OpenCDA-∞: A Closed-loop Benchmarking Platform for End-to-end Evaluation of Cooperative Perception

Chia-Ju Chen
UCLA
ju40268@ucla.edu

Runsheng Xu
UCLA
rxx3386@g.ucla.edu

Wei Shao
UC Davis
wei.shao@data61.csiro.au

Junshan Zhang
UC Davis
jazh@ucdavis.edu

Zhengzhong Tu
Texas A&M University
tzz@tamu.edu

Abstract

Vehicle-to-vehicle (V2V) cooperative perception systems hold immense promise for surpassing the limitations of single-agent lidar-based frameworks in autonomous driving. While existing benchmarks have primarily focused on object detection accuracy, a critical gap remains in understanding how the upstream perception performance impacts the system-level behaviors—the ultimate goal of driving safety and efficiency. In this work, we address the crucial question of how the detection accuracy of cooperative detection models natively influences the downstream behavioral planning decisions in an end-to-end cooperative driving simulator. To achieve this, we introduce a novel simulation framework, **OpenCDA**-∞, that integrates the OpenCDA cooperative driving simulator with the OpenCOOD cooperative perception toolkit. This feature bundle enables the holistic evaluation of perception models by running any 3D detection models inside OpenCDA in a real-time, online fashion. This enables a closed-loop simulation that directly assesses the impact of perception capabilities on safety-centric planning performance. To challenge and advance the state-of-the-art in V2V perception, we further introduce the **OPV2V-Safety** dataset, consisting of twelve challenging and pre-crash open scenarios designed following the National Highway Traffic Safety Administration (NHTSA) reports. Our findings reveal that OPV2V-Safety indeed challenges the prior state-of-the-art V2V detection models, while our safety benchmark yielded new insights on evaluating perception models as compared to the results on prior standard benchmarks. We envision that our end-to-end, closed-loop benchmarking platform will drive the community to rethink how perception models are being evaluated at the system level for the future development of safe and efficient autonomous systems. The code and benchmark will be made publicly available.

1 Introduction

Accurate, robust, and rapid perception of complex and dynamic environments is essential for responsible autonomous driving. Recent advances in robotic sensing equipped with advanced machine learning techniques have fueled perception performance, evidenced by successes in tasks such as 3D object detection, tracking, and semantic map segmentation. However, these advancements often falter in scenarios featuring extensive occlusions, small or distant objects, potentially leading to

Submitted to the 38th Conference on Neural Information Processing Systems (NeurIPS 2024) Track on Datasets and Benchmarks. Do not distribute.

catastrophic outcomes due to insufficient sensor data coverage, which underscores the challenges inherent in single-vehicle perception systems limited by physical constraints and occlusions.

To overcome these limitations, recent studies have pivoted towards multi-vehicle cooperative frameworks that leverage Vehicle-to-Everything (V2X) or Vehicle-to-Vehicle (V2V) communication technologies. These frameworks empower Connected and Automated Vehicles (CAVs) to share a diversity of data forms—from raw sensor outputs like LiDAR point clouds, RGB images, and radar frames to processed features and detection results—thereby collaboratively enhancing perception capabilities by amalgamating multiple vehicular perspectives. Despite their potential, these technologies' evolution is predominantly driven by the development of diverse, large-scale, and open-sourced datasets and benchmarks. For instance, initiatives such as OPV2V [1], V2X-ViT [2], and V2X-Sim [3] have mainly utilized simulation platforms like CARLA [4] and SUMO [5] to create extensive synthetic datasets tailored for cooperative perception tasks. Yet, traditional evaluations using metrics like Average Precision (AP) fall short of capturing the full spectrum of autonomous driving requirements, particularly in ensuring safe driving behaviors and robust vehicular planning.

To this end, here we introduce a novel framework that marries OpenCDA cooperative driving co-simulation platform [6] and the OpenCOOD cooperative perception toolset [1], which we dub **OpenCDA**-∞, allowing for holistic development and testing of cooperative perception models in a closed-loop, end-to-end fashion that mainly focuses on safety-centric evaluation. In other words, we can directly assess how the perception performance of V2V algorithms impacts the actual driving behavior and safety implications of the vehicles. To achieve this, we have made several enhancements to the OpenCDA simulation platform. First, we incorporated the OpenSCENARIO standard [7, 8] for precise actor controls (vehicles, pedestrians, etc.) in simulation, leading to more realistic and customizable scenarios. Second, we've added the capability to run cooperative perception models in real-time during simulation, enabling a true closed-loop evaluation. Third, we've integrated advanced modules for vehicle trajectory prediction and robust behavior planning to ensure that the ego vehicle (the one we're controlling) makes intelligent decisions based on the perceived information.

Moreover, We build the **OPV2V-Safety** dataset, comprising twelve diverse and challenging pre-crash traffic scenarios cataloged by the National Highway Traffic Safety Administration (NHTSA) [9], tailored to test the robustness of V2V perception and planning algorithms under adverse conditions. This dataset, featuring 4,377 frames, serves as a critical testbed for evaluating state-of-the-art 3D object detection techniques and multi-vehicle fusion strategies from a planning perspective. we move beyond standard detection accuracy metrics and introduce a multi-tiered safety-critical evaluation suite. Our metrics encompass not only the quality of object detection but also the robustness, efficiency, and stability of the overall cooperative perception system. This holistic approach provides a deeper understanding of how different perception models impact autonomous vehicles' system-level performance and safety. Our extensive benchmarking results on OPV2V-C, using various V2V algorithms, reveal that models that excel in traditional detection accuracy metrics do not necessarily lead to the best planning outcomes or the safest driving behaviors. This underscores the importance of our system-level evaluation approach and the value of the OPV2V-C dataset in driving the development of more robust and safety-conscious V2V autonomous driving systems.

In summary, our contributions are manifold: ❶ We propose a closed-loop, end-to-end simulation platform called **OpenCDA-**∞ that facilitates the *planning-oriented* evaluation of cooperative perception models at a system level. ❷ We extend the capabilities of OpenCDA with advanced functionalities, including realistic scenario customization and robust behavior planning, enabling real-time, online simulation, and comprehensive assessment of any perception models. ❸ We release the **OPV2V-Safety** dataset, a safety-critical testbench comprising diverse corner-case scenarios that can rigorously test existing V2V perception models and planning algorithms, which can facilitate the development of more safety-critical autonomous systems. ❹ A multi-tiered safety evaluation metric suite beyond traditional detection metrics has been provided, offering deeper insights into the safety and effectiveness of cooperative perception systems. ❺ Our extensive benchmarking results on state-of-the-art cooperative perception models highlight the importance of our benchmarking platform in regard to system-level evaluation of V2V perception.

2 Related Work

Autonomous driving datasets. Publicly available, large-scale datasets always play a fundamental role in advancing any machine learning field, and autonomous driving is no exception. The pioneering KITTI dataset [10], a trailblazer in providing multimodal sensor data, marked a significant leap towards data-driven autonomous learning with its front-facing stereo cameras and LiDAR across 22 sequences. Subsequent community efforts have escalated the scale and complexity of KITTI, including diversity in driving scenarios, sensor modalities, and data annotations that can be employed to train larger, multimodal algorithms for diverse vision and planning tasks. For example, the NuScenes [11] and Waymo Open dataset [12] are two representative multimodal datasets that consist of a significantly broader array of annotated RGB images and LiDAR point clouds, enabling more performant and robust vehicle and pedestrian detection models.

End-to-end autonomous driving. Significant progress has been made in end-to-end autonomous driving. UniAD [13] integrated full-stack driving tasks in a single network with query-unified interfaces. ReasonNet [14] improved perception by leveraging temporal and global scene information for better occlusion detection. ASAP-RL [15] proposed an efficient reinforcement learning algorithm for autonomous driving that simultaneously leverages motion skills and expert priors. Coopernaut [16] enhanced V2V cooperative driving with cross-vehicle perception and vision-based decision-makin. LMDrive [17] incorporated large language models, enabling natural language interaction and improving reasoning in complex scenarios. Approaches like Latent DRL [18] and Roach [19] utilized reinforcement learning to enhance decision-making, while ScenarioNet [20] and TrafficGen [21] generated diverse driving scenarios for testing. However, this end-to-end driving automation merely focuses on single-agent-based approaches, and a system that incorporates cooperative detection methods in a closed-loop simulator is in pressing need.

V2X/V2V cooperative systems and datasets. Despite the rapid progress in single-vehicle autonomous driving, it still encounters substantial challenges in complex real-world scenarios, such as extreme occlusions and limited long-range perception capabilities [22]. Recent advancements in Vehicle-to-Everything (V2X) (including Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I)) technologies have enabled vehicles to connect, communicate, and collaborate, significantly expanding their perception range as well as compensating each other to collaboratively handling occlusion via shared viewpoints. OPV2V [1] paves the way by constructing a novel 3D cooperative detection dataset using CARLA and OpenCDA co-simulation. Other studies like V2X-ViT [2] and V2X-Sim [3] leverage the capabilities of smart infrastructure in conjunction with connected vehicles to enable Vehicle-to-Everything (V2X) perception. In contrast to these simulated datasets, DAIR-V2X [23] and V2V4Real [24] provide large-scale real-world data for cooperative detection research, establishing benchmarks on realistic and dynamic traffic scenarios.

V2X/V2V cooperative perception models. Cooperative systems have emerged as powerful tools for addressing the inherent limitations of single-vehicle perception, enabling a paradigm shift towards multi-vehicle detection. The landscape of V2V and Vehicle-to-Everything (V2X) cooperative perception can be broadly segmented into three categories: ❶ *Early Fusion*, where raw point clouds are shared among Connected Autonomous Vehicles (CAVs), allowing the ego vehicle to draw predictions based on the assembled raw data [25]; ❷ *Late Fusion*, where detection outputs (e.g., 3D bounding boxes, confidence scores) are exchanged, which are subsequently fused into a single 'consensus' prediction [26]; and ❸ *Intermediate Fusion*, where intermediate feature maps or representations are derived from each agent's observation and then shared among the other CAVs [22, 27, 1, 28]. These categories encapsulate the diverse ways in which cooperative systems can be leveraged to enhance the breadth and depth of perception in autonomous driving.

Recent frontier cooperative detection models predominantly adopt intermediate fusion strategies where the intermediate neural features computed from each agent's sensor data are broadcasted, achieving the best trade-off between accuracy and bandwidth requirements. Specific examples include F-Cooper [27], which devises a simple max-pooling operation to fuse intermediate visual features, while V2VNet [22] employs graph neural networks to fuse shared features from connected

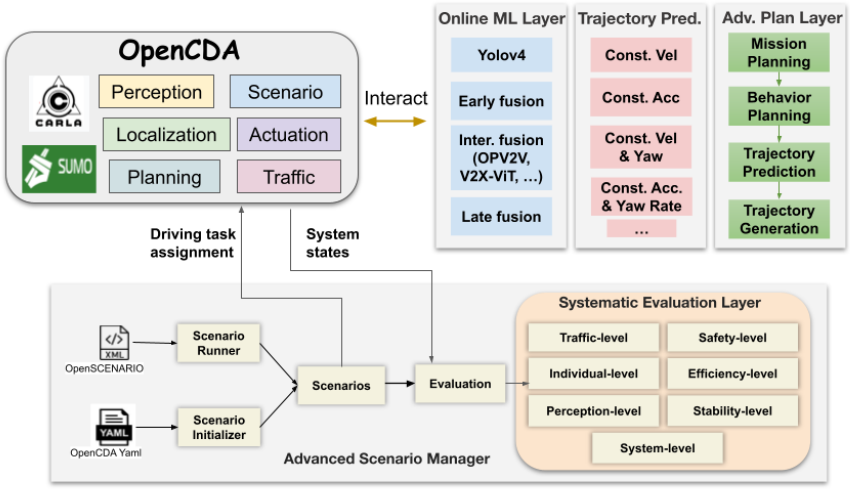


Figure 1: **OpenCDA-∞**: a closed-loop, end-to-end simulation platform that bridges two software suites: the cooperative driving simulation platform OpenCDA and the cooperative perception toolkit OpenCOOD. We further enhance this platform with advanced modules, including OpenSCENARIO customization (Sec. 3.3), online cooperative detection (Sec. 3.2), trajectory prediction and planning (Sec. 3.4). Finally, we build **OPV2V-Safety**, a challenging, pre-crash scene dataset, equipped with a spectrum of evaluation metrics for examining cooperative perception models.

vehicle nodes. Additionally, Coopernaut [16] uses Point Transformer [29] to process shared point features, CoBEVT [30] introduces an innovative local-global sparse attention mechanism that captures spatial interactions among different views and agents, and AttFuse [1] suggests an agent-specific self-attention module to fuse the received features. V2X-ViT [2] designs a unified vision transformer optimized for multi-agent, multi-scale perception, delivering robust performance even under conditions of GPS error and communication delay.

3 *OpenCDA-∞*: An Online, Closed-loop, End-to-end Simulator

3.1 OpenCDA Simulation Platform

OpenCDA [6] is a simulation-integrated framework for dynamic cooperative driving automation (CDA) research, which supports a broad range of automated vehicle interactions through a benchmarking scenario database and trending CDA algorithms. As illustrated in Fig. 1 , OpenCDA coherently integrates several core components: simulation tools, a Python-based CDA system, and an extensive scenario manager. For the simulation tools, OpenCDA utilizes CARLA [4], a free open-source driving simulator that boasts high-quality rendering capabilities powered by the Unreal Engine. The scenario manager of OpenCDA is structured into four main elements: the configuration file, initializer, event trigger, and evaluation functions. Scenarios blend static elements, such as road structures defined by CARLA's assets, with dynamic features managed by a YAML config file. Central to its design is the application layer, where CAVs exchange data and strategies, such as blending individual and communal sensing data for improved perception. OpenCDA provides both default and customizable protocols, enabling researchers to evaluate the entire CDA system or to conduct comparative analyses of specific algorithms. We refer the readers to the Appendix for more details regarding OpenCDA.

3.2 Online Cooperative Detection

Most simulation platforms available today, including OpenCDA, do not support the real-time operation of trained models; instead, they heavily rely on the offline evaluation of detection accuracy. This static approach fails to reflect the dynamic interplay that occurs in real-world driving scenarios—

where detection results continuously influence downstream planning and decision-making and, in turn, further determine the next system state for perception. This process creates a feedback loop that dynamically evolves based on real-time data and interactions. Unfortunately, current driving simulation benchmarks [1, 3, 2] fail to account for these feedback mechanisms, focusing instead on static outputs that do not measure the adaptive performance of systems under changing conditions.

To this end, we make a big step forward to fill this gap by amalgamating the current offline cooperative perception toolkit OpenCOOD [1] into the OpenCDA simulator. Specifically, compiling OpenCOOD as an additional `MLManager` component, our enhanced **OpenCDA-∞** can now not only run a diverse array of cooperative detection models on the fly but also allow the outputs of these models to directly steer the planning and decision-making processes of its autonomous agents. This enriched feature makes it possible to investigate the influence of state-of-the-art cooperative perception models at a system level, which can more faithfully simulate real-world scenarios.

We would like to re-emphasize the importance of building a simulator that runs online detection models, as we have later surfaced in Sec. 5.2 that detection accuracy, although a relatively reliable metric, does not necessarily strictly reveal the overall rank in terms of planning performance. Instead, examining other metrics beyond detection, such as safety- or efficiency-level metrics, can usually offer more informative insights into the system-level evaluation in various aspects. On the one hand, this type of real-time testing capability allows us to directly test and refine models within an end-to-end simulated environment that accurately reflects the unpredictability of real-world driving, thus largely reducing the development time before onboard deployment. On the other hand, this approach serves as a testbench that supports the crucial phase of the sim-to-real generalization research.

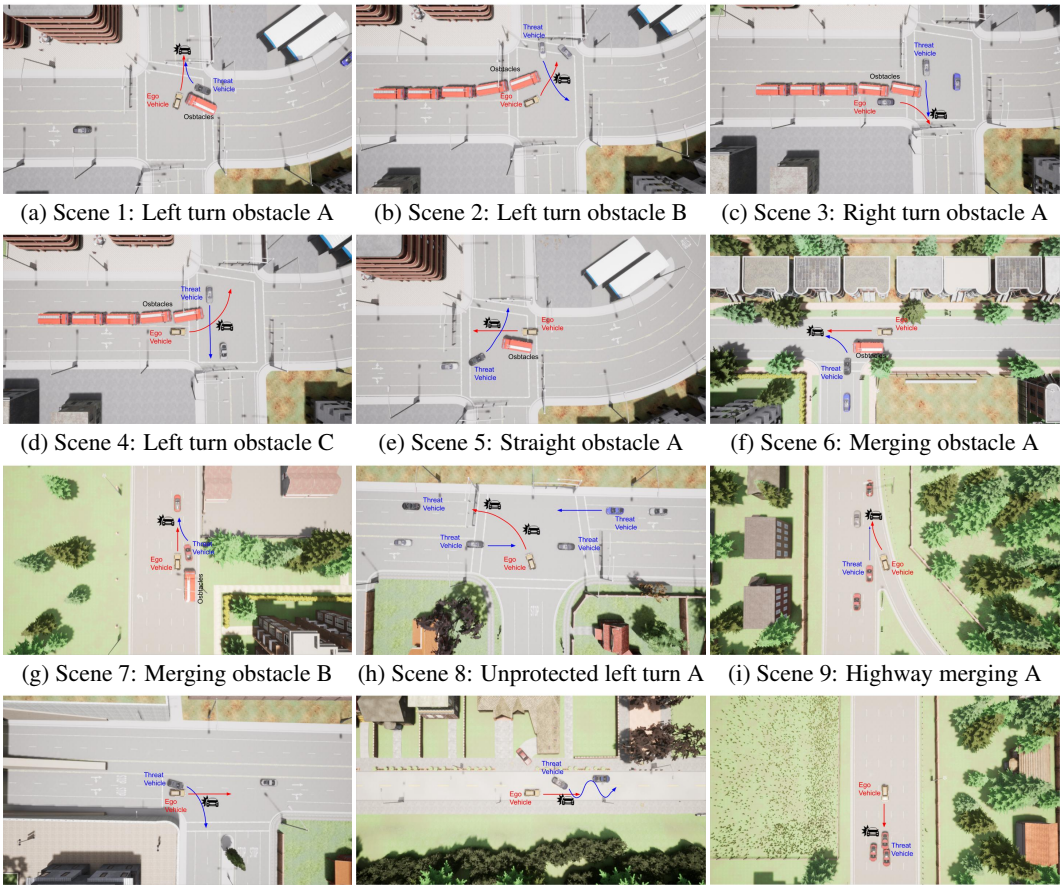
3.3 OpenSCENARIO Add-ons

OpenCDA, by default, utilizes the built-in CARLA traffic manager to simulate vehicle dynamics, automatically computing routes from initial spawn points to destinations. However, this approach provides limited control over the specific behaviors of individual actors, which can be restrictive when generating complex scenarios. OpenSCENARIO [7], a standardized XML-based language for driving scenarios, offers a structured method to create complex, reproducible, and configurable simulations that range from simple straight-road driving to intricate urban settings with multiple dynamic actors. This framework not only facilitates the scripting of detailed scenarios but also supports the encoding of high-level traffic rules and participant behaviors. We integrated this feature with CARLA through the ScenarioRunner extension, allowing us to construct highly challenging pre-crash scenarios through accurate agent behavior control. More details are in the Appendix.

3.4 Trajectory Prediction and Behavior Planning

OpenCDA originally did not support real-time trajectory prediction, limiting our ability to explore how predicted vehicle movements impact subsequent planning in automated driving systems. To address this, we have incorporated a trajectory prediction module capable of simulating realistic traffic scenarios and driver behaviors. We implemented various common trajectory prediction models in OpenCDA-∞: ❶ *Constant Velocity*: Suitable for steady traffic flow, predicting linear vehicle movements as $x = vt$. ❷ *Constant Acceleration*: Useful for scenarios of acceleration or deceleration, modeled by $v = u + at$. ❸ *Constant Speed and Yaw Rate*: Applies to vehicles moving at a constant speed but changing direction, described as $\theta = \omega t$. ❹ *Constant Acceleration and Yaw Rate*: Combines linear and angular dynamics for scenarios like exit ramps. ❺ *Physics Oracle Model*: A comprehensive model for predicting complex maneuvers in high-stakes environments.

OpenCDA utilizes a rule-based finite-state machine for planning, dynamically responding to specific traffic scenarios. This system transitions through several states, including route calculation, lane changing, overtaking, and adaptive speed regulation based on proximity to obstacles (see Appendix.A.1.1 for details). The initial planning algorithms implemented in OpenCDA constantly fail to complete the planned global route without incorporating prediction models, particularly in scenarios involving complex intersections, lane mergers, or vehicles emerging from blind spots. To



(a) Scene 1: Left turn obstacle A (b) Scene 2: Left turn obstacle B (c) Scene 3: Right turn obstacle A

(d) Scene 4: Left turn obstacle C (e) Scene 5: Straight obstacle A (f) Scene 6: Merging obstacle A

(g) Scene 7: Merging obstacle B (h) Scene 8: Unprotected left turn A (i) Scene 9: Highway merging A

(j) Scene 10: Lane-crossing turn A (k) Scene 11: Zigzagging A (l) Scene 12: Studden stopping A

Figure 3: The visualization of potential pre-crash moments in the OPV2V-Safety Benchmark.

enhance the planning capability, in OpenCDA-∞, we established a collision prediction mechanism that assesses potential future collisions within a specified lookahead time window, T seconds. Given the uncertainty of velocity and traffic conditions, we defined configurable parameters that delineate the range of possible future positions, as explained in Fig. 2, formalized as follows:

$$L = \min_{t \in T} \min_{r \in [t-\tau, t+\tau]} |\mathbf{x}_r - \mathbf{y}_r|_2, \tag{1}$$

where $T = t_1, t_2, ..., t_K$ represents a uniform time series, $\mathbf{x}_r$ and $\mathbf{y}_r$ are the respective positions of the ego and threat vehicles at time r, and τ accounts for prediction error. If L exceeds a predefined safe distance, the planning algorithm adjusts to mitigate collision risk.

4 The OPV2V-Safety Dataset and Benchmark

In this section, we will detail how we collect the **OPV2V-Safety** dataset and build the benchmark for end-to-end evaluation of cooperative perception models in our end-to-end **OpenCDA**-∞ simulator.

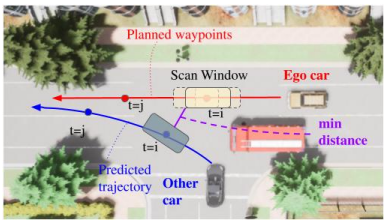


Figure 2: The diagram of our collision check model for robust planning with trajectory prediction.

4.1 Data Protocols

Scenario Setting. We generate the scenarios using the eight default towns, which are directly available in CARLA for easy reproducibility. In each scenario, we follow the safety-critical

pre-crash traffic from NHTSA to set up the ego vehicle's driving route and the threat vehicle prone to colliding. We further add another layer of complexity by positioning large trucks to obstruct the sensors (LiDAR and cameras) of the ego vehicle, simulating challenging V2V co-perception conditions. Typically, each scenario features two intelligent CAVs, including a collaborating CAV that aids the ego vehicle in detecting potential collision threats.

~~**Sensor Configuration.** Similar to previous works [1, 3], we configured each CAV to be equipped with four RGB cameras facing front, back, left, and right, respectively, so that they collectively cover the whole 360° panoramic view. Each camera has 110° field-of-view (FOV), capturing 800×600 RGB frames. Additionally, a 64-channel LiDAR mounted on the vehicle roof captures detailed point clouds with a range of 120 meters, recording data at 10 Hz along with essential metadata such as position and timestamps.~~

Scenario visualization. Following NHTSA guidelines, we carefully crafted twelve pre-crash scenarios representing diverse challenging driving conditions. These scenarios are designed to test the limits of vehicle visibility and showcase the efficacy of multi-agent cooperative perception in mitigating visibility constraints and extreme occlusion. Fig. 3 illustrates critical moments before potential crashes across all scenarios, depicting a variety of hazardous driving situations. These include complex interactions such as left and right turn obstacles, straight and merging challenges, unprotected turns, highway merges, and emergency stops. The Appendix provides a detailed specification of these scenarios, ranging from urban intersections to rural roads, each demanding proactive hazard avoidance and adherence to traffic norms by the ego vehicle.

4.2 Evaluation Metrics

We employ a multi-tiered evaluation framework in the OPV2V-C scenario benchmark to comprehensively assess cooperative perception models across several dimensions:

❶ **Model Level:** We utilize Average Precision (AP) at varying Intersection-over-Union (IoU) thresholds as standard metrics to assess 3D detection accuracy within a specified range around the ego vehicle, reporting AP@0.3, @0.5, and @0.7 for each scenario.

❷ **Safety Level:** Critical for any driving system, safety is evaluated through Collision Rate (CR), Time-to-Collision (TTC), and Off-Road (OR) incidents, which provide insights into the vehicle's ability to avoid collisions and maintain road discipline.

❸ **Efficiency Level:** Operational efficiency is measured by Time-to-Destination (TTD), Average Speed (AS), and Average Route Distance (ARD), quantifying the autonomous system's performance in achieving its objectives effectively.

❹ **Stability Level:** Stability metrics, including average acceleration (ACC) and average yaw rate (AYR), assess the smoothness and predictability of vehicle movements, enhancing passenger comfort and trust in the autonomous system.

❺ **System Level:** An aggregate score encapsulates overall performance, calculated as a weighted sum of normalized scores from all levels: $\text{OS} = \sum_{i=1}^{n} w_i \times M_i$, where each metric M_i is normalized based on its optimal value: $M_i = m_i / m_i^{max}$, if m_i is the higher the better; else, $= 1 - (m_i / m_i^{max})$, if m_i is the lower the better.

These metrics collectively provide a detailed and nuanced view of the autonomous system's capabilities, offering insights into its real-world applicability and effectiveness. Each metric has been chosen to reflect crucial aspects of autonomous operation, ensuring that our evaluations mirror the complexities and challenges of real driving scenarios.

5 Experiments

5.1 Experiment Settings

We conducted all the simulation experiments using our closed-loop simulation platform, The reference scenarios directly retrieve 3D bounding boxes from the server (i.e., 100% average precision), then run the entire simulation to get the reference metrics, such as the time-to-collision (TTC), average time spent to complete the route (TS), etc. We evaluated three types of cooperative detection methods: early fusion, intermediate fusion, and late fusion [1], all using the PointPillar [31] backbone for feature extraction. For intermediate fusion, we include two leading models, OPV2V [1] and V2X-ViT [2]. We run simulations for all the compared perception methods using the same configuration to deduce the impact of each detection module on the overall system performance.

Table 1: **Comprehensive diagnostic report of OpenCDA-∞ simulation performance on OPV2V-Safety benchmark**. We evaluated all the cooperative perception models on all the testing scenarios and reported the metrics. 1) Safety Level. CR: collision rate, TTC: average time-to-collision, SOR: stuck on road, OR: off-road. 2) Efficiency Level. TTD: time-to-destination, AS: average speed, ARD: average route distance. 3) Stability Level: ACC: average accelaration, AYR: average yaw rate. 4) System Level. OS: an overall score that summarizes all the metrics. $\uparrow$ / $\downarrow$: higher/lower the better.

		Safety Level			**Efficiency Level**			**Stability Level**		**Sys. Level**
Method	**V2V**?	CR$\downarrow$	TTC$\uparrow$	OR$\downarrow$	TTD$\downarrow$	AS$\uparrow$	ARD$\downarrow$	ACC$\downarrow$	AYR$\downarrow$	OS$\uparrow$
Early Fusion	No	0.500	N/A	0.000	15.50	23.91	83.71	0.295	0.132	0.516
	Yes	0.000	N/A	0.000	16.05	21.75	79.93	0.354	0.125	0.758
Late Fusion	No	0.417	6.31	0.000	15.80	22.53	79.34	0.298	0.132	0.535
	Yes	0.083	5.59	0.083	16.39	20.30	77.12	0.249	0.109	0.610
OPV2V [1]	No	0.417	6.45	0.000	16.55	20.86	81.85	0.278	0.120	0.532
	Yes	0.167	6.64	0.000	18.71	19.98	79.18	0.215	0.103	0.658
V2X-ViT [2]	No	0.583	5.28	0.000	16.42	21.73	81.02	0.293	0.136	0.440
	Yes	0.250	5.67	0.083	18.11	20.88	77.53	0.301	0.101	0.524

5.2 Quantitative Planning Results

Tab. 1 outlines the end-to-end simulation outcomes across different online cooperative perception methods as per evaluation metrics established in Sec. 4.2.

❶ **Safety Level**: we may observe that models without V2V communication consistently report high collision rates despite utilizing advanced planning and trajectory prediction (Sec. 3.4). The *Late Fusion* and *OPV2V* methods report slightly better but still inadequate CRs, over 40%, infeasible for practical deployment. In stark contrast, the *Early Fusion* method with V2V communication achieves a zero collision rate, significantly enhancing safety across all scenarios. The V2X-ViT model, however, shows a CR of 33.3%, indicating varying performance depending on the fusion method used. Evaluating against OR metrics, only *Late Fusion* and V2X-ViT with V2V communication record a metric score of 0.083; other methodologies report no such violations.

❷ **Efficiency Level**: Tab. 1 reveals a notable trend: the introduction of V2V communication typically results in a trade-off between safety and efficiency, often reducing Average Speed (AS) and increasing Time-to-Destination (TTD), due to early threat detection and preventive deceleration to avoid potential collision. However, Average Route Distance (ARD) tends to decrease, suggesting more efficient route planning. Moreover, it is worth mentioning that different models may outperform others across different evaluation metrics. As an illustration, while *Early Fusion* with V2V integration achieves an impressive zero CR score, emphasizing its safety, its efficiency level ARD performance doesn't quite match the performance of some other models.

❸ **Stability Level:** From our observations, models present distinct behaviors in this domain. Specifically, while both *Early Fusion* and *V2X-ViT* excel in performance in ACC, *Late Fusion* and *OPV2V* largely enjoy the benefits of cooperative perception. In terms of AYR, the integration of V2V communication facilitates a decline in scores across all fusion methods, hinting at improved stability concerning yaw rate modifications.

Table 2: **3D cooperative detection results on in-distribution OPV2V-test dataset**. We show Average Precision (AP) at IoU=0.5. The **boldfaced** and underlined entries indicate the best and second performers for enabling and disabling V2V communication, respectively.

Method	V2V?	Scenario index 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Avg.↑
Early Fusion	No	0.92	0.75	0.91	0.79	0.70	0.93	0.67	0.32	0.28	0.85	0.81	0.72	0.72	0.57	0.65	0.54	0.70
	Yes	0.88	0.91	0.95	0.89	0.71	0.95	0.67	0.43	0.58	0.93	0.80	0.73	0.75	0.64	0.83	0.79	0.78
Late Fusion	No	0.91	0.75	0.95	0.82	0.63	0.82	0.67	0.29	0.35	0.84	0.71	0.72	0.64	0.47	0.65	0.63	0.68
	Yes	0.89	0.87	0.94	0.88	0.69	0.81	0.82	0.38	0.54	0.91	0.65	0.75	0.71	0.52	0.77	0.68	0.74
OPV2V [1]	No	0.92	0.77	0.93	0.74	0.79	0.80	0.78	0.36	0.33	0.82	0.69	0.74	0.76	0.55	0.59	0.57	0.70
	Yes	0.91	0.92	0.94	0.92	0.80	0.90	0.48	0.47	0.48	0.93	0.75	0.84	0.76	0.69	0.81	0.73	0.77
V2X-ViT [2]	No	0.95	0.75	0.95	0.78	0.84	0.91	0.73	0.34	0.50	0.85	0.80	0.75	0.71	0.68	0.73	0.61	**0.74**
	Yes	0.93	0.92	0.95	0.89	0.86	0.90	0.73	0.46	0.64	0.91	0.86	0.80	0.75	0.67	0.87	0.61	**0.80**

❹ **System Level:** Our evaluation reveals a significant trade-off between safety and efficiency metrics in single-vehicle mode. Models like *Early Fusion* achieve high Average Speed (AS) scores, indicating efficient route completion but at the cost of higher collision rates (CR). Conversely, *OPV2V* showcases low CR scores but at the expense of the slowest Time-to-Destination (TTD), highlighting a fundamental conflict between safety and efficiency as also noted in prior studies [32, 33]. However, integrating V2V cooperative perception can mitigate these trade-offs. For example, *Early Fusion* with V2V not only maintains low CR but also improves TTD, showcasing the potential of V2V systems to break the limitations of single-vehicle perception. This analysis underscores the necessity for a unified system-level metric that comprehensively evaluates all performance dimensions. The composite Overall Score (OS) metric suggests that *Early Fusion* enhanced with V2V excels in overall system performance, while the popular *V2X-ViT* model scores lowest within our simulation framework despite claiming high detection capabilities in standard benchmarks.

5.3 Discussions on Detection Results

We then present comparative results using standard 3D object detection performance metrics on both the OPV2V-Test set and our newly introduced OPV2V-Safety set. It is worth noting that, in contrast to prior works like [1, 2] that adopted offline evaluation, we embed these models within our **OpenCDA**-∞ simulator for online AP assessment.

As indicated by Tab. 2, V2X-ViT consistently outperforms others, irrespective of V2V communication, aligning with their original study [2]. Interestingly, this result contrasts significantly with our planning-focused outcomes in Tab. 1 where V2X-ViT actually lags behind other approaches in the planning-oriented view. To further understand the root cause, we evaluated the AP scores on the OPV2V-Safety dataset, as in Tab. 3. It may be seen that all detection models, regardless of the fusion strategies, experience a marked drop in AP scores. Specifically, without V2V, OPV2V's AP@0.5 is only 0.24, significantly lower than scores reported in earlier works [1, 2]. Enabled V2V sees the straightforward *Early Fusion* leading with a 0.39 AP. Still, these results are rather unacceptably low compared to numbers on the in-domain test set (i.e., OPV2V-test), highlighting the challenging nature of our proposed OPV2V-Safety benchmark. Our findings call for a reevaluation of current V2V perception models and emphasize the necessity for advancements in technologies that ensure safety and reliability in cooperative autonomous driving. This rigorous analysis of detection capabilities within a realistic, dynamic environment reveals critical insights into the limitations and potential improvements for future autonomous vehicle technologies.

6 Concluding Remarks

In this paper, we introduce a comprehensive closed-loop, end-to-end simulation framework called **OpenCDA**-∞ to evaluate V2V cooperative perception systems with a focus on planning-oriented performances beyond detection accuracy. Our framework enriches OpenCDA with functionalities such as online cooperative detection, OpenSCENARIO customization, trajectory prediction, and advanced planning capabilities, enabling online evaluation of the detection model's impact on downstream planning performance. We also introduced the OPV2V-Safety benchmark, which

Table 3: **3D cooperative detection results on the proposed (out-of-distribution) OPV2V-Safety dataset**. We show Average Precision (AP) at IoU=0.5.

Method	V2V?	Scenario index 1	2	3	4	5	6	7	8	9	10	11	12	Avg.↑
Early Fusion	No	0.08	0.06	0.11	0.10	0.02	0.00	0.46	0.17	0.26	0.21	0.11	0.31	0.16
	Yes	0.35	0.56	0.16	0.37	0.29	0.29	0.42	0.38	0.53	0.54	0.33	0.43	**0.39**
Late Fusion	No	0.14	0.10	0.11	0.08	0.01	0.03	0.20	0.15	0.32	0.38	0.14	0.25	0.16
	Yes	0.37	0.43	0.16	0.42	0.33	0.44	0.30	0.41	0.46	0.42	0.23	0.38	0.36
OPV2V [1]	No	0.18	0.08	0.11	0.07	0.00	0.03	0.55	0.23	0.34	0.31	0.26	0.42	**0.22**
	Yes	0.34	0.52	0.12	0.29	0.31	0.28	0.39	0.36	0.51	0.45	0.42	0.46	0.37
V2X-ViT [2]	No	0.16	0.12	0.08	0.08	0.01	0.03	0.47	0.25	0.40	0.30	0.21	0.38	0.21
	Yes	0.33	0.33	0.18	0.18	0.27	0.16	0.38	0.32	0.54	0.61	0.29	0.54	0.34

includes twelve complex scenarios carefully designed to challenge current cooperative systems under severe occlusions and challenging conditions. We provide a suite of evaluation metrics to assess performance across model safety, efficiency, stability, and overall system-level score. Our experiments demonstrate the effectiveness of our simulation framework in providing detailed insights into the diagnosis report of V2V perception models, highlighting their effects on planning-centric metrics like safety and efficiency levels. We hope these contributions mark a significant step forward in advancing the safety and planning-oriented benchmarks and modeling for cooperative driving systems.

References

[1] 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. 2, 3, 4, 5, 7, 8, 9, 10

[2] 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 *ECCV*, pages 107–124. Springer, 2022. 2, 3, 4, 5, 8, 9, 10

[3] Yiming Li, Ziyan An, Zixun Wang, Yiqi Zhong, Siheng Chen, and Chen Feng. V2x-sim: A virtual collaborative perception dataset for autonomous driving. *arXiv preprint arXiv:2202.08449*, 2022. 2, 3, 5, 7

[4] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. Carla: An open urban driving simulator. In *Conference on robot learning*, pages 1–16. PMLR, 2017. 2, 4

[5] Cristina Olaverri-Monreal, Javier Errea-Moreno, Alberto Díaz-Álvarez, Carlos Biurrun-Quel, Luis Serrano-Arriezu, and Markus Kuba. Connection of the sumo microscopic traffic simulator and the unity 3d game engine to evaluate v2x communication-based systems. *Sensors*, 18(12):4399, 2018. 2

[6] Runsheng Xu, Yi Guo, Xu Han, Xin Xia, Hao Xiang, and Jiaqi Ma. Opencda: an open cooperative driving automation framework integrated with co-simulation. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pages 1155–1162. IEEE, 2021. 2, 4

[7] CARLA authors. Openscenario support. https://carla-scenariorunner.readthedocs.io/en/latest/openscenario_support/, 2023. 2, 5

[8] He Chen, Hongpinng Ren, Rui Li, Guang Yang, and Shanshan Ma. Generating autonomous driving test scenarios based on openscenario. In *2022 9th International Conference on Dependable Systems and Their Applications (DSA)*, pages 650–658. IEEE, 2022. 2

[9] Wassim G Najm, John D Smith, Mikio Yanagisawa, et al. Pre-crash scenario typology for crash avoidance research. Technical report, United States. National Highway Traffic Safety Administration, 2007. 2

[10] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *CVPR*, pages 3354–3361. IEEE, 2012. 3

[11] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *CVPR*, pages 11621–11631, 2020. 3

[12] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *CVPR*, pages 2446–2454, 2020. 3

[13] Yihan Hu, Jiazhi Yang, Li Chen, Keyu Li, Chonghao Sima, Xizhou Zhu, Siqi Chai, Senyao Du, Tianwei Lin, Wenhai Wang, et al. Planning-oriented autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17853–17862, 2023. 3

[14] Hao Shao, Letian Wang, Ruobing Chen, Steven L Waslander, Hongsheng Li, and Yu Liu. Reasonnet: End-to-end driving with temporal and global reasoning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13723–13733, 2023. 3

[15] Letian Wang, Jie Liu, Hao Shao, Wenshuo Wang, Ruobing Chen, Yu Liu, and Steven L Waslander. Efficient reinforcement learning for autonomous driving with parameterized skills and priors. *arXiv preprint arXiv:2305.04412*, 2023. 3

[16] Jiaxun Cui, Hang Qiu, Dian Chen, Peter Stone, and Yuke Zhu. Coopernaut: end-to-end driving with cooperative perception for networked vehicles. In *CVPR*, pages 17252–17262, 2022. 3, 4

[17] Hao Shao, Yuxuan Hu, Letian Wang, Guanglu Song, Steven L Waslander, Yu Liu, and Hongsheng Li. Lmdrive: Closed-loop end-to-end driving with large language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15120–15130, 2024. 3

[18] Marin Toromanoff, Emilie Wirbel, and Fabien Moutarde. End-to-end model-free reinforcement learning for urban driving using implicit affordances. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7153–7162, 2020. 3

[19] Zhejun Zhang, Alexander Liniger, Dengxin Dai, Fisher Yu, and Luc Van Gool. End-to-end urban driving by imitating a reinforcement learning coach. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 15222–15232, 2021. 3

[20] Quanyi Li, Zhenghao Mark Peng, Lan Feng, Zhizheng Liu, Chenda Duan, Wenjie Mo, and Bolei Zhou. Scenarionet: Open-source platform for large-scale traffic scenario simulation and modeling. *Advances in neural information processing systems*, 36, 2024. 3

[21] Lan Feng, Quanyi Li, Zhenghao Peng, Shuhan Tan, and Bolei Zhou. Trafficgen: Learning to generate diverse and realistic traffic scenarios. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3567–3575. IEEE, 2023. 3

[22] 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 *ECCV*, pages 605–621. Springer, 2020. 3

[23] 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 *CVPR*, pages 21361–21370, 2022. 3

[24] 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 *CVPR*, pages 13712–13722, 2023. 3

[25] Qi Chen, Sihai Tang, Qing Yang, and Song Fu. Cooper: Cooperative perception for connected autonomous vehicles based on 3d point clouds. In *2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*, pages 514–524. IEEE, 2019. 3

[26] Zaydoun Yahya Rawashdeh and Zheng Wang. Collaborative automated driving: A machine learning-based method to enhance the accuracy of shared information. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 3961–3966. IEEE, 2018. 3

[27] 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. 3

[28] 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. *arXiv preprint arXiv:2211.07214*, 2022. 3

[29] Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip HS Torr, and Vladlen Koltun. Point transformer. In *ICCV*, pages 16259–16268, 2021. 4

[30] 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. 4

[31] Alex H Lang, Sourabh Vora, Holger Caesar, Lubing Zhou, Jiong Yang, and Oscar Beijbom. Pointpillars: Fast encoders for object detection from point clouds. In *CVPR*, pages 12697–12705, 2019. 8

[32] Scott Fujimoto, Edoardo Conti, Mohammad Ghavamzadeh, and Joelle Pineau. Benchmarking batch deep reinforcement learning algorithms. *arXiv preprint arXiv:1910.01708*, 2019. 9

[33] Zuxin Liu, Hongyi Zhou, Baiming Chen, Sicheng Zhong, Martial Hebert, and Ding Zhao. Constrained model-based reinforcement learning with robust cross-entropy method. *arXiv preprint arXiv:2010.07968*, 2020. 9

Checklist

The checklist follows the references. Please read the checklist guidelines carefully for information on how to answer these questions. For each question, change the default **[TODO]** to [Yes] , [No] , or [N/A] . You are strongly encouraged to include a **justification to your answer**, either by referencing the appropriate section of your paper or providing a brief inline description. For example:

- Did you include the license to the code and datasets? [Yes] See Section A.3.1.1.
- Did you include the license to the code and datasets? [No] The code and the data are proprietary.
- Did you include the license to the code and datasets? [N/A]

Please do not modify the questions and only use the provided macros for your answers. Note that the Checklist section does not count towards the page limit. In your paper, please delete this instructions block and only keep the Checklist section heading above along with the questions/answers below.

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] See Sec. 1.
   (b) Did you describe the limitations of your work? [Yes] See Appendix A.4.
   (c) Did you discuss any potential negative societal impacts of your work? [Yes] See Appendix A.5.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes] We have read the ethics review guidelines and ensured that our paper conforms to them.
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments (e.g. for benchmarks)...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] We have included the code, data, and instructions in the supplemental material.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] We have specified all the details in Appendix A.1.2.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] See Appendix A.1.2.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [Yes] See Appendix.A.2.4 and the supplemental material.
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] See the supplemental material.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [Yes] See Appendix.A.2.2.
5. If you used crowdsourcing or conducted research with human subjects...

(a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]

(b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]