

# A Survey of Intelligent Sensing Technologies in Autonomous Driving



SHAO Hong<sup>1</sup>, XIE Daxiong<sup>1,2</sup>, HUANG Yihua<sup>1</sup>

(1. ZTE Corporation, Shenzhen 518057, China;

2. State Key Laboratory of Mobile Network and Mobile Multimedia, Shenzhen 518057, China)

**Abstract:** Intelligent perception technology of sensors in autonomous vehicles has been deeply integrated with the algorithm of autonomous driving. This paper provides a survey of the impact of sensing technologies on autonomous driving, including the intelligent perception reshaping the car architecture from distributed to centralized processing and the common perception algorithms being explored in autonomous driving vehicles, such as visual perception, 3D perception and sensor fusion. The pure visual sensing solutions have shown the powerful capabilities in 3D perception leveraging the latest self-supervised learning progress, compared with light detection and ranging (LiDAR)-based solutions. Moreover, we discuss the trends on end-to-end policy decision models of high-level autonomous driving technologies.

**Keywords:** autonomous vehicles; neuron network; automotive electronics; sensor fusion

DOI: 10.12142/ZTECOM.202103007

<https://kns.cnki.net/kcms/detail/34.1294.TN.20210817.1538.004.html>, published online August 18, 2021

Manuscript received: 2021-06-22

**Citation** (IEEE Format): H. Shao, D. X. Xie, and Y. H. Huang, "A survey of intelligent sensing technologies in autonomous driving," *ZTE Communications*, vol. 19, no. 3, pp. 56 – 63, Sept. 2021. doi: 10.12142/ZTECOM.2021030007.

## 1 Introduction

The next generation vehicles are transforming from mechanical-centric to software-defined. Since the Grand Challenge orchestrated by the Defense Advanced Research Projects Agency (DARPA)<sup>[1]</sup>, autonomous driving (AD) technologies have been accelerating. Autonomous driving is considered to be a revolutionary technology that profoundly affects human society and transportation. To categorize these systems, the Society of Automobile Engineers (SAE) has defined six levels of automation ranging from 0 (no automation) to 5 (full automation). The deployment of full autonomy is still expected in years. Automotive manufactures are more inclined to gradually increase the level of autonomy from Advanced Driver Assistance Systems (ADAS) to full autonomous driving. The ADAS ranges on the spectrum of passive to

active safety functions, such as forward collision warning (FCW), lane departure warning (LDW), blind spot monitoring (BSM), autonomous emergency braking (AEB), lane keeping assistance (LKA), adaptive cruise control (ACC), forward collision-avoidance (FCA), traffic jam assist (TJA), and traffic jam pilot (TJP).

Autonomous driving cars need to understand the surrounding environment and then take actions continuously. Autonomous vehicles rely on different sensors that work together to perceive the internal and external car environments. The most involved sensors in the car are radars, light detection and ranging (LiDAR) systems, cameras, ultrasonic and far-infrared sensors, etc.<sup>[2]</sup> These long-range and short-range sensors provide relevant data to interpret the surrounding scenes near the vehicle with a variety of solutions, such as from 8 Vision 1 Radar

(8V1R) to 15 Vision 5 Radar 3 LiDAR (15V5R3L). Sensor fusion processing is also deeply integrated into the algorithms of autonomous driving. Smart sensors combined with the extreme compute performance deployed in a car make the car more and more like a robot. This will be continuously increasing the complexity and bringing challenges for the automotive electronic architecture.

The goal of this paper is to provide a survey of sensing technologies on autonomous vehicles. Ref. [2] reviewed the most popular sensor technologies and their characteristics but did not analyze how the progress of the algorithms affected the configuration of vehicle sensors. Here we track some important improvements of the neural network and deep learning algorithms linked with perception in autonomous driving. The well-known mask region based convolutional neural network (Mask R-CNN) algorithm<sup>[3]</sup> achieves the best instance segmentation accuracy in 2D visual recognition. The popular vision algorithm You Only Look Once (YOLO)<sup>[4]</sup> is less accurate but much faster than Mask R-CNN and suitable for autonomous driving. YOLO is also extended to LiDAR 3D point clouds<sup>[5]</sup>. The fusion of multiple sensors like vision and LiDAR<sup>[6]</sup> has taken more advantages before Pseudo LiDAR technology<sup>[7]</sup> emerges and the latter is showing the power of pure vision in 3D perception. Unsupervised learning approaches of depth estimation<sup>[8]</sup> have further accelerated the utilization of pure vision in autonomous driving. Moreover, sensors should not only perceive the current environment, but also constantly predict the environmental context in the next few seconds. For example, Uber uses a convolutional neural network (CNN) model to predict possible trajectories of the surrounding actors<sup>[9]</sup>.

The structure of this paper is arranged as follows. Section 2 explains the impact of intelligent sensing technology on automotive electronic architecture. Section 3 provides a detailed overview of the sensing algorithms. Section 4 discusses a decision-making model. Finally, Section 5 concludes this paper.

## 2 Impacts on Electrical/Electronic (E/E) Architecture

Autonomous driving requires processing of dozens of sensors with high performance computation. This brings new impacts on the traditional automotive electrical and electronic architecture. Firstly, centralized processing will replace the distributed processing to provide high computational power. Secondly, sensor data transmission will require higher communication bandwidth and time-sensitive networking (TSN) becomes the promising technology for it.

### 2.1 Centralized Computing

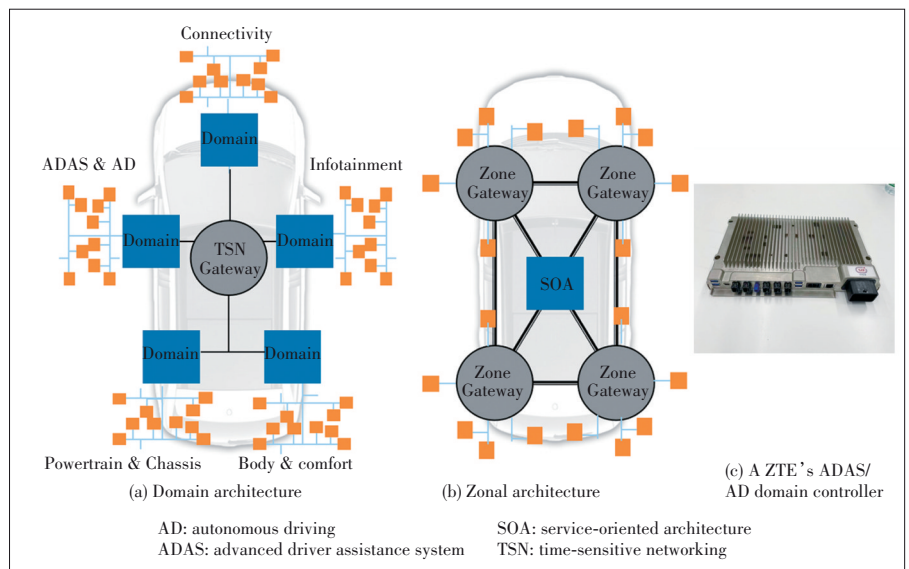
Traditional cars are composed of one-box one-function modular electronic control units (ECUs). However, due to the complexity of autonomous vehicles, the approach where ECUs are tightly coupled with firmware from hardware will encounter difficulties to meet the requirements of high computation power and software integration in intelligent perception. Regarding the increasing number of sensors and actuators in autonomous vehicles, there are several impacts on the legacy automotive E/E architecture such as complexity, harness, high bandwidth, and artificial intelligence (AI) computing.

The distributed modular ECU system needs to upgrade to an integrated centralized computing system for autonomous driving. The future trend is combining sensors and ECUs into the domain controller (Fig. 1). Then all domain controllers will be further merged into one centralized vehicle computing platform with functional redundancy to achieve functional centralization.

Fig. 1 shows the domain architecture and the zonal architecture with a centralized vehicle computing platform to further optimize the harness layout. The domain architecture consists of separated domain controllers according to the vehicle functions. The zonal architecture consists of gateways connected with the redundant computing platform that supports service oriented architecture (SoA) to process the vehicle functions. Fig. 1(c) shows a ZTE's ADAS/AD domain controller, which can be used in L2 and L3 autonomous driving scenarios.

### 2.2 Time-Sensitive Networking

Another impact is the high data rate sensors and actuators, such as raw data cameras, LiDARs and radars, which will need high bandwidth and deterministic real-time communication within the car. As early as 2006, the IEEE802.1 established the audio video bridging (AVB) working group and suc-



▲ Figure 1. Domain and zonal electrical and electronic (E/E) architecture

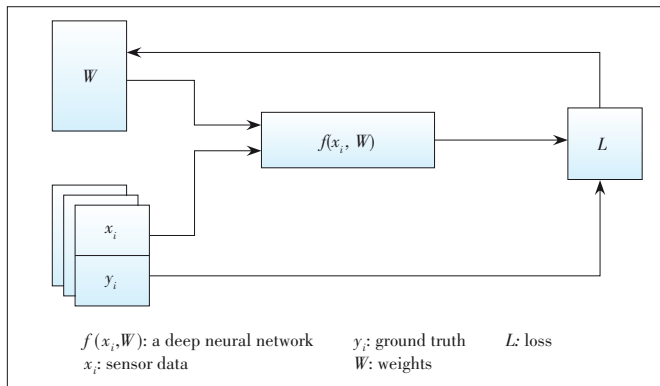
cessfully solved real-time synchronous data transmission in the following years. This immediately attracted the attention of the automotive industry. In 2012, The AVB working group was renamed by the TSN working group, focusing on enabling low-latency and high-quality transmission of streaming data. TSN aims to establish a “universal” time-sensitive mechanism for the Ethernet protocol to ensure the time determinism of network data transmission and the delay reaches the microsecond level. So it will be the ideal candidate communication backbone technology for the new automotive E/E architecture. As shown in Fig. 1, in the automotive backbone which requires high bandwidth and deterministic real-time communication, the TSN gateway is used to ensure that it can transmit between different domains with low latency and small jitters. The TSN node will also transmit a redundant frame for the high-performance sensors (e.g., high-resolution cameras). The Ethernet TSN can reach 1 Gbit/s or more, while the controller area network (CAN) and FlexRay are 1 Mbit/s and 20 Mbit/s respectively.

### 3 Sensing Algorithms

At present, most intelligent perception tasks are achieved by deep neural network models. Here we analyze several basic algorithms and their development.

#### 3.1 Deep Learning Models

For intelligent sensing, a very basic deep learning model is well known as illustrated in Fig. 2. A deep neural network  $f(x_i, W)$  consists of multiple layers. For example, it could be combinations of CNN, fully connected (FC) networks, residual networks (ResNet), long short term memory (LSTM) networks, even transformer networks, etc. An input data set  $\{(x_i, y_i)\}_{i=1}^n$  is sampled from the collected data which could be historical driving data or synthesized data constructed from a simulator. Within the input data set where  $x_i$  is denoted as the sensor data that could be images, videos, or LiDAR point clouds and  $y_i$  denoted as the ground truth of the target or labels that are extracted from the car environment. The target  $y_i$  can be spatial information (like drivable area, lanes, roads, etc.), semantic in-



▲ Figure 2. Basic deep learning model

formation (traffic lights, traffic signs, turn indicators, on-road marking, etc.), or moving objects (pedestrians, cyclists, cars, etc.). We can manually label them or generate them by a simulator.

The difference between the predicted result of the neural network  $f$  and the ground truth  $y_i$  is the loss function denoted as  $L$ . The training goal is to optimize  $L$  through iterations of the data set and adjust the weights  $W$ .

The weights  $W$  could be a very large tensor including all the weights of each deep network layer. The neural network  $f$  can do the inference once we get the optimal weight  $W^*$ :

$$W^* = \underset{W}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n L(f(x_i; W), y_i), \quad (1)$$

where  $W = \{W^{(0)}, W^{(1)}, \dots\}$ .

Reasonably defining the loss function  $L$  is the most important work for designing deep learning model. For example, for the classification case, we use the binary cross-entropy to measure the loss:

$$L = \frac{1}{n} \sum_{i=1}^n y_i \log(f(x_i; W)) + (1 - y_i) \log(1 - f(x_i; W)). \quad (2)$$

And for the regression case, we use mean square error (MSE) to measure the loss:

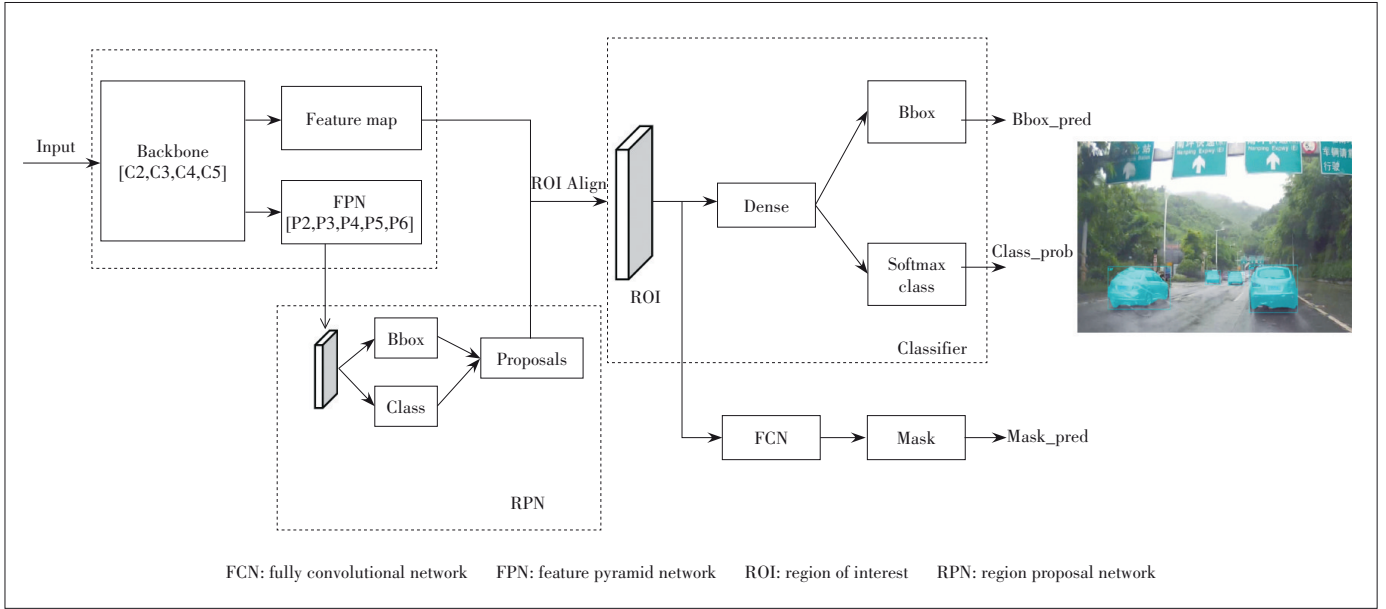
$$L = \frac{1}{n} \sum_{i=1}^n (y_i - f(x_i; W))^2. \quad (3)$$

We can also put extra terms in the loss function to reduce its generalization error but not its training error. These strategies are as known as regularization.

#### 3.2 Visual Perception Understanding

Computer vision has almost become the foundation of intelligent perception for autonomous driving. Many visual recognition problems are related to autonomous driving, such as object detection, segmentation and instance detection. We have built a practice Mask R-CNN<sup>[3]</sup> for visual perception, as shown in Fig. 3.

We use a ResNet50<sup>[10]</sup> to construct the backbone of the Mask R-CNN. The original images are resized to a fixed size before entering the backbone network. The feature maps extracted from the backbone are C2, C3, C4, and C5, which construct the feature pyramid networks (FPN) in order to detect the objects from different scales. FPN has a bottom-up and top-down structure to connect the corresponding layer and generate a new feature map [P2, P3, P4, P5, P6], where P5 corresponds to C5, P4 corresponds to C4+ UpSampling2D of P5, P3 corresponds to C3+ UpSampling2D of P4, P2 corresponds to C2+ UpSampling2D of P3, and P6 corresponds to MaxPooling2D of P5. Then this feature map is put into a region proposal net-



▲ **Figure 3. A practiced mask region based convolutional neural network (Mask R-CNN)**

work (RPN) to generate the region of interest (ROI) proposals and tune the coordinates.

This network completes three tasks simultaneously: 1) target localization, which directly predicts a target bounding box on the image. Here it is called Bbox\_pred; 2) target classification, which is denoted as Class\_prob; 3) pixel-level target segmentation, denoted as the Mask\_pred for each ROI.

The total loss of the network is:

$$L = L_{cls} + L_{box} + L_{mask}, \quad (4)$$

where  $L_{cls}$  is the classification loss,  $L_{box}$  is the bounding-box loss, and  $L_{mask}$  is the average binary cross-entropy loss of pixel-wise mask for each instance.

Although the accuracy of Mask R-CNN is relatively high, its region proposal pipelines are still time-consuming. For autonomous driving, since the perception task requires real-time performance, we need to make the inference efficient and maintain good accuracy. So a single shot detector model like YOLO<sup>[4]</sup> will be a better choice than Mask-R-CNN. It can process videos at real time.

### 3.3 3D Perception

3D object detection in autonomous driving is a common task. A direct and reliable approach is employing the LiDAR sensor to provide the 3D point cloud reconstruction of the surrounding environment. The object detection and classification of LiDAR point clouds may be conducted in 2D bird-view by projecting the 3D point clouds into 2D or directly conducted in 3D space. Unlike visual systems, LiDAR point clouds lack of rich RGB information, the density of the point cloud is critical for small object detection.

For high resolution LiDAR, YOLO3D<sup>[5]</sup> introduced a 3D ob-

▼ **Table 1. YOLO3D network architecture<sup>[5]</sup>**

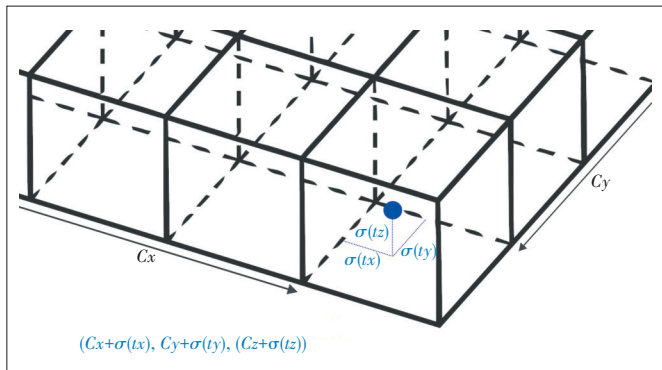
Layer	Filter	Size	Feature Maps
Conv2d	32	(3, 3)	608×608×2
Maxpooling		(size 2, stride 2)	
Conv2d	64	(3, 3)	
Maxpooling		(size 2, stride 2)	
Conv2d	128	(3, 3)	
Conv2d	64	(3, 3)	
Conv2d	128	(3, 3)	
Maxpooling		(size 2, stride 1)	
Conv2d	256	(3, 3)	
Conv2d	128	(3, 3)	
Conv2d	256	(3, 3)	
Maxpooling		(size 2, stride 2)	
Conv2d	512	(3, 3)	
Conv2d	256	(1,1)	
Conv2d	512	(3, 3)	
Conv2d	256	(1,1)	
Conv2d	512	(3, 3)	
Maxpooling		(size 2, stride 2)	
Conv2d	1 024	(3, 3)	
Conv2d	512	(1, 1)	
Conv2d	1 024	(3, 3)	
Conv2d	512	(1, 1)	
Conv2d	1 024	(3, 3)	
Conv2d	1 024	(3, 3)	
Conv2d	1 024	(3, 3)	
Conv2d	1 024	(3, 3)	
Conv2d	1 024	(1,1)	38×38×33
Reshape			38×38×3×11

ject detection algorithm that direct expands from the 2D algorithm YOLO<sup>[4]</sup>. This approach directly projects the LiDAR point cloud to the bird-view space for real-time classification and detecting 3D Object Bounding Box. The structure of the network is shown in Table 1.

The network output prediction is expanded by the YOLO regression to 3D dimensions regression output and target classification. It will return the object bounding box center ( $x$ ,  $y$ , and  $z$ ), the 3D dimensions (length, width, and height), the orientation in the bird-view space, the confidence, and the object class label. The YOLO3D grid expanding from YOLO is shown in Fig. 4.

The loss also extends from the YOLO 2D boxes ( $x$ ,  $y$ ,  $l$ ,  $h$ ) to 3D oriented boxes ( $x$ ,  $y$ ,  $z$ ,  $w$ ,  $l$ ,  $h$ ) and the orientation. The total loss includes the confidence score and the cross-entropy loss over the object classes.

The YOLO3D network is trained end to end because it is a single shot detector that ensures its real-time 3D performance in the inference path.



▲ Figure 4. YOLO3D grid cells, assuming one layer high<sup>[5]</sup>

### 3.4 Sensor Fusion

If the density of the LiDAR point cloud is sparse, small objects like pedestrians and cyclists will be hard to recognize. The LiDAR needs to do a sensor fusion with the camera. The sensor fusion can take advantage of both LiDAR point clouds and camera images, which can preserve more semantic information to achieve higher object detection accuracy. Therefore, the autonomous driving cars are commonly equipped with multiple kinds of sensors like both LiDAR and camera.

The multi-view 3D (MV3D) network<sup>[6]</sup> gives an example of sensor fusion framework that takes 3D LiDAR point clouds and RGB images as input to predict 3D objects (Fig. 5).

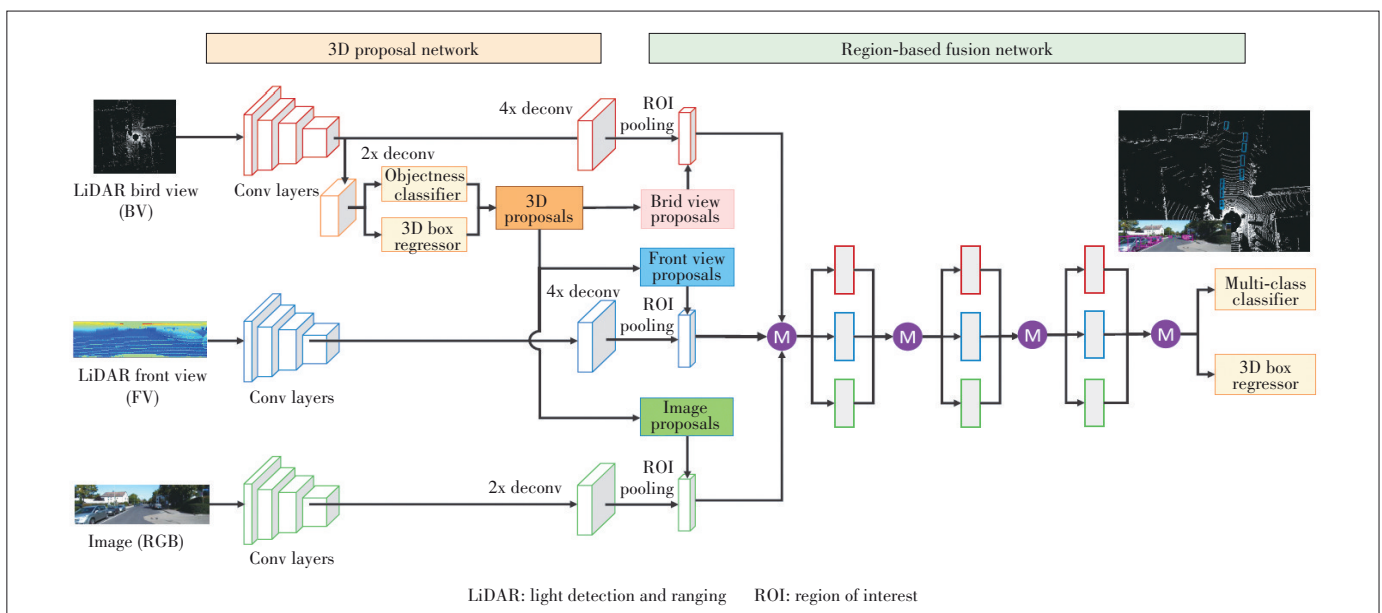
The MV3D network consists of two sub-nets: a 3D proposal network and a region-based fusion network. The 3D proposal network generates highly accurate 3D candidate boxes from the bird's eye view of a point cloud. The region-based fusion network deeply fuses multi-view features to predict the position, size, and orientations of the 3D target.

A multiple layer feature fusion is adopted to increase the selected ROI from the fusion between different view features. The fusion network can significantly improve the position accuracy and recognition accuracy of 3D perception.

### 3.5 Pure Vision vs. LiDAR

In autonomous driving cars, LiDAR or pure vision based solution has become a controversial topic. We mainly consider the perception algorithm to put aside the cost, weather environment and other factors.

There is a vast difference in perception between pure vision and the LiDAR solution. While in the LiDAR case, the vehicle detects the 3D object to avoid collisions of pedestrians, bicycles, vehicles, etc. and it also compares the features of real-time 3D point clouds with a pre-built high definition map, uti-



▲ Figure 5. Multi-view 3D object detection network (MV3D)<sup>[6]</sup>

lizing the simultaneous localization and mapping (SLAM) algorithm to precisely localize the vehicle position and execute lane follow function. However, for pure vision cases, the camera creates 2D information and it is difficult to reliably and accurately reconstruct the 3D environment of each pixel, so the SLAM will not be conducted to directly predict the lane from the camera images. This limits the pure vision solution under the L3 autonomy.

With the progress of monocular or binocular camera 3D perception, the accuracy of depth estimation is continuously improved and even pseudo-LiDAR can be constructed, as shown in Fig. 6<sup>[7]</sup>.

With the input of stereo or monocular images, the network can predict the depth map and back-project it into a 3D point cloud in the LiDAR coordinate system called a pseudo-LiDAR, so we can reuse the LiDAR-based algorithms in the pure visual solution and also implement high-level autonomy.

### 3.6 Self-Supervised Learning

The supervised learning requires the provision of a data set with depth information as a ground truth. However, the ground-truth of depth information of visual data is more difficult to obtain, so pure visual depth perception technology men-

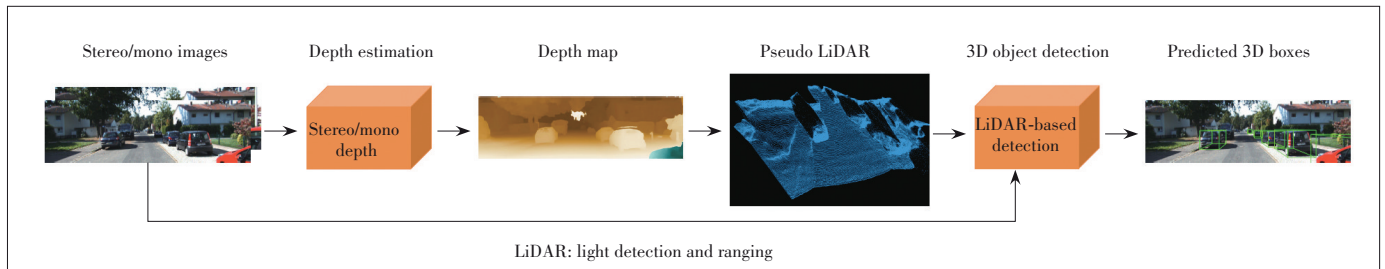
tioned in the previous section is limited to a restricted training data set, while the self-supervised method developed later does not require depth information annotation and directly uses video frames to complete the training, which is a great improvement.

The self-supervised learning method is to reconstruct the related pixels through geometric constraints between two frames of the multi-view as the supervision input so that there is no need to rely on the annotation of depth. In the backbone network, it is the same as the original network of supervised learning.

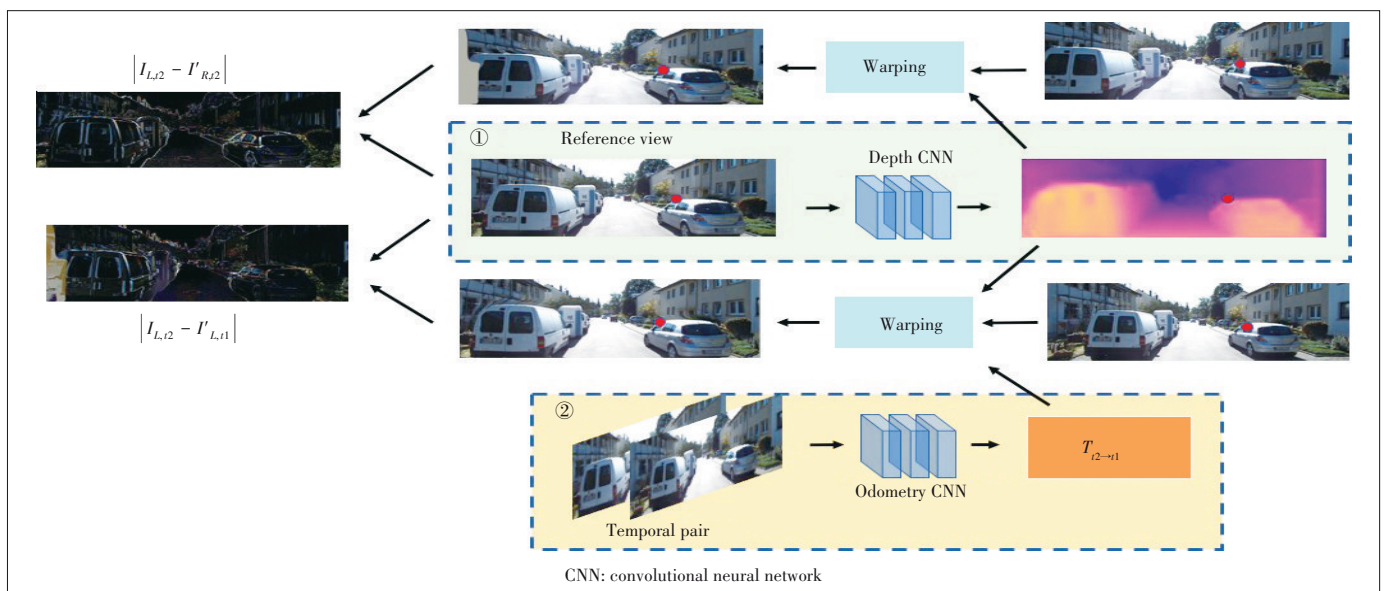
A self-supervised learning example<sup>[8]</sup> is shown in Fig. 7. It can estimate the depth and movement of the camera using stereo video sequences.

Ref. [8] enables the use of both spatial and temporal photometric warp errors, and constrains the scene depth and camera motion in a common real-world scale.

As shown in Fig. 8, a convolutional neural network for single view depth ( $CNN_D$ ) and a convolutional neural network for visual odometry ( $CNN_{VO}$ ) are used. For self-supervised learning, the fundamental supervision signal comes from the task of image reconstruction and the image reconstruction loss is used as a supervision signal to train  $CNN_D$  and  $CNN_{VO}$ .



▲ Figure 6. Image-based 3D object detection<sup>[7]</sup>



▲ Figure 7. A self-supervised learning example with the use of stereo video sequences<sup>[8]</sup>

### 4 End-to-End Decision Model

A typical autonomous intelligent vehicle system can be simply divided into three parts: the perception part, the planner and the controller. The perception part extracts the features from the environment and the planner outputs a driving trajectory for driving in the 3D space. The controller then executes this trajectory as the steering angle and acceleration within the physical constraint of the vehicle. We call the decision model end-to-end when it takes in sensing data and outputs how we should drive with fully end-to-end training approach to mimic human driving.

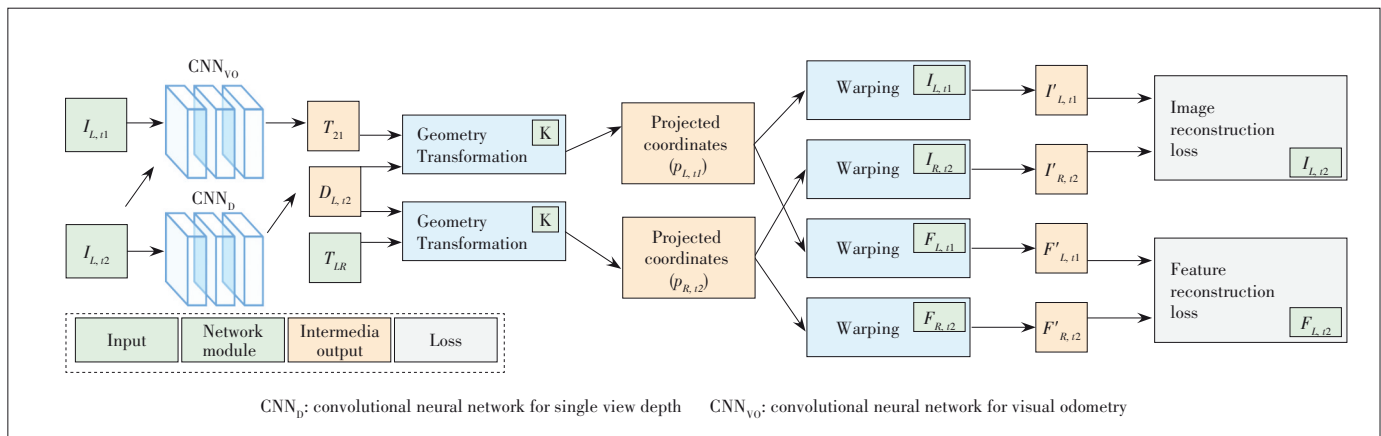
We usually implement autonomous driving by using a rule-based planner to make trajectory decisions. Engineers will manually write the planner for such an autonomous driving system. Due to the complexity of the driving problem, the manual rule-based planner may never enable the level of full autonomy because the edge cases, such as temporary road signals and traffic accidents, are constantly increasing in new scenarios. To build an end-to-end intelligent planner based on neural network is an idealist goal people want to achieve<sup>[9]</sup>.

The reinforcement learning algorithms such as AlphaGo, AlphaZero and muZero<sup>[11]</sup> have shown powerful capabilities in policy searching for building an end-to-end decision model.

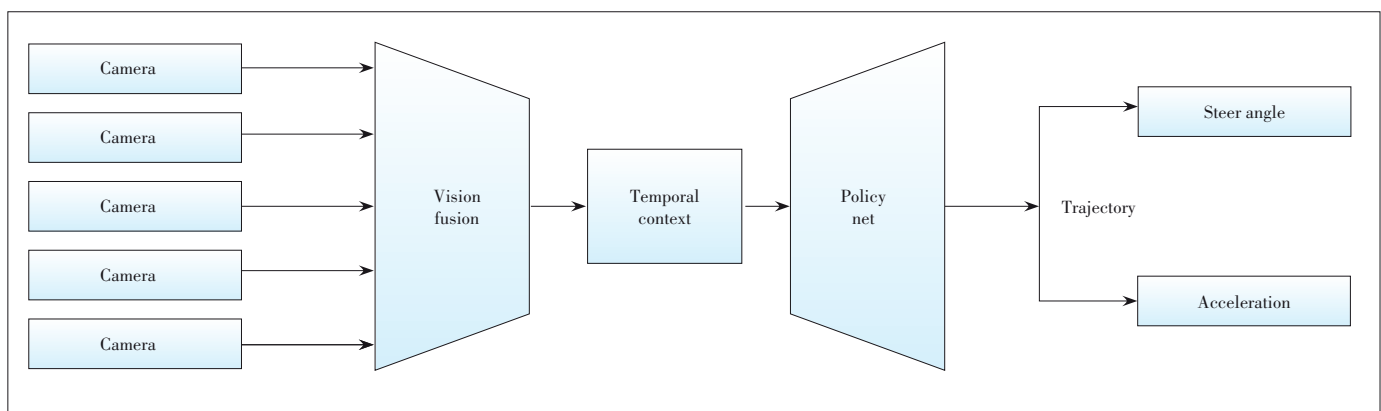
However, reinforcement learning is still limited to being trained in a simulation environment or a game environment. Interacting with the true environment of autonomous driving is extremely expensive, slow and dangerous, which is completely unrealistic.

Recent development shows model-based offline reinforcement learning approaches are trying to learn a policy model from the environment dynamics. Just learning from observational data, the model may perform well in a real environment. Intuitively, we may extend this kind of model to build an end-to-end planner. Then human driving behaviors are collected and a model with observational data is trained to predict the trajectories for mimicking human driving. Fig. 9 shows the prototype of an end-to-end decision model we are studying. The videos from multiple cameras are input to a convolutional fusion network and a feature map is output. The features go into a temporal block, such as LSTM and GRU, to extract the sequence features, and then connect to the policy network and predict possible multiple trajectories.

However, the learning policies from purely observational data may not normally work because the data only cover a small region of the observed space<sup>[12]</sup>. Once a car deviates from the predicted best “human driving” trajectory, it is difficult for the car to recover from deviation and it will drift away from the



▲ Figure 8. A self-supervised learning framework<sup>[8]</sup>



▲ Figure 9. End-to-end autonomous driving decision-making network

ideal trajectory. The reason is that, unlike the learning in simulation environment where interaction and self-correcting are allowed, there is no actual interactive driving data for training on in this case.

In order to solve the problem, Ref. [12] proposes to train a policy by unrolling a learned model of the environment dynamics over multiple time steps while explicitly penalizing such costs as an uncertainty cost that represents its divergence from the states (trajectories) on which it is trained.

## 5 Conclusions

We discuss the application of intelligent sensors in autonomous vehicles and their impacts on automotive E/E architecture. The distributed ECU system will be replaced by centralized architecture to provide more computation power and integration. Moreover, for the sensors with high data rates, a TSN backbone plays a key role for E/E architecture. The algorithm of sensing perception based on neural networks is highly integrated with autonomous driving. The fusion of multiple sensors may enable better accuracy and robustness. Moreover, pure visual perception shows a powerful capability of 3D estimation versus LiDAR, and the visual-based pseud-LiDAR can reuse the existing LiDAR-based algorithms and improve the autonomy to a high level. Self-supervised learning is a more promising technology for cars in 3D perception.

It is also pointed out that the rule-based policy will never get over the edge cases, so the end-to-end policy seems to be a better approach to high-level autonomous driving and still needs further studying.

## References

- [1] BEBEL J C, HOWARD N, PATEL T. An autonomous system used in the darpa grand challenge [C]//7th International IEEE Conference on Intelligent Transportation Systems. Washington, USA: IEEE, 2004: 487 - 490. DOI: 10.1109/ITSC.2004.1398948
- [2] MARTI E, DE MIGUEL M A, GARCIA F, et al. A review of sensor technologies for perception in automated driving [J]. IEEE intelligent transportation systems magazine, 2019, 11: 94 - 108. DOI: 10.1109/MITS.2019.2907630
- [3] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [EB/OL]. (2018-01-24)[2021-04-28]. <https://arxiv.org/abs/1703.06870v3>
- [4] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [EB/OL]. (2016-12-25)[2021-04-28]. <https://arxiv.org/abs/1612.08242v1>
- [5] ALI W, ABDELKARIM S, ZAHNAN M, et al. YOLO3D: end-to-end real-time 3D oriented object bounding box detection from LiDAR point cloud [EB/OL]. (2018-08-07)[2021-04-28]. <https://arxiv.org/abs/1808.02350v1>
- [6] CHEN X Z, MA H M, WAN J, et al. Multi-view 3D object detection network for autonomous driving [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. DOI: 10.1109/CVPR.2017.691
- [7] WANG Y, CHAO W-L, GARG D, et al. Pseudo-LiDAR from visual depth estimation: bridging the gap in 3D object detection for autonomous driving [EB/OL]. (2020-02-22)[2021-04-28]. <https://arxiv.org/abs/1812.07179v6>
- [8] ZHAN H Y, GARG R, WEERASEKERA C S, et al. Unsupervised learning of monocular depth estimation and visual odometry with deep feature reconstruction [EB/OL]. (2020-02-22)[2021-04-28]. <https://arxiv.org/abs/1803.03893v3>
- [9] CUI H, RADOSAVLJEVIC V, F-CCHOU, et al. Multimodal trajectory predictions for autonomous driving using deep convolutional networks [EB/OL]. (2019-03-01)[2021-04-28]. <https://arxiv.org/abs/1809.10732v2>
- [10] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [EB/OL]. (2015-12-10)[2021-04-28]. <https://arxiv.org/abs/1512.03385>
- [11] SCHRITTWIESER J, ANTONOGLU I, HUBERT T, et al. Mastering Atari, Go, chess and shogi by planning with a learned model [J]. Nature, 2020, 588: 604 - 609. DOI: 10.1038/s41586-020-03051-4
- [12] HENAFF M, CANZIANI A, LECUN Y. Model-predictive policy learning with uncertainty regularization for driving in dense traffic [EB/OL]. (2019-01-08)[2021-04-28]. <https://arxiv.org/abs/1901.02705v1>

## Biographies

**SHAO Hong** (shao.hong@zte.com.cn) received his M.S. degree from Harbin Institute of Technology, China and joined ZTE Corporation in 1994. He is currently the leader of the Innovation Team in the Corporation Development Department of ZTE Corporation. His research interests include deep learning, autonomous vehicles and robotics.

**XIE Daxiong** is currently the chairman of the board of supervisors of ZTE Corporation and the director of State Key Laboratory of Mobile Network and Mobile Multimedia Technology, China. He is a professorate senior engineer and a member of the Mobile and Satellite Expert Group of Communication Science and Technology Commission of Ministry of Industry and Information Technology of China. With ZTE Corporation, he led the successful research and development of CDMA, GoTa, WCDMA and other large-scale digital telecommunication systems.

**HUANG Yihua** received his M.S. degree from Huazhong University of Science and Technology, China. He is currently the manager of Corporation Development Department of ZTE Corporation. His research interests include new technology innovation, digital transformation and innovation strategy.