ITM Web of Conferences **80**, 01033 (2025) https://doi.org/10.1051/itmconf/20258001033
ACAAI 2025

Deep Learning for Autonomous Driving: A Survey of Methods, Paradigms, and Future Trends

Yixiong Xiao *

Changsha No.15 High School, Changsha, Hunan Province, China

Abstract. With the continuous growth of transportation demand and the increasingly prominent issue of road safety, improving traffic efficiency and driving safety has become a core problem that urgently needs to be solved. Although the emergence of autonomous driving technology has brought significant convenience to public transportation, traditional artificial intelligence methods overly rely on manual labeling and fixed models, making it difficult to cope with complex and changing real-world road environments. In scenarios of extreme weather, sudden traffic accidents, or uncertain driving behavior, its performance is particularly inadequate. As the core supporting technology of automatic driving, deep learning has the ability to independently extract key information from massive data, conduct self-training and model iteration, and provides a solid foundation for the intelligence and self-evolution of auto drive system. This paper systematically reviews the typical applications of deep learning in autonomous driving, including environmental perception, behavior prediction, decision-making and control, and analyzes the advantages and limitations of deep learning based-methods such as convolutional neural networks, recurrent neural networks, Transformers. Moreover, the current challenges faced, such as insufficient data, poor model generalization were summarized, and future development directions were discussed.

1 Introduction

Nowadays, since the development of the world, transportation has been developing along with the economy, and people's pursuit of beautiful visions such as efficiency, timeliness, and convenience in transportation has become increasingly demanding [1]. According to statistics from the Ministry of Public Security, the number of newly licensed drivers in China will reach 22.26 million in 2024. According to the website of the Ministry of Public Security reposted, in 2024, the national public security traffic management departments handled a total of 667507 general and above road traffic accidents, resulting in 104372 deaths, 494174 injuries, and direct economic losses of 3.37 billion yuan, a decrease of 13.7%, 4.6%, 12.1%, and an increase of 1.4% respectively compared to the previous year. Obviously, beginners lack more experience compared to experienced ones, and the time cost required for a novice

* Corresponding author: 13574121782@163.com

to grow into an experienced one is extremely high. Moreover, traffic accidents are not exclusive to beginners, and experienced people who often walk by the river are inevitably "wet shoes". It can be seen that the time cost paid by the public to avoid traffic accidents has not achieved the effect of absolute risk avoidance.

In recent years, the deployment of artificial intelligence in the market has led to the gradual introduction of autonomous driving technology as a byproduct, bringing many conveniences to the public [2]. Whether it's difficult parking or complex congestion and avoidance issues, autonomous driving technology greatly assists people in making decisions and executing them, greatly aiding those who are prone to making incorrect judgments in these areas.

However, traditional artificial intelligence technology overly relies on manually annotated and fed training data, and the models are too fixed to cope with the ever-changing real road conditions. For example, in extreme weather conditions or when facing different road conditions, how to reasonably speculate on the motives of the other party's vehicles, and how to judge the impact of natural traffic accidents on driving routes are problems that traditional technology is difficult to solve properly and urgently needs to be changed. As an important core technology supporting self-evolution, deep learning algorithms are applied and optimized in the field of autonomous driving. It enables the auto drive system to independently process massive road condition data, especially to extract key information from special scenes such as traffic accidents, conduct information processing, self-feeding training, and then realize self-iteration and evolution of the model. Therefore, research on the self-evolution of deep learning algorithms in the field of autonomous driving is not only in line with the urgent needs of current transportation development, but also of great significance for improving the safety and reliability of autonomous driving. As shown in Figure 1, this paper will focus on this topic and review and explore related research.

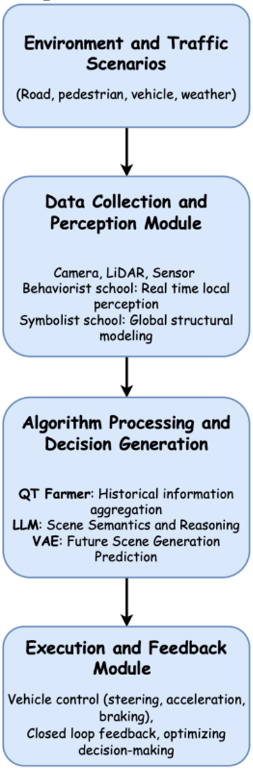

Fig. 1. Process of autonomous driving.

2 Tasks in Autonomous Driving

The core goal of the auto drive system is to realize the perception, understanding and decision-making of the complex traffic environment, so as to ensure the safe and efficient operation of vehicles in the dynamic environment [3]. According to this objective, the main tasks of autonomous driving are usually broken down into three categories by current research: behavior prediction, environmental perception, and decision-making and control.

The ability to identify and recognize both static and dynamic features in the surrounding environment, including as lane markings, traffic signs, road geometry, and traffic participants like cars and pedestrians, is a key component of autonomous driving. To achieve high-precision object recognition and semantic segmentation, some devices are used to integrate multimodal sensor data, including cameras, LiDAR, millimeter wave radar. The data collected by these devices can be used as input for deep learning-based tasks.

Based on above discussion, behavior prediction aims to predict the future movements and traffic activities. For example, behavior prediction can judge if people may cross the street. In addition, the program requires long-term inference based on historical trajectory data. Generally, deep learning is used in this field, including Recurrent Neural Networks (RNNs), Long Short-Term Memory Networks (LSTMs), and Transformer. These methods have greatly improved the precision of prediction.

Finally, converting perception and prediction results into useful activities are required. They are used into critical decision-making and control. In this whole process, the control module transforms these advanced judgments into precise steering, acceleration, and braking actions. However, decision-making module must perform path planning and action selection based on dynamic traffic conditions. In this aspect, reinforcement learning techniques demonstrate particular advantages. The reason is that they continuously interact with the environment to derive optimal strategies. Thus, this process can generate adaptive driving behavior in challenging scenarios.

In conclusion, three essential components of the auto drive system are environmental awareness, behavior prediction, and decision-making control. These three components work closely together to form a complete closed-loop process, thus providing a solid application foundation for the implementation and improvement.

3 Application and Schools of Deep Learning Algorithms in Autonomous Driving

Major automakers and government research institutions have steadily increased their research expenditures in the autonomous driving. These organizations have developed numerous comprehensive technological systems [4]. These studies mainly focus on three core aspects: (1) firstly, extracting and collecting data from the environment. It is worth mentioning that collecting data includes the acquisition and preprocessing of multimodal sensor data; (2) Secondly, specific algorithms are used to calculate and process data, thereby forming decision instructions. (3) Finally, issue the decision command and await execution. Subsequently, complete the action through the execution module. This closed-loop process has achieved remarkable results in practical applications. This process enhances environmental awareness and decision-making accuracy effectively [5].

During technical implementation, different research schools exhibit significant variations in the emphasis accorded to each type of connection. The two most representative mainstream schools, including Behaviorist and Symbolist. The main idea of the Symbolist school is to use neural networks to mimic human thought and judgment processes. Specifically, Symbolist school uses symbolic reasoning and logical principles to process environmental data. This school emphasizes the integration of global information and long-term planning capabilities. Also, this school is applicable to security verification and policy derivation. It differs from the behaviorist school, which employs a perception-action paradigm to respond environmental changes. By processing raw data, this school prioritizes short-term dynamic adaptability and operational flexibility [6].

As the development of technology, researches tend to integrate the advantages of the two schools. This not only reflects the diversified trend of technological development, but also provides a theoretical basis and practical direction for the optimization of deep learning algorithms in autonomous driving in the future.

4 Extract and Collect Data from the Environment

High quality data is the core driving force for breakthroughs in autonomous driving technology. There are two major limitations to existing datasets: single vehicle single traversal - relying solely on single vehicle data collection at a specific time, lacking multi vehicle collaboration and scene reproducibility; Single modality - often focusing on single sensor data from vision or LiDAR, making it difficult to support multimodal fusion algorithm training in complex scenes. Obviously, the comprehensive collection of information and data is crucial to the reliability of the auto drive system. The emphasis on data collection varies among different schools of thought.

4.1 Emphasis and Characteristics of Behaviorist Schools in Information Collection

The behaviorist school mainly focuses on immediate perception and dynamic adaptation, and its data collection has three major characteristics [7], as follows:

1) **Instant sensory data**: emphasizes the use of perception action models to respond to the environment in real-time. Early Tesla vehicles collected surrounding information by built-in cameras. These devices captures both static objects and dynamic objects (such as pedestrians, vehicles, and animals). Specifically, static objects include traffic lights and obstacles. Dynamic objects include pedestrians, vehicles, and animals. These data can help the vehicle to make rapid judgments and decisions.
2) **Dynamic change data**: This data monitors changes in the vehicle's surrounding environment, including the acceleration and speed. This kind of information is also important for behavior prediction and decision-making.
3) **Direct sensory input data**: Instead of using pre-processed data, the behaviorist school tends to use the raw data from sensors. Normally, some deep learning methods, such as Convolutional Neural Networks (CNN), are used. After extracting raw data, it will be used as input of deep learning methods. In this way, the model can simulate people's reactions to the environment.

4.2 Emphasis and Characteristics of Symbolist Schools in Information Collection

The symbolist school emphasize three main methods of data collecting technique, as follows:

1) **Environmental structured data**: The data collected by sensor is needed to convert as the structured representation. For example, to create high-precision maps, Waymo employs lidar. Moreover, raw data can be compiled into tables.
2) **Historical and global data**: In addition to monitoring current environmental data, historical trajectories and global data are important too. This kind of data can integrate

with data from other vehicles. By this way, a macro-level transportation network can be established. This method provides a support for long-term planning and decision-making.

3) **Semantic information data**: After collecting sensor data, the Symbolist School will further extract semantic information. It will use symbols and logical rules to define environmental aspects. Then, algorithms will be used to support higher-level reasoning and decision-making.

In summary, compared with the behaviorist school, the semiotic school places more emphasis on global information integration and long-term planning.

4.3 Comparison and Practical Significance of Data Collection between Schools

As discussion above, by comparison, it can be seen that:

1) The behaviorist school emphasizes immediate response and local adaptability, suitable for rapid decision-making in complex dynamic environments, but has relatively limited global planning capabilities;
2) The Symbolist school emphasizes global information integration and logical reasoning, which is suitable for long-term planning and macro decision-making, but its ability to respond to real-time dynamic changes may be limited.

In practical applications, mainstream auto drive system tends to integrate the advantages of the two genres, which not only retain the real-time perception ability of behaviorism, but also rely on the overall planning ability of symbolism to achieve more reliable and intelligent driving decisions [8].

5 Algorithms for Decisions

In the auto drive system, data collection is only the first step. How to process these data and form executable decisions is the core of the whole closed-loop process [9]. Existing technologies typically use deep learning and probabilistic models to input multimodal data into algorithm modules for feature extraction, information fusion, and decision generation, thereby guiding the selection and execution of vehicle actions. In practical applications, different schools also have different priorities in the decision-making process, and modern auto drive system often integrate multiple methods to improve the accuracy and reliability of decision-making.

5.1 Fusion Algorithm Architecture

Taking the current typical three-stage architecture as an example: **QT Farmer** aggregates long-term historical information + **LLM scenario inference** + **VAE generation planner**, achieving significant improvement in closed-loop performance [10]. Each of them performs its own duties, forming a relatively complete information processing loop:

1) **Basic principle of QT Farmer**: This module is divided into three stages: visual feature interaction, historical context fusion, and internal information integration. The visual feature interaction stage utilizes Scene Query and Perception Query to extract global semantic and dynamic factor information of the current frame, respectively, and achieves information exchange through cross attention mechanism. The historical context fusion stage enhances spatiotemporal continuity through time encoding, enabling the model to understand the evolution of actions over time. The internal information integration stage

refines and outputs all information to provide high-quality features for subsequent decision-making.

2) **LLM (Large Language Model) principle**: Based on the Transformer architecture, LLM utilizes self-attention mechanism to identify the weight relationships of various parts in the input information, while capturing sequence context through multi-layer stacking and positional encoding. During the training process, a strategy of "pre training + fine-tuning" is adopted to enable the model to grasp general rules while also being competent for inference tasks in specific scenarios. This module is mainly used for environmental semantic understanding and strategy reasoning, such as predicting potential risks or selecting the optimal path.
3) **VAE (Variational Autoencoder) principle**: VAE is a generative model that combines deep learning and probabilistic modeling. It enhances the model's generation ability by introducing probability distributions, avoiding the problem of traditional AE only remembering samples. In autonomous driving scenarios, VAE can generate reasonable future scene images based on historical data and environmental conditions. Even in extreme weather, complex congestion, or accidents, it can assist vehicles in accurately determining paths and making decisions.

Obviously, the use of QT Form + LLM + VAE in the above approach is a typical processing method, which does not belong to any particular genre. As for specific genres, there is no absolute division of genres, and no car company is solely pursuing a certain genre. Google's Waymo, to a certain extent, has algorithms and core ideas that are close to symbolism. It uses the SLAM model, collects data using LiDAR, and uses it to build maps. The core logic is to allow vehicles to locate themselves in unfamiliar places, constantly understand their location, and construct the layout of surrounding maps while using cameras to find feature points in the environment.

In terms of algorithms, it uses Kalman filtering to process updates of its own position, allowing it to know where it is at all times. In the behaviorist genre, early Tesla was very close to the behaviorist genre, which focused on using CNN.

The logic is to use the car's built-in camera to observe and collect images, while CNN uses convolutional layers to extract object images, recognize their edges, textures, and other features, pool for dimensionality reduction, and fully connected layers for object classification and recognition.

5.2 Characteristics of Different Schools

Although modern systems typically adopt a fusion architecture, the characteristics of traditional schools in the decision processing stage still deserve attention:

Symbolism school: emphasizes logical reasoning and global planning. For example, Waymo uses SLAM technology to build high-precision maps, collects environmental information through LiDAR and cameras, updates vehicle positions using Kalman filtering, and achieves global environmental modeling and precise positioning.

Behaviorism school: emphasizes real-time perception and action response, for example, early Tesla used CNN to directly process camera images, recognize surrounding objects from features such as edges and textures, make quick action decisions, and adapt to dynamic environmental changes.

6 Conclusion

This paper introduces the self-evolution of deep learning algorithms in autonomous driving. Results demonstrate that autonomous iteration and learning can effectively cope with complex and changing road conditions, improving system safety and reliability. Deep learning algorithms offer significant advantages in processing massive amounts of road data and information from unique scenarios, providing crucial support for the development of intelligent transportation.

References

1. Nguyen, H., Kieu, L. M., Wen, T., & Cai, C. (2018). Deep learning methods in transportation domain: A review. IET Intelligent Transport Systems, 12(9), 998-1004
2. Rosenzweig, J., & Bartl, M. (2015). A review and analysis of literature on autonomous driving. E-Journal Making-of Innovation, 1-57
3. Zhao, J., Zhao, W., Deng, B., Wang, Z., Zhang, F., Zheng, W., ... & Burke, A. F. (2024). Autonomous driving system: A comprehensive survey. Expert Systems with Applications, 242, 122836
4. Chen, L., Wu, P., Chitta, K., Jaeger, B., Geiger, A., & Li, H. (2024). End-to-end autonomous driving: Challenges and frontiers. IEEE Transactions on Pattern Analysis and Machine Intelligence
5. Liao, B., Chen, S., Yin, H., Jiang, B., Wang, C., Yan, S., ... & Wang, X. (2025). DiffusionDrive: Truncated diffusion model for end-to-end autonomous driving. In Proceedings of the Computer Vision and Pattern Recognition Conference (pp. 12037-12047)
6. Deng, Y., Tu, Z., Yao, J., Zhang, M., Zhang, T., & Zheng, X. (2025). Target: Traffic rule-based test generation for autonomous driving systems. IEEE Transactions on Software Engineering
7. Cui, C., Ma, Y., Cao, X., Ye, W., Zhou, Y., Liang, K., ... & Zheng, C. (2024). A survey on multimodal large language models for autonomous driving. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 958-979)
8. Marcu, A. M., Chen, L., Hünermann, J., Karnsund, A., Hanotte, B., Chidananda, P., ... & Sinavski, O. (2024). LingoQA: Visual question answering for autonomous driving. In European Conference on Computer Vision (pp. 252-269). Cham: Springer Nature Switzerland
9. Wu, D., Han, W., Liu, Y., Wang, T., Xu, C. Z., Zhang, X., & Shen, J. (2025). Language prompt for autonomous driving. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 39, No. 8, pp. 8359-8367)
10. Zheng, W., Song, R., Guo, X., Zhang, C., & Chen, L. (2024). GenAD: Generative end-to-end autonomous driving. In European Conference on Computer Vision (pp. 87-104). Cham: Springer Nature Switzerland