WILEY | Hindawi

Research Article

An Advanced System-Level Testing for Roadside Multimodal Sensing and Processing in IoV

Yancong Wang [ID],[1] **Jian Wang** [ID],[1] **Xuyan Bao** [ID],[2] **Bingyan Yu** [ID],[2] **and Yuming Ge** [ID][2]

[1]*College of Computer Science and Technology, Jilin University, Changchun 130012, China*
[2]*Key Laboratory of Internet of Vehicle Technical Innovation and Testing (CAICT), China Academy of Information and Communications Technology, Beijing 100191, China*

Correspondence should be addressed to Xuyan Bao; baoxuyan@caict.ac.cn

Currently, there are mature test methods for specific sensing devices or processing devices in the Internet of Vehicles (IoV). However, when a system is combined with these different types of devices and algorithms for real scenarios, the existing device-level test results cannot reflect the comprehensive functional or performance requirements of the IoV applications at the system level. Therefore, novel application-oriented system-level evaluation indexes and test methods are needed. To this end, we extract the data processing functional entities into specific and quantifiable evaluation indexes by considering the IoV application functions and performance requirements. Then, we build a roadside sensing and processing test system in a real test zone to collect and process these evaluation indexes into accurate multidimensional ground-truth. According to the actual test results of multiple manufacturers' solutions, our proposed test method is verified to effectively evaluate the performance of the system-level solutions in real IoV application scenarios. The unprecedented evaluation indexes, system-level test method, and the actual test results in this paper can provide an advanced reference for academics and industry.

1. Introduction

In the Internet of Vehicles (IoV), in-vehicle entertainment, traffic efficiency, and safety applications rely on the real-time and dynamically perceived surrounding vehicles and road information [1]. The information is first sensed by the sensors as raw data, and then, data is processed by the computing devices.

The sensors are assembled on the vehicle or the roadside, including cameras, millimetre-wave radars (mmWave radars), and lidars [2]. The camera has robust image recognition capability and a lower price, which provides video and image information in IoV [3]. However, the camera is so susceptible to light that its raw data has poor quality in strong light or dark environments [4]. Besides, the dynamic sensing range of the camera is limited, and a single camera

cannot provide three-dimensional information [5]. The mmWave radar is used for speed and distance detection in IoV. It has a more comprehensive range of perception and is less susceptible to environmental influences because of its excellent penetration ability [6]. But mmWave radar does not have a robust recognition capability as lidar. False detection and missing detection may occur due to the uncertainty of the effective echo [7]. Lidar has advantages in sensing stability, response time, distance measure accuracy, and so on, which achieves the high precise sensing of the traffic environment. Lidar is not affected by the light but the specular reflection on rainy days. The amount of data obtained by lidar is far greater than mmWave radar, which requires higher-performance computing capacity to support, and its deployment cost is also relatively high [8, 9]. A single sensor lacks the sensing ability required by IoV applications in

complex scenarios [10]. Therefore, multimodal sensing information fusion from multiple sensors is emerging as a promising technology, which provides a more reliable and wider range of perception capabilities without being affected by the environment [11, 12]. But multimodal sensing for a single vehicle cannot increase the vehicle's perception range. By means of cooperation between vehicles and roadside infrastructures, multimodal sensing information can be shared beyond the single vehicle's light of sight and used for complex traffic scenarios [13].

With the increment of sensing information from massive sensors, the amount of multimodal data processed in IoV increases exponentially. On the one hand, data processing performance depends on the computing capacity of the hardware equipment. To meet the requirement of vehicular application's computation intensiveness and delay sensitivity, multiaccess edge computing (MEC) is proposed to process the sensing data closer to the vehicle than cloud computing [14]. MEC has a relatively robust computing and storage capacity to process the data fusion algorithm timely and a multiaccess communication capacity to transmit the processed data, reducing the end-to-end delay of the applications [15]. On the other hand, the selection of the data fusion algorithm also influences the data processing performance. The algorithms for fusing the collected multimodal sensing data in IoV are mainly divided into three categories: raw data fusion (early-fusion), feature-level fusion, and target-level fusion [16, 17]. Raw data fusion refers to completing the superposition and fusion of multimodal data (images, point clouds, etc.) before the target feature extraction, which preserves the raw information to the greatest extent and generates high-precision sensing data [18, 19]. However, few manufacturers currently select raw data fusion algorithms because of the problems such as complex time-space synchronization and large consumption of computing capacity for implementation. In contrast, target-level fusion algorithms are the most common and widely used in the industry. The target-level fusion algorithm fuses the structured data generated after raw data is returned by each sensor independently, which is easy to implement and has agile deployment [20]. The disadvantage of target-level fusion algorithms is the loss of the raw data accuracy in the independent data processing, resulting in a certain degree of reduction in the accuracy of fusion sensing. Some manufacturers also choose feature-level fusion algorithms which extract the features of raw information from different sensors and then comprehensively analyse and process the feature information [21–24]. Lidar and camera fusion are generally used as feature-level fusion, which reduces the difficulty of implementation, further enriches semantic information, and ensures the accuracy of fusion sensing to a certain extent [25, 26].

It is necessary to test the functionality and performance of sensing and data processing devices before applying them in real scenarios. At present, the test methods for sensors like cameras, mmWave radars, and lidars are relatively mature and have established the corresponding national or industry test standards. For cameras, the related standards involve the China group standard T/ITS 0184-2021 "Testing specifica-

tions for intelligent analysis function of road cameras" and China group standard T/ITS 0171-2021 "Intelligent transportation system—Technical requirements for roadside cameras interface." The test standards of the mmWave radar are China group standard T/ITS 0128-2021 "Intelligent transportation system—traffic condition detector by millimetre-wave radar" and China group standard T/ITS 0172-2021 "Intelligent transportation system—Technical requirements for the interface of millimetre-wave radar traffic condition detector." The standard related to lidar testing is China group standard T/ITS 0173-2021 "Intelligent transportation system—Technical requirements for roadside lidar interface." The processing device we mainly focused on as roadside MEC has also developed a method to test the capacity of its southern API linked to the sensing devices and measure its computing capacity for data processing. However, these existing test methods for the single device cannot be directly applied to a system-level evaluation in real scenarios. Specifically, Reference [27] proposes a new method of lidar simulation, which implements the rapid creation of point cloud data with accurate point-level labels using a computer game and a method for automatic calibration between point clouds and captured images. Reference [28] focuses on the key function test at the system level of autonomous driving software, which is based on the simulated vehicle model that realizes the function. In Reference [29], considering the relationship between different types of vehicle kinematic simulation software, Minnerup and Knoll combine different simulation software to meet more simulation scenarios' requirement. A large-scale complex traffic network test environment is simulated in [30] based on Microsoft AirSim. The vehicle kinematic model, virtual reality environment model, automatic driving software, and radar sensors are combined to form their autonomous driving test platform in a virtual environment. Reference [31] realizes a virtual environment for testing autonomous driving by constructing the traffic environment of the test vehicle; the simulation components of the camera, radar, and lidar; and the data required by these sensors. Zhang et al. [32] implement a new traffic scene modelling method based on image sequences and road GIS data in the Road-View system, which is used for performance evaluation and testing of assisted autonomous driving software. Reference [33] uses lidar and cameras to scan the real traffic scenarios and generates the reasonable traffic flows of vehicles and pedestrians from the acquired trajectory data, which can be used for test scenario simulation. Reference [34] gives a machine learning model to implement environmental perception to test the sensing devices. One of the reasons is that the definitions of test indexes changed. For instance, the sensing delay in the device-level test only represents the time taken by the single sensor to collect the raw data, which is not suitable for describing the comprehensive sensing performance in the system-level test. Another reason is the test method needs to be redesigned from "device-level" to "system-level" when devices and algorithms are coupled into a system.

To measure the better performance of roadside sensing and processing system in real scenarios, we first propose a

novel multidimensional evaluation index system based on IoV application functions and performance requirements. Then, we build a roadside sensing and processing test system in a real test zone to obtain the ground-truth of the evaluation indexes. Finally, multiple manufacturers participated in our test, and the test results show our test method can effectively detect application-oriented system-level performance in real IoV scenarios. The contributions of our work include the proposed evaluation indexes, system-level test method, the collected ground-truth, and actual test results mentioned above, which have a positive effect on academic research and industrial development.

The remaining of this paper is organized as follows. Section 2 introduces the IoV general architecture. The evaluation index system is proposed in Section 3. Section 4 elaborates on the test method and shows the test results with analysis. Finally, in Section 5, we summarize this work and introduce future work.

2. Architecture

In this section, a general architecture that supports multimodal sensing and processing to meet application functional and performance requirements on the IoV is introduced in Figure 1. It is a four-layer architecture, from bottom to top, roadside equipment and end-user layer, MEC platform layer, central cloud platform layer, and IoV application layer.

At the lowest layer of the architecture, the multimodal data derives from roadside/vehicular sensing devices and intelligent traffic management infrastructures. The roadside/vehicular sensing devices (cameras, mmWave radars, lidars, etc.) collect raw data such as images, videos, and point clouds in traffic scenarios. The intelligent traffic management infrastructures (traffic signals, electronic identifications, etc.) periodically upload self-generated real-time data to upper-layer devices in IoV [35, 36]. In the following, we only focus on the multimodal sensing data collected by various sensors.

The multimodal data is first uploaded to roadside MEC equipment before accessing the upper MEC layer. Roadside sensing devices and intelligent traffic management infrastructures are generally wired link to roadside MEC equipment via star network topology to upload multimodal data. The sensing data of vehicular sensing devices are first wirelessly transmitted to the roadside unit (RSU) through the onboard unit (OBU) and then forwarded to the roadside MEC equipment. After collecting raw data from sensors, roadside MEC equipment generates structured sensing data through time-space synchronization and data fusion execution, including status information of traffic participants and real-time traffic target and event detection information.

Roadside MEC equipment is deployed near end-users to process the sensing data with low latency. The processed data, for one way, is broadcast to various traffic participants via RSU, providing more diverse and detailed information on collaborative perception, decision-making, and control to these traffic participants. For another, the preprocessed data is uploaded to the upper layer, providing data to support various applications on the regional MEC platform and the central cloud platform.

The processed sensing information is ultimately used to support the implementation of various IoV applications. We will discuss the IoV applications in detail in the next section. The realization of applications depends on the combined implementation of different data processing functional entities. The data processing functional entities in IoV are divided into traffic participant detection, traffic participant localization, traffic participant tracking, traffic participant recognition, traffic flow detection, and so on. In general, these functional entities are derived through research from the standards and technical discussions. According to the scenarios defined in the standards and recent industry development, we have extracted the common requirements as the functional entities for roadside perception systems in various scenarios.

In this work, we focus on the road sensing and processing system, which involves the roadside MEC equipment and the road sensing devices (framed by the red dashed line in Figure 1).

3. Evaluation Index System

Corresponding to the data processing application functional entities in the above architecture, we first convert them to specific, quantifiable indexes and then test these indexes in real scenarios for obtaining the ground-truth of our proposed roadside sensing and processing test system. In Subsection 3.1, we first introduce the classified IoV applications' functional and performance requirements for sensing and data processing capabilities. Subsection 3.2 describes the specific evaluation indexes used in the test system described in the next section.

3.1. IoV Applications. We divide the applications into the following four categories according to the capability requirements of IoV applications for perception information provided by multimodal data.

(1) *Network Connection Applications.* The realization of these applications requires the network connection to obtain external device (traffic signal, positioning facility, etc.) information without sensing information. Typical applications include traffic light countdown, near-field payment, floating car data collection, and road section traffic control instruction distribution.

(2) *Basic Perception Applications.* Sensing and computing devices are required to provide some basic perception information (traffic flow data, weather parameters, infrastructure status information, etc.) for these applications. Applications such as traffic flow detection, traffic incident identification, and some safety warnings based on vehicle-road collaboration are sorted into this category.

(3) *Enhanced Perception Applications.* Compared with (2), these applications require the perception of
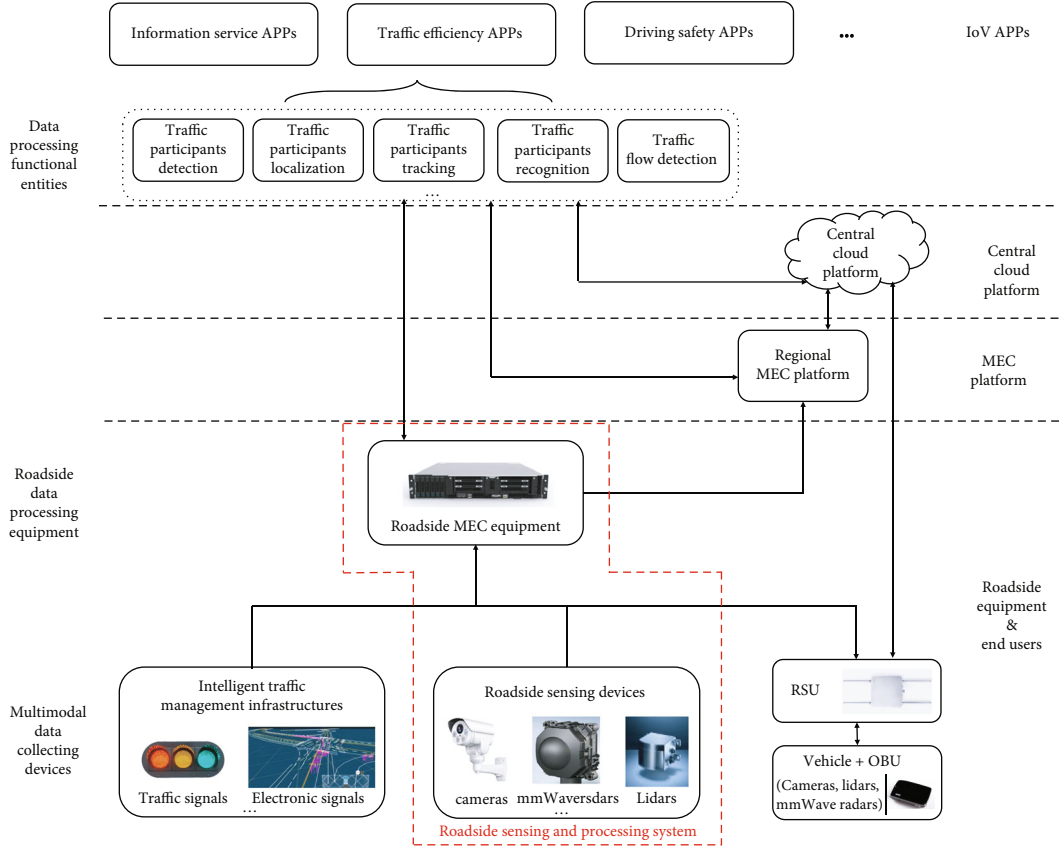
FIGURE 1: Architecture supports multimodal data sensing and processing in IoV applications.

more detailed, more precise, and more efficient traffic microbehaviours (sensing range, sensing delay, attribution, and status recognition of traffic participants), for example, intersection collision warning, vulnerable road user collision warning, and sensing data sharing.

(4) *Collaborative Decision and Control Applications.* These applications have the highest demands on perception, which requires all-weather, no blind spots, and robust multimodal data for perception. The applications enable millisecond-level system latency, centimetre-level positioning accuracy, extremely high detection accuracy of traffic incidents and traffic flow, and continuous coverage sensing, for instance, collaborative lane change, collaborative merge on the highway, and crossing the intersection without signal based on vehicle-road collaboration.

3.2. Evaluation Index System. Considering the data processing functional entities and the application's performance requirements, we convert the functional entities into specific, quantifiable evaluation indexes shown in Figure 2. Here, we mainly consider the applications with perception demands. The evaluation indexes are divided into five dimensions: indexes of system essential capability, target recognition capability, target positioning capability, and traffic flow

detection capability. We will describe the basis for index selection and the definition of the indexes below.

From the perspective of the existing industry maturity, the index system is complete. The indexes of system essential capability, vehicular kinematics, and target classification are common basic indexes, which quantify the specific performance of the roadside perception system. The performance of functional indexes such as the traffic flow detection capability index is restricted by common basic indexes. Generally speaking, a roadside perception system with high accuracy of common basic indexes is positively correlated with its performance in functional indexes.

(1) System essential capability indexes include sensing range, sensing delay, response time, sensing frequency, and the maximum number of detected targets

(i) Sensing range, strongly related to the value of positioning accuracy and considered the maximum boundary that continuously outputs the target state information under the positioning accuracy required by the application layer from the perspective of supporting business continuity in real scenarios, is defined as the maximum distance at
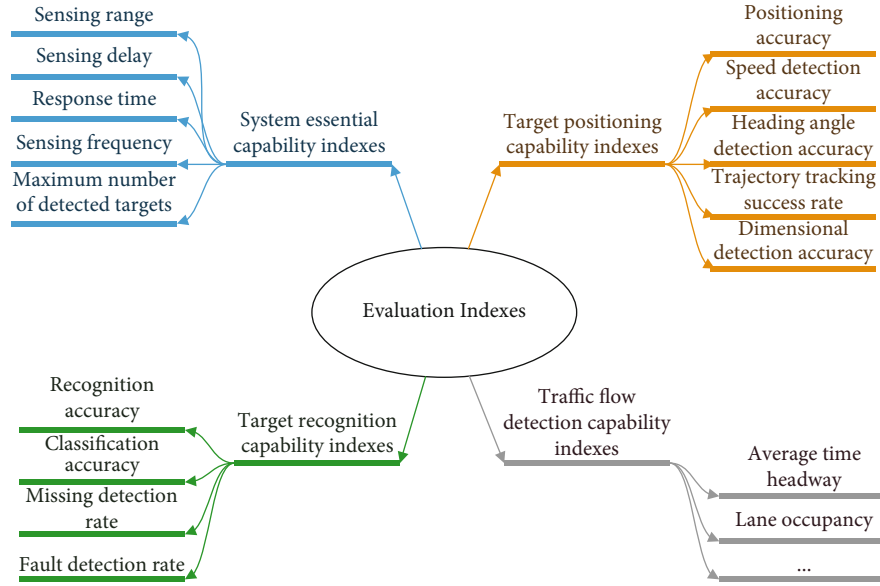
FIGURE 2: Evaluation indexes for roadside sensing and processing system in IoV.

which the roadside sensing and processing system can stably detect traffic participants or traffic incidents with given positioning accuracy

(ii) Sensing delay, which represents the perception and computing performance of the roadside sensing and processing system, is defined as the delay from raw data acquisition by sensors to generating structured data by data fusion processing. Sensing delay is observed from the perspective of the roadside equipment, the time difference from the start of the first sensing device obtained the time the first frame of sensing data, and the result calculated by roadside MEC equipment. Sensing delay can also be considered as the inverse of the sensing frequency

(iii) Response time, evaluating the time deviation when the vehicle and roadside sensors observed the target at the same position and used to support various vehicle-road collaborative applications, is defined as the time deviation between the moment the roadside sensing and processing system detects the presence of traffic participants at any position and the moment when they are present. Response time is observed from two perspectives of the vehicle and the roadside equipment. From the vehicle's perspective, the vehicle records the ground-truth of its absolute position and the time corresponding to the position. From the roadside equipment's perspective, the time of the vehicle's position is also recorded. Response time is the time difference between the recorded time from both vehicle and the roadside equipment's perspective. For IoV applications, response time is more important and practical than sensing delay. Because when the vehi-

cle arrives at a certain position, it is more important to record the accurate time from roadside equipment rather than observing the data from the perspective of the vehicle or roadside equipment

(iv) Sensing frequency, which is used to evaluate the number of sensing message frames generated by the roadside sensing and processing system per unit of time and to guide the vehicle to design the receiving mechanism of roadside messages, is defined as the instantaneous frequency mean of the sensing messages sent by the roadside sensing and processing system within a given sampling period

(v) The maximum number of detected targets is defined as the maximum number of targets detected by the sensing and processing system in a given period

(2) Target recognition capability indexes include recognition accuracy, classification accuracy, missing detection rate, and fault detection rate, which are used to quantify the perception recognition capability of the roadside sensing and processing system

(i) Recognition accuracy is defined as the ratio of the number of correctly recognized samples by the roadside sensing and processing system to the actual total number of samples of a particular class of traffic participants in the given test samples

(ii) Classification accuracy is defined as the ratio of the number of samples that the roadside sensing and processing system perceive correctly to the total

number of samples of this particular class of traffic participants in the given test samples

(iii) Missing detection rate is defined as the ratio of the number of samples that the roadside sensing and processing system does not perceive to the actual total number of samples of a particular class of traffic participants in the given test samples

(iv) Fault detection rate is defined as the ratio of the number of samples that the roadside sensing and processing system perceived does not match their accurate class to the actual total number of samples of a particular class of traffic participants in the given test samples

(3) Target positioning capability indexes include positioning accuracy, speed detection accuracy, heading angle detection accuracy, trajectory tracking success rate, and dimensional detection accuracy. Positioning accuracy, speed detection accuracy, and heading angle detection accuracy are used to describe the deviation of the motion state of traffic participants perceived by the roadside sensing and processing system from its actual motion state, which reflects the perception accuracy of the roadside sensing and processing system to the traffic microbehaviours at any given moment. Trajectory tracking success rate and dimensional detection accuracy are used to evaluate the roadside sensing and processing system's capability to detect the trajectory tracking of traffic participants and the spatial size of the targets

(i) Positioning accuracy is defined as the Euclidean distance between the latitude and longitude of the traffic participants perceived by the roadside sensing and processing system and their actual latitude and longitude

(ii) Speed detection accuracy is defined as the deviation between the value of the traffic participants' speed detected by the roadside sensing and processing system and the value of their actual speed

(iii) Heading angle detection accuracy is defined as the deviation between the value of the traffic participants' heading angle detected by the roadside sensing and processing system and the value of their actual heading angle

(iv) Trajectory tracking success rate is defined as the ratio of the number of targets that the roadside sensing and processing system stably tracks, shown as targets' ID unchanged to the actual number of targets within a certain time period

(v) Dimensional detection accuracy is defined as the deviation between the value of the spatial size of the traffic participants measured by the roadside sensing and processing system and the value of their actual spatial size

(4) Traffic flow detection capability indexes are used to evaluate the capability of the roadside sensing and processing system to detect traffic flow information, which describes the traffic volume and concentration like average time headway and lane occupancy. The concepts of the traffic flow indexes are depicted in the typical traffic flow theory, and we will not describe them in detail here. The most accurate way to measure these indexes is manual counting. In addition, it is not easy to record the time parameter and test these indexes at the system level, because there are many combination calculations of the targets' classification and targets' status indexes involved in testing these indexes. And the calculations are based on accurately identifying the targets. Therefore, we selected the relative error of the traffic flow of each lane in the truncation plane in the same direction within the specified time as the traffic flow detection capacity index in this paper, which is more suitable for system-level testing

4. Test Method and Results

After giving the evaluation index system, in this chapter, we will introduce the roadside sensing and processing test system built by adopting these indexes detected in real scenarios as the ground-truth. Comparing the test results with the ground-truth of the test system, the performance of different solutions (sensing device deployments and algorithm schemes) based on the proposed architecture can be tested and evaluated. The test zone, composition, and provided ground-truth of the roadside sensing and processing test system are presented in Subsection 4.1. We introduce the test method in Subsection 4.2. The test results are shown and analysed in Subsection 4.3.

4.1. Test System. Our proposed test system is the first time to collect the system-level ground-truth for roadside sensing and processing system in real scenarios, providing a reliable solution and dataset reference for system-level testing of roadside sensing and processing system and simulation in industry and academia.

The specific intersection used in the city is selected as the test zone for the following reasons. The test zone with dense traffic flow, multitype traffic participants, and weak signal blockage can simulate the typical application scenarios in the IoV to the maximum extent. In addition, rich infrastructure resources, a fully connected fibre-optic network, and the abundant power supplement in the test zone can provide the essential environment for meeting the common deployment requirements. The test zone's live picture and simulation scene are shown in Figures 3(a) and 3(b), respectively.
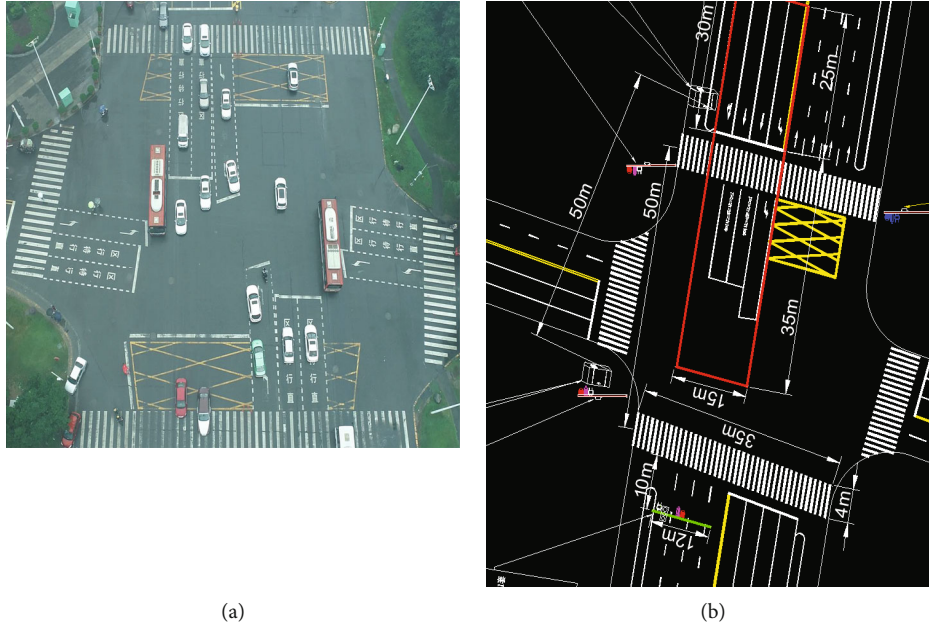
(a)

(b)

FIGURE 3: The specific in-use intersection as the test zone: (a) live picture; (b) simulation scene.

We build the roadside sensing and processing test system in the test zone and collect the ground-truth of the evaluation indexes. The roadside sensing and processing test system consists of the vehicular ground-truth test system and the roadside ground-truth test system. The vehicular ground-truth test system provides reference ground-truth for single-target-oriented evaluation indexes, such as system essential capability indexes, vehicular kinematic indexes, and traffic incident detection capability indexes. The roadside ground-truth test system provides reference ground-truth for multi-target-oriented evaluation indexes, such as target recognition capability indexes, trajectory tracking success rate, and traffic flow detection capability indexes.

In the vehicular ground-truth test system, the vehicle is equipped with high-precision integrated inertial navigation RT-Range, providing the reference ground-truth of the vehicular kinematic indexes such as the vehicle's position, pitch angle, heading angle, and other kinematic information which is shown in Figure 4. The vehicle is equipped with onboard RT-Range, including the RT-XLAN antenna, RT3000, and a host computer. The positioning accuracy of the vehicle can be premise within 0.02 metres (≤ 0.02 m) under a good signal environment. In the actual test process, the vehicle's positioning deviation in an area with very few buildings will be limited to within 0.1 metres (≤ 0.1 m), caused by comprehensive factors such as signal blockage and error compensation from gyroscope after the vehicle outages. The main parameters of the indexes are described in Table 1.

Figure 5 shows that the roadside ground-truth test system consists of various sensors and an offline service processor. Lidar, sets of cameras, and backup roadside computing equipment are integrated on the roadside. Lidar has 128 laser beams, and cameras collect point clouds and videos, respectively. The raw sensing data will be back transmitted

to the offline service processor for offline AI processing instead of processing on the roadside equipment, which is aimed at obtaining more accurate roadside target information. On this basis, the processed image is calibrated by manually reviewing the videos. Considering the uncertainty of AI processing, it is not easy to quantify the deviation of the roadside ground-truth test system. In the actual evaluation, we sample four real scenarios at the intersection and manually calibrated all data as the absolute ground-truth. Then, we compare the result generated by the roadside ground-truth test system and the absolute ground-truth, which shows that our roadside ground-truth test system can meet the following index requirements: (1) positioning accuracy ≤ 0.05 m, (2) classification accuracy > 95%, (3) missing detection rate < 3%, (4) fault detection rate < 3%, and (5) tracking success rate > 90%.

The reasons why our proposed ground-truth test system is reliable for testing the DUT are as follows. For the vehicular ground-truth test system, the equipped integrated inertial navigation RT-Range is higher precise than the mass-produced DUT. For the roadside ground-truth test system, all raw data collected by different sensors is transmitted back to the offline service processor for data processing. However, the DUT needs to process a large amount of data in real time, which will force the DUT to appropriately discard a part of the data. In addition, we manually calibrated the processed data by random sampling. The test results verify that the accuracy of our ground-truth system is much higher than that of the DUTs.

4.2. Test Method. We further extract the evaluation indexes except for the traffic flow detection capacity indexes in Subsection 3.2 into seven categories of indicators which are shown in Figure 6: (1) sensing range, (2) sensing delay, (3) sensing frequency, (4) vehicular kinematic indexes, (5)
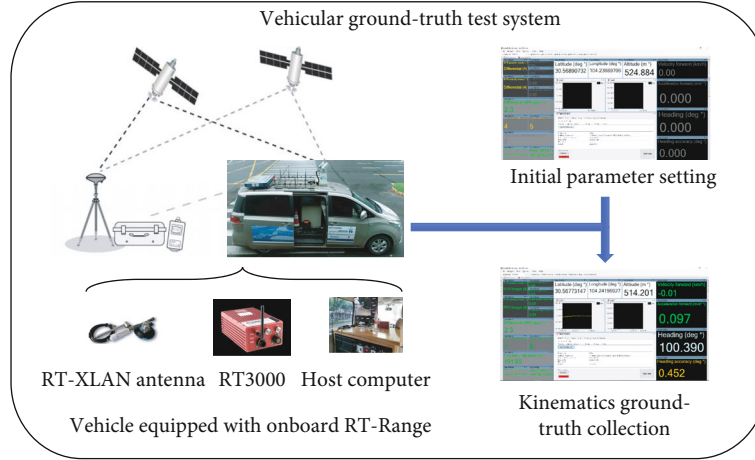
FIGURE 4: The vehicular ground-truth test system.

TABLE 1: Main parameters of the indexes.

Indexes	Parameters
Positioning accuracy	SPS ≤ 1.5 m, RTK ≤ 0.01 m
Heading angle detection accuracy	≤0.1°
System frequency	≥100 Hz
Attitude accuracy	≤0.03°
Speed detection accuracy	≤0.05 km/h
Tracking angle	≤0.07
Offset angle	≤0.15°
Acceleration range	±10g
Acceleration offset	≤2μg
Angular velocity range	≥100°/s
Angular velocity offset	≤2°/hr
Communication coverage	≥500 m
Relative distance error	≤0.03 m
Relative speed error	≤0.02 m/s
Tracking target number	≥4
Line number	≥8

trajectory tracking success rate, (6) identification accuracy indexes, and (7) traffic flow detection capability. During the testing process, four indexes ((1)–(4)) can be tested simultaneously by a single-target mobile terminal with high precision positioning for its state, as by the vehicular ground-truth test system, which will be called the ground-truth vehicle below. Testing the other two indexes ((5), (6)) and the traffic flow detection capacity indexes requires the test system to perceive and recognize all traffic participants at any time, which can be tested by the roadside ground-truth test system. The relationship between evaluation indexes and roadside sensing and processing test system is shown in Figure 6. We will specifically introduce our test method in the actual testing process below.

(1) *Sensing Range*. The ground-truth vehicle records its position in real time, driving from outside into the system's sensing range. Suppose the DUT can detect the vehicular ground-truth test system in 10 consecutive frames under the given positioning accuracy. In that case, the sensing range is calculated as the Euclidean distance between the position of the first frame and the sensor calibration position.

(2) *Response Time*. A reference line perpendicular to the direction of the lane line is first to be selected; then, the ground-truth vehicle drives to the reference line within the sensing range. The moment when the vehicle reaches the reference line is recorded, and the moment when the vehicle reaches the reference line is extracted from the sensing message output by the DUT. Calculate the time difference between these two moments as the sensing delay.

(3) *Sensing Frequency*. Recording the timestamps of adjacent sensing data frames, calculating the instantaneous time interval ϵ, the instantaneous system frequency is $1/\epsilon$.

(4) *Positioning Accuracy in Vehicular Kinematic Indexes*. The Euclidean distance between the position output by the ground-truth vehicle and the position where the DUT sensed the vehicle in a given period.

(5) *Trajectory Tracking Success Rate*. Selecting the data frame with the largest number of targets and several frames before and after, generating the ground-truth sample B of the trajectory, and initializing the number of trajectories that the DUT can track stably as $A = 0$, DUT is time-aligned with the data frame of the ground-truth. If the target distance between the DUT and the ground-truth is less than the threshold and associated with the target's ID, for each target, if there is only one ID corresponding to the target ID of ground-truth, we set $A = A + 1$. The success rate of multitarget trajectory tracking is $A/B \ast 100\%$.

(6) *Identification Accuracy Indexes*. Setting the sample size threshold according to the traffic flow of the real scenario, the DUT detects synchronously with the
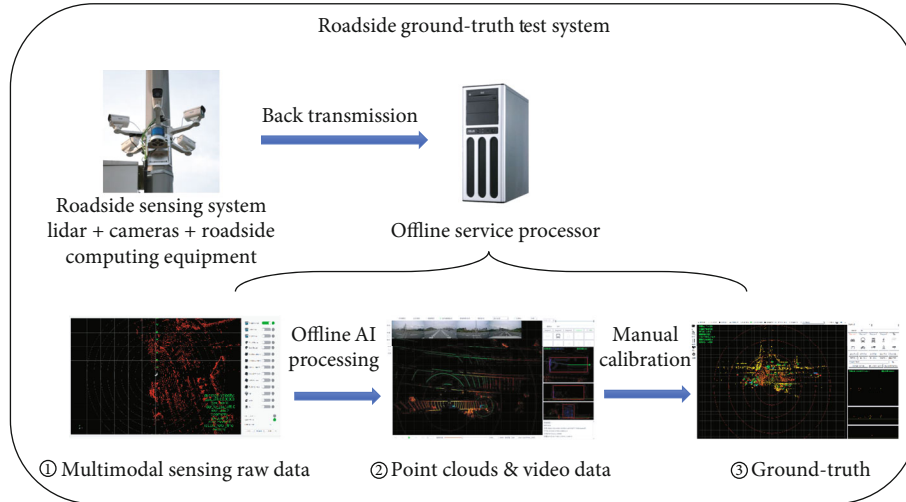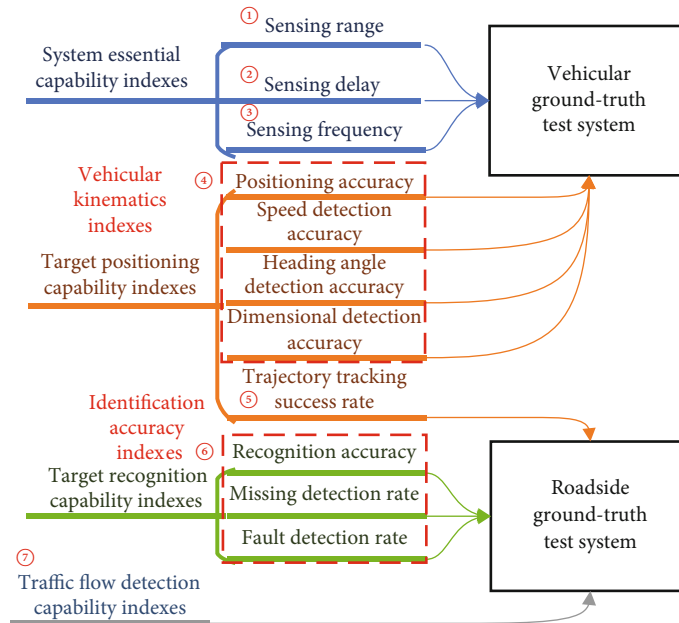
Figure 5: The roadside ground-truth test system.

Figure 6: The relationship between evaluation indexes and roadside sensing and processing test system.

Table 2: Different manufacturers' solutions.

Solution	1	2	3	4	5	6	7	8	9	10	11	12	13
Lidars	✓	✓											
Integrated radar-video machine			✓	✓									
Camera+mmWave radar					✓	✓							
Camera+lidar							✓	✓	✓				
Camera+lidar mmWave radar										✓	✓	✓	✓
Roadside MEC			✓			✓		✓	✓	✓	✓		

roadside ground-truth system to collect enough samples. The ground-truth of each frame is generated by the roadside ground-truth system, which is compared with the DUT detected result.

(7) Traffic flow detection capability is reflected by the relative error of the traffic flow of each lane in the truncation plane in the same direction within the

	1	2	3	4	5	6	7	8	9	10	11	12	13
	0	33.6	0	0	0	0	46.8	0	0	0	0	0	0
	111.1	34.3	95.04	160.4	0	0	77.22	0	83.8	0	0	0	0
	111.1	36.53	119.8	237.6	0	0	77.22	0	103.8	0	45.12	0	79.09
	112.1	40.62	196.8	237.6	0	71.07	78.43	0	127.7	0	47.21	43.18	79.09

Solutions

- ■ ≤ 50 cm
- ■ ≤ 100 cm
- ■ ≤ 150 cm
- ■ ≤ 200 cm

(a)

	1	2	3	4	5	6	7	8	9	10	12	13
■	−86	−113	215.3	−532	164.2	153	−6	620.5	−109	2600	−251	90.17

Solutions

(b)

	3	4	5	6	7	8	9	10	11	12
■	69.33	37	131.33	321.4	107.5	93.5	132.33	134	49	51

Solutions

(c)

FIGURE 7: Continued.

	1	2	3	4	5	6	7	8	9	10	11	12	13
	9.97	9.86	11.57	26.55	5.15	6.99	9.15	2.99	9.65	0	14.54	9.99	9.97
	9.73	9.85	0	26.31	5.03	9.66	9.39	2.96	8.55	8.596	14.93	8.99	10

Solutions

■ Offpeak period
■ Peak period

(d)

	1	2	3	4	5	6	7	8	9	10	11	12	13
	21	18	41	53	30	17	52	55	73	0	29	53	22
	21	22	0	50	30	18	62	62	127	32	47	59	20

Solutions

■ Offpeak period
■ Peak period
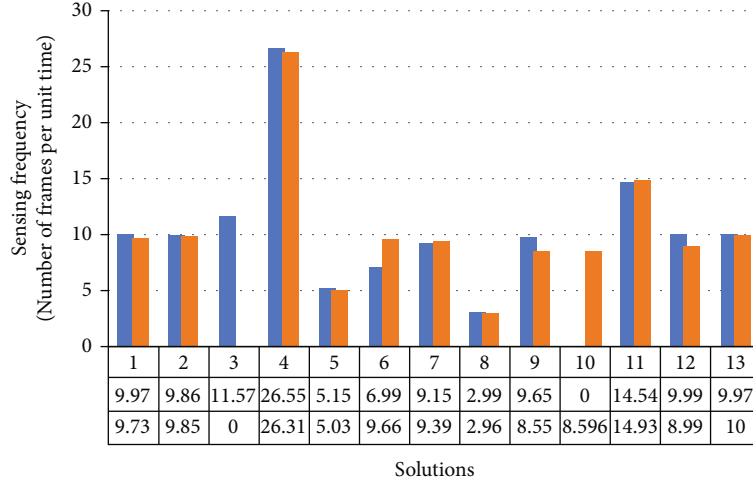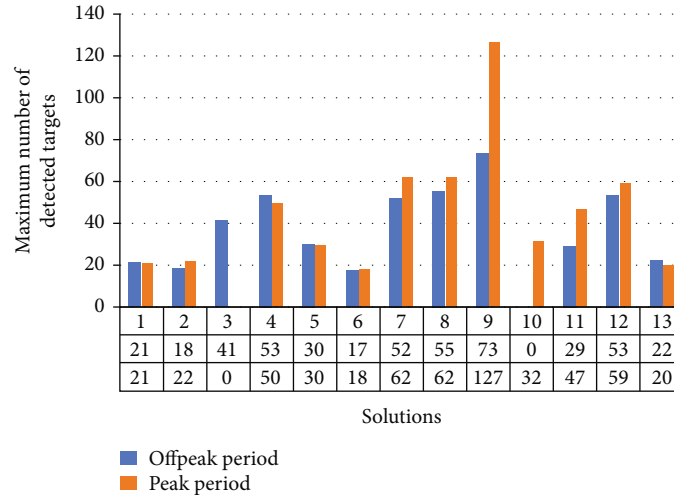
(e)

FIGURE 7: System essential capability indexes for different solutions: (a) sensing range (m); (b) response time (ms); (c) sensing delay (ms); (d) sensing frequency (no./s); (e) the maximum number of detected targets (no.).
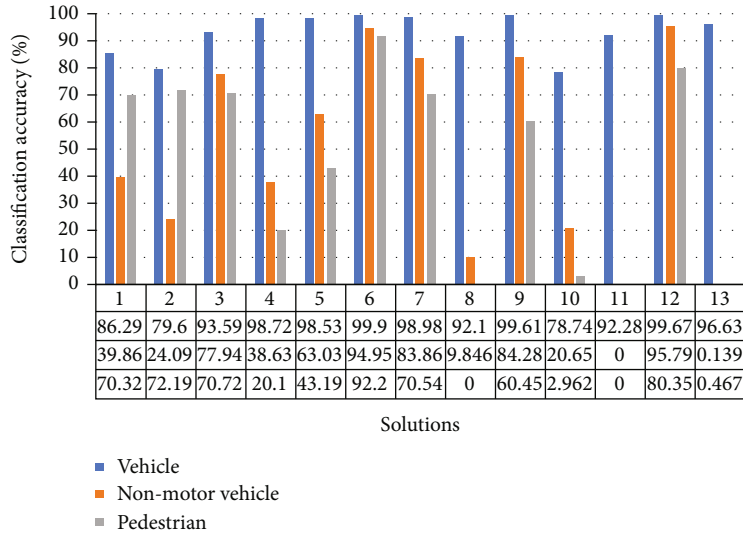
specified time, shown in Formula (1). In the actual test, we set a truncation plane that covers four lanes (Lane 1: left turn lane, Lanes 2 and 3: straight lane, and Lane 4: right turn lane) in the test zone. The sensing and processing systems under test collect the traffic flow data of the truncation plane and each lane within a specified time (10 minutes). The road-side ground-truth test system records the traffic flow statistics on the lanes during the test. If there exists any disagreement with the test results of the roadside ground-truth test system, we will manually recheck the test results by videos:

$$P_{\text{traffic flow}} = |T_{\text{actual}} - T_{\text{test}}| \div T_{\text{actual}} \times 100\%, \quad (1)$$
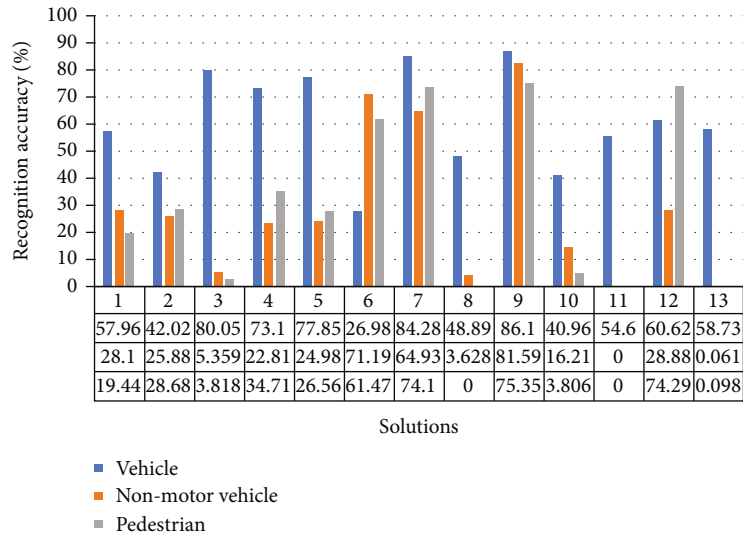
where $P_{\text{traffic flow}}$ (%) represents the relative error between the detected traffic flow and the actual traffic flow. T_{actual} is

defined as the number of vehicles passing the test truncation plane in a given time in the same direction. T_{test} denotes the number of vehicles passing the test truncation plane detected by the sensing and processing system under test in a given time in the same direction. $|x|$ is expressed as the absolute value of x.
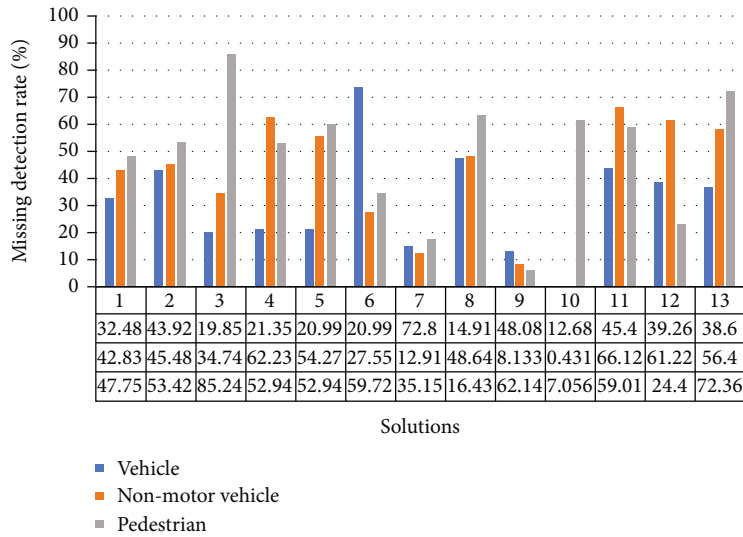
4.3. Test Results and Analysis. There are 13 roadside sensing and processing solutions from different manufacturers that participated in our test. Considering technical protection and other issues, we anonymize and obfuscate the manufacturers' names and devices' information. But the solutions selected by each manufacturer and the corresponding test data shown here are real and accurate. In the test process, almost all manufacturers have chosen the late-fusion data fusion algorithms, and some of them directly deploy the MEC server on the sensor side. The multimodal sensor combination schemes selected by different manufacturers are
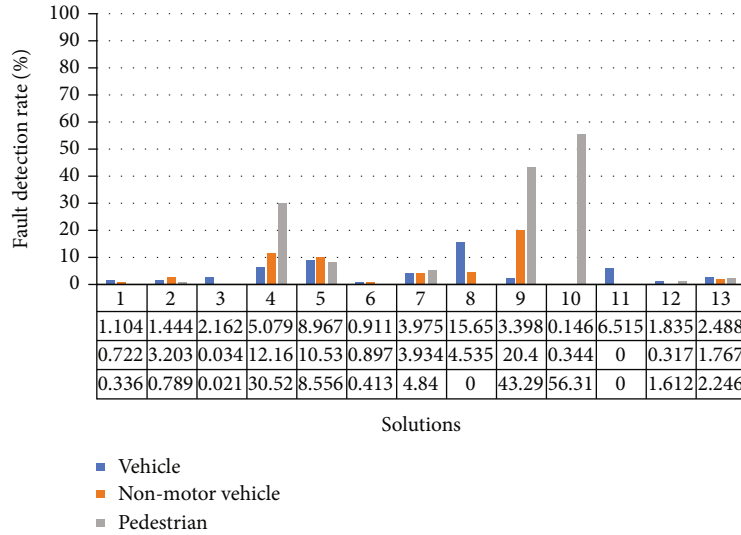
	1	2	3	4	5	6	7	8	9	10	11	12	13
	86.29	79.6	93.59	98.72	98.53	99.9	98.98	92.1	99.61	78.74	92.28	99.67	96.63
	39.86	24.09	77.94	38.63	63.03	94.95	83.86	9.846	84.28	20.65	0	95.79	0.139
	70.32	72.19	70.72	20.1	43.19	92.2	70.54	0	60.45	2.962	0	80.35	0.467

Solutions

■ Vehicle
■ Non-motor vehicle
■ Pedestrian

(a)

	1	2	3	4	5	6	7	8	9	10	11	12	13
	57.96	42.02	80.05	73.1	77.85	26.98	84.28	48.89	86.1	40.96	54.6	60.62	58.73
	28.1	25.88	5.359	22.81	24.98	71.19	64.93	3.628	81.59	16.21	0	28.88	0.061
	19.44	28.68	3.818	34.71	26.56	61.47	74.1	0	75.35	3.806	0	74.29	0.098

Solutions

■ Vehicle
■ Non-motor vehicle
■ Pedestrian

(b)

	1	2	3	4	5	6	7	8	9	10	11	12	13
	32.48	43.92	19.85	21.35	20.99	20.99	72.8	14.91	48.08	12.68	45.4	39.26	38.6
	42.83	45.48	34.74	62.23	54.27	27.55	12.91	48.64	8.133	0.431	66.12	61.22	56.4
	47.75	53.42	85.24	52.94	52.94	59.72	35.15	16.43	62.14	7.056	59.01	24.4	72.36

Solutions

■ Vehicle
■ Non-motor vehicle
■ Pedestrian

(c)

FIGURE 8: Continued.

FIGURE 8: Target recognition capability indexes for different solutions: (a) recognition accuracy (%); (b) classification accuracy (%); (c) missing detection rate (%); (d) fault detection rate (%).

divided into five categories: lidars (solutions 1, 2), integrated radar-video machine (solutions 3, 4), camera+mmWave radar (solutions 5, 6), camera+lidar (solutions 7, 8, and 9), and camera+lidar+mmWave radar (solutions 10, 11, 12, and 13), which is shown in Table 2.
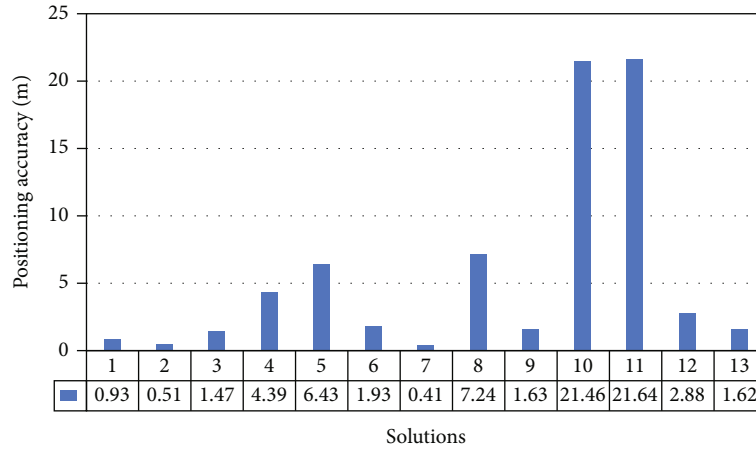
Corresponding to the evaluation indexes introduced in Subsection 3.2, we present the test results and analysis of these indexes below by categories.

We carry out two rounds of tests on sensing range, response time, and sensing delay in four different lanes. Particularly, the sensing range is strongly related to positioning accuracy. Therefore, we test the sensing range under the condition of satisfying different positioning accuracies (≤50 cm, ≤100 cm, ≤150 cm, and ≤200 cm). For sensing frequency and the maximum number of detected targets, we test them both during the peak period and the off-peak period. The average value of the collected test data is used as the test results of system essential capability indexes, sensing range, response time, sensing delay, and sensing frequency, and the maximum number of detected targets for different solutions is shown in Figures 7(a)–7(e), respectively.
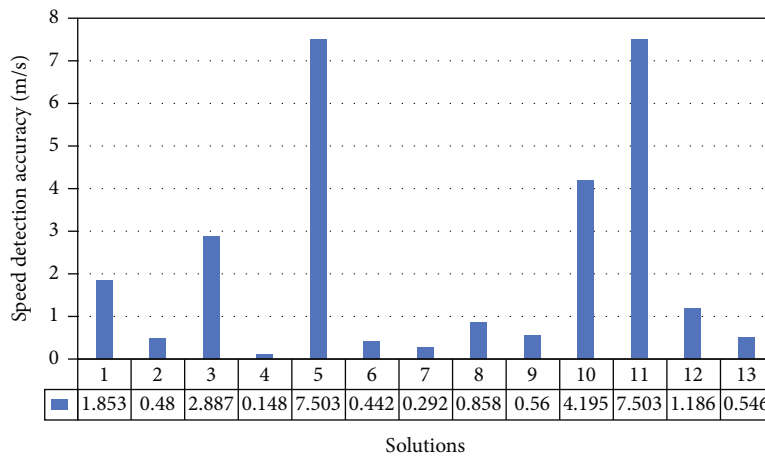
From the result in Figure 7(a), we find that when the given positioning accuracy requirement is within 200 centimetres, mmWave radar-based systems have a farther sensing range of about 200 metres, and lidar-based systems' sensing range is about 100 metres. But when the given positioning accuracy requirement is within 50 centimetres, only lidar-based systems (solutions 2, 7) can generate a sensing range of about 50 metres. With the expansion of the positioning accuracy range, the sensing range also continues to expand. Figure 7(b) shows that it is about 70% of the solutions that could achieve a response time of around 200 milliseconds, and more than 50% of solutions could achieve response delays of nearly 150 milliseconds. Since the test method of the response time is strongly related to the posi-

tioning accuracy, the value can be positive or negative. In Figure 7(c), the sensing delays of all other solutions with valid test data except for solution 6 are all within 100 milliseconds. By calculating the data in Figure 7(d), there are about 10.53 and 10.33 average sensing message frames generated by the roadside sensing and processing system per unit of time during the off-peak and the peak periods, respectively. Among the test results, solution 6 had the most prominent one. According to the results in Figure 7(e), approximately 38.67 and 45.83 targets can be detected as the maximum number by the roadside sensing and processing system in the given period on average during the off-peak and the peak periods, respectively. The outstanding solution 9 even detects the number of targets as nearly three times the average during the peak period.
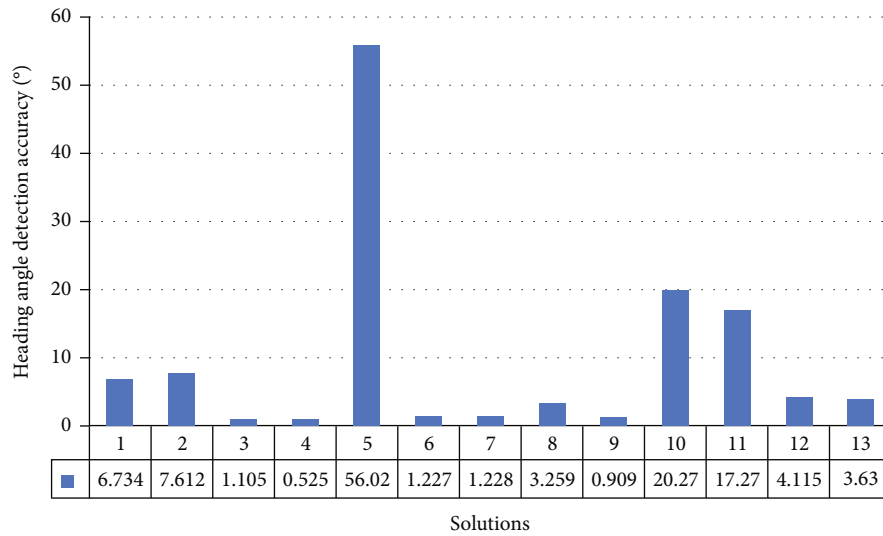
Three types of traffic participants: vehicles, nonmotor vehicles, and pedestrians are considered the different targets in the target recognition capability tests. The test results of recognition accuracy, classification accuracy, missing detection rate, and fault detection rate are shown in Figures 8(a)–8(d), respectively. In accordance with the data shown in the figures, we can conclude that for the target as a vehicle, the average recognition accuracy of all the solutions is slightly over 60%. It is about 85% of the solutions' classification accuracy which is over 90%, which is more accurate than the situation of the target as nonmotor vehicle and pedestrian. The recognition and classification capabilities of each solution in terms of nonmotor vehicles and pedestrians are quite different, and the overall performance needs to be further improved. It is worth noting that the false detection rate of all traffic participants is basically controlled by 10%. But the missing detection rate of all traffic participants is fairly high. The reasons for this result relate to weather, the number of traffic participants, point cloud occluded by large vehicles, and so on, which is not easy to control in a real scenario.
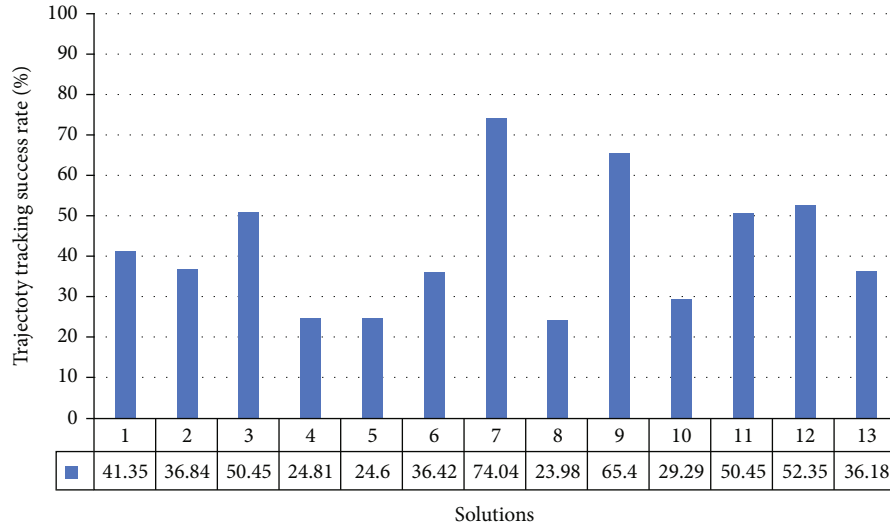
	1	2	3	4	5	6	7	8	9	10	11	12	13
■	0.93	0.51	1.47	4.39	6.43	1.93	0.41	7.24	1.63	21.46	21.64	2.88	1.62

Solutions

(a)

	1	2	3	4	5	6	7	8	9	10	11	12	13
■	1.853	0.48	2.887	0.148	7.503	0.442	0.292	0.858	0.56	4.195	7.503	1.186	0.546

Solutions

(b)

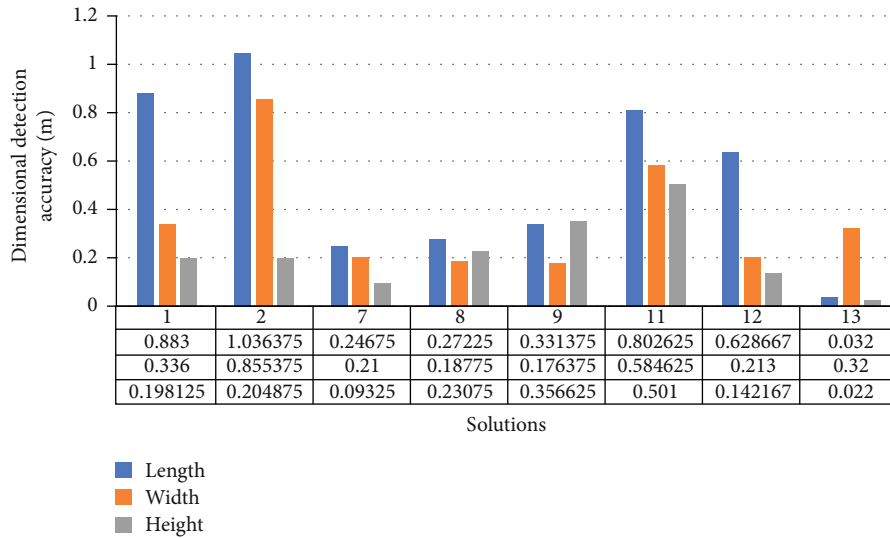|	1	2	3	4	5	6	7	8	9	10	11	12	13
■	6.734	7.612	1.105	0.525	56.02	1.227	1.228	3.259	0.909	20.27	17.27	4.115	3.63

Solutions

(c)

Figure 9: Continued.

(d)

(e)

FIGURE 9: Target positioning capability indexes for different solutions: (a) positioning accuracy (m); (b) speed detection accuracy (m/s); (c) heading angle detection accuracy (°); (d) trajectory tracking success rate (%); (e) dimensional detection accuracy (m).

The average value of two-round tests for positioning accuracy, speed detection accuracy, heading angle detection accuracy, and dimensional detection accuracy in four different lanes is given in Figures 9(a)–9(e), respectively. Figure 9(a) illustrates that more than 50% of the solutions can position within the accuracy of 2 metres. The positioning accuracy of some solutions (solutions 1, 2, and 7) can be within 1 metre. In Figure 9(b), more than 60% of the solutions have relatively high accuracy of speed detection; the deviation between the detected vehicle speed and its actual speed can be limited to within 1 m/s. Figure 9(c) provides some interesting data regarding the heading angle detection accuracy of some solutions (solutions 5, 10, and 11). These mmWave radar-based solutions have a short-term outage of detecting the vehicle's heading angle caused by the stationary state of the vehicle (such as stopping at a traffic light) in continuous tracking because of the zero-

Doppler filtering detection principle. As shown in Figure 9(d), the range of the trajectory tracking success rate distribution is approximately between 25% and 65%. The reason for the unsatisfied results is caused by the occlusion of the vehicles, which leads to the interruption of trajectory tracking. Figure 9(e) only demonstrates the dimensional detection accuracy results of the valid solutions. We can see that the detection accuracy of vehicle height and width is relatively higher than vehicle length. In addition, both integrated radar-video machine solutions (solutions 3, 4) and camera+mmWave radar solutions (solutions 5, 6) fail to detect the dimensional information of the vehicles. The solutions based on camera+lidar (solutions 7, 8, and 9) provide the best performance of dimensional detection accuracy of all the solutions.

The traffic flow detection capability index, the relative error of detected traffic flow (%) in 4 lanes for different
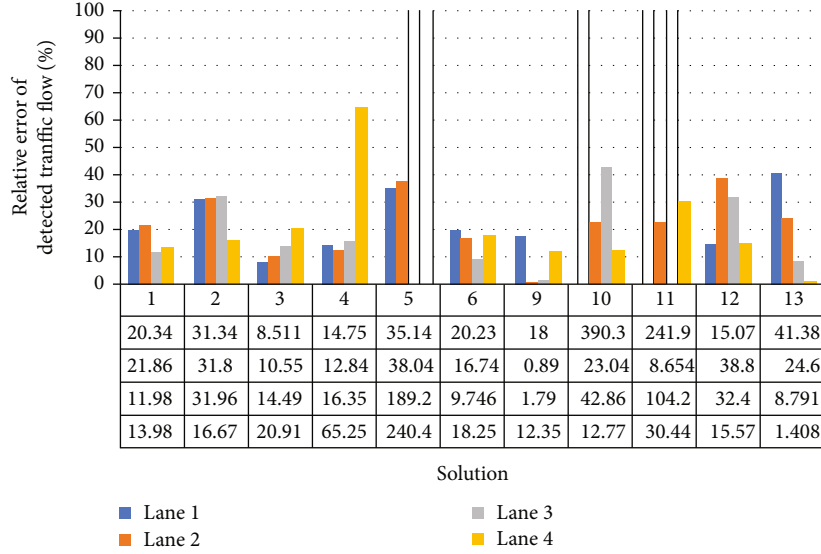
FIGURE 10: Traffic flow detection capability for different solutions.

solutions, is shown in Figure 10. Solutions based on camera +lidar (solutions 7, 8) fail to give valid data and are not depicted in the figure. The test results reveal that the traffic flow detection performance of the straight lanes (Lanes 2 and 3) and the left turn/right turn lane (Lanes 1 and 4) is generally quite different. All solutions perform the most robust traffic flow detection capability on the straight lane (Lane 2) with an average of about 20% relative error of detected traffic flow. In addition, five test results are too large to be shown in the figure, which is due to oversegmentation during the detection.

The test results show that various solutions give different performances for different detection indexes. Therefore, we only give this real test data to readers as a reference but cannot draw a conclusion about which type of solution has the best performance comprehensively. Moreover, based on the analysis of the actual test process and results, we summarize the following main factors that affect the performance of the roadside sensing and processing solutions in IoV: (1) problems in calibration, including static calibration, camera calibration, external parameter calibration of lidar, and timing synchronization; (2) the bottleneck of the equipment, like the bottleneck of hardware in different multimodal sensor combination schemes and the restricted computing capacity of roadside MEC; (3) advantages and disadvantages of different multimodal sensing information fusion algorithm designs, such as early-fusion, feature-level fusion, target-level fusion, and algorithms. At present, the perception equipment of traditional intelligent transportation is only suitable for some safety warning scenarios for vehicle-road collaborative applications. These kinds of scenarios are mainly based on the perception capability of the vehicle and are less dependent on the perception capability of the roadside equipment. However, for other vehicle-road collaborative applications, especially for complex functional scenarios like collaborative traffic and high-level autonomous driving, there is still room for improvement in technology and product maturity. From the perspective of the index's decisiveness, common basic indexes such as system essential capability, vehicular kinematics, and target classification are more decisive, which directly affects the performance of functional indexes. In the next step, considering the different categories of application scenarios, we will develop a graded classification standard for roadside perception systems based on our proposed evaluation index system to promote the improvement of roadside perception system-related products' performance and technology maturity.

5. Conclusions

We propose a system-level test method to evaluate roadside multimodal sensing and processing solutions in real IoV scenarios. To this end, we summarize the evaluation index system corresponding to the sensing data processing application functional entities of IoV applications and build the roadside sensing and processing test system to collect the ground-truth in real scenarios. The test results show that our proposed test system can effectively evaluate the performance of the IoV sensing and processing system in real scenarios. In the future, we will continue to research the test method to improve the accuracy of the ground-truth under the multi-target-oriented test and improve the integration level of test equipment and flexible deployment capabilities. In addition, we will build a test prototype that simulates the real scenario based on the absolute ground-truth, which provides a low-cost and high-efficiency laboratory test method for academic research and industrial development.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

References

[1] W. Duan, J. Gu, M. Wen, G. Zhang, Y. Ji, and S. Mumtaz, "Emerging technologies for 5G-IoV networks: applications, trends and opportunities," *IEEE Network*, vol. 34, no. 5, pp. 283–289, 2020.

[2] D. F. Llorca, I. G. Daza, N. H. Parra, and I. P. Alonso, "Sensors and sensing for intelligent vehicles," *Sensors*, vol. 20, no. 18, pp. 5115–5138, 2020.

[3] K. Yabuuchi, M. Hirano, T. Senoo, N. Kishi, and M. Ishikawa, "Real-time traffic light detection with frequency patterns using a high-speed camera," *Sensors*, vol. 20, no. 14, pp. 4035–4053, 2020.

[4] J. G. Wang and L. B. Zhou, "Traffic light recognition with high dynamic range imaging and deep learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 4, pp. 1341–1352, 2018.

[5] M. G. Cho, "A study on the obstacle recognition for autonomous driving RC car using LiDAR and thermal infrared camera," in *2019 Eleventh International Conference on Ubiquitous and Future Networks (ICUFN)*, Zagreb, Croatia, 2019.

[6] T. Zhou, M. Yang, K. Jiang, H. Wong, and D. Yang, "MMW radar-based technologies in autonomous driving: a review," *Sensors*, vol. 20, no. 24, pp. 7283–7304, 2020.

[7] B. S. Jahromi, T. Tulabandhula, and S. Cetin, "Real-time hybrid multi-sensor fusion framework for perception in autonomous vehicles," *Sensors*, vol. 19, no. 20, pp. 4357–4380, 2019.

[8] J. Dickmann, J. Klappstein, M. Hahn et al., "Automotive radar the key technology for autonomous driving: from detection and ranging to environmental understanding," in *2016 IEEE RADAR CONFERENCE (RADARCONF)*, pp. 782–787, Philadelphia, PA, USA, 2016.

[9] J. Dickmann, N. Appenrodt, H. L. Bloecher et al., "Radar contribution to highly automated driving," in *2014 11th European Radar Conference*, pp. 412–415, Rome, Italy, 2014.

[10] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 6–23, 1997.

[11] Z. Wei, F. Zhang, S. Chang, Y. Liu, H. Wu, and Z. Feng, "MmWave radar and vision fusion for object detection in autonomous driving: a review," *Sensors*, vol. 22, no. 7, pp. 2542–2565, 2022.

[12] D. Jeon and H. Choi, "Multi-sensor fusion for vehicle localization in real environment," in *2015 15th International Conference on Control, Automation and Systems (ICCAS)*, pp. 411–415, Busan, Korea, 2015.

[13] Z. Wang, Y. Wu, and Q. Niu, "Multi-sensor fusion in automated driving: a survey," *IEEE Access*, vol. 8, pp. 2847–2868, 2020.

[14] N. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, "Mobile edge computing: a survey," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 450–465, 2018.

[15] A. Waheed, M. A. Shah, S. M. Mohsin et al., "A comprehensive review of computing paradigms, enabling computation offloading and task execution in vehicular networks," *IEEE Access*, vol. 10, pp. 3580–3600, 2022.

[16] P. Kolar, P. Benavidez, and J. Mo, "Survey of datafusion techniques for laser and vision based sensor integration for autonomous navigation," *Sensors*, vol. 20, no. 8, pp. 2180–2229, 2020.

[17] Y. Cui, R. Chen, W. Chu et al., "Deep learning for image and point cloud fusion in autonomous driving: a review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 2, pp. 722–739, 2022.

[18] V. D. Silva, J. Roche, and A. Kondoz, *Fusion of LiDAR and Camera Sensor Data for Environment Sensing in Driverless Vehicles*, IEEE, 2017.

[19] R. O. Chavez-Garcia and O. Aycard, "Multiple sensor fusion and classification for moving object detection and tracking," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 2, pp. 525–534, 2016.

[20] L. Guan, Y. Chen, G. Wang, and X. Lei, "Real-time vehicle detection framework based on the fusion of LiDAR and camera," *Electronics*, vol. 9, no. 3, pp. 451–469, 2020.

[21] S. Chang, Y. Zhang, F. Zhang et al., "Spatial attention fusion for obstacle detection using MmWave radar and vision sensor," *Sensors*, vol. 20, no. 4, pp. 956–977, 2020.

[22] F. Nobis, M. Geisslinger, M. Weber, J. Betz, and M. Lienkamp, "A deep learning-based radar and camera sensor fusion architecture for object detection," in *2019 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, Bonn, Germany, 2019.

[23] S. Chadwick, W. Maddern, and P. Newman, "Distant vehicle detection using radar and vision," in *Proceedings-IEEE International Conference on Robotics and Automation*, pp. 8311–8317, Montreal, QC, Canada, 2019.

[24] P. Sermanet, K. Kavukcuoglu, S. Chintala, and Y. LeCun, "Pedestrian detection with unsupervised multi-stage feature learning," in *2013 IEEE conference on computer vision and pattern recognition (CVPR)*, pp. 3626–3633, Portland, Oregon, USA, 2013.

[25] T. Kato, Y. Ninomiya, and I. Masaki, "An obstacle detection method by fusion of radar and motion stereo," *IEEE Transactions on Intelligent Transportation Systems*, vol. 3, no. 3, pp. 182–188, 2002.

[26] C. Lin, D. Tian, X. Duan, J. Zhou, D. Zhao, and D. Cao, "3D-DFM: anchor-free multimodal 3-D object detection with dynamic fusion module for autonomous driving," *IEEE Transactions on Neural Networks and Learning Systems*, vol. PP, pp. 1–11, 2022.

[27] X. Yue, B. Wu, S. A. Seshia, K. Keutzer, and A. L. Sangiovanni-Vincentelli, "A lidar point cloud generator: from a virtual world to autonomous driving," in *Proceedings of 2018 ACM on International Conference on Multimedia Retrieval*, Yokohama, Japan, 2018.

[28] G. Wang, W. Deng, S. Zhang, J. Wang, and S. Yang, *A comprehensive testing and evaluation approach for autonomous vehicles*, SAE Technical Papers, 2018.

[29] P. Minnerup and A. Knoll, *Testing autonomous driving systems against sensor and actuator error combinations*, IEEE Intelligent Vehicles Symposium, 2014.

[30] S. Yao, J. Zhang, Z. Hu, Y. Wang, and X. Zhou, "Autonomous-driving vehicle test technology based on virtual reality," *The Journal of Engineering*, vol. 2018, no. 16, pp. 1768–1771, 2018.

[31] K. Kufieta and M. Ditze, *A virtual environment for the development and validation of highly automated driving systems*, The 17th Internationales Stuttgarter Symposium, 2017.

[32] C. Zhang, Y. Liu, D. Zhao, and Y. Su, *RoadView: a traffic scene simulator for autonomous vehicle simulation testing*, Proceedings of the 17th IEEE International Conference on Intelligent Transportation Systems, 2014.

[33] W. Li, C. Pan, R. Zhang et al., "AADS: augmented autonomous driving simulation using data-driven algorithms," *Science robotics*, vol. 4, no. 28, p. eaaw0863, 2019.

[34] F. Rosique, P. J. Navarro, C. Fernández, and A. Padilla, "A systematic review of perception system and simulators for autonomous vehicles research," *Sensors*, vol. 19, no. 3, pp. 648–677, 2019.

[35] M. Braun, S. Krebs, F. Flohr, and D. M. Gavrila, "EuroCity persons: a novel benchmark for person detection in traffic scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 8, pp. 1844–1861, 2019.

[36] M. Baek, D. Jeong, D. Choi, and S. Lee, "Vehicle trajectory prediction and collision warning via fusion of multisensors and wireless vehicular communications," *Sensors*, vol. 20, no. 1, pp. 288–314, 2020.