SCIENCE CHINA
Information Sciences

• REVIEW •

August 2019, Vol. 62 081101:1–081101:27
https://doi.org/10.1007/s11432-018-9850-9

Autonomous driving: cognitive construction and situation understanding

Shitao CHEN[1,2], Zhiqiang JIAN[1,2], Yuhao HUANG[1,2], Yu CHEN[1,2], Zhuoli ZHOU[1,2] & Nanning ZHENG[1,2*]

[1]*Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an* 710049*, China;*
[2]*National Engineering Laboratory for Visual Information Processing and Applications, Xi'an* 710049*, China*

Received 1 August 2018/Revised 27 December 2018/Accepted 15 March 2019/Published online 12 July 2019

Abstract Autonomous vehicle is a kind of typical complex artificial intelligence system. In current research of autonomous driving, the most widely adopted technique is to use a basic framework of serial information processing and computations, which consists of four modules: perception, planning, decision-making, and control. However, this framework based on data-driven computing performs low computational efficiency, poor environmental understanding and self-learning ability. A neglected problem has long been how to understand and process environmental perception data from the sensors referring to the cognitive psychology level of the human driving process. The key to solving this problem is to construct a computing model with selective attention and self-learning ability for autonomous driving, which is supposed to possess the mechanism of memorizing, inferring and experiential updating, enabling it to cope with traffic scenarios with high noise, dynamic, and randomness. In addition, for the process of understanding traffic scenes, the efficiency of event-related mechanism is more significant than single-attribute scenario perception data. Therefore, an effective self-driving method should not be confined to the traditional computing framework of 'perception, planning, decision-making, and control'. It is necessary to explore a basic computing framework that conforms to human driver's attention, reasoning, learning, and decision-making mechanism with regard to traffic scenarios and build an autonomous system inspired by biological intelligence. In this article, we review the basic methods and main progress in current data-driven autonomous driving technologies, deeply analyze the limitations and major problems faced by related algorithms. Then, combined with authors' research, this study discusses how to implement a basic cognitive computing framework of self-driving with selective attention and an event-driven mechanism from the basic viewpoint of cognitive science. It further describes how to use multi-sensor and graph data with semantic information (such as traffic maps and a spatial correlation of events) to realize the associative representations of objects and drivable areas, as well as the intuitive reasoning method applied to understanding the situations in different traffic scenarios. The computing framework of autonomous driving based on a selective attention mechanism and intuitive reasoning discussed in this study can adapt to a more complex, open, and dynamic traffic environment.

Keywords autonomous driving, event-driven mechanism, cognitive construction, situation understanding, intuitive reasoning

Citation Chen S T, Jian Z Q, Huang Y H, et al. Autonomous driving: cognitive construction and situation understanding. Sci China Inf Sci, 2019, 62(8): 081101, https://doi.org/10.1007/s11432-018-9850-9

1 Introduction

In recent years, with the development of science and society, many research groups around the world have started exploring related technologies in the field of autonomous driving. In 2004, the Defense

* Corresponding author (email: nnzheng@mail.xjtu.edu.cn)

© Science China Press and Springer-Verlag GmbH Germany, part of Springer Nature 2019

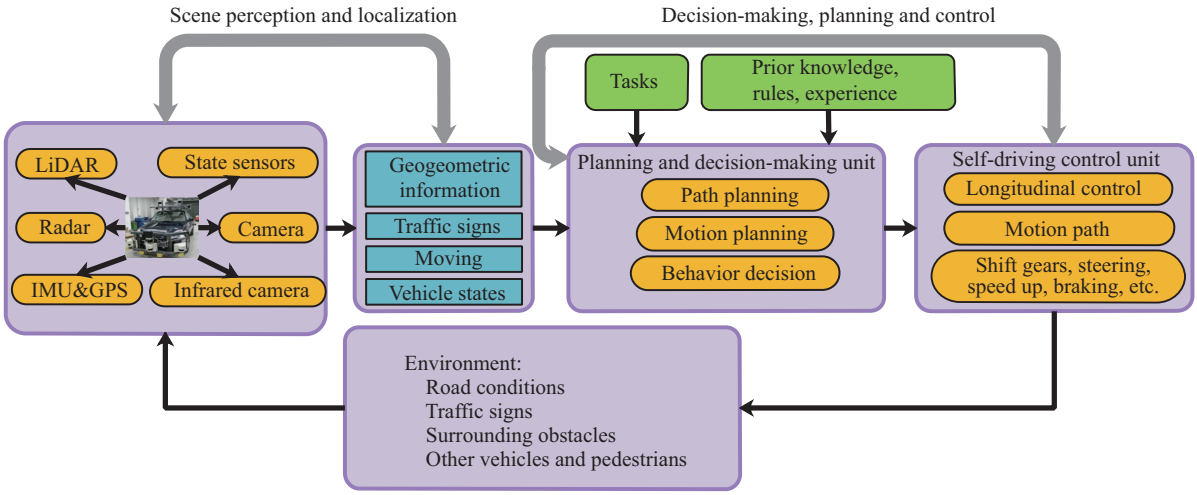

Figure 1 Data-driven computing framework of self-driving cars. The scene perception obtained by the sensors is used in localization, after planning and decision-making step, the control signal received by the self-driving control unit is utilized to control the self-driving cars. Meanwhile, environments like road conditions are feedback into the perception module for further processing.

Advanced Research Projects Agency (DARPA) pioneered 'DARPA Grand Challenge', but no vehicle was able to complete the mission of crossing a desert. In 2005, the Stanford team won the 'DARPA Grand Challenge' [1]. They proposed a computing framework for autonomous driving based on 'perception, planning, decision-making, and control' and verified its feasibility for the first time. Subsequently, Google followed the same framework but introduced more advanced perception algorithms and differential GPS positioning technologies into self-driving vehicles. It has laid the foundation for current technical route of autonomous vehicles, namely, the data-driven computing framework, as shown in Figure 1. This framework may generate a approximate model of the scene through a large number of data descriptions of the environment, then accordingly adjust the decision of the intelligent system and realize autonomous driving in a simple traffic scene or for a specific driving task.

However, the existing data-driven computing framework based on 'perception, planning, decision-making, and control' is showing increasing problems of low computational efficiency, poor adaptability to the environment, and insufficient self-learning ability. In data-driven mode, the correlations among the perceived results are ignored. The calculation of a large amount of data has caused a huge increase in the redundancy and complexity of the system, reducing the efficiency of intelligent systems. Owing to the inconsistencies between observed and actual scenes, which are caused by sensor noise and errors of the perception algorithm, the reliability of current autonomous vehicles needs to be guaranteed through the simplicity of the scene. At the same time, the description of a traffic scene based on the modeling method cannot express the spatio-temporal continuity of the scene, which reduces the early-warning capability of autonomous vehicles and the reliability of the driving path planning, and as a result, the vehicle is unable to adapt to open traffic scenes.

Although the data-driven computing framework can safely realize autonomous driving in a specific environment, to a certain extent, such 'autonomous cars' can only be regarded as automatic systems to complete a specific task, rather than intelligent systems with highly autonomous performance. An autonomous car based on the 'perception, planning, decision-making, and control' framework basically relies on the positioning system and the predetermined route. Similar to an industrial robot system that relies on sensor information to avoid obstacles, it is difficult to conduct a logical reasoning analysis or achieve situational cognition regarding its environment. Too much reliance on sensor data and neglect of the cognitive processes also make current self-driving cars unable to adapt to open traffic scenarios with high dynamic and strong randomness, which may even lead to traffic accidents and other serious consequences. It is therefore necessary to explore a new computing framework for current self-driving cars.

Traffic scenes have their own unique complexity and dynamics. In general, traffic scenes include roads, traffic signs, static and dynamic objects, and climatic conditions, among other factors. Because the road conditions and climate are usually complicated and varied, the traffic scenes faced by a self-driving vehicle are open. If a self-driving vehicle is expected to achieve fully autonomous driving in a complex traffic scene, it must have the ability to learn and make predictions. The cognitive process of human brain is supposed to be well referred to the autonomous vehicle's environmental understanding. It helps the vehicle possess the abilities of self-learning and self-referring in different environments, so as to build models which can understand and explain the world. Meanwhile, the vehicle can expand its knowledge during the process of continuous learning and evolution, and capture and build new models and knowledge extremely quickly to adapt to traffic scenarios of different characteristics. Autonomous vehicles face many different scenes and road conditions, such as high-speed scenes, low-speed urban roads, and unstructured roads. Under different scenarios, the requirements of the vehicle regarding its understanding of the environment are also different. For example, in a high-speed scene, the road environment is relatively structured, and the perception requirements for a complex environment are not high. Thus, we do not need to construct complex models to understand the environment. However, owing to the high speeds of autonomous vehicles, the complexity and real-time performance of the algorithm used need to be sufficiently high. Similarly, in urban roads, although a vehicle will not drive too quickly, it is necessary to construct a more complex model to accurately identify and estimate vehicle and pedestrian targets that may appear in an urban area. Nevertheless, merely using a large number of annotated scene data to perform special model design and training on different scenarios that may be encountered is not realistic (especially that the training for self-driving requires a large number of negative samples). Therefore, self-driving cars need a model that can learn independently under different road conditions and build the most suitable model for the corresponding traffic scenes on its own.

In summary, to achieve autonomous driving under all types of scenes, the following three points are necessary. First, a self-driving system must be an artificial intelligence system that cannot make mistakes and can respond safely to changing scenarios. Second, a self-driving system must be able to understand pre-actions and understand the driving intentions contained in the behavior. Third, a self-driving system must be capable of abstracting the situational information contained in the scene. Therefore, the realization of a fully autonomous self-driving system requires solving the following two basic scientific problems:

(1) How to make self-driving vehicles understand and remember traffic situations like a human being, such that they can possess the mechanism of memorizing, reasoning, and experiential updating, by which to cope with high dynamic and strong random traffic scene changes.

(2) How to develop an evolutionary and developing learning system for self-driving, where the learning process is similar to the training process of a human driver and can extend the generalized knowledge learned to new scenes that were previously unknown.

The key to solving the above two problems lies in how to introduce the human cognitive model into the framework of the current self-driving system. It is necessary to use multi-sensor data and graph data with semantic information to map the relationship between the objects and drivable area, as well as carry out intuitive reasoning for an understanding of the traffic scenarios, and eventually construct a cognitive computing framework for self-driving cars with selective attention and an event-driven mechanism.

In recent years, new discoveries and an unprecedented amount of data on the biological science of the brain and neural system have emerged. Meanwhile, the concepts of modern physics and computations have been widely applied to an analysis of cognitive processes. These developments [2–4] provide us with a biological and psychological basis for understanding "how the brain perceives the environment". It also provides important insight, allowing us to explore human-like intelligent driving. Human driving is a continuous process accompanied with a constant perception and understanding of the traffic scenes, which can be divided into two parts. The first is to extract the events in the scenes of continuous time segments and obtain semantic information. The second uses the internal relationship between different events to construct the cognition of the scenarios, and then form the graph data of the semantic information. It is optimal and complete to describe a space with events rather than data. Therefore, replacing a data-

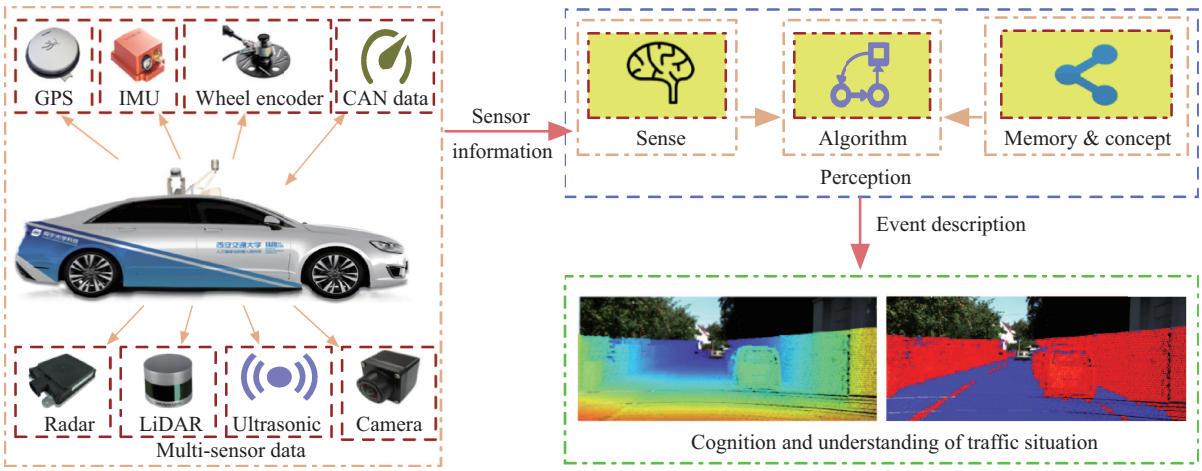

Figure 2 Cognition and understanding of traffic situation based on the information from the results of multi-sensor fusion. Using a variety of sensors to obtain data of different feature spaces and using different algorithms for analysis, and finally get cognitive results for the situation.

driven computing framework with the event-driven computing framework can enable self-driving cars to drive safely in open scenarios. The core of the event-driven computing framework lies in finding the intrinsic relations between the diverse data descriptions of the scenes in a continuous time segment and abstracting these descriptions, thereby accordingly adjusting the upper decision-making module.

Implementing a self-driving framework in event-driven mode may improve the efficiency and robustness of the system, thereby helping self-driving cars adapt to various scenarios and achieve an efficient response to a variety of circumstances. In this way, the driving tasks in complex traffic scenes can be complete with safety and reliability, which means that autonomous vehicles can achieve human-level intelligence when driving. This is a significant step toward a transition from "autonomous driving" to "self-driving".

The second part of this study mainly discusses the current progress and existing problems of the data-driven computing framework, and introduces the basic implementation methods of each module within this framework. As shown in Figure 2, in the third part of this study, we start from the basic viewpoints of cognitive science and a human driving behavior analysis, and further discuss how to transform the scene perception of multi-source information fusion into the cognition of the scenario and an understanding of the event-relation analysis. Then, combined with research conducted by the authors' team, the fourth part explores how to implement selective attention to the perception of complex traffic scenes and an intuitive reasoning of traffic scene understanding on a computer. In addition, the cognitive computing framework for self-driving cars based on a selective attention and event-driven mechanism is also given at the end of Section 4.

2 Classic computing framework of autonomous vehicles

The classic computing framework of autonomous vehicles relies on a serial information processing model based on 'perception, planning, decision-making, and control', which is called a data-driven computing framework for self-driving. The data-driven computing framework is the mainstream direction of the current self-driving research, which uses perception as the underlying module of the vehicle, and utilizes autonomous vehicle sensors including cameras, LiDAR, millimeter wave radar, and ultrasonic radar as information sources to understand traffic scenarios, such as dynamic and static objects detection, traffic identification, scene semantic understanding, and positioning through different perception algorithms. Planning is based on the realization of the environmental perception to find a set of drivable paths that meet the vehicle's kinematic constraints through search, prediction, trajectory generation, and other algorithms. As the core of autonomous driving system, the decision-making module is based on the results of the scene perception and motion planning to judge all drivable routes and determine the action

to take for vehicles. Control is used as the interface between the computing framework and the physical system, which guides the autonomous vehicle to drive according to the predetermined trajectory and speed through classic control theory and vehicle model construction. The data-driven computing framework for autonomous driving takes data as the guide and analyzes the characteristics and patterns of the data to implement the algorithms. Therefore, the framework takes data processing as the core, which has shown satisfactory results in structured traffic scenarios and autonomous driving of specific tasks with a continuous enrichment of the computing resources and data volume. Nevertheless, this computing model for autonomous driving cannot achieve an understanding or response to open, dynamic, and fragile complex traffic scenarios.

2.1 From sensation to perception

It is necessary for autonomous vehicles to constantly interact with their surroundings and adjust their behavior based on changes in the environment. The realization of this process is inextricably linked to the environmental perception of the vehicles, enabling them to recognize different parts of the environment, attaining an understanding of the scene in which they are located, and make decisions based on this understanding. Therefore, an accurate perception of the surrounding environment and detection of the drivable area are of significant importance to the safe and efficient driving of autonomous vehicles in complex and dynamic traffic scenarios.

According to the basic idea of neuroscience, the formation process of perception is as follows: A pulse signal in a neural network is formed through a stimulation of the external environment to the sensory organs, which is transmitted to the brain and interacts with the memory block to generate a feedback that is called perception, as shown in Figure 3. Abstractly, perception is a shallow understanding of a scene produced through both the sensation and cognitive effects. Because the memory blocks of the brain are diverse, the same sensory information produces different perceptions under different cognitive effects, which also leads to a diversity of human perception regarding the environment. In addition, the multi-aspect shallow understanding of the environment constitutes an exhaustive perception of the environment. Following the concept of perception, it can be concluded that the sensor data processing in an autonomous vehicle system is the perception of self-driving cars, including object detection, image segmentation, localization, and drivable area detection. Sensor data are information of the scene provided to autonomous vehicles, and different data processing algorithms are similar to a sensor interaction with a different memory block, which ultimately provides a variety of perception results. Through diverse perception methods, autonomous vehicles are capable of understanding different aspects of the environment.

Object detection is the extraction and classification of objects in a scene including 2D and 3D object detection. Here, 2D object detection obtains the scene data from visual sensors and analyzes the data using digital image processing algorithms to obtain 2D bounding boxes of the objects and their categories in the image. The classic 2D object detection algorithms apply HOG [5] or LBP [6, 7] features for calculation, whereas Ref. [8] used the stereo image and neural network for pedestrian detection, which contain a large amount of uncertainty. Except these methods, Ref. [9] proposed an incremental framework for traffic sign detection based on temporal constraint. The current state-of-the-art 2D object detection methods are based on deep learning, including methods with the region proposal [10–12] represented by the Faster-RCNN [12] and end-to-end methods [13–16] represented by SSD [13] and YOLO [14, 15], as shown in Figure 4. The former proposes some candidate boxes in the original image and extracts the features by applying a convolutional neural network. The features contribute to the filtering and optimization of the candidate boxes and the determination of the objects categories, which leads to the final results. The methods using the region proposal achieve high accuracy at the cost of low speed. The latter divides an image into grids of the same size. Each grid corresponds to diverse fixed-size anchors and is associated with specific features owing to the spatial invariance of the convolutional neural network, which determines whether there are objects in the relevant grid and what categories they belong to. If objects are in the grid, the bounding boxes will be calculated through a regression with its corresponding

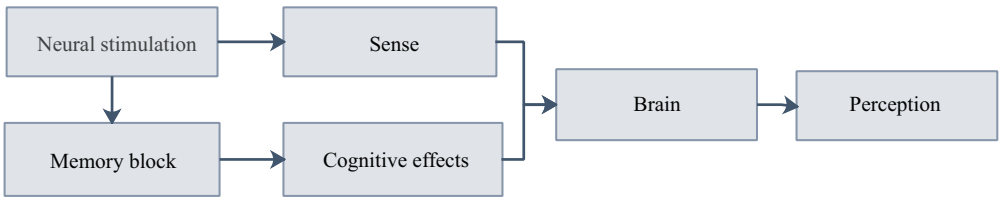

Figure 3 The generation process from sensation to perception. Perception is the cognitive result of the external things formed by the brain. It is the combination of two parts of information. The first part is the input information of the outside world, and the second part is the memory.

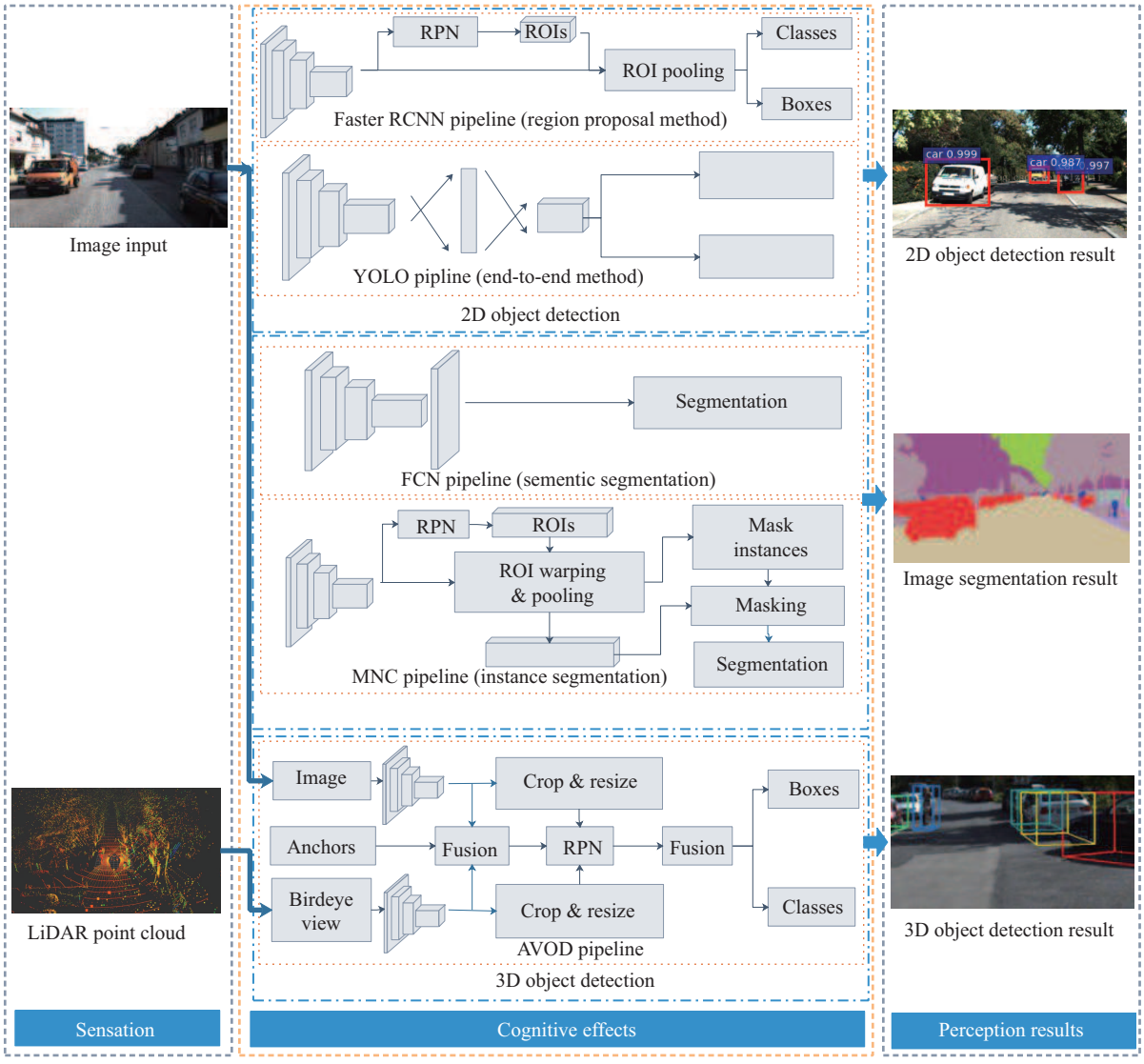

Figure 4 Objects detection and image segmentation in traffic scenarios. Three types of traffic scene perception algorithms are listed here, the input of which is mainly images and LiDAR point cloud. Different types of algorithms have different outputs for specific tasks, and these outputs will also play different roles in the perception of autonomous driving systems.

anchors. In addition, the final results can be obtained by merging overlapping bounding boxes. The end-to-end methods achieve both high accuracy and speed. The 3D object detection fuses and processes the LiDAR point cloud and the image to acquire 3D bounding boxes and categories of objects in the boxes, as shown in Figure 4. The state-of-the-art algorithms for 3D object detection include MV3D [17] and AVOD [18], both of which generate 3D candidate boxes by RPN [12], and then merge the bird-eye view of the LiDAR point cloud and image data to extract features through a convolutional neural network.

Bounding boxes are obtained by filtering and optimizing the candidate boxes according to the features and so do the categories. However, these algorithms achieve an unsatisfactory level of accuracy (the test accuracy rates on the KITTI dataset [19] are 62.35% and 71.88%), which indicates that further research is needed.

Image segmentation is used to distinguish pixels belonging to different parts of an image in order to separate the observed scene. Classic image segmentation algorithms mainly divide the regions in the image by using edge operator, cluster analysis, and wavelet transform. However, because an image segmented by classic algorithms is not clearly divided, and the divided parts of the image lack attributes, the current mainstream traffic scene image segmentation methods are semantic segmentation and instance segmentation, as shown in Figure 4. Semantic segmentation judges the category of each division when segmenting pixels. As a typical semantic segmentation algorithms, FCN [20] applies fully convolutional networks for feature extraction, and then deconvolutes to obtain the final result and category information, which proves the effectiveness of deep learning in the field of image semantic segmentation. Based on this, the follow-up algorithms are divided into two parts according to their architectures: algorithms based on the codec architecture, and an expanded convolution architecture. The former, including SegNet [21] performs feature extraction through the convolutional layer and encodes through the pooling layer with the encoding information recorded. Finally, it decodes through an unsampling process according to the encoding information to obtain the semantic segmentation result. The latter, including Deep Lab [22,23], convolutes by introducing an expanded convolution kernel and discarding the pooling layer, and finally reuses the conditional random field to obtain a semantic segmentation result. Wang et al. [24] also proposed a joint method of priori convolutional neural networks at superpixel level and soft restricted context transfer for scenes labeling, whereas Oliveira et al. [25] utilized a customized and efficient CNN for road segmentation. Instance segmentation distinguishes the instances of the division in the image. Compared with the classic image segmentation, instance segmentation segments each individual object in an image. The current instance segmentation algorithms include MNC [26], Mask-RCNN [27], and FCIS [28]. By introducing the bounding boxes proposed by object detection into the image segmentation model, the final instance segmentation result is obtained by combining boxes with fully convolutional networks.

The localization information of a self-driving car is usually obtained by receiving satellite signals of the global positioning system (GPS). Nevertheless, in order to avoid the problem in which positioning cannot be performed when the satellite signal is lost, the fusion of GPS and a perception-based positioning system has become the mainstream of vehicle localization. The method of localization through environmental perception is usually to initialize the positioning information according to the inertial navigation unit, and then performs key points extraction on the visual or LiDAR sensor data, which are matched within a certain range of the map according to the initial positioning information to obtain the accurate localization of autonomous vehicles [29–31]. However, it is of high cost and large computation for the key points matching methods to be applied in a large-scale scene. Therefore, how to use the perceptual data for localization in highly dynamic scenarios is still an unsolved problem.

Drivable area detection is the basic guarantee for the safe driving of autonomous vehicles. As shown in Figure 5, Liu et al. [32] proposed an unsupervised algorithm for drivable area detection, which mainly fuses the image information of the camera and point cloud information of the LiDAR for detection. During the process of information fusion, the algorithm considers the topological structure and geometric relationship between different modal data obtaining good test results of the drivable area without relying on the training data, which realizes an unsupervised method for drivable area detection with fast speed and high adaptability. In particular, due to the existence of traffic rules in structured roads, the problem of the drivable area is also transformed into a road detection problem. By extracting the lane boundaries in the LiDAR point cloud and images, or segmenting the senor data, the drivable area can be obtained. Traditional methods extract lane boundaries by Hough transform [33], which shows a relatively low level of accuracy. Therefore, the current mainstream method of lane boundary extraction is to obtain the bird-eye view of the road by inverse perspective transformation of the scene image, and then use a deep learning algorithm to perform lane boundary discrimination. Finally, lane boundaries are calculated by

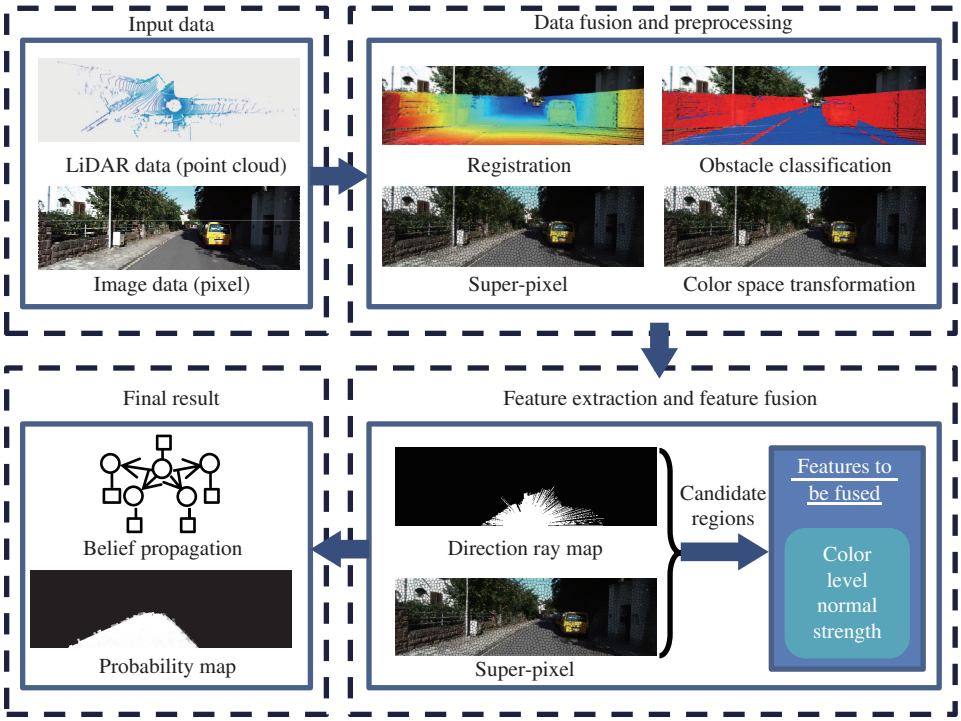

Figure 5 Unsupervised method for drivable area detection. The scene is analyzed in different feature spaces, and the different features obtained are merged. Finally belief propagation is used to analyze the fused features to obtain probability of the drivable area.

regression [34–36]. Compared with the classical road detection algorithms which only use images [37], nowadays, road extraction is usually implemented by using a fusion algorithm with LiDAR point cloud and camera images [38–41].

2.2 Motion planning and decision-making

2.2.1 *Motion planning – trajectory generation of autonomous vehicles*

As the upper structure of the perception module, the motion planning module of an autonomous driving system aims at finding a drivable path from the current position to the target based on knowledge of the environment. A drivable path means that the kinematics module of vehicles should be satisfied and that collisions should be avoided when self-driving cars follow the path. The mainstream approach to motion planning consists of search, prediction, and trajectory generation. Search is to find a path from the start to the goal on the given map in order to navigate the autonomous vehicle in complicated environments. Prediction is to forecast the objects' movement in the environment aiming at the improvement of autonomous vehicles' reliability and safety. In addition, trajectory generation is to construct a path that satisfies the vehicle kinematics model on purpose of meeting the demand of the executive agency.

Currently, A* [42–44] is one of the most widely spread algorithms for searching, which discretizes the environment into a grid map in which each grid is labeled by a value related to whether the grid is occupied by obstacles, the cost of moving from the start to the grid, and the heuristic function of the grid (the heuristic function is a custom function that measures the distance from the grid to the target in the map). As the grid map is constructed, the final result is generated by searching the path from the start to the target with the least cost in the map. Furthermore, A* algorithm always finds the best path in the map for its completeness and optimality. Compared to the discrete search algorithm represented by A*, rapidly exploring random tree (RRT) [45, 46] is also widely applied in the self-driving car as one of the classic continuous search algorithm. The RRT algorithm regards the start as the root to generate a tree by randomly sampling the drivable areas in the environment as leaf nodes. When the goal is

included by the tree, tree search makes sense to draw the route from the start to the goal. However, the path searched by RRT algorithm is a feasible solution rather than an optimal one. Besides, probabilistic roadmap (PRM) [47], artificial potential field [48], and neural network [49] are also applied for search. Because paths obtained by search algorithms are not kinematically constrained, it is impossible to input them into an executive agency directly, which highlights the necessity of trajectory generation.

The prediction module forecasts whether the autonomous vehicle will collide with other objects in the environment when driving along a planned path, contributing to an improvement of vehicle safety and the avoidance of emergency braking. Based on the environmental perception, the motion state of each object is modeled as a finite state machine. By detecting the states of the objects in the continuous time segment as an input, the Bayesian model is applied to calculate the possibility of objects' next motion states [50]. In addition to the methods of prediction by probability models, methods based on deep learning and reinforcement learning directly forecast the motion of objects by detecting their intention [51–53]. Also Kim et al. [54] used the LSTM for trajectory prediction over occupancy grid map. The results of the prediction ultimately function to the decision-making module to provide discriminative information for planned path selection.

Trajectory generation aims to construct curves that satisfy the kinematic constraints from the start to the target. Quintic spline curve generation and its optimization methods [55,56] are classic trajectory generation methods that generate curves by fitting the quintic polynomial based on the vehicle's position, velocity, and acceleration at the start and goal. The Quintic spline generation algorithm achieves expected results in vehicle following, lane changing, and overtaking on the structured roads for self-driving cars. In addition, B-spline curve generation [57] and G3 curve generation [58] are also common methods similar to the quintic spline curve generation. Otherwise, the hybrid A* algorithm is also one of the most important methods in trajectory generation [59], which regards the environment as a continuous space and performs a cost map search under the constraints of the kinematic model. Compared with the classic A* algorithm, the hybrid A* algorithm meets the executive agency's demand at the expense of completeness and optimality.

The motion planning module of autonomous vehicles proposes a plurality of drivable path plans from the start to the goal based on the environmental perception, allowing decision making module to make a selection, which provides all of the appropriate ways to solve the driving mission with the premise of ensured safety.

2.2.2 *Decision-making – optimal control of driving behavior*

When the autonomous vehicle completes its perception of a traffic scene, it can obtain the driving intentions through a motion planning algorithm. Based on the combination of the perception results, the current position of the autonomous vehicle, and the driving intention obtained from the motion planning, the self-driving system makes the most reasonable driving choice for an autonomous vehicle. This process is defined as behavior decision. According to the behavior of human drivers, decision-making behavior can be divided into several categories such as the state maintenance, lane change, acceleration and deceleration, waiting, and U-turn, among other factors. In a real traffic scene, the movement of vehicles and pedestrians participating in the traffic has a certain uncertainty, which requires high demands on the autonomous vehicle to accurately determine the traffic situation and make reasonable decisions. The decision-making module of the autonomous vehicle directly affects the control module, determines the movement and driving state of the autonomous vehicle, and is therefore directly related to the driving safety of the autonomous vehicles. How to make decisions effectively and accurately while ensuring safety is the key to the decision-making module of autonomous vehicles. The decision-making algorithm for self-driving can be divided into rule-based decision model and statistical information-based decision model [60].

From the perspective of human driving behavior, the decision-making module can be regarded as a discrete state transition. The rule-based decision-making algorithm uses the behavioral reasoning mechanism to take the autonomous vehicle's perception results and motion planning as input, and outputs

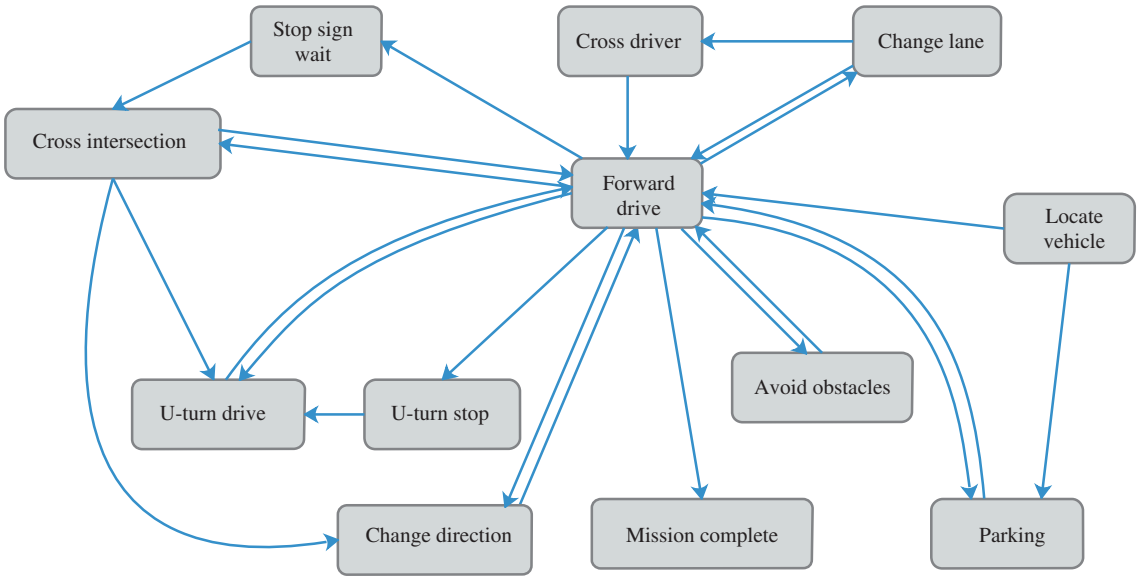

Figure 6 Illustration of decision-making finite state machine. The finite state machine is consist of several driving behavioral states. With the different input information, it defines the transition relationships of the current and the next behavioral state.

the behavior decision of the autonomous vehicle. Finite state machine [61] is a classic representation of the rule-based algorithm, and is a mathematical model that exhibits a set of finite states and state transitions. State transition is determined by the input information and the current state of the state machine. The use of the finite state machine for autonomous vehicle behavior decision-making not only conforms to the logic of human driving, but also achieves high computational efficiency and allows decisions to be made in real time. Montemerlo et al. [62] divided the driving state of autonomous vehicles into 13 categories and formed a finite state machine to switch between different normal driving scenes and states, and effectively prevented the autonomous vehicle from falling into an abnormal state. Figure 6 shows a finite state machine that makes driving behavior decisions consisting of driving behavior states and transition relationships. Rule-based decision-making algorithms like finite state machine show the promising performance in most of the scenarios, yet they still need some improvement in spacial occasions.

Because the motion state of vehicles and pedestrians in real traffic scenes has particular uncertainties, the statistical-based decision-making model often uses the Markov decision process [63] to deal with the behavior decision-making problem of autonomous vehicles. The Markov decision process is applicable to the decision-making process with a limited set of actions in a partially random and partially controllable state. It is the process of making decisions on stochastic dynamic systems with the Markov property through current state observations. Markov property, also known as the state transition probability with no aftereffect, can be interpreted as the conditional probability distribution of the future state, which is only related to the current state, and is independent of the past state. Ulbrich et al. [64] used the partially observable Markov decision process (POMDP) to make real-time decisions on autonomous vehicles, ensuring the real-time predictability of the decision-making modules. Brechtel et al. [65] proposed a combination of the Markov decision process and continuous state hierarchical Bayesian transformation model to solve the self-driving behavioral decision-making by considering the time prediction, environmental perception, sensor noise and uncertainty of partial occlusions. Because the Markov decision process is optimized using the reward function, combined with reinforcement learning, the Markov decision process can be solved by gaining the optimal reward function value [66]. Except this, Morton et al. [67] proposed a LSTM for the modeling of driver behavior.

Autonomous vehicle control needs to follow the designed motion planning trajectory at the expected longitudinal and lateral speeds according to the results of the behavior decision-making. The control module serves as the interface between the "perception, planning, decision-making, and control" data-

driven framework and the physical model of the autonomous vehicle. A smooth and precise control is a guarantee for the safe and comfortable driving of an autonomous vehicle. The control module is composed of lateral and longitudinal control. The former controls the steering of the autonomous vehicle through the steering wheel angle. The latter controls the speed of the autonomous vehicle by adjusting the throttle and brake. According to different control schemes, autonomous vehicle control is mainly divided into proportional-integral-differential (PID) control, pure pursuit and model predictive control algorithms.

The PID control algorithm is a classical algorithm used in the control field. It consists of a proportionalizer (P), an integrator (I), a differentiator (D), and a feedback structure. Beyond the proportional sub-module, the PID control eliminates the static deviation through the integrator and the differentiator accelerates the response speed of the system, which allows the controller to achieve a real-time performance. In general, the lateral and longitudinal control of the autonomous vehicle adopt two independent PID controllers. The lateral controller controls the lateral deviation through the steering wheel angle, and the longitudinal controller controls the longitudinal deviation through the driving speed of the autonomous vehicle. PID-based autonomous vehicle control has a wide range of applications owing to its low resource consumption and relatively simple architecture. Because these methods do not consider the road or kinematics model of the vehicle, they are prone to oscillation. As shown in Figure 7, Xu et al. [68] integrated multiple typical and effective controllers, proposed a control computing framework, which uses the adaptive parameter method reducing the sensitivity of the controller's parameters.

By imitating the driving behavior of a human driver, the pure pursuit algorithm [69] acquires the control signal by setting a lookahead distance and calculating the deviation between the current position of the autonomous vehicle and the desired position, which allows the autonomous vehicle to track the path through the pursuing point. The algorithm takes the midline of the rear axle of the autonomous vehicle as the tangent point, and the longitudinal direction of the vehicle body as the tangent line to control the steering angle of the autonomous vehicle, which enables the autonomous vehicle to track a curved path to the pursuing point. Through a continuous iteration of the two steps of the pursuing point selection and the autonomous vehicle following, the pure pursuit algorithm can complete the control of the autonomous vehicle in a smooth manner. The lookahead distance create the key parameters of the pure pursuit algorithm, which is generally a fixed value, the flexibility of the algorithm needs to be further improved.

Model-based control [70,71] uses system state prediction, online optimization, and feedback correction to further improve the robustness and stability of autonomous vehicle control. The model predictive control predicts the state change in state of the autonomous vehicle within a certain period of time according to the current speed, pose and other states of the autonomous vehicle, and uses the optimization to adjust the input control signal during the period to allow the deviation between the real autonomous vehicle motion state and the expectation of the motion state to achieve the smallest value. During the next period, the control algorithm re-executes the autonomous vehicle motion state estimation and optimization process. Model predictive control has better robustness and control effect, and the algorithm can effectively deal with the constraints of controlled variables and carry out nonlinear programming. Therefore, model predictive control is also the mainstream algorithm of current autonomous vehicle control.

3 Multi-source information fusion and event-driven situational cognition

3.1 Multi-source information fusion

Multi-source information fusion combines and verifies the observation information of different data sources in a certain manner, and finally obtains a consistent interpretation of the observed objects. For different information sources, the form of the data and content are different, and some sources may even produce contradictory representations of the observed objects. Therefore, organically combining the perceptual results of various information sources to generate cognition of the observed objects is still a problem that

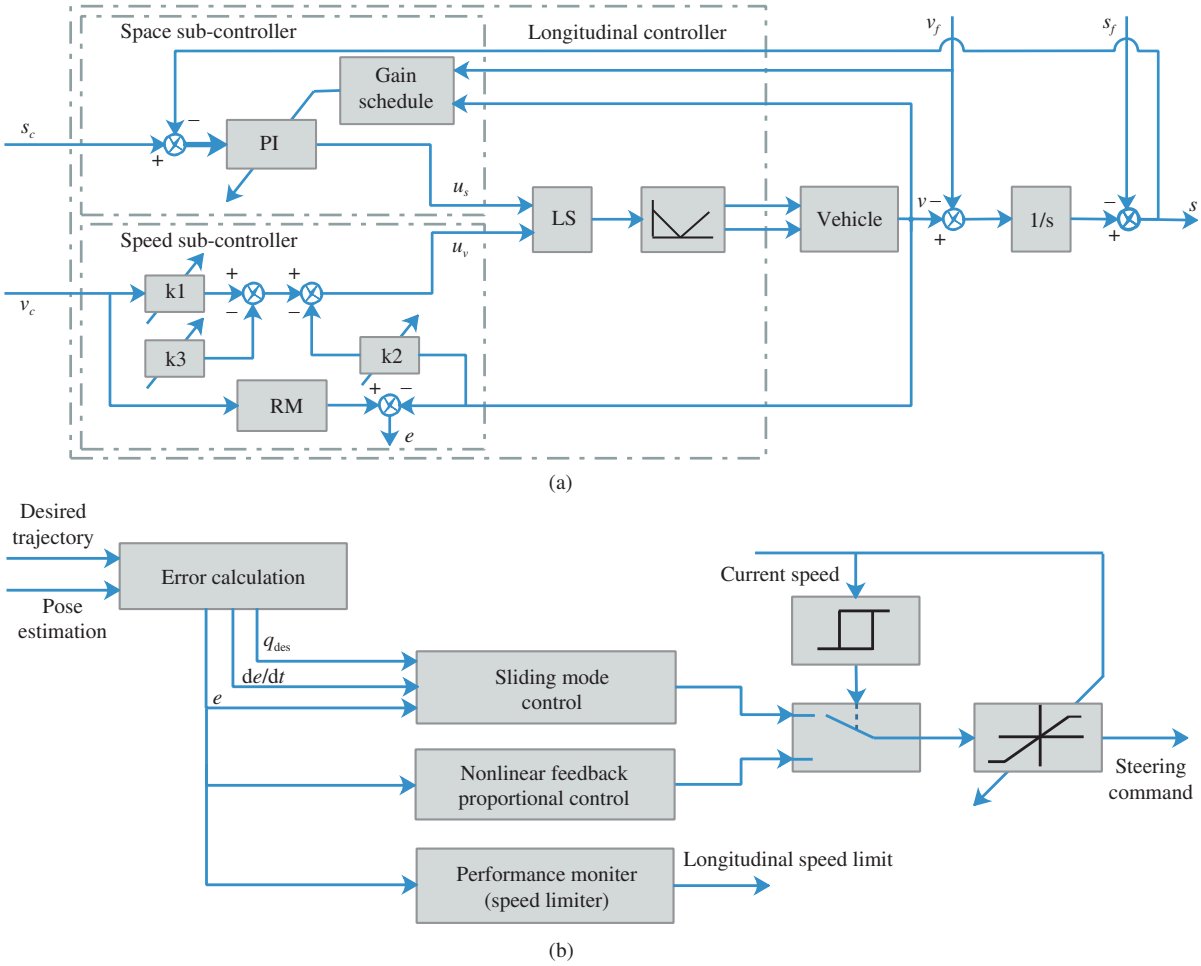

Figure 7 Longitudinal and lateral control computing framework. (a) describes the longitudinal control computing framework, which consists of speed sub-controller and space sub-controller. Speed controller takes speed as the input and transforms the command to the control of throttle and brake, whereas the space-sub controller acts as a basic distance insurance (autonomous emergency braking). (b) shows the details of the lateral control computing framework. It uses the sliding mode control and nonlinear feedback proportional control with an amplitude limited unit to obtain the steering command, and the performance monitor is operated as the constraint of the longitudinal speed.

has not yet to be solved.

Autonomous vehicles need to process observations from different modalities and observation angles, and then convert the observations into the perception of the traffic scenarios. Although there is an inevitable correlation between such data, there are also large differences. How to integrate the information from different sources and realize the complementary advantages between them is a key issue that needs to be urgently solved. A real dynamic traffic scene is very complicated. The main problems of the autonomous vehicles sensor in perceiving the traffic scene are as follows.

(1) Complex illumination and weather conditions. The various illumination conditions and corresponding scene changes, such as backlighting, darkness, mottled roads, and access tunnels, will produce significant disturbances in the image obtained by the camera. Poor weather conditions, such as rain, snow, and fog, will interfere with the reception of point cloud data of the LiDAR, which will cause the LiDAR to fail.

(2) Distinguishing road scenes. Road scenes consist of structured and unstructured roads. The roads with obvious boundaries and traffic signs are structural scenes, such as highways, urban roads and other types, whereas unstructured road such as rural dirt roads has no clear boundaries. Allowing autonomous vehicles to drive reasonably and legally according to the established lanes and traffic signs on structured roads, or safely and smoothly on unstructured roads, is another challenge for self-driving cars.

Figure 8 Cognitive process of understanding traffic scenes. It is made up of pre-processing, feature extraction, post-processing such as classification, regression steps and finally obtains the description of the scene.

(3) Interaction with other driving vehicles and pedestrians. During the driving process, the autonomous vehicles will have some interactions with other vehicles and pedestrians, such as transportation, overtaking, and pedestrian avoidance. The movement of other vehicles and pedestrians in open traffic scenes is uncertain, and the accurate acquisition of the vehicle position and speed, and prediction of the pedestrian position, have higher requirements for the perception of autonomous vehicles.

(4) Noise interference and loss of signal. The observation data obtained by a sensor are inevitably interfered with noise, which results in data with low accuracy. For example, the vehicle inertial navigation system (IMU) is often affected by noise and generate cumulative errors. The loss of observation data is also a problem often encountered by autonomous vehicles, because GPS signals are affected by tunnels and other occlusions, resulting in signal loss.

It can be concluded from the above discussion that the traffic scene perception of an autonomous vehicle based on a single sensor has difficulty ensuring robustness of an autonomous vehicle in dynamic and random traffic scene. However, as a system with extremely high security requirements, autonomous vehicles are not allowed to make false judgments. Therefore, in the face of real complex traffic scenarios, autonomous vehicles need to integrate multi-source information, and combine the observation data of various information sources to check and calibrate with each other to complete the robust traffic scene perception and cognition. Multi-source information fusion also provides redundant design for autonomous systems, and thus when individual information sources fail, the system can still operate safely, which increases its fault tolerance. Meanwhile, because the space observed by the information source is not identical, multi-source information fusion can also increase the scope of the system's observations [72]. Multi-source information fusion can not only combine the sensing results of multiple sensors to form a more complete perception of a traffic scene, it can also obtain the understanding of the traffic scene through mutual verification, and finally converts the observation data of the sensor into the interpretation and description of the traffic scene. This process is consistent with human cognition.

3.1.1 *Hierarchical fusion structure of multiple sensors*

Perception is the description of a scene generated by current observations and prior knowledge, whereas cognition is the secondary processing of the perceived results and obtaining a semantic understanding of the observed information. As shown in Figure 8, the cognition process of a traffic scene can be divided into three stages: obtaining the observation data of the environment and performing the corresponding pre-processing, extracting the features capable of characterizing the observed data, and using classification or other operations to generate a description of the traffic scene and semantic cognition based on the features of the extracted observation data and prior knowledge. According to the cognitive stage, multi-source information fusion can be divided into three levels, namely data, feature and decision-level fusion [73].

Data-level fusion combines the observation data from multiple sensors to obtain a higher quality as well as wider range of observations. Analogous to the cognitive process of human beings toward the environment, data-level fusion is similar to the cooperation between two eyes, and the spatial depth information and the object texture information are finally acquired through the combination of homogeneous and different content information. Therefore, data-level fusion requires strict spatio-temporal rectification, and all observation data needs to be strictly synchronized with a good spatial correspondence. The information lost of observed data in data-level fusion is minimal, but the required computing resources are relatively large. Owing to the strict spatio-temporal matching requirements of data-level fusion, general data-level fusion is mainly used for the fusion of homogeneous sensors, such as image fusion, and LiDAR point cloud stitching. Zhang et al. [74] summarized the multi-modality images and multi-focus images

fusion through sparse representation, which is a great reference for data-level fusion.

Feature-level fusion combines the features of sensor observation data, that is, splicing the feature vectors of all observation data to obtain a more complete feature description of the observations, and then applying post-processing such as feature classification. An analogy of feature-level fusion is that when a person obtains the color of an object through their eyes and combines it with the odor obtained by the nose and then uses the brain to achieve the recognition of the object. This is a combination and verification of different types of qualitative information. Feature-level fusion requires an efficient use of the valid features of all observation data, removing redundant features and completing dimension reduction. It requires a relatively low temporal and spatial registration of all observation data, and thus is often used to deal with the fusion of heterogeneous sensors, such as camera and LiDAR fusion based object detection.

Decision-level fusion combines the descriptions of the traffic scenes obtained by the feature extraction and classification of each source. Decision-level fusion is a high-level fusion. This process is similar to the process of comprehensively judging the description of objects obtained by the human brains through various sensory organs. It is based on the perceptual results of observation data, and is synthesized and verified, finally forming the cognition of the observed objects. Due to the inconsistency of the information sources to the observation attributes of the observed objects, the contradiction and conflict of each information source are problems that needs to be solved during the decision-level fusion. Decision-level fusion requires the lowest temporal and spatial registration of the sensors, and can handle asynchronous information of heterogeneous sensors, such as GPS and IMU fusion for localization. Liu et al. [75] utilized the Adaptive Kalman Filter with attenuation factor for GPS and IMU fusion and Significantly reduce the environmental noises, whereas Behrendt et al. [76] used the stereo images and vehicle odometry for traffic sign detection and achieve the state of the art performance.

Based on the above three levels of multi-source information fusion, some methods have been proposed based on multi-level fusion, which enriches the framework of multi-source information fusion. Haberjahn et al. [77] and Scheunert et al. [78] proposed a multi-level multi-source information fusion structure that combines different levels of fusion information to finally gain a perception of the traffic scenarios.

Multi-source information fusion has different requirements for temporal matching according to different fusion levels. For example, decision-level fusion does not require the time of the observed data received by all information sources to be fully aligned. However, spatial matching is a problem that must be solved through multi-source information fusion for each fusion level. Spatial registration is also called spatial calibration, which is the process of obtaining the transformation relationship between information source coordinate systems by the coordinate deviation of the calibration objects under different information sources. Through the use of spatial calibration, different information source coordinates can be unified to a fixed coordinate. Rodríguez-Garavito et al. [79] and Park et al. [80] used the ground to solve the spatial registration of the LiDAR and camera, Wang et al. [81] and Wang et al. [82] used perspective transformation to solve the millimeter-wave radar and image spatial transformation relationship, and Zhu et al. [83] used a reference plane to calculate the extrinsic parameter calibration of the LiDAR. Beyond these methods, Jiang et al. [84] used the parallel features in the road scene and an online search algorithm to solve the spatial geometric transformation of the LiDAR and camera.

3.1.2 *Approaches to multi-source information fusion*

There are different fusion methods used for multi-source information fusion, which can be divided into two categories: methods based on statistics, reasoning and probability, and cognitive-based fusion methods. The former mainly includes Kalman filter [85], particle filter [86], Bayesian theory [87], and DS evidence theory [88–90]. Kalman filtering is an algorithm based on the linear system state equation for the optimal estimation. It uses the current time observation data and the previous time prediction value for the iterative estimation.

When the system is a linear dynamic system and the system noise can be represented by Gaussian white noise, Kalman filtering can provide an optimal estimation in a statistical sense. Using the Kalman filter as

the fusion method, the distributed fusion structure can be realized, thus there is no unnecessary connection between the information sources, thereby enhancing the fault tolerance of the system. Particle filtering implements a state estimation using non-parametric methods and is suitable for the state estimation in nonlinear non-Gaussian environments. Multi-source information fusion based on Bayesian theory first calculates the conditional probability of the observed attributes of each information source, and then calculates the posterior probability of the event. As an extension of the Bayesian theory, D-S theory uses the trust function and evidence synthesis algorithm to judge the possibility of established events.

The representative methods of cognitive-based fusion are fuzzy set theory based on fuzzy logic inference [91], and an artificial neural network [92] method. In recent years, with the rapid development of deep learning, many data fusion algorithms based on convolutional neural networks have also received extensive attention. Convolutional neural networks have powerful nonlinear expression capabilities, which can simulate complex nonlinear mapping relationships. The convolutional neural network extracts the feature descriptions of the observed objects in different information sources through the learning of the network weights and completes the multi-source information fusion. Owing to its powerful feature representation, the fusion level of multi-source information fusion methods based on deep learning is often on feature level. Chen et al. [17] used a front view camera and the bird-eye view and front view of the LiDAR point cloud as the input data, as well as a fusion network to conduct the feature fusion of the observation data, finally obtaining 3D object bounding boxes; Ku et al. [18] used a front view camera and LiDAR in bird-eye view as input, and proposed an object detection algorithm based on feature-level fusion. The multi-source information fusion based on deep learning obtains the spatial transformation relationship between different information sources through learning, and does not require a strict spatial registration, thereby greatly simplifying the processing steps of the information fusion.

3.2 From data-driven perception to event-driven cognition

The current data-driven computing framework for autonomous driving adjusts other modules of an intelligent system through perception-aware clusters of the environment, which cannot fully adapt to highly dynamic and open traffic scenarios with a limited ability to represent a scene. In addition, this computing framework needs to process data from a different modality and different observation angles simultaneously, including plenty of redundant data and noise, which may lead to large computational complexity and a high probability of error. Meanwhile, the current machine learning algorithms are an essential process of optimizing the parameters of nonlinear functions and obtaining optimal parameters through learning. What is found of the algorithm is the pattern correlation between data instead of the reasons for the occurrence and effects of the events or the causality between the events. Therefore, the feature extraction and cost calculation of machine learning are completely different from the concept formation and internal analysis of the human cognition process, which is often based on a semantic interpretation. To obtain cognitive results similar to those of a human, it is necessary for the machine learning model to make the extracted feature capable of a mathematical and semantic explanation, matching the results of a neurophysiological experiment. Based on this, there is an inevitable correlation between the data obtained by multiple sensors because they are different observations of the same scene. And it can not only improve the reliability and efficiency of the computing framework for autonomous driving, but also promote the intelligence of the autonomous vehicle to extract an event model from the data and drive the upper modules by the model. At the same time, the transformation process from a data-driven to an event-driven framework is the process from perception to cognition, which is the result of thinking based on perception as its upper structure for self-driving cars. According to cognition science, cognition is the process of the brain's secondary processing of perceptual information, which forms an abstract description of a scene and an analysis of the information generating a deep understanding of the scene. An autonomous driving system gains semantic information by analyzing the spatio-temporal and causal correlation between the data. On this basis, an event is described, the graph data of the semantic information are constructed, and the transformation from a data-driven to an event-driven framework is completed, as shown in Figure 9. The left side of the figure is the computational framework of the

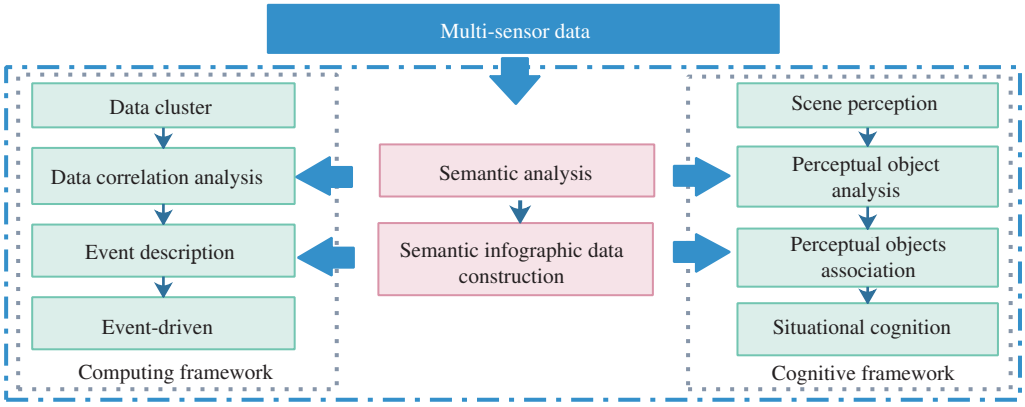

Figure 9 From data-driven scene perception to event-driven situational cognition. This is our assumption for an event-driven model. The computing framework of the model is to extract the events among them through the correlation between the data, and describe the scene through the events. It corresponds to the psychological cognitive process of humans as the cognitive framework's description.

event-driven model, and the right side is the corresponding cognitive process.

In the cognitive framework, the transformation from a data-driven to an event-driven framework is equivalent to a change from scene perception to situational cognition. Different from the scene perception, which represents concrete instances of traffic scenarios at a specific time and in a specific space, situational cognition is a generalization of many specific events within a certain period of time and space, indicating the intertwined factors that constitute and are contained in the scene, forming a cognitive map as an abstraction and interpretation of each instance's relationship in the scene [93]. Situational cognition enables autonomous vehicles to fully and effectively recognize the scene in order to understand the distribution and connection of objects in space and thus make optimal decisions [94,95], and Li et al. [96] formed a driving situation graph cluster for the decision-making module. For example, the establishment of a high-precision map is a manifestation of situational cognition, which is the result of the cognition of traffic scenes. It abstracts and integrates information such as road boundaries, traffic signs, and traffic lights to form an abstract representation of a traffic scene. This abstract representation contains important information needed for autonomous driving, ignoring irrelevant information, and correlating the information to generate prior knowledge. However, because the current high-precision maps are created through manual labeling, autonomous vehicles can only understand static scenes. For dynamic scenes, autonomous vehicles can characterize them through perceptual objects [3], as shown in Figure 9. Perceptual objects are the basic unit of selective attention mechanism in cognitive psychology and one of the basic units of spatial representation, which is defined as the concept of objects with specific topological structures formed in the brain [97]. Figure 10 shows how autonomous vehicles can locate themselves through the perceptual objects in the scene, the localization result is shown in Figure 11. Traffic scenes are described through objects instead of features, which can reduce the noise and help autonomous vehicles to better understand the environment. When decomposing a scene into separated objects, interpreting and reasoning the relationship between them similar to human cognition [98] is the only way for autonomous vehicles to intuitively understand the physical world and generate situational cognition. In addition, through this method, we can greatly reduce the storage capacity of the point cloud map, as shown in the Table 1. The number of point clouds originally used to describe the map is huge. Through our method, only the point cloud after the sampling and the point cloud of perceptual objects are used to describe the map, and the number of point clouds is reduced to about one-thirtieth of the original.

3.3 Situational cognition based on multi-source information fusion

By using multi-source information fusion, the data on different modalities from various observation angles in a traffic scene are correlated and abstracted, which can form a situational cognition of the traffic scene.

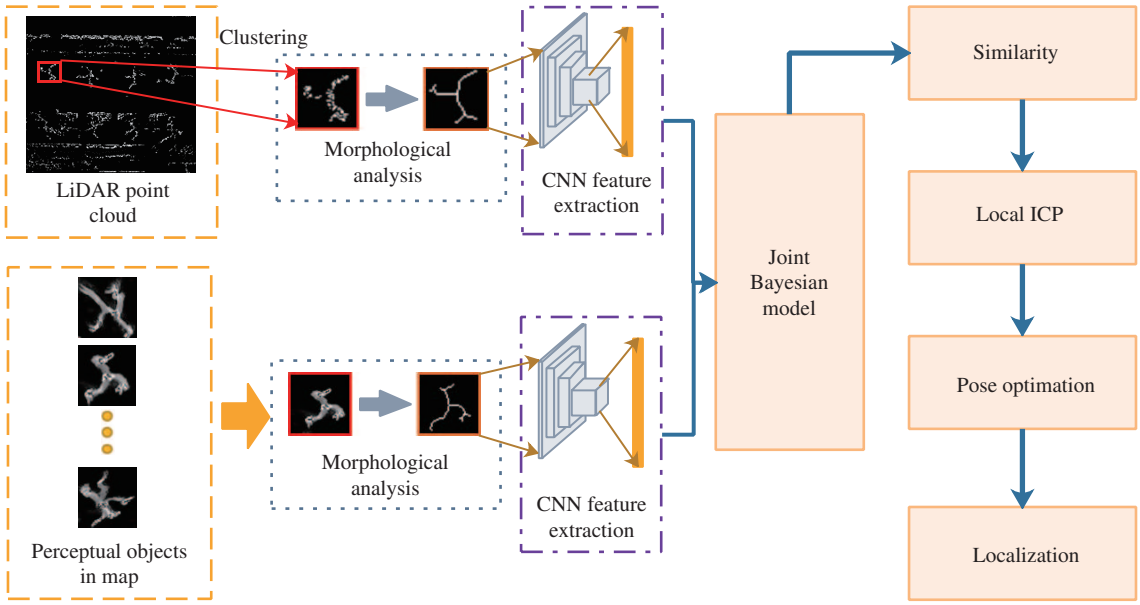

Figure 10 Localization method with perceptual objects. Through the similarity analysis of the perceptual objects in the map and the scene, we can get the correspondence between the map and the point cloud in the scene, so as to obtain the localization information by using the point cloud registration method.

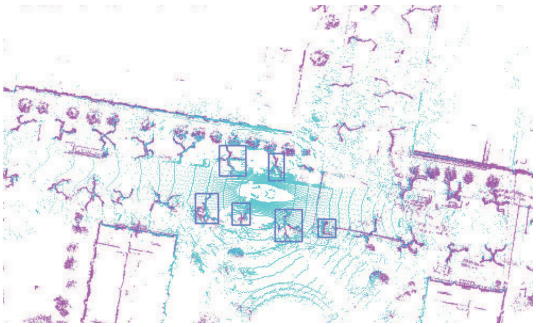

Figure 11 Visualization of the localization results using perceptual objects. The purple point cloud is the map, and the blue point cloud is the scene currently perceived by the vehicle. It can be seen that the two have a high degree of coincidence, which proves the accuracy of the positioning.

Table 1 Number of point clouds for the original map, the sampled map, and the perceived object

Type	Number of points
Raw map	1344843
Sampled map	22483
Perceptual objects	17756

Multi-source information fusion has a wide range of application scenarios in autonomous vehicles, such as object detection and tracking, scene segmentation, and vehicle localization. Through multi-source information fusion, the data from different observation spaces of each information source are correlated to construct an event model, which can achieve event-driven situational cognition.

Object detection and tracking in traffic scenes are significant aspects of the research in autonomous driving. Commonly used sensors include the camera, LiDAR, millimeter wave radar and ultrasonic radar. In recent research, multi-source fusion shows the promising results in the field of object detection. Figure 12 shows a framework of multi-source fusion-based object detection, where multi-source information fusion enables a combination of perceptual results such as texture and category information of the image, the distance and spatial information of the LiDAR point cloud, and the position, linear velocity, and angular velocity information of the radar to obtain more complete and accurate object state information.

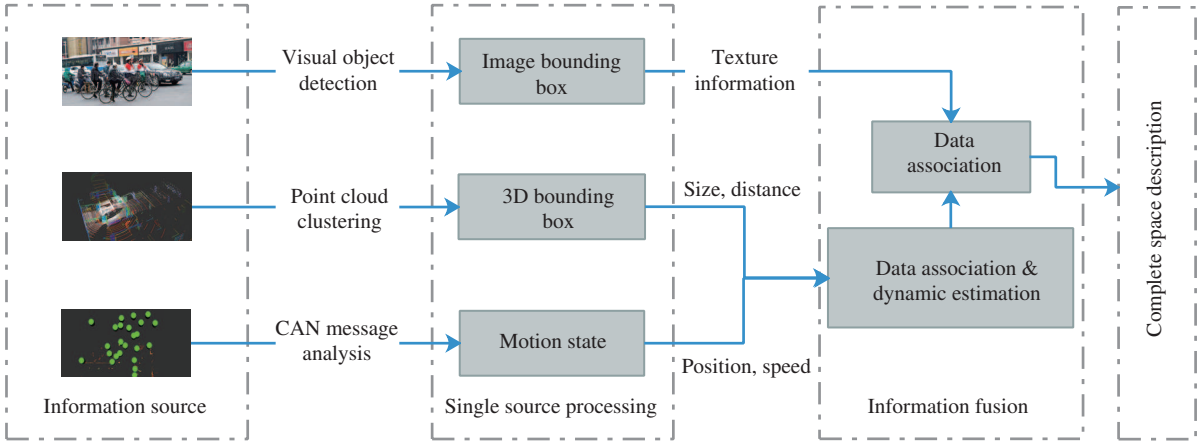

Figure 12 Decision-level multi-source fusion-based object detection framework. It is made up of single source processing and information fusion stages, where each single-source sensors obtain the object separately at the single source processing step and followed by the fusing process which finally obtains the complete description of the scene.

Based on this, through the analysis of the state of the object, the motion trajectory and motion state of the object over a certain period can be obtained, achieving a stable tracking of the object. An analysis of motion information from a single-frame objects motion state to the objects state during a certain period is the transition process from scene perception to situational cognition. Because decision-level fusion requires the low transmission bandwidth and can handle the fusion of asynchronous and heterogeneous information fusion, decision-level fusion has great advantages in the field of object detection and tracking.

For object detection and tracking, decision-level fusion is divided into object and tracking-level fusion. Object level fusion uses a single source for object recognition, and then applies the fusion algorithm to combine the object detection results of each information source for object tracking. The tracking-level fusion is based on the perceptual results of object identification and tracking through a single information source, and the tracking results of each information source are merged, finally forming the recognition of the object.

Multi-source information fusion needs to analyze the association of each information source, that is, the problems of observation data matching from different sources. Bar-Shalom et al. [99] and Svensson et al. [100] proposed a joint probability model for the object data association, whereas Blackman [101] and Kim et al. [102] proposed a data association algorithm based on multiple hypothesis tracking (MHT). In addition, the Hungarian algorithm [103], which solves the problem of the linear assignment, is also a common algorithm for solving data associations between information sources.

Object detection and tracking algorithms based on multi-source information fusion have been widely used in the field of autonomous driving. Cho et al. [104] and Chavez-Garcia et al. [105] proposed a multi-source information fusion object tracking algorithm based on LiDAR, radar and camera; Göhring et al. [106] proposed a vehicle follow-up system based on the fusion of radar and LiDAR; Fayad et al. [107] and Kim et al. [108] used a decision-level based multi-sensor fusion algorithm for pedestrian detection and object tracking, and Govaers et al. [109] developed an object detecting and tracking algorithm based on tracking level fusion and distributed Kalman scheme to obtain the optimal tracking results.

Scene segmentation distinguishes the observation data in the traffic scene to complete the distinction between different objects within the scene, which is the key to understanding a traffic scene. The semantic segmentation of traffic scenes is used to distinguish pixel category information of different categories in an image. In recent years, instance segmentation has put forward new requirements for scene segmentation. The instance segmentation needs to complete the object detection, semantic segmentation and classification tasks at the same time. It not only distinguishes the categories of objects in a traffic scene, it also needs to distinguish different instances in the basic categories. Zhang et al. [110] performed instance segmentation using images and achieveed good results on the KITTI dataset. By fusing the image texture, distance and reflectivity information of the LiDAR point cloud data can complete the description of the

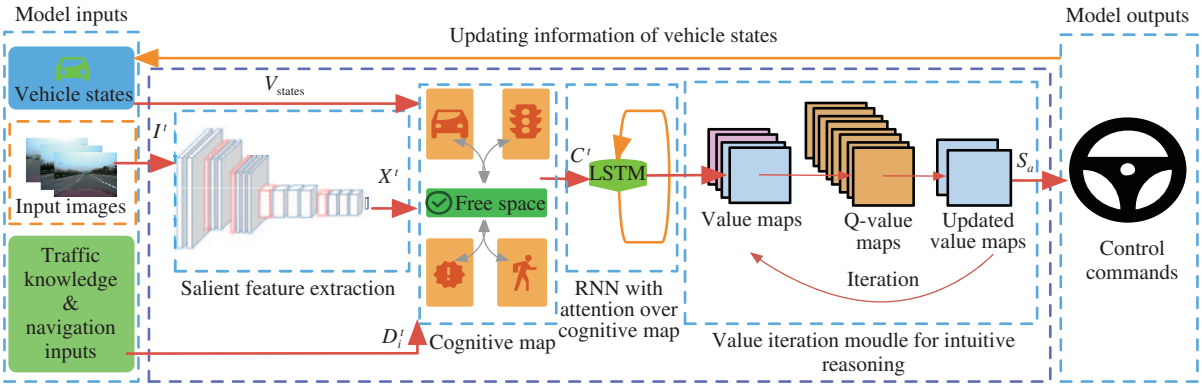

Figure 13 Cognitive computing framework of self-driving cars with selective attention mechanism and intuitive reasoning. The cognitive process is divided into four parts. The first part is the feature extraction by convolutional neural network, and the second part is the combination of features and prior knowledge to form a cognitive map. The third part is to filter the cognitive map by implementing the attention mechanism through LSTM. The last part is to reflect the cognitive map to the behavior space through the value iteration model, and output the final behavior decision.

traffic scene from a higher dimension, and thus a more accurate segmentation result can be obtained. Also, this kind of fused RGB-D data has a promising future in the field of SLAM [111, 112].

Localization is the key to determining whether an autonomous vehicle can operate safely on a road. Integrating information from GPS, IMU and wheel speed sensors can not only allow cumulative errors of the wheel speed sensor and IMU to be avoided, it also enables an autonomous vehicle to drive smoothly when the GPS signal is blurred or even completely lost, which greatly increases the robustness and fault tolerance the localization of the autonomous vehicles. Caron et al. [113] used Kalman filter to combine GPS and IMU to obtain more accurate localization information for autonomous vehicles. The spatial information obtained from the fusion of GPS and IMU is global information, which only describes the spatial location without describing the relationship with the traffic scene. By integrating the data of the LiDAR point cloud and the camera image with GPS and IMU, the local localization can be completed by obtaining the global positioning information, and the spatial correlation between the autonomous vehicles and the perceptual objects in the traffic scene is then obtained. This forms a necessary component of the scenario cognition of the traffic scene. Suhr et al. [114] used particle filtering to combine GPS, IMU, wheel speed sensor, camera and digital map, and implemented a low-cost car positioning system on an embedded platform. Wan et al. [115] used GNSS and IMU for integration with LiDAR information and realized the localization of urban and highway roads, achieving state-of-the-art result.

Multi-source information fusion is used to simulate the human cognition process. By combining and verifying the scene perception results of homogeneous and heterogeneous information sources, the situational cognition of traffic scenes is finally formed. Therefore, with multi-source information fusion, the autonomous vehicle can not only fully and reasonably utilize the observation data from autonomous vehicle sensors, it also obtains a deep understanding of the traffic scene.

4 Cognitive computing framework of autonomous vehicle based on selective attention model and event-driven mechanism

The event-driven computing framework for autonomous driving includes a selective attention mechanism and intuitive reasoning in different traffic scenarios. Based on this, the present study proposes a cognitive computing model, as shown in Figure 13. In this model, the convolutional neural network extracts the feature of the traffic scene to form a cognitive map. The cognitive map contains descriptions of various objects in the traffic scene and prior information such as the vehicle states and traffic knowledge. Second, a recurrent neural network is used to construct the attention model of the driving behavior, learn the attention mechanism of human drivers on traffic scenes, and extract the necessary pivotal information from cognitive maps. The proposed pivotal information is updated through reinforcement learning of the

value map in the value iteration model [116], and a control signal of the vehicle actuator is finally formed according to the value map.

4.1 Selective attention mechanism of traffic scene perception

Attention is a cognitive process that selectively focuses on gaining certain aspects of information to weaken or ignore other aspects. The human brain has limited processing power and resources, but the amount of information obtained through the sensory organs is enormous, and thus attention is a selective mechanism for humans to efficiently manage the information they acquire. In the current research on data-driven self-driving algorithms, the processing of data such as images and point clouds in a traffic scene generally provides the same weight during the initial processing stage, which means processing the acquired data indiscriminately. Although computers have powerful computing resources and sufficient capability to process all received data, paying attention to all clues from a traffic scene may cause the "behavior" to collapse. Thus, by introducing the attention model, it is possible to ignore irrelevant information or clues and focus the calculations on information and clues associated with driving behavior decisions. Imitating the human attention mechanism, the attention model gives different weights and attention to the data through an analysis, thereby accomplishing the directional perception and processing of the received data. This orientation perception and processing includes the semantic description of the perceptual results of the traffic scene and is the basis for the "next" driving intention. It is reprocessed to form a part of the perception of the traffic scene.

Attention is divided into bottom and top-down attention [117, 118]. Bottom-up attention is the direct response to external information features such as color, shape, and other stimuli, driven by external events. Top-down attention is based on the established goals and filtered information using prior knowledge, which is driven by internal attention. General attention studies focus on the study of the bottom-up attention mechanism, whereas research on the top-down attention mechanism is relatively rare owing to insufficient recognition of the attention mechanism. Bottom-up attention research focuses on the study of salient feature models. The human visual system can quickly capture a target with significant features in an acquired image. Based on this, the salient feature detector constructs the visual attention algorithm by constructing the salient features in the image. Itti et al. [119] proposed a visual structure significance model, which divides an image into intensity, color, and direction channel, and generates feature maps by different scale operators. The resulting feature maps are combined to obtain a significant visual feature map. The local entropy model [120], multi-scale quaternion Fourier transform [78], and other significant feature-based algorithms also effectively represent the visual attention model.

Attention mechanism has a wide range of applications in deep learning. Mnih et al. [121] used an attention model to ignore unrelated objects, thereby enabling precise multi-object detection and classification tasks in the presence of noise. Hu et al. [122] considered the structure of the convolutional neural network from the perspective of the feature channel, and proposed SeNet based on the Excitation and Squeeze module. The network structure explicitly establishes the dependencies between feature channels and uses a learning method to obtain the weight of each channel. It enhances the effective features and suppresses invalid features according to the important features. Fu et al. [123] proposed a network structure based on the recursive attention mechanism. The network realizes the judgment of multi-scale discriminative regions through the attention model and enables the network to automatically find the most discriminative regions and classify them. The network has achieved satisfactory results in the classification of refined objects. The video question and answer (VQA) problem takes the image and text questions as input, and outputs the answer to the question. It requires a semantic level understanding of the content of the images in the video. Jang et al. [5] used a temporal and spatial based attention model to understand the semantics of video content and achieve satisfactory results. The traffic scene where an autonomous vehicle is located is complicated and contains noise, whereas the information of certain scenes, such as traffic signs and the preceding vehicles on the current lane are of more significance. Through the attention model, the noise interference can be effectively avoided, and the robustness of the significant area recognition can be increased.

Therefore, introducing the attention mechanism in the self-driving computing framework can enhance the processing of pivotal information in traffic scenarios and suppress the interference caused by non-critical information, which can greatly strengthen the robustness and accuracy of the perceptual results, and form the cognition of a traffic scene.

4.2 Intuitive reasoning for traffic situation understanding based on reinforcement learning and transfer learning

It is necessary for autonomous vehicles to achieve the ability of intuitive reasoning. Intuitive reasoning and logical reasoning are two important ways for humans to make cognitive decisions regarding the objective world, which complement each other and can greatly improve the event response and decision-making ability of autonomous vehicles. Kahneman et al. [4] proposed the two-system theory in 2002, arguing that intuition and logical reasoning are two different cognitive systems. Intuitive reasoning has the characteristics of simplicity, spontaneity, rapid parallelism and skill, whereas the characteristics of logical reasoning are controllable, complex, deductive, slow, continuous and regular. The characteristics of intuition and logical reasoning make the former more suitable for solving empirical problems, whereas the latter is suitable for dealing with logical problems and correcting when intuitively cognitive impairment occurs. Kuo et al. [124] showed that the neurons in the insula and anterior cingulate cortex of the brain are more active during intuitive reasoning, wheras the median, inferior parietal, and anterior torso are more active in logical reasoning, demonstrating the independence of intuition and logical reasoning systems from the perspective of neuroscience. Zheng et al. [125] also pointed out that a human's intuitive reasoning is closely related to the abstract ability and strong generation ability of the human brain to prior knowledge, rather than simple mechanical memory. It was further pointed out that human intuition can rapidly allow decisions to be made with regard to risk avoidance based on the world model of the human brain owing to this high degree of generalization. The authors [125] also provided a basic method of computational realization of intuitive reasoning. In conclusion, it is insufficient to rely on logical reasoning for autonomous vehicles to drive in open traffic scenarios, and intuitive reasoning must therefore be introduced into the computing framework of self-driving cars.

Human intuitive reasoning can be seen as the initial iteration position to find the global optimal solution in the problem solution space. Because the solution space is always complex, non-convex, or even unable to be structurally described in practical problems, the selection of the initial iteration position is critical, and directly determines whether the final iteration result is globally optimal. Therefore, a traditional machine learning algorithm tends to converge to a local minimum in this case.

In order to realize intuition in intelligent systems, methods based on reinforcement learning have been proposed and have had a significant impact on the field of artificial intelligence. Reinforcement learning feeds back the decision-making of the agent through a reward and punishment mechanism, and continuously strengthens the correctness of the agent's decision-making through training in order to form an intuitive response to a specific task. In the case of ignoring the neuron activity process of intuition in the brain, the framework of reinforcement learning is constructed directly according to the formation of intuition and has shown expected results. Figure 14 shows the method of vehicle following implemented through reinforcement learning [126]. This method enables autonomous vehicles to follow their front vehicle according to the output of reinforcement learning model.

On this basis, there has been a huge breakthrough in the artificial intelligence field, which is called deep reinforcement learning. Deep reinforcement learning strengthens the experience buffer for the corresponding scene through the reward mechanism to analyze and understand a similar scene. The key to deep reinforcement learning is to adjust the parameters of the deep network through the experience in the memory buffer to obtain the optimal strategy. Mnih et al. [127] proposed a deep reinforcement learning framework named DQN, which successfully achieved expert-level performance on the Atari2600 video game. In addition, Silver et al. [128] designed a system that defeated human players in the game 'Go' with a similar framework in 2016, which fully proves that deep reinforcement learning can achieve human-level intelligence when the state space is limited. In 2017, Gupta et al. [129] proposed a method

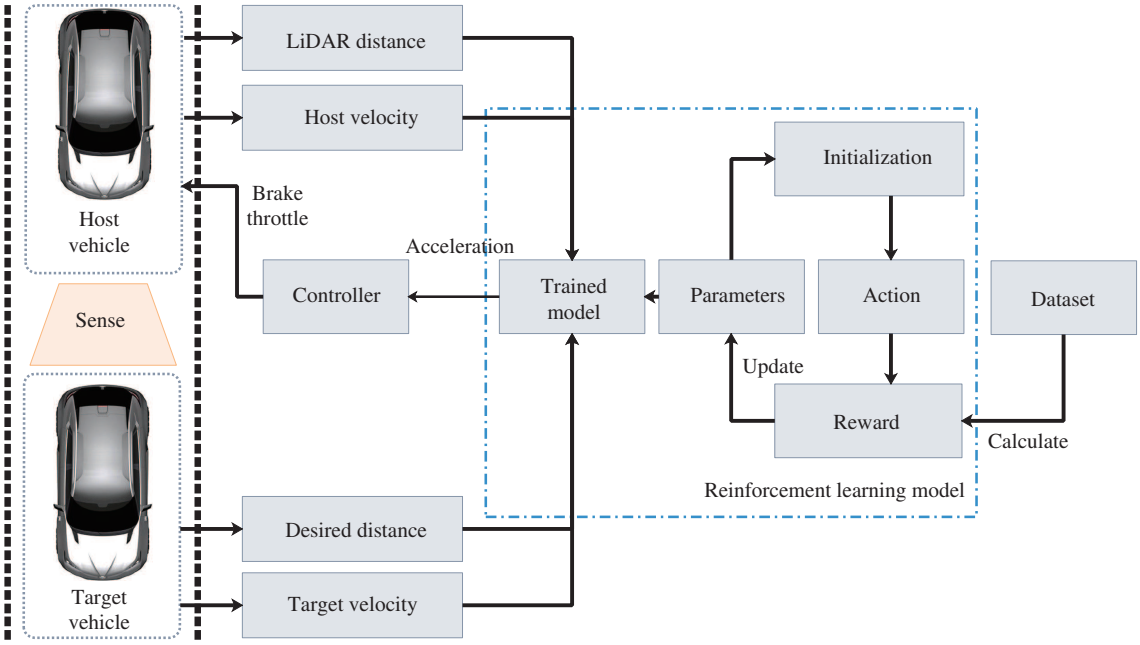

Figure 14 Reinforcement learning framework for adaptive cruise control. The desired speed and distance and the current speed and distance are input into the trained reinforcement learning model, and the acceleration is output to the control module to achieve vehicle following.

for indoor robot mapping and navigation through deep reinforcement learning, showing the adaptability of deep reinforcement learning in semi-open scenes. In an open scene, optimizing the decision-making of autonomous vehicles through the deep reinforcement learning framework is an important orientation of current autonomous driving technology research. It is essential for autonomous vehicles to achieve universal intuition in open traffic scenarios.

In addition, studies on artificial intelligence have also been inspired by the way in which intuitive reasoning is applied to solve problems. Humans usually apply the general knowledge gained in certain environments to new, previously unknown fields. This way of reasoning through relevance is called transfer learning in the field of artificial intelligence. The current research on transfer learning frameworks for strong generalization capabilities is ongoing. One type of architecture called progressive network, can use the knowledge gained in one game for other games, greatly reducing the learning time [130]. It has been successfully applied to the quick transfer of the knowledge of agents in a simulated environment to real robotic arms, with progress made in the transfer learning of intelligent systems [131]. Transfer learning has two areas of significance for autonomous vehicles. First, transfer learning can apply the knowledge learned by the vehicles in a simulation environment to an actual scene, thus simplifying the learning process and improving the learning efficiency. Second, transfer learning can help autonomous vehicles achieve perception with a strong generalization ability and adapt to high dynamic scenarios. Extending the model trained by deep learning in a simulation environment to a real scene based on transfer learning is one of the methods used to realize intuitive reasoning for autonomous vehicles, which is shown in Figure 15. Through this framework, autonomous vehicles can adapt what they learn in the simulation environment to the real traffic scenarios and better update their knowledge.

5 Conclusion

Artificial intelligence is profoundly changing the world, and autonomous vehicles will become close companions of human in the future. It is an important and promising direction to explore the brain-inspired self-driving technology for the new generation of artificial intelligence. To process and understand the data from sensors with reference to the cognitive psychological level of human driving process can greatly

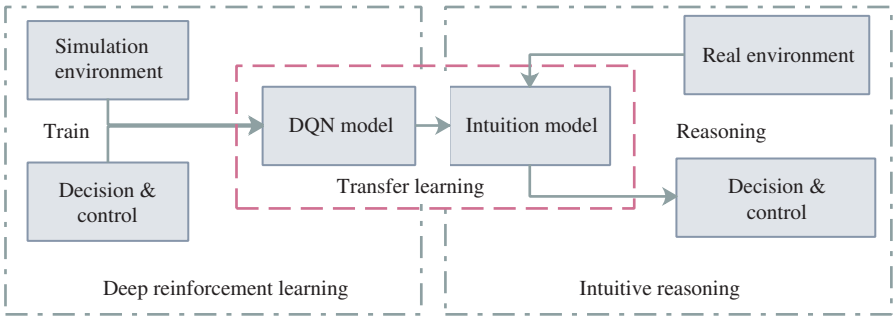

Figure 15 The generation of autonomous vehicle's intuition based on "reinforcement-transfer" learning. The autonomous driving model is trained in the simulation environment using the deep reinforcement learning, and then the model is transplanted to autonomous vehicles through migration learning.

improve the cognizing ability, decision-making ability and adaptability to complex situations of self-driving system. A self-driving system based on cognitive construction enables autonomous vehicles to push themselves to higher levels of intelligence through intuitive reasoning and empirical learning. Based on a review of the development of current self-driving technology and challenges it faces, this study deeply discusses some basic scientific issues of the self-driving approach based on cognitive construction, as well as the methods, computing models and technical routes to solve these problems. Furthermore, we expound the important role of selective attention and event-driven mechanism in the realization of robust cognitive computing in complex traffic situation. And the intuitive reasoning self-driving method based on reinforcement learning and transfer learning is also discussed in this study. It is no doubt that realizing autonomous driving is an exciting but daunting challenge. We hope that industry and academia may work together to enrich the development and practice of the self-driving technology and advance it greatly.

Acknowledgements This work was partially supported by National Natural Science Foundation of China (Grant Nos. 61773312, 61790563).

References

1 Thrun S, Montemerlo M, Dahlkamp H, et al. Stanley: the robot that won the DARPA grand challenge. J Field Robot, 2006, 23: 661–692
2 Miller G A. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev, 1956, 63: 81–97
3 Kahneman D, Treisman A, Gibbs B J. The reviewing of object files: object-specific integration of information. Cogn Psychol, 1992, 24: 175–219
4 Kahneman D, Frederick S. Representativeness revisited: attribute substitution in intuitive judgment. In: Heuristics and Biases: the Psychology of Intuitive Judgment. New York: Cambridge University Press, 2002. 49–81
5 Jang Y, Song Y, Yu Y, et al. TGIF-QA: toward spatio-temporal reasoning in visual question answering. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), Honolulu, 2017. 2758–2766
6 Ojala T, Pietikainen M, Harwood D. Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In: Proceedings of the 12th IAPR International Conference on Pattern Recognition, Jerusalem, 1994. 582–585
7 Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on featured distributions. Pattern Recogn, 1996, 29: 51–59
8 Zhao L, Thorpe C E. Stereo- and neural network-based pedestrian detection. IEEE Trans Intell Transp Syst, 2000, 1: 148–154
9 Yuan Y, Xiong Z T, Wang Q. An incremental framework for video-based traffic sign detection, tracking, and recognition. IEEE Trans Intell Transp Syst, 2017, 18: 1918–1929
10 Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014. 580–587
11 Girshick R. Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, 2015. 1440–1448
12 Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks. In: Proceedings of the 28th International Conference on Neural Information Processing Systems, Montreal, 2015. 91–99
13 Liu W, Anguelov D, Erhan D, et al. Ssd: single shot multibox detector. In: Proceedings of European Conference on Computer Vision. Berlin: Springer, 2016. 21–37

14 Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. 779–788
15 Redmon J, Farhadi A. Yolo9000: better, faster, stronger. 2017. ArXiv: 1612.08242
16 Wu B C, Iandola F, Jin P H, et al. Squeezedet: unified, small, low power fully convolutional neural networks for real-time object detection for autonomous driving. In: Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, 2017
17 Chen X Z, Ma H M, Wan J, et al. Multi-view 3D object detection network for autonomous driving. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. 3
18 Ku J, Mozifian M, Lee J, et al. Joint 3D proposal generation and object detection from view aggregation. 2017. ArXiv: 1712.02294
19 Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? the kitti vision benchmark suite. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012. 3354–3361
20 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015. 3431–3440
21 Badrinarayanan V, Kendall A, Cipolla R. Segnet: a deep convolutional encoder-decoder architecture for image segmentation. 2015. ArXiv: 1511.00561
22 Chen L-C, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected crfs. 2014. ArXiv: 1412.7062
23 Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans Pattern Anal Mach Intell, 2018, 40: 834–848
24 Wang Q, Gao J Y, Yuan Y. A joint convolutional neural networks and context transfer for street scenes labeling. IEEE Trans Intell Transp Syst, 2018, 19: 1457–1470
25 Oliveira G L, Burgard W, Brox T. Efficient deep models for monocular road segmentation. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2016. 4885–4891
26 Dai J F, He K M, Sun J. Instance-aware semantic segmentation via multi-task network cascades. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. 3150–3158
27 He K M, Gkioxari G, Dollár P, et al. Mask r-cnn. In: Proceedings of IEEE International Conference on Computer Vision (ICCV), 2017. 2980–2988
28 Li Y, Qi H Z, Dai J F, et al. Fully convolutional instance-aware semantic segmentation. 2016. ArXiv: 1611.07709
29 Bosse M, Zlot R. Continuous 3D scan-matching with a spinning 2D laser. In: Proceedings of IEEE International Conference on Robotics and Automation, 2009. 4312–4319
30 Baldwin I, Newman P. Laser-only road-vehicle localization with dual 2D push-broom lidars and 3D priors. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2012. 2490–2497
31 Pfrunder A, Borges P V K, Romero A R, et al. Real-time autonomous ground vehicle navigation in heterogeneous environments using a 3D lidar. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2017. 2601–2608
32 Liu Z Y, Yu S Y, Wang X, et al. Detecting drivable area for self-driving cars: an unsupervised approach. 2017. ArXiv: 1705.00451
33 Satzoda R K, Sathyanarayana S, Srikanthan T, et al. Hierarchical additive hough transform for lane detection. IEEE Embedded Syst Lett, 2010, 2: 23–26
34 Huang Y H, Chen S T, Chen Y, et al. Spatial-temproal based lane detection using deep learning. In: Proceedings of IFIP International Conference on Artificial Intelligence Applications and Innovations, 2018. 143–154
35 Lee S, Kweon I S, Kim J, et al. Vpgnet: vanishing point guided network for lane and road marking detection and recognition. In: Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV), 2017. 1965–1973
36 Pan X G, Shi J P, Luo P, et al. Spatial as deep: spatial CNN for traffic scene understanding. 2017. ArXiv: 1712.06080
37 Zhang G, Zheng N N, Cui C, et al. An efficient road detection method in noisy urban environment. In: Proceedings of 2009 IEEE Intelligent Vehicles Symposium. New York: IEEE, 2009. 556–561
38 Lv X, Liu Z Y, Xin J M, et al. A novel approach for detecting road based on two-stream fusion fully convolutional network. In: Intelligent Vehicles. New York: IEEE, 2018
39 Chen Z, Chen Z J. Rbnet: a deep neural network for unified road and road boundary detection. In: Proceedings of International Conference on Neural Information Processing. Berlin: Springer, 2017. 677–687
40 Munoz-Bulnes J, Fernandez C, Parra I, et al. Deep fully convolutional networks with random data augmentation for enhanced generalization in road detection. In: Proceedings of IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), 2017. 366–371
41 Lv X, Liu Z Y, Xin J M, et al. A novel approach for detecting road based on two-stream fusion fully convolutional network. In: Proceedings of 2018 IEEE Intelligent Vehicles Symposium (IV). New York: IEEE, 2018. 1464–1469
42 Warren C W. Fast path planning using modified A* method. In: Proceedings of IEEE International Conference on Robotics and Automation, 1993. 662–667
43 Zeng W, Church R L. Finding shortest paths on real road networks: the case for A*. Int J Geographical Inf Sci, 2009, 23: 531–543
44 Šišlák D, Volf P, Pěchouček M. Accelerated A* path planning. In: Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems, Budapest, 2009. 1133–1134
45 LaValle S M. Rapidly-Exploring Random Trees: a New Tool for Path Planning. Technical Report (TR 98-11), Iowa State University, 1998

46 Kuffner J J, LaValle S M. Rrt-connect: an efficient approach to single-query path planning. In: Proceedings of IEEE International Conference on Robotics and Automation, 2000. 995–1001
47 Bohlin R, Kavraki L E. Path planning using lazy prm. In: Proceedings of IEEE International Conference on Robotics and Automation, 2000. 521–528
48 Barraquand J, Langlois B, Latombe J C. Numerical potential field techniques for robot path planning. IEEE Trans Syst Man Cybern, 1992, 22: 224–241
49 Yang S X, Luo C. A neural network approach to complete coverage path planning. IEEE Trans Syst Man Cybern B, 2004, 34: 718–724
50 Ferrer G, Sanfeliu A. Bayesian human motion intentionality prediction in urban environments. Pattern Recogn Lett, 2014, 44: 134–140
51 Ghori O, Mackowiak R, Bautista M, et al. Learning to forecast pedestrian intention from pose dynamics. In: Proceedings of 2018 IEEE Intelligent Vehicles Symposium (IV), 2018
52 Ma W-C, Huang D-A, Lee N, et al. Forecasting interactive dynamics of pedestrians with fictitious play. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. 4636–4644
53 Pfeiffer M, Schaeuble M, Nieto J, et al. From perception to decision: a data-driven approach to end-to-end motion planning for autonomous ground robots. In: Proceedings of 2017 IEEE International Conference on Robotics and Automation (ICRA), 2017. 1527–1533
54 Kim B, Kang C M, Lee S H, et al. Probabilistic vehicle trajectory prediction over occupancy grid map via recurrent neural network. 2017. ArXiv: 1704.07049
55 Takahashi A, Hongo T, Ninomiya Y, et al. Local path planning and motion control for agv in positioning. In: Proceedings of IEEE/RSJ International Workshop on Intelligent Robots and Systems, 1989. 392–397
56 Piazzi A, Bianco C G L. Quintic g/sup 2/-splines for trajectory planning of autonomous vehicles. In: Proceedings of the IEEE Intelligent Vehicles Symposium, 2000. 198–203
57 Komoriya K, Tanie K. Trajectory design and control of a wheel-type mobile robot using b-spline curve. In: Proceedings of IEEE/RSJ International Workshop on Intelligent Robots and Systems, 1989. 398–405
58 Holger B, Dennis N, Marius Z J, et al. From G2 to G3 continuity: continuous curvature rate steering functions for sampling-based nonholonomic motion planning. In: Proceedings of Intelligent Vehicles. New York: IEEE, 2018
59 Petereit J, Emter T, Frey C W, et al. Application of hybrid A* to an autonomous mobile robot for path planning in unstructured outdoor environments. In: Proceedings of the 7th German Conference on Robotics, 2012. 1–6
60 Veres S M, Molnar L, Lincoln N K, et al. Autonomous vehicle control systemsa review of decision making. In: Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering, 2011. 225: 155–195
61 Lee D, Yannakakis M. Principles and methods of testing finite state machines-a survey. Proc IEEE, 1996, 84: 1090–1123
62 Montemerlo M, Becker J, Bhat S, et al. Junior: the stanford entry in the urban challenge. J Field Robot, 2008, 25: 569–597
63 Feinberg E A, Shwartz A. Handbook of Markov Decision Processes: Methods and Applications. Berlin: Springer Science & Business Media, 2012
64 Ulbrich S, Maurer M. Probabilistic online pomdp decision making for lane changes in fully automated driving. In: Proceedings of the 16th International IEEE Conference on Intelligent Transportation Systems (ITSC), 2013. 2063–2067
65 Brechtel S, Gindele T, Dillmann R. Probabilistic decision-making under uncertainty for autonomous driving using continuous pomdps. In: Proceedings of IEEE 17th International Conference on Intelligent Transportation Systems (ITSC), 2014. 392–399
66 van Otterlo M, Wiering M. Reinforcement learning and markov decision processes. In: Proceedings of Reinforcement Learning. Belin: Springer, 2012. 3–42
67 Morton J, Wheeler T A, Kochenderfer M J. Analysis of recurrent neural networks for probabilistic modeling of driver behavior. IEEE Trans Intell Transp Syst, 2017, 18: 1289–1298
68 Xu L H, Wang Y Z, Sun H B, et al. Integrated longitudinal and lateral control for Kuafu-II autonomous vehicle. IEEE Trans Intell Transp Syst, 2016, 17: 2032–2041
69 Coulter R C. Implementation of the Pure Pursuit Path Tracking Algorithm. Technical Report, Carnegie-Mellon UNIV Pittsburgh PA Robotics INST, 1992
70 Camacho E F, Alba C B. Model Predictive Control. Berlin: Springer Science & Business Media, 2013
71 Rasekhipour Y, Khajepour A, Chen S K, et al. A potential field-based model predictive path-planning controller for autonomous road vehicles. IEEE Trans Intell Transp Syst, 2017, 18: 1255–1267
72 Varshney P K. Multisensor data fusion. Electron Commun Eng J, 1997, 9: 245–253
73 Hall D L, Llinas J. An introduction to multisensor data fusion. Proc IEEE, 1997, 85: 6–23
74 Zhang Q, Liu Y, Blum R S, et al. Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: a review. Inf Fusion, 2018, 40: 57–75
75 Liu Y H, Fan X Q, Lv C, et al. An innovative information fusion method with adaptive Kalman filter for integrated INS/GPS navigation of autonomous vehicles. Mech Syst Signal Process, 2018, 100: 605–616
76 Behrendt K, Novak L, Botros R. A deep learning approach to traffic lights: detection, tracking, and classification. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA), 2017. 1370–1377
77 Haberjahn M, Kozempel K. Multi level fusion of competitive sensors for automotive environment perception. In: Pro-

ceedings of 16th International Conference on Information Fusion (FUSION), 2013. 397–403
78 Scheunert U, Lindner P, Richter E, et al. Early and multi level fusion for reliable automotive safety systems. In: Proceedings of Intelligent Vehicles Symposium. New York: IEEE, 2007. 196–201
79 Rodríguez-Garavito C H, Ponz A, García F, et al. Automatic laser and camera extrinsic calibration for data fusion using road plane. In: Proceedings of the 17th International Conference on Information Fusion (FUSION), 2014. 1–6
80 Park Y, Yun S, Won C S, et al. Calibration between color camera and 3D LIDAR instruments with a polygonal planar board. Sensors, 2014, 14: 5333–5353
81 Wang X, Xu L H, Sun H B, et al. On-road vehicle detection and tracking using MMW radar and monovision fusion. IEEE Trans Intell Transp Syst, 2016, 17: 2075–2084
82 Wang T, Xin J M, Zheng N N. A method integrating human visual attention and consciousness of radar and vision fusion for autonomous vehicle navigation. In: Proceedings of IEEE 4th International Conference on Space Mission Challenges for Information Technology (SMC-IT), 2011. 192–197
83 Zhu Z, Liu J L. Unsupervised extrinsic parameters calibration for multi-beam lidars. In: Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering, Paris, 2013. 1110–1113
84 Jiang J J, Xue P X, Chen S T, et al. Line feature based extrinsic calibration of lidar and camera. In: Proceedings of 2018 IEEE International Conference on Vehicular Electronics and Safety (ICVES), 2018. 1–6
85 Sun S L, Deng Z L. Multi-sensor optimal information fusion Kalman filter. Automatica, 2004, 40: 1017–1023
86 Särkkä S, Vehtari A, Lampinen J. Rao-blackwellized particle filter for multiple target tracking. Inf Fusion, 2007, 8: 2–15
87 Yang G S, Lin Y, Bhattacharya P. A driver fatigue recognition model based on information fusion and dynamic Bayesian network. Inf Sci, 2010, 180: 1942–1954
88 Li Y B, Chen J, Ye F, et al. The improvement of DS evidence theory and its application in IR/MMW target recognition. J Sens, 2016, 2016: 1–15
89 Wu H D, Siegel M, Stiefelhagen R, et al. Sensor fusion using dempster-shafer theory [for context-aware hci]. In: Proceedings of the 19th IEEE Instrumentation and Measurement Technology Conference, 2002. 7–12
90 Murphy R R. Dempster-Shafer theory for sensor fusion in autonomous mobile robots. IEEE Trans Robot Automat, 1998, 14: 197–206
91 Subramanian V, Burks T F, Dixon W E. Sensor fusion using fuzzy logic enhanced Kalman filter for autonomous vehicle guidance in citrus groves. Trans ASABE, 2009, 52: 1411–1422
92 Klein L A, Klein L A. Sensor and data fusion: a tool for information assessment and decision making. In: Proceedings of SPIE, 2004
93 Eslami S M A, Rezende D J, Besse F, et al. Neural scene representation and rendering. Science, 2018, 360: 1204–1210
94 Chen S T, Shang J H, Zhang S Y, et al. Cognitive map-based model: toward a developmental framework for self-driving cars. In: Proceedings of IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), 2017. 1–8
95 Chen S T, Zhang S Y, Shang J H, et al. Brain-inspired cognitive model with attention for self-driving cars. IEEE Trans Cogn Dev Syst, 2019, 11: 13–25
96 Li D Y, Gao H B. A hardware platform framework for an intelligent vehicle based on a driving brain. Engineering, 2018, 4: 464–470
97 Chen L. The topological approach to perceptual organization. Visual Cognition, 2005, 12: 553–637
98 Eslami S M A, Heess N, Weber T, et al. Attend, infer, repeat: fast scene understanding with generative models. In: Proceedings of the 30th Conference on Neural Information Processing Systems, Barcelona, 2016. 3225–3233
99 Bar-Shalom Y, Daum F, Huang J. The probabilistic data association filter. IEEE Control Syst, 2009, 29: 82–100
100 Svensson L, Svensson D, Guerriero M, et al. Set JPDA filter for multitarget tracking. IEEE Trans Signal Process, 2011, 59: 4677–4691
101 Blackman S S. Multiple hypothesis tracking for multiple target tracking. IEEE Aerosp Electron Syst Mag, 2004, 19: 5–18
102 Kim C, Li F X, Ciptadi A, et al. Multiple hypothesis tracking revisited. In: Proceedings of the IEEE International Conference on Computer Vision, 2015. 4696–4704
103 Kuhn H W. The Hungarian method for the assignment problem. Naval Res Logistics, 1955, 2: 83–97
104 Cho H, Seo Y-W, Kumar B V K V, et al. A multi-sensor fusion system for moving object detection and tracking in urban driving environments. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA), 2014. 1836–1843
105 Chavez-Garcia R O, Aycard O. Multiple sensor fusion and classification for moving object detection and tracking. IEEE Trans Intell Transp Syst, 2016, 17: 525–534
106 Göhring D, Wang M, Schnürmacher M, et al. Radar/lidar sensor fusion for car-following on highways. In: Proceedings of the 5th International Conference on Automation, Robotics and Applications (ICARA), 2011. 407–412
107 Fayad F, Cherfaoui V. Object-level fusion and confidence management in a multi-sensor pedestrian tracking system. In: Proceedings of IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, 2008. 58–63
108 Kim D Y, Jeon M. Data fusion of radar and image measurements for multi-object tracking via Kalman filtering. Inf Sci, 2014, 278: 641–652
109 Govaers F, Koch W. An exact solution to track-to-track-fusion at arbitrary communication rates. IEEE Trans Aerosp Electron Syst, 2012, 48: 2718–2729

110 Zhang Z Y, Fidler S, Urtasun R. Instance-level segmentation for autonomous driving with deep densely connected mrfs. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. 669–677

111 Sun Y X, Liu M, Meng M Q H. Improving RGB-D SLAM in dynamic environments: a motion removal approach. Robot Auton Syst, 2017, 89: 110–122

112 Sun Y X, Liu M, Meng M Q H. Motion removal for reliable RGB-D SLAM in dynamic environments. Robotics Autonomous Syst, 2018, 108: 115–128

113 Caron F, Duflos E, Pomorski D, et al. GPS/IMU data fusion using multisensor Kalman filtering: introduction of contextual aspects. Inf Fusion, 2006, 7: 221–230

114 Suhr J K, Jang J, Min D, et al. Sensor fusion-based low-cost vehicle localization system for complex urban environments. IEEE Trans Intell Transp Syst, 2017, 18: 1078–1086

115 Wan G W, Yang X L, Cai R L, et al. Robust and precise vehicle localization based on multi-sensor fusion in diverse city scenes. 2017. ArXiv: 1711.05805

116 Tamar A, Wu Y, Thomas G, et al. Value iteration networks. In: Advances in Neural Information Processing Systems, 2016. 2154–2162

117 Katsuki F, Constantinidis C. Bottom-up and top-down attention: different processes and overlapping neural systems. Neuroscientist, 2014, 20: 509–521

118 Miller E K. Neurobiology: straight from the top. Nature, 1999, 401: 650–651

119 Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Machine Intell, 1998, 20: 1254–1259

120 Kadir T, Brady M. Saliency, scale and image description. Int J Comput Vision, 2001, 45: 83–105

121 Ba J, Mnih V, Kavukcuoglu K. Multiple object recognition with visual attention. 2014. ArXiv: 1412.7755

122 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. 2017 ArXiv: 1709.01507

123 Fu J L, Zheng H L, Mei T. Look closer to see better: recurrent attention convolutional neural network for fine-grained image recognition. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, 2017. 3

124 Kuo W J, Sjostrom T, Chen Y P, et al. Intuition and deliberation: two systems for strategizing in the brain. Science, 2009, 324: 519–522

125 Zheng N N, Liu Z Y, Ren P J, et al. Hybrid-augmented intelligence: collaboration and cognition. Front Inf Technol Electron Eng, 2017, 18: 153–179

126 Zhao D B, Hu Z H, Xia Z P, et al. Full-range adaptive cruise control based on supervised adaptive dynamic programming. Neurocomputing, 2014, 125: 57–67

127 Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning. Nature, 2015, 518: 529–533

128 Silver D, Huang A, Maddison C J, et al. Mastering the game of Go with deep neural networks and tree search. Nature, 2016, 529: 484–489

129 Gupta S, Davidson J, Levine S, et al. Cognitive mapping and planning for visual navigation. In: Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. 7272–7281

130 Rusu A A, Rabinowitz N C, Desjardins G, et al. Progressive neural networks. 2016. ArXiv: 1606.04671

131 Rusu A A, Vecerik M, Rothorl T, et al. Sim-to-real robot learning from pixels with progressive nets. 2016. ArXiv: 1610.04286