



# CooTest: An Automated Testing Approach for V2X Communication Systems

An Guo

guoan218@smail.nju.edu.cn  
State Key Laboratory for Novel  
Software Technology  
Nanjing University  
Nanjing 210023, China

Xinyu Gao

xinyugao@smail.nju.edu.cn  
State Key Laboratory for Novel  
Software Technology  
Nanjing University  
Nanjing 210023, China

Zhenyu Chen\*

zychen@nju.edu.cn  
State Key Laboratory for Novel  
Software Technology  
Nanjing University  
Nanjing 210023, China

Yuan Xiao

yuan.xiao@smail.nju.edu.cn  
State Key Laboratory for Novel  
Software Technology  
Nanjing University  
Nanjing 210023, China

Jiakai Liu

liujk@smail.nju.edu.cn  
State Key Laboratory for Novel  
Software Technology  
Nanjing University  
Nanjing 210023, China

Xiuting Ge

dg20320002@smail.nju.edu.cn  
State Key Laboratory for Novel  
Software Technology  
Nanjing University  
Nanjing 210023, China

Weisong Sun

weisongsun@smail.nju.edu.cn  
State Key Laboratory for Novel  
Software Technology  
Nanjing University  
Nanjing 210023, China

Chunrong Fang\*

fangchunrong@nju.edu.cn  
State Key Laboratory for Novel  
Software Technology  
Nanjing University  
Nanjing 210023, China

## Abstract

Perceiving the complex driving environment precisely is crucial to the safe operation of autonomous vehicles. With the tremendous advancement of deep learning and communication technology, Vehicle-to-Everything (V2X) collaboration has the potential to address limitations in sensing distant objects and occlusion for a single-agent perception system. However, despite spectacular progress, several communication challenges can undermine the effectiveness of multi-vehicle cooperative perception. The low interpretability of Deep Neural Networks (DNNs) and the high complexity of communication mechanisms make conventional testing techniques inapplicable for the cooperative perception of autonomous driving systems (ADS). Besides, the existing testing techniques, depending on manual data collection and labeling, become time-consuming and prohibitively expensive.

In this paper, we design and implement CooTest, the first automated testing tool of the V2X-oriented cooperative perception module. CooTest devises the V2X-specific metamorphic relation and equips communication and weather transformation operators that can reflect the impact of the various cooperative driving factors to

produce transformed scenes. Furthermore, we adopt a V2X-oriented guidance strategy for the transformed scene generation process and improve testing efficiency. We experiment CooTest with multiple cooperative perception models with different fusion schemes to evaluate its performance on different tasks. The experiment results show that CooTest can effectively detect erroneous behaviors under various V2X-oriented driving conditions. Also, the results confirm that CooTest can improve detection average precision and decrease misleading cooperation errors by retraining with the generated scenes.

## CCS Concepts

• **Software and its engineering** → **Software testing and debugging**.

## Keywords

Software testing, Autonomous driving system, Cooperative perception, Metamorphic testing

## ACM Reference Format:

An Guo, Xinyu Gao, Zhenyu Chen, Yuan Xiao, Jiakai Liu, Xiuting Ge, Weisong Sun, and Chunrong Fang. 2024. CooTest: An Automated Testing Approach for V2X Communication Systems. In *Proceedings of the 33rd ACM SIGSOFT International Symposium on Software Testing and Analysis (ISSTA '24)*, September 16–20, 2024, Vienna, Austria. ACM, New York, NY, USA, 13 pages. <https://doi.org/10.1145/3650212.3680373>

## 1 Introduction

Autonomous driving has attracted increasing attention due to its great potential to reduce the burden on drivers and improve traffic

\*Zhenyu Chen and Chunrong Fang are the corresponding authors.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

ISSTA '24, September 16–20, 2024, Vienna, Austria

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0612-7/24/09

<https://doi.org/10.1145/3650212.3680373>

safety [38, 59]. As an exemplar of safety-critical intelligent software, autonomous driving systems rely on perception components to understand information about the surrounding environment and communicate perception results to downstream decision modules to enable smooth autonomous operation. Recent advancements in deep learning and sensor technology have improved the performance of modern perception systems [36, 50]. Despite the remarkable progress, single-agent perception systems still yield inaccurate or incomplete perception results due to inherent single-view constraints (e.g., occlusion and distant sparse sensor observation), which can further lead to incorrect system behavior or even severe accidents [5, 8]. An illustration of this issue involves an Uber autonomous vehicle that collided with a pedestrian crossing the street. The vehicle’s perception system initially misidentified the pedestrian at a distance owing to single-view constraints. When the system eventually recognized the pedestrian, it was too late to avoid the collision [3].

Vehicle-to-Everything (V2X) communication has recently received significant attention as a solution to avoid the inherent limitations of single-agent sensing [16, 41]. As a key technology in V2X autonomous driving, multi-agent cooperative perception strategies enable multiple connected agents to share complementary perceptual information with each other, thus providing more accurate and comprehensive perception [49, 56]. Various prominent manufacturers and organizations, such as Ford and Baidu, are actively developing V2X platforms to enhance and ensure the performance of autonomous driving perception [4, 6]. However, when deployed in real-world environments, such communication systems across multiple intelligent agents could encounter intricate communication challenges, such as lossy communication [43] and communication latency [49]. In practice, corrupted or misaligned shared perception information from cooperative vehicles could hardly assist the ego agent in understanding the surroundings or even misleading it into making incorrect decisions. Based on the above reasons, it is crucial to design a test approach for V2X cooperative perception systems.

Software engineering researchers have proposed several testing techniques to assess the potential risks of perception systems in real-world scenarios [15, 21, 48]. However, existing testing techniques only focus on testing the single-agent perception system rather than considering specific communication issues during the cooperation process. Moreover, practitioners are required to manually collect data obtained from connected vehicles during real-world road testing while covering various scenarios against diverse adverse communication conditions demands substantial resource consumption [37]. Furthermore, manual labeling for collected data (e.g., images, point clouds, etc.) from various connected vehicles is a laborious and time-consuming task requiring a massive workforce to manually examine the visualization results [32].

To bridge this gap, we propose a systematic V2X-oriented automated testing approach and implement it into a tool, namely CooTest. The primary objective of CooTest is to automate the testing of cooperative perception systems in self-driving cars and subsequently facilitate their performance improvement through retraining. CooTest is built upon the theory of metamorphic testing to reduce manual data collection and labeling efforts. It implements

the metamorphic relation with two families of transformation operators that can reflect the characteristics of practical V2X cooperative perception scenes. CooTest leverages a V2X-oriented guided approach in the testing process to efficiently generate testing cases derived from seed testing data utilizing transformation operators. Furthermore, CooTest leverages transformation-specific metamorphic relations between the transformed data and seed data to detect the cooperative perception faults automatically.

We conduct experiments involving six state-of-the-art cooperative perception models with distinct information fusion schemes to assess the efficacy of CooTest. In the experiment, all transformation operators implemented by CooTest proved effective on models across various fusion schemes during the testing process. Subsequent experimental findings underscore that CooTest with V2X-oriented transformation guidance can generate more failed tests on cooperative perception models than random strategy guidance. Moreover, our experiments reveal that retraining the cooperative perception model with generated data can substantially decrease the misleading cooperation errors and enhance the performance by 6.7% on average, as measured by average precision (AP).

The main contributions of this paper are summarized as follows:

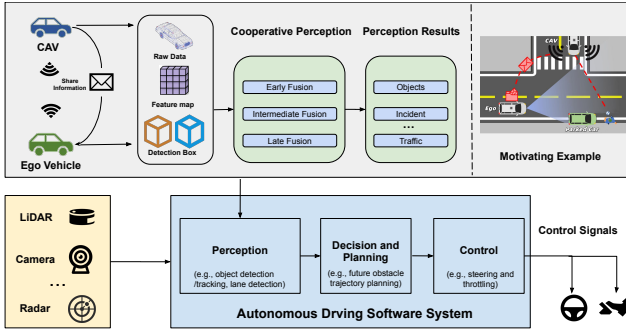
- We propose an automated testing approach for cooperative perception systems based on the metamorphic testing theory. We implement metamorphic relations with two families of transformation operators designed to simulate influential factors in application scenarios, thereby generating test scenes. Under the V2X-oriented transformation guidance, CooTest can detect cooperative-related faults for cooperative perception systems and enhance them efficiently.
- We implement the proposed approach into the automated testing tool, CooTest. To the best of our knowledge, CooTest is the first systematic and automated testing tool for V2X-oriented cooperative perception systems. To support the open science community, we have made the source code available<sup>1</sup> and released the generated transformed data.
- We utilize CooTest to assess six typical cooperative perception models and find various misleading cooperation errors, some of which may lead to potentially fatal collisions. The results show that the synthetic test cases can be used for retraining and making cooperative perception systems more robust to various corner cases.

## 2 Cooperative Driving Systems

### 2.1 V2X in Autonomous Driving

Autonomous driving systems is a complex system composed of various subsystems that operate sequentially. Figure 1 shows the architecture of an autonomous driving software system based on modern designs like Autoware [1] and Apollo [2]. The perception module is responsible for processing environmental information collected by various sensors to achieve functionalities such as object detection and tracking. Upon receiving the processed perception information, the decision and planning module estimates future trajectories for the detected obstacles across both temporal and

<sup>1</sup><https://github.com/meng2180/CooTest>



**Figure 1: The architecture of cooperative perception systems in autonomous driving.**

spatial dimensions. Finally, the control module calculates and executes the planned trajectory with lateral and longitudinal control. The perception system forms the fundamental basis of autonomous driving, which is essential for understanding the environment and ensuring the smooth operation of the entire system.

Considering autonomous driving systems are operated in safety-critical contexts and perception systems play a crucial role, most industrial manufacturers leverage the V2X technique to overcome inherent single-view perception limitations in single-agent sensing (e.g., occlusions and short-range perceiving capability) and improve the overall performance [16]. V2X denotes vehicles receiving shared information from other traffic elements via wireless communication technology [58]. The received information is rectified, matched, and integrated with ego vehicle sensing data into comprehensive and complementary perception information. This integrated data is then utilized by the perception system as input to make more accurate environmental predictions. Such a process involving multi-agent collaborative perception lays the foundation of the V2X autonomous driving workflow and significantly impacts the final ADS behaviors, as shown in Figure 1. V2X cooperative autonomous driving has demonstrated outstanding performances in intelligent traffic systems, encompassing tasks such as road intersection management [14] and the design of vehicle platooning [34].

## 2.2 V2X Cooperative Perception with On-road Agents

Owing to the inherent constraints associated with the camera and LiDAR devices, the issues of occlusions and long-distance perception pose considerable challenges for autonomous systems designed for single vehicles [57]. These challenges can give rise to severe consequences within intricate traffic scenarios. In contrast, cooperative perception systems can unlock the potential for multi-vehicle detection, effectively addressing the constraints associated with single-vehicle perception. By leveraging the V2X communication technology, different connected vehicles and infrastructures can share their sensing information and thus provide multiple viewpoints for the same obstacle to compensate each other [26, 30].

V2X cooperative perception methods can be divided into three categories based on the information fusion stage: *early fusion*, *late fusion*, and *intermediate fusion* [23]. Early fusion techniques directly

share raw data with connected agents located within the communication range. Then, the ego vehicle utilizes the aggregated data to predict detected objects [10]. In contrast, late fusion methods transmit detection outputs and integrate received proposals to formulate a coherent prediction [45]. To meet the bandwidth and detection accuracy requirements, researchers have delved into intermediate fusion [9, 49]. Intermediate fusion encodes the sparse raw data into dense vector representations for sharing.

V2X cooperative perception system holds the potential to enhance the safety and reliability of autonomous vehicles through the exchange of information with surrounding vehicles. However, it may face challenges from environmental factors like adverse weather (e.g., rain, snow, fog, etc.). Additionally, cooperative perception may introduce communication delays and unavoidable errors in relative pose among vehicles during data transmission. Therefore, it is critical to ensure the quality of the V2X communication systems before deploying them in a real-world scenario.

## 3 Approach

In this section, we present the design and implementation of CooTest, a tool devised for automatically testing the V2X-oriented cooperative perception module within ADS. In Figure 2, the developer initiates the process by selecting seeds from real-world data. Subsequently, CooTest adopts the metamorphic relation that we defined specifically for the cooperative perception systems. CooTest can generate new test scenes based on the seed by applying communication and weather transformation to simulate practical V2X application scenarios. Guided by V2X-oriented guidance, CooTest efficiently selects test data most likely to expose potential defects in the model. The transformed test scenes are then generated to assess the robustness of the cooperative perception model in autonomous driving. Finally, CooTest generates test reports containing AP and misleading cooperation errors based on models' predicted bounding boxes and gives feedback on model perception robustness. Moreover, CooTest can enhance the model's cooperative perception performance through retraining with generated transformed scenes.

### 3.1 Formalization and Definition

Cooperative errors in complex cooperative perception systems can arise from multiple components, including the ego vehicle, cooperative vehicles or infrastructures, and the information fusion process. The most basic reliability requirement of V2X systems is that the perception performance of the ego vehicle should not be impaired by the cooperative perception system, especially for close-range object detection [27]. However, due to communication challenges (Section 2.2), shared information from cooperative vehicles may provide limited assistance to the ego agent in understanding surroundings, and in certain situations, it might even share misleading information. In this case, even if the sensors equipped on the ego vehicle provide correct environment understanding, the ego vehicle's perception system will still prioritize the inaccurate shared information from collaborators over the correct one. In this paper, we refer to this error due to incorrect cooperative perception as *misleading cooperation error*. Before giving a formal definition of misleading cooperation error, we give a formal representation of V2X-oriented cooperation perception.

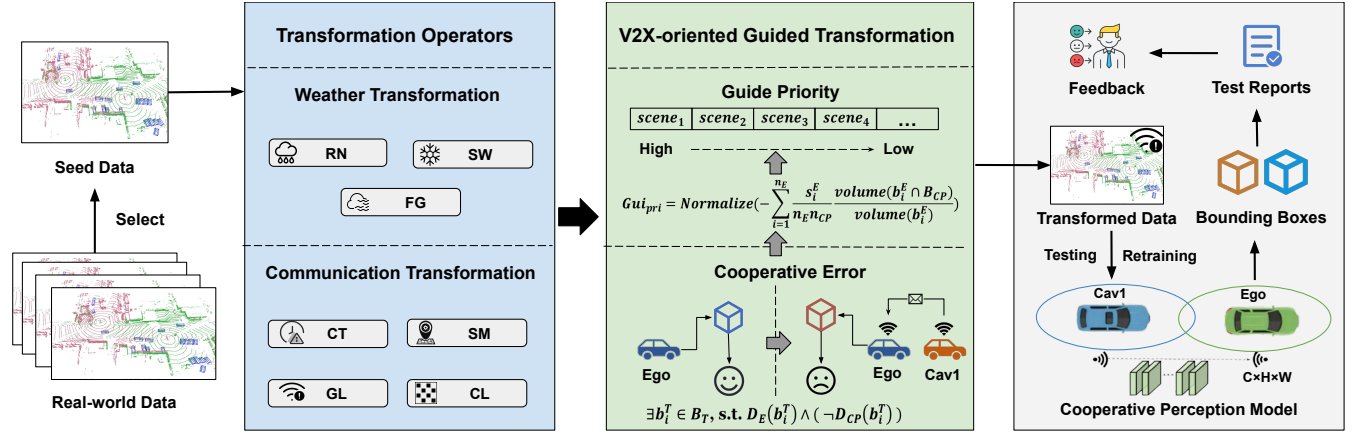


Figure 2: The Workflow of CooTest.

**3.1.1 Formal Representation.** We first formalize the V2X-oriented cooperative perception system and the ego vehicle perception system. Consider  $n$  cooperative vehicles and an ego vehicle in the scene. Let  $X_E$  and  $X_i$  represent the observation of the ego vehicle and the  $i$ -th connected vehicle, respectively. The cooperative perception systems  $\mathbb{CP}$  can be represented as:

$$\Phi_{CP}(X_E, \{X_j\}_{j=1}^n) = \{(s_i^{CP}, b_i^{CP}) | i \in [n_{CP}]\}$$

Similarly, the ego vehicle perception system can be represented as:

$$\Phi_E(X_E) = \{(s_i^E, b_i^E) | i \in [n_E]\}$$

where  $x \in \{E, CP\}$  and  $\Phi_x$  is the perception network of perception system.  $b_i^x, s_i^x$  represent the  $i$ -th 3D bounding box and its confidence of the perception results, respectively.  $[n_x] = 1, 2, \dots, n_x$  and  $n_x$  represent the number of 3D bounding boxes detected by the perception system. We leverage  $S_x = \cup_{i=1}^{n_x} s_i^x, x \in \{E, CP\}$  representing the set of all  $s_i^x$ .

**3.1.2 Misleading Cooperation Error (MCE).** Misleading cooperation error occurs (1) when an ego vehicle accurately perceives its surroundings during operation, and (2) the collaborative perception system is misled by the inaccurate shared information to make incorrect perception predictions. Here, we define the misleading cooperation error of a cooperative perception system  $\mathbb{CP}$ . Let  $b_i^T$  represent the  $i$ -th ground-truth bounding box and  $n_T$  represent the number of ground-truth 3D bounding boxes. We use  $B_x = \cup_{i=1}^{n_x} b_i^x, x \in \{T, E, CP\}$  representing the set of all  $b_i^x$ . A correct cooperative perception method should detect the bounding box, which is ground truth and successfully detected by the ego vehicle. Thus, a misleading cooperation error occurs if  $\mathbb{CP}$  cannot detect the ground-truth bounding box, which is successfully detected by the ego vehicle.

**Definition.** Consider a cooperative perception system  $\mathbb{CP}$ .  $\mathbb{CP}$  makes misleading cooperation error if

$$\exists b_i^T \in B_T, \text{ s.t. } D_E(b_i^T) \wedge (\neg D_{CP}(b_i^T))$$

$D_E$  and  $D_{CP}$  are the criteria to decide whether the ego vehicle system and the cooperative perception system  $\mathbb{CP}$  successfully detect the bounding box, respectively. Such misleading cooperation errors can diminish the precision of the cooperative perception

system, thereby influencing the safety and reliability of autonomous driving. To this end, this paper focuses on misleading cooperation error and their impact on the overall perception performance of cooperative systems.

## 3.2 Metamorphic Testing for Cooperative Perception System

A significant challenge in testing the intricate cooperative perception system lies in manually creating the system's specifications. Moreover, annotating LiDAR-captured point clouds requires a significant amount of time. Cooperative perception entails all connected autonomous vehicles participating in data collection, thus demanding a greater labeling effort compared to single-agent sensing. To overcome these issues, we leverage metamorphic relations (MRs) [11, 12] between the cooperative perception results across the original test data and its transformed test data to create the test oracle. The central idea is the ability to establish relationships among the model's detection outcomes under specific transformations. The violation of MRs often signals potential defects.

CooTest is designed to realize specific MRs designed for V2X communication systems. Specifically, given a seed test suite  $T_{orig}$  collected from connected vehicles, and a transformation operator  $\gamma \in \mathcal{O}$  which can generate various transformed data from each original data  $S \in T_{orig}$ , the MR to test cooperative perception system  $\mathbb{CP}$  with additional transformed point clouds adopted in this paper can be formalized as follows:

$$\forall S \in T_{orig} \wedge \forall \gamma \in \mathcal{O}, \zeta\{\mathbb{CP}[\gamma(S)], \mathbb{CP}[S]\} \quad (1)$$

where  $\zeta$  is a criterion asserting the equality of  $\mathbb{CP}$  results. The given MR is defined such that no matter how the transformed data is synthesized by applying specifically designed transformation operators denoted as  $\gamma$  on seed data  $S$ , the resulting object detection outcomes are anticipated to maintain consistency with those from the original data. Thus, if a  $\mathbb{CP}$  infers a set of 3D bounding boxes  $B_o$  for a seed data  $S$  and infers another set of boxes  $B_c$  for a corresponding transformed data  $S_c$  generated by applying transformation  $\gamma$  on  $S$ ,  $B_o$  and  $B_c$  should be identical; otherwise, erroneous predictions can be revealed by checking the failure of this MR.



However, asserting the strict equality of outputs from  $\mathbb{CP}$  is often too stringent because there is usually no single correct detection result for a given test case, which may result in many false positives. To introduce more permissiveness into the MRs, the equality criteria  $\zeta$  is derived from AP [17]. The AP score is computed by considering both precision and recall values, as explained in Section 4.3. By utilizing V2X-oriented MRs, CooTest can straightforwardly validate their satisfaction within the testing process of  $\mathbb{CP}$  by verifying against the test oracle. Based on this, CooTest further identifies misleading cooperative errors by automatically verifying the accuracy of the ego vehicle's perception results. In this context, the design and implementation of transformation operators become critical for optimizing the performance of CooTest.

### 3.3 Transformation Operators

V2X communication systems operating in the real world frequently face unexpected environmental conditions. These conditions could give rise to subtle variations in the signals during the communication process, causing the ego vehicle to receive low-quality shared information and potentially resulting in unexpected behaviors. Therefore, V2X communication systems are required to be robust against different corruption signals in various conditions.

To simulate practical application scenarios as much as possible, we investigate the common challenge for different V2X communication systems and design seven different realistic V2X-oriented transformation operators: communication latency (CT), spatial misalignment (SM), global feature lossy communication (GL), channel-specific lossy communication (CL), rain (RN), snow (SW), and fog (FG). These transformations can be classified into two categories: communication transformation and weather transformation. We implement these transformation operators and integrate them into CooTest. The detailed definition and design are as follows:

**3.3.1 Communication Transformation.** Due to the nature of the connectivity in the context of V2X, factors such as channel errors, network congestion, and delay violations can contribute to communication issues during data transmission within the wireless network [43]. These communication issues can be divided into three categories, including *communication latency*, *spatial misalignment*, and *lossy communication* [39]. Therefore, we design the corresponding communication transformation operators for each category and leverage them to simulate different real-world V2X-oriented communication issues. The specific operators are designed as follows: **Communication Latency.** V2X communication demands well-synchronized data transmission across connected vehicles to ensure the quality and timeliness of shared sensory information. In practical scenarios, limited communications bandwidth [49] or transmission failure [56] may cause a time delay in a participant, resulting in communication latency. We simulate the communication delay during the data transmission process, which can be expressed as:

$$\Delta M_{ego}(q) = M_{ego}^i - M_{cav_i} \quad (2)$$

where  $M_{ego}^i$  and  $M_{cav_i}$  respectively represent the timestamp when the ego vehicle receives the data packet  $q$  sent by the  $i$ -th connected vehicle and the corresponding data packet  $q$  collected by the  $i$ -th connected vehicle, respectively.  $\Delta M_{ego}$  can represent a time delay

in transmitting data from the  $i$ -th connected autonomous vehicle to the ego vehicle.

**Spatial Misalignment.** V2X communication system requires the determination of a positional transformation matrix between participants to ensure that local perception results from multiple participants can be matched efficiently. However, obstacles or GPS signal interference [40] in the real world will inevitably interfere with obtaining precise positioning information, leading to data misalignment during aggregation and performance degradation in cooperative perception. The transformation from the connected vehicle coordinates to the ego vehicle can be expressed as:

$$p_{cav}^{t, projected} = T_{cav \rightarrow ego} \cdot p_{cav}^t \quad (3)$$

where  $p_{cav}^t$  is the pose  $[x, y, z, 1]^T$  in  $i$ -th connected autonomous vehicle at the time  $t$ , and  $T_{cav \rightarrow ego} \in \mathbb{R}^{4 \times 4}$  is coordinate transformation matrix from connected vehicle to ego vehicle. In our experiments, we leverage rotation and translation to simulate spatial misalignment between the ego vehicle and connected vehicles.

**Lossy Communication.** Within urban traffic scenarios, various random factors, such as obstacle presence, rapid and fluctuating vehicle speeds, as well as varying distances between vehicles, may lead to lossy communication when transmitting a set of shared feature values [7, 43]. To simulate the realistic and complex lossy communication mechanisms in the real world, we designed two transformations based on the proposed communication models [33] to simulate different lossy types in real-world V2X communication. The specific details are as follows:

- **Global Feature Lossy Communication.** The shared feature after V2X metadata sharing is reshaped from a 3D tensor to a 2D matrix first. Subsequently, the reshaped feature undergoes random selection based on the global random probability  $p_g$  and is replaced by random noise within the range of the original shared feature values.
- **Channel-specific Lossy Communication.** Different from simulated lossy communication on the reshaped global feature, the channel-specific lossy type simulates lossy communication across distinct channels. Specifically, given a shared feature  $C \times H \times W$ ,  $\lfloor p_c * C \rfloor$  channels are randomly selected by the random probability  $p_c$ . Subsequently, these selected channels are replaced by random noise, maintaining values within the range of the original shared feature.

Eventually, the simulated lossy feature is transformed back to its initial shape of  $C \times H \times W$  before being received by the ego vehicle.

**3.3.2 Weather Transformation.** Adverse weather conditions can induce measurement distortions in the sensors of each connected vehicle, presenting a substantial challenge for cooperative perception systems reliant on redundant information [25]. For instance, droplets in the environment can decrease the detected object reflectivity and cause scattered points due to laser backscattering and absorption, leading to a reduction in the perceived quality of LiDAR point clouds. Therefore, we design the corresponding weather transformation operators to simulate different real-world weather.

In this study, we consider three representative adverse weather conditions: **Rain**, **Snow**, and **Fog**, as they are common in real-world driving scenarios. To achieve this, we utilize domain-specific physical models as outlined in [31] to simulate the characteristics of each

weather condition. The realism of these models has been validated in several benchmark studies within the community [19, 35]. During each transformation, we fix the environmental parameters and run the transformation for each connected vehicle to ensure that the connected vehicles are in an identical weather environment.

---

**Algorithm 1** V2X-oriented Guided Scene Transformation
 

---

**Input :** The tested model  $\mathbb{CP}$ , the transformation operator set  $\mathbb{A}$ , the set of seed scene  $\mathbb{T}_{orig}$ , the number of generated tests  $num_{gen}$

**Output :** Generated high-quality tests  $genTests$

```

1:  $genTests \leftarrow \emptyset$ 
2: for  $S$  in  $\mathbb{T}_{orig}$  do
3:   for  $a$  in  $\mathbb{A}$  do
4:      $S' = ApplyTransform(S, a)$ 
5:      $S_E, n_E, n_{CP}, B_E, B_{CP} = GetPred(\mathbb{CP}, S')$ 
6:      $Gui_{pri} = CalculateGuiPri(S_E, n_E, n_{CP}, B_E, B_{CP})$ 
7:      $Spc = CreateTransScene(S', Gui_{pri})$ 
8:     if  $len(genTests) < num_{gen}$  then
9:        $genTests.update(Spc)$ 
10:    else
11:      if  $Spc.Gui_{pri} > genTests[-1].Gui_{pri}$  then
12:         $genTests[-1] = Spc$ 
13:         $genTests.SortByGuiPri()$ 
14:      end if
15:    end if
16:  end for
17: end for
18: return  $genTests$ 
    
```

---

### 3.4 V2X-oriented Guided Transformation

In theory, the number of tests that each transformation operator can generate is endless. Therefore, we must impose limitations and guidance to ensure the quality of tests and improve the testing efficiency. To address this, inspired by the differential testing in traditional software testing, which leverages behavioral differences between multiple systems to detect errors [42], we propose a novel V2X-oriented guided transformation (VGT) strategy designed for cooperative perception systems to bootstrap the test generation process. The basic idea behind this guidance strategy is to maximize the difference in environmental understanding between the ego vehicle perception and cooperative perception, thus generating test inputs that can confuse cooperative perception models, which can be formalized as follows:

$$Gui_{pri} = Normalize(-\sum_{i=1}^{n_E} \frac{s_i^E}{n_{ENCP}} \frac{volume(b_i^E \cap B_{CP})}{volume(b_i^E)}) \quad (4)$$

$Gui_{pri}$  can be considered as a measure of the possibility of the cooperative perception system  $\mathbb{CP}$  making misleading cooperation errors. The larger  $Gui_{pri}$  means that the cooperative perception system  $\mathbb{CP}$  is less capable of detecting the correct bounding box that the ego vehicle can detect. Specifically, we compute the ratio of the intersection volume of the bounding box detected by both  $\mathbb{CP}$  and the ego vehicle to the volume of the bounding box  $b_i^E$  detected

by the ego vehicle. We weigh the ratio by the confidence  $s_i^E$  of the bounding box detected by the ego vehicle to emphasize the case that the ego vehicle detects the bounding box with high confidence while  $\mathbb{CP}$  does not detect this bounding box. The above value is summed and then averaged by the number  $n_E$  of the bounding boxes detected by the ego vehicle. We divide the value by the number  $n_{CP}$  of bounding boxes detected by  $\mathbb{CP}$  to reduce the effect that there exist two or multiple bounding boxes detected by  $\mathbb{CP}$  overlap with one bounding box at the same time. Finally, we normalize the negative value of the result to quantify the rate of predicted misleading cooperation errors. As such, we get the resulting value  $Gui_{pri}$ .

This guidance strategy enables CooTest to uncover more model defects with the same testing set size, thereby improving testing efficiency. It involves sequentially traversing the seed scenes and randomly selecting operators for transformation, ensuring diversity in the algorithm's output. Algorithm 1 presents the process of V2X scene transformation of CooTest in detail. The algorithm takes the tested model  $\mathbb{CP}$ , a set of communication transformation operators  $\mathbb{A}$  with the corresponding parameters, a seed set  $\mathbb{T}_{orig}$ , and the number of generated tests  $num_{gen}$  as input. Transformed tests can retain the same oracle information as the seed after a single transformation by configuring parameters for the transformation operators (Line 4). The core implementation process of this algorithm involves initially transforming the seed  $S$  and subsequently calculating the guidance metric value based on the predictions of bounding boxes and confidence through Equation 4 (Lines 5-7). Then, the generated test cases will be added to  $genTests$  (Lines 8-9). When the generated quantity surpasses  $num_{gen}$  (Lines 10-15), CooTest retains tests that are more likely to detect recognition errors.

## 4 Experimental Design

In this section, we introduce the experimental design, including the dataset and cooperative perception models under test, transformation settings, and evaluation metrics in the experiments. To conduct the experiments, we implemented the workflow of CooTest. All experiments are performed on Ubuntu 21.10 desktop with GeForce RTX 4070, one 16-core processor at 3.80GHz, and 32GB RAM.

CooTest is designed to systematically test the cooperation driving system, especially to verify its robustness for the V2X cooperative perception module. To this end, we empirically explore the following three research questions (RQ):

- RQ1: How effective are different transformation operators for detecting the erroneous behaviors of cooperative perception tasks using MRs?
- RQ2: How effectively can CooTest generate tests under the V2X-oriented guided transformation?
- RQ3: Can CooTest improve the cooperative perception model through retraining with transformed tests?

### 4.1 Dataset and Cooperative Perception Models

**V2V4Real** [55] is a large-scale real-world dataset for Vehicle-to-Vehicle cooperative perception. The data is collected by cooperative vehicles equipped with sensors, navigating through various scenarios. It comprises driving logs over three days, encompassing

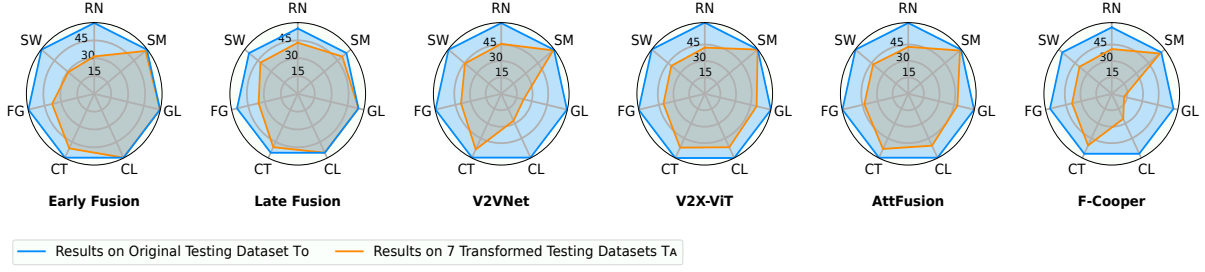


Figure 3: Testing results on different models with seven transformation operators.

a total of 347 km on highways and 63 km within city roads. Additionally, it consists of 20,000 LiDAR frames, 40,000 RGB frames, 240,000 annotated 3D bounding boxes covering five classes, and comprehensive HDMaps encompassing all the driving routes. In this paper, we apply the proposed technique to the original test split of the cooperative object detection dataset to detect errors related to the V2X cooperative perception module.

**Cooperative Perception Models.** To assess the performance of our techniques comprehensively, we employ six cooperative perception models with diverse fusion schemes in our experiment, including one early fusion model, one late fusion model, and four state-of-the-art intermediate fusion models. All of these collaborative perception models are LiDAR-based and take multi-view point clouds as input. A brief introduction is provided as follows.

**Early Fusion** model [55] directly transmits the raw point clouds to other collaborators, and the ego vehicle will aggregate all the data to its own coordinate frame. This process ensures the preservation of complete information.

**Late Fusion** model [55] detects objects using the sensor observations of connected vehicles and subsequently shares the detection results with others. The receiving ego vehicle then applies non-maximum suppression to generate the final outputs.

**V2VNet** model [49] proposes multi-round message passing based on graph neural networks to achieve better perception performance. It consists of three main stages: a convolutional network block to produce an intermediate representation, a cross-vehicle aggregation stage, and an output network to compute the final outputs.

**V2X-ViT** model [56] introduces a novel vision transformer designed specifically for V2X perception. It incorporates a customized heterogeneous multi-head self-attention module crafted for graph attribute-aware multi-agent 3D visual feature fusion. This module is adept at capturing the inherent heterogeneity within V2X systems.

**F-Cooper** model [9] employs the maxout fusion to aggregate shared intermediate features. Data inputs undergo processing by the voxel feature encoding layers independently to generate voxel features. Next, spatial features are obtained locally on individual vehicles and fused together to generate the ultimate feature maps.

**AttFusion** model [57] is designed to capture interactions among features of neighboring connected vehicles, enabling the network to focus on key observations. This pipeline is flexible and can be seamlessly integrated with current deep learning-based detectors.

Table 1: Transformation operators and parameters used by CooTest for generating new V2X-oriented scenes.

Types	Operators	Parameters	Parameter Ranges
Weather	RN	$r_n$	(0,10)
	SW	$s_w$	(0.1,2.4)
	FG	$f_g$	(200,1000)
Communication	CT	$c_t$	(0,300)
	CL	$p_c$	(0,1)
	GL	$p_g$	(0,1)
	SM	$s_m = (t_x, r_y, t_z, r_z)$	$t_x \in (-0.2, 0.2), t_y \in (-0.2, 0.2), t_z \in (-0.2, 0.2), r_z \in (-2, 2)$

## 4.2 Transformations Settings

We leverage CooTest to improve the robustness of the cooperative detection model in adverse V2X-oriented driving conditions by making realistic communication and weather transformations. As described in Section 3.3, we mainly apply seven different realistic data transformation operators: i.e., CT, SM, GL, CL, RN, SW, and FG. In order to make our transformations realistic, each transformation is designed to represent corresponding cooperative driving semantics in real-world scenarios. The RN, SW, and FG operators use rainfall intensity (mm/h), snowfall intensity (mm/h), and fog visibility (m) as weather transformation semantics, respectively. The CL operator employs milliseconds (ms) as the unit for the transmission delay in vehicle-to-vehicle communication. For the SM operator, CooTest translates a specific distance (m) along the axis and rotates a certain angle around the z-axis. For lossy communication transformations, we leverage the global random probability  $p_g$  and the probability of choosing a channel  $p_c$  to simulate various lossy communication situations when transmitting a set of shared feature values. In our experiment design, candidate parameters for each transformation are shown in Table 1. Whenever the transformation operator is applied to the original data, CooTest randomly selects the parameters for a specific transformation.

## 4.3 Evaluation Metric

The primary task of cooperative perception is to collaboratively detect objects across multi-view 3D point clouds collected by the ego vehicle and the connected vehicles. Identified objects are annotated with a 3D bounding box. The central concept for evaluating the

3D object detection performance is Intersection over Union (IoU), which compares the coordinates of ground truth and predicted 3D bounding boxes [44]. The calculation of IOU can be represented as:

$$IoU = \frac{\text{area}(B_p \cap B_{gt})}{\text{area}(B_p \cup B_{gt})} \quad (5)$$

Specifically, IoU measures the overlap (intersection) area between a ground-truth 3D bounding box in bird's-eye view  $B_{gt}$  and a predicted 3D bounding box  $B_p$  in bird's-eye view over their union. In our experiments evaluating cars, we adhere to the conventions established in other 3D object detection research [18], requiring a bounding box overlap of at least 50%.

Then we employ the AP, which is widely utilized in object detection research for autonomous driving [44], to measure the testing performance of CooTest. To achieve fair comparison results, we use the 11 recall positions proposed in the Pascal VOC benchmark [17], which is defined as follows:

$$AP|R = \frac{1}{|R|} \sum_{r \in R} \rho_{\text{interp}}(r) \quad (6)$$

We apply exactly 11 equally spaced recall levels, i.e.,  $R_{11} = \{0, 0.1, 0.2, \dots, 1\}$ . The interpolation function is defined as  $\rho_{\text{interp}}(r) = \max_{r' : r' \geq r} \rho(r')$ , where  $\rho(r)$  gives the precision at recall  $r$ . A higher AP means a better performance of cooperative perception systems.

**Table 2: The AP of different models on the transformed testing sets under different perceiving ranges.**

Model	Short		Middle		Long		Overall	
	Before	After	Before	After	Before	After	Before	After
Early Fusion	76.1	<b>66.1</b>	42.4	<b>29.6</b>	47.7	<b>36.7</b>	59.7	<b>49.6</b>
Late Fusion	73.5	<b>68.2</b>	43.7	<b>34.5</b>	36.3	<b>31.1</b>	55.1	<b>48.7</b>
V2VNet	80.6	<b>72.1</b>	52.6	<b>34.9</b>	42.6	<b>26.8</b>	64.6	<b>50.9</b>
V2X-ViT	82	<b>72.7</b>	51.7	<b>36</b>	43.2	<b>28.9</b>	65	<b>53.5</b>
AttFusion	79.8	<b>72.3</b>	53.1	<b>38.2</b>	43.6	<b>32</b>	64.7	<b>54.5</b>
F-Cooper	80.8	<b>60.4</b>	45.6	<b>23.3</b>	32.8	<b>14.9</b>	60.7	<b>45.4</b>

## 5 Result Analysis and Discussion

### 5.1 Answer to RQ1

To verify whether the transformed test dataset can effectively detect the erroneous behaviors, CooTest applies seven transformations introduced in Section 3.3 for each cooperative perception model on the original test set  $T_O$  to generate test data. To avoid the inconsistency between the size of the transformed test dataset  $T_A$  and the original test dataset  $T_O$ , CooTest generates the same size of the transformed dataset as  $T_O$  for each transformation operator. Then the transformed datasets  $T_A = \{T_{CT}, T_{SM}, T_{GL}, T_{CL}, T_{RN}, T_{SW}, T_{FG}\}$  are put into the pre-trained model  $M$  to obtain 3D bounding box prediction outputs. Then we calculate the AP of car detection and total misleading cooperation errors according to the prediction results. We further analyze the impact of these transformation operators on the performance of cooperative perception systems under different perceiving ranges. Specifically, we divided the detection target into three ranges (refer to [39, 55]) based on the distance relative to

the ego vehicle, including 0 ~30m (Short), 30 ~50m (Middle), and 50 ~100m (Long). Then we calculate the AP value on all cooperative models for each level respectively. The drop of AP on transformed test sets indicates that CooTest can effectively detect the potential defects of the model under test.

**Results.** Figure 3 shows the AP of all tested models on the seed sets  $T_O$  and transformed testing sets  $T_A$ . The AP of all tested models decreases significantly on most transformed testing sets. The AP decline percentages of the models after applying transformation operators are 20.6% (RN), 21.7% (SW), 22.3% (FG), 16% (CL), 20.3% (GL), 8.3% (CT), and 1.8% (SM), respectively. Therefore, both lossy communication and adverse weather can greatly impair the accuracy of collaborative perception models. Table 2 further elaborates the average precision under different perceiving ranges before and after applying operators. After applying transformation operators, the AP decreases for short, middle, and long-distance cooperative perception are 10.2%, 15.4%, and 12.6%, respectively. These results show that all the proposed transformation operators have a positive effect on the model testing process.

**Table 3: The AP and MCE of models on the transformed testing sets generated with different guidance approaches.**

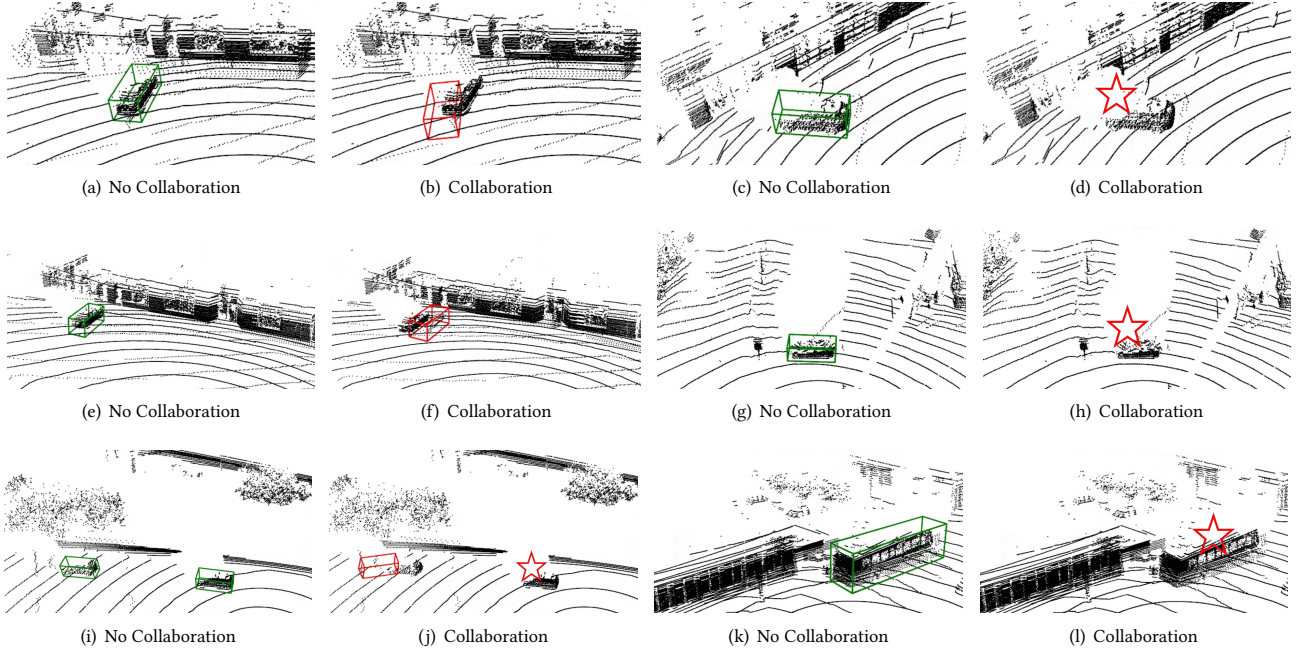
Model	VGT(10%)		Ran.(10%)		VGT(15%)		Ran.(15%)	
	AP	MCE	AP	MCE	AP	MCE	AP	MCE
Early Fusion	30.8	<b>1030</b>	51.1	723	38.7	<b>1670</b>	51.5	989
Late Fusion	43.6	<b>1202</b>	48.7	454	<b>45.0</b>	<b>1488</b>	48.4	652
V2VNet	15.9	<b>587</b>	53.1	443	25.4	<b>1100</b>	51.6	668
V2X-ViT	43.8	<b>756</b>	54.6	416	<b>46.3</b>	<b>1300</b>	54.5	626
AttFusion	43.3	<b>935</b>	54.9	412	<b>46.1</b>	<b>1416</b>	55.5	586
F-Cooper	39.5	<b>1844</b>	46.0	1210	<b>41.6</b>	<b>2433</b>	46.5	1841

**Discussion.** According to the results of the analysis, transformation operators can reduce the AP of all models within a certain range. Specifically, operators have a minor impact on the perception of close objects and a more significant impact on the perception of distant objects due to the sparse nature of the transformed point cloud. Moreover, the experimental results show that the same transformation operator has different effects on various cooperative perception models. Based on the observations in Figure 3, the late fusion strategy demonstrates greater resistance to interference from weather conditions and communication noise compared to other fusion schemes. We also find V2X-ViT and AttFusion demonstrate superior robustness to lossy communication operators (including CL and GL) compared to the other two intermediate fusion models. This resilience may be attributed to the presence of attention modules that enable them to handle shared lossy information effectively.

### 5.2 Answer to RQ2

To evaluate the effectiveness of the proposed V2X-oriented guided transformation (VGT) in detecting erroneous behaviors, we select the random strategy as the baseline. Specifically, we divide the original test set into two halves ( $T_{h1}$  and  $T_{h2}$ ), using one half ( $T_{h1}$ ) as the initial seeds to generate transformed test cases through both random testing and V2X-oriented guided testing. In the experiments, we apply the random method and the V2X-guided transformation





**Figure 4: Visualization samples showing misleading cooperation errors detected by CooTest using transformed scenes. The cars detected by the ego vehicle are marked in green, while the cars detected by cooperative perception systems are marked in red.**

method to select and retain the same batch of generated data. For each guided strategy, we retain 10% and 15% of the transformed data to evaluate the effectiveness of the guidance strategies. Then we calculate the AP in generated test cases of cooperative perception models under test for each configuration. Furthermore, we count the number of misleading cooperation errors (see Section 3.1) and visualize some typical instances of misleading cooperation errors detected by CooTest to provide a direct understanding.

**Results.** Table 3 presents the testing results of different models on the tests generated by two guidance approaches. Obviously, compared to the random guidance strategy, the proposed V2X-oriented guidance method can help to find more misleading cooperation errors in all configurations. Specifically, V2X-oriented guidance identifies at least 32.2% more misleading cooperation errors compared to random guidance for each configuration. Furthermore, we find that a large number of misleading cooperation errors could lead to a sharp decline in AP. On average, V2X-oriented guidance proves to be 29.2% and 20.9% more effective than random guidance when retaining 10% and 15% of transformed data, respectively. These results underscore that CooTest can achieve higher testing efficiency by utilizing the V2X-oriented guided transformation method.

To prove the effectiveness of CooTest in detecting misleading cooperation errors, we intercept some motivating examples found by CooTest, as shown in Figure 4. We provide a list of six real examples of misleading cooperation errors in the model under test, encompassing diverse and challenging communication scenes. As depicted in Figure 4, CooTest is capable of identifying misleading cooperation errors covering multiple object types (e.g., car, bus, etc.) in various transformation scenes. Specifically, surrounding objects accurately perceived by the ego vehicle are marked with green

cuboids. However, when involving connected vehicle collaboration, incorrect perception predictions are made, including inaccurate position estimation (marked with the red cuboid) and the inability to detect ground truth objects (marked with the red five-pointed star). One possible reason lies in that the ego vehicle’s perception system prioritizes inaccurate shared information from collaborators, even when the ego vehicle has correct environmental understanding.

**Discussion.** To investigate the effectiveness of misleading cooperation errors found by CooTest, we conduct an analysis of the number of misleading cooperation errors and AP values identified by the V2X-guided transformation method. The analysis results reveal that misleading cooperation errors can adversely affect the performance of the cooperative perception model. Additionally, our emphasis on misleading cooperation errors stems from the intuitive expectation that the cooperative perception system should aid the ego vehicle in detecting occlusion and distant objects even in various adverse communication scenarios. However, existing cooperative perception systems struggle to distinguish misleading information even if the ego vehicle accurately perceives its surroundings. For example, from Figure 4(g) and Figure 4(h), it is evident that the ego vehicle correctly detects objects in challenging snowy scenes. However, it fails to detect the ground truth object after collaborating with connected autonomous vehicles. We posit that mitigating such errors will lead to a further enhancement in the perception performance of the cooperative perception system.

### 5.3 Answer to RQ3

Here we investigate whether retraining the cooperative perception models with test scenes generated by CooTest helps enhance the model’s robustness. We apply the generated test suites in RQ2 for

**Table 4: The test result comparison after retraining the model with the transformed data generated by CooTest.**

Model	Short				Middle				Long				Overall			
	AP		MCE		AP		MCE		AP		MCE		AP		MCE	
	Before	After	Before	After	Before	After	Before	After	Before	After	Before	After	Before	After	Before	After
<b>Early Fusion</b>	65.9	<b>71.4</b>	3708	<b>1974</b>	30.0	<b>35.6</b>	2716	<b>1638</b>	36.5	<b>38.7</b>	613	<b>354</b>	49.5	<b>53.3</b>	7037	<b>3966</b>
<b>Late Fusion</b>	68.6	<b>73.5</b>	2326	<b>2001</b>	35.0	<b>47.0</b>	1718	<b>1078</b>	31.5	<b>47.4</b>	538	<b>317</b>	49.4	<b>60.6</b>	4582	<b>3396</b>
<b>V2VNet</b>	71.5	<b>79.0</b>	1810	<b>655</b>	34.4	<b>42.3</b>	2070	<b>913</b>	27.0	<b>31.4</b>	613	<b>373</b>	50.7	<b>57.0</b>	4493	<b>1941</b>
<b>V2X-ViT</b>	71.9	<b>79.8</b>	1434	<b>513</b>	35.9	<b>45.4</b>	1985	<b>719</b>	29.3	<b>33.3</b>	696	<b>366</b>	53.3	<b>59.2</b>	4115	<b>1598</b>
<b>AttFusion</b>	71.4	<b>78.8</b>	1874	<b>1131</b>	37.8	<b>45.0</b>	1648	<b>1111</b>	32.3	<b>35.6</b>	632	<b>458</b>	53.9	<b>60.2</b>	4154	<b>2700</b>
<b>F-Cooper</b>	59.9	<b>71.3</b>	7425	<b>1313</b>	23.7	<b>37.5</b>	3784	<b>1217</b>	15.4	<b>29.9</b>	1084	<b>424</b>	45.4	<b>52.0</b>	12293	<b>2954</b>

model training. To demonstrate the improvement of robustness clearly, we leverage the remaining data  $T_{h2}$  (introduced in RQ2) to construct the validation set. Specifically, we leverage each transform operator to generate a transformed data set with the same size as  $T_{h2}$ , and then combine all these transformed data to create a large-scale validation set. We keep the experimental configuration of retraining consistent with the initial training model process, and the retraining epoch is set as 5. To further ensure that the retrained models perform effectively in head scenes, we tested all retrained models on the original test scenes.

**Results.** Table 4 presents the AP and misleading cooperation error numbers under different perceiving ranges of all models after retraining with the transformed data generated by CooTest. As shown in Table 4, it is evident that regardless of the fusion scheme employed by the cooperative perception model, CooTest can consistently enhance the AP and mitigate misleading cooperation errors across varying perceiving ranges after retraining. Compared with the perception performance before retraining, the overall improvement range of AP is 7.7% ~22.7%, and the reduction percentage range of misleading cooperation error is 25.9% ~76%. Additionally, experiments show that among the six models, only the F-cooper model exhibits a slight decrease in performance in head scenes. In contrast, the performance of the other models remains almost the same or slightly improved. Overall, the average accuracy of all retrained models in the original test scenes increased by 3.1%.

**Discussion.** Our experiments show that using transformed data generated by CooTest for retraining can significantly improve the model robustness on the cooperative perception task. Meanwhile, we find it possible to reduce misleading cooperation errors by leveraging retraining techniques to increase the cooperative perception system's overall performance. However, we find that the retrained model still suffers from a variable number of misleading cooperation errors. One possible reason is that cooperative perception systems involve multiple agents collaborating in complex traffic scenarios. The complex model structure and specific communication mechanisms make it difficult to directly apply retraining techniques to completely repair all cooperative perception errors. Considering the significant impact of misleading cooperation errors on the system, more research on the continuous enhancement of the cooperative perception system, such as debugging and repairing, is needed.

## 5.4 Threats to Validity

**Data Selection.** Data selection stands as one of the fundamental threats to validity. The limitation of the experiment dataset may threaten the generality of the results. The quality of the transformed data is fundamentally contingent on the characteristics of the original dataset. Consequently, when it comes to input data featuring driving scenes not present in the training set, the retrained model may face challenges in accurate predictions. To alleviate this threat, CooTest adopts a diverse range of transformation operators and a widely used large-scale dataset for experiments.

**Data Simulation.** One threat to validity comes from the data simulation. On the one hand, various additional factors can potentially introduce communication issues, thereby influencing the behaviors of cooperative perception systems. On the other hand, the generated data may not be perfectly reproducible in the real world due to a multitude of unpredictable factors (e.g., the back-scattering coefficient of the target, the parameters of the LiDAR sensors, etc). However, our transformations are meticulously designed to be realistic, introduced in Section 4.2. Besides, we take into consideration unexpected environmental conditions and changes for V2X communication systems to the greatest extent possible.

**Parameter Settings.** An additional potential threat may stem from the variability in configurable transformation parameters. Some of our findings might not necessarily extend to different parameter sets during the evaluation phase. To mitigate this concern, we meticulously selected a wide array of parameters that closely replicate the actual cooperative driving environment within our experimental design. We delve into the candidate parameters for each transformation in Section 4.2.

**The Testing Functionality Selection.** CooTest conducts only offline testing [24, 47] and identifies cooperative issues in LiDAR-based cooperative object detection. While these issues are significant for perception module developers, this study does not assess their impact on subsequent tasks such as decision-making, planning, and control. This limitation underscores a crucial direction for future research.

## 6 Related Work

### 6.1 Testing and Verification of Autonomous Driving Perception Systems

As a foundational component in supporting ADS industrial applications, the ability of perception systems to understand the environment significantly influences the overall system quality [20, 36]. Recent research efforts have primarily concentrated on robustness testing of single-agent perception systems across various driving tasks [15, 19, 21, 48, 51]. Wang et al. [48] propose MetaOD to generate data by inserting object instances into the background image to test camera-based perception systems. Then they assert equality criteria between the original and synthesized images to reveal object detection model defects. Moreover, Xie et al. [51] effectively assess the robustness of object tracking systems in scenarios characterized by varying speeds of tracked objects and temporary camera failures. There are also a few works that focus on the testing of LiDAR-based perception systems. LiRTest [21] utilizes metamorphic relations to generate testing point clouds and identifies failures in the 3D object detection models. Christian et al. [15] propose an innovative approach for testing LiDAR-based perception systems by applying semantic mutations to labeled real-world data.

The key difference between the aforementioned related works and CooTest lies in their focus on single-vehicle perception quality assurance techniques, whereas our approach involves designing various transformation operators specifically for testing the V2X-oriented cooperative perception module. We are the first to establish the concept of faults attributed to misleading cooperation errors and formulate a V2X-oriented guided transformation strategy to improve the test efficiency of cooperative perception systems.

### 6.2 Metamorphic Testing

Metamorphic Testing (MT) [12, 46] is a software testing methodology that alleviates the test oracle problem. It discerns software bugs by detecting violations of domain-specific MRs defined across outputs from multiple executions of the program with different test inputs. In conventional software testing, researchers propose the utilization of MT to debugging [29], proving [13], fault localization [53], program slicing [54], program repair [28], etc.

With the swift advancement of deep learning, metamorphic testing has found application in testing both classification and regression tasks [22, 48, 52, 60]. Xie et al. [52] introduce an MT technique to test machine learning classification algorithms, demonstrating its effectiveness in identifying faults within a widely recognized open-source classification program. Besides, Wang et al. [48] employs MT as an adaptive and effective testing strategy by asserting the equivalence of object detection results between original and synthetic images to expose defects in object detection systems. Guo et al. [21] introduce LiDAR-based specific metamorphic relations and employ them to generate various corner test cases.

The metamorphic testing methods mentioned above cannot be directly applied in the V2X field due to the unique communication characteristics of cooperative perception systems. In this paper, CooTest develops a specific V2X-oriented metamorphic testing module for cooperative perception systems, where the communication and weather MRs are defined in Section 3.2. As a result, erroneous

cooperative perception results can be unveiled by checking the failure of these MRs.

## 7 Conclusion

In this paper, we propose and evaluate CooTest, the first systematic and automated module testing tool for V2X communication systems. In order to enrich the test samples and uncover corner cases for the models under test, CooTest has introduced seven innovative transformation operators. These operators are utilized to generate synthetic test data by applying transformations to seed data. We use V2X domain-specific MRs to find misleading cooperation errors in the cooperative perception models without detailed specifications.

CooTest has been evaluated using six state-of-the-art cooperative detection models employing various fusion methods. The experimental outcomes demonstrated that the transformed test data generated by CooTest efficiently detected erroneous behaviors and notably enhanced the robustness of V2X cooperative perception models through retraining. This substantiates CooTest's significant role in ensuring the quality and reliability of autonomous driving perception systems.

## 8 Data Availability

Our data are available on the website: <https://github.com/meng2180/CooTest>

## Acknowledgement

We would like to thank anonymous reviewers for their insightful comments. This work is supported partially by the National Natural Science Foundation of China (61932012, 62372228), and Science, Technology and Innovation Commission of Shenzhen Municipality (CJGJZD20200617103001003).

## References

- [1] 2015. Autoware. <https://github.com/Autoware-AI/autoware.ai>
- [2] 2017. Apollo. <https://github.com/ApolloAuto/apollo>
- [3] 2018. LIDAR Maker Velodyne Shifts Away Blame In Fatal Uber Self-Driving Crash. <https://jalopnik.com/lidar-maker-velodyne-blame-to-uber-in-fatal-self-driving-1824027977>.
- [4] 2019. FORD ACCELERATES CONNECTIVITY STRATEGY IN CHINA; TARGETS PRODUCTION OF FIRST C-V2X-EQUIPPED VEHICLE IN 2021. [https://media.ford.com/content/fordmedia/fap/cn/en/news/2019/03/26/Ford\\_Accelerates\\_Connectivity\\_Strategy\\_in\\_China\\_And\\_Targets\\_Production\\_of\\_First\\_C-V2X-Equipped\\_Vehicle\\_in\\_2021.html](https://media.ford.com/content/fordmedia/fap/cn/en/news/2019/03/26/Ford_Accelerates_Connectivity_Strategy_in_China_And_Targets_Production_of_First_C-V2X-Equipped_Vehicle_in_2021.html).
- [5] 2019. Tesla's Autopilot was engaged when Model 3 crashed into truck, report states. <https://www.theverge.com/2019/5/16/18627766/tesla-autopilot-fatal-crash-delray-florida-ntsb-model-3>
- [6] 2023. ASTRI and Baidu Apollo Collaborate to Promote the Implementation of C-V2X Technology in Hong Kong. <https://www.astri.org/news-detail/astri-and-baidu-apollo-collaborate-to-promote-the-implementation-of-c-v2x-technology-in-hong-kong/>.
- [7] Evgeny Belyaev, Alexey V. Vinel, Adam Surak, Moncef Gabbouj, Magnus Jonsson, and Karen O. Egiazarian. 2015. Robust Vehicle-to-Infrastructure Video Transmission for Road Surveillance Applications. *IEEE Trans. Veh. Technol.* 64, 7 (2015), 2991–3003. <https://doi.org/10.1109/TVT.2014.2354376>
- [8] Long Chen, Yuchen Li, Chao Huang, Bai Li, Yang Xing, Daxin Tian, Li Li, Zhongxu Hu, Xiaoxiang Na, Zixuan Li, et al. 2022. Milestones in autonomous driving and intelligent vehicles: Survey of surveys. *IEEE Transactions on Intelligent Vehicles* 8, 2 (2022), 1046–1056. <https://doi.org/10.1109/TIV.2022.3223131>
- [9] Qi Chen, Xu Ma, Sihai Tang, Jingda Guo, Qing Yang, and Song Fu. 2019. F-cooper: feature based cooperative perception for autonomous vehicle edge computing system using 3D point clouds. In *Proceedings of the 4th Symposium on Edge Computing*. ACM, Arlington, Virginia, USA, 88–100. <https://doi.org/10.1145/3318216.3363300>



- [10] Qi Chen, Sihai Tang, Qing Yang, and Song Fu. 2019. Cooper: Cooperative Perception for Connected Autonomous Vehicles Based on 3D Point Clouds. In *Proceedings of the 39th International Conference on Distributed Computing Systems*. IEEE, Dallas, TX, USA, 514–524. <https://doi.org/10.1109/ICDCS.2019.00058>
- [11] Tsong Yueh Chen, S. C. Cheung, and Siu-Ming Yiu. 2020. Metamorphic Testing: A New Approach for Generating Next Test Cases. *CoRR* abs/2002.12543, 1 (2020), 1–11. <https://arxiv.org/abs/2002.12543>
- [12] Tsong Yueh Chen, Fei-Ching Kuo, Huai Liu, Pak-Lok Poon, Dave Towey, TH Tse, and Zhi Quan Zhou. 2018. Metamorphic testing: A review of challenges and opportunities. *ACM Computing Surveys (CSUR)* 51, 1 (2018), 1–27. <https://doi.org/10.1145/3143561>
- [13] Tsong Yueh Chen, T. H. Tse, and Zhiquan Zhou. 2002. Semi-proving: an integrated method based on global symbolic evaluation and metamorphic testing. In *Proceedings of the International Symposium on Software Testing and Analysis*. ACM, Roma, Italy, 191–195. <https://doi.org/10.1145/566172.566202>
- [14] Andy HF Chow, ZC Su, EM Liang, and RX Zhong. 2021. Adaptive signal control for bus service reliability with connected vehicle technology via reinforcement learning. *Transportation Research Part C: Emerging Technologies* 129 (2021), 103264. <https://doi.org/10.1016/j.trc.2021.103264>
- [15] Garrett Christian, Trey Woodlief, and Sebastian Elbaum. 2023. Generating Realistic and Diverse Tests for LiDAR-Based Perception Systems. In *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*. IEEE, 2604–2616. <https://doi.org/10.1109/ICSE48619.2023.00217>
- [16] Guangzhen Cui, Weili Zhang, Yanqiu Xiao, Lei Yao, and Zhanpeng Fang. 2022. Cooperative perception technology of autonomous driving in the internet of vehicles environment: A review. *Sensors* 22, 15 (2022), 5535. <https://doi.org/10.3390/S22155535>
- [17] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. 2010. The pascal visual object classes (voc) challenge. *International journal of computer vision* 88 (2010), 303–338. <https://doi.org/10.1007/S11263-009-0275-4>
- [18] Duarte Fernandes, António Silva, Rafael Névoa, Cláudia Simões, Dibet Gonzalez, Miguel Guevara, Paulo Novais, João Monteiro, and Pedro Melo-Pinto. 2021. Point-cloud based 3D object detection and classification methods for self-driving applications: A survey and taxonomy. *Information Fusion* 68 (2021), 161–191. <https://doi.org/10.1016/j.inffus.2020.11.002>
- [19] Xinyu Gao, Zhijie Wang, Yang Feng, Lei Ma, Zhenyu Chen, and Baowen Xu. 2023. Benchmarking Robustness of AI-enabled Multi-sensor Fusion Systems: Challenges and Opportunities. In *Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. 871–882. <https://doi.org/10.1145/3611643.3616278>
- [20] Dominique Gruyer, Valentin Magnier, Karima Hamdi, Laurence Claussmann, Olivier Orfila, and Andry Rakotonirainy. 2017. Perception, information processing and modeling: Critical stages for autonomous driving applications. *Annual Reviews in Control* 44 (2017), 323–341. <https://doi.org/10.1016/j.arcontrol.2017.09.012>
- [21] An Guo, Yang Feng, and Zhenyu Chen. 2022. LiRTTest: augmenting LiDAR point clouds for automated testing of autonomous driving systems. In *ISSTA '22: 31st ACM SIGSOFT International Symposium on Software Testing and Analysis, Virtual Event, South Korea, July 18 - 22, 2022*, Sukyoung Ryu and Yannis Smaragdakis (Eds.). ACM, 480–492. <https://doi.org/10.1145/3533767.3534397>
- [22] An Guo, Yang Feng, Yizhen Cheng, and Zhenyu Chen. 2024. Semantic-guided fuzzing for virtual testing of autonomous driving systems. *J. Syst. Softw.* 212 (2024), 112017. <https://doi.org/10.1016/j.jss.2024.112017>
- [23] Yushan Han, Hui Zhang, Huifang Li, Yi Jin, Congyan Lang, and Yidong Li. 2023. Collaborative perception in autonomous driving: Methods, datasets, and challenges. *IEEE Intelligent Transportation Systems Magazine* (2023). <https://doi.org/10.1109/ITS.2023.3298534>
- [24] Fitash Ul Haq, Donghwan Shin, Shiva Nejati, and Lionel Claude Briand. 2021. Can Offline Testing of Deep Neural Networks Replace Their Online Testing? *Empir. Softw. Eng.* 26, 5 (2021), 90. <https://doi.org/10.1007/S10664-021-09982-4>
- [25] Robin Heinzler, Philipp Schindler, Jürgen Seekircher, Werner Ritter, and Wilhelm Stork. 2019. Weather influence and classification with automotive lidar sensors. In *2019 IEEE intelligent vehicles symposium (IV)*. IEEE, 1527–1534. <https://doi.org/10.1109/IVS.2019.8814205>
- [26] Laurens Hober, Andreas Festag, Ignacio Llatser, Luciano Altomare, Filippo Vintainer, and Andras Kovacs. 2015. Enhancements of V2X communication in support of cooperative autonomous driving. *IEEE communications magazine* 53, 12 (2015), 64–70. <https://doi.org/10.1109/MCOM.2015.7355568>
- [27] Senkang Hu, Zhengru Fang, Yiqin Deng, Xianhao Chen, and Yuguang Fang. 2024. Collaborative Perception for Connected and Autonomous Driving: Challenges, Possible Solutions and Opportunities. *CoRR* abs/2401.01544 (2024). <https://doi.org/10.48550/ARXIV.2401.01544>
- [28] Mingyue Jiang, Tsong Yueh Chen, Fei-Ching Kuo, Dave Towey, and Zuohua Ding. 2017. A metamorphic testing approach for supporting program repair without the need for a test oracle. *J. Syst. Softw.* 126 (2017), 127–140. <https://doi.org/10.1016/j.jss.2016.04.002>
- [29] Hao Jin, Yanyan Jiang, Na Liu, Chang Xu, Xiaoxing Ma, and Jian Lu. 2015. Concolic Metamorphic Debugging. In *39th IEEE Annual Computer Software and Applications Conference, COMPSAC 2015, Taichung, Taiwan, July 1-5, 2015. Volume 2*, Sheikh Iqbal Ahamed, Carl K. Chang, William C. Chu, Ivica Crnkovic, Pao-Ann Hsiung, Gang Huang, and Jingwei Yang (Eds.). IEEE Computer Society, 232–241. <https://doi.org/10.1109/COMPSAC.2015.79>
- [30] Chanyoung Jung, Daegyu Lee, Seungwook Lee, and David Hyunchul Shim. 2020. V2X-communication-aided autonomous driving: System design and experimental validation. *Sensors* 20, 10 (2020), 2903. <https://doi.org/10.3390/S20102903>
- [31] Velat Kilic, Deepti Hegde, Vishwanath Sindagi, A. Brinton Cooper, Mark A. Foster, and Vishal M. Patel. 2021. Lidar Light Scattering Augmentation (LISA): Physics-based Simulation of Adverse Weather Conditions for 3D Object Detection. *CoRR* abs/2107.07004, 1 (2021), 1–12. <https://arxiv.org/abs/2107.07004>
- [32] Jinhan Kim, Jeongil Ju, Robert Feldt, and Shin Yoo. 2020. Reducing dnn labelling cost using surprise adequacy: An industrial case study for autonomous driving. In *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. 1466–1476. <https://doi.org/10.1145/3368089.3417065>
- [33] Jinlong Li, Runsheng Xu, Xinyu Liu, Jin Ma, Zicheng Chi, Jiaqi Ma, and Hongkai Yu. 2023. Learning for vehicle-to-vehicle cooperative perception under lossy communication. *IEEE Transactions on Intelligent Vehicles* (2023). <https://doi.org/10.1109/TIV.2023.3260040>
- [34] Ming Li, Xinkai Wu, Xiaozheng He, Guizhen Yu, and Yunpeng Wang. 2018. An eco-driving system for electric vehicles with signal control under V2X environment. *Transportation Research Part C: Emerging Technologies* 93 (2018), 335–350. <https://doi.org/10.1016/j.trc.2018.06.002>
- [35] Shuangzhi Li, Zhijie Wang, Felix Juefei-Xu, Qing Guo, Xingyu Li, and Lei Ma. 2022. Common Corruption Robustness of Point Cloud Detectors: Benchmark and Enhancement. *CoRR* abs/2210.05896 (2022). <https://doi.org/10.48550/ARXIV.2210.05896>
- [36] You Li and Javier Ibanez-Guzman. 2020. Lidar for autonomous driving: The principles, challenges, and trends for automotive lidar and perception systems. *IEEE Signal Processing Magazine* 37, 4 (2020), 50–61. <https://doi.org/10.1109/MSP.2020.2973615>
- [37] Luyang Liu, Hongyu Li, Jian Liu, Cagdas Karatas, Yan Wang, Marco Gruteser, Yingying Chen, and Richard P Martin. 2017. Bigroad: Scaling road data acquisition for dependable self-driving. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*. 371–384. <https://doi.org/10.1145/3081333.3081344>
- [38] Liangkai Liu, Sidi Lu, Ren Zhong, Baofu Wu, Yongtao Yao, Qingyang Zhang, and Weisong Shi. 2020. Computing systems for autonomous driving: State of the art and challenges. *IEEE Internet of Things Journal* 8, 8 (2020), 6469–6486. <https://doi.org/10.1109/JIOT.2020.3043716>
- [39] Si Liu, Chen Gao, Yuan Chen, Xingyu Peng, Xianghao Kong, Kun Wang, Runsheng Xu, Wentao Jiang, Hao Xiang, Jiaqi Ma, et al. 2023. Towards Vehicle-to-everything Autonomous Driving: A Survey on Collaborative Perception. *arXiv preprint arXiv:2308.16714* (2023). <https://doi.org/10.48550/ARXIV.2308.16714>
- [40] Wei Liu, Xin Xia, Lu Xiong, Yishi Lu, Letian Gao, and Zhuoping Yu. 2021. Automated vehicle sideslip angle estimation considering signal measurement characteristic. *IEEE Sensors Journal* 21, 19 (2021), 21675–21687. <https://doi.org/10.1109/JSEN.2021.3059050>
- [41] Yunsheng Ma, Juanwu Lu, Can Cui, Sicheng Zhao, Xu Cao, Wenqian Ye, and Ziran Wang. 2024. MACP: Efficient Model Adaptation for Cooperative Perception. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 3373–3382. <https://doi.org/10.1109/WACV57701.2024.00334>
- [42] William M McKeeman. 1998. Differential testing for software. *Digital Technical Journal* 10, 1 (1998), 100–107. <https://www.hpl.hp.com/hpjournal/dtj/vol10num1/vol10num1art9.pdf>
- [43] Moustafa M Nasralla, Chaminda TER Hewage, and Maria G Martini. 2014. Subjective and objective evaluation and packet loss modeling for 3D video transmission over LTE networks. In *2014 international conference on telecommunications and multimedia (temu)*. IEEE, 254–259. <https://doi.org/10.1109/TEMU.2014.6917770>
- [44] Rafael Padilla, Sergio L Netto, and Eduardo AB Da Silva. 2020. A survey on performance metrics for object-detection algorithms. In *2020 international conference on systems, signals and image processing (IWSSIP)*. IEEE, 237–242. <https://doi.org/10.1109/IWSSIP48289.2020.9145130>
- [45] Zaydoun Yahya Rawashdeh and Zheng Wang. 2018. 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)*. IEEE, 3961–3966. <https://doi.org/10.1109/ITSC.2018.8569832>
- [46] Sergio Segura, Gordon Fraser, Ana B Sanchez, and Antonio Ruiz-Cortés. 2016. A survey on metamorphic testing. *IEEE Transactions on software engineering* 42, 9 (2016), 805–824. <https://doi.org/10.1109/TSE.2016.2532875>
- [47] Andrea Stocco, Brian Pulfer, and Paolo Tonella. 2023. Model vs system level testing of autonomous driving systems: a replication and extension study. *Empir. Softw. Eng.* 28, 3 (2023), 73. <https://doi.org/10.1007/S10664-023-10306-X>
- [48] Shuai Wang and Zhenhong Su. 2020. Metamorphic Object Insertion for Testing Object Detection Systems. In *35th IEEE/ACM International Conference on Automated Software Engineering, ASE 2020, Melbourne, Australia, September 21-25,*



2020. IEEE, 1053–1065. <https://doi.org/10.1145/3324884.3416584>
- [49] Tsun-Hsuan Wang, Sivabalan Manivasagam, Ming Liang, Bin Yang, Wenyuan Zeng, and Raquel Urtasun. 2020. V2vnet: Vehicle-to-vehicle communication for joint perception and prediction. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II* 16. Springer, 605–621. [https://doi.org/10.1007/978-3-030-58536-5\\_36](https://doi.org/10.1007/978-3-030-58536-5_36)
- [50] Li-Hua Wen and Kang-Hyun Jo. 2022. Deep learning-based perception systems for autonomous driving: A comprehensive survey. *Neurocomputing* 489 (2022), 255–270. <https://doi.org/10.1016/j.NEUCOM.2021.08.155>
- [51] Xiaoyuan Xie, Ying Duan, Songqiang Chen, and Jifeng Xuan. 2022. Towards the Robustness of Multiple Object Tracking Systems. In *2022 IEEE 33rd International Symposium on Software Reliability Engineering (ISSRE)*. IEEE, 402–413. <https://doi.org/10.1109/ISSRE55969.2022.00046>
- [52] Xiaoyuan Xie, Joshua Wing Kei Ho, Christian Murphy, Gail E. Kaiser, Baowen Xu, and Tsong Yueh Chen. 2011. Testing and validating machine learning classifiers by metamorphic testing. *J. Syst. Softw.* 84, 4 (2011), 544–558. <https://doi.org/10.1016/j.jss.2010.11.920>
- [53] Xiaoyuan Xie, W. Eric Wong, Tsong Yueh Chen, and Baowen Xu. 2011. Spectrum-Based Fault Localization: Testing Oracles are No Longer Mandatory. In *Proceedings of the 11th International Conference on Quality Software, QSIQ 2011, Madrid, Spain, July 13–14, 2011*, Manuel Núñez, Robert M. Hierons, and Mercedes G. Merayo (Eds.). IEEE Computer Society, 1–10. <https://doi.org/10.1109/QSIQ.2011.20>
- [54] Xiaoyuan Xie, W. Eric Wong, Tsong Yueh Chen, and Baowen Xu. 2013. Metamorphic slice: An application in spectrum-based fault localization. *Inf. Softw. Technol.* 55, 5 (2013), 866–879. <https://doi.org/10.1016/j.infsof.2012.08.008>
- [55] Runsheng Xu, Xin Xia, Jinlong Li, Hanzhao Li, Shuo Zhang, Zhengzhong Tu, Zonglin Meng, Hao Xiang, Xiaoyu Dong, Rui Song, et al. 2023. 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*. 13712–13722. <https://doi.org/10.1109/CVPR52729.2023.01318>
- [56] Runsheng Xu, Hao Xiang, Zhengzhong Tu, Xin Xia, Ming-Hsuan Yang, and Jiaqi Ma. 2022. V2x-vit: Vehicle-to-everything cooperative perception with vision transformer. In *European conference on computer vision*. Springer, 107–124. [https://doi.org/10.1007/978-3-031-19842-7\\_7](https://doi.org/10.1007/978-3-031-19842-7_7)
- [57] Runsheng Xu, Hao Xiang, Xin Xia, Xu Han, Jinlong Li, and Jiaqi Ma. 2022. 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, 2583–2589. <https://doi.org/10.1109/ICRA46639.2022.9812038>
- [58] Ruozhou Yu, Dejun Yang, and Hao Zhang. 2021. Edge-assisted collaborative perception in autonomous driving: A reflection on communication design. In *2021 IEEE/ACM Symposium on Edge Computing (SEC)*. IEEE, 371–375.
- [59] Ekim Yurtsever, Jacob Lambert, Alexander Carballo, and Kazuya Takeda. 2020. A survey of autonomous driving: Common practices and emerging technologies. *IEEE access* 8 (2020), 58443–58469. <https://doi.org/10.1109/ACCESS.2020.2983149>
- [60] Quanjun Zhang, Juan Zhai, Chunrong Fang, Jiawei Liu, Weisong Sun, Haichuan Hu, and Qingyu Wang. 2024. Machine Translation Testing via Syntactic Tree Pruning. *ACM Trans. Softw. Eng. Methodol.* 33, 5 (2024), 125:1–125:39. <https://doi.org/10.1145/3640329>

Received 2024-04-12; accepted 2024-07-03