2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Toward Robust Collaborative Perception under Adverse Weather Conditions via Dual-branch Network

Yuquan Yang[1], Hui Zhang*[1,2], ZiYin Zhang[1], Wenyu Lu[1] and Xiaohua Xu[1]

***Abstract*— Recent advances in collaborative perception systems have led to significant improvements in 3D object detection performance. While widely deployed LiDAR and camera systems often experience performance degradation under adverse weather conditions, weather-robust 4D radar offers a promising alternative to address this challenge. However, effectively fusing 4D radar measurements with degraded LiDAR data remains a critical challenge. In this work, we decompose the weather-induced degradation in LiDAR perception into feature attenuation requiring enhancement and feature contamination requiring suppression, based on the underlying physical interactions. Building upon this decomposition, we propose a dual-branch network to handle each degradation pattern in a specialized manner. One branch focuses on enhancement based on spatial and channel attention, guided by 4D radar cues. The other branch focuses on suppression based on intra-modal structural consistency and cross-modal consistency. To achieve adaptive branch integration, we propose a dynamic decision network to generate a decision weight map for each branch and capture the complex interaction between branches. To validate the effectiveness of our method, we conduct extensive experiments on V2X-R, the only publicly available collaborative LiDAR-4D radar dataset. Extensive experimental results demonstrate that our method achieves improvements of 3.65% and 10.80% in mAP@0.7 under fog and snow conditions, respectively, outperforming previous state-of-the-art approaches.**

I. INTRODUCTION

Perceiving complex driving environments is essential for ensuring the safety of intelligent agents, such as autonomous vehicles [1], [2], [3]. With advancements in localization and perception technologies, even single-agent systems have achieved impressive performance in 3D object detection tasks [4], [5], [6]. However, these systems face inherent limitations, such as occlusions and reduced detection accuracy for distant objects. Collaborative perception has emerged as a promising approach to address these challenges. By leveraging vehicle-to-everything (V2X) communication, agents can share their perception data, enabling more comprehensive and extended environmental awareness [7], [8], [9].

Current research in collaborative perception predominantly relies on LiDAR data [10], [11], [12] or LiDAR-camera fusion [13], [14], [15] for downstream 3D object detection tasks. While camera data can provide rich semantic cues to complement LiDAR, both modalities are highly susceptible to environmental degradation, particularly under adverse weather conditions. Under such conditions, sensor inputs suffer from complex degradation, resulting in missing or noisy features. Crucially, due to the shared vulnerability of both modalities to environmental degradation, they may fail concurrently in critical scenarios, undermining the robustness of the collaborative perception system. Despite recent advances in robust perception [16], [17], the inherent characteristic limitations of these sensors constrain their adaptability under adverse weather conditions.

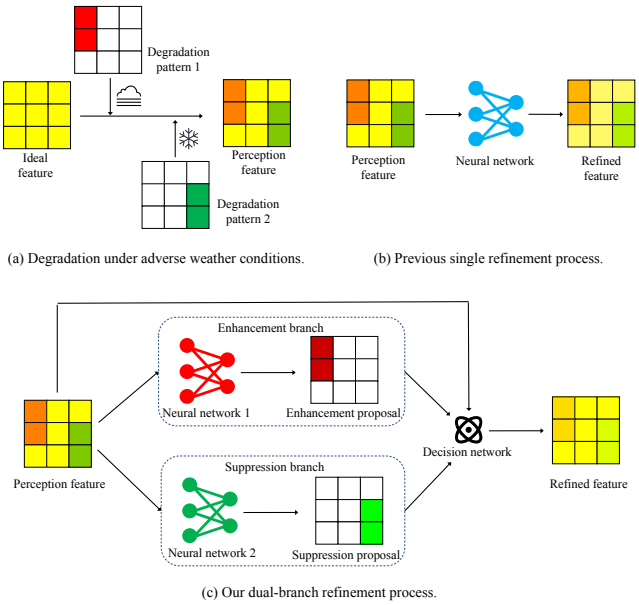

Fig. 1. (a) Perception feature results from a combination of degradation patterns induced by distinct underlying physical interactions. (b) Previous single refinement process, which may converge to a balanced compromise under the degradation from distinct patterns. (c) Our dual-branch refinement process, which explicitly decomposes and handles distinct weather-induced degradation patterns.

In recent years, 4D radar has garnered increasing attention due to its robustness under diverse weather conditions, emerging as a promising modality for detection tasks [18], [19], [20]. Despite this advantage, the inherently limited resolution of 4D radar makes it unsuitable for standalone perception tasks [18]. Therefore, fusing LiDAR and 4D radar data is regarded as an effective strategy to improve perception reliability under adverse weather conditions [21], [22]. In multi-agent collaborative settings, weather-induced degradation in LiDAR perception may accumulate during transmission and aggregation, resulting in more complex er-

* Corresponding author
[1]School of Computer Science and Technology, University of Science and Technology of China, Hefei, 230027, China.
Email: (Yuquan Yang, ZiYin Zhang, Wenyu Lu){yuquany, zhangziyin5163, luwenyu9}@mail.ustc.edu.cn,
(Hui Zhang, Xiaohua Xu){fzhh, xiaohuaxu}@ustc.edu.cn.
[2]Anhui Provincial Key Laboratory of High Performance Computing Hefei, 230027, China.

ror distributions that impair overall system performance. The existing collaborative LiDAR-4D radar fusion method [22] leverages a 4D-radar-conditioned diffusion model to implicitly mitigate weather-induced degradation in LiDAR features. Such an approach is prone to converging to suboptimal solutions during the diffusion process. On the other hand, weather-induced degradation in LiDAR perception usually varies significantly across different weather conditions, due to distinct underlying physical interactions. Adverse weather conditions degrade LiDAR perception through the interaction of laser beams with airborne particles [23]. Depending on their size and density, these particles may attenuate the laser beam through absorption and scattering, resulting in weakened or missing returns, or may generate spurious reflections that contaminate the point cloud with noise [24], [25]. Indeed, both effects may coexist as particles can simultaneously attenuate the laser beam and generate spurious reflections. As a result, LiDAR perception is often degraded by a combination of these effects, rather than by a single one, as shown in Fig. 1(a). Previous LiDAR-4D radar fusion methods [19], [20] under adverse weather conditions typically treat weather-induced degradation as a single pattern, attempting to address it with a single network architecture, which may fail to adapt to scenarios with varying degradation patterns, as shown in Fig. 1(b).

To address these challenges, we decompose the weather-induced degradation in LiDAR perception into feature attenuation requiring enhancement and feature contamination requiring suppression, based on distinct underlying physical interactions. Building upon this decomposition, we propose a dual-branch network to handle each degradation pattern in a specialized manner, as shown in Fig. 1(c). Specifically, one branch enhances weakened or missing regions of the LiDAR features using attention mechanisms guided by 4D radar structural cues. The other focuses on suppressing unreliable regions by leveraging intra-modal structural consistency in 4D radar and cross-modal consistency between the two modalities. To achieve adaptive branch integration, we propose a dynamic decision network to generate a decision weight map for each branch and capture the complex interaction between branches, yielding the final refined LiDAR feature map. To validate the effectiveness of our method, we conduct extensive experiments on V2X-R [22], the only publicly available collaborative LiDAR-4D radar dataset. The main contributions can be summarized as follows:

- We propose a dual-branch network architecture for robust LiDAR perception under adverse weather conditions by decomposing weather-induced degradation into two distinct patterns.
- We introduce a dynamic decision network that generates a decision weight map for each branch while capturing the complex interaction between branches.
- Extensive experiments on the only publicly available collaborative LiDAR-4D radar dataset V2X-R demonstrate that our method outperforms previous state-of-the-art approaches under adverse weather conditions.

II. RELATED WORK

A. Collaborative Perception

Collaborative perception enables multiple agents to share complementary perception information, thereby achieving a more comprehensive understanding of the environment. Depending on the stage of information transmission and fusion, collaborative perception modes can mainly be categorized into early fusion [26], intermediate fusion [27], and late fusion [28]. Intermediate fusion, sharing features among agents, has gained increasing attention for its favorable trade-off between performance and transmission bandwidth. While extensive research has explored aspects such as accuracy improvement [29], [30], pose estimation error [31], [32] and trade-off between perception performance and communication costs [9], [33], [34], performance degradation under adverse weather remains a critical yet overlooked challenge. Although a few works have begun to tackle this challenge, the fundamental problem remains. For instance, DSRC [17] attempts to learn weather-agnostic representations by focusing on density-insensitivity and semantic awareness. The majority of intermediate fusion approaches are based on LiDAR or LiDAR-camera frameworks, in which the inherent sensitivity of these sensors to adverse weather conditions remains a critical challenge for collaborative perception.

B. 3D Object Detection under Adverse Weather Conditions

Widely used sensors, such as LiDAR and cameras, are prone to degradation under adverse weather conditions. Physics-based weather simulations [35], [36] have been proposed to alleviate data scarcity. SPG [37] proposes a general completion framework to bridge domain gaps across different weather conditions. SRKD [16] introduces a general knowledge distillation framework that transfers performance from sunny to rainy conditions. Due to the potential degradation of sensors under adverse weather conditions, recent studies have explored 4D-radar-enhanced object detection to improve robustness. InterFusion [19] identifies cross-modal relationships between 4D radar and LiDAR features, and fuses information from both modalities through a self-attention mechanism. 3D-LDR [20] leverages the complementary strengths of LiDAR and 4D radar in 3D space through a learnable gating mechanism. L4DR [21] enhances model adaptability to diverse weather conditions by separately extracting cross-modal and intra-modal features from LiDAR and 4D radar inputs, followed by a multi-scale feature fusion strategy. However, these single-agent approaches may encounter challenges when they are extended to collaborative multi-agent scenarios.

III. METHOD

A. Collaboration Framework

Consider N agents collaboratively detecting 3D objects within the scenario. Each agent is equipped with both LiDAR and 4D radar sensors. Let X_i^k denote the observation of modality $k \in \{l, r\}$ of the i-th agent, where l and r denote the LiDAR and 4D radar modalities, respectively. The

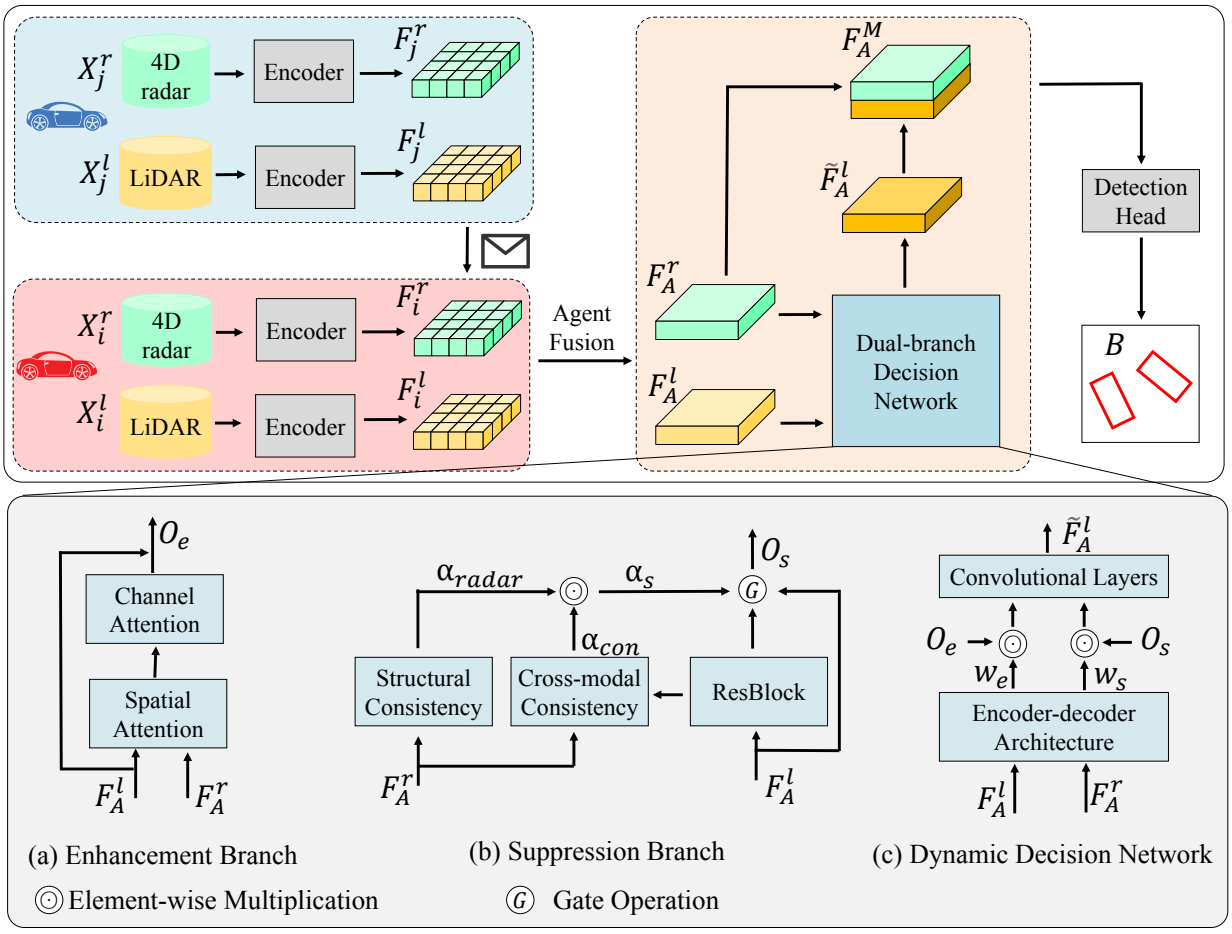

Fig. 2. Overview of our proposed method. The key designs are the dual-branch network, consisting of (a) and (b), and the dynamic decision network (c).

proposed collaborative 3D object detection for the i-th agent works as follows:

$$F_i^k = \phi_{Enc}(X_i^k), k \in \{l, r\}, \tag{1a}$$
$$F_A^k = \phi_A(F_i^k, \{F_{j\to i}^k\}), k \in \{l, r\}, \tag{1b}$$
$$\tilde{F}_A^l = \phi_R(F_A^l, F_A^r) \tag{1c}$$
$$F_A^M = \phi_M(\tilde{F}_A^l, F_A^r), \tag{1d}$$
$$B = \phi_{Det}(F_A^M), \tag{1e}$$

where F_i^k is the encoded feature of the corresponding observation of X_i^k, $F_{j\to i}^k$ is the feature of modality k transmitted from agent j to i, F_A^k is the multi-agent feature of modality k after agent fusion, $\tilde{F}_A^l$ is a refined version of F_A^l after the dual-branch network, F_A^M is the fused multi-agent multi-modal feature, and B is the detected bounding boxes.

In step (1a), agent i encodes its observations from two modalities into respective BEV features. In step (1b), for each modality, an agent-fusion network ϕ_A fuses agent i's feature F_i^k with the features from other agents to generate a multi-agent feature F_A^k. In step (1c), the multi-agent LiDAR feature F_A^l is refined by a refinement network ϕ_R, which leverages the weather-robust 4D radar feature F_A^r for guidance to generate a refined version $\tilde{F}_A^l$. In step (1d), a modal-fusion network ϕ_M fuses the refined LiDAR feature with the radar feature to generate a multi-agent multi-modal representation F_A^M, which is processed by a detection head to yield the final detection boxes B in step (1e). In the collaboration framework, the key lies in mitigating weather-induced degradation in step (1c).

B. Model Formulation

Under ideal conditions, LiDAR sensors can capture accurate scene information. During the collaboration process, the ideal multi-agent LiDAR feature is denoted by $\mathcal{F}_A^l$. However, under adverse weather conditions, the LiDAR sensors suffer significant degradation. This degradation process can be modeled as a single function $\mathcal{D}$, parameterized by a hidden weather factor θ:

$$F_A^l = \mathcal{F}_A^l + \mathcal{D}_\theta(\mathcal{F}_A^l). \tag{2}$$

Due to the fact that weather-induced degradation patterns arise from distinct underlying physical interactions, employing a single network module to simultaneously approximate fundamentally different inverse processes may force the model to learn a compromised solution, thereby impairing overall performance. Instead, we model weather-induced degradation as the combined effect of two interacting

degradation patterns:

$$\mathcal{F}_A^l = F_A^l + \mathcal{D}_a(\mathcal{F}_A^l) + \mathcal{D}_c(\mathcal{F}_A^l) + \Delta(\mathcal{D}_a, \mathcal{D}_c), \quad (3)$$

where $\mathcal{D}_a$ and $\mathcal{D}_c$ are the separate degradation functions to model the effect of attenuation and contamination, respectively, and $\Delta(\mathcal{D}_a, \mathcal{D}_c)$ represents the latent interaction between the two degradation patterns. Therefore, our objective is to learn the two separate functions and capture the latent interaction.

C. Dual-branch Network

According to the underlying physical interactions of LiDAR degradation, we decompose the weather-induced degradation into attenuation and contamination, and address them through a dual-branch network architecture to handle each degradation pattern in a dedicated and efficient manner.

To mitigate LiDAR feature degradation caused by attenuation, we enhance responses in weakened or missing regions using a cascaded attention module that exploits 4D radar feature F_A^r as structural priors to recover the geometric contours and structural details of valid objects, as shown in Fig. 2(a). Specifically, we concatenate the LiDAR feature F_A^l and the 4D radar feature F_A^r along the channel dimension. The concatenated feature is subsequently processed by a cascaded spatial attention module and a channel attention module to adaptively recalibrate feature responses. Finally, a 1×1 convolution projects the attention-weighted feature into a residual correction term, and the enhancement proposal O_e is obtained through residual learning. The enhancement process can be formulated as:

$$O_e = F_A^l + \text{Conv}(CA(SA([F_A^l, F_A^r]))), \quad (4)$$

where $SA(\cdot)$ and $CA(\cdot)$ denote spatial and channel attention modules adapted from CBAM [38], respectively. The residual formulation allows the enhancement branch to specifically target the attenuation effect without impairing the integrity of the original LiDAR feature.

To mitigate LiDAR feature degradation caused by contamination, we introduce a consistency check that leverages both intra-modal structural consistency in 4D radar and cross-modal consistency between the 4D radar feature and the LiDAR feature to identify and suppress spurious or unreliable responses, as shown in Fig. 2(b). Specifically, we apply a residual block [39] to derive a correction F_l from F_A^l, where F_l provides local adjustments for potentially contaminated areas:

$$F_l = \text{ResBlock}(F_A^l). \quad (5)$$

Subsequently, we compute the structural consistency α_{radar} leveraging the intrinsic regularity of the 4D radar feature, and the cross-modal consistency α_{con} based on the concatenated LiDAR and 4D radar features:

$$\alpha_{radar} = \sigma\left(W_r * F_A^r\right), \alpha_{con} = \sigma(W_c * [F_A^r, F_A^l]), \quad (6)$$

where σ is the sigmoid function, W_r and W_c are trainable convolutional kernels, and $*$ denotes the convolution operation. The two consistency maps are combined to form a spatial consistency map α_s:

$$\alpha_s = \alpha_{con} \odot \alpha_{radar}, \quad (7)$$

where $\odot$ is element-wise multiplication. The suppression proposal O_s is then computed as:

$$O_s = (1 - \alpha_s) \odot F_A^l + \alpha_s \odot F_l. \quad (8)$$

D. Dynamic Decision Network

After both branches generate specialized feature proposals tailored to their respective patterns, a decision strategy is required to regulate their latent integration based on contextual cues. Since the degradation patterns are often spatially heterogeneous and dynamic, a simple, global decision strategy would fail to adapt to local degradation patterns.

Therefore, we introduce a dynamic decision network to adaptively integrate these feature proposals, as shown in Fig. 2(c). The dynamic decision network employs a spatial weight generator $\mathcal{G}$, implemented as a lightweight encoder-decoder architecture with strided convolutions for downsampling and transposed convolutions for upsampling, to generate a decision weight map for each branch based on F_A^l and F_A^r:

$$[w_e, w_s] = \mathcal{G}([F_A^l, F_A^r]), \quad (9)$$

where w_e and w_s are the weight maps for the enhancement branch and the suppression branch, respectively. $\mathcal{G}$ captures multi-scale cross-modal contextual cues to account for spatially varying degradation patterns, enabling localized and adaptive integration decisions.

Unlike simple linear combinations that assume a fixed or additive interaction, we aim to capture the complex interaction between branches. To this end, we apply a lightweight combination network $\mathcal{C}$, consisting of several convolutional layers, which integrates the weighted branch proposals and generates the global residual correction term. This correction is then applied to the original LiDAR input feature F_A^l via a global residual connection, yielding the refined output $\tilde{F}_A^l$:

$$\tilde{F}_A^l = F_A^l + \mathcal{C}([w_e \odot O_e, w_s \odot O_s]). \quad (10)$$

By making dynamic decisions and capturing the complex interaction, our method adaptively modulates enhancement and suppression.

E. Overall Loss Function

Although the desired correction $\mathcal{C}([w_e \odot O_e, w_s \odot O_s])$ in Equation (10) is expected to align with the modeled degradation $\mathcal{D}_c(\mathcal{F}_A^l) + \mathcal{D}_a(\mathcal{F}_A^l) + \Delta(\mathcal{D}_c, \mathcal{D}_a)$ in Equation (3), LiDAR features collected under clear weather conditions are not perfectly ideal in practice. Hence, we adopt an end-to-end learning strategy without supervision on latent feature corrections, ensuring a unified behavior under all scenarios.

In the downstream detection task, we leverage the focal loss [40] for classification (denoted as $\mathcal{L}_{cls}$) and the smooth L1 loss for bounding box regression (denoted as $\mathcal{L}_{reg}$). The total loss is:

$$\mathcal{L}_{\text{total}} = \beta_{\text{cls}}\mathcal{L}_{\text{cls}} + \beta_{\text{reg}}\mathcal{L}_{\text{reg}}, \quad (11)$$

where $\beta_{\text{cls}}, \beta_{\text{reg}}$ are hyperparameters.

TABLE I

3D DETECTION PERFORMANCE COMPARISON UNDER CLEAR, FOG AND SNOW CONDITIONS.

Method	Clear			Fog			Snow		
	mAP@0.3	mAP@0.5	mAP@0.7	mAP@0.3	mAP@0.5	mAP@0.7	mAP@0.3	mAP@0.5	mAP@0.7
Attfuse	94.03	92.69	82.45	85.48	80.60	56.76	87.78	84.30	63.26
V2X-ViT	91.82	89.73	72.75	86.23	81.17	52.43	87.76	84.81	57.06
CoBEVT	87.71	87.21	74.61	82.14	79.25	55.86	81.22	80.27	61.93
CoAlign	93.84	92.75	82.64	85.94	81.33	61.62	88.10	85.22	67.54
InterFusion	83.94	80.36	48.53	73.49	69.37	40.69	79.55	75.19	40.53
L4DR	94.32	92.92	80.15	85.85	81.26	60.94	86.78	84.45	67.93
MDD	92.13	90.26	79.21	84.86	81.53	59.52	87.41	84.25	61.99
Ours	**95.07**	**94.23**	**85.69**	**87.34**	**83.52**	**65.27**	**92.25**	**90.72**	**78.73**

IV. EXPERIMENTS

A. Dataset and Metric

We evaluate our method on V2X-R [22], the only publicly available collaborative LiDAR-4D radar dataset. V2X-R is a simulated V2X collaborative perception dataset, jointly developed using the CARLA [41] autonomous driving simulator and the OpenCDA [42] collaborative driving automation framework. It contains LiDAR and 4D radar point clouds under simulated adverse weather conditions, including fog and snow, with a total of 37,727 data frames and 170,859 annotated 3D vehicle bounding boxes. We evaluate the performance of our method based on the final 3D object detection accuracy, which is measured using mean Average Precision (mAP) at Intersection-over-Union (IoU) thresholds of 0.3, 0.5, and 0.7.

B. Baseline Methods

The implementation of the aforementioned collaboration framework relies on the agent fusion and modal fusion networks. Due to the lack of dedicated methods for collaborative LiDAR-4D radar fusion, we build upon the implementations of MDD [22] to develop our collaborative perception framework. For state-of-the-art multi-agent single-modal methods such as AttFuse [14], V2X-ViT [29], CoBEVT [30] and CoAlign [31], we generalize these methods to support multi-modal fusion by processing 4D radar data in a manner similar to LiDAR and concatenating the features from both modalities. For state-of-the-art multi-modal methods such as InterFusion [19], L4DR [21], we derive their multi-agent versions by incorporating a self-attention-based agent-fusion module to aggregate features.

C. Implementation Details

Following the experimental setup in OPV2V [14], we designate a vehicle as the ego agent for evaluation. Detection performance is evaluated in the vicinity of the ego agent, within a spatial range of $x \in [0, 140]$ and $y \in [-40, 40]$. We assume a uniform communication range of 70m for all agents. All the agents outside this broadcasting radius of the ego agent will not participate in collaborative perception. Furthermore, since the 4D radar sensor only provides front-view data, all evaluation results are in the camera field of view of the ego agent. To ensure a fair comparison, all methods employ PointPillar [43] as the backbone to extract features from the point cloud. It converts the point cloud data into voxels with a resolution of 0.4m for both height and width. We implement the proposed and baseline methods using the PyTorch framework [44] and train them on eight RTX 3090 24G GPUs using the Adam optimizer [45] with $lr = 1e-3, \beta_1 = 0.9, \beta_2 = 0.999$. Following prior work [22], we set the hyperparameters as $\beta_{\text{cls}} = 1, \beta_{\text{reg}} = 2$.

D. Quantitative Evaluation

Table I presents a comprehensive comparison of 3D object detection performance between our method and several state-of-the-art methods under clear, fog and snow conditions. Under fog conditions, fog droplets cause strong scattering and signal attenuation, resulting in blurred distant objects and false points at close range. Under snow conditions, snowflakes induce spurious reflections that lead to data contamination, reducing valid surface points and increasing false returns. In these adverse weather conditions, our method achieves improvements of 3.65% and 11.19% in mAP@0.7 over existing multi-agent collaboration methods under fog and snow conditions, respectively. Compared with the well-designed LiDAR-4D radar fusion methods, our method achieves performance improvements of 4.33% and 10.80% in mAP@0.7 under fog and snow conditions, respectively. Notably, our method also outperforms all baseline methods under clear weather conditions. This consistent performance improvement across both adverse and clear weather conditions demonstrates that our method is not only robust to weather-induced degradation but also effective at mitigating potential feature degradation caused by other factors in clear environments.

E. Qualitative Evaluation

As shown in Fig. 3, we present the collaborative detection results of CoAlign, L4DR, and our method under fog and snow conditions. Both baseline methods exhibit noticeable artifacts in certain scenarios, such as missed detections and false positives highlighted by the blue bounding boxes. These cases indicate limitations of these methods in handling weather-induced degradation, particularly in preserving reliable feature representations under adverse conditions. In contrast, our method produces more accurate predictions across these scenarios, demonstrating superior robustness to weather-induced degradation.

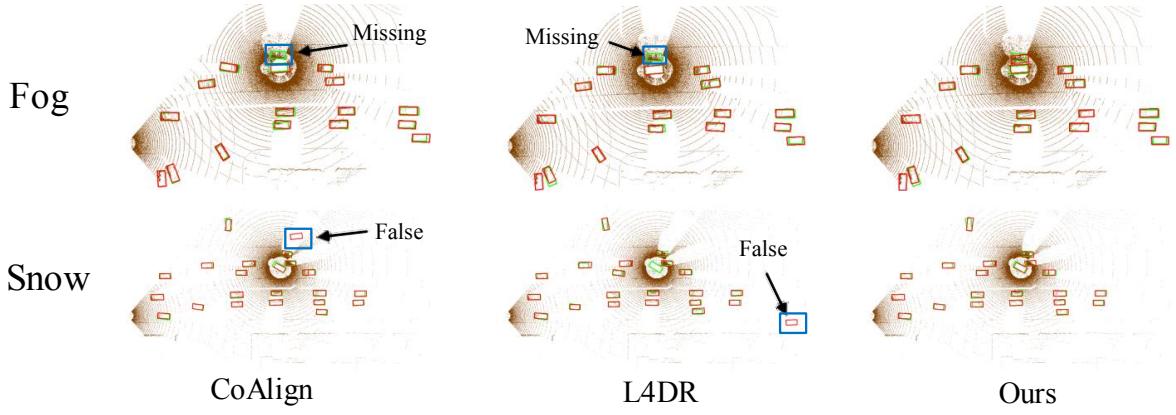

Fig. 3. Visualization of CoAlign, L4DR and our method under fog and snow conditions. Green and red boxes denote ground truth and predictions, respectively. We highlight the obvious errors in blue bounding boxes.

TABLE II

ABLATION STUDY RESULTS ON PROPOSED COMPONENTS, INCLUDING ENHANCEMENT BRANCH (EB), SUPPRESSION BRANCH (SB) AND DYNAMIC DECISION NETWORK (DDN).

Modules			Fog	Snow
SB	EB	DDN	mAP@0.3/0.5/0.7	mAP@0.3/0.5/0.7
			85.48/80.60/56.76	87.78/84.30/63.26
✓			81.25/76.21/57.88	90.25/87.70/69.83
	✓		86.29/82.48/64.55	87.49/84.62/64.53
✓	✓	✓	**87.34/83.52/65.27**	**92.25/90.72/78.73**

TABLE III

ABLATION STUDY RESULTS ON DIFFERENT DECISION METHODS, INCLUDING CONDITIONAL DEPLOYMENT (CD), GLOBAL WEIGHTED GATING (GWG), SPATIAL WEIGHTED GATING (SWG) AND DDN.

Decision method	Fog	Snow
	mAP@0.3/0.5/0.7	mAP@0.3/0.5/0.7
CD	86.29/82.48/64.55	90.25/87.70/69.83
GWG	83.41/80.38/62.74	90.13/88.11/71.99
SWG	85.13/81.77/63.62	90.74/89.09/78.16
DDN	**87.34/83.52/65.27**	**92.25/90.72/78.73**

F. Ablation Studies

Table II presents the ablation study results on proposed components, aiming to evaluate the individual and combined contributions of the designed branches. When only the suppression branch is added, the model shows a noticeable improvement under snow conditions, while it even drops below the baseline method under fog conditions. This is likely because the gate-based suppression branch effectively mitigates degradation caused by contamination, but is unable to handle the induced attenuation problem. However, the enhancement branch is effective in compensating for weakened or missing regions, but limited in handling degradation caused by contamination. When integrating the dynamic decision network, performance improves under both weather conditions, particularly under snow conditions. The success of combining both branches demonstrates that weather-induced degradation patterns are not uniform, but stem from the interaction of multiple factors. This further validates the effectiveness and robustness of our method under complex conditions.

Table III presents the ablation study results on different decision methods, aiming to evaluate the ability of our designed dynamic decision network. A simple way to integrate the two branches is conditional deployment according to specific weather conditions. However, it is less desirable in practice. To effectively integrate the branches, we explore several strategies. The global weighted gating integrates branches using a single weight for the entire feature, ignoring spatial variations and leading to unsatisfactory results. By incorporating spatial context, the spatial weighted gating generates a spatial weight map, enabling spatially adaptive integration of branches. Compared to the global weighted gating strategy, the spatial weighted gating strategy achieves a notable improvement, although it is still suboptimal. Specifically, the weighted gating strategies may enforce a compromise between branches even in regions where both could contribute positively. Instead, we use a spatial weight map for each branch to allow both branches to be activated simultaneously, thereby promoting more effective complementarity. The experiments demonstrate that our proposed method achieves the best overall performance.

V. CONCLUSIONS

In this paper, we decompose the weather-induced degradation into two distinct patterns based on the underlying physical interactions, and develop a dual-branch network tailored to each pattern. To integrate both branches, we introduce a dynamic decision network to generate a weight map for each branch and capture the complex interaction between branches. Extensive experimental results demonstrate that our method not only effectively addresses degradation caused by adverse weather conditions, but also handles potential degradation caused by other factors under clear conditions, consistently outperforming previous state-of-the-art methods. Future work will explore the trade-off between communication efficiency and perceptual robustness in more extreme weather conditions and real-world deployment settings.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (NSFC) under Grants No.62172383 and No.62231015, in part by Anhui Provincial Key R&D Program under Grant No.S202103a05020098, and in part by Research Launch Project of University of Science and Technology of China(USTC) under Grant No.KY0110000049.

REFERENCES

[1] S. Wang, Z. Yu, X. Jiang, S. Lan, M. Shi, N. Chang, J. Kautz, Y. Li, and J. M. Alvarez, "Omnidrive: A holistic llm-agent framework for autonomous driving with 3d perception, reasoning and planning," *CoRR*, 2024.

[2] H. Zhang, W. Zhou, Y. Zhu, X. Yan, J. Gao, D. Bai, Y. Cai, B. Liu, S. Cui, and Z. Li, "Visionpad: A vision-centric pre-training paradigm for autonomous driving," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 17165–17175, 2025.

[3] Z. Xing, X. Zhang, Y. Hu, B. Jiang, T. He, Q. Zhang, X. Long, and W. Yin, "Goalflow: Goal-driven flow matching for multimodal trajectories generation in end-to-end autonomous driving," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 1602–1611, 2025.

[4] G. Zhang, J. Chen, G. Gao, J. Li, S. Liu, and X. Hu, "Safdnet: A simple and effective network for fully sparse 3d object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14477–14486, 2024.

[5] X. Li, B. Fan, J. Tian, and H. Fan, "Gafusion: Adaptive fusing lidar and camera with multiple guidance for 3d object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21209–21218, 2024.

[6] M. Zheng, Z. Xu, Q. Xia, H. Wu, C. Wen, and C. Wang, "Seg2Box: 3D Object Detection by Point-Wise Semantics Supervision," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, pp. 10591–10598, 2025.

[7] J. Xu, Y. Zhang, Z. Cai, and D. Huang, "CoSDH: Communication-Efficient Collaborative Perception via Supply-Demand Awareness and Intermediate-Late Hybridization," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 6834–6843, 2025.

[8] H. Yu, W. Yang, J. Zhong, Z. Yang, S. Fan, P. Luo, and Z. Nie, "End-to-end autonomous driving through v2x cooperation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, pp. 9598–9606, 2025.

[9] Z. Ding, J. Fu, S. Liu, H. Li, S. Chen, H. Li, S. Zhang, and X. Zhou, "Point cluster: A compact message unit for communication-efficient collaborative perception," in *The Thirteenth International Conference on Learning Representations*, 2025.

[10] Y. Yuan, Y. Xia, D. Cremers, and M. Sester, "SparseAlign: A Fully Sparse Framework for Cooperative Object Detection," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 22296–22305, 2025.

[11] D. Qu, Q. Chen, T. Bai, H. Lu, H. Fan, H. Zhang, S. Fu, and Q. Yang, "SiCP: Simultaneous individual and cooperative perception for 3d object detection in connected and automated vehicles," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 8905–8912, IEEE, 2024.

[12] K. Yang, D. Yang, J. Zhang, M. Li, Y. Liu, J. Liu, H. Wang, P. Sun, and L. Song, "Spatio-temporal domain awareness for multi-agent collaborative perception," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 23383–23392, 2023.

[13] H. Xiang, R. Xu, and J. Ma, "HM-ViT: Hetero-modal vehicle-to-vehicle cooperative perception with vision transformer," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 284–295, 2023.

[14] R. Xu, H. Xiang, X. Xia, X. Han, J. Li, and J. Ma, "OPV2V: An Open Benchmark Dataset and Fusion Pipeline for Perception with Vehicle-to-Vehicle Communication," in *2022 International Conference on Robotics and Automation (ICRA)*, pp. 2583–2589, IEEE, 2022.

[15] W. Zimmer, G. A. Wardana, S. Sritharan, X. Zhou, R. Song, and A. C. Knoll, "Tumtraf v2x cooperative perception dataset," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 22668–22677, 2024.

[16] X. Huang, H. Wu, X. Li, X. Fan, C. Wen, and C. Wang, "Sunshine to rainstorm: Cross-weather knowledge distillation for robust 3d object detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, pp. 2409–2416, 2024.

[17] J. Zhang, Y. Wang, L. Qian, P. Sun, Z. Li, S. Jiang, M. Liu, and L. Song, "DSRC: Learning density-insensitive and semantic-aware collaborative representation against corruptions," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, pp. 9942–9950, 2025.

[18] R. Xu, Z. Xiang, C. Zhang, H. Zhong, X. Zhao, R. Dang, P. Xu, T. Pu, and E. Liu, "SCKD: Semi-supervised cross-modality knowledge distillation for 4d radar object detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, pp. 8933–8941, 2025.

[19] L. Wang, X. Zhang, B. Xv, J. Zhang, R. Fu, X. Wang, L. Zhu, H. Ren, P. Lu, J. Li, and others, "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, 2022.

[20] Y. Chae, H. Kim, and K.-J. Yoon, "Towards robust 3d object detection with lidar and 4d radar fusion in various weather conditions," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15162–15172, 2024.

[21] X. Huang, Z. Xu, H. Wu, J. Wang, Q. Xia, Y. Xia, J. Li, K. Gao, C. Wen, and C. Wang, "L4DR: Lidar-4dradar fusion for weather-robust 3d object detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, pp. 3806–3814, 2025.

[22] X. Huang, J. Wang, Q. Xia, S. Chen, B. Yang, X. Li, C. Wang, and C. Wen, "V2X-R: Cooperative LiDAR-4D Radar Fusion with Denoising Diffusion for 3D Object Detection," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 27390–27400, 2025.

[23] H. Lian, P. Sun, Z. Meng, S. Li, P. Wang, and Y. Qu, "LIDAR point cloud augmentation for dusty weather based on a physical simulation," *Mathematics*, vol. 12, no. 1, p. 141, 2023. Publisher: MDPI.

[24] M. Bijelic, T. Gruber, F. Mannan, F. Kraus, W. Ritter, K. Dietmayer, and F. Heide, "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*, pp. 11682–11692, 2020.

[25] J. Kim, B.-j. Park, and J. Kim, "Empirical analysis of autonomous vehicle's lidar detection performance degradation for actual road driving in rain and fog," *Sensors*, vol. 23, no. 6, p. 2972, 2023. Publisher: MDPI.

[26] Q. Chen, S. Tang, Q. Yang, and S. Fu, "Cooper: Cooperative Perception for Connected Autonomous Vehicles Based on 3D Point Clouds," in *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, pp. 514–524, IEEE, 2019.

[27] Q. Chen, X. Ma, S. Tang, J. Guo, Q. Yang, and S. Fu, "F-cooper: Feature based cooperative perception for autonomous vehicle edge computing system using 3D point clouds," in *Proceedings of the 4th ACM/IEEE Symposium on Edge Computing*, pp. 88–100, 2019.

[28] R. Xu, W. Chen, H. Xiang, X. Xia, L. Liu, and J. Ma, "Model-Agnostic Multi-Agent Perception Framework," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1471–1478, IEEE, 2023.

[29] R. Xu, H. Xiang, Z. Tu, X. Xia, M.-H. Yang, and J. Ma, "V2X-ViT: Vehicle-to-everything cooperative perception with vision transformer," in *European conference on computer vision*, pp. 107–124, Springer, 2022.

[30] R. Xu, Z. Tu, H. Xiang, W. Shao, B. Zhou, and J. Ma, "CoBEVT: Cooperative Bird's Eye View Semantic Segmentation with Sparse Transformers," in *Proceedings of The 6th Conference on Robot Learning*, pp. 989–1000, PMLR, 2023.

[31] Y. Lu, Q. Li, B. Liu, M. Dianati, C. Feng, S. Chen, and Y. Wang, "Robust Collaborative 3D Object Detection in Presence of Pose Errors," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4812–4818, IEEE, 2023.

[32] Z. Lei, Z. Ni, R. Han, S. Tang, C. Feng, S. Chen, and Y. Wang, "Robust collaborative perception without external localization and clock devices," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 7280–7286, IEEE, 2024.

[33] Y. Hu, S. Fang, Z. Lei, Y. Zhong, and S. Chen, "Where2comm: Communication-efficient collaborative perception via spatial confidence maps," *Advances in neural information processing systems*, vol. 35, pp. 4874–4886, 2022.

[34] Y. Hu, J. Peng, S. Liu, J. Ge, S. Liu, and S. Chen, "Communication-efficient collaborative perception via information filling with codebook," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15481–15490, 2024.

[35] M. Hahner, C. Sakaridis, M. Bijelic, F. Heide, F. Yu, D. Dai, and L. Van Gool, "Lidar snowfall simulation for robust 3d object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16364–16374, 2022.

[36] V. Kilic, D. Hegde, A. B. Cooper, V. M. Patel, and M. Foster, "Lidar light scattering augmentation (lisa): Physics-based simulation of adverse weather conditions for 3d object detection," in *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, IEEE, 2025.

[37] Q. Xu, Y. Zhou, W. Wang, C. R. Qi, and D. Anguelov, "SPG: Unsupervised domain adaptation for 3d object detection via semantic point generation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 15446–15456, 2021.

[38] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 3–19, 2018.

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[40] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, 2017.

[41] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Conference on robot learning*, pp. 1–16, PMLR, 2017.

[42] R. Xu, Y. Guo, X. Han, X. Xia, H. Xiang, and J. Ma, "OpenCDA: an open cooperative driving automation framework integrated with co-simulation," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pp. 1155–1162, IEEE, 2021.

[43] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "Pointpillars: Fast encoders for object detection from point clouds," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 12697–12705, 2019.

[44] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, and others, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.

[45] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations (ICLR)*, 2015.