From Chaos to Clarity: Strengthening 3D Collaborative Autonomous Vehicle Perception with Noise-Aware Training

Everett Richards[1,*], Allie Lopez[2,*], Jose Morales[3], Ziming Zhang[3]
[1]*San Diego State University,* [2]*College of St. Scholastica,* [3]*Worcester Polytechnic Institute*

Abstract—**Reliable perception is critical for autonomous vehicles (AVs), particularly in collaborative systems where multiple agents share sensor data to improve environmental awareness. However, most collaborative object detection frameworks assume ideal sensor conditions and degrade sharply when exposed to noise, occlusion, or hardware variability—issues common in real-world deployment. In this work, we present a noise-aware training framework that improves the robustness and efficiency of 3D object detectors by injecting Gaussian noise into LiDAR point clouds during training. Using the BM2CP architecture and the DAIR-V2X dataset, we evaluated a range of training regimes, both constant and progressive, in more than 1,400 inference trials. We find that models trained with curriculum-style exposure to increasing noise levels degrade more gracefully under inference-time corruption, generalize better across sensor quality, and converge faster than traditional baselines. For instance, a model trained with heavy noise for just 20 epochs outperforms a baseline trained for 50 epochs when evaluated on degraded input, highlighting the accelerated rate of convergence enabled by noise-infused training. These findings demonstrate that robustness is a tunable and scalable property, offering a practical path toward safer and more cost-effective AV perception in noisy, uncertain, or resource-constrained environments.**

Index Terms—**Collaborative Perception, LiDAR, Gaussian Noise, Robustness, 3D Object Detection, Autonomous Vehicles**

I. Introduction

Autonomous vehicles (AVs) must reliably detect pedestrians, cyclists, and other vehicles under diverse and unpredictable conditions. However, even small perception errors–due to occlusion or degraded sensors–can lead to catastrophic failures. Ensuring robust, complete environmental awareness remains a major hurdle in AV safety.

To achieve comprehensive and reliable perception, autonomous vehicles increasingly rely on collaborative methods that combine sensor data from multiple agents [1], [2]. By sharing information, vehicles can overcome occlusions and blind spots that limit single-agent perception. While much of the literature focuses on optimizing communication strategies or reducing latency through techniques like feature-level fusion [3] or edge computing [4], less attention has been paid to the quality and reliability of the data itself. In real-world deployments, collaborative systems must contend with noisy, incomplete, or low-fidelity sensor input, especially from cost-constrained hardware [5]. This motivates the need for perception models that are robust not only to communication constraints, but also to degraded or uncertain data inputs.

One ongoing challenge in 3D collaborative vehicle perception is the extreme cost of high-quality LiDAR sensors. For example, the self-driving taxi company Waymo spends more than $7,500 to manufacture a single LiDAR sensor [6]. This high cost makes it nearly impossible for automotive manufacturers to mass produce safe and affordable autonomous vehicles. Collaborative perception algorithms work best when there are many sensor-equipped vehicles in the same physical area, so the widespread deployment of such safety-critical algorithms is contingent on affordable LiDAR of sufficient quality. We aim to address this challenge by developing a model that works well for low-cost LiDAR modules (which produce noisier data than their more expensive counterparts), thereby making collaborative perception more feasible in the real world.

Another challenge in vehicle perception is that existing models are often overfitted to simulated data setsets that do not reflect real-world driving conditions [7]. We address the overfitting problem by working to maintain high performance despite poor data quality, making a model more robust to natural data variance.

We address these challenges by training collaborative perception models to anticipate and adapt to noisy sensor inputs. By injecting Gaussian noise into LiDAR point clouds during training, our models learn to generalize across sensor fidelity levels, improving their performance in real-world, cost-constrained environments.

Our primary contributions are as follows:

- Propose a noise-aware training framework that improves the robustness of collaborative perception to noisy data.
- Show that models trained with progressive Gaussian noise exhibit significantly slower performance degradation under increasing inference-time noise.
- Demonstrate that noise-injected models converge faster than baselines, enabling quicker deployment.
- Validate our approach on a real-world dataset (DAIR-V2X) using the collaborative perception framework BM2CP.

[*] These authors contributed equally to this work

II. Related Works

Collaborative perception has been widely explored in recent years as a solution to the limitations of single-agent systems, particularly occlusion and incomplete scene coverage [8], [9]. These methods aggregate sensor data from multiple vehicles to enhance detection accuracy but often introduce additional latency due to inter-agent communication.

Single-Agent Perception. Traditional perception systems rely on either LiDAR or cameras–or a fusion of both–mounted directly on a single vehicle. While these systems are efficient and cost-effective, their performance is limited by their fixed field of view and vulnerability to occlusion [10]. Recent models such as ICanC [11], RT3D [12], and AutoVision [13] have improved single-agent perception across different modalities, but they cannot overcome fundamental perspective limitations.

Collaborative Perception. Collaborative multi-agent perception leverages data from nearby vehicles to provide more complete situational awareness. BM2CP [14], MDNet [15], LCV2I [5], and Where2Comm [3] adopt multimodal fusion strategies that incorporate both LiDAR and camera data. Liu et al. [16] and Richards et al. [4] explore modality-agnostic frameworks for perception-aware communication.

Among these, Biased Multi-Modal Collaborative Perception (BM2CP) stands out for its hybrid voxel fusion design, which integrates depth information from LiDAR and projected camera features. It selectively incorporates data from neighboring vehicles to fill in uncertain or occluded regions. This makes BM2CP particularly robust to sensor dropout, although it assumes high-quality sensor input and lacks robustness guarantees under degraded conditions. Our work builds on BM2CP by relaxing this assumption and training the model to handle noisy inputs.

Feng et al. [5] propose LCV2I, a lightweight collaborative perception model that fuses low-resolution LiDAR and camera data. By using regional feature enhancement and adaptive transmission, LCV2I achieves 60% lower bandwidth and 20% lower latency than BM2CP, though with a small tradeoff in detection accuracy. However, neither LCV2I nor BM2CP explicitly models sensor degradation or trains for robustness across varying data quality. Prior work on noise injection in 3D deep learning shows that adding Gaussian perturbations can improve model resilience to corruption [17], and curriculum-style learning strategies have been shown to enhance generalization in noisy environments [18].

Our method complements this body of work by introducing a noise-aware training strategy that improves resilience to real-world sensor variance, particularly relevant for low-cost, low-resolution LiDAR deployment.

III. Methods

A. Model and Dataset

We build on BM2CP [14], a collaborative 3D object detection framework that fuses LiDAR and camera data from multiple vehicles. BM2CP supports partial modality fusion, allowing it to function even when certain sensors are unavailable or degraded. We retain BM2CP's original architecture and fusion strategies without modification, focusing solely on noise injection during the data loading phase. For training and evaluation, we use the DAIR-V2X dataset [2], a real-world collection of synchronized LiDAR and camera data from connected autonomous vehicles.

B. Noise Injection Strategy

To simulate data degradation and promote robustness, we inject zero-mean isotropic Gaussian noise into the LiDAR point clouds. This noise is applied independently to each point's (x, y, z) coordinates. For a given standard deviation σ, the noise tensor is sampled from $\mathcal{N}(0, \sigma^2)$ and added to each point cloud as it is loaded. We implement this by modifying BM2CP's data loader to conditionally inject noise based on an environment variable, enabling the same mechanism to be used during both training and inference.

Noise is added at inference time to benchmark the model's robustness under noisy deployment conditions, such as those arising from low-cost or poorly calibrated LiDAR hardware.

C. Training Regimes

We explore four training regimes with increasing noise levels over time. Each regime consists of 50 training epochs, divided into five 10-epoch stages. Table I shows the noise standard deviation used in each stage for every regime. The No Noise regime is used as a baseline, as it represents standard supervised learning under clean conditions. The next three **progressive regimes** are designed to simulate a curriculum learning approach, gradually increasing the model's resilience to more extreme sensor noise. We also introduce four **constant regimes** trained at fixed noise levels, to serve as baselines for our progressive noise regimes.

Training Regime	**Epoch Range**				
	1-10	**11-20**	**21-30**	**31-40**	**41-50**
No Noise (Baseline)	0.00	0.00	0.00	0.00	0.00
Light Noise	0.00	0.01	0.02	0.03	0.05
Moderate Noise	0.00	0.03	0.05	0.08	0.12
Heavy Noise	0.00	0.05	0.10	0.15	0.20
Constant 0.05	0.05	0.05	0.05	0.05	0.05
Constant 0.10	0.10	0.10	0.10	0.10	0.10
Constant 0.20	0.20	0.20	0.20	0.20	0.20

TABLE I: Noise schedules for each training regime. Each model was trained for 50 epochs in five stages, with noise standard deviation (σ) increasing according to the values shown. Progressive regimes simulate curriculum-style learning, while constant regimes apply fixed noise throughout.

D. Evaluation Metrics

In each experiment, we evaluate the performance using the average precision (AP) at intersection-over-union (IoU) thresholds of 0.3, 0.5, and 0.7, which are standard in 3D object detection literature. AP is measured as the ratio between true positives and all positives. IoU is measured as the ratio of intersecting area to union area between the predicted and ground truth bounding boxes for given objects.

IV. Experimental Setup

A. Hardware and Environment

All experiments were conducted using 10 NVIDIA RTX Quadro 6000 GPUs (24GB VRAM each) distributed across a multi-GPU cluster. Training and inference were implemented using PyTorch 1.10.1 with CUDA 12.1. Each model was trained on a dedicated GPU, with the four training regimes run in parallel. A single 10-epoch training phase required approximately 6 hours, while full 50-epoch training for one model took about 30 hours.

Inference jobs were distributed across all 10 GPUs, running 20 concurrent instructions (2 per GPU). Each instruction corresponds to a specific model checkpoint and Gaussian noise level, and took approximately 45 minutes to complete. The full set of 1,400 inference trials was completed in roughly 60 hours.

B. Dataset Preparation

We use the DAIR-V2X dataset [2], a real-world collaborative perception dataset that includes time-synchronized LiDAR and camera data from connected autonomous vehicles. Our experiments focus on the vehicle-to-vehicle (V2V) scenario using paired ego and partner agent data.

We follow the standard DAIR-V2X split, consisting of 50,000 frames for training, 10,000 for validation, and 11,254 for testing. LiDAR point clouds are voxelized and projected according to the BM2CP preprocessing pipeline, with no additional filtering or downsampling.

C. Training Configuration

Each model was trained for 50 epochs using a batch size of 2 and a learning rate of 1×10^{-3} with the Adam optimizer. No learning rate scheduling or weight decay was applied. All models used the same configuration to enable fair comparisons across regimes. We used a fixed random seed for each training run to ensure reproducibility.

D. Inference Protocol

To evaluate robustness, we conducted inference tests for each model under Gaussian noise applied to the LiDAR input. We swept the standard deviation of the added noise from 0.00 to 0.80 in increments of 0.02, resulting in 40 noise levels per model checkpoint. This range of noise levels far exceeds normal point cloud tolerances, allowing for comprehensive stress testing. Inference was run at 10-epoch intervals (e.g., 10, 20, 30, 40, 50 epochs), yielding 200 evaluations per regime. With 7 regimes, this resulted in a total of 1,400 evaluations.

Average precision (AP) was computed at three standard IoU thresholds: 0.3, 0.5, and 0.7. These metrics are defined in Section III. Each trial was evaluated on the full DAIR-V2X test set with the same conditions to ensure comparability.

V. Results & Discussion

Our experiments demonstrate three key findings: (1) noise-aware training improves model robustness under degraded sensor inputs, (2) regimes that were gradually exposed to increasing levels of noise performed better than those that received constant noise levels across training epochs, and (3) regimes trained with noise converge faster, requiring fewer epochs to outperform baseline models.

Fig. 1 visualizes how performance degrades as Gaussian noise increases. At low noise levels (0.00–0.10 meters), all regimes perform similarly, with Light or No Noise training slightly outperforming others. However, as noise increases beyond 0.2 meters, the performance of the baseline model collapses rapidly, while models trained with progressively more noise maintain significantly higher accuracy. At 0.5 meters, the Heavy Noise model achieves 41.9 AP@0.3, 36.0 AP@0.5, and 14.0 AP@0.7, dramatically outperforming all other regimes.

Table II highlights this trend numerically. Across all IoU thresholds, we observe a consistent crossover pattern: Light Noise performs best at low noise, Moderate Noise leads at mid-range levels, and Heavy Noise dominates at high degradation. The uniformity of this progression across all three IoUs reinforces the generalizability of our approach. Moreover, these results demonstrate that robustness is not only a byproduct of noise, but something that can be directly learned and scaled with training-time exposure.

Furthermore, Fig. 2 illustrates the relative performance between progressive and constant noise regimes. Progressive training regimes perform better under higher levels of training noise, since they have the opportunity to form key associations on clean data before noise is introduced. However, low-amplitude constant regimes are more effective than low-amplitude progressive regimes when subjected to noisy data, indicating that training on too little noise hinders the development of noise-robust associations.

Additionally, model performance generally improves as the number of training epochs increases, as seen in Fig. 3. Noisier models also converge more rapidly. Fig. 4 compares the Heavy Noise model trained for only 20 epochs to the baseline model trained for 50 epochs. Despite using just 40% of the training time, the Heavy model consistently outperforms the baseline under moderate and high noise levels. This suggests that early exposure to noisy data promotes faster learning of generalizable features. In practical terms, this enables faster fine-tuning in deployment settings where compute or time budgets are limited.

Together, these findings suggest that noise-aware training provides a scalable, generalizable strategy for robust collaborative perception. By embracing input uncertainty during training, models become more resilient to sensor imperfections and better suited for deployment on low-cost or degraded hardware. This opens the door to safer, more reliable, and more affordable AV systems, especially in challenging real-world conditions.

Training Regime	AP@0.3						AP@0.5						AP@0.7					
	0.00	0.10	0.20	0.30	0.40	0.50	0.00	0.10	0.20	0.30	0.40	0.50	0.00	0.10	0.20	0.30	0.40	0.50
No Noise	69.7	68.2	60.0	41.6	19.8	7.5	64.4	63.3	54.8	36.7	16.1	5.1	49.4	47.0	34.3	16.7	4.5	0.8
Light Noise	**70.2**	69.2	64.5	49.6	27.4	10.6	**65.3**	64.3	59.5	44.7	23.1	7.8	**50.8**	49.1	40.3	22.7	7.5	1.5
Moderate Noise	69.7	**69.3**	67.6	60.7	43.0	21.9	64.7	**64.5**	62.7	55.3	37.7	17.7	49.6	**49.5**	45.6	32.5	16.0	4.8
Heavy Noise	69.2	69.1	**68.3**	**65.3**	**57.5**	**41.9**	64.2	64.1	**63.1**	**60.0**	**51.7**	**36.0**	47.9	48.6	**46.9**	**41.1**	**28.5**	**14.0**
Constant 0.05	**67.8**	**67.8**	63.3	53.7	37.4	18.7	**63.1**	**63.1**	58.5	48.0	31.3	13.5	**49.2**	**48.7**	40.7	25.2	9.9	2.5
Constant 0.10	66.2	65.6	64.4	57.4	39.2	19.6	61.4	61.0	59.4	52.0	33.8	15.6	46.4	46.6	43.0	31.3	14.8	4.6
Constant 0.20	67.2	67.5	**66.3**	**64.3**	**56.7**	**39.8**	62.2	62.5	**61.6**	**59.2**	**51.0**	**34.2**	45.3	46.3	**45.4**	**41.2**	**29.6**	**15.1**

TABLE II: Average Precision (AP) at IoU thresholds of 0.3, 0.5, and 0.7 for all training regimes under increasing inference-time noise. Results are reported at six levels of Gaussian noise (standard deviation in meters). Bold values indicate the highest performance at each noise level across all regimes.

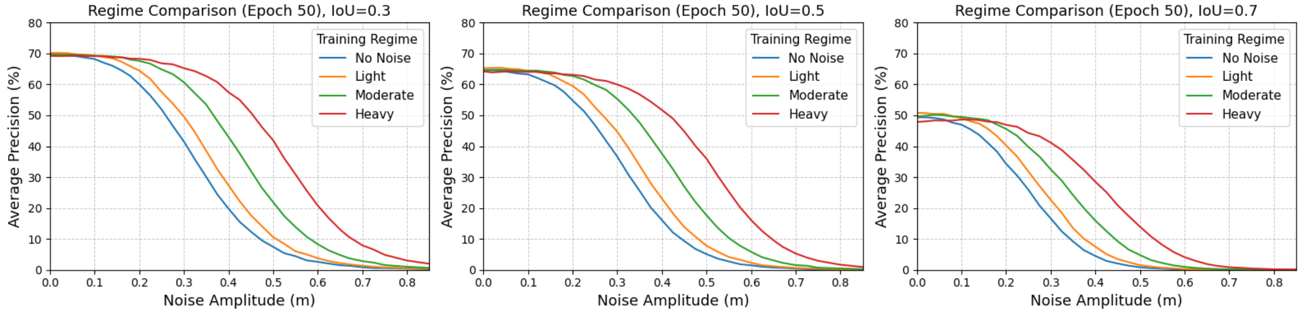

Fig. 1: Model robustness under increasing LiDAR noise at three IoU thresholds. Each curve represents a training regime evaluated at 50 epochs. From left to right, subplots show AP at IoU = 0.3, 0.5, and 0.7. Progressive noise regimes degrade more gracefully than constant or no-noise baselines.

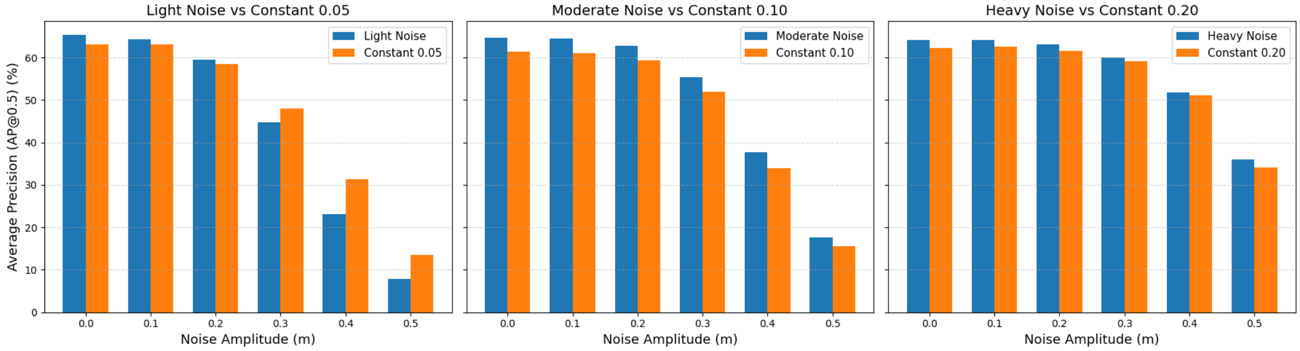

Fig. 2: Progressive noise exposure outperforms constant noise training. At each inference-time noise level, progressive regimes (blue) achieve higher AP@0.5 than constant-noise counterparts (orange), particularly at moderate to high degradation levels.

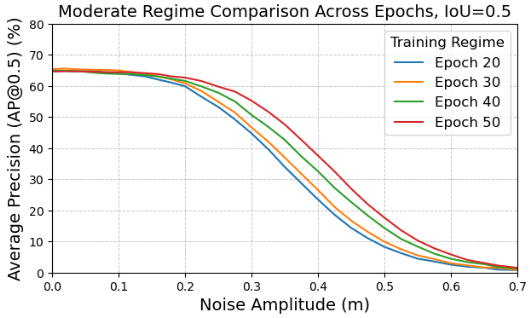

Fig. 3: Robustness improves over training epochs. At each inference-time noise level, we plot model performance across multiple training checkpoints (epochs). Models trained for more epochs tend to perform better than those trained for fewer epochs, especially at high inference-time noise levels.

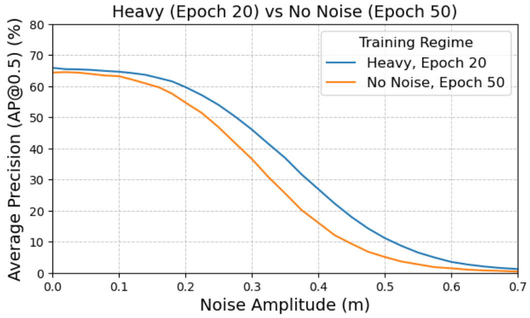

Fig. 4: Noise-trained models converge faster than baselines. Average Precision (AP@0.5) comparison between the Heavy Noise regime trained for 20 epochs and the baseline No Noise regime trained for 50 epochs. Despite reduced training time, the noise-aware model maintains superior robustness under moderate to high inference-time noise.

VI. Conclusion

We present a noise-aware training framework for improving the robustness and efficiency of collaborative 3D object detection systems. By injecting Gaussian noise into LiDAR point clouds during training, we enable perception models to generalize across a wide range of sensor conditions, including those produced by low-cost or degraded LiDAR hardware. Using the BM2CP framework and the real-world DAIR-V2X dataset, we evaluate both progressive (curriculum-style) and constant noise training regimes, comparing their ability to resist inference-time corruption.

Our findings show that models trained with progressively increasing noise degrade more gracefully under sensor corruption and converge faster than baseline models trained without noise. Notably, progressive noise regimes consistently outperform both no-noise baselines and constant-noise regimes at high degradation levels. In some cases, models trained for only 20 epochs with noise exposure match or exceed the performance of fully trained baselines, highlighting substantial gains in training efficiency.

Furthermore, our results reinforce the idea that robustness is not an incidental byproduct but a learnable and tunable property of the training process. The use of curriculum-style noise exposure enables smoother optimization and stronger generalization, especially in safety-critical scenarios where data quality cannot be guaranteed. This positions noise-aware training as a scalable strategy for deploying collaborative perception in cost-constrained or uncertain real-world environments.

In future work, we plan to explore alternative noise modalities, including non-Gaussian and adversarial perturbations, and extend our curriculum framework to additional 3D perception architectures beyond BM2CP. We also aim to develop formal metrics and adaptive scheduling policies for robustness-aware training. Together, these directions will contribute to a deeper understanding of how deep learning models can be made resilient to the imperfections that define real-world autonomy.

Acknowledgements

This work was supported in part by the National Science Foundation under Grant No. 2349370, and by the Data Science department at Worcester Polytechnic Institute. The authors of this paper would like to thank Elke Rundensteiner, Kelsey Briggs, and Ana Maria Young for their invaluable leadership and support through the National Science Foundation Research Experience for Undergraduates (NSF REU) in Applied Artificial Intelligence for Advanced Applications.

References

[1] R. Xu, H. Xiang, X. Xia, X. Han, J. Li, and J. Ma, “Opv2v: An open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication,” 2022. [Online]. Available: https://arxiv.org/abs/2109.07644

[2] H. Yu, Y. Luo, M. Shu, Y. Huo, Z. Yang, Y. Shi, Z. Guo, H. Li, X. Hu, J. Yuan, and Z. Nie, “Dair-v2x: A large-scale dataset for vehicle-infrastructure cooperative 3d object detection,” 2022. [Online]. Available: https://arxiv.org/abs/2204.05575

[3] Y. Hu, S. Fang, Z. Lei, Y. Zhong, and S. Chen, “Where2comm: Communication-efficient collaborative perception via spatial confidence maps,” 2022. [Online]. Available: https://arxiv.org/abs/2209.12836

[4] E. Richards, B. Thapa, and L. Mashayekhy, “Edge-enabled collaborative object detection for real-time multi-vehicle perception,” in *2025 IEEE International Conference on Edge Computing and Communications (EDGE), Helsinki, Finland*, 2025, pp. XX–YY.

[5] X. Feng, H. Sun, and H. Zheng, “Lcv2i: Communication-efficient and high-performance collaborative perception framework with low-resolution lidar,” 2025. [Online]. Available: https://arxiv.org/abs/2502.17039

[6] S. Crowe, “Waymo to stop selling lidar sensors,” 2021. [Online]. Available: https://www.therobotreport.com/waymo-ending-lidar-sales-to-other-companies/#:∼:text=Waymo%20began%20manufacturing%20its%20own,how%20many%20customers%20it%20had.

[7] D. Talwar, S. Guruswamy, N. Ravipati, and M. Eirinaki, “Evaluating validity of synthetic data in perception tasks for autonomous vehicles,” in *2020 IEEE International Conference On Artificial Intelligence Testing (AITest)*, 2020, pp. 73–80.

[8] Y. Han, H. Zhang, H. Li, Y. Jin, C. Lang, and Y. Li, “Collaborative perception in autonomous driving: Methods, datasets, and challenges,” *IEEE Intelligent Transportation Systems Magazine*, vol. 15, no. 6, pp. 131–151, 2023.

[9] S. Hu, Z. Fang, Y. Deng, X. Chen, and Y. Fang, “Collaborative perception for connected and autonomous driving: Challenges, possible solutions and opportunities,” 2025. [Online]. Available: https://arxiv.org/abs/2401.01544

[10] J. Mao, S. Shi, X. Wang, and H. Li, “3d object detection for autonomous driving: A comprehensive survey,” 2023. [Online]. Available: https://arxiv.org/abs/2206.09474

[11] D. Ma, R. Zhong, and W. Shi, “Icanc: Improving camera-based object detection and energy consumption in low-illumination environments,” 2025. [Online]. Available: https://arxiv.org/abs/2503.00709

[12] Y. Zeng, Y. Hu, S. Liu, J. Ye, Y. Han, X. Li, and N. Sun, “Rt3d: Real-time 3-d vehicle detection in lidar point cloud for autonomous driving,” *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3434–3440, 2018.

[13] L. Heng, B. Choi, Z. Cui, M. Geppert, S. Hu, B. Kuan, P. Liu, R. Nguyen, Y. C. Yeo, A. Geiger, G. H. Lee, M. Pollefeys, and T. Sattler, “Project autovision: Localization and 3d scene perception for an autonomous vehicle with a multi-camera system,” in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 4695–4702.

[14] B. Zhao, W. ZHANG, and Z. Zou, “Bm2cp: Efficient collaborative perception with lidar-camera modalities,” in *Proceedings of The 7th Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, 2023, pp. 1022–1035.

[15] J. He, X. Deng, J. Gui, T. Zhang, and X. He, “Mdnet: Multimodal cooperative perception via spatial alignment of modal decision-making,” *IEEE Internet of Things Journal*, vol. 12, no. 11, pp. 16 142–16 154, 2025.

[16] Y. Liu, G. Liu, L. Liang, H. Ye, C. Guo, and S. Jin, “Deep reinforcement learning-based user scheduling for collaborative perception,” 2025. [Online]. Available: https://arxiv.org/abs/2502.10456

[17] Z. Song, L. Liu, F. Jia, Y. Luo, G. Zhang, L. Yang, L. Wang, and C. Jia, “Robustness-aware 3d object detection in autonomous driving: A review and outlook,” 2024. [Online]. Available: https://arxiv.org/abs/2401.06542

[18] P. Soviany, R. T. Ionescu, P. Rota, and N. Sebe, “Curriculum learning: A survey,” 2022. [Online]. Available: https://arxiv.org/abs/2101.10382