimal performance. Considering the overall detection performance across diverse weather conditions and evaluation settings, $K = 0.7$ provides the most robust performance among different configurations. Therefore, we adopt $K = 0.7$ as the default configuration for the TRD block.

Temporal LiDAR Observation Gain. To validate the effectiveness of the proposed TMF-based LiDAR temporal compensation, we compare the GT-region LiDAR response ratio, defined as the proportion of BEV cells from the GT-region exhibiting LiDAR responses, before and after temporal enhancement in Figure 7. The response ratio increases from 11.31% to 56.91% across all weather conditions, showing substantial enhancement of object-region observations. The ratio further improves from 8.82% to 79.17%, yielding a gain of 70.35% under adverse weather conditions. These results indicate that TMF effectively exploits reliable historical observations to alleviate object-region sensing degradation and improve robust 3D object detection under adverse weather conditions.