COLLABORATIVE PERCEPTION FOR AUTONOMOUS DRIVING: FROM VIRTUAL TO REAL DATASETS

Judit Salavedra Pujol, Mingrui Wang, Javier Ruiz-Hidalgo, Josep R. Casas

Universitat Politècnica de Catalunya (UPC)

ABSTRACT

Cooperative Perception (CP) supports autonomous driving by enabling connected autonomous vehicles (CAVs) to share sensor data, thereby increasing situational awareness and safety. However, training CP models requires diverse traffic scenarios, and real-world datasets are limited due to the high cost of multi-agent setups and the labor-intensive nature of labeling. Synthetic datasets offer a scalable alternative, but models trained exclusively on virtual data often face domain adaptation challenges. This paper presents a systematic benchmark that evaluates two advanced 3D LiDAR-based object detectors on synthetic and real datasets using three strategies: training from scratch, transferring from synthetic to real data, and training with mixed datasets. The results show that pre-training on synthetic data followed by fine-tuning on real datasets yields the best performance (+2% AP@0.5, +2-4% AP@0.7 for car detection), while mixed training improves cross-domain generalization. These findings underscore the value of synthetic data as a complementary resource for developing robust CP systems.

***Index Terms*—** Autonomous Driving (AD), V2V Cooperative Perception (CP), 3D LiDAR-based Object Detection, Transfer Learning (TL), Simulation-to-Reality Domain Gap

1. INTRODUCTION

Perception of the environment is crucial for autonomous driving (AD) systems. Perception modules enable vehicles to sense their surroundings and enhance awareness by continuously collecting and processing raw data from onboard sensors. Growing concerns about road safety have intensified efforts to develop accurate perception systems for reliable real-world operation.

Single-vehicle systems perform well on real benchmarks but face inherent limitations, such as occlusions, limited sensor range, and sparser data at greater distances, which can significantly degrade performance in complex environments.

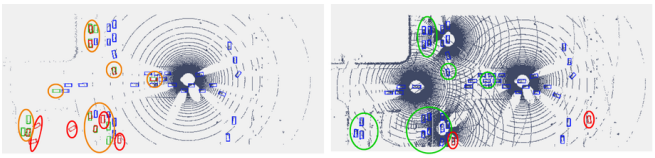
This work was partially supported by the C3DRUM project PID2024-161868OB-I00 funded by the Spanish MCIN/AEI/10.13039/501100011033 and FEDER(EU) and by the 6G-EWOC project from the Smart Networks and Services Joint Undertaking (SNS JU) under European Union's Horizon Europe research and innovation programme Grant Agreement No. 101139182.

Cooperative perception (CP) offers a promising solution to the limited perceptual capabilities of individual vehicles. By leveraging information shared among multiple road agents, such as sensor-equipped vehicles and infrastructure [1], CP improves collective awareness and driving safety. Common communication modalities include *V2V*, *V2I*, and *V2X* (V: vehicle, I: infrastructure, X: everything).

In a collaborative scenario, agents exchange and integrate information from other agents at different stages of the perception pipeline. The objective is not only to exchange isolated messages with localized observations (e.g. Bird-Eye-View detection coordinates and class labels), but also to jointly generate a unified detection map. Depending on the processing stage at which data is shared, approaches are categorized as follows [1, 2]: *Early fusion*, where agents communicate their raw sensor data directly (accurate, but extremely demanding in terms of bandwidth); *Late fusion*, where the perceptual results of individual agents are shared [3] (simpler communication, but at the cost of losing relevant information); and *Intermediate fusion*, where each agent extracts features from its on-board sensors and transmits them to other users (provides a balance by enabling efficient communication while retaining rich information).

Before real-world deployment, CP systems must be safely evaluated in virtual environments. Several datasets have recently been introduced to support various AD tasks. However, while many focus on single-vehicle scenarios, large datasets specifically designed for CP remain limited. Early CP datasets rely on data generated by traffic simulation platforms which provide large amounts of labeled data while reducing the cost and complexity of deploying multiple connected, fully equipped agents. Despite the strong performance of state-of-the-art (SotA) models in virtual scenarios, accuracy drops when moving to the real world due to discrepancies between simulation and reality. With the recent introduction of some large real CP datasets, increasing attention is being paid to bridging this synthetic-to-reality gap.

Building on current research, this study addresses a critical gap in CP. While the simulation-to-reality (sim2real) domain gap for single-vehicle 2D perception has been studied, it remains largely unexplored for 3D LiDAR-based perception, especially in multi-agent settings due to the scarcity and high collection costs of real collaborative datasets. We examine the



a) Single car (EGO) b) CAVs = 4

Fig. 1. 3D car detection results for a LiDAR-based scene in OPV2V. Bounding boxes indicate true positives (blue), false positives (red), and ground-truth without valid predictions (green). Circles highlight false positives (red), missed detections (orange), and improvements achieved with CP (green).

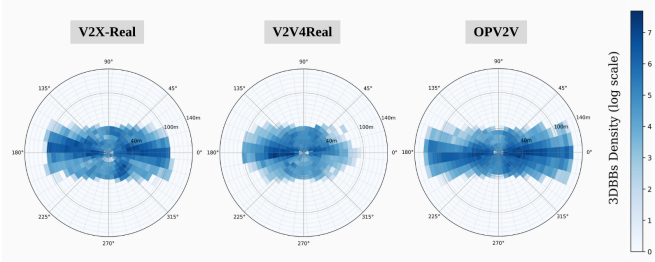


Fig. 2. Polar density maps of car 3DBBs over training sets. The radial and polar axes represent the distance and angle of 3DBBs relative to the ego vehicle. Color intensity represents density per bin (on a logarithmic scale).

utility and effectiveness of virtual data, analyzing the impact of the existing sim2real gap to provide a practical engineering roadmap for training in CP research.

To achieve this, this study compares synthetic and real datasets based on their data distributions and characteristics, and trains two SotA 3D LiDAR-based object detection models that support CP settings. We conduct extensive experiments on a large open-source synthetic dataset, OPV2V [4], and the only two currently available large open-source real datasets that support V2V communication: V2V4Real [5] and V2X-Real [6]. By testing under different training configurations, this study quantifies how synthetic data can complement real data in CP, revealing previously undocumented domain-specific generalization patterns. Specifically, we compare the performance of models trained from scratch on real data as a baseline with models fine-tuned after pre-training on virtual data and with models trained jointly on real and synthetic data. In addition, we systematically perform cross-data inference to assess the impact of the discrepancy between simulation and reality on performance, establishing a comprehensive triangular cross-dataset benchmark.

The paper is organized as follows: the next section reviews SotA models and datasets in CP and discusses strategies to address the sim2real gap. Section 3 provides a comparative analysis of the adopted datasets. Section 4 describes the experimental setup, training strategies, and results. Section 5 summarizes our key findings.

2. RELATED WORK

By aggregating information from multiple viewpoints, CP enables a more comprehensive and robust understanding of the scene, which is crucial for improving driving safety [4–7]. Fig. 1 presents an example where CP enhances 3D object detection capabilities compared to single-vehicle perception.

For our experiments, we use two SotA 3D LiDAR-based object detection models: *S-AdaFusion* [7] and *TS-IFF* [8]. Both extend PointPillars [9] and are implemented within the OpenCOOD platform [4], which natively supports CP. The common architecture underlying both models, comprises five stages: (i) *metadata sharing* between CAVs, *LiDAR preprocessing* to remove noisy points, and *projection* of point clouds into a unified coordinate system; (ii) *feature encoding* of raw point clouds through pillarization to generate 2D pseudo-images; (iii) *feature extraction* from pseudo-images to obtain mid-level features capturing richer semantic information; (iv) *intermediate feature fusion* to aggregate multi-agent features and create a shared representation of the environment; and (v) a *detection head* that performs object classification and 3D bounding box (3DBB) regression using predefined anchors. These models differ only in the mid-level fusion stage: S-AdaFusion applies spatially adaptive fusion, while TS-IFF uses a two-stage adaptive fusion mechanism.

To train and evaluate CP systems, extensive and diverse datasets are required. Real datasets are costly to collect, especially for CP research, as they require multiple agents deployed in real environments with high-quality sensors, and labeling is labor-intensive. Synthetic datasets have proven effective, less expensive, scalable and allow automatic labeling.

We examine the sim2real gap using both large-scale synthetic and real datasets. For synthetic data, we use OPV2V [4], one of the largest and widely used datasets for validating CP models. It provides simulated 64-beam LiDAR data from multiple CAVs across diverse environments, including CARLA's default towns (DT) and a digital replica of Culver City (CC). For real data, we use the only two public large datasets that support V2V communication: V2V4Real [5] and V2X-Real [6]. V2V4Real is the first large dataset designed for V2V settings, with 32-channel LiDARs on 2 vehicles driving on both highways and city streets. V2X-Real is the first dataset to support V2X, involving both CAVs and roadside units (RSUs). Data was recorded by 128-beam LiDAR sensors on 2 vehicles navigating through urban intersections and corridors, and from 128-beam LiDARs on 2 RSUs at urban intersections.

Motivated by the limitations of existing open-source datasets, especially in real settings that rarely provide multi-agent recordings beyond a small number of agents, this study focuses on V2V settings involving up to 2 CAVs.

Table 1. Comparison of statistics for the OPV2V, V2V4Real and V2X-Real datasets: $\mu \pm \sigma$ are calculated from preprocessed point clouds projected into the coordinate system of the ego-vehicle, discarding points and 3DBBs outside the perception range.

			3DBBs Density (#3DBBs/frame)	Point-cloud size (#points/frame $\times 10^3$)		LiDAR points inside 3DBBs (#points/3DBB)		Dimensions of 3DBBs [m]		
				1 CAV	2 CAVs	1 CAV	2 CAVs	L	W	H
OPV2V	Train		15 ± 8	53.8 ± 1.9	96.7 ± 20.2	341 ± 945	$597 \pm 1{,}243$	4.6 ± 0.5	2.0 ± 0.1	1.6 ± 0.1
	Val		24 ± 14	52.8 ± 2.6	88.3 ± 22.8	325 ± 921	$498 \pm 1{,}136$	4.6 ± 0.5	2.0 ± 0.1	1.6 ± 0.1
	Test	DT	14 ± 8	53.8 ± 1.9	99.6 ± 18.1	346 ± 999	$582 \pm 1{,}269$	4.5 ± 0.5	2.0 ± 0.1	1.6 ± 0.1
		CC	38 ± 10	54.6 ± 0.6	84.7 ± 24.5	269 ± 805	384 ± 962	4.5 ± 0.5	2.0 ± 0.1	1.6 ± 0.1
V2V4Real	Train		10 ± 7	32.0 ± 8.7	56.4 ± 17.0	220 ± 251	247 ± 285	4.6 ± 0.5	2.0 ± 0.2	1.8 ± 0.3
	Val		10 ± 5	26.3 ± 3.6	49.0 ± 11.2	189 ± 251	224 ± 315	4.9 ± 0.7	2.2 ± 0.2	1.9 ± 0.3
	Test		12 ± 9	35.7 ± 5.8	63.9 ± 14.4	328 ± 325	423 ± 529	4.7 ± 0.6	2.1 ± 0.3	1.8 ± 0.3
V2X-Real	Train	64	13 ± 6	69.8 ± 1.1	98.0 ± 33.6	340 ± 527	430 ± 618	4.6 ± 0.4	2.0 ± 0.2	1.7 ± 0.2
		128		202.8 ± 11.2	285.0 ± 98.4	611 ± 921	$787 \pm 1{,}066$			
	Val	64	14 ± 6	68.9 ± 3.4	107.4 ± 36.0	303 ± 416	423 ± 554	4.6 ± 0.4	2.1 ± 0.2	1.7 ± 0.2
		128		204.9 ± 11.1	318.8 ± 104.0	550 ± 690	847 ± 972			
	Test	64	19 ± 10	70.1 ± 1.3	128.7 ± 24.4	373 ± 568	537 ± 754	4.6 ± 0.4	2.1 ± 0.2	1.7 ± 0.2
		128		205.1 ± 13.4	376.6 ± 73.5	648 ± 904	$963 \pm 1{,}222$			

2.1. Bridging the Simulation-to-Real Datasets Gap

Most SotA models are developed and evaluated on specific datasets, assuming that training and deployment data share the same distribution [10, 11]. However, existing datasets cannot capture the full complexity and variability of the real world. Virtual datasets enable controlled experimentation and efficient labeling, but lack realism [5]. In contrast, real datasets naturally include sensor errors and reflect realistic traffic dynamics, but are limited in scope and variety. As a result, these datasets often fail to capture real-world variations and models tend to overfit to the domain-specific characteristics of their training data [11]. This performance degradation is primarily caused by the *domain gap* resulting from discrepancies in sensor types, scene layouts, agent diversity, perception architectures and data distributions [10].

Deep *transfer learning* (TL) has emerged as a powerful approach to improve the generalization capability of CP models, enhancing adaptability and robustness under different operating conditions. TL methods are broadly categorized into supervised, unsupervised [5, 12], weakly supervised and semi-supervised, along with domain adaptation and domain generalization [10, 13] strategies, each offering different ways to address data scarcity and domain shift challenges.

Recent work has also explored the use of synthetic data in *fine-tuning* strategies or as a *data augmentation* tool to improve generalization, with most efforts targeting single-vehicle 2D object detection, such as [14]. The authors show that pre-trained weights improve generalization on unseen datasets, while training exclusively on synthetic data provides no significant benefit. Combining synthetic and real data consistently enhances performance. Similarly, [15] highlights that models trained only on synthetic images do not generalize, whereas balanced integration of synthetic data (e.g. 25–33%) improves performance on real benchmarks. Beyond 2D perception, [16] evaluates the impact of synthetic data in both 2D and 3D single-vehicle object detection, and shows that combining real and synthetic images improves generalization. However in 3D LiDAR, it may increase cross-dataset robustness at the expense of reduced in-domain performance.

3. DATASET CHARACTERISTICS AND ANALYSIS

Table 1 presents a comparative overview of the statistics for LiDAR point clouds and 3DBBs in the datasets used in this study: OPV2V, V2V4Real and V2X-Real. To ensure a consistent and fair comparison, all statistics are derived from preprocessed LiDAR point clouds, where 3DBBs and LiDAR points outside the ego-vehicle's perception range are filtered out, as is done during training and evaluation.

Point cloud and 3DBB LiDAR density. V2X-Real's 128-beams provide the highest point density and spatial coverage, while its 64-beam version produces sparser clouds comparable to OPV2V's virtual 64-beam data. V2V4Real uses 32-beams, resulting in the lowest density. Consistently, 3DBB LiDAR point density is highest with V2X-Real's setup, and increases under CP due to the additional viewpoint.

Density, spatial distribution and size of 3DBBs. OPV2V provides more annotations per frame than V2V4Real, but its

Table 2. **S-AdaFusion** 3D Object Detection Performance under different training strategies

		Test							
		OPV2V				V2V4Real		V2X-Real	
		DT		CC					
		AP@.5	AP@.7	AP@.5	AP@.7	AP@.5	AP@.7	AP@.5	AP@.7
Train Strategy	OPV2V (S)	90.5	81.6	84.1	72.7	37.2	13.7	39.6	16.1
	V2V4Real (S)	35.7	23.3	29.5	19.3	62.4	34.8	54.0	36.0
	OPV2V (P) → V2V4Real (F)	9.4	7.2	6.1	4.2	**64.3**	**38.9**	63.0	41.2
	OPV2V (100%) + V2V4Real (100%)	90.6	83.0	85.3	74.2	**65.1**	**39.6**	63.7	48.7
	V2X-Real (S)	73.9	56.8	65.1	48.7	59.3	31.3	76.6	60.7
	OPV2V (P) → V2X-Real (F)	71.3	58.1	58.5	46.6	59.1	33.1	**78.2**	**64.2**
	OPV2V (100%) + V2X-Real (100%)	89.4	81.5	86.2	75.2	**66.2**	36.3	74.7	61.3

Note: *DT* = Default Towns; *CC* = Culver City; **Training Strategies:** *S* = From scratch, *P* = Pre-training, *F* = Fine-tuning; **Symbols:** → = two-phase training, + = mixed dataset; **Colours:** *grey* = trained on synthetic data only, *green* = trained using at least V2V4Real, *blue* = trained using at least V2X-Real.

Table 3. **TS-IFF** 3D Object Detection Performance under different training strategies (See note in Table 2)

		Test							
		OPV2V				V2V4Real		V2X-Real	
		DT		CC					
		AP@.5	AP@.7	AP@.5	AP@.7	AP@.5	AP@.7	AP@.5	AP@.7
Train Strategy	OPV2V (S)	90.0	79.9	85.2	70.6	36.9	12.6	44.3	18.4
	V2V4Real (S)	26.6	17.6	22.9	15.9	60.6	33.5	52.8	35.3
	OPV2V (P) → V2V4Real (F)	25.0	18.3	21.0	14.9	**62.8**	**35.0**	61.8	43.8
	OPV2V (100%) + V2V4Real (100%)	88.4	78.8	86.1	72.2	53.1	26.3	46.1	32.0
	V2X-Real (S)	75.6	52.9	64.7	43.8	54.2	30.1	72.6	56.3
	OPV2V (P) → V2X-Real (F)	78.5	60.3	68.1	51.4	62.3	34.4	**73.9**	**58.9**
	OPV2V (100%) + V2X-Real (100%)	88.6	78.2	85.2	71.4	48.8	22.7	56.3	41.1

subsets vary considerably. The validation set contains 10-35 cars per frame, while the CC test set is even denser with around 38 cars per frame. In contrast, V2V4Real consistently contains fewer 3DBBs. V2X-Real is more similar to OPV2V and has slightly more 3DBBs than V2V4Real. The spatial distribution of 3DBBs also shows important differences. Polar density maps in Fig. 2 show that OPV2V and V2X-Real provide a balanced distribution of objects over radial distances and angles relative to the ego vehicle, with similar patterns and annotation densities. In contrast, V2V4Real displays an irregular distribution, with most annotations located behind the ego vehicle. For 3DBBs, all datasets have comparable length, width, and height ranges, suggesting that GT dimensions are unlikely to contribute to domain gap effects.

4. EXPERIMENTAL SETUP AND RESULTS

We use the detection models codebases and datasets released by the original authors, built on top of the OpenCOOD [4] framework. Each experiment runs on a single GPU, either a GeForce GTX 1080 Ti or a GeForce RTX 2080 Ti, each with 11 GB of memory. The software environment uses CUDA v11.3 and the cuDNN v8.3.2 library.

Experiments follow three training strategies: (i) *training on single datasets*; (ii) *sim2real knowledge transfer*; and (iii) *training on mixed datasets*, with batches balanced at a 1:1 sim2real ratio. To ensure a controlled environment, all experiments use the OpenCOOD deterministic protocol (fixed random seed) and identical hyperparameters. Our analysis first evaluates the Average Precision (AP) of S-AdaFusion (Table 2) across the main strategies, both on the training set and by cross-dataset inference at 0.5 and 0.7 Intersection-over-Union thresholds. We then analyse repeated experiments on TS-IFF (Table 3) to confirm these trends. Additional details and experiments are available in the extended report [17].

Single-dataset training. We first analyze *single-dataset setups* to establish a baseline with models trained on individual datasets without prior knowledge from others. The model

trained on synthetic data (OPV2V) performs strongly on synthetic test sets (90.5% AP on DT). However, on both real test sets (V2V4Real and V2X-Real), performance drops to nearly 40% AP@.5 and 15% AP@.7, illustrating the sim2real domain gap. Models trained from scratch on real data also perform well within the same source domain, but with lower scores than those of models trained and tested on synthetic data, reflecting the higher complexity of real data. Cross-domain evaluation, reveals contrasting generalization patterns. The model trained on V2V4Real shows weaker generalization, indicating strong overfitting to its own training distribution, while the V2X-Real model generalizes more effectively. This improved transferability can be attributed to a greater variety of traffic patterns captured in V2X-Real. Additionally, V2V4Real contains fewer and longer scenes, likely increasing redundancy in the training data. The analysis in Section 3 further supports this: OPV2V and V2X-Real have more similar car distributions in terms of density and spatial arrangement, while V2V4Real contains fewer vehicles per frame and a stronger concentration on the rear side.

Sim2real knowledge transfer. Next, we analyze the *two-stage strategy* which uses synthetic data as a strong initialization (pre-training) before fine-tuning on real datasets. Fine-tuning on V2V4Real or V2X-Real improves accuracy on their respective test sets, with AP@.5 and AP@.7 increasing by about $+2\%$ and $+4\%$, respectively, compared to training from scratch. Domain transfer across real datasets is slightly beneficial: fine-tuning on V2V4Real improves V2X-Real performance by $7-9\%$ AP relative to the model trained solely on V2V4Real and fine-tuning on V2X-Real maintains or slightly increases accuracy on V2V4Real. A consequence of this strategy is catastrophic forgetting, as model performance on synthetic data drops significantly after real fine-tuning, as observed after fine-tuning on V2V4Real. However, since our goal is strong performance on real data, synthetic pre-training is valuable providing a better initialization than random weights and offering transferable knowledge that enhances learning across real domains.

Mixed-dataset training. Finally, we explore *joint training* on mixed synthetic and real data from scratch to leverage complementary information from both domains, encouraging the model to learn domain-independent features. To assess the impact of the balance of synthetic and real data, we explore different ratios of OPV2V and real samples: 25%, 50% and 100% synthetic samples, all combined with a 100% real dataset. Results for the *100%+100%* setup are shown in Table 2, as it achieves the highest accuracy across all ratios. Additional results for *50%* and *25%* ratios can be found in [17]. Joint training with *OPV2V+V2V4Real* yields consistent performance gains on most datasets compared to training on V2V4Real alone. The combination of *OPV2V+V2X-Real* improves cross-domain robustness and benefits OPV2V and V2V4Real scores, although we observe a small in-domain drop on V2X-Real itself.

Evaluation of TS-IFF across training strategies. To verify the consistency of observed trends and assess model sensitivity across training strategies, we repeated all experiments using the TS-IFF model (results in Table 3). The same general trends appear for single-dataset training and fine-tuning: strong in-domain performance but poor transfer when training on OPV2V, and consistent improvements by around $+2\%$ AP on both real test sets by fine-tuning. However, when mixing synthetic and real data, instead of gaining robustness, TS-IFF struggles, likely due to its higher model complexity, as it employs complex multi-stage mid-level modules and dynamic weights. When jointly trained on disparate distributions (virtual 64 vs. real 128/32-beam), its complexity makes joint training less stable and results in weaker performance.

Other observations and insights. Some additional performance trends are evident in the reported results. First, performance on the *CC test set* is consistently lower than on the DT set across all models and training strategies. This is likely because CC scenes are denser with significantly more objects per frame (an average of 38 cars vs. 14 cars per frame in DT), making detection more challenging. Another notable trend across all experiments is that performance on V2V4Real is consistently lower than on V2X-Real, even with the best training settings. With S-AdaFusion, models trained on V2X-Real achieve $75-78\%$ AP@0.5, while V2V4Real does not exceed 66% AP@.5. This discrepancy may be due to differences in sensor resolution: V2X-Real 128-beam LiDAR vs. V2V4Real 32-beam. This leads to much denser point clouds and richer information in V2X-Real, which directly supports stronger detection performance as supported by Table IX in [17]. Applying the same trained model to identical scenes, merely reducing the input resolution from 128 to 64 beams causes a drop of $10-20\%$ AP@.5.

5. CONCLUSIONS

The experimental results provide clear evidence that synthetic data, while unable to replace real data, can play a valuable role in advancing CP in real settings. Our systematic empirical evaluation reveals three key insights: (i) pure synthetic training fails to generalize to real domains, a two-stage strategy using synthetic pre-training followed by real fine-tuning proves to be the most reliable approach, consistently outperforming training from scratch on real datasets ($+2\%$ AP@.5); (ii) joint training on mixed domains shows that virtual and real data can serve as complementary sources, improving robustness and cross-domain generalization for simpler models like S-AdaFusion, yet causing optimization instability in complex models like TS-IFF; and (iii) performance differences across datasets demonstrate the substantial impact of LiDAR resolution, realistic point cloud data, 3DBB distribution, and scene complexity, highlighting the importance of large, diverse datasets and the need for virtual data that better replicate real technologies.

6. REFERENCES

[1] 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.

[2] Yushan Han, Hui Zhang, Huifang Li, Yi Jin, Congyan Lang, and Yidong Li, "Collaborative perception in autonomous driving: Methods, datasets, and challenges," *IEEE Intelligent Transportation Systems Magazine*, vol. 15, no. 6, pp. 131–151, 2023.

[3] Chenguang Liu, Jianjun Chen, Yunfei Chen, Yubei He, Zhuangkun Wei, Hongjian Sun, Haiyan Lu, and Qi Hao, "Coop-WD: Cooperative perception with weighting and denoising for robust V2V communication," *arXiv preprint arXiv:2505.03528*, 2025.

[4] 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)*. IEEE, 2022, pp. 2583–2589.

[5] Runsheng Xu, Xin Xia, Jinlong Li, Hanzhao Li, Shuo Zhang, Zhengzhong Tu, Zonglin Meng, Hao Xiang, Xiaoyu Dong, Rui Song, et al., "V2v4real: A real-world large-scale dataset for vehicle-to-vehicle cooperative perception," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 13712–13722.

[6] Hao Xiang, Zhaoliang Zheng, Xin Xia, Runsheng Xu, Letian Gao, Zewei Zhou, Xu Han, Xinkai Ji, Mingxi Li, Zonglin Meng, et al., "V2x-real: a large-scale dataset for vehicle-to-everything cooperative perception," in *European Conference on Computer Vision*. Springer, 2024, pp. 455–470.

[7] Donghao Qiao and Farhana Zulkernine, "Adaptive feature fusion for cooperative perception using LiDAR point clouds," in *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2023, pp. 1186–1195.

[8] Mingrui Wang, Dongjie Li, Josep R Casas, and Javier Ruiz-Hidalgo, "Adaptive fusion of lidar features for 3d object detection in autonomous driving," *Sensors*, vol. 25, no. 13, pp. 3865, 2025.

[9] Alex H Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom, "PointPillars: Fast encoders for object detection from point clouds," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2019, pp. 12689–12697.

[10] Baolu Li, Zongzhe Xu, Jinlong Li, Xinyu Liu, Jianwu Fang, Xiaopeng Li, and Hongkai Yu, "V2X-DG: Domain generalization for vehicle-to-everything cooperative perception," *arXiv preprint arXiv:2503.15435*, 2025.

[11] Senkang Hu, Zhengru Fang, Yiqin Deng, Xianhao Chen, Yuguang Fang, and Sam Kwong, "Toward full-scene domain generalization in multi-agent collaborative Bird's Eye View segmentation for connected and autonomous driving," *IEEE Transactions on Intelligent Transportation Systems*, 2024.

[12] Hongbo Yin, Daxin Tian, Chunmian Lin, Xuting Duan, Jianshan Zhou, Dezong Zhao, and Dongpu Cao, "CUDA-X: Unsupervised domain-adaptive vehicle-to-everything collaboration via knowledge transfer and alignment," *IEEE Transactions on Neural Networks and Learning Systems*, 2025.

[13] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon, "Dynamic graph cnn for learning on point clouds," *ACM Transactions on Graphics (tog)*, vol. 38, no. 5, pp. 1–12, 2019.

[14] Miguel Antunes, Luis M Bergasa, Javier Araluce, Rodrigo Gutiérrez, J Felipe Arango, and Manuel Ocaña, "Including transfer learning and synthetic data in a training process of a 2d object detector for autonomous driving," in *Iberian Robotics conference*. Springer, 2022, pp. 465–478.

[15] Jérémy Jaspar, Emmanuel Viennet, David Gualandris, Jean-Louis Sauvaget, and Françoise Fogelman-Soulié, "Using synthetic images to improve and test object detection in the context of the autonomous vehicle," in *2024 12th European Workshop on Visual Information Processing (EUVIP)*. IEEE, 2024, pp. 1–6.

[16] Enes Özeren and Arka Bhowmick, "Evaluating the impact of synthetic data on object detection tasks in autonomous driving," *arXiv preprint arXiv:2503.09803*, 2025.

[17] J. Salavedra, "Collaborative perception for autonomous driving: from virtual to real datasets.," Research report, GPI-UPC Research Rept., 2025.