RCP-Bench: Benchmarking Robustness for Collaborative Perception Under Diverse Corruptions

Shihang Du[1], Sanqing Qu[1], Tianhang Wang[1], Xudong Zhang[1],
Yunwei Zhu[1], Jian Mao[1], Fan Lu[1], Qiao Lin[2], Guang Chen[1]*
[1]Tongji University, [2]EACON Technology Co., ltd.

Abstract

Collaborative perception enhances single-vehicle perception by integrating sensory data from multiple connected vehicles. However, existing studies often assume ideal conditions, overlooking resilience to real-world challenges, such as adverse weather and sensor malfunctions, which is critical for safe deployment. To address this gap, we introduce RCP-Bench, the first comprehensive benchmark designed to evaluate the robustness of collaborative detection models under a wide range of real-world corruptions. RCP-Bench includes three new datasets (i.e., OPV2V-C, V2XSet-C, and DAIR-V2X-C) that simulate six collaborative cases and 14 types of camera corruption resulting from external environmental factors, sensor failures, and temporal misalignments. Extensive experiments on 10 leading collaborative perception models reveal that, while these models perform well under ideal conditions, they are significantly affected by corruptions. To improve robustness, we propose two simple yet effective strategies, RCP-Drop and RCP-Mix, based on training regularization and feature augmentation. Additionally, we identify several critical factors influencing robustness, such as backbone architecture, camera number, feature fusion methods, and the number of connected vehicles. We hope that RCP-Bench, along with these strategies and insights, will stimulate future research toward developing more robust collaborative perception models. Our benchmark toolkit is available at https://github.com/LuckyDush/RCP-Bench.

1. Introduction

Perception is a fundamental capability for robots and autonomous vehicles to accurately interpret their surroundings. Despite significant advancements in perception tasks through deep learning, including object detection [3, 68] and segmentation [19, 27], single-vehicle perception using onboard sensors still faces inherent limitations, such as restricted range, limited field of view, and vulnerability to occlusions [17, 30, 50], resulting in blind spots that hinder full environmental understanding [5]. Collaborative perception, where perception data are shared among Connected Autonomous Vehicles (CAVs), has shown promise in overcoming these challenges. By leveraging multiple viewpoints, collaborative perception extends perceptual range, enhances spatial coverage, and improves resilience to occlusions [16].

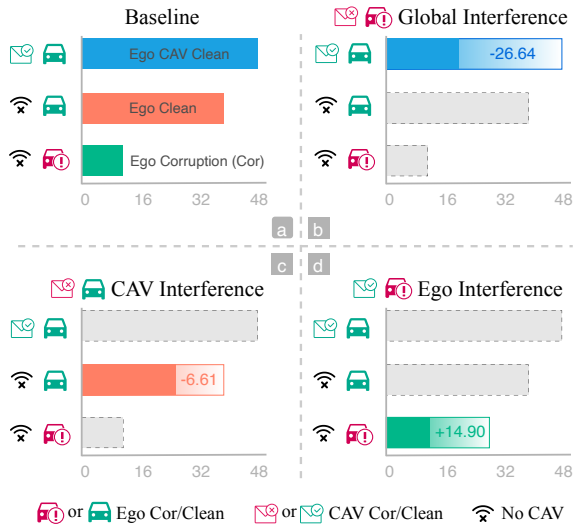


Figure 1. Comparison of collaborative object detection results (averaged across 10 leading methods) in OPV2V-C across three interference scenarios: Global, Ego, and CAV Interference. The results indicate that perception models are particularly vulnerable to Global Interference, benefit from collaborative compensation under Ego Interference, and experience performance degradation under CAV Interference.

Numerous studies have investigated practical challenges that may impact collaborative performance, such as bandwidth limitations for shared information [15, 32, 33, 44], transmission issues [26, 38, 47, 60], pose estimation errors [24, 34, 43, 66], and heterogeneity among cooperating vehicles [28, 48, 55]. However, current models and evaluations generally assume idealized conditions, such as clear weather and fully functional sensors. In real-world driving scenarios, adverse conditions such as external weather [9, 10, 18] and sensor failures [2, 12] are unavoidable. Although recent efforts address individual robustness

*Corresponding author: guangchen@tongji.edu.cn

challenges, including communication delays [23, 36] and diverse weather conditions [14, 25], a comprehensive assessment of robustness for collaborative perception in real-world scenarios remains an important open challenge.

To bridge this gap, we introduce RCP-Bench, the first benchmark designed to systematically evaluate the reliability of collaborative perception methods under a wide range of corruptions. Specifically, RCP-Bench comprises three new datasets, including OPV2V-C, V2XSet-C and DAIR-V2X-C. These datasets simulate six distinct cases and incorporate 14 types of corruption, incorporating variations in weather conditions, sensor noise or failures, and temporal misalignments, as depicted in Fig. 2. Given that collaborative perception depends on data shared across collaborating CAVs, where corruptions in individual vehicles can propagate throughout the collaborative process, creating diverse and challenging scenarios. RCP-Bench provides a holistic assessment by evaluating robustness across three aspects:

- Global Interference: Both the ego vehicle and collaborating CAVs are affected by diverse corruptions, testing the overall robustness of collaborative perception.
- Ego Interference: Only the ego vehicle experiences corruption, assessing the compensatory advantages of collaboration over single-vehicle perception.
- CAV Interference: Only the collaborating CAVs are subject to corruptions, evaluating the risk of disruptions compared to single-vehicle perception.

Leveraging RCP-Bench, we conduct extensive experiments on 10 collaborative perception models. As shown in Fig. 1, while these models perform well under ideal conditions, their performance significantly degrades in the presence of corruptions. To enhance the robustness, we propose two straightforward strategies tailored for collaborative perception: RCP-Drop and RCP-Mix. Unlike traditional methods [1, 8] that drop features within specific network layers or blocks during training, RCP-Drop operates by randomly discarding data from selected collaborating vehicles, effectively simulating real-world data loss encountered in communication-limited or sensor-failure scenarios. RCP-Mix, inspired by MixStyle [64, 65], probabilistically combines feature statistics, i.e., the means and standard deviations of feature maps, between the ego vehicle and collaborating CAVs, promoting resilience to distribution shifts.

Our contributions can be summarized as follows:

- We introduce RCP-Bench, the first benchmark systematically evaluating collaborative perception robustness across diverse real-world corruptions.
- We conduct a comprehensive evaluation of 10 leading collaborative perception models across six scenarios, covering 14 corruption types.
- We propose two novel strategies to enhance the robustness of collaborative perception models and highlight key factors affecting resilience.

2. Related Work

Collaborative Perception: Collaborative perception [5, 6, 16, 34, 40, 54, 56, 58, 61, 62] allows vehicles to share perceptual data with connected vehicles, realizing a more comprehensive understanding of the surrounding environment. Existing studies have proposed various techniques to enhance perception performance [42]. For example, Who2com [33] and When2com [32] propose strategies for selecting collaborators and determining the timing of collaboration. Where2comm [15] and What2comm [57] introduce methods for selecting relevant information to be shared. UMC [44] enhances spatiotemporal continuity in collaborative perception by leveraging historical data and multi-scale feature fusion. CoBEVT [51] enables real-time collaborative semantic segmentation by capturing sparse local and global spatial interactions across views and agents. Nevertheless, most of these methods assume ideal conditions, addressing robustness under diverse and challenging conditions remains essential to advancing collaborative perception reliability in practice.

Robustness against Corruptions: Perception models are often vulnerable to various sensor corruptions, which can compromise their effectiveness in real-world applications. To evaluate and improve robustness under such conditions, several benchmarks have been developed. ImageNet-C [12] is the first public benchmark to assess object recognition models under a range of corruptions, including noise, blur, adverse weather, and digital interference. For object detection, Michaelis et al. [35] introduce three benchmarks to examine robustness against similar impairments. Recently, benchmarks tailored to driving scenarios have emerged. RoboDepth [20] assesses robustness in monocular depth estimation under corrupted conditions, while RoboBEV [49] introduces a comprehensive benchmark for evaluating BEV perception tasks. Similarly, MapBench [11] is designed to test the robustness of high-definition mapping methods against various sensor corruptions. Although these benchmarks have advanced robustness assessments in single-vehicle perception, they do not account for multi-agent contexts where corruptions may propagate, compounding effects across interconnected systems. Developing collaborative perception specific benchmarks is crucial for evaluating and improving robustness in these multi-vehicle scenarios.

Robustness in Collaborative Perception: With the advancement of collaborative perception research [13, 31, 37, 45], increasing attention is being paid to robustness and safety, especially concerning communication challenges between cooperating agents. These challenges include issues in data transmission, such as information loss during communication [26] and communication interruptions [38], timing issues like delays [23, 47, 60] and the reception of outdated or missing collaborative information. While these studies have highlighted potential pitfalls within the collab-

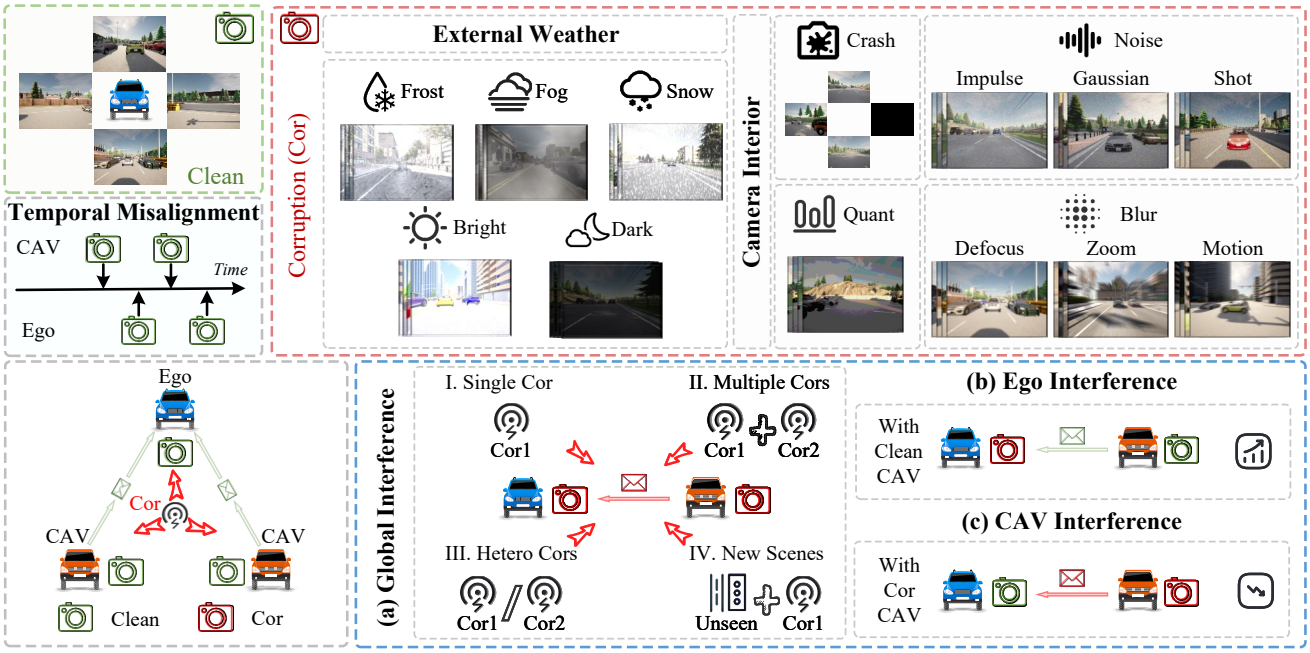


Figure 2. Definitions of corruption types and evaluation scenarios in RCP-Bench. Our benchmark includes 14 types of camera corruptions, classified into three categories: *External Weather*, *Camera Interior*, and *Temporal Misalignment*. To comprehensively evaluate robustness, we define three interference scenarios: *Global Interference* (corruptions affecting both the ego vehicle and collaborative vehicles), *Ego Interference* (corruptions specific to the ego vehicle), and *CAV Interference* (corruptions isolated to one or more collaborative vehicles). Within the Global Interference scenario, we further examine four cases: *Single Corruption*, *Multiple Corruptions*, *Heterogeneous Corruptions*, and *New Scenes with Corruption*.

orative process itself, robustness under diverse environmental and sensor conditions remains underexplored. Recently, a few studies have begun to examine collaborative performance under adverse environmental conditions [14, 25]. Nevertheless, these efforts are primarily limited to extreme sensor failures, without systematically analyzing the benefits and resilience of collaboration across a wider range of realistic sensor corruptions. To the best of our knowledge, this paper is the first to comprehensively examine the robustness and advantages of different collaborative states under diverse sensor corruptions, thereby providing an empirical basis for assessing and improving collaborative perception robustness across various real-world challenges.

3. RCP-Bench

3.1. Benchmark Design

Collaborative perception relies on information collected by both the ego vehicle and surrounding connected autonomous vehicles (CAVs), each contributing to the final perception outcome. To rigorously evaluate the benefits of shared information and assess model robustness under various corruptions, this study isolates these two information sources and defines three corruption scenarios, as shown in Fig. 2: (1) Global Interference scenario affecting both the ego vehicle and CAVs simultaneously, (2) Ego Interference scenario specific to the ego vehicle, and (3) CAV Interference scenario isolated to one or more CAVs.

Given mutual interference among CAVs, independent control of lidar-based corruptions is impractical; therefore, this study focuses exclusively on camera-based corruptions. We categorize these into three groups based on real-world conditions: (1) External environmental factors, like rain, fog, or light; (2) Internal camera issues, such as lens scratches or sensor degradation; (3) Misalignment in capture timing, such as temporal desynchronization between the ego and CAV cameras. We identify 14 distinct types of corruption, following the methodology in [7], and divide each type into five severity levels, resulting in 70 unique corruption conditions. To evaluate collaborative perception model robustness, we create three corruption benchmarks using widely used datasets: DAIR-V2X [59], OPV2V [53], and V2XSet [52]. These corruptions are introduced into the validation sets, producing modified versions: DAIR-V2X-C, OPV2V-C, and V2XSet-C.

3.2. Typical Corruption Scenarios

Global Interference: In practical applications, it is common for corruptions to affect both the ego vehicle and collaborative vehicles simultaneously. To evaluate model robustness across various types of corruption, we first assess 14 distinct corruptions individually, referring to this as the *Single Corruption* case. To better reflect real-world complexity, we further define three combined corruption cases. The first, *Multiple Corruptions*, involves multiple types of corruptions occurring concurrently. For simplicity, we focus on situations where temporal misalignment coincides with either external weather conditions or internal camera

issues. In the second case, *Heterogeneous Corruptions*, the types of corruption affecting the ego vehicle differ from those impacting the collaborative vehicles, allowing us to examine situations with varying corruption profiles across vehicles. Finally, in the *New Scenes with Corruption* case, we extend prior research on the generalization of collaborative perception by analyzing performance when corruptions are introduced simultaneously in unfamiliar environments.

Ego Interference: In real-world scenarios, the ego vehicle may experience unique corruptions that do not affect nearby collaborative vehicles, presenting an opportunity to use uncorrupted data from these vehicles as a compensatory mechanism. To explore this potential, we introduce the concept of "collaborative compensation." Here, when the ego vehicle encounters interference or corruption, we investigate whether reliable information from neighboring collaborative vehicles can help mitigate or even correct these effects. Since some forms of interference, such as adverse weather, are unavoidable and impact all vehicles, we specifically focus on evaluating the advantages of collaborative compensation for the remaining ten types of corruption that are unique to the ego vehicle.

CAV Interference: Similar to Ego Interference, collaborative vehicles may experience unique corruptions that do not directly affect the ego vehicle, creating a potential risk rather than a benefit. To examine this, we introduce the concept of "collaborative disruption." In this scenario, we investigate whether corrupted data from surrounding collaborative vehicles can degrade the perception performance of an otherwise well-functioning ego vehicle. This analysis provides insights into the robustness of collaborative perception systems, highlighting the potential adverse effects of relying on collaborative data when unintentional corruptions are present in the shared information.

3.3. Evaluation Metrics

In evaluating 3D object detection performance, Average Precision (AP) is commonly used as the primary metric, particularly at Intersection-over-Union (IoU) thresholds of 0.3, 0.5, and 0.7. Building on the evaluation metrics proposed in [7, 41, 67], we introduce two metrics specifically designed to assess the robustness of cooperative perception systems: Corrupted Average Precision ($\mathrm{AP}_{\mathrm{cor}}$) and Relative Corruption Error (RCE). To further examine the effects of collaboration under Ego Interference and CAV Interference, we define "Positive Collaborative" and "Negative Collaborative", and introduce two additional metrics: the Positive Collaborative Coefficient (PosC) and the Negative Collaborative Coefficient (NegC). These four new evaluation metrics are primarily analyzed based on AP@0.5, additional results of AP@0.3 and AP@0.7 are provided in the appendix.

Corrupted Average Precision: We denote the performance on the clean, uncorrupted test set as $\mathrm{AP}_{\mathrm{clean}}$. For each corruption type c at each severity l, the performance is measured as $\mathrm{AP}_{c,l}$. The corruption robustness, AP_{cor} is then defined as the average performance across all corruption types and severity levels:

$$\mathrm{AP}_c = \frac{1}{5}\sum_{l=1}^{5}\mathrm{AP}_{c,l}\ ,\ \mathrm{AP}_{cor} = \frac{1}{|\mathcal{C}|}\sum_{c\in\mathcal{C}}\mathrm{AP}_c\ , \tag{1}$$

where $\mathcal{C}$ is the set of corruptions in evaluation.

Relative Corruption Error: Since $\mathrm{AP}_{\mathrm{cor}}$ measures only the absolute performance under corrupted conditions, we define Relative Corruption Error (RCE) as a relative indicator that quantifies the extent to which model performance is retained under corruption:

$$\mathrm{RCE}_c = \frac{\mathrm{AP}_{clean} - \mathrm{AP}_c}{\mathrm{AP}_{clean}}\ ,\ \mathrm{mRCE} = \frac{1}{|\mathcal{C}|}\sum_{c\in\mathcal{C}}\mathrm{RCE}_c\ . \tag{2}$$

Positive Collaborative Coefficient: To quantify the ability of collaborative information to mitigate corruptions affecting the ego vehicle under Ego Interference, we define the Positive Collaborative Coefficient (PosC), which measures how effectively the model leverages collaborative information from other vehicles to counteract corruptions:

$$\mathrm{PosC}_c = \frac{\mathrm{AP}_{c,5} - \mathrm{AP}^{ego}_{c,5}}{1 - \mathrm{AP}^{ego}_{c,5}}\ ,\ \mathrm{mPosC} = \frac{1}{|\mathcal{C}|}\sum_{c\in\mathcal{C}}\mathrm{PosC}_c\ , \tag{3}$$

where $\mathrm{AP}^{\mathrm{ego}}_{\mathrm{c,5}}$ represents the performance of the ego vehicle under a specific type of corruption at severity 5 in the absence of collaboration.

Negative Collaborative Coefficient: To assess the adverse impact of corrupted collaborative information under CAV Interference, we introduce the Negative Collaborative Coefficient (NegC), which quantifies the negative effects arising from collaborative perception:

$$\mathrm{NegC}_c = \frac{1 - \mathrm{AP}_{c,5}}{1 - \mathrm{AP}^{ego}_{clean}}\ ,\ \mathrm{mNegC} = \frac{1}{|\mathcal{C}|}\sum_{c\in\mathcal{C}}\mathrm{NegC}_c\ , \tag{4}$$

where $\mathrm{AP}^{\mathrm{ego}}_{\mathrm{clean}}$ denotes the perception performance of the ego vehicle using only its own clean, uncorrupted data.

4. Benchmarking Results

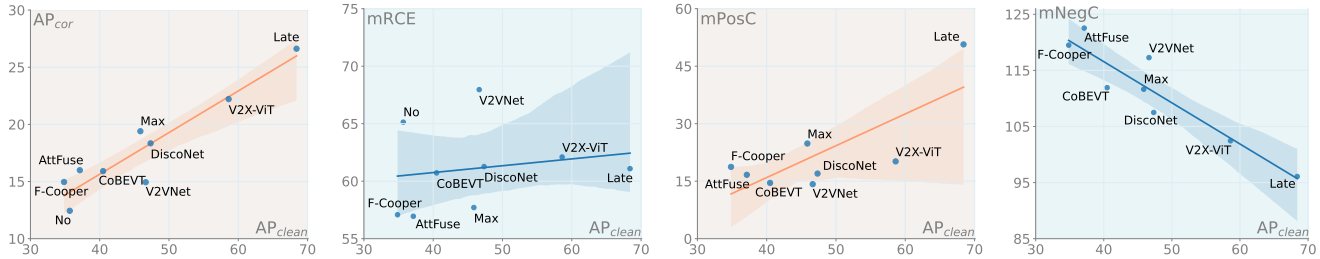
In this section, we primarily present the results on OPV2V-C. Additional results for DAIR-V2X-C and V2XSet-C are provided in the appendix.

4.1. Benchmark Setup

Our RCP-Bench evaluates a total of 10 leading collaborative perception models and their variants for the detection task, including AttFuse [53], F-Cooper [4], V2X-ViT [52], DiscoNet [29], V2VNet [46], CoBEVT [51], and BM2CP [63]. To ensure a fair comparison, we use official model configurations and public checkpoints from open-source codebases where available or re-train the models using default settings if necessary. For each corruption type, we report metrics by averaging results across five severity levels. To provide a

Table 1. Benchmarking results for Global Interference on OPV2V-C. We report the performance under each corruption AP_c and the overall corruption robustness AP_{cor} averaged over all corruption types. The results are evaluated based on AP@0.5.

Cortype		AttFuse [53]	F-Cooper [4]	V2X-ViT [52]	DiscoNet [29]	V2VNet [46]	CoBEVT [51]	Max	Late	NoFusion
$\mathrm{AP}_{clean}\uparrow$		37.13	34.85	58.61	47.34	46.64	40.49	45.87	**68.41**	35.68
Weather	Bright	21.39	9.05	43.06	33.55	36.42	23.18	27.66	**51.76**	25.22
	Dark	14.00	16.69	15.58	18.48	7.84	14.12	13.33	**22.03**	8.94
	Fog	10.30	11.15	15.23	15.02	**15.58**	10.78	11.73	12.99	4.47
	Frost	7.12	9.54	**11.21**	9.62	2.82	6.14	10.19	11.06	4.21
	Snow	6.46	4.77	11.19	10.42	7.55	9.36	**14.33**	12.37	4.75
Noise	Gaussian	9.17	8.98	7.45	3.10	2.15	5.97	11.13	**13.80**	5.23
	Impulse	9.17	9.01	8.11	2.08	2.03	5.65	10.56	**14.91**	5.55
	Shot	8.58	10.55	8.48	0.71	2.27	4.51	11.26	**15.60**	6.24
Blur	Zoom	18.85	21.44	26.57	21.68	15.89	18.95	23.52	**31.10**	12.90
	Motion	22.07	24.87	31.19	30.17	19.24	24.55	27.56	**40.52**	17.22
	Defocus	18.64	17.65	**21.57**	21.27	8.79	17.01	19.09	17.07	6.55
Failure	Crash	22.30	16.62	21.96	16.21	16.19	23.23	21.08	**32.02**	13.61
Color	Quant	26.01	21.03	41.48	35.09	33.68	25.86	33.27	**48.13**	23.60
Alignment	Tempormis	29.74	28.05	47.91	39.31	38.75	33.38	36.85	**49.28**	35.68
$\mathrm{AP}_{cor}\uparrow$		15.99	14.96	22.21	18.34	14.94	15.91	19.40	**26.62**	12.44



(a) AP_{clean} *vs.* AP_{cor} (b) AP_{clean} *vs.*mRCE (c) AP_{clean} *vs.*mPosC (d) AP_{clean} *vs.*mNegC

Figure 3. The relationships between baseline performance (AP_{clean}) and various metrics, including collaborative robustness (AP_{cor}, mRCE), collaborative compensation (mPosC) and collaborative disruption (mNegC).

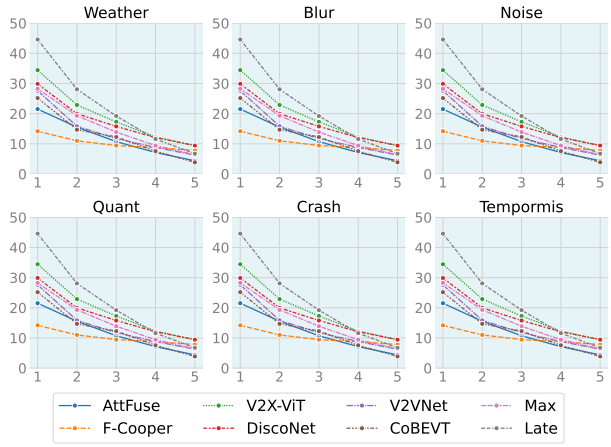


Figure 4. The corrupted average precision (AP_c) of existing methods across five severity levels of corruption.

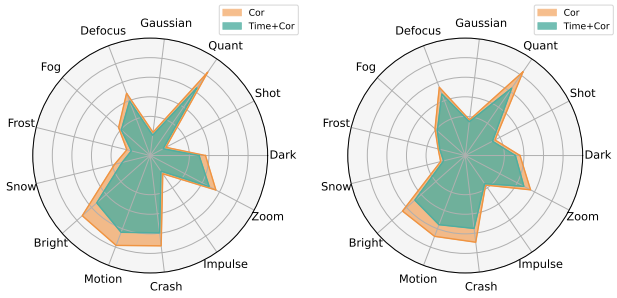
comprehensive robustness evaluation, we also include three non-learning baseline methods: *Max*, which directly maximizes the feature map for collaboration; *Late*, which fuses final predictions shared from CAVs; and *NoFusion*, representing the ego vehicle perception model without collaboration. The *NoFusion* model is further used to compute the PosC and NegC metrics as defined in Eq. 3 and Eq. 4.

4.2. Benchmarking Results for Global Interference

Single Corruption: The performance of collaborative perception models under various corruption types is presented in Table 1, with results across different corruption severity levels shown in Fig. 4. The results indicate that existing models experience varying levels of performance degradation depending on the corruption type, with conditions like snow, frost, and noise having particularly severe impacts. The *NoFusion* model consistently performs the worst, showing significantly lower AP_{cor} across all corruption types, underscoring the limitations of single-vehicle perception. Interestingly, the simple *Max* fusion method outperforms several more complex models in interference scenarios, suggesting that increased model complexity does not necessarily lead to improved robustness against interference. Moreover, the *Late* fusion methods demonstrate greater adaptability and robustness compared to existing intermediate feature fusion techniques.

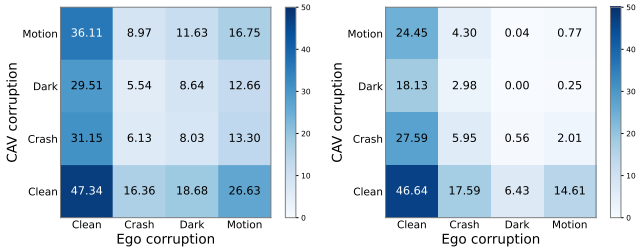
To further understand the relationship between baseline performance and robustness, we analyze the results in Fig. 3a. The findings suggest that models with higher baseline performance may handle disturbances better, likely due to more robust feature extraction. Similarly, Fig. 3b shows a positive correlation between a model's mRCE and its clean-data performance, indicating that higher-performing models may also experience larger relative performance drops under corruption. With mRCE values generally between 55% and 70%, current models exhibit suboptimal robustness, underscoring the need for further improvement.

Multiple Corruptions: As shown in Fig. 5, different types



(a) CoBEVT (b) AttFuse

Figure 5. Radar charts for CoBEVT [51] and AttFuse [53] showing performance under single corruptions (labeled as "Cor" in the figure) and multiple corruptions that include additional temporal misalignment (labeled as "Time" in the figure).



(a) DiscoNet (b) V2VNet

Figure 6. Results of V2VNet [46] and DiscoNet [29] in terms of AP_{cor} under heterogeneous corruptions between Ego and CAVs.

of corruptions have varying impacts on CoBEVT [51] and AttFuse [53], with noise and frost disturbances causing the most significant performance degradation. Adding temporal misalignment to these camera corruptions further intensifies the effects of disturbances such as brightness, motion, and quantization. These results suggest that, in real-world scenarios, the simultaneous occurrence of multiple disturbances presents even greater challenges for model robustness. Please refer to the appendix for more results.

Hetero Corruptions: Fig. 6 shows that corruptions impact the Ego vehicle more severely than the CAVs in both DiscoNet [29] and V2VNet [46]. When the Ego vehicle is "Clean", the system performs effectively despite various "CAV corruption" conditions. However, when both Ego and CAVs are subjected to different types of disturbances, the model's performance declines significantly, particularly when one of them experiences crash corruption, which has the most pronounced impact on overall system performance. More results are included in the appendix.

New Scenes with Corruption: While collaborative perception models are typically evaluated in familiar environments, real-world deployment demands robustness in diverse, unseen scenes. Table 2 evaluates the model performance in familiar OPV2V scenes (Default) versus previously unseen scenes (Culver). The AP_{clean} performance of V2X-ViT [52] declines by 47.8%, indicating that the scene variation significantly impacts model effectiveness. When disturbances and new scenes occur simultaneously, performance drops further, with AP_{cor} decreasing by 55% and RCE increasing by 9.5%. These results highlight the challenges that current collaborative methods encounter in handling disturbed environments with unfamiliar scenes.

Table 2. Performance comparison in the familiar OPV2V scenes (Default) versus the unseen new scenes (Culver).

Model	$\mathrm{AP}_{clean}\uparrow$		$\mathrm{AP}_{cor}\uparrow$		mRCE↓	
	Default	Culver	Default	Culver	Default	Culver
AttFuse	37.13	19.48(↓47.5%)	14.93	7.29(↓51.2%)	41.51	41.93(↑1.0%)
F-Cooper	34.85	26.52(↓23.9%)	13.95	8.58(↓38.5%)	37.54	48.25(↑28.5%)
V2X-ViT	58.61	30.62(↓47.8%)	20.24	9.11(↓55.0%)	46.94	51.38(↑9.5%)
DiscoNet	47.34	28.66(↓39.5%)	16.72	8.31(↓50.3%)	49.94	53.65(↑7.4%)
V2VNet	46.64	26.80(↓42.5%)	13.11	6.79(↓48.3%)	53.86	56.96(↑5.8%)
CoBEVT	40.49	23.97(↓40.8%)	14.56	7.10(↓51.2%)	46.42	49.39(↑6.4%)
Max	45.87	24.97(↓45.6%)	18.05	7.56(↓58.1%)	41.90	47.95(↑14.4%)

4.3. Benchmarking Results for Ego Interference

When the ego vehicle encounters unique corruptions, uncorrupted data from collaborators can offer collaborative compensation. Table 3 verifies this through PosC_c and mPosC metrics. Results indicate that the simple *Late* fusion significantly enhances mPosC compared to other intermediate fusion methods, suggesting that complex fusion models (e.g., attention-based [51–53], graph-based [29, 46]) are effective only when the ego vehicle captures essential information. The compensation stability varies across models and disturbances: V2X-ViT [52] and V2VNet [46] show fluctuations, whereas F-Cooper [4] and CoBEVT [51] exhibit more consistency. Fig. 3c further reveals a positive correlation between collaborative compensation mPosC and baseline performance AP_{clean}, suggesting that models with higher baseline results provide stronger compensatory support.

4.4. Benchmarking Results for CAV Interference

In scenarios where collaborative vehicles experience unique corruptions not directly affecting the ego vehicle, shared data from these vehicles can introduce "collaborative disruption." Table 3 presents the NegC_c and mNegC metrics to quantify this effect. A NegC value over 100% indicates that corrupted data from collaborators degrades the ego vehicle's performance. The results show that the *Late* fusion is less vulnerable to these disruptions and often provides positive benefits even under corruption, while intermediate feature fusion methods tend to reduce performance. Additionally, Fig. 3d shows that higher baseline performance (AP_{clean}) correlates with reduced collaborative disruption, suggesting that models with stronger baseline performance may better handle corrupted collaborative data.

5. RCP-Drop & RCP-Mix

To enhance robustness under various corruption scenarios, we develop two simple yet effective strategies specifically for collaborative perception: RCP-Drop and RCP-Mix.

Dropout [1, 8] is a widely used regularization technique in deep learning, aimed at reducing model overfitting by

Table 3. Benchmarking results for Ego Interference and CAV Interference on OPV2V-C. We report collaborative compensation and disruption performance under each corruption type, including NegC_c and PosC_c, as well as averages across all corruption types, mNegC and mPosC. Results are evaluated based on AP@0.5. Further details are available in the appendix.

Cor Types	AttFuse [53]		F-Cooper [4]		V2X-ViT [52]		DiscoNet [29]		V2VNet [46]		CoBEVT [51]		Max		Late	
	NegC_c	PosC_c	NegC_c	PosC_c	NegC_c	PosC_c	NegC_c	PosC_c	NegC_c	PosC_c	NegC_c	PosC_c	NegC_c	PosC_c	NegC_c	PosC_c
Bright	121.27	8.54	115.97	9.68	86.68	28.90	95.77	18.29	103.03	27.29	107.82	6.24	108.71	20.68	**82.45**	**50.82**
Dark	126.73	12.96	122.28	21.77	104.24	14.80	109.59	18.68	127.29	6.43	116.00	14.97	121.52	22.14	**99.91**	**50.81**
Zoom	117.63	19.66	116.43	21.74	93.72	29.29	100.96	20.66	107.62	23.63	107.98	17.63	104.68	25.41	**88.50**	**51.99**
Motion	116.67	21.12	115.52	22.50	94.50	25.84	99.33	22.52	117.46	9.83	106.19	19.19	102.83	25.67	**90.30**	**51.09**
Defocus	118.17	25.27	116.87	24.62	**99.47**	27.14	106.20	20.90	121.30	10.81	107.82	18.68	107.80	27.13	101.55	**50.36**
Gaussian	127.32	13.30	122.73	15.47	112.39	10.16	115.02	12.32	123.35	5.81	117.93	11.545	115.84	24.71	**99.75**	**50.83**
Impulse	127.04	15.75	122.98	14.94	112.06	11.78	115.13	11.82	123.21	5.62	117.86	11.11	115.33	24.33	**99.53**	**50.84**
Shot	126.87	13.12	122.01	17.78	112.33	9.86	115.39	11.64	123.82	5.95	118.17	11.81	114.04	25.70	**99.89**	**50.82**
Crash	116.22	22.35	115.73	20.51	97.50	13.75	107.04	10.67	112.58	11.98	105.05	18.96	108.12	24.21	**91.53**	**50.14**
Quant	118.03	17.63	113.91	22.49	99.86	22.69	**98.60**	21.65	110.00	20.66	110.00	12.82	110.14	25.83	98.76	**48.70**
mNegC↓/mPosC↑	121.59	16.97	118.44	19.15	101.27	19.42	106.31	16.92	116.97	12.80	111.48	14.29	110.90	24.58	**95.22**	**50.64**

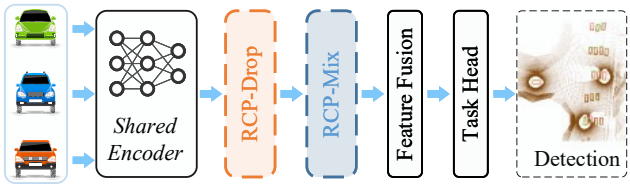


Figure 7. Overview of the RCP-Drop & RCP-Mix strategies.

randomly dropping features within certain network layers or blocks during training. In the collaborative perception context, the ego vehicle may encounter incomplete shared perceptual data from CAVs due to communication failures, or CAVs leaving the collaborative queue. Existing collaborative perception models, while beneficial in multi-vehicle setups, often underperform in single-vehicle scenarios due to their learned dependency on supplementary data from other vehicles. To prevent models from being overly reliant on CAVs data and to better equip them for single-vehicle operation, as shown in Fig. 7 we introduce RCP-Drop, which simulates these conditions by selectively discarding information from certain CAVs. Technically, it is defined as:

$$F_{agg}^{ego} = \mathcal{A}(F^{ego}, \{\mathbb{1}(p_i < p_{drop}) \cdot F_i^{cav}\}_{i=1}^{N}), \quad (5)$$

where F^{ego} represents the feature originating from the ego vehicle, F_{agg}^{ego} denotes the aggregated feature after perception information exchange, and F_i^{cav} is the feature map shared from the i-th collaborative vehicle. Here, $\mathcal{A}$ is a specific feature fusion function developed in existing methods [29, 44, 53], p_i is sampled from a uniform distribution between 0 and 1, and p_{drop} is the threshold determining whether data from each collaborative vehicle is discarded.

Rather than keeping the threshold fixed during training, we develop a dynamic adjustment strategy for p_{drop}. Specifically, we start with a high dropout probability, which reduces the number of participating collaborative vehicles and enhances the single-vehicle backbone's feature extraction capabilities. Gradually, we lower the dropout probability to increase the participation of collaborative vehicles, optimizing the collaborative perception performance over time. In this study, we set $p_{drop} = 1.0 - (t/T)^2$, where t and T are the current and maximum training epoch.

Existing research shows that visual data can be characterized by feature statistics (i.e., the mean and standard deviation of feature maps). For the same semantic concepts, distinct styles (such as variations in color and texture) correspond to different feature statistics. Inspired by Mixstyle [64, 65], we propose RCP-Mix to enhance model adaptability to varied environmental conditions and styles encountered in collaborative perception. Unlike Mixstyle, which performs random feature statistic mixing across instances within a mini-batch, RCP-Mix probabilistically combines feature statistics between the ego vehicle and collaborating CAVs. Technically, RCP-Mix is defined as:

$$\begin{aligned} \mu_{all} &= \frac{1}{N}\sum_{i=1}^{N}\mu_i; \quad \sigma_{all} = \frac{1}{N}\sum_{i=1}^{N}\sigma_i, \\ \mu_i^{mix} &= \lambda\mu_i + (1-\lambda)\mu_{all}, \\ \sigma_i^{mix} &= \lambda\sigma_i + (1-\lambda)\sigma_{all}, \\ F_i^{mix} &= \frac{F_i - \mu_i}{\sigma_i} * \sigma_i^{mix} + \mu_i^{mix}, \end{aligned} \quad (6)$$

where μ_i, σ_i are the means and standard deviations of the feature map F_i, λ is the weight sampled from the Beta distribution, i.e., $\lambda \sim Beta(\alpha, \alpha)$, we set $\alpha = 0.1$ by default.

Table 4. Efficacy of RCP-Drop and RCP-Mix on collaborative robustness, collaborative compensation and collaborative disruption.

	$\mathrm{AP}_{clean}\uparrow$	$\mathrm{mAP}_{cor}\uparrow$	mNegC ↓	mPosC ↑
CoBEVT [51]	40.49	15.91	111.48	14.29
+BN [39]	42.74	27.87	112.94	18.28
+BN+RCP-Mix	**48.25**	**31.83**	**109.11**	**23.44**
+BN+RCP-Drop	47.84	30.68	109.45	20.07
AttFuse [53]	37.13	15.99	121.59	16.97
+BN [39]	36.13	22.35	126.04	18.85
+BN+RCP-Mix	43.12	**28.85**	**114.97**	**22.51**
+BN+RCP-Drop	**45.18**	27.56	115.64	18.17

Experimental Analysis: We selected CoBEVT [51] and AttFuse [53] with online Batch Normalization (BN) Stats Adapt [39] as baselines, integrating them with our RCP-Drop and RCP-Mix strategies. Table 4 demonstrates that, despite their simplicity, both RCP-Drop and RCP-Mix effectively counteract performance degradation from data corruptions. In particular, for AP_{clean}, these strategies lead to substantial performance improvements compared to the marginal gains or occasional declines observed with the vanilla BN adapt. Moreover, across the three inter-

Table 5. Ablation on the use of different backbone.

Method	backbone	Weather	Blur	Noise	Quant	Crash	AP	AP_{cor}
Max ○	ResNet101	11.79	22.3	22.77	22.87	18.64	35.12	18.14
Max ●	EfficientNet-b0	15.45	23.39	10.98	33.27	21.08	45.87	18.05
V2X-ViT○	ResNet101	10.90	24.66	16.63	28.75	10.41	46.21	16.73
V2X-ViT ●	EfficientNet-b0	19.25	26.45	8.01	41.48	21.96	58.61	20.24

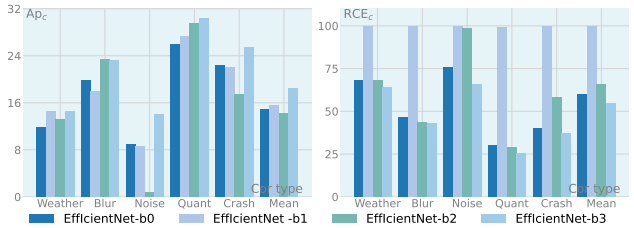


Figure 8. Impact of different backbone sizes on model robustness.

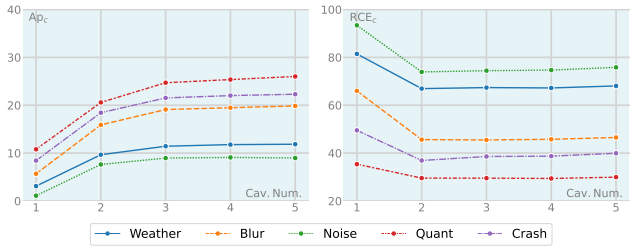


Figure 9. Effect of CAV number on model robustness.

ference scenarios, Global Interference, Ego Interference, and CAV Interference, our RCP-Drop and RCP-Mix consistently show robust improvements in model resilience under corruption conditions (AP_{cor}), collaborative compensation (mPosC), and reduced collaborative disruption (mNegC). These results further underline the efficacy of our approaches in maintaining model performance and collaborative stability in challenging conditions.

6. Observation & Discussion

In this section, we analyze and discuss the impact of various model configurations and techniques on robustness.

Backbones: Table 5 compares the performance of different backbones, showing that models using EfficientNet-b0 outperform ResNet101 [22] in all conditions except noise disturbances. leading to its use as the backbone for all models in this study. Fig. 8 further compares AttFuse with various backbone sizes, revealing that larger backbones generally improve robustness. Notably, EfficientNet-b0 and EfficientNet-b3 [21] show greater stability and robustness across multiple conditions, particularly excelling under "Noise" and "Crash" scenarios, while EfficientNet-b2 performs poorly in these cases.

Feature Fusion: Table 6 shows that multiscale fusion enhances model performance on clean datasets and improves robustness against various perturbations. Additionally, the simple Max Fusion demonstrates greater robustness than attention mechanisms, particularly under noisy conditions, suggesting that more complex mechanisms may be less effective at mitigating certain types of perturbations.

Table 6. Ablation on the use of multiscale fusion.

Method	multiscale	Weather	Blur	Noise	Quant	Crash	AP	AP_{cor}
Max ○	✗	15.45	23.39	10.98	33.27	21.08	45.87	18.05
Max ●	✓	30.65	34.92	28.61	49.86	38.15	65.38	33.22
AttFuse○	✗	11.85	19.85	8.97	26.01	22.30	37.13	14.93
AttFuse ●	✓	23.36	31.90	9.01	43.43	30.07	62.58	24.08

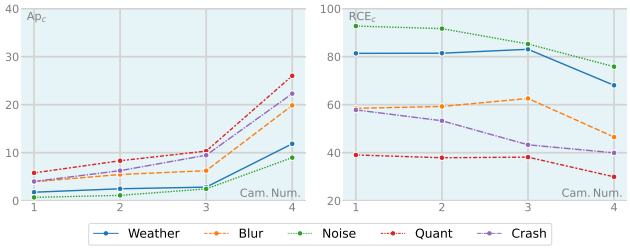


Figure 10. Effect of camera number on model robustness.

CAV and Camera Number: Fig. 9 shows that increasing the number of collaborative CAVs generally enhances model performance. Notably, two collaborating vehicles provide a significant improvement over a single vehicle; however, the rate of improvement diminishes once the number of collaborators exceeds four, indicating a diminishing marginal effect. Fig. 10 shows that as camera count increases, both model performance and robustness improve significantly, with a particularly notable boost from three to four cameras.

7. Conclusion

In this study, we introduce RCP-Bench, the first comprehensive benchmark designed to evaluate the robustness of camera-based collaborative perception models under diverse real-world disturbances. RCP-Bench encompasses 14 corruption types at 5 severity levels across 3 large-scale datasets. We systematically evaluate three collaborative scenarios across six cases to assess robustness in the Global Interference scenario, explore collaborative advantages in the Ego Interference scenario, and examine disturbance risks in the CAV Interference scenario. To improve robustness, we propose two straightforward strategies, RCP-Drop and RCP-Mix, tailored for collaborative perception. Our analysis of backbone architecture, feature fusion, CAV number, and camera number provides key insights for enhancing model resilience. We hope this work offers valuable insights and inspires future research toward more robust collaborative perception models.

Acknowledgment: This work was supported by the National Key Research and Development Program of China (No. 2024YFE0211000), in part by the National Natural Science Foundation of China (No. 62372329), in part by Shanghai Scientific Innovation Foundation (No. 23DZ1203400), in part by Tongji-Qomolo Autonomous Driving Commercial Vehicle Joint Lab Project, and in part by Xiaomi Young Talents Program.

References

[1] Pierre Baldi and Peter J Sadowski. Understanding dropout. *Advances in neural information processing systems*, 26, 2013. 2, 6

[2] Alexander Carballo, Jacob Lambert, Abraham Monrroy, David Wong, Patiphon Narksri, Yuki Kitsukawa, Eijiro Takeuchi, Shinpei Kato, and Kazuya Takeda. Libre: The multiple 3d lidar dataset. In *2020 IEEE intelligent vehicles symposium (IV)*, pages 1094–1101. IEEE, 2020. 1

[3] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer, 2020. 1

[4] Qi Chen, Xu Ma, Sihai Tang, Jingda Guo, Qing Yang, and Song 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*, pages 88–100, 2019. 4, 5, 6, 7

[5] Runjian Chen, Yao Mu, Runsen Xu, Wenqi Shao, Chenhan Jiang, Hang Xu, Zhenguo Li, and Ping Luo. Coˆ 3: Cooperative unsupervised 3d representation learning for autonomous driving. *arXiv preprint arXiv:2206.04028*, 2022. 1, 2

[6] Jiaxun Cui, Hang Qiu, Dian Chen, Peter Stone, and Yuke Zhu. Coopernaut: End-to-end driving with cooperative perception for networked vehicles. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17252–17262, 2022. 2

[7] Yinpeng Dong, Caixin Kang, Jinlai Zhang, Zijian Zhu, Yikai Wang, Xiao Yang, Hang Su, Xingxing Wei, and Jun Zhu. Benchmarking robustness of 3d object detection to common corruptions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1022–1032, 2023. 3, 4

[8] Golnaz Ghiasi, Tsung-Yi Lin, and Quoc V Le. Dropblock: A regularization method for convolutional networks. *Advances in Neural Information Processing Systems*, 31, 2018. 2, 6

[9] 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 *Proceedings of the IEEE/CVF international conference on computer vision*, pages 15283–15292, 2021. 1

[10] Martin Hahner, Christos Sakaridis, Mario Bijelic, Felix Heide, Fisher Yu, Dengxin Dai, and Luc Van Gool. Lidar snowfall simulation for robust 3d object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16364–16374, 2022. 1

[11] Xiaoshuai Hao, Mengchuan Wei, Yifan Yang, Haimei Zhao, Hui Zhang, Yi Zhou, Qiang Wang, Weiming Li, Lingdong Kong, and Jing Zhang. Is your hd map constructor reliable under sensor corruptions? *arXiv preprint arXiv:2406.12214*, 2024. 2

[12] Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. *arXiv preprint arXiv:1903.12261*, 2019. 1, 2

[13] Shixin Hong, Yu Liu, Zhi Li, Shaohui Li, and You He. Multi-agent collaborative perception via motion-aware robust communication network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15301–15310, 2024. 2

[14] Senkang Hu, Zhengru Fang, Xianhao Chen, Yuguang Fang, and Sam Kwong. Towards full-scene domain generalization in multi-agent collaborative bird's eye view segmentation for connected and autonomous driving. *arXiv preprint arXiv:2311.16754*, 2023. 2, 3

[15] 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. 1, 2

[16] Yue Hu, Yifan Lu, Runsheng Xu, Weidi Xie, Siheng Chen, and Yanfeng Wang. Collaboration helps camera overtake lidar in 3d detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9243–9252, 2023. 1, 2

[17] Xiaohui Jiang, Shuailin Li, Yingfei Liu, Shihao Wang, Fan Jia, Tiancai Wang, Lijin Han, and Xiangyu Zhang. Far3d: Expanding the horizon for surround-view 3d object detection. *arXiv preprint arXiv:2308.09616*, 2023. 1

[18] 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

[19] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023. 1

[20] Lingdong Kong, Shaoyuan Xie, Hanjiang Hu, Lai Xing Ng, Benoit Cottereau, and Wei Tsang Ooi. Robodepth: Robust out-of-distribution depth estimation under corruptions. *Advances in Neural Information Processing Systems*, 36, 2024. 2

[21] Brett Koonce and Brett Koonce. Efficientnet. *Convolutional neural networks with swift for Tensorflow: image recognition and dataset categorization*, pages 109–123, 2021. 8

[22] Brett Koonce and Brett Koonce. Resnet 50. *Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization*, pages 63–72, 2021. 8

[23] Zixing Lei, Shunli Ren, Yue Hu, Wenjun Zhang, and Siheng Chen. Latency-aware collaborative perception. In *European Conference on Computer Vision*, pages 316–332. Springer, 2022. 2

[24] Zixing Lei, Zhenyang Ni, Ruize Han, Shuo Tang, Chen Feng, Siheng Chen, and Yanfeng Wang. Robust collaborative perception without external localization and clock devices. *arXiv preprint arXiv:2405.02965*, 2024. 1

[25] 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. *arXiv preprint arXiv:2403.11371*, 2024. 2, 3

[26] Jinlong Li, Runsheng Xu, Xinyu Liu, Jin Ma, Zicheng Chi, Jiaqi Ma, and Hongkai Yu. Learning for vehicle-to-vehicle cooperative perception under lossy communication. *IEEE Transactions on Intelligent Vehicles*, 8(4):2650–2660, 2023. 1, 2

[27] Xiangtai Li, Henghui Ding, Haobo Yuan, Wenwei Zhang, Jiangmiao Pang, Guangliang Cheng, Kai Chen, Ziwei Liu,

and Chen Change Loy. Transformer-based visual segmentation: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 1

[28] Xiang Li, Junbo Yin, Wei Li, Chengzhong Xu, Ruigang Yang, and Jianbing Shen. Di-v2x: Learning domain-invariant representation for vehicle-infrastructure collaborative 3d object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 3208–3215, 2024. 1

[29] Yiming Li, Shunli Ren, Pengxiang Wu, Siheng Chen, Chen Feng, and Wenjun Zhang. Learning distilled collaboration graph for multi-agent perception. *Advances in Neural Information Processing Systems*, 34:29541–29552, 2021. 4, 5, 6, 7

[30] Gen Liu, Jin Han, and Wenzhong Rong. Feedback-driven loss function for small object detection. *Image and Vision Computing*, 111:104197, 2021. 1

[31] Si Liu, Zihan Ding, Jiahui Fu, Hongyu Li, Siheng Chen, Shifeng Zhang, and Xu Zhou. V2x-pc: Vehicle-to-everything collaborative perception via point cluster. *arXiv preprint arXiv:2403.16635*, 2024. 2

[32] Yen-Cheng Liu, Junjiao Tian, Nathaniel Glaser, and Zsolt Kira. When2com: Multi-agent perception via communication graph grouping. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pages 4106–4115, 2020. 1, 2

[33] Yen-Cheng Liu, Junjiao Tian, Chih-Yao Ma, Nathan Glaser, Chia-Wen Kuo, and Zsolt Kira. Who2com: Collaborative perception via learnable handshake communication. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6876–6883. IEEE, 2020. 1, 2

[34] Yifan Lu, Quanhao Li, Baoan Liu, Mehrdad Dianati, Chen Feng, Siheng Chen, and Yanfeng Wang. Robust collaborative 3d object detection in presence of pose errors. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4812–4818. IEEE, 2023. 1, 2

[35] Claudio Michaelis, Benjamin Mitzkus, Robert Geirhos, Evgenia Rusak, Oliver Bringmann, Alexander S Ecker, Matthias Bethge, and Wieland Brendel. Benchmarking robustness in object detection: Autonomous driving when winter is coming. *arXiv preprint arXiv:1907.07484*, 2019. 2

[36] Hang Qiu, Pohan Huang, Namo Asavisanu, Xiaochen Liu, Konstantinos Psounis, and Ramesh Govindan. Autocast: Scalable infrastructure-less cooperative perception for distributed collaborative driving. *arXiv preprint arXiv:2112.14947*, 2021. 2

[37] Deyuan Qu, Qi Chen, Tianyu Bai, Hongsheng Lu, Heng Fan, Hao Zhang, Song Fu, and Qing Yang. Sicp: Simultaneous individual and cooperative perception for 3d object detection in connected and automated vehicles. *arXiv preprint arXiv:2312.04822*, 2023. 2

[38] Shunli Ren, Zixing Lei, Zi Wang, Siheng Chen, and Wenjun Zhang. Robust collaborative perception against communication interruption. In *the 2nd IJCAI Workshop on Artificial Intelligence for Autonomous Driving (2022)*, 2022. 1, 2

[39] Steffen Schneider, Evgenia Rusak, Luisa Eck, Oliver Bringmann, Wieland Brendel, and Matthias Bethge. Improving robustness against common corruptions by covariate shift adaptation. *Advances in neural information processing systems*, 33:11539–11551, 2020. 7

[40] Rui Song, Chenwei Liang, Hu Cao, Zhiran Yan, Walter Zimmer, Markus Gross, Andreas Festag, and Alois Knoll. Collaborative semantic occupancy prediction with hybrid feature fusion in connected automated vehicles. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17996–18006, 2024. 2

[41] Ziying Song, Lin Liu, Feiyang Jia, Yadan Luo, Caiyan Jia, Guoxin Zhang, Lei Yang, and Li Wang. Robustness-aware 3d object detection in autonomous driving: A review and outlook. *IEEE Transactions on Intelligent Transportation Systems*, 2024. 4

[42] Sanbao Su, Yiming Li, Sihong He, Songyang Han, Chen Feng, Caiwen Ding, and Fei Miao. Uncertainty quantification of collaborative detection for self-driving. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5588–5594. IEEE, 2023. 2

[43] Nicholas Vadivelu, Mengye Ren, James Tu, Jingkang Wang, and Raquel Urtasun. Learning to communicate and correct pose errors. In *Conference on Robot Learning*, pages 1195–1210. PMLR, 2021. 1

[44] Tianhang Wang, Guang Chen, Kai Chen, Zhengfa Liu, Bo Zhang, Alois Knoll, and Changjun Jiang. Umc: A unified bandwidth-efficient and multi-resolution based collaborative perception framework. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8187–8196, 2023. 1, 2, 7

[45] Tianhang Wang, Fan Lu, Zehan Zheng, Guang Chen, and Changjun Jiang. Rcdn: Towards robust camera-insensitivity collaborative perception via dynamic feature-based 3d neural modeling. *arXiv preprint arXiv:2405.16868*, 2024. 2

[46] Tsun-Hsuan Wang, Sivabalan Manivasagam, Ming Liang, Bin Yang, Wenyuan Zeng, and Raquel Urtasun. V2vnet: Vehicle-to-vehicle communication for joint perception and prediction. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, pages 605–621. Springer, 2020. 4, 5, 6, 7

[47] Sizhe Wei, Yuxi Wei, Yue Hu, Yifan Lu, Yiqi Zhong, Siheng Chen, and Ya Zhang. Robust asynchronous collaborative 3d detection via bird's eye view flow. *arXiv preprint arXiv:2309.16940*, 2023. 1, 2

[48] Hao Xiang, Runsheng Xu, and Jiaqi Ma. Hm-vit: Heterogeneous vehicle-to-vehicle cooperative perception with vision transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 284–295, 2023. 1

[49] Shaoyuan Xie, Lingdong Kong, Wenwei Zhang, Jiawei Ren, Liang Pan, Kai Chen, and Ziwei Liu. Robobev: Towards robust bird's eye view perception under corruptions. *arXiv preprint arXiv:2304.06719*, 2023. 2

[50] Chang Xu, Jinwang Wang, Wen Yang, and Lei Yu. Dot distance for tiny object detection in aerial images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1192–1201, 2021. 1

[51] Runsheng Xu, Zhengzhong Tu, Hao Xiang, Wei Shao, Bolei Zhou, and Jiaqi Ma. Cobevt: Cooperative bird's eye view semantic segmentation with sparse transformers. *arXiv preprint arXiv:2207.02202*, 2022. 2, 4, 5, 6, 7

[52] Runsheng Xu, Hao Xiang, Zhengzhong Tu, Xin Xia, Ming-Hsuan Yang, and Jiaqi Ma. V2x-vit: Vehicle-to-everything cooperative perception with vision transformer. In *European*

conference on computer vision, pages 107–124. Springer, 2022. 3, 4, 5, 6, 7

[53] 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. 3, 4, 5, 6, 7

[54] Runsheng Xu, Weizhe Chen, Hao Xiang, Xin Xia, Lantao Liu, and Jiaqi Ma. Model-agnostic multi-agent perception framework. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1471–1478. IEEE, 2023. 2

[55] Runsheng Xu, Jinlong Li, Xiaoyu Dong, Hongkai Yu, and Jiaqi Ma. Bridging the domain gap for multi-agent perception. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6035–6042. IEEE, 2023. 1

[56] Dingkang Yang, Kun Yang, Yuzheng Wang, Jing Liu, Zhi Xu, Rongbin Yin, Peng Zhai, and Lihua Zhang. How2comm: Communication-efficient and collaboration-pragmatic multi-agent perception. *Advances in Neural Information Processing Systems*, 36, 2024. 2

[57] Kun Yang, Dingkang Yang, Jingyu Zhang, Hanqi Wang, Peng Sun, and Liang Song. What2comm: Towards communication-efficient collaborative perception via feature decoupling. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 7686–7695, 2023. 2

[58] Kun Yang, Dingkang Yang, Ke Li, Dongling Xiao, Zedian Shao, Peng Sun, and Liang Song. Align before collaborate: Mitigating feature misalignment for robust multi-agent perception. In *European Conference on Computer Vision*, pages 282–299. Springer, 2025. 2

[59] Haibao Yu, Yizhen Luo, Mao Shu, Yiyi Huo, Zebang Yang, Yifeng Shi, Zhenglong Guo, Hanyu Li, Xing Hu, Jirui Yuan, et al. 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. 3

[60] Haibao Yu, Yingjuan Tang, Enze Xie, Jilei Mao, Ping Luo, and Zaiqing Nie. Flow-based feature fusion for vehicle-infrastructure cooperative 3d object detection. *Advances in Neural Information Processing Systems*, 36, 2024. 1, 2

[61] Jingyu Zhang, Kun Yang, Hanqi Wang, Peng Sun, and Liang Song. Efficient vehicular collaborative perception based on saptial-temporal feature compression. *IEEE Transactions on Vehicular Technology*, 2024. 2

[62] Jingyu Zhang, Kun Yang, Yilei Wang, Hanqi Wang, Peng Sun, and Liang Song. Ermvp: Communication-efficient and collaboration-robust multi-vehicle perception in challenging environments. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12575–12584, 2024. 2

[63] Binyu Zhao, Wei Zhang, and Zhaonian Zou. Bm2cp: Efficient collaborative perception with lidar-camera modalities. *arXiv preprint arXiv:2310.14702*, 2023. 4

[64] Kaiyang Zhou, Yongxin Yang, Yu Qiao, and Tao Xiang. Domain generalization with mixstyle. In *International Conference on Learning Representations (ICLR)*, 2021. 2, 7

[65] Kaiyang Zhou, Yongxin Yang, Yu Qiao, and Tao Xiang. Mixstyle neural networks for domain generalization and adaptation. *International Journal of Computer Vision*, 132 (3):822–836, 2024. 2, 7

[66] Yang Zhou, Jiuhong Xiao, Yue Zhou, and Giuseppe Loianno. Multi-robot collaborative perception with graph neural networks. *IEEE Robotics and Automation Letters*, 7(2):2289–2296, 2022. 1

[67] Zijian Zhu, Yichi Zhang, Hai Chen, Yinpeng Dong, Shu Zhao, Wenbo Ding, Jiachen Zhong, and Shibao Zheng. Understanding the robustness of 3d object detection with bird's-eye-view representations in autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21600–21610, 2023. 4

[68] Zhengxia Zou, Keyan Chen, Zhenwei Shi, Yuhong Guo, and Jieping Ye. Object detection in 20 years: A survey. *Proceedings of the IEEE*, 111(3):257–276, 2023. 1