2022 9th International Conference on Dependable Systems and Their Applications (DSA)

A Survey of Autonomous Driving Scenarios and Scenario Databases

Hongping Ren[1], Hui Gao[2], He Chen[1], and Guangzhen Liu[1,*]
[1]Institute of Software Chinese Academy of Sciences, Beijing, China
[2]Nanjing Institute of Software Technology, Nanjing, China
hongping@iscas.ac.cn, gaohui@nj.iscas.ac.cn, chenhe@iscas.ac.cn, guangzhen@iscas.ac.cn
*corresponding author

2022 9th International Conference on Dependable Systems and Their Applications (DSA) | 978-1-6654-8877-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/DSA56465.2022.00107

Abstract—**With the development of autonomous driving technology, traditional road testing methods can no longer meet the needs of autonomous driving testing. These methods lack sufficiency, comprehensiveness and efficiency. Using the autonomous driving scenario databases for testing can greatly shorten the test time and cost, and can improve the safety and reliability of the test. By systematically sorting out a large number of related publications, this paper summarizes the research progress of autonomous driving scenarios and scenario databases. The article firstly compares and analyzes the different definitions of autonomous driving scenarios, clarifies the connotation of the scenarios, summarizes the types of elements of the scenarios, and introduces the scenario layered model; secondly, we outline the description standards of scenario. We mainly summarize the two scenario data formats, OpenDRIVE and OpenSCENARIO, which are commonly used in the world. Thirdly, the scenario data collection and research work carried out at home and abroad is reviewed from the perspective of scenario data sources, and different datasets are compared; In addition, the definition of the scenario database, the construction process of the scenario database and the typical scenario databases are summarized; Finally, the problems and future development trends of autonomous driving scenarios and scenario databases are discussed and prospected.**

Keywords-autonomous driving scenarios; scenario dataset; scenario database; scenario description.

I. INTRODUCTION

At present, scenario-based autonomous driving testing has become the main route of autonomous driving testing technology. This testing method has flexible configuration, high testing efficiency, strong testing repeatability, safe testing process, and low testing cost. It can realize automatic testing and save a lot of manpower and material resources.

Autonomous driving scenarios are an important basis for the development, verification, validation, and evaluation of autonomous driving technologies. The most important criterion for evaluating autonomous driving technology is to test whether it can handle enough scenarios. Therefore, research on autonomous driving scenarios is crucial to improving the R&D efficiency of autonomous vehicles, improving technical standards, and promoting industrial innovation and development. Nowadays, scenario-based autonomous driving testing has become an indispensable and important part in the field of autonomous vehicles. Countries around the world have paid considerable attention to autonomous driving scenario technology. Many scientific research institutions and researchers have carried out extensive research on it.

Based on the analysis of domestic and foreign studies and related research literature, this paper summarizes the current autonomous driving scenarios and scenario databases from the following aspects:

1) Autonomous driving scenario: As the main body of the autonomous driving simulation scenario database, the scenario is of great significance. This paper sorts out the scenario definition, scenario elements, and scenario layered model through a series of analyses.
2) Scenario description standard: Scenario automated testing and scenario reuse (scenario data exchange) are the general trend of automated driving testing, which all require the support of standard scenario files. This paper introduces the ASAM OpenX standard, and summarizes the current international common scenario data formats OpenDRIVE and OpenSCENARIO.
3) Scenario dataset: The scenario datasets are classified from the sources of scenario data, and common public datasets are compared and summarized.
4) Scenario database: The scenario database is the basic database in the process of autonomous vehicle development and testing. This section sorts out the scenario database from the aspects of scenario database definition, construction of database, and typical scenario databases. The common public databases are compared and summarized.
5) Summary and prospect: The current research results and development of autonomous driving scenarios and scenario databases are discussed and summarized, and the future development trends are also analyzed.

II. SCENARIO IN AUTONOMOUS DRIVING

A. *Scenario Definition*

Having a common definition and a structured framework for "scenario" is the premise of scenario research. In the early days, many researchers described the environmental information around the vehicle as a scenario, which was mainly used in the analysis and research of human driving and traffic accidents [1-3]. As scenarios play an increasingly important role in the field of autonomous driving [4],

2767-6684/22/$31.00 ©2022 IEEE
DOI 10.1109/DSA56465.2022.00107

scholars have proposed different definitions based on various perspectives.

Geyer and Baltzer [5] suggest that the term scenario can be understood as some kind of storyline, which includes the expected actions of the driver, but does not specify every action in detail. The exact way it is realized is strongly influenced by a lot of factors, such as the freewill of the driver.

Go and Carroll [6] argue that the elements of a scenario are same, although the usage and detailed definitions of scenarios are different in different domains. They see scenario as a description that contains actors, background information on the actors and assumptions about their environment, actors' goals or objectives, and sequences of actions and events. Some applications may omit one of the elements or they may simply or implicitly express it.

Ulbrich and Menzel [7] propose a more general definition of scenario. According to them, "A scenario describes the temporal development between several scenes in a sequence of scenes. Every scenario starts with an initial scene. Actions & events as well as goals & values may be specified to characterize this temporal development in a scenario. Other than a scene, a scenario spans a certain amount of time."

According to Elrofai [8], scenario is a typical manoeuvre on the road with the complete set of relevant conditions and trajectories of other traffic participants that have an interaction with the ego vehicle over a relevant time period, typically in the order of seconds. A ride on the road can in this way be described by a continuous sequence of scenarios, where scenarios might overlap in time.

Similarly, Zhu Bing et al. [9] regard the scenario as an organic combination of driving situations and driving scenarios of autonomous vehicle. They suggested scenario is an overall dynamic description of the elements of the autonomous vehicle and its driving environment over a period of time. These elements are determined by the function of the autonomous vehicle that is expected to be tested, where the driving situation is the driving environment, and the driving situation is the dynamic elements and the relationships between them.

The term 'scenario' has not been defined uniformly, which makes it difficult to achieve a consistent understanding regarding the role of scenarios in the development process. Based on the previous publication, however, the core elements of scenario in autonomous driving can be understood as: scenario describes the external roads, traffic facilities, weather and traffic participants, as well as the driving tasks and states of the vehicle itself. It is an organic combination and comprehensive reflection of the driving environment, traffic participants and driving behavior within a certain time and space, and it will change dynamically.

B. Scenario Elements and Layered Model

Real-world scenarios are complex and endless. Decomposing the scenario and extracting the element types contained in the scenario is the basic method for dimensionality reduction and abstraction of the real scenario. When conducting scenario research, it is necessary to clarify the elements covered by the scenario and integrate the scenario elements. By deconstructing the scenario to a certain extent, it is helpful to grasp the essence and connotation of the scenario, and better conduct related research.

Currently, researchers still have different views on the types and specific contents of scenario elements. Geyer et al. [5] propose that scenario elements should include predetermined driving tasks, static scenario elements and dynamic scenario elements. Groh et al. [10] divide scenario elements into three categories: static elements, which refer to objects that remain stationary for a long enough time that the action is almost imperceptible, such as lane lines, static traffic signs, etc.; dynamic elements, which refer to certain objects within a certain period of time. Objects that act, such as pedestrians, cyclists, etc.; environmental elements, refer to some microscopic elements in the environment, such as weather, lighting, wind speed, etc. Sauerbier et al. [11] propose that scenario elements should include test vehicles, traffic environment elements, driving task information and specific driving behaviors. Zhu Bing et al. [9] propose that the scenario elements can be divided into two parts: the test vehicle's own elements and the external traffic environment elements. Among them, the basic information of the test vehicle includes three categories of basic elements of the test vehicle, target information and driving behavior; the traffic environment elements include four categories of weather and light, static road information, dynamic road information and traffic participant information.

At present, scholars in the field of autonomous driving mainly classify scenario elements according to the six-layer scenario model proposed by the German PEGASUS project [12-13]. The German PEGASUS project is jointly initiated by related companies and research institutions in the German automotive industry, with the purpose of formulating a series of test standards related to autonomous vehicles.

Initially, the German Pegasus project proposed a 5-layer layered model of the scenario from the perspective of deconstruction and reconstruction of the test scenario [14]: the first layer is the road layer, which describes the geometry, size, the topology relationship between the various geometric structures of the road, road surface quality status and pavement boundaries information; the second layer is the traffic infrastructure layer, which describes the various types of fixed motions attached to the road layer that can constrain the movement of autonomous vehicles and other dynamic traffic participants in the scenario according to certain traffic rules; the third layer is the temporary manipulation layer of the first layer and the second layer, which describes the road sections and related traffic facilities that are temporarily built in the scenario and will not be moved in a short period of time; the fourth layer is the objects layer, which describes all kinds of dynamic, static and movable traffic participants in the scenario, as well as the interaction and movement behavior between traffic participants; the fifth layer is the environment layer, which describes the environmental conditions such as weather and light.

The above five-layer model realizes the clear layered division of scenario elements, which is conducive to standardized and general scenario analysis and automatic generation. This makes it cited by a large number of scholars and institutions. The shortage of this model is that it ignores the scenario construction requirements for ICV positioning and communication capability testing. In response to this problem, the PEGASUS project proposed in a subsequent report [15] that the data communication layer was added as the sixth layer on the basis of the original five-layer model, and the elements included mainly V2X information, digital map information, etc. A six-layer model is used to describe the scenario elements, as shown in the following TABLE I:

TABLE I. THE SIX-LAYER MODEL

Layer	Name	Scenario Description
L1	Road layer	Road topology, pavement quality, pavement boundaries, etc.
L2	Traffic infrastructure layer	Structural boundaries, traffic signs, signal lights, etc.
L3	Temporary manipulation layer of L1 and L2	Temporary facilities on roads, such as temporary road closures, road construction sites, etc.
L4	Objects layer	Traffic participants, and the status, behavior, etc. of traffic participants
L5	Environment layer	Weather, electromagnetic strength, temperature, etc.
L6	Data communication layer	V2X information, digital map information, etc.

Scenario elements are defined as the basic units that constitute a scenario. From the perspective of autonomous driving test requirements, scenario elements can be divided into three types of elements: environmental elements, traffic participant elements, and self-vehicle tasks. From the basic attributes of scenario elements, they can be divided into static elements and dynamic elements. From the topological relationship of scenario elements, it can be divided into road elements, traffic participant elements, meteorological elements, network elements, etc.

In short, a scenario is a combination of various scenario elements in a certain time and space, and there is a strong coupling relationship between them. Changes in one element may affect all other elements. The study of scenario elements plays an indispensable role in the construction of autonomous driving scenarios.

III. SCENARIO DESCRIPTION STANDARD

A. ASAM OpenX Standards

At present, many autonomous driving tests are based on scenarios. As the trend of scenario-based autonomous driving becomes more and more obvious, scenario description is crucial for testing and verifying the safety of autonomous vehicles. Moreover, scenario automation testing and scenario reuse are the general trend of automatic driving testing, which all require the support of standard scenario files. The design of the scenario database at this stage is largely dependent on the software involved, and if they are not transmitted in a standardized way, the scenario database will be very expensive to migrate between different test tools. Moreover, in the actual development process, various data formats and interfaces used by OEMs, suppliers and simulation tool vendors make it difficult to unify standards. Therefore, it is very important to determine a unified standard for scenario description.

In order to solve the problems of scenario-based simulation testing of intelligent networked vehicles, inconsistent interfaces in actual development and applications, and various data formats, the German Association for Standards of Automation and Measurement Systems (ASAM) has launched the OpenX series of standards in the field of autonomous driving. It has gained worldwide attention as a long-term standard to promote future industrial development.

The ASAM OpenX series of standards includes seven aspects as a whole, as shown in the following Figure 1:

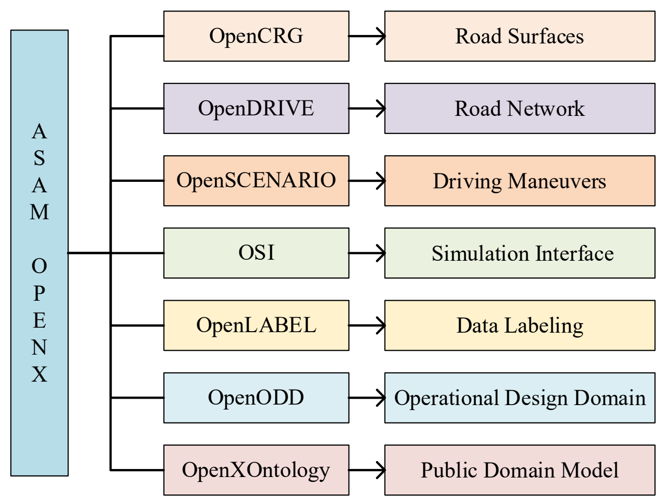

Figure 1. ASAM OpenX Standards

In the whole process of the automatic driving test, OpenDRIVE corresponds to the static map scenario and is responsible for describing map information, including high-precision map information; OpenSCENARIO corresponds to the dynamic behavior scenario, that is, the test scenario describing the automatic driving; OpenDRIVE and OpenSCENARIO are for the simulation scenario. Different data formats are unified; OpenCRG corresponds to the road surface, focusing on vehicle dynamics and vehicle feedback on road information, realizing the interaction between road physical information and static road scenarios; OpenLABEL studies scenario labels and sensor raw data, and the raw data Provides a unified calibration method with the scenario; OSI corresponds to the simulation interface, such as the sensor simulation interface or the simulation interface of various information, connects the automatic driving function and simulation tools, and integrates a variety of sensors; OpenODD defines the design operation domain;

OpenXOntology connects the OpenX standard serially through a public domain model represented by an ontology.

OpenDRIVE and OpenSCENARIO are international scenario data formats, and they are the most commonly used open-source file formats for scenario descriptions. OpenSCENARIO only defines dynamic content in the scenario, static content is not included in OpenScenario, but refers to other standards such as OpenDRIVE and OpenCRG. For simulation tools that use scenario data in other formats, they should also be equipped with corresponding scenario file parsing interfaces to realize conversion with standard data formats.

In ASAM OpenX, OpenSCENARIO describes dynamic content, OpenDRIVE describes the road network for static content, and OpenCRG describes the road network for static content. The three standards complement each other and cover both static and dynamic aspects of autonomous vehicle simulation applications.

B. OpenDRIVE

The OpenDRIVE format is an open source road network logical description format standard, which provides a switching format specification to describe the static road network for driving simulation applications. The main task of OpenDRIVE is road description, including objects along the road. OpenDRIVE specification covers the description of how to model roads, lanes, intersections, but does not cover dynamic content. The road network is modelled along the reference line, which is the core piece of every road. The reference line is in the center of the road; the lanes are attached to this reference line. The signs next to the road are placed in the s/t-coordinate system [16]. The OpenDRIVE format uses extensible markup language (XML) syntax with the file extension xodr. These data describe the geometry of roads, lanes and objects. The road networks can either be synthetic or based on real data.

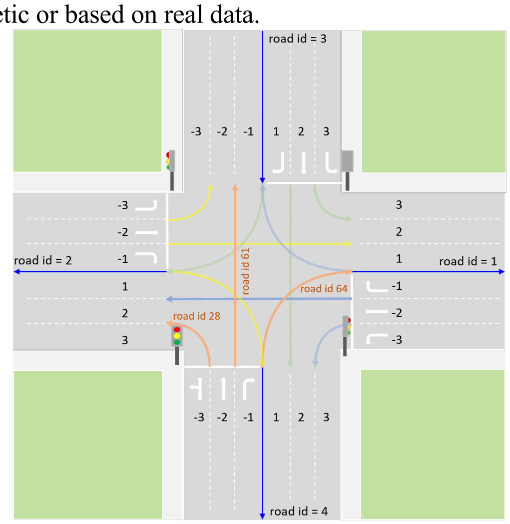

Figure 2. Example of a junction [16]

In OpenDRIVE several roads form a road network and can be connected. OpenDRIVE can be seen as a construction kit of different road sections. The overall road network is composed of individual sections interconnected with each other. Besides, the lanes between roads are linked. The Figure 2 shows the connecting roads inside the junction area that connect the incoming and outgoing roads.

The OpenDRIVE description format contains all static objects of a road network that allow realistic simulation of vehicles driving on roads.

C. OpenSCENARIO

OpenSCENARIO defines a file format for the description of the dynamic content of driving and traffic simulators. OpenSCENARIO describes complex, synchronized maneuvers that involve multiple entities like vehicles, pedestrians and other traffic participants. Other content, such as the description of the ego vehicle, driver appearance, pedestrians, traffic and environment conditions, is included in the standard as well [17]. Scenario descriptions in ASAM OpenSCENARIO are organized in a hierarchical structure and serialized in an XML file format with the file extension xosc.

OpenSCENARIO V1.x consists of seven main parts, as shown in the following Figure 3:

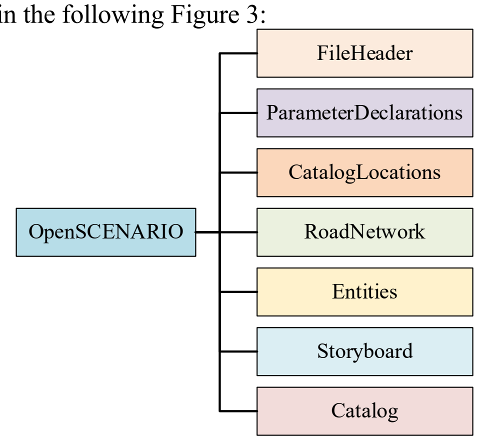

Figure 3. OpenSCENARIO V1.x

The FileHeader describes the version, author, date, related description and so on. The related parameters are declared in ParameterDeclarations. There are eight different kinds of elements that may be outsourced to a Catalog. All kinds of objects may be defined within a Catalog. The standard describes vehicle maneuvers in a storyboard, which is subdivided in stories, acts and sequences. The maneuver could be a lane change, overtaking another car or driving in a traffic jam. Stories consist of acts, which are triggered when a specific condition is met. Events and actions describe the detailed driving behavior of the vehicle. The standard uses the notion of sequences to define the maneuvers of multiple vehicles and response them. A story can describe the driving maneuvers of one single vehicle or specify the dynamic behavior of several entities.

OpenSCENARIO V2.0 is designed to be a complete set of OpenSCENARIO V1.x functions, and needs to support various complex scenarios required in the development and testing of autonomous vehicles. OpenSCENARIO V2.0 can create maneuver descriptions and tests to define scenarios at a higher level of abstraction. It can provide alternative expression methods to the current XML format of

OpenSCENARIO V1.x via a domain-specific language (DSL).

In addition, in OpenSCENARIO, the complete scenario description can realize test automation through parameterization without creating a large number of scenario files.

IV. Scenario Dataset

The data sources of autonomous driving scenarios can generally be divided into three categories: real data, simulated data and empirical data [9], as shown in the following Figure 4:

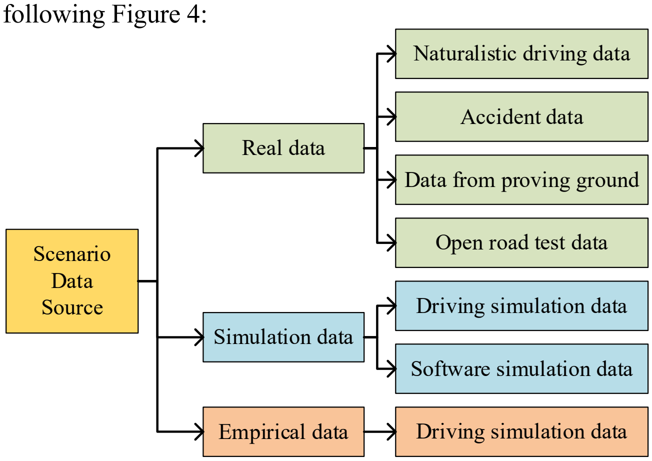

Figure 4. Scenario Data Source

1) Real data

a) *Naturalistic driving data*

Naturalistic driving data contains very rich scenario information, which can truly reflect the traffic operation state when traffic conditions occur. Collecting naturalistic driving data can accumulate a large amount of test data for autonomous vehicles, which is an effective means to improve autonomous driving technology. Currently commonly used public naturalistic driving datasets include HighD, NGSIM, inD, RoundD, ACFR Five Roundabouts, INTERACTION, Mirror-Traffic, Trajectory Dataset (APOLLOSCAPE) etc. This paper sorts out and compares these datasets in terms of dataset size, scenario type, road type, traffic participants, and scenario complexity, as shown in the following TABLE II.

b) *Accident data*

Accident data is a typical safety test condition, which can test the stress ability of the automatic driving system.

Accident data mainly comes from accident databases of various countries and organizations. Although accident data is relatively scarce, in order to analyze the causes of accidents and improve road traffic safety, many countries and organizations have established traffic accident databases, such as China's CIDAS database, Germany's GIDAS database, Japan's ITARDA database, and the United States National Highway Traffic Security Administration's (NHTSA's) GES database, the European Union's ASSESS database, etc. Based on the rich traffic accident database, researchers can extract representative vehicle accidents and typical accident scenario elements to build a traffic accident-based autonomous driving scenario database.

TABLE II. Comparison of Public Naturalistic Driving Datasets

Dataset	From	Number of Vehicles	Source	Road Type	Traffic Participants	Scenario Complexity
HighD [18]	RWTH Aachen University	110000	highway	straightaway	car, truck	Medium
NGSIM [19]	Federal Highway Administration	9206	highway, urban road	straightaway, ramp	car, truck, motorcycle	Medium
inD [20]	RWTH Aachen University	13500	urban road	crossroads without signal lamp	car, truck, pedestrians, bicyclists	High
RounD [21]	RWTH Aachen University	13746	urban road	roundabouts	vehicles (car, truck, trailer, van, bus), pedestrians, bicyclists, motorcycles	High
ACFR Five Roundabouts [22-23]	The University of Sydney	23000	urban road	roundabouts	bike, car, heavy vehicle, pedestrian	High
INTERACTION [24]	University of California, Berkeley	40054	highway, urban road	straightaway, crossroads, roundabouts, ramp	bike, car, pedestrian	High
Mirror-Traffic[25]	Tsinghua University Automotive Research Institute & JIVIC	1827	highway, urban road	straightaway, road curve, ramp, crossroad	vehicles, pedestrian	High
Trajectory Dataset (APOLLOSCAPE) [26]	Baidu Research	-	urban road	straightaway, crossroad	vehicles, pedestrian, motorcyclist, bicyclist	High

c) *Data from proving ground*

Many countries require that automated vehicles must be tested in self-driving testing center before they enter the consumer market, thus promoting the construction of purpose-built proving ground for smart vehicles. These data are the scenario elements obtained from autonomous driving tests of different intelligence levels in a purpose-built proving ground [27]. At present, many countries have established special testing centers for self-driving vehicles, such as the M-city at the University of Michigan, the self-driving vehicles proving ground in Texas, the Mira testing center in the UK, the J-Town in Japan, and the IDIADA proving ground in Spain, Sweden ASTAZero road safety test environment, K-city in South Korea, etc. In recent years, China has established a number of national-level demonstration zones with the support of the Ministry of Industry and Information Technology, such as the ICV pilot demonstration zone in Shanghai, the i-VISTA proving ground in Chongqing, and the national ICV application in Changchun Demonstration Area, Chang'an University Self-driving Testing Center, etc. With these proving grounds, it is possible to flexibly build rich traffic scenarios, conduct self-driving vehicle development testing and performance testing, and accumulate a large amount of self-driving test scenario data.

d) *Open road test data*

The open road test data refers to the scenario elements collected during self-driving vehicles are tested on the open roads. Besides traditional scenario data, through open road testing, the interaction data between autonomous vehicles and other road users can also be obtained, laying a data foundation for the industrialization of autonomous vehicles. Baidu's driverless car has completed an open road test of about 140,000 km in Beijing. Waymo driverless cars in the United States have accumulated more than 20 million km of self-driving data in more than 20 cities [28]. These data are helpful for discovering the defects of automatic driving functions, and then improving and upgrading the self-driving system.

2) Simulation data

a) *Driving simulation data*

The driving simulator data mainly comes from the driving simulator, which is the scenario element information obtained by using the driving simulator to test. Driving simulators can be driven by real people in simulation software to collect data. Compared with real road driving, driving simulators can not only ensure safety but also reflect authenticity [29], and can conduct driver-in-the-loop tests in dangerous and extreme situations. The University of Michigan in the United States, the University of Leeds in the United Kingdom, and Tsinghua University in the United Kingdom all use driving simulators to collect a large amount of human-vehicle-environment closed-loop system data.

b) *Software simulation data*

Simulation software The simulation data mainly comes from the simulation software. By artificially setting the driving task or driving route, the test vehicle is driven in the virtual simulation scenario, so as to generate the simulation data [30]. The simulation environment can be generated by importing a real scenario or modeling the vehicle driving environment. The vehicle driving environment modeling includes traffic flow modeling, surrounding static environment modeling, and meteorological environment modeling. Among them, most of the traffic flow modeling uses open source traffic flow simulation software, such as VISSIM, SUMO and so on.

3) Empirical data

a) *Driving simulation data*

Empirical data refers to the scenario elements information obtained by summarizing the experience and knowledge of previous tests [31]. The most typical data source is standard scenario data [32].

Among the above-mentioned scenario data sources, two typical ones are naturalistic driving data and accident data.

The content of naturalistic driving data is diverse, including typical driving conditions under different venues, weather, and traffic conditions [33]. These data are representative and can meet the testing needs of autonomous driving systems under various driving conditions. After the naturalistic driving data is collected, it needs to be cleaned, filtered, extracted, calibrated, deconstructed and reconstructed. By extracting and analyzing data, searching for the target of interest, and using cluster analysis to build a scenario database. Accident data is a typical safety test condition, which can test the stress ability of the automatic driving system [34]. Through data mining and other methods, traffic accident data can restore the accident process to obtain test scenarios, which are generally used as a beneficial supplement to natural driving data.

Scenarios generated based on real data can meet most of the requirements for self-driving testing but cannot fully cover all the scenarios. Simulation data and expert experience are generally used as supplements to real data.

V. SCENARIO DATABASE

A. *Scenario Database Definition*

The autonomous driving scenario database is a data collection of different scenarios, which is expressed in the form of a database. It is composed of a series of autonomous driving scenarios that meet certain test requirements, which facilitates unified and effective organization, management and application of scenarios. The scenario database is the basic database in the development and testing of autonomous vehicles. It is an important database for accelerating the development of automated driving safety tests. All functions of autonomous driving must be tested and run through tens of millions of scenarios. Building a scenario database is a necessary path to use scenario data to development, management and application. The construction of the scenario database can effectively drive the R&D and testing of autonomous driving, and in turn, the R&D and testing of autonomous driving can provide feedback for the scenario database and enrich the scenario database.

The construction of the scenario database is mainly realized through the virtual simulation environment and tool chain. The development of autonomous driving tests and the construction of scenario databases are mainly concentrated in some large companies at home and abroad, such as Waymo in the United States and Baidu in China. At present, some car companies and universities have also carried out a lot of research and development work in the construction of scenario databases.

B. Construction of Scenario Database

In the scenario-driven R&D test of autonomous driving, the scientific and orderly construction of test scenarios can effectively support the R&D and testing of autonomous driving. The whole construction process is from the data layer to the scenario layer, and then to the application layer:

The data layer is mainly to construct the required data by collecting simulation scenarios from real open or closed sites, simulation software, and experience databases, and perform a series of preprocessing operations on the collected data, and then use automated processing tools to The data is extracted and digitized, and finally the data format and parameters are unified and imported into the scenario layer.

The scenario layer mainly processes the data imported from the data layer, such as scenario understanding, feature extraction and mining, etc., and then imports the scenario database in a unified format. On this basis, it is necessary to cluster, generate, and optimize around the scenario to form different scenarios.

The application layer confirms the authenticity, representativeness and validity of the scenario by calling the scenario in the scenario database for virtual scenario verification and real vehicle scenario verification. The application layer will also feedback the test results to the scenario database, revise the analysis and mining methods of the scenario, or reconstruct the production scenario according to the requirements, and update and supplement the scenario database.

C. Typical Scenario Databases

Autonomous driving tests are characterized by its long time period and wide coverage of scenarios. A large-scale autonomous driving scenario database is the basis for autonomous driving testing. Only by increasing the number and richness of scenarios can it be possible to effectively discover some special scenarios and boundary conditions, thus effectively promoting the iteration and implementation of technologies. Many domestic enterprises have made some breakthroughs in this area and established their own scenario databases, such as the "Case Scenario Database" of China Automotive Technology and Research Center, and the "China Typical Driving Scenario Database V3.0" built by China Automotive Engineering Research Co., Ltd., Baidu's "Apollo scenario database", Tencent's "TAD Sim scenario database", as shown in the following TABLE III.

VI. SUMMARY AND PROSPECT

With the continuous development of autonomous driving technology research, the difficulty of autonomous driving testing has also increased. Developing test scenarios and building a database of autonomous driving scenarios is a promising direction in the future. At present, many domestic researchers have conducted a lot of research on this, and many companies in the market have gradually developed their own autonomous driving scenario databases and achieved certain results. But there are still many problems, and current research cannot keep up with the rapid development of autonomous driving technology. In the future, there is still a huge research space for the development of autonomous driving scenarios and scenario databases. Here are the following research areas that can be considered and explored:

- The autonomous driving scenario has the problems of inconsistent scenario design elements, inconsistent scenario partition models, and inconsistent scenario format. The industry has not yet reached a consensus on the definition and architecture of autonomous driving scenarios. Different test methods and technical means require different scenarios. In the future, it is necessary to start from the actual testing, fully analyze the requirements of different testing methods for the scenario, improve the scenario architecture, and form the definition of the autonomous driving scenario on this basis.

TABLE III. TYPICAL SCENARIO DATABASES

Scenario Database	From	Type of Scenario	Source
Case Scenario Database	China Automotive Technology and Research Center (CATARC)	Naturalistic driving scenario, Accident scenario, Standard regulatory scenario, Empirical scenario	highway, urban road, country road, parking lot
China Typical Driving Scenario Database V3.0	China Automotive Engineering Research Institute Co., Ltd.	Naturalistic driving scenario, Accident scenario, Standard regultory scenario, Function failure scenario, Empirical scenario, Typical accident scenario, Safe scenario of the intended functionality	highway, urban road, urban expressway, country road, national trunk highway
Apollo Scenario Database	Baidu	Naturalistic driving scenario, Standard regulatory scenario, High collision risk scenario(dangerous scenario)	urban road, urban expressway
TAD Sim Scenario Database	Tencent	Naturalistic driving scenario, Accident scenario, Standard regulatory scenario, Empirical scenario	highway, urban road, mountain road, park road

- The ASAM OpenX series of scenario description standards have been widely used or referenced. For the description of dynamic scenarios, OpenSCENARIO V1.x is further developed into

OpenSCENARIO V2.0, which aims to define scenarios from a higher level of abstraction. However, from the perspective of the current technology maturity, the abstract representation capability of OpenSCENARIO V2.0 needs to be further verified, and the comprehensiveness of the dynamic scenario description method needs to be further improved.

- At present, most of the data sources of the scenario databases come from naturalistic driving data and accident data. Based on the existing and new research on various driving scenarios, data analysis and theoretical analysis are carried out to construct the scenario database. However, these scenarios are known scenarios. For unknown scenarios, due to their unpredictability, autonomous driving vehicles cannot complete all verifications before they go on the road. In the future, it is necessary to conduct research on mining unknown scenarios to improve the reliability and safety of autonomous driving tests.
- The research on the automatic driving scenario database is in the stage of continuous exploration and development. The scenario generation process is relatively difficult, and the autonomous driving test also has great requirements on the data scale of the scenario database. Therefore, the richness, diversity, and effectiveness of the scenario database all require the joint efforts of research developers. In the future, we can focus on the rapid generation and extraction of scenarios, and study accelerated testing methods and scenario database construction methods for accelerated testing.
- Scenario-based autonomous driving testing is still in the initial stage of technological development, and requires the joint advancement of researchers in multiple fields. Future research should focus on breaking through core common technologies, formulate and improve autonomous driving test standards, and establish relevant standard systems. These will provide guidance for autonomous driving testing work and scenario database construction, and provide strong support for the development of autonomous driving technology and industrial implementation.

References

[1] Najm W G, Sen B, Smith J D. Analysis of light vehicle crashes and pre-crash scenarios based on the 2000 general estimates system[R]. United States. National Highway Traffic Safety Administration, 2003.

[2] Najm W G, Ranganathan R, Srinivasan G. Description of light-vehicle pre-crash scenarios for safety applications based on vehicle-to-vehicle communications[R]. United States. National Highway Traffic Safety Administration, 2013.

[3] Bálint A, Fagerlind H, Kullgren A. A test-based method for the assessment of pre-crash warning and braking systems[J]. Accident Analysis & Prevention, 2013, 59: 192-199.

[4] Wolf C T. Explainability scenarios: towards scenario-based XAI design[C]//Proceedings of the 24th International Conference on Intelligent User Interfaces. 2019: 252-257.

[5] Geyer, S., Baltzer, M., Franz, B., Hakuli, S., Kauer, M., Kienle, M., Meier, S., Weißgerber, T., Bengler, K., Bruder, R., et al. (2014). Concept and development of a unified ontology for generating test and use-case catalogues for assisted and automated vehicle guidance. Intelligent Transport Systems, IET, 8(3):183-189.

[6] Go K, Carroll J M. The blind men and the elephant: Views of scenario-based system design[J]. interactions, 2004, 11(6): 44-53.

[7] Ulbrich S, Menzel T, Reschka A. Defining and substantiating the terms scene, situation, and scenario for automated driving[C]//2015 IEEE 18th international conference on intelligent transportation systems. IEEE, 2015: 982-988.

[8] Elrofai H, Paardekooper J P, de Gelder E. Scenario-based safety validation of connected and automated driving[J]. Netherlands Organization for Applied Scientific Research, TNO, Tech. Rep, 2018.

[9] Bing Z H U, Pei-xing Z, Jian Z. Review of scenario-based virtual validation methods for automated vehicles[J]. China Journal of Highway and Transport, 2019, 32(6): 1.

[10] Groh K, Thomas Kühbeck M S. Towards a scenario-based assessment method for highly automated driving functions[C]//8. Tagung Fahrerassistenz. 2017.

[11] Sauerbier J, Bock J, Weber H. Definition of scenarios for safety validation of automated driving functions[J]. ATZ worldwide, 2019, 121(1): 42-45.

[12] PEGASUS. Scenario Description and Knowledge-Based Scenario Generation [EB/OL]. [2021-4-25]. https://www.pegasusprojekt.de.

[13] Scholtes M, Westhofen L, Turner L R. 6-layer model for a structured description and categorization of urban traffic and environment[J]. IEEE Access, 2021, 9: 59131-59147.

[14] Bagschik G, Menzel T, Maurer M. Ontology based scene creation for the development of automated vehicles[C]//2018 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2018: 1813-1820.

[15] SCHIEMENTZ M. Test case variation and execution[R]. Berlin: Research Project PEGASUS, 2019.

[16] ASAM OpenDRIVE: User Guide [EB/OL].[2021-08-03]. https://www.asam.net/index.php?eID=dumpFile&t=f&f=4422&token=e590561f3c39aa2260e5442e29e93f6693d1cccd.

[17] ASAM OpenSCENARIO: User Guide [EB/OL].[2022-05-13]. https://www.asam.net/index.php?eID=dumpFile&t=f&f=4908&token=ae9d9b44ab9257e817072a653b5d5e98ee0babf8.

[18] Krajewski R, Bock J, Kloeker L. The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems[C]//2018 21st International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2018: 2118-2125.

[19] Alexiadis V, Colyar J, Halkias J. The next generation simulation program[J]. Institute of Transportation Engineers. ITE Journal, 2004, 74(8): 22.

[20] Bock J, Krajewski R, Moers T. The ind dataset: A drone dataset of naturalistic road user trajectories at german intersections[C]//2020 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2020: 1929-1934.

[21] R. Krajewski, T. Moers, J. Bock, L. Vater and L. Eckstein, "The rounD Dataset: A Drone Dataset of Road User Trajectories at Roundabouts in Germany," 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), 2020, pp. 1-6, doi: 10.1109/ITSC45102.2020.9294728.

[22] A. Zyner, S. Worrall, and E. Nebot, "ACFR Five Roundabouts Dataset: NaturalisticDriving at Unsignalised Intersections," IEEE Intelligent Transportation Systems Magazine, Early access:https://doi.org/10.1109/MITS.2019.2907676.

[23] A. Zyner, S. Worrall, and E. Nebot, "ACFR Five Roundabouts Dataset: NaturalisticDriving at Unsignalised Intersections," IEEE Intelligent Transportation Systems Magazine, Early access:https://doi.org/10.1109/MITS.2019.2907676.

[24] Zhan W, Sun L, Wang D. Interaction dataset: An international, adversarial and cooperative motion dataset in interactive driving

scenarios with semantic maps[J]. arXiv preprint arXiv:1910.03088, 2019.

[25] MIRROR. Holographic intelligent scenarios derived from real traffic data [EB/OL]. http://www.scenarios.cn/html/dataset_en.html.

[26] Ma Y, Zhu X, Zhang S. Trafficpredict: Trajectory prediction for heterogeneous traffic-agents[C]//Proceedings of the AAAI conference on artificial intelligence. 2019, 33(01): 6120-6127.

[27] Woo J W, Lee S B. Test-bed design for evaluation of intelligent transportation systems and intelligent vehicle systems[C]//13th International Conference on Advanced Communication Technology (ICACT2011). IEEE, 2011: 1511-1514.

[28] Google. Waymo is first to put fully self-driving cars on US roads without a safety driver [EB/OL]. [2017-11-07]. https://www.theverge.com/2017/11/7/16615290/waymo-self-drivingsafety-drive-chandler-autonomous.

[29] Zeng Desong, Gao Chen, Tan Beihai, Yu Rong. Research on the Construction of Autonomous Driving Simulation Test Scenario Library [J]. Modern Computer, 2021, 27(31): 1-12+32.

[30] Zhao Xiangmo, Cheng Jingjun, Xu Zhigang, Wang Wenwei, Wang Runmin, Wang Guanqun, Zhu Yu, Wang Guiping, Zhou Yu, Chen Nanfeng. Indoor Rapid Test Platform for Autonomous Vehicles Based on Vehicle-in-the-Loop Simulation [J]. Journal of China Highway and Transportation, 2019,32(06):124-136.DOI:10.19721/j.cnki.1001-7372.2019.06.013.

[31] Zhang Y D, Le-Nan W U. Survey on development of expert system[J]. Computer Engineering & Applications, 2010.

[32] Kang Y, Yin H, Berger C. Test your self-driving algorithm: An overview of publicly available driving datasets and virtual testing environments[J]. IEEE Transactions on Intelligent Vehicles, 2019, 4(2): 171-185.

[33] Ries L, Langner J, Otten S. A driving scenario representation for scalable real-data analytics with neural networks[C]//2019 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2019: 2215-2222.

[34] Gambi A, Huynh T, Fraser G. Generating effective test cases for self-driving cars from police reports[C]//Proceedings of the 2019 27th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering. 2019: 257-267.