

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/370114593>

Multi-Modal Fusion Technology Based on Vehicle Information: A Survey

Article in IEEE Transactions on Intelligent Vehicles · January 2023

DOI: 10.1109/TIV.2023.3268051

CITATIONS

0

READS

31

7 authors, including:



Xinyu Zhang

Tsinghua University

89 PUBLICATIONS 865 CITATIONS

SEE PROFILE



Jun Li

Chinese Academy of Agricultural Sciences

63 PUBLICATIONS 126 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Cloud Model Approach for Lateral Control of Intelligent Vehicle Systems [View project](#)



A study on key technologies of unmanned driving [View project](#)

Multi-modal Fusion Technology Based on Vehicle Information: A Survey

Xinyu Zhang, Yan Gong*, Jianli Lu, Jiayi Wu, Zhiwei Li, Dafeng Jin, and Jun Li

Abstract—Multi-modal fusion is a basic task of autonomous driving system perception, which has attracted many scholars' attention in recent years. The current multi-modal fusion methods mainly focus on camera data and LiDAR data, but pay little attention to the kinematic information provided by the sensors of the vehicle, such as acceleration, vehicle speed, angle of rotation. These information are not affected by complex external scenes, so it is more robust and reliable. In this paper, we introduce the existing application fields of vehicle information and the research progress of related methods, as well as the multi-modal fusion methods based on information. We also introduced the relevant information of the vehicle information dataset in detail to facilitate the research as soon as possible. In addition, new future ideas of multi-modal fusion technology for autonomous driving tasks are proposed to promote the further utilization of vehicle information.

Index Terms—Multi-modal fusion, perception, autonomous driving.

I. INTRODUCTION

IN the early research such as semantic segmentation, object detection, and other fields, most of the research directions are based on sensors, such as camera and LiDAR [1], [2]. The camera has the characteristics of a high frame rate and high resolution, which provides accurate information in normal environments, such as good weather and lighting. However, in bright light or dark conditions, such as severe camera exposure or thunderstorm weather, the camera does not provide much information as in normal environments [3]. LiDAR has a strong ability to acquire three-dimensional information and provide accurate distance measurement, which is not affected by lighting conditions, but greatly affected by weather, e.g., rainy day and foggy day [3]. Each sensor has its advantages and disadvantages, so they are inevitably limited by sensors.

Due to the limitations of a single sensor, multi-sensor fusion is extensively studied to reduce the effects of the hardware itself and the weather [4]. In recent years, multi-modal fusion methods for autonomous driving perception tasks have

developed rapidly [5]. There are many types of sensors used in autonomous driving, including monocular vision, stereo vision, LiDAR, vehicle dynamic information obtained from an inertial measurement unit used to measure the three-axis attitude angle (or angular rate) and acceleration of an object, and its own position information obtained using positioning and maps [6]. In the prediction process, single-modal information only reflects the value of information from a single point of view, while multi-modal information combines information from two or more modalities for prediction, reflecting the value of different information and making the information complementary, making up for the shortcomings of a single sensor [7], [8]. Therefore, recent works use multi-modal fusion technology to fuse the data of different modalities to play their respective advantages and improve the precision accuracy, and robustness of the results.

At present, in the field of vehicle automatic driving research, most methods conduct multi-modal fusion research on 3D point cloud captured by LiDAR and images captured by the camera, such as fusion camera-LiDAR for 3D detection and 3D trajectory generation [9], [10], [11], [12], [13], [14], [15], [16], [17]. The research on the use of vehicle information is still lacking. However, the LiDARs and cameras are sensors external to the vehicle in most tasks, which are affected by factors such as sensor cost, sensor hardware limitations, and weather. Some information from the car itself, such as vehicle velocity, acceleration, turning angle, and engine current, are accurately obtained in real-time from the vehicle itself and smart portable devices [18], which are no need to install additional sensors on the vehicle to obtain this vehicle information, greatly reducing sensor cost. However, vehicle information, which is not affected by ambient weather and can be obtained in real-time, is not widely used.

In this paper, we present various aspects of vehicle information that have not been fully utilized. First, we mainly analyze the existing research on vehicle steering angle and speed. Appropriate fusion technology is required for the application of different information, so we describe some fusion technologies [19]. Then some datasets containing vehicle information are introduced to provide sufficient data sources for research. Afterward, we describe some challenges in applying vehicle information. Finally, we introduce future research directions in detail and prove that vehicle information has an important auxiliary role for existing task research through experiments. For instance, the steering angle can be used for steering control or lane recognition during the turning process of autonomous vehicles, which not only avoids the influence of strong light but also gives the task prior data. In the process of object tracking,

This work was supported in part by the National High Technology Research and Development Program of China under Grant 2018YFE0204300, in part by the National Natural Science Foundation of China under Grant 62273198 and Grant U1964203. *Corresponding author: Yan Gong.*

Xinyu Zhang is with the School of Transportation Science and Engineering, Beihang University, Beijing 100191, China, and also with the State Key Laboratory of Automotive Safety and Energy, Tsinghua University, Beijing 100084, China. Yan Gong*, Jianli Lu, Jiayi Wu, Dafeng Jin, and Jun Li are with the State Key Laboratory of Automotive Safety and Energy, and the School of Vehicle and Mobility, Tsinghua University, Beijing, 100084 China. Zhiwei Li is with Beijing University of Chemical Technology, Beijing, 100029, China. E-mail: xyzhang@tsinghua.edu.cn, gongyan2020@foxmail.com, lujianli364@163.com, wujiayi96@outlook.com, 2022500066@buct.edu.cn, jindf@tsinghua.edu.cn, lijun19580326@126.com.

the change degree of the object position can be reflected by the vehicle speed. The faster the vehicle speed is, the greater the change degree of the object is. Therefore, vehicle speed has a certain auxiliary effect on object tracking. The speed can also be used to classify driving behaviors and estimate the expected mileage of the vehicle. We believe that the vehicle information contained in the vehicle sensors, such as speed, steering angle, brake, and accelerator, is very essential for many tasks of automatic driving. Based on this information, extensive research can be carried out in the above areas. In view of this, we have conducted an extensive investigation on the research direction of multi-modal fusion based on vehicle information and put forward the existing challenges and future research prospects. Our contributions can be summarized as follows:

- As far as we know, our paper is the first investigation focusing on the multi-modal fusion of vehicle information, including steering angle prediction, multiple auxiliary tasks, lane change prediction, driving behavior classification, vehicle speed prediction, trajectory prediction, etc.
- We also introduced and tested two future research directions of vehicle information fusion to prove the role of vehicle information, such as lane detection combined with steering angle and images, and driving behavior classification combined with vehicle speed and images.
- Our paper organizes different datasets related to vehicle information, and proposes some problems in the research field of vehicle information fusion. In addition, we introduce the fusion method of vehicle information and other information for different data spaces, which provides great convenience for other scholars to quickly start the field.

The rest of the organization structure of this paper is as follows: Section II reviews the research status based on vehicle information. Section III introduces some multi-modal fusion technologies. Section IV introduces the mainstream datasets containing vehicle information. Then, we describe the existing challenges and research prospects in Section V and Section VI. Finally, Section VII is the conclusion.

II. REPRESENT METHODS

The following mainly introduces the application status and representative methods of steering angles and vehicle speeds, the development over time is shown in Fig. 1. In terms of steering angle, we recommend some research on steering angle prediction, multiple auxiliary tasks, lane change prediction, and perception under noise image data. In terms of speed, we present some research on driving style classification, speed prediction, trajectory prediction, mileage, time, energy consumption forecasts, sideslip angle, and vehicle deceleration prediction.

A. Steering angle

1) *Steering angle prediction:* In automatic driving technology, the vehicle steering angle is very important for controlling the driving direction of the vehicle. Autonomous vehicles need to accurately control the steering angle of the vehicle to

maintain safety. To achieve this, autonomous vehicles usually use sensors, image recognition technology, and road maps to collect road information, and use this information to decide how to control the steering angle. For example, when the vehicle enters a curve, the automatic driving technology can recognize the road characteristics and ensure that the vehicle enters the curve stably by adjusting the steering angle of the vehicle. Similarly, automatic driving technology can also achieve vehicle following while maintaining a safe distance between vehicles. In the past few years, convolutional neural networks (CNNs) and other deep learning techniques have made significant strides in autonomous driving, particularly in perception and control. These advancements have a significant impact on the ability of autonomous vehicles to accurately perceive their environment and make decisions based on that information. This section focuses on the combination of vehicle steering angle information and other sensor information based on deep learning to assist automatic driving.

Bojarski et al. [20] proposed an end-to-end network for steering angle prediction. The model directly uses the image information taken from the camera to predict, and map to the steering angle command. CNN automatically learns the road characteristics and only uses the steering angle data obtained by human manipulation of the vehicle as the training signal. This method fully utilizes the steering information captured by the vehicle's sensors. The steering control of conventional vehicles needs to clarify the location of the lane line, while the end-to-end system does not care whether there is a clear and accurate lane line. Some roads lacking driving instruction information, such as forests, Unpaved roads, or more complex and rugged mountain roads have achieved a good steering control effect. Besides, Rausch et al. [21] proposed an end-to-end system to decide the steering control of vehicles. Compared with other traditional auto-drive systems, it is composed of a unified framework, which eliminates the intermediate steps such as target detection and path planning and directly outputs the control strategy of vehicles. It is worth mentioning that they used the simulator as the experimental environment, and the input data is to use the camera system of the vehicle they are driving under the simulator to take images and record the steering wheel angle corresponding to each frame. The goal is to establish the mapping relationship between the image and the steering angle. Santana et al. [22] proposed a method based on virtual simulation, which used virtual simulation technology to generate a large number of road scenes with different complexities, as well as vehicle trajectories and sensor data in these scenes. These data are then used to train a vehicle steering prediction model that utilizes a convolutional neural network to learn the mapping between input image sequences and output steering angles. The results show that this virtual simulation-based approach can significantly improve the accuracy and generalization ability of the vehicle steering prediction model. Compared with traditional methods, it can not only effectively reduce the time and cost of data collection, but also effectively solve the problems of data collection, such as the limitations of factors such as weather, time, and location.

In addition to problems with vehicle steering during normal

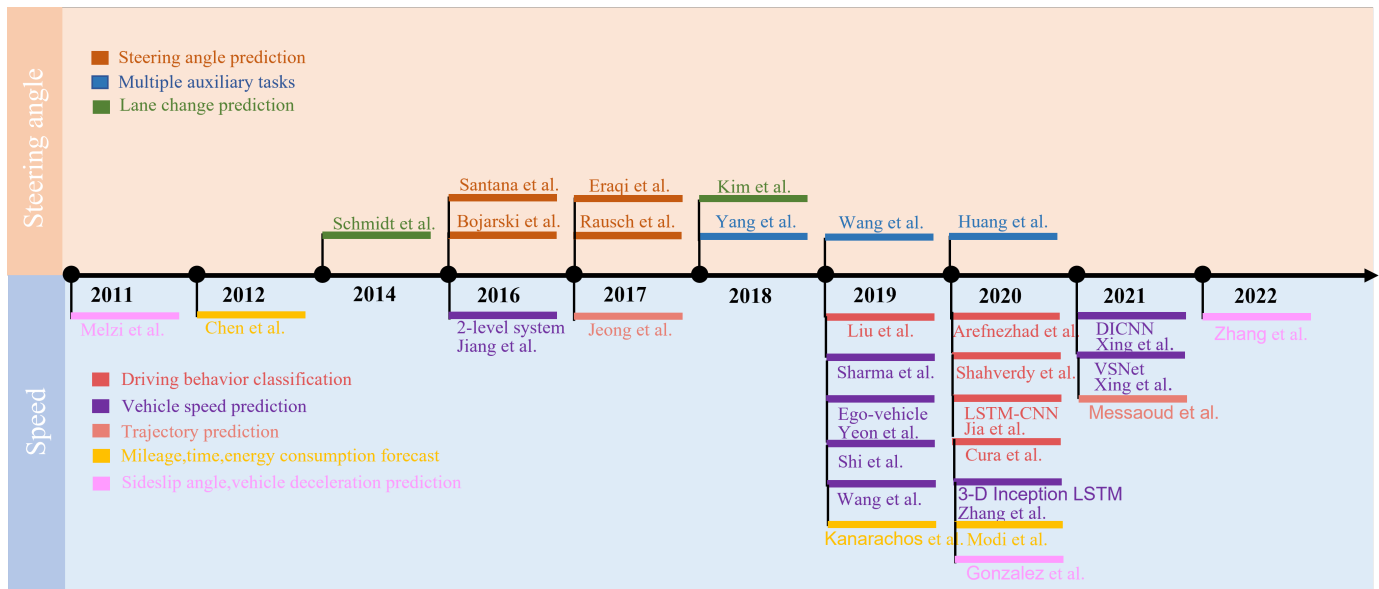


Fig. 1. The development of multi-modal fusion methods based on vehicle information.

driving, when the vehicle approaches an intersection, it may happen that the control commands are not clear, or require more specific instructions to guide the vehicle, but are not efficient. Aiming at the problem of autonomous driving of vehicles under a given path. Eraqi et al. [23] proposed a new method aimed at considering the time dependence of input data, that is, in addition to the current input image, the previous motion of the vehicle and the motion of surrounding objects are also considered. The proposed method uses a CNN to extract visual features from an input image and an LSTM network to capture the temporal dependencies of the input data. CNNs and LSTMs are integrated using a novel architecture called a "ConvLSTM" network. The method is evaluated on the Udacity autonomous vehicle dataset and shows that it outperforms state-of-the-art methods in both the accuracy and smoothness of steering angle prediction. They also provide a visualization technique to understand how the model is making predictions, which is useful for debugging and improving the model.

2) *Multiple auxiliary tasks:* During driving, purely relying on visual input to judge steering will be inaccurate, and vehicle information can improve the estimation of vehicle behavior and ensure that driving obeys certain physical rules. For example, a U-turn at 10 mph versus 30 mph would have a different turn angle and control strategy, even though visual observations would appear to be similar. Inferring a vehicle's motion state purely from images in a changing scene is ambiguous, but sensors on the vehicle can provide more accurate information, such as current speed and steering angle.

Furthermore, auxiliary tasks applied in vehicle information fusion can help to comprehend the environment surrounding the vehicle. For instance, deep learning can be employed to train drivers to make decisions based on salient information and comprehend various driving scenarios. Nevertheless, understanding the relationship between distinct features and driving behaviors could be challenging for end-to-end

models, particularly when it involves critical factors such as pedestrians, traffic lights, and drivable areas. To tackle this challenge, multiple auxiliary task learning can be utilized to focus on essential regions and enhance the overall understanding of driving scenes. Yang et al. [24] proposed a method for end-to-end vehicle control using a multimodal architecture is proposed, combining information from multiple sensors, including cameras and LiDAR, to better perceive the environment and make accurate control decisions. A multi-task learning framework is also introduced that enables the model to simultaneously learn to perform multiple related tasks, including steering angle prediction, and accelerator and brake control. Accurate and safe decisions can be made based on information from multiple sensors. Wang et al. [25] proposed an end-to-end learning method based on multiple auxiliary tasks for better autonomous driving of vehicles. The proposed method is mainly based on deep neural networks, and more information can be learned by training multiple related auxiliary tasks (such as lane line tracking, front vehicle tracking, traffic signal recognition, etc.) in the vehicle's front camera images. Rich functions realize more precise automatic driving. At the same time, a multi-level perception model is also used to process different information flows, and finally realize the integration of vehicle driving decisions. The experimental results show that the method has achieved good performance in both simulation and real environments, and can effectively realize the automatic driving of vehicles, which has high practical value. Huang et al. [26] proposed a new end-to-end autonomous driving method that integrates data from multiple sensors such as vision, radar, and LiDAR to achieve more comprehensive environmental perception and decision-making. The method is based on a structured deep learning method called factor graph-based deep learning, which is capable of end-to-end learning and decision-making on multi-modal data, modeling and inferring the behavior of traffic participants, and through analysis and inference The behavior of different

traffic participants improves the understanding and decision-making ability of the automatic driving system for complex scenes. In addition, multi-task learning is also possible by combining multiple technical means with multiple modality data. For example, semantic segmentation can identify the location of road boundaries and the location of surrounding vehicles on the road, which makes it easier for vehicles to perceive the stop/go or steering behavior of surrounding vehicles and assist vehicle control, significantly reducing the implicit difficulty of learning everything. Optical flow [27] can recognize the motion of objects. Transfer learning [28] exploits the ability to share common features between tasks, while LSTM [29] can be used to extract temporal information. Combining them into a network and workflow can compensate for the limitations of a single task, leading to the development of smarter autonomous driving systems.

3) *Lane change prediction*: The application of vehicle steering angle in lane change prediction is an important part of autonomous driving technology. The following discusses the application of vehicle steering angle in lane change prediction.

The lane change prediction model predicts the lane change intention by analyzing the vehicle's speed, acceleration, steering angle, and other information. By continuously monitoring the steering angle when the vehicle's steering angle changes sharply, the automated driving system can predict when a lane change is likely to occur and make the necessary adjustments to ensure the vehicle can safely change lanes during the maneuver. The control of vehicle steering angle is also a key factor in lane change prediction. When the vehicle is ready to change lanes, the automatic driving system needs to consider factors such as vehicle speed, acceleration, and body center of gravity to accurately calculate and control the vehicle steering angle to achieve accurate and safe lane change operations. At the same time, the automatic driving system needs to feed back the steering angle information to adjust the lane change operation. This feedback information can help the vehicle drive system to calculate the vehicle motion information more accurately and ensure the safety and stability of the lane change operation.

The main scenario where vehicle steering information is used is lane changes [30]. A German study reported that the probability of changing lanes on urban roads is 55%, and the use rate of turn signals on highways is 75%. In early predictions, steering wheel angle, as a directly measurable vehicle parameter, appeared to be a promising predictor of lane changes. Some researchers have proposed mathematical models of steering angles to help predict lane change maneuvers. Kim et al. [31] proposed a vehicle lane change prediction method, which collects sensor data such as steering angle and acceleration of the vehicle, and extracts the relevant features of these data. Using support vector machines and decision trees, a classification model is trained to identify vehicle lane change intention. The model first uses steering angle and acceleration data to predict whether to change lanes and then predicts which lane the vehicle will change to based on other features such as speed and distance. The authors conducted multiple trials on real roads to evaluate the performance of the model. The results show that the proposed model can accurately

identify and predict how vehicles change lanes. Schmidt et al. [32] proposed a mathematical model to predict how vehicles change lanes, which is based on the angle of the steering wheel. Steering wheel angle is an important indicator of driver behavior, which can be used to predict lane change behavior. First, models the steering dynamics of the vehicle, considering the influence of steering angle, speed, vehicle mass, and other factors on the vehicle's steering behavior. Then, they used a logistic regression-based classifier to classify the intent of vehicle change lanes and took the steering wheel angle as the input variable of the classifier. Finally, the experiment verifies the effectiveness of the model. The research results can be used in the development of intelligent transportation systems, the model can provide an important reference for the study of vehicle steering behavior, and provide a new idea for driver behavior modeling.

4) *Perception under noisy image*: Numerous studies have established that visual noise can significantly impair a vehicle's perception abilities. For instance, the driverless car accident involving a Tesla vehicle was attributed to a malfunction of the perception module under intense illumination [23]. During the experiment, due to the sunlight, the overexposure of the road surface on the image, and the lens flare caused by direct sunlight on the camera sensor will affect the judgment of the automatic driving system on the road conditions and the corresponding steering control. Moreover, the varying lighting conditions during testing, caused by the changing sun position, could also affect the comparison of different model outcomes [33]. As is shown in Fig. 2.

Several studies have shown that even without image data, the steering behavior of a vehicle can be predicted from other sensor data. The inertial measurement unit (IMU) can measure information such as the acceleration velocity, and the global positioning system (GPS) can be used to measure information such as the position and speed of the vehicle, which can be used to predict the steering behavior of the vehicle. Several studies exploit these data for feature extraction and classification to predict vehicle steering behavior. In addition, the vehicle dynamics model can be used to describe the state of motion of the vehicle, including information such as vehicle speed, acceleration, steering wheel angle, etc. This method can be used to predict when there is a lack of sensor data.

B. Speeds

1) *Driving behavior classification*: At present, drivers' abnormal driving behavior is the main cause of traffic accidents, such as drunk driving, fatigue driving, and aggressive driving [34]. If there are some methods to predict driving behavior, it can remind drivers and reduce the occurrence of accidents. Some methods of monitoring driver behavior directly monitor drivers' faces and bodies, but people can subjectively control their expressions, so this method may be deceptive and invade people's privacy. Driving behavior predicted by vehicle information cannot be subjectively changed by people, so the driving behavior predicted by vehicle information is more accurate and convincing [35]. Besides classifying driving behavior can help prevent traffic accidents, it also provides

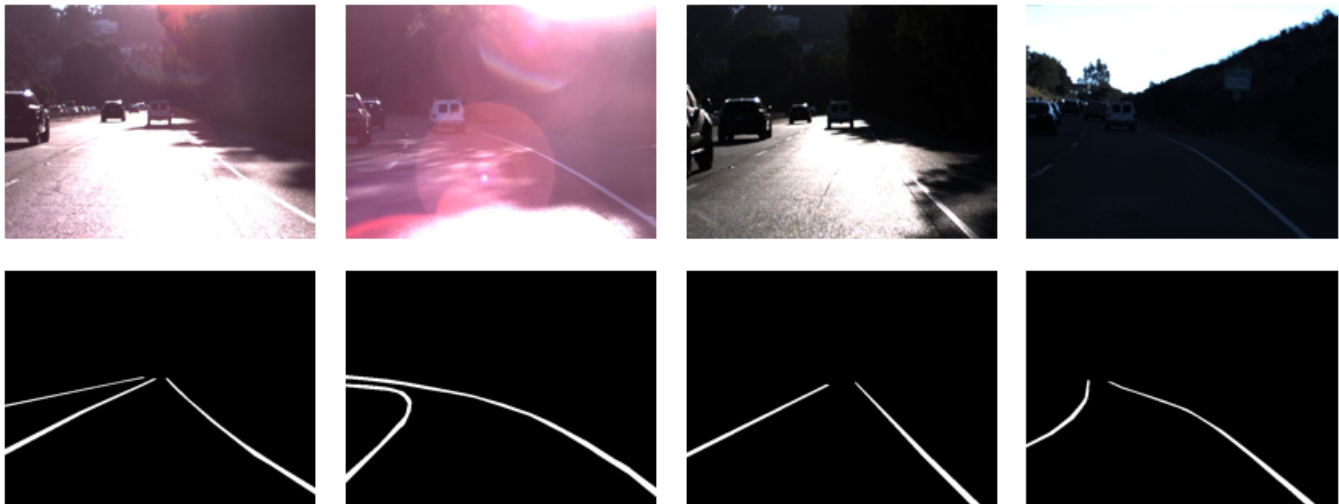


Fig. 2. Visual information is affected by noise. Fusion of the extracted lane line information and steering information to assist vehicle control or to perform lane line detection and segmentation.

some support for optimizing fuel consumption, because different driving behavior have different fuel consumption. For example, aggressive driving often consumes more fuel than normal driving [36]. At present, some insurance companies also have to classify drivers' usual driving behavior, in order to divide the insurance cost by different standards, if the driver is usually aggressive driving, then his cost will be higher than the average person [37].

At the beginning of the development of using vehicle information to classify driving behavior, fewer signals were used. Sayed et al. [38] proposed that only according to steering angles, based on the artificial neural network to predict whether the driver is drowsy if steering angles have not changed much, there is no sharp change, then the driver is not awake. But in fact, there is no sufficient reason to think that the driver is sleepy. For example, the driving journey does not need large steering, the signal used in this method is single, and the accuracy is only 90%, which needs to be further improved.

In the wake of the development of neural networks and the research of classifying driving behavior, there are more and more methods. Arefnezhad et al. [39] proposed a method that uses five signals, yaw rate, lateral acceleration, lateral deviation from the center of the road, steering wheel angle, and steering wheel angular velocity to classify whether the driver is not drowsy, moderately sleepy, or extremely sleepy, and the drowsiness level is subdivided. Arefnezhad et al. [39] compared with Sayed et al. [38], they made full use of the vehicle information, classified from different angles, and considered it more comprehensively.

Compared with the previous two methods, Shahverdy et al. [40] were not limited to classifying drivers' sleepy state, they further predicted whether the driver is normal driving, aggressive driving, distracted driving, drunk driving, and the behavior of classification is more diverse. Most methods directly input vehicle information into neural networks without association with images. The novelty of this paper is that the acceleration,

gravity, throttle, speed, and Revolutions Per Minute (RPM) information are converted into images through the recurrence plot technology, which can be applied to various image-based neural networks. The network architecture is shown in Fig. 3.

When the driver performs some abnormal operations (such as sudden braking and acceleration), the acceleration and deceleration will change sharply, and the acceleration and deceleration change trends of different driving behaviors are different, so Jia et al. [41] innovated from the extreme value of acceleration and deceleration of vehicle information, and classify different driving behaviors by detecting the extreme value of acceleration and deceleration.

In terms of the automatic transmission system of the car, if the gear is automatically changed according to driver's driving purpose, it is able to improve the efficiency of shifting, reduce fuel consumption, and save resources to protect the environment. Liu et al. [42] proposed the use of vehicle speeds, opening degree of the accelerator pedal and brake pedal force to identify driver intent, and developed a shift strategy based on driver intention, which showed that this strategy is able to reduce the number of shift gears, thereby reducing fuel consumption.

Many methods were based on the hybrid network of 1D-CNN and LSTM or only on LSTM, while Cura et al. [43] found that the performance of only using 1D-CNN is equivalent to that of only using LSTM, and even the ability to distinguish aggressive driving is stronger.

2) *Vehicle speeds prediction*: With the increase in population density, most of the air pollution in urban areas is caused by motor vehicle exhaust emissions. At present, many studies have optimized powertrain control by predicting vehicle status, so as to reduce fuel consumption, meet strict emission regulations, and reduce environmental pollution. In addition, in terms of autonomous driving, the biggest challenge facing cars is autonomous lateral and longitudinal control, including steering and speed control. Therefore, it is particularly important to accurately predict speeds.

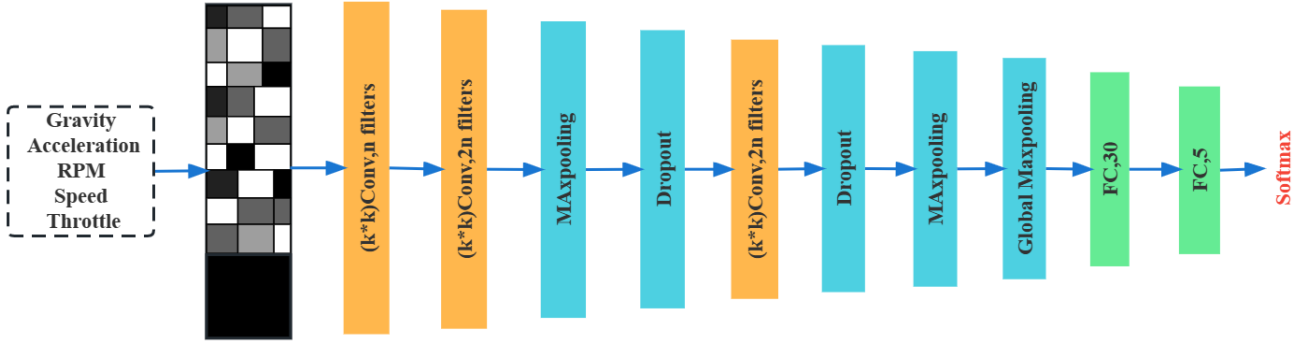


Fig. 3. Network structure for judging driving state based on recurrence plot technique.

In terms of driving habits, people usually slow down when turning corners and accelerate appropriately when going straight. Therefore, this behavior can continue to be imitated in autonomous vehicles, predicting whether the speed should increase or decrease according to the curvature of the road ahead [44], so as to help the vehicle have better control.

Many methods of predicting vehicle speeds use the information of the vehicle in a certain period of time to predict speeds in the future. They only consider the information of the vehicle itself and do not consider the factors that affect the speed control on the side such as the road traffic conditions and the distance from the surrounding vehicles. The influence of this external environment increases the pressure on speed prediction. Jiang et al. [45] first used a neural network model to get the average traffic speeds of the road section based on the previous traffic data, and then established the relationship between individual vehicle speeds and average traffic speeds through Hidden Markov Model (HMM), so as to predict the speed. Yeon et al. [46] broke the limitation of predicting vehicle speeds under specific conditions. They not only used the internal information of the vehicle such as speed, acceleration, and engine speed to predict the vehicle speed but also used information such as the relative speeds and distance with the vehicle in front, as well as the position of the vehicle. Through the internal and external information in 30 seconds, they jointly predicted vehicle speeds in the next 15 seconds. These methods fully consider the external environment information and consider the factors that affect speed prediction more comprehensively.

What data to use to predict speeds needs sufficient reasons. When using too little data, it is not enough to reflect the vehicle state. When using too much data, irrelevant data will interfere with the model. Xing et al. [47] collected a variety of experimental data, mining the deep value of the data, such as the VCU speed and VBox speed, so that the measured speed is more accurate, and Xing et al. [47] added the information of the opening of the driving pedal and the opening of the brake pedal to the model, fully considering the intention of the driver to accelerate and decelerate. Xing et al. [48] also made a full selection on the input signal. In most studies, people choose the input signal to predict vehicle speeds according

to their own feelings, without too much consideration for the quality of the signal, which is subjective. Xing et al. [48] used Pearson correlation to calculate the relationship between various vehicle signals and vehicle speeds, and selected the 15 signals with the highest scores as the input of the network, ensuring the data quality and making the predicted vehicle speed more accurate to a certain extent.

Most of the above methods are based on fuel vehicles. In terms of electric vehicles, electric vehicles have gradually entered people's lives because of their advantages such as no pollution, low noise, and no need for frequent maintenance. Zhang et al. [49] [50] broke the limitation of using only vehicle information to predict vehicle speeds, used image and vehicle speeds fusion to jointly predict vehicle speeds, and further optimized the energy management strategy of electric vehicles on this basis.

Most studies on speed prediction are carried out under normal geographical conditions. Affected by air pressure, temperature, and humidity, oxygen content, etc., different altitudes have different effects on vehicle dynamic performance, and people's driving habits are also different. Therefore, specific models are needed to predict vehicle speeds in high-altitude areas. Shi et al. [51] considered the influence of altitude on the effect of speed prediction, and established a speed prediction neural network model in high-altitude areas.

Many methods based on vehicle information are not connected with perception tasks. Wang et al. [25] not only predicted steering angles and speeds at the same time, but also shared the predicted network on semantic segmentation and object detection tasks. It can efficiently carry out three tasks at the same time, making a breakthrough in-vehicle information and perception tasks. The network architecture is shown in Fig. 4.

3) *Trajectory prediction*: If the trajectory prediction of vehicles on the road is performed in real time on the traffic monitoring system, or a set of trajectory prediction system is applied to vehicles, then when the trajectory between vehicles will conflict, the trajectory prediction system can reduce a large number of traffic accidents by giving owners a certain degree of reminder [52].

On the highway, the speeds of vehicle driving are faster, and the lane change process is more dangerous, if the model

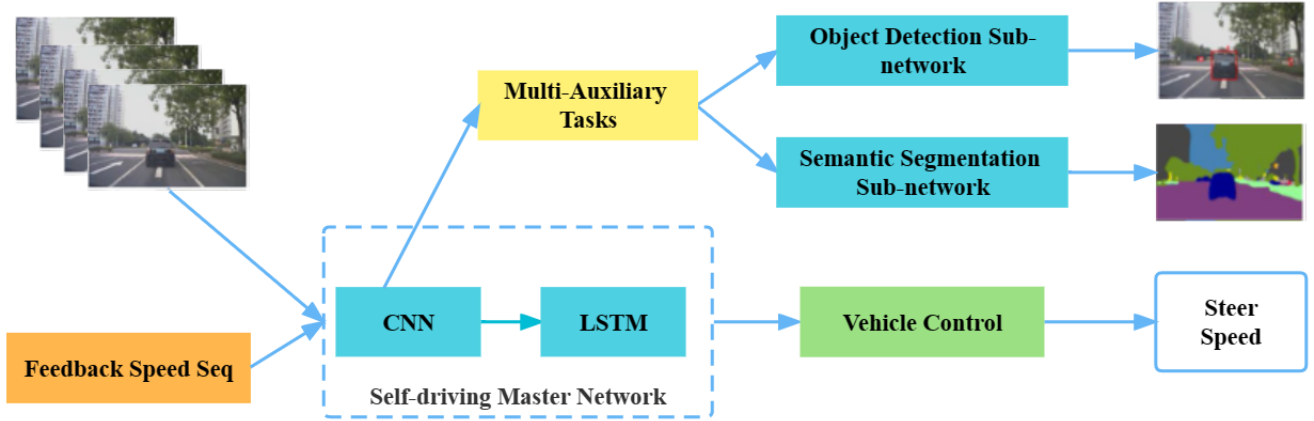


Fig. 4. Multi-task end-to-end network architecture.

can predict in time whether the vehicle is safe during the lane change, it can affect driving behavior and avoid accidents. Tomar et al. [53] proposed a lane change model, which predicts the next state during lane changing through the relative speed between vehicles, safe distance, whether the vehicle is accelerating or decelerating, and the current vehicle state (collision, near collision, absolutely safe and safe), so as to ensure the safety during lane changing.

Traditional prediction trajectory methods use physics-based motion models, and while their results are accurate in the short term, if they exceed one second, the results are less reliable and cannot be further referenced. Therefore, Jeong et al. [54] used the advantages of neural network to fuse the vehicle information such as vehicle speeds, acceleration, steering, and road conditions, and predicted the position of the vehicle in 0-4 seconds, which provided a reference for the auto drive system and road safety.

4) *Mileage, time, energy consumption forecasts*: Although the fuel consumption meter of vehicles can now display the instantaneous fuel consumption, average fuel consumption, and mileage of vehicles, the mileage of vehicles is calculated based on the remaining fuel in vehicles' fuel tanks divided by the average fuel consumption of vehicles. The average fuel consumption is calculated according to the mileage and fuel consumed over a period of time, and the real-time performance is not high. Moreover, if both the remaining time and the remaining mileage can be displayed on the vehicle display screen according to the real-time information, it is more informative, so that the driver can arrange the itinerary according to a variety of information. Therefore, using parameters that affect fuel usages such as vehicle speeds and road slope to predict the remaining mileage and time can provide greater help to the driver. Evaluating the real-time energy consumption of vehicles can monitor road pollution and provide data support for pollution reduction technology.

In terms of fuel vehicles, Chen et al. [55] proposed a method of predicting the remaining mileage and time of the vehicle based on the comprehensive vehicle and road information, not only based on the fuel capacity of the vehicle itself, the current

speed, the engine speed, the weight of the vehicle and other information, but also based on the road slope and other road information, through multi-faceted information prediction is more real-time and accurate than only the reference to the average fuel consumption. In terms of electric vehicles, the real-time energy consumption is estimated by vehicle speeds, road altitude, etc., and the proposed model has the lowest error compared with existing technologies [56].

In order to mitigate the pollution caused by road traffic, effective environmental optimization technologies alone may not be sufficient. Real-time and convenient monitoring methods are also essential, as accurate monitoring can enable more effective optimization strategies. Kanarachos et al. [57] proposed a method that can evaluate vehicle energy consumption through the neural network only by using the GPS location, speed, altitude, and other information of the smartphone, without the need to use complex and cumbersome monitoring methods, which is accurate and simple, and brings new methods to monitor fuel consumption.

5) *Sideslip angle and vehicle deceleration prediction*: In the case of understeering or oversteering (such as drift), the vehicle's stability control system is particularly important, and the stability control system restores the vehicle to a stable state by interfering with the engine or wheels. In terms of lateral stability control, the slip angle and roll angle are the key parameters of vehicle stability, but these parameters require very expensive equipment measurement. Melzi et al. [58] based on the vehicle's lateral acceleration, yaw rate, speed, and steering angle, the lateral slip angle is estimated by the neural network, which greatly reduces the cost and the accuracy of the method is proved to be high [59]. The roll angle and the slip angle are estimated at the same time, and the cost is greatly reduced.

If the driver is given a certain amount of warning before a collision occurs, a large number of traffic accidents will be reduced, and even in the case of automatic driving, it also takes a certain period of time to respond. Brake deceleration is the ability of the vehicle to rapidly reduce its travel speed until it stops while on the move. Zhang et al. [60] proposed that

according to the relative distance between vehicles, the relative speed of vehicles, the acceleration of the vehicle in front of the vehicle, and the lateral distance, the braking deceleration of the vehicle should be effectively predicted, and the ability of the vehicle to immediately slow down and stop should be judged.

III. FUSION METHODOLOGY

For multi-modal fusion, it is extremely important to decide what information fusion and fusion method to use. The recurrence plot and spectrogram, which can transform one-dimensional information into two-dimensional images and are convenient for two-dimensional convolutional neural networks, are first introduced in detail. Then we present efficient tensor fusion and adaptive multi-modal fusion technology.

A. Recurrence plot

Recurrence Plot (RP) was first proposed in 1987 for qualitative analysis of nonlinear dynamical systems. A general recurrence plot is an important method to analyze the periodicity, chaos, and non-stationary of time series data, giving prior knowledge about similarity, informativeness, and predictability, and a recurrence plot is especially useful for shorter time series. Recursion plots are also suitable for identifying dynamic pattern changes in time series, such as financial data time series [61], [62] and ecosystem time series [63]. Because RP can convert time series into two-dimensional information, it is especially suitable for two-dimensional convolutional neural networks, which provide important support for the transformation of one-dimensional information and two-dimensional information. Shahverdy et al. [40] used this technology to convert the vehicle information into RP, and then input it into CNNs.

B. Spectrogram

In the field of signal processing, there are three main domain aspects to analyze the signal, namely the time domain, the frequency domain, and the time-frequency domain, corresponding to the time domain map, the frequency domain map, and the time-frequency map, that is the spectrogram. The time domain and frequency domain can represent information in only 2D of signals, while a spectrogram uses 2D images to represent information in three dimensions. In spectrograms, the horizontal axis is time and the vertical axis is frequency, and the coordinate point value is the speech data energy. Since three-dimensional information is expressed in two dimensions, the size of the energy value is expressed by color, and the color depth indicates that the strength of speech energy of the point [64], which is a comprehensive description of the audio in the time domain and frequency domain characteristics. Spectrogram is based on the short-term Fourier transform (STFT) and is very helpful in analyzing the time-frequency characteristics of a signal. STFT is the most classic time-frequency domain analysis method. STFT by a long period of signal framing, and windowing, and then do Fourier transform (FFT) for each frame, and finally stack

the results of each frame along another dimension, forming a spectrogram. Arandjelovic et al. [65], [66] converted the sound into a log-spectrum, and then through a convolutional neural network processing, to determine whether the video matches the sound and locate the sounding object.

The steering angle, vehicle speed, acceleration, and other information about the vehicle change with time. The spectrogram reflects the time domain and frequency domain information of the signal. It can convert one-dimensional information into two-dimensional information, so it can be more conveniently applied to convolutional neural networks.

C. Tensor fusion

As a mainstream fusion method of multi-modal information fusion, tensor fusion is used in a variety of applications, and autonomous driving is no exception. There are many ways of tensor fusion, such as early simple feature stitching, late decision fusion, tensor outer product fusion, and so on.

As can be seen from the above, in the field of automatic driving, in addition to images and other data, the data information of vehicles, such as steering angle and vehicle speed, are also very important. Through the tensor outer product method, various such vehicle information and image data information can be fused, to better solve practical problems. This fusion method can fuse the features between various data information more fully and flexibly, and the effect is better than simple feature stitching and late decision-making.

Zadeh et al. [67] proposed tensor outer product fusion applied to their Tensor Fusion Network (TFN), and many subsequent optimizations and variant models are also based on this network model. Tensor outer product is the outer product operation of the feature vector extracted from each mode to get a high-dimensional fusion matrix Z , and then project the high-dimensional fusion matrix Z into the low-dimensional space through a linear layer. Each element of each eigenvector is fully fused. After the fusion of two modes, a second-order tensor is formed, and after the fusion of three modes, a third-order tensor is formed.

TFN needs external product operation for the feature vectors of each mode, when there are many feature vectors to be fused, the network will carry out the high-dimensional tensor calculation, and the calculation cost will be very high. For example, when we need to fuse the three information of steering angle, vehicle speed, and image at the same time, after fusion, we will get a third-order tensor Z . If we want to use the linear layer to project it into the low dimensional space, we need a fourth-order weight matrix W and Z to complete the calculation.

In view of the high computational cost of the TFN, many network models have proposed different solutions. Through rigorous mathematical derivation, Liu et al. [68] decomposed the parameter W and fusion tensor Z of the TFN. Then, the high-order tensor operation is decomposed into linear operation, so that the calculation cost will not increase exponentially with the increase of mode, and propose a Low-rank Multi-modal Fusion. The hierarchical plural fusion network (HPFN) proposed in the paper [69] not only adopts a Low-rank Multi-modal Fusion strategy similar to the paper [68] to solve the

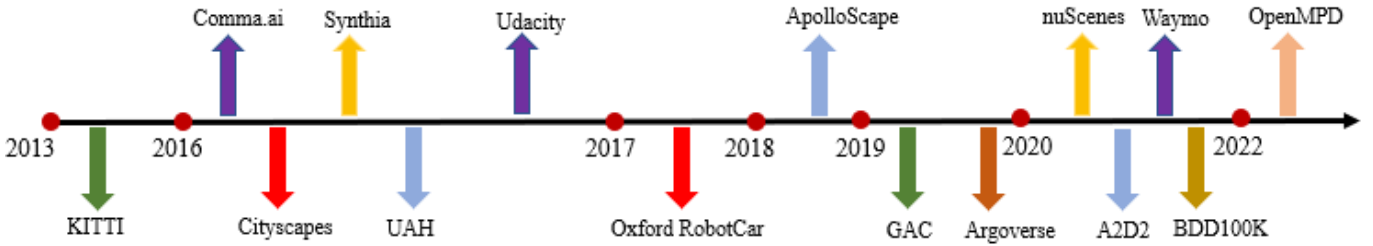


Fig. 5. Dataset timeline with vehicle information.

problem of high computational cost, but also realizes local feature fusion through the added sliding window mechanism. The feature fusion of multiple time periods and the fusion sequence can be controlled to make the feature fusion more sufficient. Zhu et al. [70] improved the Low-rank Multi-modal Fusion network and solved the problem that TFN ignores the correlation between various modes by adding a self-attention mechanism. The specific operation is to change each mode into a new single-mode feature vector through the self-attention mechanism module, and then fuse through Low-rank Multi-modal Fusion.

D. Adaptive fusion

Compared with the above fusion methods, adaptive fusion is more flexible and natural, because the network using this fusion method will not determine a specific fusion operation, such as feature stitching, tensor outer product, etc., but let the network decide "how" to more effectively integrate a given set of multi-modal features. Sahu et al. [71] proposed two adaptive fusion network structures: 1) Auto-Fusion, which encodes the information of all modes and splices them into a tensor, then restores the features with the decoder and finally calculates the loss between the features. This method not only integrates the feature vectors but also learns the useful features. It solves the problem that the final predictor bears the additional responsibility of identifying useful signals. 2) GAN-Fusion, the network first finds a trunk mode, then fuses the other modal information except for the trunk mode, and fuses the fused information against the information of the trunk mode, thus we can obtain the new feature vector for each pattern. In the same operation, all modes are used as the trunk mode at the same time, so we can obtain the new feature vector of each mode, and then splice these feature vectors. That is to complete the final integration.

IV. DATASETS

Many datasets only contain image and LiDAR information for semantic segmentation, object detection, and other tasks, but few datasets contain a large amount of vehicle information at the same time. This chapter shows some datasets covering images, LiDAR, vehicle information, and so on, which makes the research of multi-modal fusion based on vehicle information more convenient. The various datasets that have emerged over time are shown in Fig. 5, and the detailed information contained in them is shown in Table I.

A. UAH-DriveSet

UAH-DriveSet [83] is a public driving analysis dataset released in 2016. Six drivers of different ages and genders drive different vehicles on different roads (motorway and secondary road) to simulate three different driving behaviors (normal driving, drowsy driving, and aggressive driving), so as to collect data, including video, original vehicle data, and processed data. The UAH-DriveSet is available at: <http://www.robosafe.com/personal/eduardo.romera/uah-driveset>.

Each folder of the dataset includes video and data text files. Data files include measurements obtained by the mobile phone sensor; variables processed by DriveSafe in real-time (car position relative to lane center, car angle relative to lane curvature, distance to ahead vehicle in current lane, etc); a single event generated during driving detected by the DriveSafe (brake, accelerate, steer, etc); Drivesafe's rating of driver behavior. Vehicle bus data mainly includes vehicle speed, acceleration, longitude, and latitude, etc.

The dataset also provides a reader tool, by selecting a route that can be synchronized in the interface to display video and recorded data changes in real time, which can facilitate researchers to observe the performance of vehicles in certain driving behavior, providing researchers with good analysis tools.

B. BDD100K

The BDD100K dataset [80], released in 2018 by the University of Berkeley AI Lab (BAIR), is a classic large-scale and diverse dataset of driving videos.

The dataset includes high-resolution (720p) image information, as well as GPS/IMU information to save driving trajectories. The dataset contains 100,000 driving videos (40 seconds each) from different areas (New York, San Francisco Bay Area, and other areas) and different scenarios (city roads, residential areas, and highways), different weather, and different times. Due to the diversity of its external environment, the robustness of the applied model is improved. Dataset videos are divided into a 70K training set, a 10K validation set, and a 20K test set. The dataset can be used for tasks such as object detection, lane line detection, semantic instance segmentation, multiple object tracking, and segmentation.

C. A2D2

Audi Autonomous Driving Dataset (A2D2) [74] is a public driving dataset released by Audi in 2020. Data and other

TABLE I

PUBLIC DATASETS FOR AUTONOMOUS DRIVING. WE DESCRIBE IN DETAIL THE NUMBER OF IMAGES CONTAINED IN THE DATA SET AND WHETHER IT CONTAINS LiDAR DATA, RADAR DATA, GPS DATA, VEHICLE BUS DATA, AND CONTROL-RELATED DATA.

Dataset	Samples	Image Type	LiDAR	RADAR	GPS	Bus Data	Control Actions
Udacity	34,000	RGB	YES	NO	YES	YES	Steeringwheel, Speed
Waymo [72]	1,000,000	RGB	YES	NO	YES	NO	Steeringwheel
Comma.ai [73]	video	RGB	NO	NO	YES	YES	Steeringwheel, Speed
GAC [25]	3,240,000	RGB	NO	NO	—	—	Steeringwheel, Speed
A2D2 [74]	433,833	RGB	YES	NO	YES	YES	Steeringwheel, Speed, Acceleration
Cityscapes [75]	25,000	RGB	NO	NO	YES	YES	—
KITTI [76]	14,999	RGB	YES	NO	YES	NO	—
Synthia [77]	213,400	RGB	YES	NO	NO	—	—
Oxford RobotCar [78]	20,000,000	RGB	YES	YES	YES	YES	Speed
nuScenes [79]	1,400,000	RGB	YES	YES	YES	YES	Steeringwheel, Speed, Acceleration
BDD100K [80]	120,000,000	RGB	NO	NO	YES	NO	—
ApolloScape [81]	143,906	RGB	YES	NO	YES	YES	—
Argoverse [82]	video	RGB	YES	NO	YES	YES	—
UAH [83]	video	RGB	NO	NO	YES	YES	Steeringwheel, Speed, Acceleration
OpenMPD [84]	15,000	RGB	YES	NO	NO	YES	Steeringwheel, Speed, Acceleration

details are available from the <http://www.a2d2.audi>. The A2D2 dataset also has a certain diversity, taking into account not only the diversity of road environments (highways, country roads, and urban roads in southern Germany), but also the diversity of weather (cloudy, rainy, and sunny days). Datasets include image data and 3D point clouds, as well as 3D bounding boxes, semantic segmentation data and label, instance segmentation data and label, and vehicle information from the car bus, which can be used for a wide variety of tasks due to the diversity of their annotations. Vehicle bus data mainly includes steering angle, brake, throttle, odometry, etc. A total of 41,227 frames with semantic segmentation labels and point cloud labels in the dataset, of which 12,497 frames contained 3D callout boxes for targets in the field of view of the front camera, and also included 392,556 consecutive frames of unlabeled sensor data.

D. Udacity

In collecting vehicle dynamics information, ROS provides a flexible ecosystem and includes powerful tools and libraries. Many researchers use the ROS framework for dataset creation. ROS nodes can perform synchronous data collection from the CAN bus and RGB cameras. Through the ROS node information, the dynamic information of the vehicle during the driving process, such as speed, steering angle, braking, torque, and accelerator value, can be obtained.

Udacity's autonomous driving dataset, 1920 × 1200 resolution images captured using a Point Grey research camera, the data collected is divided into two datasets. The first consists of daytime conditions in Mountain View, California, and adjacent cities. The dataset contains more than 65,000 annotated objects in 9,423 frames, and the annotation method combines machines and humans. The labels are cars, trucks, and pedestrians. The second dataset is generally similar to the former, except that the annotation content of traffic lights is

increased, the number of datasets has also increased to 15,000 frames, and the annotation method is completely manual. In addition to the images taken by the vehicle, the content of the dataset also includes the attributes and parameters of the vehicle itself, such as latitude and longitude, brakes, accelerator, steering, and rotational speed. Many researchers have performed the above fusion research on the vehicle information and visual data on the Udacity dataset. The Udacity dataset is available at: <https://github.com/udacity/self-driving-car/tree/master/datasets>.

E. Comma.ai

The Comma.ai [73] Dataset is a video dataset for autonomous driving containing a total of 7.25 hours of video, which contains 10 videos recorded at 20 Hz by a video camera mounted on an Acura ILX 2016 windshield Camera recording. In addition to this, the dataset also contains some measurements of the vehicle itself, such as car speed, acceleration, steering angle, GPS coordinates, gyroscope angle, etc. The measurements are all converted to a uniform 100 Hz time base, the Comma.ai dataset is provided by Comma.ai company was released in 2016.

F. OpenMPD

OpenMPD [84] is our multi-modal autonomous driving dataset released in 2022. Compared with the existing dataset, OpenMPD includes more complex driving scenes (such as night, road construction, U-turn, etc.) and complex roads (such as intersections, tunnels/culvert, viaducts, etc.) to provide diverse data for autonomous driving tasks. Our vehicle is equipped with six cameras and four LiDARs to achieve 360-degree full coverage while acquiring multi-modal data. In particular, we use 128-beam LiDAR to collect high-resolution dense point cloud data. We have extracted 15K keyframes for

annotation, including 2D/3D semantic segmentation, 2D/3D bounding box, which can be used for semantic segmentation, object detection, object tracking, and other multi-task, multi-view research. We also collect vehicle information such as vehicle speed, acceleration, steering angle, and engine speed from CAN bus. For data and more information, please visit <http://www.openmpd.com/>.

V. CHALLENGES

Most current autonomous driving methods primarily involve learning control policies directly from sensory data without explicit intermediate representations. This approach, which maps images to driving behavior, often lacks robust generalization ability. Although much research has been done on the joint learning of vehicle information and images, there are still many challenges to be further investigated. In this section, we will discuss some open problems from vehicle control, scene generalization, and datasets, as follows.

The automatic driving function of the vehicle is mainly realized by the longitudinal motion control and the lateral motion control. In automatic vehicle control, longitudinal control is still a challenging problem [85]. In addition to the visual information collected by camera radar, with the increase in the number of onboard sensors, more and more vehicle motion state parameters can be collected. For example, longitudinal acceleration, air resistance, tire load, ground friction, ground inclination, etc. By combining image perception information, vehicles can achieve better longitudinal control in constant speed cruise, adaptive cruise, and anti-collision systems. Therefore, researchers can try to combine rich self-sensor information from different driving scenes to achieve vehicle control.

In terms of research scenarios, most studies currently focus on daytime driving scenarios [86]. In works such as [20] and [25], only a few of their studies involve nighttime driving, while as stated in [87], most of researchers ignored nighttime driving in their scenarios. If automatic driving only relies on daytime driving, its range of motion is limited. However, the information on self-owned sensors is not affected by bad weather and light conditions. Therefore, we suggest that researchers can widely use vehicle information in future research to improve scene generalization ability.

In terms of datasets, there are relatively few publicly available datasets with vehicle information. Most researchers collect fixed data for specific scenarios or conditions, and large-scale autonomous driving datasets with vehicle kinematic parameters are lacking. A common solution is to perform data augmentation on the limited data to obtain additional training data. However, it still has certain limitations compared to real datasets covering various lighting conditions and complex road conditions. In addition, many researchers have built autonomous driving simulation scenarios on simulators and obtained simulated data for model training, but in the transition from simulation to the real world, the error rate is almost doubled, and the complex and diverse real-world environments present even greater challenges. Therefore, it is necessary to build larger real-world datasets and more realistic simulated environments for training and testing in the future.

VI. FUTURE RESEARCH PROSPECTS

Compared with additional sensors such as LiDAR, vehicle information is not limited by the external environment and sensor hardware, and it is low-cost and can provide accurate and real-time data. However, many studies have not made full use of these advantages, especially in robot autonomous control, automatic driving, and so on. The following describes our future research directions in terms of vehicle speeds and steering angles.

A. Vehicle speed

Accurately distinguishing the driving behavior of drivers plays a significant role in driving assistance systems, road safety, energy optimization, and so on. The behavior classification method based on driver dynamics directly uses the camera to aim at drivers' faces and bodies, which violates drivers' privacy. The behavior classification method based on vehicle dynamics only uses vehicle information for analysis, which lacks images that can provide rich information about vehicles and roads. The change degree of vehicle information directly reflects the driver's driving behavior. For example, the sharp change of speeds in a short time reflects that the driver is in an aggressive state. The roadside image can reflect the degree of emptiness, congestion, and distance from obstacles. If the vehicle information is combined with the roadside image, it can make the information complementary and improve the classification performance.

1D information can be converted into 2D information through Gramian Angular Field, Markov Transition Field, Recurrence Plot, and Short-time Fourier Transform. We first convert the temporal speeds within five seconds into 2D and then get their own features with the roadside image through ResNet50, then pool in space to obtain features for fusion, and then fuse them through MSELoss. The MSELoss between the vehicle speed and the image features are calculated to force them to align. Finally, the MSELoss value is mapped to three driving behaviors through the full connection layer. We classify the driver's behavior (normal, drowsy, aggressive) according to the road (all road, motorway road, secondary road) in the public UAH-DriveSet, as shown in Table II, Table III and Table IV. Table II shows our comparison with other methods on all roads, which is 2.1% higher than the current best DriveBFR. Table III shows our comparison with other methods on motorway road, which is 7.4% higher than the current best Stacked-LSTM. Table IV shows our comparison with other methods on a secondary road, which is 3.3% higher than the best Stacked-LSTM. Experiments show that speed and image can be effectively fused to improve classification performance.

For autonomous vehicles, due to the fixed frequency of camera shooting, the vehicle can only track the object blindly without knowing the degree of the object position change. The vehicle information such as vehicle speed intuitively reflects the change degree of the object position. The faster the speed between adjacent frames, the greater the change of the object position, and the lower the object similarity between images. Therefore, the vehicle speed can be used as auxiliary

TABLE II

ON THE UAH DATASET, THE PERFORMANCE OF OUR METHOD OF FUSING VEHICLE SPEED AND IMAGE IS COMPARED WITH OTHER DRIVING BEHAVIOR CLASSIFICATION METHODS. THE ACC, PRE, AND REC REPRESENT ACCURACY, PRECISION, AND RECALL. THE "-" MEANS THAT IT IS NOT INDICATED IN THE METHOD.

Type	Model	F1	Acc	Pre	Rec
Machine Learning	MLP [88]	48.0%	-	-	-
	Decision Tree [88]	80.0%	-	-	-
	Generic Model [89]	-	81.3%	82.2%	80.7%
	Personalized Model [89]	-	89.5%	89.9%	89.2%
	Road-Aware Model [89]	-	83.6%	84.4%	83.2%
	Personalized Model With Road Information [89]	-	91.6%	92.0%	91.5%
	Discriminant Analy [89]	-	51.1%	-	-
	Decision tree [89]	-	63.1%	-	-
	KNN [89]	-	66.8%	-	-
	SVM [89]	-	67.4%	-	-
	Ensemble Learning [89]	-	81.3%	-	-
Neural Network	DriveBFR [90]	95.0%	-	-	-
	Stacked-LSTM [88]	91.0%	-	-	-
	HDC-ANN [91]	94.0%	-	-	-
	HDC-SNN [91]	92.0%	-	-	-
	Ours	97.1%	97.3%	97.0%	97.3%

TABLE III
PERFORMANCE COMPARISON WITH OTHER DRIVING BEHAVIOR CLASSIFICATION METHODS ON MOTORWAY ROAD.

Type	Model	F1	Acc	Pre	Rec
Machine Learning	Logistic Regression [92]	53.0%	54.0%	53.0%	54.0%
	Gradient Boosting [92]	68.0%	67.0%	70.0%	67.0%
	Random Forest [92]	63.0%	63.0%	63.0%	63.0%
	MLP [88]	51.0%	-	-	-
	Decision-Tree [88]	75.0%	-	-	-
Neural Network	Neural Network [92]	25.0%	29.0%	29.0%	27.0%
	Stacked-LSTM [88]	86.0%	-	-	-
	Our	93.4%	93.5%	93.5%	93.5%

TABLE IV
PERFORMANCE COMPARISON WITH OTHER DRIVING BEHAVIOR CLASSIFICATION METHODS ON SECONDARY ROAD.

Type	Model	F1	Acc	Pre	Rec
Machine Learning	Logistic Regression[93]	51.0%	55.0%	54.0%	52.0%
	Random Forest [93]	61.0%	63.0%	63.0%	62.0%
	Gradient Boosting [93]	64.0%	65.0%	65.0%	64.0%
	MLP [88]	64.0%	-	-	-
	Decision-Tree [88]	92.0%	-	-	-
Neural Network	Stacked-LSTM [88]	95.0%	-	-	-
	Our	98.3%	98.4%	98.3%	98.3%

information for object tracking, which can be combined with the image to improve the object detection ability.

B. Steering angle

Steering angle prediction can be incorporated into a wider range of computer vision-based techniques such as lane line detection, obstacle detection, etc. In the field of autonomous driving, one of the necessary capabilities of vehicles is the need to accurately identify lane lines. According to different

road conditions, reasonable lane keeping to prevent deviation from the safe driving area is inseparable from accurate steering angle prediction. During the turning process, since the lane line is in a curved state and the lane line spans different areas of the image, the convolution kernel in the current convolutional neural network has limitations in extracting feature regions, and cannot efficiently extract curved and cross-regional features. The turning angle describes the size of the turning span, that is, the bending degree of the lane line, which intuitively reflects the state of the lane line. Therefore, the turning angle can be used as auxiliary data in images to improve the recognition ability of lane lines.

To prove the correctness of our theory, we have carried out a lot of work and experiments in this field. First of all, we extract more than 6,000 roadside images and steering angle data corresponding to each image from Udacity, a public dataset in the field of automatic driving, and annotated the data. Secondly, we use the classic lane line detection algorithm Lanenet [94] as our baseline. Based on this baseline, we incorporate the steering angle data into the feature maps generated by downsampling. First, we perform a maxpool operation on the feature map to obtain the one-dimensional feature information of the feature map, and we extract steering angle features through Multi-Layer Perceptron (MLP), and then combine the one-dimensional feature information of the feature map with the steering angle feature vector after feature extraction to get the channel attention weights of the feature map, the weights are multiplied by the original feature map to get the new feature map that we fuse the steering angle. And classified into two groups by light intensity and whether the lane line is curved, to observe the index changes after adding steering angle in different scenes, as shown in Table V. The experimental results show that the steering angle and image fusion can improve the detection accuracy of lane lines in various scenes, and the improvement effect is most obvious in

the case of bad lighting and lane line curving. Among them, when the lane line is curved, the mIOU with a steering angle is 4.4% higher than that without a steering angle, and in the case of bad lighting, the mIOU with a steering angle is 4.7% higher than that without a steering angle.

TABLE V
EXPERIMENTAL COMPARISON OF LANE LINE DETECTION BASED ON
IMAGE AND STEERING ANGLE FUSION.

Group	Steering angle	mIOU	Acc	F1
All	yes	79.4%	99.3%	93.5%
All	no	76.3%	97.4%	89.9%
Good light	yes	78.4%	99.1%	89.4%
Good light	no	76.1%	98.4%	87.8%
Bad light	yes	77.2%	99.0%	92.1%
Bad light	no	72.5%	96.4%	88.3%
Curve lane	yes	76.5%	99.2%	93.0%
Curve lane	no	72.1%	96.4%	89.1%
Straight lane	yes	77.7%	99.4%	92.4%
Straight lane	no	76.9%	99.3%	91.8%

In addition, we have reproduced the open-source lane line algorithm in recent years on our dataset, and carried out a comparative experiment. Table VI shows that when we integrate the steering angle into our baseline, all indicators exceed other lane line detection algorithms. Among them, mIOU is 2.1% higher than CondLane, the current best lane detection algorithm, and 4.3% higher than SCNN.

TABLE VI
COMPARE THE LATEST OPEN-SOURCE LANE LINE DETECTION
ALGORITHM.

Method	Backbone	mIOU	Acc	F1
SCNN [95]	VGG16	75.1%	96.2%	89.1%
RESA [96]	ResNet50	75.3%	97.0%	90.1%
UFLD [97]	ResNet18	75.2%	96.5%	89.3%
LaneATT [98]	ResNet18	76.2%	97.1%	90.4%
LaneATT [98]	ResNet34	76.5%	97.6%	91.0%
LaneATT [98]	ResNet122	77.1%	98.2%	91.2%
CondLane [99]	ResNet18	76.6%	97.2%	89.7%
CondLane [99]	ResNet34	76.9%	97.5%	90.7%
CondLane [99]	ResNet101	77.3%	98.2%	91.1%
LaneNet+Angle (ours)	VGG16	79.4%	99.3%	93.5%

VII. CONCLUSIONS

A great many of the existing multi-modal fusion methods use image and point cloud data, which are acquired through external devices, it can not avoid the influence of the external environment and the limitations of the sensor itself. However, the vehicle information (steering angle, speed, etc) will not have the above factors, which can provide real-time and accurate information and an intuitive representation of the vehicle's status. In this paper, we first elaborate and analyze some studies of application steering angle and vehicle speed. Secondly, the multi-modal fusion technology is described in detail. Then, we recommend some public datasets that contain both images and information about the vehicle. Finally, We

describe the current challenges and future research prospects. Our survey can provide meaningful references for researchers interested in vehicle information and multi-modal fusion fields.

REFERENCES

- [1] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," *arXiv preprint arXiv:1704.06857*, 2017.
- [2] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 11, pp. 3212–3232, 2019.
- [3] M. Bijelic, T. Gruber, F. Mannan, F. Kraus, W. Ritter, K. Dietmayer, and F. Heide, "Seeing through fog without seeing fog: Deep multi-modal sensor fusion in unseen adverse weather," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11 682–11 692.
- [4] C. Shu and Y. Luo, "Multi-modal feature constraint based tightly coupled monocular visual-lidar odometry and mapping," *IEEE Transactions on Intelligent Vehicles*, pp. 1–10, 2022.
- [5] J. Nie, J. Yan, H. Yin, L. Ren, and Q. Meng, "A multimodality fusion deep neural network and safety test strategy for intelligent vehicles," *IEEE Transactions on Intelligent Vehicles*, vol. 6, no. 2, pp. 310–322, 2021.
- [6] W. Zhou, S. Dong, J. Lei, and L. Yu, "Mtanet: Multitask-aware network with hierarchical multimodal fusion for rgb-t urban scene understanding," *IEEE Transactions on Intelligent Vehicles*, pp. 1–1, 2022.
- [7] R. Bramer, I. Boada, A. Bardera, J. Rodriguez, M. Feixas, J. Puig, and M. Sbert, "Multimodal data fusion based on mutual information," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 9, pp. 1574–1587, 2011.
- [8] L. Gao, L. Qi, E. Chen, and L. Guan, "Discriminative multiple canonical correlation analysis for information fusion," *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 1951–1965, 2017.
- [9] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, "Multi-view 3d object detection network for autonomous driving," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2017, pp. 1907–1915.
- [10] A. Asvadi, L. Garrote, C. Premebeda, P. Peixoto, and U. J. Nunes, "Multimodal vehicle detection: fusing 3d-lidar and color camera data," *Pattern Recognition Letters*, vol. 115, pp. 20–29, 2018.
- [11] 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.
- [12] J. Schlosser, C. K. Chow, and Z. Kira, "Fusing lidar and images for pedestrian detection using convolutional neural networks," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 2198–2205.
- [13] Z. Wang, W. Zhan, and M. Tomizuka, "Fusing bird view lidar point cloud and front view camera image for deep object detection," *arXiv preprint arXiv:1711.06703*, 2017.
- [14] A. Rangesh and M. M. Trivedi, "No blind spots: Full-surround multi-object tracking for autonomous vehicles using cameras and lidars," *IEEE Transactions on Intelligent Vehicles*, vol. 4, no. 4, pp. 588–599, 2019.
- [15] K. Samal, H. Kumawat, P. Saha, M. Wolf, and S. Mukhopadhyay, "Task-driven rgb-lidar fusion for object tracking in resource-efficient autonomous system," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 1, pp. 102–112, 2021.
- [16] S. Gu, T. Lu, Y. Zhang, J. M. Alvarez, J. Yang, and H. Kong, "3-d lidar+ monocular camera: An inverse-depth-induced fusion framework for urban road detection," *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 3, pp. 351–360, 2018.
- [17] K. Wang, T. Zhou, X. Li, and F. Ren, "Performance and challenges of 3d object detection methods in complex scenes for autonomous driving," *IEEE Transactions on Intelligent Vehicles*, pp. 1–18, 2022.
- [18] A. Sathyanarayana, S. O. Sadjadi, and J. H. Hansen, "Leveraging sensor information from portable devices towards automatic driving maneuver recognition," in *2012 15th International IEEE Conference on Intelligent Transportation Systems*. IEEE, 2012, pp. 660–665.
- [19] D. Yang, H. Zhang, E. Yurtsever, K. A. Redmill, and . zgner, "Predicting pedestrian crossing intention with feature fusion and spatio-temporal attention," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 2, pp. 221–230, 2022.

- [20] M. Bojarski, D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang *et al.*, "End to end learning for self-driving cars," *arXiv preprint arXiv:1604.07316*, 2016.
- [21] V. Rausch, A. Hansen, E. Solowjow, C. Liu, E. Kreuzer, and J. K. Hedrick, "Learning a deep neural net policy for end-to-end control of autonomous vehicles," in *2017 American Control Conference (ACC)*. IEEE, 2017, pp. 4914–4919.
- [22] E. Santana and G. Hotz, "Learning a driving simulator," *arXiv preprint arXiv:1608.01230*, 2016.
- [23] H. M. Eraqi, M. N. Moustafa, and J. Honer, "End-to-end deep learning for steering autonomous vehicles considering temporal dependencies," *arXiv preprint arXiv:1710.03804*, 2017.
- [24] Z. Yang, Y. Zhang, J. Yu, J. Cai, and J. Luo, "End-to-end multi-modal multi-task vehicle control for self-driving cars with visual perceptions," in *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018, pp. 2289–2294.
- [25] D. Wang, J. Wen, Y. Wang, X. Huang, and F. Pei, "End-to-end self-driving using deep neural networks with multi-auxiliary tasks," *Automotive Innovation*, vol. 2, no. 2, pp. 127–136, 2019.
- [26] Z. Huang, C. Lv, Y. Xing, and J. Wu, "Multi-modal sensor fusion-based deep neural network for end-to-end autonomous driving with scene understanding," *IEEE Sensors Journal*, vol. 21, no. 10, pp. 11 781–11 790, 2020.
- [27] D. Fortun, P. Bouthemy, and C. Kervrann, "Optical flow modeling and computation: A survey," *Computer Vision and Image Understanding*, vol. 134, pp. 1–21, 2015.
- [28] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big data*, vol. 3, no. 1, pp. 1–40, 2016.
- [29] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "Lstm: A search space odyssey," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 10, pp. 2222–2232, 2016.
- [30] Z.-N. Li, X.-H. Huang, T. Mu, and J. Wang, "Attention-based lane change and crash risk prediction model in highways," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 12, pp. 22 909–22 922, 2022.
- [31] W. Kim, C. M. Kang, Y. S. Son, S.-H. Lee, and C. C. Chung, "Vehicle path prediction using yaw acceleration for adaptive cruise control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 12, pp. 3818–3829, 2018.
- [32] K. Schmidt, M. Beggiato, K. H. Hoffmann, and J. F. Krems, "A mathematical model for predicting lane changes using the steering wheel angle," *Journal of safety research*, vol. 49, pp. 85–e1, 2014.
- [33] K. Wang, L. Pu, J. Zhang, and J. Lu, "Gated adversarial network based environmental enhancement method for driving safety under adverse weather conditions," *IEEE Transactions on Intelligent Vehicles*, pp. 1–1, 2022.
- [34] G. A. M. Meiring and H. C. Myburgh, "A review of intelligent driving style analysis systems and related artificial intelligence algorithms," *Sensors*, vol. 15, no. 12, pp. 30 653–30 682, 2015.
- [35] C. Xu, W. Zhao, C. Wang, T. Cui, and C. Lv, "Driving behavior modeling and characteristic learning for human-like decision-making in highway," *IEEE Transactions on Intelligent Vehicles*, pp. 1–12, 2022.
- [36] T. Lee and J. Son, "Relationships between driving style and fuel consumption in highway driving," in *16th Asia Pacific Automotive Engineering Conference*, no. 2011-28-0051, 2011.
- [37] W. Weidner, F. W. Transchel, and R. Weidner, "Telematic driving profile classification in car insurance pricing," *Annals of actuarial science*, vol. 11, no. 2, pp. 213–236, 2017.
- [38] R. Sayed and A. Eskandarian, "Unobtrusive drowsiness detection by neural network learning of driver steering," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 215, no. 9, pp. 969–975, 2001.
- [39] S. Arefnezhad, S. Samiee, A. Eichberger, M. Frühwirth, C. Kaufmann, and E. Klotz, "Applying deep neural networks for multi-level classification of driver drowsiness using vehicle-based measures," *Expert Systems with Applications*, vol. 162, p. 113778, 2020.
- [40] M. Shahverdy, M. Fathy, R. Berangi, and M. Sabokrou, "Driver behavior detection and classification using deep convolutional neural networks," *Expert Systems with Applications*, vol. 149, p. 113240, 2020.
- [41] S. Jia, F. Hui, S. Li, X. Zhao, and A. J. Khattak, "Long short-term memory and convolutional neural network for abnormal driving behaviour recognition," *IET Intelligent Transport Systems*, vol. 14, no. 5, pp. 306–312, 2020.
- [42] Y. Liu, P. Zhao, D. Qin, G. Li, Z. Chen, and Y. Zhang, "Driving intention identification based on long short-term memory and a case study in shifting strategy optimization," *IEEE Access*, vol. 7, pp. 128 593–128 605, 2019.
- [43] A. Cura, H. Küçük, E. Ergen, and İ. B. Öksüzözü, "Driver profiling using long short term memory (lstm) and convolutional neural network (cnn) methods," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 10, pp. 6572–6582, 2020.
- [44] S. Sharma, G. Tewolde, and J. Kwon, "Lateral and longitudinal motion control of autonomous vehicles using deep learning," in *2019 IEEE International Conference on Electro Information Technology (EIT)*. IEEE, 2019, pp. 1–5.
- [45] B. Jiang and Y. Fei, "Vehicle speed prediction by two-level data driven models in vehicular networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 7, pp. 1793–1801, 2016.
- [46] K. Yeon, K. Min, J. Shin, M. Sunwoo, and M. Han, "Ego-vehicle speed prediction using a long short-term memory based recurrent neural network," *International Journal of Automotive Technology*, vol. 20, no. 4, pp. 713–722, 2019.
- [47] J. Xing, L. Chu, C. Guo, S. Pu, and Z. Hou, "Dual-input and multi-channel convolutional neural network model for vehicle speed prediction," *Sensors*, vol. 21, no. 22, p. 7767, 2021.
- [48] J. Xing, L. Chu, Z. Hou, W. Sun, and Y. Zhang, "Energy management strategy based on a novel speed prediction method," *Sensors*, vol. 21, no. 24, p. 8273, 2021.
- [49] Y. Zhang, Z. Huang, C. Zhang, C. Lv, C. Deng, D. Hao, J. Chen, and H. Ran, "Improved short-term speed prediction using spatiotemporal-vision-based deep neural network for intelligent fuel cell vehicles," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 9, pp. 6004–6013, 2020.
- [50] C. Zhang, Y. Zhang, Z. Huang, C. Lv, D. Hao, C. Liang, C. Deng, and J. Chen, "Real-time optimization of energy management strategy for fuel cell vehicles using inflated 3d inception long short-term memory network-based speed prediction," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 2, pp. 1190–1199, 2021.
- [51] F. Shi, Y. Wang, J. Chen, J. Wang, Y. Hao, and Y. He, "Short-term vehicle speed prediction by time series neural network in high altitude areas," in *IOP Conference Series: Earth and Environmental Science*, vol. 304, no. 3. IOP Publishing, 2019, p. 032072.
- [52] K. Messaoud, I. Yahiaoui, A. Verroust-Blondet, and F. Nashashibi, "Attention based vehicle trajectory prediction," *IEEE Transactions on Intelligent Vehicles*, vol. 6, no. 1, pp. 175–185, 2021.
- [53] R. S. Tomar, S. Verma, and G. S. Tomar, "Prediction of lane change trajectories through neural network," in *2010 International Conference on Computational Intelligence and Communication Networks*. IEEE, 2010, pp. 249–253.
- [54] D. Jeong, M. Baek, and S.-S. Lee, "Long-term prediction of vehicle trajectory based on a deep neural network," in *2017 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2017, pp. 725–727.
- [55] Y.-L. Chen, Y.-W. Kuo, S.-S. Wang *et al.*, "Artificial neural network for predictions of vehicle drivable range and period," in *2012 IEEE International Conference on Vehicular Electronics and Safety (ICVES 2012)*. IEEE, 2012, pp. 329–333.
- [56] S. Modi, J. Bhattacharya, and P. Basak, "Estimation of energy consumption of electric vehicles using deep convolutional neural network to reduce drivers range anxiety," *ISA transactions*, vol. 98, pp. 454–470, 2020.
- [57] S. Kanarachos, J. Mathew, and M. E. Fitzpatrick, "Instantaneous vehicle fuel consumption estimation using smartphones and recurrent neural networks," *Expert Systems with Applications*, vol. 120, pp. 436–447, 2019.
- [58] S. Melzi and E. Sabbioni, "On the vehicle sideslip angle estimation through neural networks: Numerical and experimental results," *Mechanical Systems and Signal Processing*, vol. 25, no. 6, pp. 2005–2019, 2011.
- [59] L. P. González, S. S. Sánchez, J. García-Guzmán, M. J. L. Boada, and B. L. Boada, "Simultaneous estimation of vehicle roll and sideslip angles through a deep learning approach," *Sensors*, vol. 20, no. 13, p. 3679, 2020.
- [60] S. Zhang, D. Li, F. Du, T. Wang, and Y. Liu, "Prediction of vehicle braking deceleration based on bp neural network," in *Journal of Physics: Conference Series*, vol. 2183, no. 1. IOP Publishing, 2022, p. 012025.
- [61] J. Belaire-Franch, D. Contreras, and L. Tordera-Lledó, "Assessing non-linear structures in real exchange rates using recurrence plot strategies," *Physica D: Nonlinear Phenomena*, vol. 171, no. 4, pp. 249–264, 2002.
- [62] F. Strozzi, J.-M. Zaldívar, and J. P. Zbilut, "Recurrence quantification analysis and state space divergence reconstruction for financial time series analysis," *Physica A: statistical mechanics and its applications*, vol. 376, pp. 487–499, 2007.

- [63] A. Facchini, C. Mocenni, N. Marwan, A. Vicino, and E. Tiezzi, "Nonlinear time series analysis of dissolved oxygen in the orbello lagoon (italy)," *Ecological modelling*, vol. 203, no. 3-4, pp. 339-348, 2007.
- [64] L. Wyse, "Audio spectrogram representations for processing with convolutional neural networks," *arXiv preprint arXiv:1706.09559*, 2017.
- [65] R. Arandjelovic and A. Zisserman, "Look, listen and learn," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 609-617.
- [66] —, "Objects that sound," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 435-451.
- [67] A. Zadeh, M. Chen, S. Poria, E. Cambria, and L.-P. Morency, "Tensor fusion network for multimodal sentiment analysis," *arXiv preprint arXiv:1707.07250*, 2017.
- [68] Z. Liu, Y. Shen, V. B. Lakshminarasimhan, P. P. Liang, A. Zadeh, and L.-P. Morency, "Efficient low-rank multimodal fusion with modality-specific factors," *arXiv preprint arXiv:1806.00064*, 2018.
- [69] M. Hou, J. Tang, J. Zhang, W. Kong, and Q. Zhao, "Deep multimodal multilinear fusion with high-order polynomial pooling," *Advances in Neural Information Processing Systems*, vol. 32, 2019.
- [70] H. Zhu, Z. Wang, Y. Shi, Y. Hua, G. Xu, and L. Deng, "Multimodal fusion method based on self-attention mechanism," *Wireless Communications and Mobile Computing*, vol. 2020, 2020.
- [71] G. Sahu and O. Vechtomova, "Adaptive fusion techniques for multimodal data," *arXiv preprint arXiv:1911.03821*, 2019.
- [72] P. Sun, H. Kretzschmar, X. Dotiwalla, 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.
- [73] H. Schafer, E. Santana, A. Haden, and R. Biasini, "A commute in data: The comma2k19 dataset," *arXiv preprint arXiv:1812.05752*, 2018.
- [74] 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.
- [75] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3213-3223.
- [76] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231-1237, 2013.
- [77] G. Ros, L. Sellart, J. Materzynska, D. Vazquez, and A. M. Lopez, "The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3234-3243.
- [78] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 year, 1000 km: The oxford robotcar dataset," *The International Journal of Robotics Research*, vol. 36, no. 1, pp. 3-15, 2017.
- [79] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: 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.
- [80] 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.
- [81] X. Huang, P. Wang, X. Cheng, D. Zhou, Q. Geng, and R. Yang, "The apolloscape open dataset for autonomous driving and its application," *arXiv preprint arXiv:1803.06184*, 2018.
- [82] 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.
- [83] E. Romera, L. M. Bergasa, and R. Arroyo, "Need data for driver behaviour analysis? presenting the public uah-driveset," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2016, pp. 387-392.
- [84] X. Zhang, Z. Li, Y. Gong, D. Jin, J. Li, L. Wang, Y. Zhu, and H. Liu, "Openmpd: An open multimodal perception dataset for autonomous driving," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 3, pp. 2437-2447, 2022.
- [85] S. Sharma, G. Tewolde, and J. Kwon, "Behavioral cloning for lateral motion control of autonomous vehicles using deep learning," in *2018 IEEE International Conference on Electro/Information Technology (EIT)*. IEEE, 2018, pp. 0228-0233.
- [86] H. Wang, Y. Chen, Y. Cai, L. Chen, Y. Li, M. A. Sotelo, and Z. Li, "Sfnet-n: An improved sfnet algorithm for semantic segmentation of low-light autonomous driving road scenes," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 11, pp. 21 405-21 417, 2022.
- [87] Z. Chen and X. Huang, "End-to-end learning for lane keeping of self-driving cars," in *2017 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2017, pp. 1856-1860.
- [88] K. Saleh, M. Hossny, and S. Nahavandi, "Driving behavior classification based on sensor data fusion using lstm recurrent neural networks," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2017, pp. 1-6.
- [89] D. Yi, J. Su, C. Liu, M. Qaddus, and W.-H. Chen, "A machine learning based personalized system for driving state recognition," *Transportation Research Part C: Emerging Technologies*, vol. 105, pp. 241-261, 2019.
- [90] J. Vyas, D. Das, and S. Chaudhury, "Drivebfr: Driver behavior and fuel efficiency-based recommendation system," *IEEE Transactions on Computational Social Systems*, 2021.
- [91] K. Schlegel, F. Mirus, P. Neubert, and P. Protzel, "Multivariate time series analysis for driving style classification using neural networks and hyperdimensional computing," in *2021 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2021, pp. 602-609.
- [92] R. Ghandour, A. J. Potams, I. Boulkaibet, B. Neji, and Z. Al Barakeh, "Driver behavior classification system analysis using machine learning methods," *Applied Sciences*, vol. 11, no. 22, p. 10562, 2021.
- [93] R. Ghandour, A. J. Potams, I. Boulkaibet, B. Neji, Z. Al Barakeh, and A. S. Karar, "Machine learning methods for driver behaviour classification," in *2021 4th International Conference on Bio-Engineering for Smart Technologies (BioSMART)*. IEEE, 2021, pp. 1-4.
- [94] D. Neven, B. De Brabandere, S. Georgoulis, M. Proesmans, and L. Van Gool, "Towards end-to-end lane detection: an instance segmentation approach," in *2018 IEEE intelligent vehicles symposium (IV)*. IEEE, 2018, pp. 286-291.
- [95] X. Pan, J. Shi, P. Luo, X. Wang, and X. Tang, "Spatial as deep: Spatial cnn for traffic scene understanding," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.
- [96] T. Zheng, H. Fang, Y. Zhang, W. Tang, Z. Yang, H. Liu, and D. Cai, "Resa: Recurrent feature-shift aggregator for lane detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 4, 2021, pp. 3547-3554.
- [97] Z. Qin, H. Wang, and X. Li, "Ultra fast structure-aware deep lane detection," in *European Conference on Computer Vision*. Springer, 2020, pp. 276-291.
- [98] L. Tabelini, R. Berriel, T. M. Paixao, C. Badue, A. F. De Souza, and T. Oliveira-Santos, "Keep your eyes on the lane: Real-time attention-guided lane detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 294-302.
- [99] L. Liu, X. Chen, S. Zhu, and P. Tan, "Condlanenet: a top-to-down lane detection framework based on conditional convolution," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 3773-3782.



Xinyu Zhang received the B.E. degree from the School of Vehicle and Mobility at Tsinghua University, Beijing, China, in 2001 and the M.S. degree at Tsinghua University, in 2004. He was a visiting scholar at University of Cambridge. He is currently an associate professor with the School of Vehicle and Mobility, and the head of the Mengshi Intelligent Vehicle Team at Tsinghua University. He is the author of more than 50 articles. His research interests include intelligent driving and multimodal information fusion.



Yan Gong is a graduate student majoring in Computer Science and Technology in Northeastern University, China. He has been a research assistant with the School of Vehicle and Mobility at Tsinghua University, China. He has authored and co-authored many academic papers. His research interests include pattern recognition, multi-modal fusion and deep learning.



Jun Li is an academican of the CAE Member and a professor of the School of Vehicles and Transportation of Tsinghua University. He has long presided over the independent product research and development and scientific and technological innovation of FAW Technology Center, built a domestic leading independent research and development system and capability, and completed a number of major product replacement research and development and national level projects.



Jianli Lu is a graduate student majoring in computer technology in China University of Mining and Technology (Beijing), China. She has been a research assistant with the School of Vehicle and Mobility at Tsinghua University, China. Her research interests include computer vision, multi-modal fusion and deep learning.



Jiayi Wu is a master in Shenyang University of Technology, majoring in computer science and technology. He has been a research assistant with the School of Vehicle and Mobility at Tsinghua University, China. His main research fields are computer vision, automatic driving, multi-modal fusion perception and deep learning.



Zhiwei Li, tutor of master students in Beijing University of Chemical Technology. In 2020, he studied as a postdoctoral fellow with Academician Jun Li at Tsinghua University. His main research interests include computer vision, intelligent perception and autonomous driving, and robot system architecture.



Dafeng Jin is an associate professor of Tsinghua University and the director of the Institute of Automotive Body Design and Engineering. The research fields are mainly active/semi-active suspension and product complexity design.