4D Radar Meets LiDAR and Camera: Cooperative Perception under Adverse Weather

Melih Yazgan[1,2†] Iramm Hamdard[1,2†] Qiyuan Wu[2] J. Marius Zoellner[1,2]
[1]FZI Research Center for Information Technology, [2]Karlsruhe Institute of Technology
last.name@fzi.de
[†]These authors contributed equally.

Abstract

Cooperative perception is important for autonomous driving but remains fragile when cameras and LiDAR degrade in adverse weather. We address this challenge by integrating 4D imaging radar as a weather-robust modality into collaborative perception and introducing a Doppler-guided spatial attention mechanism for multi-agent fusion. Our approach extends two representative backbones: a radar-camera pipeline where radar substitutes LiDAR, and a LiDAR-radar pipeline where radar complements LiDAR. To support evaluation, we release radar-augmented benchmarks, OPV2V-R and Adver-City-R, with physics-based LiDAR degradation. Experiments show strong robustness gains in fog and rain, including substantial improvements when radar replaces degraded LiDAR. Additional validation on MAN TruckScenes demonstrates transfer beyond simulation. Overall, our results highlight 4D imaging radar as a robust modality for all-weather collaborative perception. Dataset and code are available at: https://url.fzi.de/SlimComm.

1. Introduction

The safe deployment of autonomous vehicles hinges on reliable perception. While autonomous driving has made remarkable progress in recent years, particularly in environmental perception [16, 18, 33], failures in perception directly translate into safety risks. Robust perception remains a prerequisite for large-scale deployment [26], yet adverse weather such as rain and fog can severely degrade camera- and LiDAR-based systems [32]. Cameras suffer from reduced visibility, while LiDAR suffers from scattering and attenuation, leading to degraded detection and tracking performance.

To mitigate these vulnerabilities, modern autonomous platforms adopt multi-sensor fusion strategies that exploit complementary strengths across modalities [6, 32, 37].

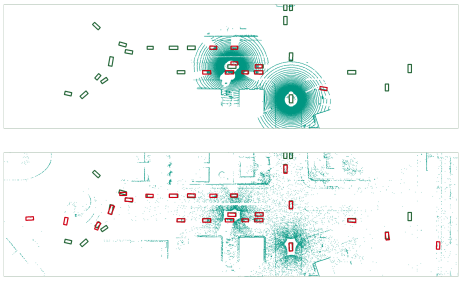

Figure 1. Impact of adverse weather on LiDAR and radar perception. Top: LiDAR point clouds degrade heavily under fog and rain, losing density and range. Bottom: Radar remains stable, with Doppler cues highlighting movers. This complementary robustness motivates our radar-augmented collaborative fusion. (Simulated via CARLA [2] and the LISA [12] framework).

Among them, four-dimensional (4D) imaging radar has attracted growing interest because it remains robust in poor visibility and directly provides Doppler velocity measurements [24, 45]. Its resilience makes 4D radar a promising complementary modality for all-weather perception, yet it remains underexplored in collaborative frameworks. A parallel trend is collaborative perception, where connected vehicles and infrastructure share sensor information to overcome occlusion and extend perceptual range. Recent surveys and benchmarks highlight its potential and open challenges [8, 9, 40]. However, the majority of collaborative perception methods have been evaluated predominantly under clear-weather conditions, with limited systematic treatment of multi-modal sensor degradation under adverse weather [42]. LiDAR, the dominant modality in collaborative frameworks, is also fragile in adverse weather, as illustrated in Fig. 1. Progress is further bottlenecked by the absence of suitable datasets. Widely used resources such as OPV2V [35], Adver-City [11], and DAIR-V2X [44] each miss essential requirements such as synchronized 4D radar, realistic weather degradation, or

full 360° coverage. Reviews of collaborative-perception datasets confirm that no public benchmark currently satisfies the combined need for 4D radar, 360° layouts, and physics-based weather simulation [31, 41]. SlimComm [43] previously integrated radar into OPV2V and Adver-City and benchmarked LiDAR-radar collaborative perception, but left weather-induced sensor degradation unaddressed. Our contributions are threefold:

1. **Radar-augmented collaborative fusion:** We integrate 4D radar into two representative collaborative baselines: substituting LiDAR in BM2CP to form a radar-camera pipeline, and complementing LiDAR in Where2comm to form a LiDAR-radar pipeline.
2. **Doppler-guided spatial attention:** We introduce an attention mechanism that leverages radar Doppler velocity to create a dynamic Bird's-Eye-View (BEV) mask, effectively emphasizing moving objects across multi-agent features.
3. **Adverse-weather benchmarks and evaluation:** We release OPV2V-R and Adver-City-R with physics-based weather degradation. We demonstrate strong robustness against severe weather, communication latency, and spatial misalignment, and validate our modules on the real-world MAN TruckScenes [3] single-vehicle dataset to confirm transferability beyond simulation.

2. Related Work

2.1. Radar Fusion in Single-Vehicle Perception

Autonomous vehicles typically rely on cameras, LiDAR, and radar as core exteroceptive sensors. Cameras provide rich semantics but degrade under poor visibility, while LiDAR delivers precise geometry but suffers from scattering and attenuation in rain and fog [27, 45]. Radar, by contrast, is inherently resilient to visibility degradation and directly provides Doppler velocity cues. The emergence of 4D radar has therefore motivated a range of fusion methods that exploit radar either in combination with cameras, LiDAR, or as a standalone modality.

Early approaches such as CRFNet [21] and RAMP-CNN [4] integrated radar detections with images, while CenterFusion [20], RCFusion [47], and CRAFT [13] demonstrated improved detection under challenging conditions. More recent works leverage 4D radar directly: InterFusion [30] proposed interaction-based LiDAR-radar fusion, LXL [34] excluded LiDAR entirely with a radar occupancy-assisted depth strategy, and MSSF [17] introduced a multi-stage radar-camera sampling framework. RadarPillars [19] further advanced radar-only detection by enriching velocity features with intra-pillar offsets and applying radar-tailored attention.

Taken together, these approaches confirm 4D radar as a weather-resilient and cost-effective modality for perception. However, they remain confined to the *single-vehicle* setting and do not address multi-agent collaboration.

2.2. Collaborative Perception under Adverse Weather

Collaborative perception extends perceptual range and mitigates occlusions by allowing connected vehicles to exchange features. Most prior works have focused on LiDAR-camera pipelines [10, 15, 28], showing that collaboration improves robustness but still inherits LiDAR's fragility in adverse weather. Jiang et al. [10] addressed uncertainty in rain via probabilistic feature aggregation, while Li et al. [15] proposed domain generalization to unseen weather. Tsakmakopoulou and Moustakas [28] showed that V2V communication helps mitigate fog degradation in cooperative pipelines by integrating their proposed method on top of S-AdaFusion [25]. Complementing these, Wang et al. [29] surveyed adverse-weather cooperative strategies, emphasizing the need for efficiency under communication constraints. Taking a different approach, Huang et al. [9] introduced AttFuse w/MDD, a denoising diffusion model that uses weather-robust 4D radar features to guide the reconstruction of noisy LiDAR features. While advanced techniques like denoising diffusion show promise for improving accuracy, they can also introduce significant computational overhead. In contrast, our work prioritizes an efficient solution, and no prior collaborative perception method has explicitly leveraged radar Doppler cues to guide attention across agents in a lightweight manner.

2.3. Datasets for Collaborative Radar Perception

Progress in collaborative perception relies heavily on dedicated datasets. Large-scale single-vehicle datasets such as nuScenes [1], K-Radar [22], and MAN TruckScenes [3] include radar and adverse-weather conditions but lack cooperative settings. Collaborative datasets such as OPV2V [35], DAIR-V2X [44], and Adver-City [11] support multi-agent fusion, but OPV2V lacks realistic weather, DAIR-V2X has restricted modality coverage, and Adver-City does not model LiDAR degradation. TUMTraf-V2X [48] provides real-world V2X cooperative perception with camera and LiDAR, but does not include radar as a cooperative modality.

Radar-focused cooperative datasets remain limited. V2X-R [9] provides simulated radar perception but without full 360° coverage and with Doppler encoded only as a boolean flag. V2X-Radar [39] introduces real-world radar data, but its public release omits cooperative elements and suffers from fidelity issues in ego-velocity and Doppler. SlimComm [43] contributed radar-augmented versions of OPV2V and Adver-City, enabling early LiDAR-radar cooperative benchmarks, but without modeling adverse weather.

Summary. Existing works confirm the promise of radar-camera fusion for single-vehicle perception, with spe-

cialized designs such as RadarPillars introducing velocity offsets, pillar-level attention, and uniform backbone scaling. However, these approaches are tailored to radar-only pipelines and do not address collaboration. Collaborative methods, by contrast, still rely primarily on LiDAR and remain fragile in adverse weather. LiDAR-radar cooperative fusion has been attempted in SlimComm [43], yet without Doppler-guided attention and without systematic weather modeling. Our work addresses these gaps by (i) proposing the Doppler-guided attention for collaborative radar fusion, applicable to both radar-camera and LiDAR-radar backbones, and (ii) extending the SlimComm datasets with physics-based weather degradation, resulting in OPV2V-R and Adver-City-R benchmarks for all-weather collaborative perception.

3. Methodology

Our approach extends two representative collaborative perception backbones: BM2CP [46], which fuses LiDAR and camera, and Where2comm [7], which is LiDAR-only and communication-efficient. We integrate 4D radar into both: (i) in BM2CP, radar substitutes LiDAR, forming a radar-camera pipeline, and (ii) in Where2comm, radar complements LiDAR, forming a LiDAR-radar pipeline. In both cases, we further introduce a Doppler-guided spatial attention mechanism that emphasizes moving objects while preserving static context. Fig. 2 illustrates our design on BM2CP; integration into Where2comm follows the same principles but with radar as an additional modality.

3.1. Radar Aggregation with Historical Frames

4D radar inherently suffers from point cloud sparsity, often leading to inaccurate and low-quality bounding box predictions [36]. To address this, we leverage multiple historical frames to densify radar point clouds.

Each radar return provides spatial coordinates (x, y, z) and a relative radial velocity v_{rel} (Doppler velocity). Since Doppler encodes relative motion only along the sensor line-of-sight, we compute the ego-compensated velocity v_r in the global frame as

$$v_r = v_{\text{rel}} + (\mathbf{v}_{\text{ego}} \cdot \mathbf{u}), \tag{1}$$

where $\mathbf{v}_{\text{ego}}$ is the ego-vehicle velocity and $\mathbf{u} = (u_x, u_y, u_z)$ is the unit vector from the sensor to the target. The resulting v_r denotes the point's absolute velocity along the line-of-sight. A point is classified as dynamic if the absolute value $|v_r| > \epsilon$, with ϵ a small threshold.

In data augmentation, static points from historical frames remain unchanged since they do not move over time. In contrast, historical dynamic points are compensated along the line-of-sight using their ego-compensated velocity v_r and the time gap Δt:

$$(x', y', z') = (x, y, z) + v_r \cdot \Delta t \cdot \mathbf{u}. \tag{2}$$

Subsequently, the compensated historical point clouds are transformed into the current frame's coordinate system and concatenated with the current point cloud to obtain the final aggregated data. The aggregation result is presented in Fig. 3, while the clustered dynamic/static map is shown in the Point-Level Dynamic Map of Fig. 5.

3.2. Velocity-Conditioned PillarVFE

Following RadarPillars [19], we enrich radar point representations with velocity features to capture direction-aware motion cues beyond raw Doppler. The compensated absolute radial velocity v_r provides stable motion estimates, while its Cartesian components $(v_{r,x}, v_{r,y}) = (u_x, u_y) \cdot v_r$ explicitly encode ground-plane motion direction, making object dynamics more interpretable for the network.

As shown in Fig. 4, we collect these descriptors into a velocity vector

$$\mathbf{v} = (v_{\text{rel}}, v_r, v_{r,x}, v_{r,y}) \in \mathbb{R}^4, \tag{3}$$

normalize it, and concatenate with spatial coordinates to form the radar feature

$$\mathbf{f} = (x, y, z, \mathbf{v}) \in \mathbb{R}^7. \tag{4}$$

A lightweight two-layer MLP with BatchNorm and PReLU refines the velocity embedding:

$$\hat{\mathbf{v}} = f_\theta(\mathbf{v}). \tag{5}$$

This design preserves Doppler sign, compensates scale differences, and enables the network to non-linearly reweight motion cues. The resulting embedding $\hat{\mathbf{v}}$ is combined with geometry and offsets before pillar max-pooling.

3.3. Doppler Mask Generation

For each voxel, we compute a motion saliency score from the compensated velocity features. Depending on the configuration, we apply a soft score $\sigma(\tau(|v_r| - \epsilon))$, where σ is the logistic function and τ controls steepness. Voxel scores are then reduced across all points within a voxel (by mean or max), yielding a per-voxel confidence value in $[0, 1]$. These values are scattered onto the BEV grid to form a dense saliency map $\mathbf{M}$. To capture the spatial extent of objects and consolidate fragmented activations, we optionally apply morphological dilation (max-pooling), producing the final dynamic mask $\tilde{\mathbf{M}}$. An explicit visualization of this process is provided in Fig. 5.

3.4. Multi-Stage Mask-Guided Attention

The dilated Doppler mask $\tilde{\mathbf{M}}$ is injected into the network at three stages, ensuring that motion cues are exploited at multiple abstraction levels:

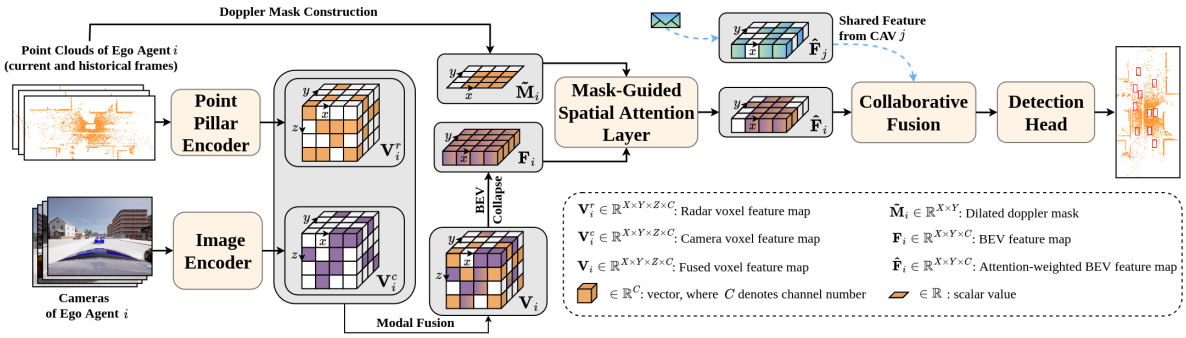

Figure 2. Overview of the proposed architecture, illustrated on BM2CP where radar replaces LiDAR. Radar features are encoded via a velocity-conditioned PillarVFE, and a Doppler-derived mask guides residual spatial attention in BEV space. In Where2comm, the same radar modules are used, but radar complements LiDAR features rather than substituting them.

Pre-Fusion Residual Gating. BEV backbone features $\mathbf{S}$ are modulated before multi-agent fusion:

$$\mathbf{S} \leftarrow \mathbf{S} \odot (1 + \gamma_{\text{pre}} \tilde{\mathbf{M}}), \tag{6}$$

with $\gamma_{\text{pre}} \geq 0$ learnable. This early gating primes the backbone to focus on dynamic regions from the outset.

Figure 3. Comparison of point clouds before augmentation (top only with current points) and after augmentation (bottom with both current and compensated historical points). With motion compensation, historical points are aligned to the current frame, resulting in a denser point cloud.

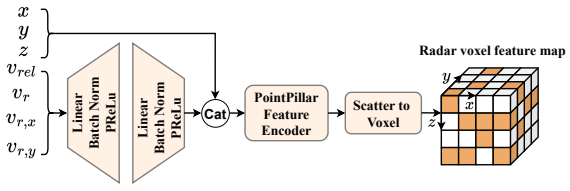

Figure 4. Overview of Doppler-aware motion encoding. Radar returns provide Doppler velocities that are ego-compensated and decomposed into planar components. These cues are embedded by a velocity MLP and further used to generate a voxel feature map.

Channel Gating. A CBAM-style channel gate reweights feature channels using global average and max pooling followed by a shared MLP. While spatial gating highlights *where* to focus, this stage emphasizes *which* feature types (e.g., density, edges) are most informative.

Residual Mask-Guided Spatial Attention. On the fused BEV $\mathbf{F}$, we concatenate average- and max-pooled maps with $\tilde{\mathbf{M}}$, apply a 7×7 convolution and normalization, and obtain an attention map $\mathbf{A}$:

$$\mathbf{A} = \sigma\Big(\text{Norm}(\text{Conv}_{7\times 7}([\,\text{AvgPool}(\mathbf{F}),\, \text{MaxPool}(\mathbf{F}),\, \tilde{\mathbf{M}}\,]))\Big). \tag{7}$$

A residual weighting formulation preserves static context while emphasizing motion-relevant regions:

$$\hat{\mathbf{F}} = \mathbf{F} \odot \big(1 + \gamma \mathbf{A}\big), \qquad \gamma \geq 0. \tag{8}$$

This final stage aligns semantic content with Doppler-derived motion saliency. This process is illustrated in detail in Fig. 5.

3.5. Backbone Integration

The above radar encoder, Doppler mask, and attention modules are designed as plug-and-play components. Their integration differs slightly between backbones:

BM2CP (radar-camera). The LiDAR branch is replaced by a radar backbone. Pre-fusion gating is applied after radar-camera projection into BEV space, channel gating operates within the backbone, and residual spatial attention refines fused BEV features before collaborative detection.

Where2comm (LiDAR-radar). Radar voxels are concatenated with LiDAR features, while a Doppler mask is computed in parallel. Pre-fusion gating modulates the combined BEV features before backbone processing and communication, channel gating follows feature compression,

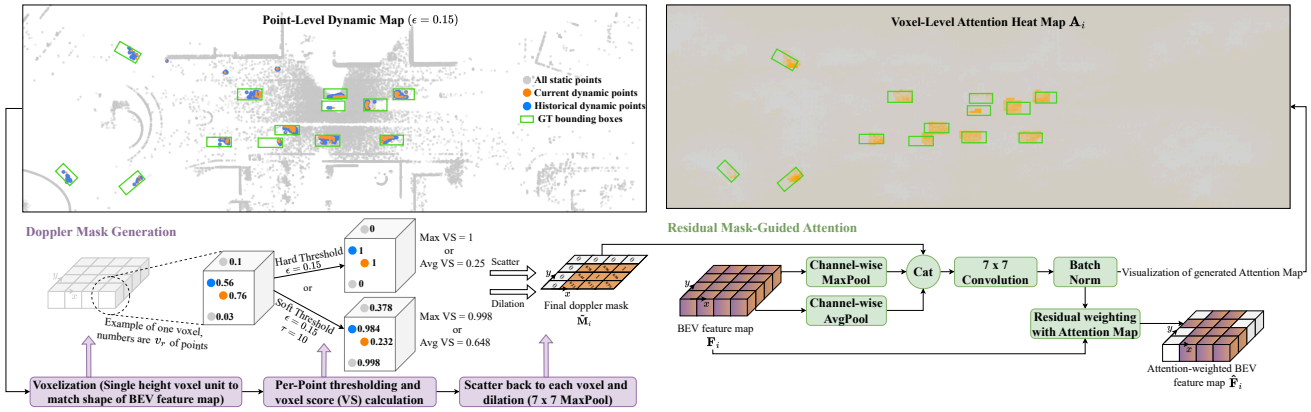

Figure 5. Overview of Doppler Mask Generation and Mask-Guided Spatial Attention. The Point-Level Dynamic Map illustrates how augmented historical data refine the representation of dynamic points, particularly in distant regions. The Voxel-Level Attention Heatmap demonstrates how the Doppler mask emphasizes moving objects and effectively guides spatial attention within the fusion backbone.

and spatial attention operates before the spatial confidence map is generated.

4. Experiments

We conducted a comprehensive evaluation of our radar-augmented collaborative perception framework. Our objectives are threefold: (i) assess model robustness under adverse weather, (ii) analyze the impact of different weather types on detection accuracy, and (iii) evaluate generalization from clear-weather training to adverse-weather conditions. In addition, we quantify the contribution of radar-camera fusion, LiDAR-radar fusion, and our mask-guided multi-stage Doppler attention mechanism.

4.1. Sensor-Level Weather Augmentation

In CARLA [2], adverse weather primarily affects the camera stream, while LiDAR and radar remain idealized. LiDAR is simulated via geometric ray-casting without scattering, and radar lacks wave-particle interaction modeling. As a result, rain or fog has no effect on these modalities in synthetic datasets such as Adver-City [11], limiting realism.

To address this, we augment LiDAR with the LISA framework [12], which models scattering-based degradation via Monte Carlo simulation. Rain is simulated using Mie scattering and the Marshall-Palmer raindrop distribution, producing attenuation, range-dependent point loss, and spurious near-range returns. Following [45], we set rainfall intensities to $12\,\mathrm{mm/h}$ (light) and $40\,\mathrm{mm/h}$ (heavy). For fog, LISA applies homogeneous attenuation and gamma-distributed range noise, reducing effective LiDAR range. In compound conditions, rain and fog models are applied sequentially.

Radar, by contrast, is minimally affected by weather. Studies [45] report only negligible impacts for radar in fog/rain, versus moderate-to-severe effects for LiDAR and cameras. Real-world datasets such as MAN TruckScenes [3] confirm this resilience, with radar detection stability across fog and rain. Given this robustness and the absence of validated radar degradation models, we omit radar augmentation to avoid introducing artifacts. Modeling radar degradation without validated physics-based models would be speculative and risk unfairly biasing the benchmark, whereas empirical studies confirm radar's stability in fog and rain.

The resulting Adver-City-R dataset, therefore, includes physically grounded weather degradation for cameras and LiDAR, while radar remains intact as a robust modality. This provides a consistent and realistic benchmark for evaluating collaborative perception in adverse conditions, underscoring radar's complementary role when LiDAR degrades, consistent with the physical sensor behaviors observed in operational environments.

4.2. Training and Evaluation Strategy

We consider two representative backbones: **BM2CP** (LiDAR-camera) and **Where2comm** (LiDAR-only). For each, we integrate 4D radar either as a replacement (BM2CP: radar-camera) or as a complement (Where2comm: LiDAR-radar), and evaluate both with and without our Doppler-guided attention.

All models are trained on the **OPV2V-R** dataset (clear weather) and tested in a zero-shot manner on **Adver-City-R**, which includes soft rain, heavy rain, and fog.

4.3. Implementation Details

All experiments were implemented in PyTorch and executed on a workstation with an NVIDIA GeForce RTX 4090 GPU. Each baseline model was trained using the official settings and hyperparameters provided in its original pub-

lication to ensure reproducibility. Both LiDAR and radar point clouds are voxelized with a horizontal resolution of 0.4×0.4 m and a vertical resolution of 4.0 m. The perception range is set to $[-40, 40]$ m laterally, $[-4, 4]$ m vertically, and $[-140, 140]$ m longitudinally. Multi-agent cooperation is limited to a 70 m maximum communication range, with up to five connected vehicles in each scenario. Standard LiDAR geometric augmentations, such as rotation and translation, were not applied, since such transformations would distort Doppler velocity patterns and disrupt the spatial alignment required for collaboration [23].

4.4. Quantitative Results and Analysis

Unless stated otherwise, we report *Average Precision* (AP) at three IoU thresholds, $\{0.3, 0.5, 0.7\}$, following the OPV2V/Adver-City evaluation protocol. For clarity, we always indicate whether the score refers to the *Sparse* (S), *Dense* (D), or *Combined* (C) split. These splits correspond to low (S) and high (D) traffic densities in the simulation scenarios.

4.4.1. Baseline under Clear Weather: OPV2V-R

To establish a baseline under clear-weather conditions, we evaluate both BM2CP and Where2comm backbones on the OPV2V-R test split. As shown in Tab. 1, the LiDAR-based BM2CP achieves the highest accuracy among our variants, consistent with the strong spatial density of LiDAR point clouds. The radar-camera variant (BM2CP-RC) performs lower, but our extension with Doppler-guided masking and spatial attention (BM2CP-RCA) recovers a notable part of this gap, confirming that motion-aware masking strengthens radar perception even under ideal conditions.

For completeness, we also report results for the Where2comm backbone and its radar-augmented variants. Unlike BM2CP, where radar substitutes LiDAR, in Where2comm radar complements LiDAR, providing an additional source of geometric information. Including these results demonstrates that our Doppler-guided masking mechanism is not tied to a single backbone design but can be integrated across different collaborative perception frameworks, yielding consistent benefits.

Table 1. Performance on OPV2V-R (clear weather). AP is reported for three IoU thresholds: 0.3, 0.5, and 0.7. RC (radar-camera), RCA (RC + mask-guided spatial attention), LR (LiDAR-radar), LRA (LR + mask-guided spatial attention).

Model Variant	AP@0.3	AP@0.5	AP@0.7
BM2CP [46] (LiDAR-camera)	**85.39**	**83.65**	**63.25**
BM2CP-RC	78.19	75.59	55.77
BM2CP-RCA	82.89	77.95	58.80
Where2comm [7] (LiDAR)	87.10	86.00	73.50
Where2comm-LR	87.49	86.21	**76.50**
Where2comm-LRA	**91.00**	**90.00**	75.00

4.4.2. Zero-shot generalization on Adver-City-R

Tab. 3 reports zero-shot performance across adverse weather conditions. Note that the *Clear* baseline in the table is evaluated on the Adver-City-R dataset. The lower absolute AP compared to Tab. 1 is attributed to the increased complexity in the scenarios [43]. Within the BM2CP family, the LiDAR-camera (LC) baseline remains superior in clear weather and rain, but performance collapses in fog (32.97 AP@0.5) due to LiDAR backscatter. Transitioning to radar-camera (RC) restores basic visibility, though our Doppler-guided RCA is the decisive factor for high precision, achieving the best overall performance in sparse fog (65.12 AP) and fog+rain (65.00 AP).

Notably, RCA surpasses the LiDAR-radar LRA model in these sparse settings, suggesting that a refined radar signal with intelligent attention can be more reliable than fusion hindered by severely degraded LiDAR features. For the Where2comm family, the LiDAR-only model drops sharply in fog ($\approx$36.00 AP); while the LRA variant dominates dense splits, the performance gap between LR and LRA confirms that Doppler-based attention is essential to align motion cues with spatial context. These results establish radar-augmented setups as competitive alternatives to tri-modal fusion in extreme conditions, proving that our attention mechanism is key to unlocking radar's full potential.

4.5. Comparison with State-of-the-Art Methods

Table 2. Comparison with State-of-the-Art methods on Adver-City-R (Combined split). We report AP@0.5 and AP@0.7 for LiDAR-only, radar-camera, and LiDAR-radar baselines. Our Where2comm-LRA achieves the highest overall robustness.

Method	Modality	AP@0.5	AP@0.7
Scope [38]	LiDAR	19.60	14.40
AttFuse [35]	LiDAR	24.00	16.00
S-AdaFusion [25]	LiDAR	38.30	30.30
SlimComm [43]	LiDAR-radar	58.00	50.74
BM2CP-RCA	radar-camera	55.36	40.10
AttFuse w/MDD [9]	LiDAR-radar	55.57	47.04
Where2comm-LRA	LiDAR-radar	**64.70**	**51.12**

To contextualize our results, Tab. 2 compares our radar-augmented models with representative LiDAR-based cooperative perception baselines, including Scope [38], AttFuse [35], and S-AdaFusion [25], as well as LiDAR-radar baselines such as AttFuse w/MDD [9] and SlimComm [43] on the Adver-City-R benchmark (Combined split). Crucially, while SlimComm inherently supports 360° perception, we retrained and evaluated AttFuse w/MDD on Adver-City-R with the same 360° coverage and continuous Doppler velocity data used by our models. The LiDAR-only methods achieve moderate accuracy but degrade substantially in adverse weather. By contrast, our radar-based

Table 3. Zero-shot generalization on Adver-City-R (AP@0.5, %). Scores are reported for *Sparse* (S) and *Dense* (D) splits. Abbreviations: BM2CP (LiDAR-camera), RC (radar-camera), RCA (RC + mask-guided spatial attention), Where2comm (LiDAR-only), LR (LiDAR-radar), LRA (LR + mask-guided spatial attention). **Bold** = best, underline = second-best; † = best within BM2CP family; ‡ = best within Where2comm family; ⋆ = best within the respective S or D split across all models.

Weather	BM2CP [46] S/D	BM2CP-RC S/D	BM2CP-RCA S/D	Where2comm [7] S/D	Where2comm-LR S/D	Where2comm-LRA S/D
Clear	67.37† / 61.97†	60.64 / 47.18	64.81 / 54.50	69.54 / 69.15	72.51 / 70.38	**74.79**‡⋆ / **70.91**‡⋆
Soft Rain	64.39† / 59.14†	60.12 / 46.33	63.92 / 53.20	65.47 / 63.31	71.63 / 67.00	**73.81**‡⋆ / **67.97**‡⋆
Heavy Rain	65.27† / 61.06†	59.11 / 46.49	63.91 / 54.39	63.51 / 61.54	72.23 / 64.97	**75.20**‡⋆ / **68.89**‡⋆
Fog	32.97 / 34.32	60.84 / 45.55	**65.12**†⋆ / 53.37†	36.00 / 35.77	63.07 / 53.11	64.93‡ / **56.80**‡⋆
Fog+Rain	28.05 / 29.10	59.72 / 46.01	**65.00**†⋆ / 52.03†	33.60 / 31.34	60.50 / 51.61	63.12‡ / **54.00**‡⋆

variants close this gap, with RCA already outperforming all LiDAR-only baselines, and our LiDAR-radar LRA model further exceeding AttFuse w/MDD while delivering the most robust overall results among the compared methods. These results highlight radar's unique role in enabling reliable cooperative perception beyond LiDAR-only pipelines and show that our mask-guided Doppler attention consistently improves performance under adverse weather conditions. Furthermore, our Where2comm-LRA demonstrates superior computational efficiency with an inference time of 58 ms, nearly twice as fast as the AttFuse w/MDD baseline (108 ms). This shows that our Doppler-guided attention is a lightweight alternative to high-overhead diffusion-based denoising.

4.6. Robustness to Spatial and Temporal Misalignment

The framework is evaluated under simulated spatial and temporal misalignments on the Adver-City-R combined fog and heavy rain split based on protocol in AttFuse w/MDD [9]. Proposed RCA and LRA models are compared against SlimComm [43] and AttFuse w/MDD [9].

As shown in Fig. 6 (Left), all models experience performance degradation as pose and localization error increase to 0.6°/0.6 m. Where2comm-LRA maintains the highest absolute AP@0.5 throughout the entire error range. While SlimComm exhibits high relative stability indicated by a flatter slope, it consistently performs at a lower baseline. In contrast, BM2CP-RCA provides a superior balance, maintaining a significantly higher AP than both SlimComm and AttFuse w/MDD while exhibiting comparable stability to increasing pose noise.

The impact of communication latency, illustrated in Fig. 6 (Right), reveals a critical shift in model ranking. While Where2comm-LRA is superior under ideal synchronization (0 ms), it is highly sensitive to temporal shifts. SlimComm proves to be the most resilient model with a relatively stable performance curve, likely due to its sparse query mechanism, which is less dependent on perfect temporal feature alignment. However, BM2CP-RCA emerges as the most effective solution for high-latency environments. It crosses over both the LRA and AttFuse w/MDD models at the 200 ms mark, maintaining the highest AP@0.5 for all delays between 200 ms and 600 ms. This suggests that the Doppler-guided attention in RCA effectively mitigates ghosting artifacts by using instantaneous velocity cues to anchor features, even when the received messages are significantly delayed.

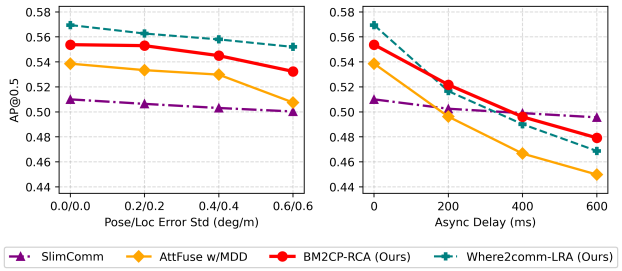

Figure 6. Robustness analysis on Adver-City-R (Combined Fog and Heavy Rain). (Left) Impact of increasing pose and localization error. (Right) Impact of asynchronous communication delay.

4.7. Ablation Study

To better understand the contribution of each component in our radar-camera pipeline, we performed an ablation analysis. As shown in Tab. 4, mask-guided spatial attention already improves over plain RC, velocity conditioning yields an additional gain, and temporal aggregation provides the strongest boost. This confirms that Doppler-based history compensation is crucial for robustness under adverse weather.

4.8. Real-World Single-Vehicle Evaluation

Motivation. Adver-City-R and OPV2V-R quantify robustness under controlled, cooperative settings with weather degradation. To test real-world applicability beyond simulation, we additionally evaluate our radar encoding and Doppler-guided modules on the real MAN TruckScenes

Table 4. The table shows the contribution of each module: **Vel** (velocity-conditioned PillarVFE), **A** (mask-guided spatial attention), and **Temp** (Doppler-compensated multi-frame aggregation).

Variant	Vel	A	Temp	AP@0.5
LC (LiDAR-camera baseline)				48.86
RC (radar-camera)				47.34
RC + A		✓		52.80
RC + Vel + A	✓	✓		53.15
RCA (Full Model)	✓	✓	✓	**55.36**

dataset [3]. Since MAN TruckScenes does not include cooperative labels, we use a single-vehicle PointPillars-based [14] backbone for this study.

Setup. MAN TruckScenes follows a nuScenes-like structure. We use all six radar and six LiDAR sensors. Ground truth is limited to the "car" class and includes only objects hit by at least one LiDAR or radar point. We select two subsets of the dataset (approximately 3960 frames), filter scenes to highway, city, residential, and rural areas, and split 80/10/10 with a fixed seed. Following the simulation protocol, all models are trained on clear-weather data and evaluated zero-shot on the foggy subset. Perception range and PointPillars settings are aligned with nuScenes. For the evaluation, we used mAP-based results to align with Secs. 4.4.1 and 4.4.2. All experiments were conducted on an NVIDIA RTX 2080 Ti GPU. Latency is reported as the mean model-only forward pass latency, averaged over 100 iterations.

Table 5. MAN TruckScenes (single-vehicle, car class only) with a PointPillars-based (PP) backbone [14]. The table shows the contribution of each module: **Vel** (velocity-conditioned PillarVFE), **A** (mask-guided spatial attention), and **Temp** (Doppler-compensated multi-frame aggregation). Results are AP@0.5 for Clear/Overcast and Foggy subsets.

Model Variant	Clear/Overcast	Foggy	Latency (ms)
PP-L (LiDAR-only)	17.70	25.82	14.0
PP-LR (LiDAR-radar)	17.40	24.52	14.7
PP-LR + A	20.15	30.98	16.9
PP-LR + A + Vel	20.15	31.92	17.0
PP-LR + A + Vel + Temp	**21.20**	**33.80**	17.1

Findings. Tab. 5 shows that naive LiDAR-radar fusion (PP-LR) does not improve over LiDAR-only (PP-L), and in fact performs slightly worse in both Clear/Overcast and Foggy conditions. This indicates that raw radar features alone are not directly beneficial in the real-world dataset. In contrast, adding our mask-guided spatial attention (+A) yields consistent gains (+2.45 AP on Clear/Overcast, +5.16 AP on Foggy) while increasing latency by only $\sim$2 ms. Velocity conditioning (+Vel) further improves robustness, reaching with just $\sim$3 ms overhead compared to the LiDAR-only baseline. Temporal aggregation (+Temp) provides the highest overall accuracy (21.20 AP on Clear/Overcast, 33.80 AP on Foggy), resulting in a total improvement of +8.0 points over the LiDAR-only baseline in Foggy conditions while adding only $\sim$3 ms latency.

Note that, unlike Adver-City-R, where weather is applied to identical scenarios, MAN TruckScenes does not control for scene type. The Foggy split contains only highway scenes, which have fewer occlusions and targets, leading to higher LiDAR AP than in Clear/Overcast. This reflects dataset composition rather than LiDAR robustness to fog; the relevant finding is the consistent gain from our Doppler-guided modules.

5. Limitations and Future Work

While our method shows strong robustness under adverse weather, several limitations remain. Most experiments are conducted on radar-augmented simulation benchmarks, and the real-world validation on MAN TruckScenes is limited to a single-vehicle setup with car-only annotations. In addition, we do not explicitly model radar-specific artifacts such as multipath reflections or sensor-dependent noise. Future research will explore backbone-aware attention and advanced temporal aggregation [5] to better handle real-world radar sparsity and noise. We also plan to extend our cooperative benchmarks to more challenging conditions, including snow, roadside infrastructure, and Vulnerable Road Users (VRUs), and to study radar motion cues for downstream tasks such as collaborative tracking and trajectory prediction.

6. Conclusion

We addressed adverse-weather cooperative perception by integrating 4D radar into BM2CP and Where2comm. Our Doppler-guided spatial attention improves robustness to degraded LiDAR and camera observations. Experiments on OPV2V-R and Adver-City-R show strong gains under fog and rain, and MAN TruckScenes confirms transfer beyond simulation. The proposed LRA model delivers robust and efficient all-weather cooperative perception.

Acknowledgements. This paper was created in the Country 2 City - Bridge project of the German Center for Future Mobility, which is funded by the German Federal Ministry for Digital and Transport.

References

[1] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yuning Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of*

the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 11618–11628, 2020. 2

[2] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. CARLA: An open urban driving simulator. In *Proceedings of the 1st Annual Conference on Robot Learning*, pages 1–16, 2017. 1, 5

[3] Felix Fent, Fabian Kuttenreich, Florian Ruch, Farija Rizwin, Stefan Juergens, Lorenz Lechermann, Christian Nissler, Andrea Perl, Ulrich Voll, Min Yan, and Markus Lienkamp. Man truckscenes: A multimodal dataset for autonomous trucking in diverse conditions, 2024. 2, 5, 8

[4] Xiangyu Gao, Guanbin Xing, Sumit Roy, and Hui Liu. Ramp-cnn: A novel neural network for enhanced automotive radar object recognition. *IEEE Sensors Journal*, 21(4): 5119–5132, 2021. 2

[5] Yuval Haitman and Oded Bialer. Doppdrive: Doppler-driven temporal aggregation for improved radar object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 26085–26094, 2025. 8

[6] Adam W Harley, Zhaoyuan Fang, Jie Li, Rares Ambrus, and Katerina Fragkiadaki. Simple-bev: What really matters for multi-sensor bev perception? In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2759–2765. IEEE, 2023. 1

[7] Yue Hu, Shaoheng Fang, Zixing Lei, Yiqi Zhong, and Siheng Chen. Where2comm: Communication-efficient collaborative perception via spatial confidence maps. *Advances in neural information processing systems*, 35:4874–4886, 2022. 3, 6, 7

[8] Tao Huang, Jianan Liu, Xi Zhou, Dinh C. Nguyen, Mostafa Rahimi Azghadi, Yuxuan Xia, Qing-Long Han, and Sumei Sun. V2x cooperative perception for autonomous driving: Recent advances and challenges, 2024. 1

[9] 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)*, pages 27390–27400, 2025. 1, 2, 6, 7

[10] Ping Jiang, Xiaoheng Deng, Weishang Wu, Lixin Lin, Xuechen Chen, Chen Chen, and Shaohua Wan. Weather-aware collaborative perception with uncertainty reduction. *IEEE Transactions on Intelligent Transportation Systems*, 25 (12):20059–20072, 2024. 2

[11] Mateus Karvat and Sidney Givigi. Adver-city: Open-source multi-modal dataset for collaborative perception under adverse weather conditions, 2025. 1, 2, 5

[12] Velat Kilic, Deepti Hegde, Vishwanath Sindagi, A Brinton Cooper, Mark A Foster, and Vishal M Patel. Lidar light scattering augmentation (lisa): Physics-based simulation of adverse weather conditions for 3d object detection. *arXiv preprint arXiv:2107.07004*, 2021. 1, 5

[13] Youngseok Kim, Sanmin Kim, Jun Won Choi, and Dongsuk Kum. Craft: Camera-radar 3d object detection with spatio-contextual fusion transformer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1160–1168, 2023. 2

[14] Alex H. Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. Pointpillars: Fast encoders for object detection from point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 8

[15] Baolu Li, Jinlong Li, Xinyu Liu, Runsheng Xu, Zhengzhong Tu, Jiacheng Guo, Xiaopeng Li, and Hongkai Yu. V2x-dgw: Domain generalization for multi-agent perception under adverse weather conditions, 2025. 2

[16] Henry Liu, Zhong Cao, Xintao Yan, Shuo Feng, and Qiujing Lu. Autonomous vehicles: A critical review (2004-2024) and a vision for the future. 2025. 1

[17] Hongsi Liu, Jun Liu, Guangfeng Jiang, and Xin Jin. Mssf: A 4d radar and camera fusion framework with multi-stage sampling for 3d object detection in autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, 26(6): 8641–8656, 2025. 2

[18] Mingyu Liu, Ekim Yurtsever, Jonathan Fossaert, Xingcheng Zhou, Walter Zimmer, Yuning Cui, Bare Luka Zagar, and Alois C. Knoll. A survey on autonomous driving datasets: Statistics, annotation quality, and a future outlook, 2024. 1

[19] Alexander Musiat, Laurenz Reichardt, Michael Schulze, and Oliver Wasenmüller. Radarpillars: Efficient object detection from 4d radar point clouds. In *2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC)*, pages 1656–1663, 2024. 2, 3

[20] Ramin Nabati and Hairong Qi. Centerfusion: Center-based radar and camera fusion for 3d object detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1527–1536, 2021. 2

[21] Felix Nobis, Maximilian Geisslinger, Markus Weber, Johannes Betz, and Markus Lienkamp. A deep learning-based radar and camera sensor fusion architecture for object detection. In *2019 Symposium on Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, pages 1–7, 2019. 2

[22] 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. 2

[23] Andras Palffy, Ewoud Pool, Srimannarayana Baratam, Julian F. P. Kooij, and Dariu M. Gavrila. Multi-class road user detection with 3+1d radar in the view-of-delft dataset. *IEEE Robotics and Automation Letters*, 7(2):4961–4968, 2022. 6

[24] Xiangyuan Peng, Miao Tang, Huawei Sun, Kay Bierzynski, Lorenzo Servadei, and Robert Wille. 4d mmwave radar for sensing enhancement in adverse environments: Advances and challenges, 2025. 1

[25] Donghao Qiao and Farhana Zulkernine. Adaptive feature fusion for cooperative perception using lidar point clouds. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 1186–1195, 2023. 2, 6

[26] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, Vijay Vasudevan, Wei Han, Jiquan Ngiam, Hang Zhao, Aleksei Timofeev, Scott Ettinger, Maxim Krivokon, Amy Gao, Aditya Joshi, Sheng Zhao, Shuyang Cheng, Yu Zhang, Jonathon Shlens, Zhifeng Chen,

and Dragomir Anguelov. Scalability in perception for autonomous driving: Waymo open dataset, 2020. 1

[27] Sven Teufel, Georg Volk, Alexander Von Bernuth, and Oliver Bringmann. Simulating realistic rain, snow, and fog variations for comprehensive performance characterization of lidar perception. In *2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring)*, pages 1–7, 2022. 2

[28] Dimitra Tsakmakopoulou and Konstantinos Moustakas. Perception for connected autonomous vehicles under adverse weather conditions. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3161–3166, 2024. 2

[29] Jizhao Wang, Zhizhou Wu, Yunyi Liang, Jinjun Tang, and Huimiao Chen. Perception methods for adverse weather based on vehicle infrastructure cooperation system: A review. *Sensors*, 24(2), 2024. 2

[30] Li Wang, Xinyu Zhang, Baowei Xv, Jinzhao Zhang, Rong Fu, Xiaoyu Wang, Lei Zhu, Haibing Ren, Pingping Lu, Jun Li, and Huaping Liu. Interfusion: Interaction-based 4d radar and lidar fusion for 3d object detection. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 12247–12253, 2022. 2

[31] Naibang Wang, Deyong Shang, Yan Gong, Xiaoxi Hu, Ziying Song, Lei Yang, Yuhan Huang, Xiaoyu Wang, and Jianli Lu. Collaborative perception datasets for autonomous driving: A review. *IEEE Sensors Journal*, 2025. 2

[32] Zhangjing Wang, Yu Wu, and Qingqing Niu. Multi-sensor fusion in automated driving: A survey. *IEEE Access*, 8: 2847–2868, 2020. 1

[33] Xiongfei Wu, Mingfei Cheng, Qiang Hu, Jianlang Chen, Yuheng Huang, Manabu Okada, Michio Hayashi, Tomoyuki Tsuchiya, Xiaofei Xie, and Lei Ma. Foundation models for autonomous driving system: An initial roadmap, 2025. 1

[34] Weiyi Xiong, Jianan Liu, Tao Huang, Qing-Long Han, Yuxuan Xia, and Bing Zhu. Lxl: Lidar excluded lean 3d object detection with 4d imaging radar and camera fusion. *IEEE Transactions on Intelligent Vehicles*, 9(1):79–92, 2024. 2

[35] Runsheng Xu, Hao Xiang, Xin Xia, Xu Han, Jinlong Li, and Jiaqi 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)*, pages 2583–2589. IEEE, 2022. 1, 2, 6

[36] Bin Yang, Runsheng Guo, Ming Liang, Sergio Casas, and Raquel Urtasun. Radarnet: Exploiting radar for robust perception of dynamic objects. In *European conference on computer vision*, pages 496–512. Springer, 2020. 3

[37] Boquan Yang, Jixiong Li, and Ting Zeng. A review of environmental perception technology based on multi-sensor information fusion in autonomous driving. *World Electric Vehicle Journal*, 16(1), 2025. 1

[38] Kun Yang, Dingkang Yang, Jingyu Zhang, Mingcheng Li, Yang Liu, Jing Liu, Hanqi Wang, Peng Sun, and Liang Song. Spatio-temporal domain awareness for multi-agent collaborative perception. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 23383–23392, 2023. 6

[39] Lei Yang, Xinyu Zhang, Jun Li, Chen Wang, Jiaqi Ma, Zhiying Song, Tong Zhao, Ziying Song, Li Wang, Mo Zhou, Yang Shen, and Chen Lv. V2x-radar: A multi-modal dataset with 4d radar for cooperative perception. *Advances in Neural Information Processing Systems (NeurIPS)*, 2025. 2

[40] Shanliang Yao, Runwei Guan, Xiaoyu Huang, Zhuoxiao Li, Xiangyu Sha, Yong Yue, Eng Gee Lim, Hyungjoon Seo, Ka Lok Man, Xiaohui Zhu, and Yutao Yue. Radar-camera fusion for object detection and semantic segmentation in autonomous driving: A comprehensive review. *IEEE Transactions on Intelligent Vehicles*, 9(1):2094–2128, 2024. 1

[41] Melih Yazgan, Mythra Varun Akkanapragada, and J. Marius Zöllner. Collaborative perception datasets in autonomous driving: A survey. In *2024 IEEE Intelligent Vehicles Symposium (IV)*, pages 2269–2276, 2024. 2

[42] Melih Yazgan, Thomas Graf, Min Liu, Tobias Fleck, and J Marius Zöllner. Real-world problems in collaborative perception: A categorized review of intermediate fusion methods. *IEEE IV*, 2024. 1

[43] Melih Yazgan, Qiyuan Wu, Iramm Hamdard, Shiqi Li, and J Marius Zoellner. Slimcomm: Doppler-guided sparse queries for bandwidth-efficient cooperative 3-d perception. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1782–1791, 2025. 2, 3, 6, 7

[44] Haibao Yu, Yizhen Luo, Mao Shu, Yiyi Huo, Zebang Yang, Yifeng Shi, Zhenglong Guo, Hanyu Li, Xing Hu, Jirui Yuan, and Zaiqing Nie. Dair-v2x: A large-scale dataset for vehicle-infrastructure cooperative 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21361–21370, 2022. 1, 2

[45] Yuxiao Zhang, Alexander Carballo, Hanting Yang, and Kazuya Takeda. Perception and sensing for autonomous vehicles under adverse weather conditions: A survey. *ISPRS Journal of Photogrammetry and Remote Sensing*, 196:146–177, 2023. 1, 2, 5

[46] Binyu Zhao, Wei ZHANG, and Zhaonian Zou. Bm2cp: Efficient collaborative perception with lidar-camera modalities. In *Conference on Robot Learning*, pages 1022–1035. PMLR, 2023. 3, 6, 7

[47] Lianqing Zheng, Sen Li, Bin Tan, Long Yang, Sihan Chen, Libo Huang, Jie Bai, Xichan Zhu, and Zhixiong Ma. Rcfusion: Fusing 4-d radar and camera with bird's-eye view features for 3-d object detection. *IEEE Transactions on Instrumentation and Measurement*, 72:1–14, 2023. 2

[48] Walter Zimmer, Gerhard Arya Wardana, Suren Sritharan, Xingcheng Zhou, Rui Song, and Alois C. Knoll. Tumtraf v2x cooperative perception dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 22668–22677, 2024. 2