

# A Survey on Datasets for Decision-making of Autonomous Vehicle

Yuning Wang, Zeyu Han, Yining Xing, Shaobing Xu\*, Jianqiang Wang\*  
School of Vehicle and Mobility, Tsinghua University, Beijing, China

**Abstract**—Autonomous vehicles (AV) are expected to reshape future transportation systems, and decision-making is one of the critical modules toward high-level automated driving. To overcome those complicated scenarios that rule-based methods could not cope with well, data-driven decision-making approaches have aroused more and more focus. The datasets to be used in developing data-driven methods dramatically influences the performance of decision-making, hence it is necessary to have a comprehensive insight into the existing datasets. From the aspects of collection sources, driving data can be divided into vehicle, environment, and driver related data. This study compares the state-of-the-art datasets of these three categories and summarizes their features including sensors used, annotation, and driving scenarios. Based on the characteristics of the datasets, this survey also concludes the potential applications of datasets on various aspects of AV decision-making, assisting researchers to find appropriate ones to support their own research. The future trends of AV dataset development are summarized.

**Index Terms**—Autonomous vehicles, decision-making, dataset, data application.

## I. INTRODUCTION

THE technology of self-driving is placed great expectations of rising revolution in people’s travel manner and future transportation system [1]. The architecture of self-driving technology can be roughly divided into three main modules: environmental perception, decision making, and motion control [2] [3]. In particular, the decision-making module plays the role of the driver’s brain with great significance.

Along with the rapid development of self-driving technology, the research on decision-making is also on the rise [3] [4] [5]. Decision-making is usually decomposed into hierarchical modules, including global route planning, behavior selection, and trajectory generation [5]. Their boundaries were explicit but are becoming obscure as the deep learning and end-to-end approaches emerge [2]. In this paper, we use a relatively generalized definition of decision-making, which includes not only motion planning but also environment and driver understanding.

The technique of decision-making can be driven by physics/optimization-based rules, or by machine intelligence learned from data, as shown in Fig. 1. As for low-level self-driving, rule-based approaches, are more widely used to

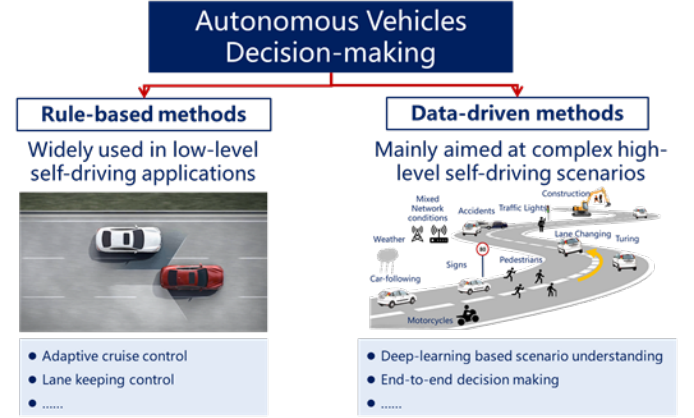


Fig. 1. Two types of autonomous vehicle decision-making algorithms

achieve self-driving capacity in simple scenarios. However, toward high-level self-driving technology, rule-based approaches are usually considered to lack generalizability in complex traffic situations [6]. To this end, data-driven approaches, including machine learning, deep learning, and reinforcement learning, have aroused researchers’ interest. These approaches outperform the rule-based approaches in some complex scenarios, as reported in [7], but their performance is actually restricted by the datasets used to train and test the algorithms. To guarantee unbiasedness and generalization of these approaches, proper and abundant data is essential and dominates data-driven decision making.

A number of large-scale datasets for decision-making have been created recently, and their features and applications vary greatly. Fig. 2 demonstrates the size explosion of decision-making datasets and the rapid growth of demands. From the aspect of data volumes, it exploded by orders of magnitude in the last 15 years, growing from one hour to over 1000 hours with much richer scenarios. Fig. 3 shows a detailed developing trend of decision-making datasets, including name, developer, and main features of each dataset. Some of them have become the milestones in this field. The data types and characteristics related to decision-making are various. Some focus on vehicle status [8], some concern about environmental understanding [9], while some pay attention to drivers’ behavior analysis [10]. The recent datasets have laid more emphasis on challenging scenarios instead of regular ones [11] [12] [13] [14]. Another obvious trend is that recently emerging datasets had much larger data volume, and the types and sources of data are also more varied. Therefore, surveying these latest datasets

Yuning Wang and Zeyu Han contribute equally to this work. This research was funded by the National Natural Science Foundation of China with award 52131201 (the key project) and Tsinghua University Toyota Joint Research Center for AI Technology of Automated Vehicle (TTAD 2022-06), the Intelligent driving and high-precision map key technology verification platform cooperation project.

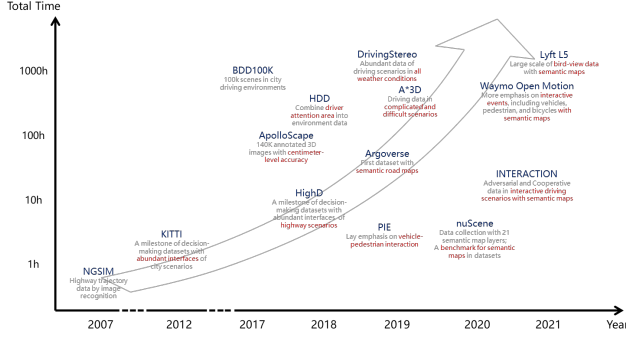


Fig. 2. The size of decision-making datasets for autonomous vehicle design

matches the increasing demand.

Recently, Yin Hang and Christian Berger presented an overview of 27 existing publicly available self-driving datasets and provided comparisons and guidelines for researchers to select datasets [15]. The methodologies of collecting datasets and the highlights of each dataset are also listed in that paper. However, the information and analysis provided are relatively less complete. As an extension, [16] analyzed virtual testing environments in addition to real on road or in-track testing. Junyao Guo et al. categorized the datasets and highlighted the datasets with complicated environments [17]. It also introduced the factors and metrics of drivability to measure the adaptability of autonomous driving in various driving conditions. Di Feng et al. demonstrated datasets for deep multi-modal object detection and semantic segmentation [18]. As a summary, most self-driving datasets and correlative surveys concentrated on environmental perception technology such as object detection and object tracking [19] [20], however, attention to datasets for decision making is insufficient. Therefore, a comprehensive survey focusing on datasets for decision-making is needed.

Compared with previous surveys of self-driving datasets, the major contributions of this paper are as follows:

- 1) This paper presents a detailed summary of self-driving datasets for decision-making, especially, many latest datasets that are not covered by other previous papers are included, which benefits researchers in this area to obtain a more focused view of related datasets.
- 2) The datasets are categorized by the data sources, i.e., drivers, vehicles, and environment. This categorization provides a criterion for classifying and comparing different datasets. Besides, various data types and sources classified in this categorization are convenient for researchers to choose datasets adaptive to their sensors and research demand.
- 3) This paper summarizes the application purposes of datasets for decision-making. This information helps researchers to choose appropriate datasets for different research.

The remainder of this paper is organized as follows. Section II-IV introduces the decision-making datasets in the order of vehicle-related datasets, environment-related datasets, and driver-related datasets, respectively. Section V demonstrates

dataset application, and Section VI discusses the future trend of decision-making datasets. The conclusion is provided in Section VII.

## II. VEHICLE DATASETS

Data from vehicles reflects the basic information of self-driving vehicles. Therefore, the related datasets have wide applications in research on self-driving decision-making. The collection of vehicle-related data is more convenient than environment or drivers related data, but the analysis is relatively complex. In this section, 11 datasets including data types and collection platforms are reviewed.

### A. Datasets Classification

There are mainly three classes of vehicle-related data, as shown in Fig. 4.

- 1) Status data: this class of data reflects the location, speed, and other external status information of the vehicle, such as distance, speed, yaw rate, and acceleration from GNSS, RTK, IMU and wheel speed sensors.
- 2) Manipulation data: this class of data contains drivers' manipulation information, including steering angle, steering rate, accelerator pedal position, and brake pedal position data.
- 3) Internal data: this class of data records the internal status of the vehicle, such as the engine speed, the liquid pressure of brake valves, the transmission gear position, and other valuable data on the CAN bus.

The status data is usually collected along with the perception data such as LiDAR and camera data, which acts as the input data of the decision-making algorithm. The manipulation data reflects the final results of decision-making and usually serves as the ground truth of the decision-making algorithm. Besides, the internal data contains underlying information, which sometimes is also used as the input of learning-based algorithm. In Table I, 11 relative self-driving datasets for decision-making are listed as a comparison.

TABLE I  
DATA CLASSES IN VEHICLE DATASETS

Dataset	Status Data	Manipulation Data	Internal Data
EU Long-term [21]	✓	×	×
KAIST [22]	✓	×	×
BDDV [23]	✓	×	×
Udacity [24]	✓	✓	×
A2D2 [8]	✓	✓	×
HDD [25]	✓	✓	×
DBNet [26]	✓	✓	×
DDD17 [27]	✓	✓	✓
DDD20 [28]	✓	✓	✓
D <sup>2</sup> CAV [29]	✓	✓	✓

A group of datasets includes status data only. Their location data (GPS/IMU) and odometry data are mostly combined with other sensors such as cameras and LiDAR to design perception and localization algorithms. The EU Long-term dataset [21] collects GPS and IMU data along with LiDAR, camera, and radar data, allowing the vehicle to perceive its surroundings

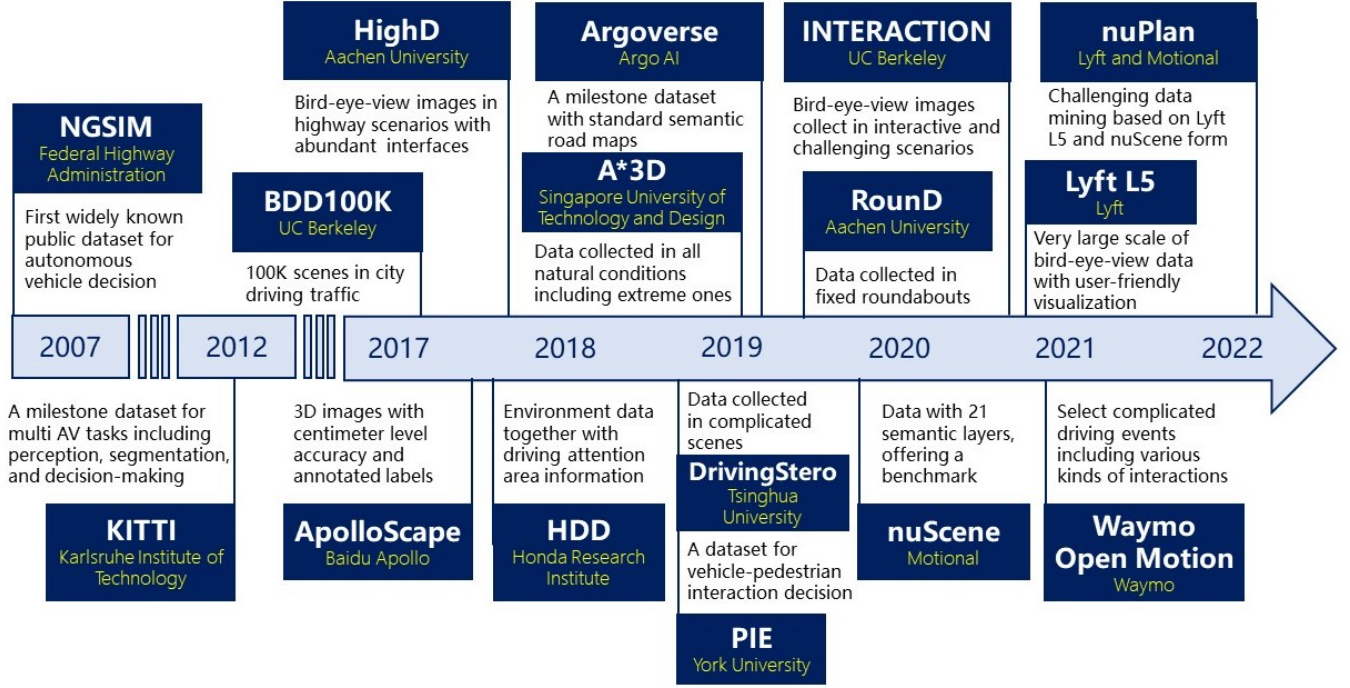


Fig. 3. The developing history of autonomous vehicle decision-making datasets

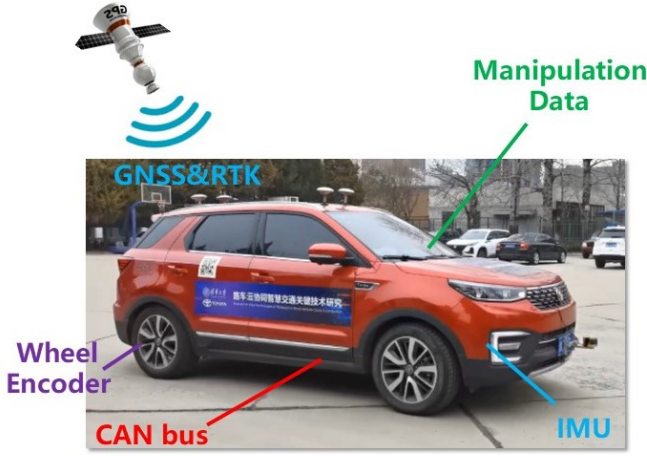


Fig. 4. Typical vehicle data

and locate itself. The data is collected by a multisensory platform and ROS as the middleware. Similarly, the KAIST dataset [22] records GPS, IMU, and wheel encoder data for SLAM algorithm design. Besides, BDDV [23] is a dataset comprised of real driving videos, vehicle course and speed, and data from GPS, IMU, gyroscope, and magnetometer. These modalities can be used to recover the trajectory and dynamics of the vehicle to serve as the ground truth of end-to-end self-driving algorithms.

However, to learn an end-to-end self-driving algorithm, it is more appropriate to learn from vehicle manipulation instead of the trajectory only. A group of datasets involves the manipulation data for this purpose. The Udacity dataset

[24] collects camera, LiDAR, GPS, IMU, and the manipulation data from CAN bus including steering angle and brake/throttle pedal opening, which can be used to train a steering control network or an image-based localization algorithm. The A2D2 dataset [8] also includes manipulation data as labels for reinforcement learning research. Besides the end-to-end algorithms, numerous research pays attention to behavioral decision learning, and several datasets emerge consequently. The HDD dataset [25] records CAN data including throttle opening, brake pressure, steering angle, yaw rate, and speed, together with location information and data from other sensors. The DBNet dataset [26] is also a large-scale dataset containing driver manipulation in a variety of traffic conditions for driving behavior learning.

Since manipulation behavior is usually correlated with of the internal status of vehicles, such as accelerating corresponds to increasing engine speed, a number of datasets collect not only the vehicle's kinematic status and drivers' manipulation data, but also the vehicle's internal data. The DDD17 dataset [27] records abundant types of internal data including the engine speed, torque input to the transmission, fuel consumption, and gear position of transmission along with other kinematic status data and manipulation data. These data including the images from an event camera enables researchers to perform end-to-end learning of steering control. Based on DDD17, the DDD20 dataset [28] expands the data duration from 12h to 39h. Note that these two datasets both do not contain LiDAR, radar, and other sensors necessary for a complete self-driving solution. Toghi et al. [29] discover the fact that the driver behavior distribution in most datasets is imbalanced, thus presenting the D<sup>2</sup>CAV dataset from human-driven vehicles performing evenly distributed behaviors, containing CAN data

such as engine speed for manipulation prediction purposes.

### B. Limitations

Taking an overview of datasets for decision-making of self-driving cars, the limitations of existing datasets can be summarized into two main aspects: the lack of driver data and the lack of standards of data type.

1) *The lack of driver data:* Most datasets containing vehicle manipulation and internal status data are created mainly for end-to-end learning research such as predicting the steering angle based on the camera or LiDAR data. However, these algorithms have a non-negligible disadvantage of interpretability. Recently, plenty of researchers have paid their attention to driver's behavior learning [2], which learns behaviors such as overtaking, car following, and lane changing from human driving data. These behaviors can be easily understood by human drivers, and can be even considered as a priori knowledge for learning the motion or manipulation of the vehicle. This promising research area calls for datasets containing not only data from perception sensors and CAN bus, but also from the drivers, including the driver behaviors, postures, and attention areas. D<sup>2</sup>CAV [29] is a comparatively comprehensive dataset containing driver behavior, but the scenarios are relatively limited. As for driver postures or attentions, they are mostly studied for driver status recognition, and there has been no dataset yet that collects them along with perception sensor data and vehicle data for driver behavior learning research.

2) *The lack of data type standards:* Data from vehicles are of great variety, especially the CAN bus data. It is often simple to collect all relevant data from the CAN bus, but specifying which types of data are more useful for self-driving decision-making, and building comprehensive data standards for datasets is complex yet important work. The lack of data type standards also limits the comparison between different decision-making datasets. The construction of data type standards for datasets is a holistic study that is noteworthy for the whole community of autonomous vehicle research.

## III. DRIVING ENVIRONMENT DATASETS

Datasets related to driving environment have been studied extensively since the 21st century to accelerate vehicle technology development [17]. With the breakthrough of hardware and neural networks, a large number of complicated datasets have been proposed with varieties of sensors, data formats, and scenarios [16] [30]. The features are so diverse that it is confusing for developers to choose the most appropriate dataset for a specific study where data mining is a must for research. Consequently, this section reviews those environment-related datasets and highlights their features.

### A. Data Types and Annotations

While using datasets for academic or engineering purposes, data types and annotation labels are two of the features that users are most concerned about, especially for driving environment-related tasks since the real driving scenarios are so complicated and diverse. Only if the input and output

interfaces are included in the datasets could the model be trained. In Table II and III, overviews of sensors used and annotation labels are given, illustrating the features of existing datasets. In Table II, Cam is the abbreviation of Cameras.

TABLE II  
SENSORS ON COLLECTION PLATFORMS OF DRIVING ENVIRONMENT DATASETS

Dataset	Cam	Radar	LiDAR	V2X	Others
ApolloScape [31]	✓	×	✓	×	
nuScenes [32]	✓	✓	✓	×	
WAYMO [12]	✓	×	✓	✓	
WoodScape [33]	✓	×	✓	×	
KITTI [34]	✓	×	✓	×	
1001 hours [35]	✓	×	✓	✓	
NGSIM [36]	×	×	×	✓	
HighD [37]	×	×	×	✓	
ROAD [38]	✓	×	×	×	
Argoverse [39]	✓	×	✓	×	Stereo Cameras
PIE [40]	✓	×	×	×	
TITAN [41]	✓	×	×	×	
DrivingStereo [42]	✓	×	✓	×	
ONCE [43]	✓	×	✓	×	
H3D [9]	✓	×	✓	×	
HDD [25]	✓	×	✓	×	
CADC [44]	✓	×	✓	×	Thermal Cameras
BDD100K [45]	✓	×	×	×	
INTERACTION [13]	×	×	×	✓	
PREVENTION [46]	✓	✓	✓	×	
DDD20 [28]	✓	×	×	×	Event Cameras
A*3D [47]	✓	×	✓	×	

As for sensors, the most chosen combination is multiple cameras and a multi-thread LiDAR. Together with images and point clouds, the driving environment can be reproduced precisely [9]. Cameras ensure the semantic target classification and LiDAR is usually responsible for depth matching [42]. Some datasets mainly aim at pattern analysis tasks where positioning is not a must, so merely using multiple cameras without LiDAR is also a common configuration [40] [41]. With the development of connected transportation, it is easier to establish a precise mapping of the real traffic environment with the help of emerging perception technologies. Krajewski et al [37] concluded four methods for data capture, e.g., using drones and infrastructure-based sensors, which referred to as V2X connection technology. This method can capture the whole scenario and is a good complement of onboard sensors. The datasets including 1001 Hours [35], INTERACTION [13], Top-view Trajectories [28], etc., took advantage of roadside units and drones to generate precise global virtual environments. In addition, to adapt to specific scenarios, some specialized sensors are utilized. Argoverse [39] used stereo cameras to form spatial features. CADC [44] includes thermal cameras to perceive objects in snowy environments. DDD20 [28] proposed that event cameras contribute hugely to capturing dynamic environment variations. In conclusion, diverse sensor combinations determine the various features and intentions of datasets.

Apart from the raw data, the annotation label is another important factor users consider for model training and development. We conclude four classes of annotations in environment-related datasets for decision-making research, which are bounding boxes in first-person view (FPV), physical states in bird eye view, high-definition (HD) map, and behavior



labels [17]. Bounding boxes in FPV are abundant in almost all recent AV datasets due to the breakthrough in the field of computer vision. Many auto-label models have been developed to offer precise label information [9] [38]. Compared with the FPV bounding boxes, the physical states in bird-view (BV) are more helpful for developing decision methods of high-level AD, since targets in the blind spots of self-vehicle's vision need to be considered [48]. Some typical datasets including INTERACTION [13] and 1001 hours [35] generate bird-view information by roadside sensors or drones directly, and other datasets like WAYMO OPEN MOTION [12], H3D [9] combined various resources to create a comprehensive digital twin of the whole driving scenarios. HD map refers to maps with lane-level recognition including drivable areas, obstacles, etc. NuScenes [32] manually input the semantic map of collected scenes for further application with the dataset. DrivingStereo [42] used stereo cameras to derive a colorized disparity map containing semantic information. ONCE [43] generated the HD map with the SLAM technology.

Behavior labels are one of the focuses in recent years since more and more deep-learning decision-making models need them for ground truth or intermediate variables. For instance, Toghi et al [29] and Jain et al [49] label events into discrete behaviors including turns, lane shifts, deceleration, etc. However, abstracting the whole events into one single label greatly compresses the amount of information, which is less accurate actually. Comparatively, HDD invented a 4-layer annotation structure, decomposing an event into target goals, behaviors, reasons, and attention area, and each layer corresponded to a finite label set [25]. Similarly, Singh et al [38] summarized an event with a triplet structure consisting of active agents, actions it performs, and the corresponding event locations in the ROAD dataset. Compared with one single label, sufficient antecedents and consequences express a traffic event better, making it easier for the training model to achieve precise decisions and possess interpretability.

TABLE III  
ANNOTATIONS OF DRIVING ENVIRONMENT DATASETS

Dataset	Bounding box in FPV	Bird-view Physical States	HD map	Behavior Labels
ApolloScape [31]	✓	×	×	×
nuScenes [32]	✓	×	✓	×
WAYMO [12]	✓	✓	✓	✓
WoodScape [33]	✓	✓	✓	×
KITTI [34]	✓	×	✓	×
1001 hours [35]	×	✓	✓	×
NGSIM [36]	×	✓	×	×
HighD [37]	×	×	×	✓
ROAD [38]	✓	×	×	✓
Argoverse [39]	✓	✓	✓	×
PIE [40]	✓	×	×	✓
TITAN [41]	✓	×	×	✓
DrivingStereo [42]	×	×	✓	×
ONCE [43]	✓	×	✓	×
H3D [9]	✓	✓	✓	×
HDD [25]	✓	×	×	✓
CADC [44]	✓	✓	×	×
BDD100K [45]	✓	×	✓	×
INTERACTION [13]	✓	✓	✓	×
PREVENTION [46]	✓	×	×	✓
DDD20 [28]	✓	×	×	✓
A*3D [47]	✓	×	✓	×

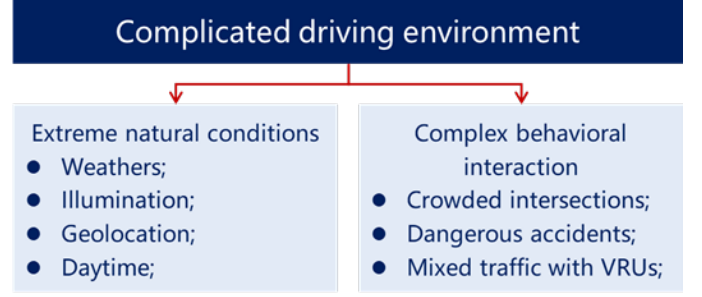


Fig. 5. Two kinds of complicated driving environment

As a summary, the sensor selection is diverse among different datasets currently, which makes it difficult to develop generalized decision-making methods across different datasets. Besides, as for annotations, the labels of behaviors are not unified.

### B. Data Scenarios

NGISM and KITTI are two of the earliest naturalistic driving datasets related to driving environment description. They collected vehicle trajectories either from FPV or BV [36] [34]. The large volumes of data have promoted the development of driving assistance system and autonomous cars greatly. However, one disadvantage is that in these two datasets, the majority of recorded scenarios are regular ones such as car-following and lane changing, failing to test the system's ability to drive in dangerous occasions [37]. Therefore, in recent years more and more researchers have been focusing on complicated driving environments.

As demonstrated in Fig. 5, extreme conditions can be divided into two aspects. One is about natural conditions, and the other is about complex behavioral interactions. Ligocki et al. [50] used infrared cameras to shoot videos in extreme conditions such as dark night and heavy rain. Yang et al. [42] proposed the DrivingStereo dataset that involved nearly all kinds of weather and illumination conditions, like cloudy and ducky days. Boreas is a multi-season self-driving dataset featured by involving all four seasons and many extreme kinds of weather [51]. In general, data in challenging nature conditions are becoming more and more mature.

Another aspect of the complicated environment is behavioral interaction. As we know, most existing AVs did not achieve high-level automation yet [54], and one of the critical problems is that decisions in complex interactive traffic scenes are not efficient enough compared to human drivers. Therefore, some studies have been dedicated to developing highly interactive datasets to support the decision-making development aimed at complicated traffic conditions. Waymo extracts complex scenes, defined as those at least involving two conflicting agents, from videos of real-world driving and further divide them into nine subclasses such as merging, unprotected turns, and intersections [12]. In their dataset, 77.5% scenarios contain more than two interactive agents. Instead of classifying by agent numbers, the Argoverse dataset collects 300 interesting scenes including intersections, lane changing, etc. [39]. From the perspective of traffic safety, the

TABLE IV  
DATASETS INCLUDING COMPLEX BEHAVIORAL INTERACTIONS

Dataset	Behavioral Interaction Descriptions
ApolloScape [31]	Three level of complexity divided by the number of moving objects in views
WAYMO [12]	More than 77.5% scenarios contain conflict areas among various agents
1001 hours [35]	Collect data in crowded places like roundabouts, large intersections, etc.
Argoverse [39]	300 interesting trajectories which satisfy at least one preset condition
PIE [40]	A interaction possibility to measure the complexity level of scenarios
H3D [9]	Challenging scenes with multiple agents and partial occlusion
HDD [25]	Scenes within a motion set, including turns, merges, etc.
INTERACTION [13]	Data collected in crowded locations such as roundabouts and large intersections
Top-view [28]	Interactions between vehicle and large numbers of pedestrians in public areas
DoTA [52]	Dangerous scenarios where collisions almost happen
PSI [53]	Drivers' post explanations for specific driving behaviors

dataset INTERACTION [13] included dangerous interactions in real world, offering materials for learning of decision making under extreme scenes. Kotseruba et al. [55] proposed the JAAD dataset which uses a first-person view to capture the interactions between ego- vehicle and pedestrians. Other datasets that include complex behavioral interactions are summarized in Table IV.

The analysis above shows that complicated traffic interactions have been broadly explored in recent large-scale datasets. However, unlike the simple classification of natural conditions, complex behavior interaction is still vague to describe. As illustrated above, the definitions and features of interaction among various datasets differ from each other. The standards for judging whether interaction exists are also different, e.g., route conflictions [12] [55], specific sets of actions [39], the possibility of hazard [13], or just presence of surroundings agents [9]. Guo et al. proposed the concept of drivability consisting of scene drivability, a binary collision indicator, and discretized behaviors to indicate the interaction complexity [17]. But the evaluation method was still coarse and incomplete. Therefore, how to design ideal datasets to make them really interactive is still a topic worth discussing.

### C. Data Labeling Techniques

Unlike the data related to vehicles or drivers, the source data of traffic environment including raw videos, LiDAR point clouds, etc., usually cannot be used by decision-making models directly. How to make correct and reasonable labels of a large amount of data efficiently is challenging.

The majority of annotation methods still rely on human works [17] [15]. However, some auto-labeling models are also developed to accelerate the process. In the ROAD dataset [38], a 3D-RetinaNet architecture was proposed for online data



Fig. 6. Top row: Synthesized realistic data; Bottom row: Real camera image

processing. Houston et al. [35] developed a labeling toolkit to process sensor data into semantic bird-view trajectories automatically. Latest deep learning methods such as the teacher-student model could realize massive processing on the basis of a small amount of manually labeled data frames [56].

A novel approach to label data is synthesized realistic technology [57]. Data sensors in the real world have blind zones of perception, resulting in some data defects. To overcome the difficulty, Yang et al. presented an effective method to generate a digital twin scenario based only on limited amounts of LiDAR and camera on real vehicles so that developers could extract desired target data at each spot and time of the traffic event, as shown in Fig. 6. To realize the mapping from real to digit, a deep learning network named SurfelGAN is leveraged to reconstruct the missing states in real the real world.

In conclusion, the labeling technique gradually becomes more and more automatic because of emerging deep-learning-based toolkits. Nevertheless, most behavior labels still rely on manual annotations.

## IV. DRIVER DATASETS

As autonomous vehicles develop to a higher level, many studies found that learning from human driver intelligence or behavior is helpful for better decision-making [58] [59]. As one source of model input, driver information adds interpretability to the deep-learning-based methods by combining driver logic frameworks into algorithms and can improve decision performance [1] [60]. Hence, datasets concentrated on drivers have gained increasing focus in the last decade. Fig. 7 shows an overview of driver datasets. Data can be generally classified into three categories: driver attention, driver posture, and driver physiological signals. The former two types of data could be used to understand driver behavior including distraction mode [61], interaction [62], etc. These labels may act as variables to select corresponding decision models. Driver physiological signals as well as posture contributes to classification of driver states such as fatigue driving [63] and driving performance rating [64] [65]. In this section, driver-related datasets are reviewed in the sequence of attention, posture, and physiological signals.

### A. Driver Attention

Datasets of driver attention mainly focus on the place or points that drivers pay attention to while driving. One way to collect drivers' attention data is to use an eye-tracking device to get the area where the drivers look at in the form

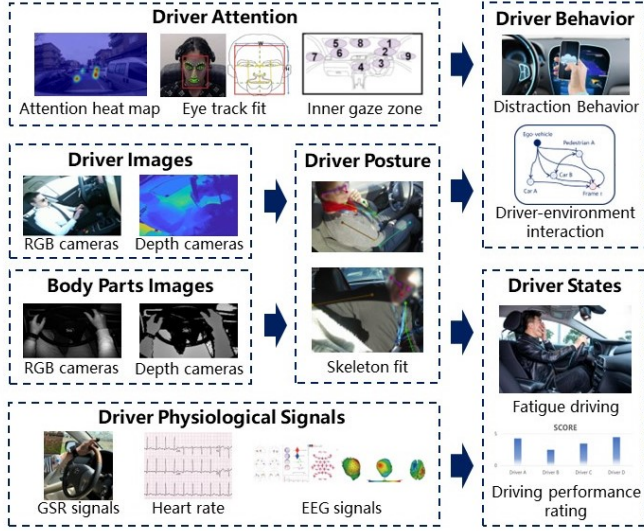


Fig. 7. An overview of driver datasets

of gaze maps. A dataset including 555000 frames collected by 8 drivers called Dr(eye)ve [10], was built in 2016 and used the SMI eye tracking glasses to get gaze maps. CoCAtt [66], a dataset built in 2021, used eye-tracking devices to get the difference in driver attention distribution between manual driving and automated driving. Some other ways can also be used to collect drivers' attention data. Speak2Label [58] applies speak-to-text conversion to automatically label the image-based gaze behavior dataset. Francisco Vicente [67] also utilized a low-cost camera (a Logitech c920 Webcam) placed on top of the steering wheel to get the driver's images. Then together with Parameterized Appearance Models and a system named SDM, images can be transferred to detect and track drivers' facial features, as well as estimate drivers' head pose.

Two kinds of eye-tracking devices are commonly used to collect attention data, as shown in Fig. 8. Glasses-formed eye-tracking device offers a first-person view. Cameras integrated into the glasses capture video the driver sees, and it rotates synchronously with the driver's head movement, ensuring sufficient view cover. One disadvantage is that the glasses hinder drivers from totally nature driving, especially for those who have worn their own glasses. Another type uses multiple cameras installed above the dashboard of vehicles. By perception from multi-cameras, eye movements can be extracted. With external installed devices, the drivers will not be disturbed. However, the fixed-cameras form is not capable of tracking eye movement when the driver looks backward or sideways.

### B. Driver Posture

Driver posture datasets are summarized in Table V. The data of driver images record the video of driver's body or driver's hands, which can be used to judge whether the driver is focusing on the road and can take over the control of the vehicle or not. Cameras fixed in cabins are the widely-used tools to collect data. For example, the datasets DriPE [70] and

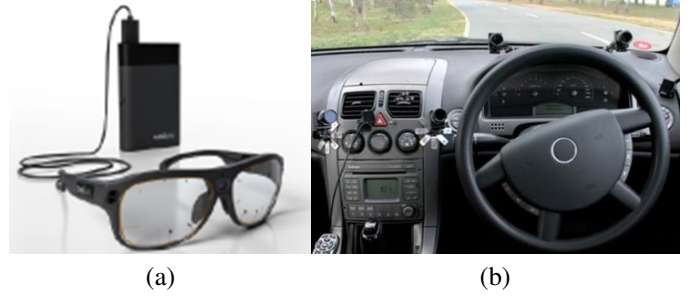


Fig. 8. (a) is the glasses-formed device, and (b) is the fixed form

TABLE V  
DATA TYPES OF DRIVING POSTURE DATASETS

Dataset	Body Parts Images	Driver Images	Driver Posture
Speak2Label [58]	×	✓	×
Brain4Cars [49]	×	✓	×
Drive&Act [68]	✓	✓	✓
Driver Anomaly Detection (DAD) [61]	✓	✓	×
3MDAD [63]	×	✓	×
VIVA [69]	×	✓	×
DriPE [70]	×	✓	✓

Driver Anomaly Detection(DAD) collected by Okan Kopuklu [61] use RGB cameras to record driver behavior. Apart from RGB camera, depth camera could enrich datasets and increase the accuracy of driver pose detection. Drive&Act [68] applied multi-modal sensors to collect various types of driver behavior data, such as color, infrared, depth and 3D body pose information. 3MDAD [63] derived data from two Microsoft Kinect cameras, both including color (RGB) sensors, infrared (IR) sensors, and microphones, and recorded data with an acquisition rate of 30fps. Focusing on the behaviors of drivers' hands, Nikhil Das [69] built a video-based driver hands dataset with different background complexities, illumination settings, users, and viewpoints. The dataset includes over 2000 annotated images from three testing vehicles and over 2000 images from YouTube which were recorded in unknown vehicles.

The labels of driver posture is important to driver-related datasets. DriPE [70] annotated driver postures manually by a bounding box and 17 key points including arms and legs each with 3 key points and 5 additional markers for the eyes, ears, and nose. The dataset Drive&Act [68] obtained 3D body pose via triangulation of 2D poses from 3 frontal views. The 3D head poses were obtained by employing the OpenFace algorithm. The hand detection dataset VIVA used the Aggregate Channel Features(ACF) object detector to extract hand pose.

### C. Driver Physiological signals

This category of driver data is deeply connected with traffic psychology. Physiological signals are commonly used to judge the driver's mental state and the level of various emotions such as nervousness, anxiety, fatigue, etc. According to the physiological indicators, the driver's performance can also be evaluated so that the distinguished ones are filtered out.



Ling et al [64] utilized electrocardiogram (ECG) and galvanic skin responses (GSR) to collect heart rates and skin temperatures. With these reflective inputs, they assessed driver cognitive loads by feature analysis. Xing et al [65] established a human-in-the-loop driving simulator based on the Ergolab system, a comprehensive platform that collects skin, muscle, breath, and blood oxygen signals. Remier et al [71] added vocal signal into the driver database by installing sensors on the front window. Besides, in Foy's research, the participants of experiments were asked to wear an EEG to collect the activation levels of prefrontal cortex.

The amount of driver physiology research is sufficient, however, no large-scale public dataset is available for human driver intelligence extraction. The main challenge is that few datasets provided time-aligned driving data of the vehicle, so no ground truth can be used as the training target of decision methods.

#### D. Limitation of Driver Datasets

Datasets for driver analysis mentioned above have two main shortages. The size of these datasets are relatively small, and the usage of driver datasets is also limited.

The sizes of the driver-related datasets are not large generally, and the scenes contained in these datasets are not abundant. For example, datasets about driver attention are mainly collected in urban scenes, which have clear traffic signs and good traffic conditions. Datasets about driver poses usually consider the influence of lighting conditions instead of traffic conditions, which also limits the usage of these datasets. Another limitation of current driver datasets is the incompleteness of data interfaces. None of the existing datasets covered all driver attention, posture, and physiological information simultaneously, which restricted potential applications on decision-making.

In conclusion, current driver-related datasets are still limited both in data volume and sources. And current datasets mainly focus on one single driver data type, which can't be used in the research in which driver action, attention and physiological information are all needed.

### V. DATASET APPLICATIONS

With the development of decision-making technology, the data types useful for decision-making algorithms are also expanded. The technical route of research on data-driven decision-making methods varies greatly, and the routes are summarized in Fig. 9. Some research tries to learn from human intelligence [72], some research directly uses the raw perception data to obtain decisions via end-to-end algorithms [73], while some research applies semantic environment understanding to guarantee better interpretability [74]. Each technical route has its own unique advantages and shortcomings, and also require different datasets. Therefore, choosing proper datasets for different research is a fundamental step but may also be confusing for beginners. Considering the previous survey research did not pay much attention to the applications of datasets for decision-making, this section illustrates the

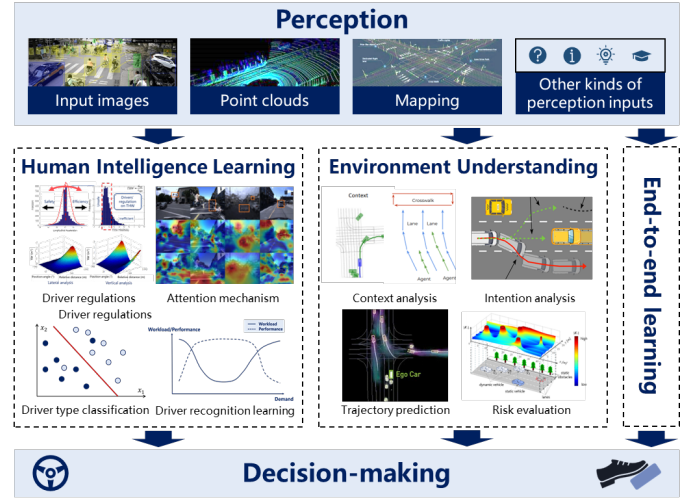


Fig. 9. Different technical routes of decision-making

potential applications of the mentioned datasets for decision-making.

Based on the different technical routes shown in Fig. 9, the applications of the introduced datasets for decision-making can be briefly divided into three aspects: datasets for driver analysis, environment understanding and decision learning. There are several subdivisions and typical datasets of each aspect as Table VI shows, which are detailed in the following.

TABLE VI  
DATASET APPLICATIONS, SUBDIVISIONS, AND TYPICAL DATASETS

Applications	Subdivisions	Typical Datasets
Driver Analysis	Driver Attention Driver Behaviors Context Analysis	Dr(eye)ve [10], CoCAtt [66], Speak2Label [58] DriPE [70], DAD [61], Drive&Act [68], VIVA [69] highD [37], DoTA [52], BDD100K [45], RUGD [75]
Environment Understanding	Intention Analysis Trajectory Prediction Behavioral Strategy	PIE [40], PSI [53], JAAD [55], HDD [25] Argoverse (2) [39] [76], PREVENTION [46] D <sup>2</sup> CAV [29], HDD [25], ROAD [38], Brain4Cars [49]
Decision Learning	trajectory planning manipulation learning	NGSIM [36], KITTI [34], DrivingStereo [42], WAYMO [12], INTERACTION [13] DBNet [26], DDD17 [27], DDD20 [28], A2D2 [8]

#### A. Datasets for Driver Analysis

Driver behavior analysis based on relative datasets could help autonomous vehicles to learn human-like decision-making. Current datasets for driver analysis can be roughly divided into three classes: driver attention, driver behavior, and driver operation. Datasets for driver attention mainly focus on the places or points to which drivers pay more attention while driving. Datasets for driver behavior mainly collect the drivers' actions and movement data, which could be used to detect and predict the status of the driver. As the third class, driver operation includes the basic driver inputs and their results, changing and overtaking, the corresponding dataset will be introduced in the part of the manipulation control dataset in Section V-C3. In the following we focus on the former two applications.

1) *Driver Attention*: Drivers usually pay more attention to places where dangerous situations might happen. By analyzing drivers' attention mechanism, the autonomous vehicle system



could focus on the riskier areas for better safety given limited perception and computing resource, or at least could detect the distraction situation of drivers and provide warning.

For example, the dataset Dr(eye)ve [10] can be applied to understand how the driver’s past experience may influence driver’s attention on different objects in the scene. The data can also be used to judge whether drivers keep their attention on the traffic or not while the autonomous driving system is working. Another dataset CoCAtt [66] can also be used to study the relationship between driver attention and autonomous driving system [66]. The speak-to-text conversion applied in the dataset Speak2Label [58] could estimate the drivers’ gaze zone roughly.

In general, datasets for driver attention mostly collect the attention data of drivers, and can be utilized by research on the analysis of driver’s attention area and distraction problem, to help the autonomous driving system to recognize important areas and be more focused.

2) *Driver Behaviors*: Considering the fact that the current autonomous driving technique is not mature enough to cover all the situations, which means drivers still need to keep their eyes on the road to be ready to take over the vehicle at any time, driver behavior datasets can help to identify the driver’s states, reflect driver’s decision, assess the complication level of the scenarios the driver is faced up with, judge whether the driver should take over the vehicle when the autonomous driving system fails to work well, etc. For manual-driving vehicles, it can also help estimate the posture, tiredness, and attention status of drivers.

The annotation of the dataset DriPE [70], including a bounding box and 17 key points, could benefit the research of driver posture estimation. Furthermore, some datasets focus on anomaly situations in favor of the practicability of the algorithms. The Driver Anomaly Detection (DAD) dataset [61] distinguishes anomalous driving situations to help researchers test the performance of the driver behavior detection algorithm under abnormal conditions. The multi-modal data collected in the Drive&Act dataset [68] enables driver behavior research by different sensors. Drivers’ hands are usually essential in driving. The VIVA dataset contains annotated images of the drivers’ hands for hand detection, provides another way to estimate drivers’ status.

In summary, there is a variety of valuable driver behaviors that are necessary to be detected to improve the security of both manual driving and autonomous driving. Related datasets and research are also vivid and provide various data and methods to train brilliant models.

## B. Datasets for Environment Understanding

Environment understanding is an indispensable link between environment perception and motion planning. For autonomous vehicles, the environmental information collected by the perception module is usually processed by the environment understanding module, to identify drivable areas, detect the intention of pedestrians, predict the future trajectory of road users, and so on.

In this subsection, environment understanding is categorized into three main subdivisions: context analysis, intention analysis, and trajectory prediction.

1) *Context Analysis*: We apply the word “context” here to indicate the factors or participants in the driving scenario related to the ego vehicle [77]. This topic includes research on scenario classification, anomaly detection, semantic segmentation, drivable area detection, and so on.

The highD dataset [37] assesses the context by criterion such as Time to Collision (TTC) and Time Headway (THW), which can be utilized for a system-level validation of highly automated driving systems. The DoTA dataset [52] contains 18 anomaly categories of driving video and multiple anomaly participants in different driving scenarios, which is significantly helpful for learning a robust detector to analyze anomaly context. Some other datasets can be applied to analyze road information and semantic segmentation. The BDD100K dataset [45] containing 100K videos and 10 context analysis tasks can be used for lane marking detection, and furthermore distinguishing drivable areas for autonomous vehicles. Besides, the dataset also benefits heterogeneous multitask learning, including lane marking, semantic instance segmentation, and so on. For driving on unstructured roads, identifying drivable areas is difficult and suffers from a lack of semantically labeled datasets compared to their urban counterparts. The RUGD dataset [75] which contains different and complicated areas of unstructured driving scenarios such as dirt, sand, grass, and trees, can help solve this unique challenge.

2) *Intention Analysis*: The intention of traffic participants is another significant criterion in decision-making. A number of datasets are designed to identify or predict the intention of pedestrians and vehicles.

Pedestrian behavior or intention anticipation is a key challenge in the design of assistive and autonomous driving systems. The PIE dataset [40] contains large-scale videos and models for pedestrian intention estimation and trajectory prediction. Specifically, this dataset concentrates on pedestrians that tend to interact with ego-vehicle and the behavior or intention is annotated by human experts. With the development of natural language processing techniques, the text-based explanation of driver’s understanding of the scenes and the reasoning process are recorded in the PSI dataset [53]. The explanations can link the critical visual scene features with the human decision-making process and results, enabling the development of explainable and human-like pedestrian intention analysis and prediction algorithms. Instead of focusing on pedestrians only, the JAAD dataset [55] is produced for the analysis on all surrounding road users. The HDD dataset that has 4-layer annotation structure [25] can help the exploration of interactions between drivers and traffic participants from cause and effect labels.

3) *Trajectory Prediction*: The trajectory prediction of other transportation participants is also an essential task for autonomous driving, for which plenty of datasets have arisen.

The Argoverse dataset [39] can be used for 3D tracking and motion forecasting of vehicles that are in unusual scenarios such as at intersections, taking left or right turns, and changing to adjacent lanes in dense traffic. Moreover, the Argoverse 2

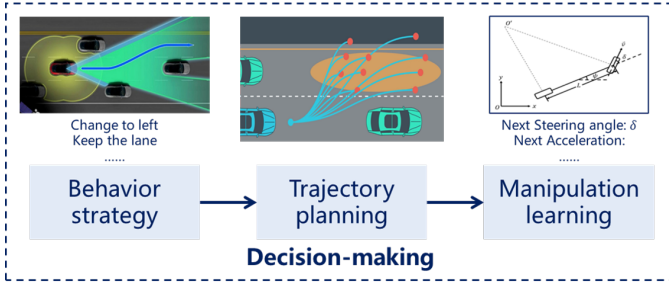


Fig. 10. Subdivisions of decision-making levels

Dataset [76] expands the categories of objects to 30 from 17 of the Argoverse 1, objects such as strollers and wheelchairs are included to provide richer samples for trajectory prediction algorithm design. The dataset PREVENTION [46] uses multi-modality of sensors, which ensures both redundancy and complementarity of information and also labels all the actions that appear in the scene. Such a dataset can be used to accelerate research on trajectory prediction under strong interaction.

### C. Datasets for Decision Learning

As powerful deep learning models spring up, learning-based methods show their promising potential to cope with complicated driving contexts and scenarios [78]. An efficient and appropriate dataset is a must for decision learning research. Decision learning can be divided into three submodules: behavioral strategy, trajectory planning, and manipulation control [79], as shown in Fig. 10. Datasets that can be applied to these three submodules are analyzed, compared, and concluded in the following.

1) *Behavioral Strategy*: As for data-driven algorithms, behavioral strategy learning is usually realized by classification machine learning methods. These methods provide discrete behaviors instead of continuous motions, thus learning autonomous driving decisions is based on the behavior annotation of vehicles in the datasets.

The D<sup>2</sup>CAV dataset [29] labels events into discrete behaviors including turns, lane shifts, deceleration, etc. The ratios of different behaviors are balanced in D<sup>2</sup>CAV to mimic reality. Comparatively, the 4-layer annotation structure in HDD dataset [25] and the triplet structure of road event in the ROAD dataset [38] provide sufficient antecedents and consequences to better express a traffic event, making it easier for the model to learn precise behavioral strategies and possess interpretability. As for modality, the Brain4Cars dataset [49] captures drivers' conditions together with road environment information, by which joint effect models could be developed to make behavior decisions.

To learn practical behavioral strategy, it is necessary for datasets to reflect the real situations in data collection. Besides, detailed annotation and processing of raw data could help algorithms to improve interpretability.

2) *Trajectory planning*: Trajectory planning is an essential step for autonomous driving. In related datasets, actual trajectories of ego vehicle could be directly used as the ground truth of model training processes.

The two most classical datasets used for trajectory planning model training is NGSIM [36] and KITTI [34]. However, most scenarios in these two datasets are regular and safe, failing to test the system's ability to drive on dangerous occasions. Therefore, in recent years, more and more research has been focusing on challenging and extreme conditions. The diverse scenarios and weathers contained in the DrivingStereo dataset [42] equips the trajectory planning research with robustness in extreme conditions. Another important characteristic of the complicated environment is interaction. The Waymo dataset [12] and the INTERACTION dataset [13] both provide abundant interactive scenarios to benefit relative research.

Datasets for trajectory planning are one of the most adequate self-driving datasets for decision-making. The scene richness and interactivity are becoming promising in dataset collection and algorithm learning.

3) *Manipulation learning*: The third level of decision learning is manipulation learning. As for current vehicle structures, the final control input are the wheel steering angle, throttle position, and brake position. To learn these basic decision strategies, onboard information needs to be recorded.

A standard data structure is like the DBNet dataset [26], which contains both onboard and environmental information for manipulation learning. Besides standard data, other internal data is also collected in related datasets to assist the algorithm training. DDD17 [27] and DDD20 [28] offer more complete onboard data, such as engine speed, odometer, etc. Similarly, A2D2 [8] records all data on the CAN bus involving even internal status like liquid pressure of brake valves. The data structure of datasets for manipulation control is rather simple. Extra types of data in addition to the fixed structure may help the algorithm learning procedure.

## VI. FUTURE TRENDS

It is obvious that datasets about drivers, environments, and vehicles are vital for developing decision-making systems of high-level autonomous vehicles. Many organizations have been devoted to collecting comprehensive data from various aspects. In this section, we summarize three future trends for potential improvements.

- 1) Collecting data from all the driver, environment, and the ego vehicle simultaneously. As aforementioned, human intelligence is an important reference for high-level autonomous vehicles [59] [60]. Models like Brain4Cars [49] have proved that drivers' attention and facial information helped to make better decisions. However, as introduced above, currently the human-factor-related data is isolated from those of vehicles and the environment. Without alignments on timelines and locations, it is difficult to use driver information as inputs for elaborate decision methods. Therefore, a dataset containing all driver, environment, and vehicle information can complement the vacancy, exploring more possibilities for AV decision-making by extraction of human's driving intelligence.
- 2) Containing data from more sensor sources. Most datasets are developed for certain tasks. To solve the target

problem in the most efficient way, the main concerns about the selection of sensors are costs and operational complexity. As a result, the data included usually just covers the requirements for the developers' research interest. This is suitable for specific fields of research, but not friendly for generalized applications. Third-party users of datasets demand various data such as Radar, event cameras, infrared, etc, especially more and more end- to-end methods have emerged. Consequently, a comprehensive and ideal dataset should contain as many commonly used sensors as possible to offer data redundancy and generalization.

- 3) Establishing a unified and compatible data format. At present, the data formats of datasets are different. Some datasets provide processed information in json, csv, txt, etc., some offer data in a self-defined struct, and others upload the raw sensor sources with annotations attached. As the complexity and size of parameter spaces of decision-making models grow larger and larger, the required amount of data also explodes to a new level. An ideal pattern is that the training and modeling process can be finished across various datasets with no or little specific adaption. However, data utilization across different datasets could not be realized at this stage because of the heterogeneous data formats. Accordingly, datasets with unified standard and interfaces will greatly promote the development of intelligent vehicles.

## VII. CONCLUSION

In this paper, we survey the state-of-the-art decision-making datasets of autonomous vehicles from four aspects: vehicle-related data, environment-related data, driver-related data, and their specific applications. After briefly reviewing the developing history of decision datasets, we present a detailed summary including many latest ones which have not been covered in previous surveys. We classify the data types, annotations, and features into specific categories. Based on comparison and feature analysis, limitations and potential applications on various decision-making of self-driving vehicles are concluded. By organizing the review in the categories of driver, vehicle, and environment datasets, this survey is dedicated to provide a clear and intuitive selection reference for researchers and developers of AV decision-making when the algorithm developing requires data. Also future trends are summarized to give advice on future dataset collection forms.

## REFERENCES

- [1] J. Wang, H. Huang, K. Li, and J. Li, "Towards the unified principles for level 5 autonomous vehicles," *Engineering*, vol. 7, no. 9, pp. 1313–1325, 2021.
- [2] W. Schwarting, J. Alonso-Mora, and D. Rus, "Planning and decision-making for autonomous vehicles," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, pp. 187–210, 2018.
- [3] B. Paden, M. Čáp, S. Z. Yong, D. Yershov, and E. Frazzoli, "A survey of motion planning and control techniques for self-driving urban vehicles," *IEEE Transactions on intelligent vehicles*, vol. 1, no. 1, pp. 33–55, 2016.
- [4] J. Duan, S. Eben Li, Y. Guan, Q. Sun, and B. Cheng, "Hierarchical reinforcement learning for self-driving decision-making without reliance on labelled driving data," *IET Intelligent Transport Systems*, vol. 14, no. 5, pp. 297–305, 2020.
- [5] C. Badue, R. Guidolini, R. V. Carneiro, P. Azevedo, V. B. Cardoso, A. Forechi, L. Jesus, R. Berriel, T. M. Paixao, F. Mutz *et al.*, "Self-driving cars: A survey," *Expert Systems with Applications*, vol. 165, p. 113816, 2021.
- [6] A. Sadat, M. Ren, A. Pokrovsky, Y.-C. Lin, E. Yumer, and R. Urtasun, "Jointly learnable behavior and trajectory planning for self-driving vehicles," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 3949–3956.
- [7] X. Liang, T. Wang, L. Yang, and E. Xing, "Cirl: Controllable imitative reinforcement learning for vision-based self-driving," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 584–599.
- [8] J. Geyer, Y. Kassahun, M. Mahmudi, X. Ricou, R. Durgesh, A. S. Chung, L. Hauswald, V. H. Pham, M. Mühlegg, S. Dorn *et al.*, "A2d2: Audi autonomous driving dataset," *arXiv preprint arXiv:2004.06320*, 2020.
- [9] A. Patil, S. Malla, H. Gang, and Y.-T. Chen, "The h3d dataset for full-surround 3d multi-object detection and tracking in crowded urban scenes," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 9552–9557.
- [10] S. Alletto, A. Palazzi, F. Solera, S. Calderara, and R. Cucchiara, "Dr (eye) ve: a dataset for attention-based tasks with applications to autonomous and assisted driving," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2016, pp. 54–60.
- [11] H. Caesar, J. Kabz, K. S. Tan, W. K. Fong, E. Wolff, A. Lang, L. Fletcher, O. Beijbom, and S. Omari, "nuPlan: A closed-loop ml-based planning benchmark for autonomous vehicles," *arXiv preprint arXiv:2106.11810v4*, 2021.
- [12] P. Sun, H. Kretschmar, X. Dotiwala, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine *et al.*, "Scalability in perception for autonomous driving: Waymo open dataset," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2446–2454.
- [13] W. Zhan, L. Sun, D. Wang, H. Shi, A. Clausse, M. Naumann, J. Kummerle, H. Königshof, C. Stillner, A. de La Fortelle *et al.*, "Interaction dataset: An international, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps," *arXiv preprint arXiv:1910.03088*, 2019.
- [14] 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," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, 2020, pp. 1–6.
- [15] H. Yin and C. Berger, "When to use what data set for your self-driving car algorithm: An overview of publicly available driving datasets," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2017, pp. 1–8.
- [16] Y. Kang, H. Yin, and C. Berger, "Test your self-driving algorithm: An overview of publicly available driving datasets and virtual testing environments," *IEEE Transactions on Intelligent Vehicles*, vol. 4, no. 2, pp. 171–185, 2019.
- [17] J. Guo, U. Kurup, and M. Shah, "Is it safe to drive? an overview of factors, metrics, and datasets for driveability assessment in autonomous driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 8, pp. 3135–3151, 2019.
- [18] D. Feng, C. Haase-Schütz, L. Rosenbaum, H. Hertlein, C. Glaeser, F. Timm, W. Wiesbeck, and K. Dietmayer, "Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1341–1360, 2020.
- [19] M. A. Kenk and M. Hassaballah, "Dawn: vehicle detection in adverse weather nature dataset," *arXiv preprint arXiv:2008.05402*, 2020.
- [20] Y. Pang, J. Cao, Y. Li, J. Xie, H. Sun, and J. Gong, "Tju-dhd: A diverse high-resolution dataset for object detection," *IEEE Transactions on Image Processing*, vol. 30, pp. 207–219, 2020.
- [21] Z. Yan, L. Sun, T. Krajník, and Y. Ruichek, "Eu long-term dataset with multiple sensors for autonomous driving," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 10 697–10 704.
- [22] Y. Choi, N. Kim, S. Hwang, K. Park, J. S. Yoon, K. An, and I. S. Kweon, "Kaist multi-spectral day/night data set for autonomous and assisted driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 934–948, 2018.
- [23] H. Xu, Y. Gao, F. Yu, and T. Darrell, "End-to-end learning of driving models from large-scale video datasets," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2174–2182.
- [24] Udacity. public driving dataset. [Online]. Available: <https://www.udacity.com/self-driving-car>



- [25] V. Ramanishka, Y.-T. Chen, T. Misu, and K. Saenko, "Toward driving scene understanding: A dataset for learning driver behavior and causal reasoning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7699–7707.
- [26] Y. Chen, J. Wang, J. Li, C. Lu, Z. Luo, H. Xue, and C. Wang, "Dbnet: A large-scale dataset for driving behavior learning," *Retrieved August*, vol. 9, p. 2020, 2019.
- [27] J. Binas, D. Neil, S.-C. Liu, and T. Delbruck, "Ddd17: End-to-end davis driving dataset," *arXiv preprint arXiv:1711.01458*, 2017.
- [28] Y. Hu, J. Binas, D. Neil, S.-C. Liu, and T. Delbruck, "Ddd20 end-to-end event camera driving dataset: Fusing frames and events with deep learning for improved steering prediction," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2020, pp. 1–6.
- [29] B. Toghi, D. Grover, M. Razzaghpour, R. Jain, R. Valiente, M. Zaman, G. Shah, and Y. P. Fallah, "A maneuver-based urban driving dataset and model for cooperative vehicle applications," in *2020 IEEE 3rd Connected and Automated Vehicles Symposium (CAVS)*. IEEE, 2020, pp. 1–6.
- [30] W. Liu, Q. Dong, P. Wang, G. Yang, L. Meng, Y. Song, Y. Shi, and Y. Xue, "A survey on autonomous driving datasets," in *2021 8th International Conference on Dependable Systems and Their Applications (DSA)*. IEEE, 2021, pp. 399–407.
- [31] X. Huang, X. Cheng, Q. Geng, B. Cao, D. Zhou, P. Wang, Y. Lin, and R. Yang, "The apolloscape dataset for autonomous driving," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 954–960.
- [32] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nusenes: A multimodal dataset for autonomous driving," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11 621–11 631.
- [33] S. Yogamani, C. Hughes, J. Horgan, G. Sistu, P. Varley, D. O'Dea, M. Uricár, S. Milz, M. Simon, K. Amende *et al.*, "Woodscape: A multi-task, multi-camera fisheye dataset for autonomous driving," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9308–9318.
- [34] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012, pp. 3354–3361.
- [35] J. Houston, G. Zuidhof, L. Bergamini, Y. Ye, L. Chen, A. Jain, S. Omari, V. Iglovikov, and P. Ondruska, "One thousand and one hours: Self-driving motion prediction dataset," *arXiv preprint arXiv:2006.14480*, 2020.
- [36] V. G. Kovvali, V. Alexiadis, and L. Zhang PE, "Video-based vehicle trajectory data collection," *Tech. Rep.*, 2007.
- [37] R. Krajewski, J. Bock, L. Kloeker, and L. Eckstein, "The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 2118–2125.
- [38] G. Singh, S. Akrigg, M. Di Maio, V. Fontana, R. J. Alitappeh, S. Saha, K. Jeddisaravi, F. Yousefi, J. Culley, T. Nicholson *et al.*, "Road: The road event awareness dataset for autonomous driving," *arXiv preprint arXiv:2102.11585*, 2021.
- [39] M.-F. Chang, J. Lambert, P. Sangkloy, J. Singh, S. Bak, A. Hartnett, D. Wang, P. Carr, S. Lucey, D. Ramanan *et al.*, "Argoverse: 3d tracking and forecasting with rich maps," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8748–8757.
- [40] A. Rasouli, I. Kotseruba, T. Kunic, and J. K. Tsotsos, "Pie: A large-scale dataset and models for pedestrian intention estimation and trajectory prediction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 6262–6271.
- [41] S. Malla, B. Dariush, and C. Choi, "Titan: Future forecast using action priors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11 186–11 196.
- [42] G. Yang, X. Song, C. Huang, Z. Deng, J. Shi, and B. Zhou, "Driving-stereo: A large-scale dataset for stereo matching in autonomous driving scenarios," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 899–908.
- [43] J. Mao, M. Niu, C. Jiang, H. Liang, J. Chen, X. Liang, Y. Li, C. Ye, W. Zhang, Z. Li *et al.*, "One million scenes for autonomous driving: Once dataset," *arXiv preprint arXiv:2106.11037*, 2021.
- [44] M. Pitropov, D. E. Garcia, J. Rebello, M. Smart, C. Wang, K. Czarnecki, and S. Waslander, "Canadian adverse driving conditions dataset," *The International Journal of Robotics Research*, vol. 40, no. 4-5, pp. 681–690, 2021.
- [45] F. Yu, H. Chen, X. Wang, W. Xian, Y. Chen, F. Liu, V. Madhavan, and T. Darrell, "Bdd100k: A diverse driving dataset for heterogeneous multitask learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2636–2645.
- [46] R. Izquierdo, A. Quintanar, I. Parra, D. Fernández-Llorca, and M. Sotelo, "The prevention dataset: a novel benchmark for prediction of vehicles intentions," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2019, pp. 3114–3121.
- [47] Q.-H. Pham, P. Sevestre, R. S. Pahwa, H. Zhan, C. H. Pang, Y. Chen, A. Mustafa, V. Chandrasekhar, and J. Lin, "A\* 3d dataset: Towards autonomous driving in challenging environments," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 2267–2273.
- [48] Y. Wang, S. Hou, and X. Wang, "Reinforcement learning-based bird-view automated vehicle control to avoid crossing traffic," *Computer-Aided Civil and Infrastructure Engineering*, vol. 36, no. 7, pp. 890–901, 2021.
- [49] A. Jain, H. S. Koppula, S. Soh, B. Raghavan, A. Singh, and A. Saxena, "Brain4cars: Car that knows before you do via sensory-fusion deep learning architecture," *arXiv preprint arXiv:1601.00740*, 2016.
- [50] A. Ligoicki, A. Jelinek, and L. Zalud, "Brno urban dataset-the new data for self-driving agents and mapping tasks," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 3284–3290.
- [51] K. Burnett, D. J. Yoon, Y. Wu, A. Z. Li, H. Zhang, S. Lu, J. Qian, W.-K. Tseng, A. Lambert, K. Y. Leung *et al.*, "Boreas: A multi-season autonomous driving dataset," *arXiv preprint arXiv:2203.10168*, 2022.
- [52] Y. Yao, X. Wang, M. Xu, Z. Pu, E. Atkins, and D. Crandall, "When, where, and what? a new dataset for anomaly detection in driving videos," *arXiv preprint arXiv:2004.03044*, 2020.
- [53] T. Chen, R. Tian, Y. Chen, J. Domeyer, H. Toyoda, R. Sherony, T. Jing, and Z. Ding, "Psi: A pedestrian behavior dataset for socially intelligent autonomous car," *arXiv preprint arXiv:2112.02604*, 2021.
- [54] U. Montanaro, S. Dixit, S. Fallah, M. Dianati, A. Stevens, D. Oxtoby, and A. Mouzakitis, "Towards connected autonomous driving: review of use-cases," *Vehicle system dynamics*, vol. 57, no. 6, pp. 779–814, 2019.
- [55] I. Kotseruba, A. Rasouli, and J. K. Tsotsos, "Joint attention in autonomous driving (jaad)," *arXiv preprint arXiv:1609.04741*, 2016.
- [56] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," *Advances in neural information processing systems*, vol. 30, 2017.
- [57] Z. Yang, Y. Chai, D. Anguelov, Y. Zhou, P. Sun, D. Erhan, S. Rafferty, and H. Kretschmar, "Surfelgan: Synthesizing realistic sensor data for autonomous driving," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11 118–11 127.
- [58] S. Ghosh, A. Dhall, G. Sharma, S. Gupta, and N. Sebe, "Speak2label: Using domain knowledge for creating a large scale driver gaze zone estimation dataset," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 2896–2905.
- [59] S. Xie, S. Chen, J. Zheng, M. Tomizuka, N. Zheng, and J. Wang, "From human driving to automated driving: What do we know about drivers?" *IEEE Transactions on Intelligent Transportation Systems*, 2021.
- [60] S. Chen, Z. Jian, Y. Huang, Y. Chen, Z. Zhou, and N. Zheng, "Autonomous driving: cognitive construction and situation understanding," *Science China Information Sciences*, vol. 62, no. 8, pp. 1–27, 2019.
- [61] O. Kopuklu, J. Zheng, H. Xu, and G. Rigoll, "Driver anomaly detection: A dataset and contrastive learning approach," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 91–100.
- [62] F. Lateef, M. Kas, and Y. Ruichek, "Saliency heat-map as visual attention for autonomous driving using generative adversarial network (gan)," *IEEE Transactions on Intelligent Transportation Systems*, 2021.
- [63] I. Jegham, A. B. Khalifa, I. Alouani, and M. A. Mahjoub, "A novel public dataset for multimodal multiview and multispectral driver distraction analysis: 3mdad," *Signal Processing: Image Communication*, vol. 88, p. 115960, 2020.
- [64] Z. Ling, Q. Xu-qiang, N. Tao, Y. Wei, and L. Yi-nong, "Driver cognitive loads based on multi-dimensional information feature analysis," *China Journal of Highway and Transport*, vol. 34, no. 4, p. 240, 2021.
- [65] D.-w. Xing, X.-s. Li, X.-l. Zheng, Y.-y. Ren, and Y. Ishiwatari, "Study on driver's preview time based on field tests," in *2017 4th International Conference on Transportation Information and Safety (ICTIS)*. IEEE, 2017, pp. 575–580.

- [66] Y. Shen, N. Wijayarathne, P. Sriram, A. Hasan, P. Du, and K. Driggs-Campbell, "Cocatt: A cognitive-conditioned driver attention dataset," *arXiv preprint arXiv:2111.10014*, 2021.
- [67] F. Vicente, Z. Huang, X. Xiong, F. De la Torre, W. Zhang, and D. Levi, "Driver gaze tracking and eyes off the road detection system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2014–2027, 2015.
- [68] M. Martin, A. Roitberg, M. Haurilet, M. Horne, S. Reiß, M. Voit, and R. Stiefelhagen, "Drive&act: A multi-modal dataset for fine-grained driver behavior recognition in autonomous vehicles," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 2801–2810.
- [69] N. Das, E. Ohn-Bar, and M. M. Trivedi, "On performance evaluation of driver hand detection algorithms: Challenges, dataset, and metrics," in *2015 IEEE 18th international conference on intelligent transportation systems*. IEEE, 2015, pp. 2953–2958.
- [70] R. Guesdon, C. Crispim-Junior, and L. Tougne, "Dripe: A dataset for human pose estimation in real-world driving settings," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 2865–2874.
- [71] B. Reimer, B. Mehler, J. Dobres, H. McAnulty, A. Mehler, D. Munger, and A. Rumpold, "Effects of an expert mode voice command system on task performance, glance behavior & driver physiology," in *Proceedings of the 6th international conference on automotive user interfaces and interactive vehicular applications*, 2014, pp. 1–9.
- [72] S. Xie, S. Chen, M. Tomizuka, N. Zheng, and J. Wang, "To develop human-like automated driving strategy based on cognitive construction: Appraisal and perspective," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2020, pp. 1–8.
- [73] S. Hecker, D. Dai, and L. Van Gool, "End-to-end learning of driving models with surround-view cameras and route planners," in *Proceedings of the european conference on computer vision (eccv)*, 2018, pp. 435–453.
- [74] S.-I. Oh and H.-B. Kang, "Object detection and classification by decision-level fusion for intelligent vehicle systems," *Sensors*, vol. 17, no. 1, p. 207, 2017.
- [75] M. Wigness, S. Eum, J. G. Rogers, D. Han, and H. Kwon, "A rugged dataset for autonomous navigation and visual perception in unstructured outdoor environments," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 5000–5007.
- [76] B. Wilson, W. Qi, T. Agarwal, J. Lambert, J. Singh, S. Khandelwal, B. Pan, R. Kumar, A. Hartnett, J. K. Pontes *et al.*, "Argoverse 2: Next generation datasets for self-driving perception and forecasting," in *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021.
- [77] E. Tivesten and M. Dozza, "Driving context influences drivers' decision to engage in visual-manual phone tasks: Evidence from a naturalistic driving study," *Journal of safety research*, vol. 53, pp. 87–96, 2015.
- [78] Y. Huang and Y. Chen, "Autonomous driving with deep learning: A survey of state-of-art technologies," *arXiv preprint arXiv:2006.06091*, 2020.
- [79] A. Furda and L. Vlacic, "Enabling safe autonomous driving in real-world city traffic using multiple criteria decision making," *IEEE Intelligent Transportation Systems Magazine*, vol. 3, no. 1, pp. 4–17, 2011.



**Zeyu Han** received the bachelor's degree in automotive engineering from School of Vehicle and Mobility, Tsinghua University, Beijing, China, in 2021. He is currently pursuing the Ph.D. degree in mechanical engineering with School of Vehicle and Mobility, Tsinghua University, Beijing, China. His research interests include risk evaluation, environment understanding of intelligent vehicles.



**Yining Xing** received the bachelor's degree in automotive engineering from School of Vehicle and Mobility, Tsinghua University, Beijing, China, in 2022. He is currently pursuing the Ph.D. degree in mechanical engineering with School of Vehicle and Mobility, Tsinghua University, Beijing, China. His research centered on risk evaluation, trajectory prediction and decision-making process of intelligent vehicles.



**Shaobing Xu** received his Ph.D. degree in Mechanical Engineering from Tsinghua University, Beijing, China, in 2016. He is currently an assistant professor with the School of Vehicle and Mobility at Tsinghua University, Beijing, China. He was an assistant research scientist and postdoctoral researcher with the Department of Mechanical Engineering and Mcity at the University of Michigan, Ann Arbor. His research focuses on vehicle motion control, decision making, and path planning for autonomous vehicles. He was a recipient of the outstanding Ph.D. dissertation award of Tsinghua University and the Best Paper Award of AVEC'2018.



**Jianqiang Wang** received the B. Tech. and M.S. degrees from Jilin University of Technology, Changchun, China, in 1994 and 1997, respectively, and Ph.D. degree from Jilin University, Changchun, in 2002. He is currently a Professor of School of Vehicle and Mobility, Tsinghua University, Beijing, China. He has authored over 150 papers and is a co-inventor of over 140 patent applications. He was involved in over 10 sponsored projects. His active research interests include intelligent vehicles, driving assistance systems, and driver behavior. He was a recipient of the Best Paper Award in the 2014 IEEE Intelligent Vehicle Symposium, the Best Paper Award in the 14th ITS Asia Pacific Forum, the Best Paper Award in 2017 IEEE Intelligent Vehicle Symposium, the Changjiang Scholar Program Professor in 2017, Distinguished Young Scientists of NSF China in 2016, and New Century Excellent Talents in 2008.



**Yuning Wang** received the bachelor's degree in automotive engineering from School of Vehicle and Mobility, Tsinghua University, Beijing, China, in 2020. He is currently pursuing the Ph.D. degree in mechanical engineering with School of Vehicle and Mobility, Tsinghua University, Beijing, China. His research centered on risk evaluation, trajectory prediction and decision-making process of intelligent vehicles.