Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 15, No. 3, 2026

A Review of Key Technologies for Deep Learning-Based Autonomous Driving

Mengfei Li

Shanghai University of Science and Technology, Shanghai, 200093, China

Abstract: With the rapid development of artificial intelligence, computer vision, and intelligent transportation technologies, autonomous driving has become an important research direction in academia and industry. Deep learning, with its powerful feature extraction, pattern recognition, and temporal modeling capabilities, has shown significant advantages in autonomous driving environmental perception, trajectory prediction, decision planning, and vehicle control. Convolutional neural networks have played a crucial role in object detection and semantic segmentation tasks, effectively improving the understanding of complex road scenarios by autonomous driving systems. Recurrent structures such as Long Short-Term Memory networks have high application value in temporal modeling and trajectory prediction. Transformer models, with their self-attention mechanism, excel in long-distance dependency modeling, multimodal fusion, and high-level decision planning. Furthermore, end-to-end autonomous driving methods, by integrating perception, decision-making, and control into a unified learning framework, provide new ideas for optimizing the overall performance of autonomous driving systems. This paper reviews the key applications of deep learning in autonomous driving, systematically analyzes its research progress in environmental perception, path decision-making, vehicle control, and end-to-end autonomous driving, and summarizes the main challenges and future development trends, aiming to provide a reference for related research.

Keywords: Deep Learning; Autonomous Driving; Environmental Perception; Trajectory Prediction; End-to-end Learning.

1. Introduction

Autonomous driving technology aims to enable vehicles to perceive, understand, make decisions, and control the external environment through sensors, computing platforms, and intelligent algorithms, thereby reducing errors caused by human operation and improving traffic efficiency and driving safety. In recent years, with the rapid development of vehicle-mounted sensors such as cameras, lidar, and millimeter-wave radar, as well as the continuous progress of high-performance chips and deep learning algorithms, autonomous driving systems have made significant progress in perception accuracy, planning capabilities, and control stability [1][2].

In autonomous driving systems, environmental perception, prediction planning, and vehicle control usually constitute a complete technical chain. Environmental perception is responsible for identifying key information such as surrounding vehicles, pedestrians, lane lines, traffic signs, and road boundaries; the prediction and planning module infers the future behavior of surrounding traffic participants based on the dynamic environment and generates reasonable driving strategies; the control module outputs steering, braking, and acceleration commands based on the planning results to realize vehicle execution. Traditional methods often rely on manually designed features and modular serial optimization, while deep learning can significantly improve the robustness and generalization ability of the system in complex scenarios through end-to-end feature learning and multi-layer representation modeling [1][2][11][12].

From the perspective of model evolution, convolutional neural networks (CNNs) have driven the development of image recognition, object detection, and semantic segmentation technologies, becoming the foundational models for environmental perception in autonomous driving; temporal models such as long short-term memory networks (LSTMs) have enhanced trajectory prediction and behavior modeling capabilities; and Transformers and their derivative architectures have shown great potential in multimodal fusion, long-term dependency modeling, and bird's-eye view representation learning.

Therefore, systematically reviewing the key technologies and representative advancements of deep learning in autonomous driving, and analyzing its current challenges and future trends, has strong theoretical significance and engineering value. This paper will discuss these aspects from five perspectives: environmental perception, decision-making and control, end-to-end learning, current challenges, and development trends.

2. Overall Process of Deep Learning-Based Autonomous Driving

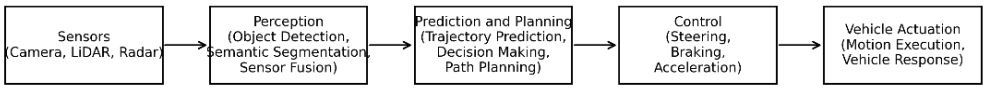

Fig 1. Overall pipeline of autonomous driving based on deep learning

Autonomous driving systems can generally be summarized as a technical process of "sensor input - environmental perception - prediction and planning - control execution". The vehicle camera, lidar and millimeter-wave radar first collect external environmental information; the perception module completes target detection, semantic segmentation, instance segmentation and fusion perception based on deep learning models; then the prediction and planning module combines

road topology, traffic rules and dynamic obstacle status to predict future trajectories and generate the optimal path; finally, the control module outputs control commands such as steering angle, braking force and acceleration according to the planning results, so that the vehicle can drive according to the expected strategy [1][2].

3. Applications of Deep Learning in Environmental Perception for Autonomous Driving

3.1. Visual Feature Extraction Based on CNN

Environmental perception is the foundation of autonomous driving systems, and its performance directly affects subsequent decision-making, planning, and control effects. In early methods, researchers typically relied on manually designed features such as edges, textures, and color histograms for scene understanding, but these methods have limited robustness in complex road environments. With the development of deep learning, CNNs, through multi-layer convolution, pooling, and nonlinear mapping, can automatically learn hierarchical visual features with discriminative capabilities, significantly improving the performance of image classification and perception tasks [3][5]. AlexNet first demonstrated the significant advantages of deep convolutional networks over traditional visual methods in the ImageNet challenge. VGGNet further improved feature representation capabilities by increasing the number of network layers, while ResNet effectively alleviated the problem of training difficulties in deep networks by utilizing residual connections. These basic models provide important network design ideas for autonomous driving visual perception tasks and have become the core backbone networks for detection, segmentation, and multi-task learning models [3][4][5].

3.2. Application of Object Detection in Autonomous Driving

The main task of object detection is to identify vehicles, pedestrians, cyclists, traffic lights, and traffic signs in images or point clouds, and to give their categories and locations. This task is directly related to the autonomous driving system's ability to understand dynamic environments. Deep learning detection models are mainly divided into two categories: two-stage detectors and single-stage detectors [6][7].

Faster R-CNN, as a typical two-stage detector, generates candidate boxes through a region proposal network, and then performs classification and regression. It has high accuracy and is suitable for scenarios with high requirements for detection accuracy [6]. The YOLO series belongs to a typical single-stage detection framework. Its core idea is to unify target localization and classification into one network, thereby greatly improving the inference speed and making it more suitable for the real-time requirements of autonomous driving. In complex traffic environments, this type of detection model can identify vehicles in front, vehicles approaching from the side, pedestrians, and traffic facilities, providing key information for collision warning, following control, and path planning.

In addition, as the complexity of autonomous driving perception scenarios increases, object detection has gradually expanded from two-dimensional images to three-dimensional spatial perception. The camera and lidar fusion detection method is becoming an important trend. Its core objective is to combine image semantic information and point cloud geometric information to improve the ability to identify occluded, small and distant targets[15].

3.3. Semantic Segmentation and Scene Understanding

Compared with object detection, semantic segmentation focuses more on pixel-level scene parsing. Its goal is to assign semantic category labels to each pixel in the image, such as roads, lane lines, pedestrian areas, buildings, green belts, and the sky. For autonomous driving, semantic segmentation helps vehicles understand drivable areas, boundary information, and the distribution of dynamic obstacles, thereby improving the accuracy of path planning and obstacle avoidance [8].

DeepLabv3+ enhances detail recovery capabilities while preserving high-level semantic information through dilated convolution and encoder-decoder structures, and has become one of the representative methods in the field of semantic segmentation. In autonomous driving scenarios, semantic segmentation models are often used for tasks such as drivable road area extraction, lane line recognition, and scene understanding, and often work in conjunction with detection models to jointly improve environmental perception capabilities.

3.4. Multi-sensor Fusion Perception

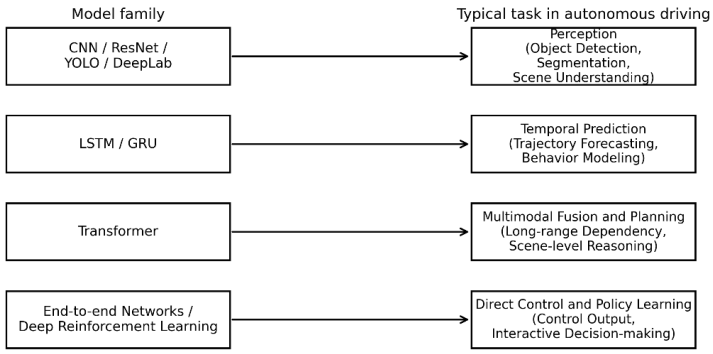

Fig 2. Mapping between deep learning models and autonomous driving tasks

While relying solely on monocular or binocular vision is relatively inexpensive, it has significant limitations in long-

distance ranging, nighttime scenes, rainy/foggy weather, and occlusion conditions. To improve the stability of the system in complex environments, more and more studies are adopting multi-sensor fusion schemes using cameras, lidar, and millimeter-wave radar. The advantage of multi-sensor fusion is that it can complement information from different modalities, thereby achieving a better balance between accuracy, robustness, and scene adaptability [15].

In recent years, fusion methods based on Transformer have shown strong potential. Models represented by TransFusion integrate LiDAR and camera features in a unified architecture, achieving good results in 3D object detection tasks. Compared with traditional fusion methods based on manual rules or simple feature stitching, these methods usually have stronger expressive power and higher detection robustness in complex road scenes.

4. Applications of Deep Learning in Decision Planning and Vehicle Control

4.1. Trajectory Prediction Based on Time Series Model

In autonomous driving systems, simply perceiving the current environment is insufficient; the system also needs to predict the future behavior of surrounding traffic participants based on their motion states and formulate safe and reasonable driving strategies accordingly. Therefore, trajectory prediction and behavior modeling have become important research topics in autonomous driving [1].

Since the motion of vehicles and pedestrians is inherently time-series, recurrent neural networks (RNNs) have been widely used for trajectory prediction tasks. However, traditional RNNs are prone to gradient vanishing and insufficient long-term dependency modeling when processing long sequences. LSTM, by introducing input gates, forget gates, and output gates, significantly enhances the model's ability to retain and update long-term information, making it more suitable for processing time-series data in complex dynamic traffic scenarios [9].

In autonomous driving, LSTM and its variants are often used for predicting the trajectory of vehicles ahead, modeling the behavior of surrounding traffic participants, and analyzing short-term path change trends. These methods can use the position, speed, and direction information from several past moments to predict the future trajectory of a target, thus providing a basis for the vehicle's following, lane changing, and obstacle avoidance decisions.

4.2. Application of Transformer in High-Level Decision-Making

Compared to recursive-based temporal models, Transformer directly establishes global dependencies within a sequence through a self-attention mechanism, avoiding the efficiency limitations caused by sequential computation of RNNs, while improving the ability to model long-distance spatiotemporal relationships [10].

In autonomous driving scenarios, Transformer has two prominent advantages. First, it can more effectively capture the interaction relationships between traffic participants in long-term temporal modeling, making it suitable for behavior prediction and high-level planning in complex scenarios. Second, it is naturally suitable for multimodal information fusion, capable of simultaneously processing heterogeneous information from multiple cameras, point clouds, radar, and maps, thereby improving the system's understanding of complex environments [14].

BEVFormer is one of the representative works in the application of Transformer in autonomous driving. This method utilizes multi-view images and temporal features to construct a bird's-eye view representation, enabling the model to perform environmental understanding and scene modeling under unified spatial coordinates, providing a more consistent intermediate representation for detection, tracking, and planning. This indicates that autonomous driving research has gradually shifted from single-task models to multi-task collaborative learning frameworks based on unified representations.

4.3. Vehicle Control and Intelligent Decision Making

After completing environmental perception and trajectory prediction, the autonomous driving system ultimately needs to be transformed into executable control commands, including steering wheel angle, throttle opening, and braking force distribution. Traditional control methods usually combine vehicle dynamics models and optimization algorithms to achieve tracking control, while deep learning methods attempt to improve the adaptive capability of control by learning the mapping relationship between environmental state and control output [2][11].

In some studies, deep neural networks have been used to learn lateral and longitudinal control strategies; in more complex scenarios, reinforcement learning has been further used to enable agents to learn optimal control behavior through interaction with the environment. These methods are particularly suitable for high-complexity tasks such as intersection passage, overtaking, merging, and dynamic obstacle avoidance, but their training cost, sample efficiency, and safety constraints remain core issues in real-world deployment [13].

5. End-to-end Autonomous Driving Approach

Traditional autonomous driving systems mostly adopt a modular architecture, designing perception, localization, prediction, planning, and control separately. The advantage of this approach is that the structure is clear, easy to debug and interpret, but errors between different modules may propagate step by step, thus affecting the final system performance [11].

End-to-end autonomous driving methods attempt to directly establish a mapping between "raw sensor input - driving control output" using a unified model, thereby reducing human design and information loss between modules. NVIDIA's classic work "End to End Learning for Self-Driving Cars" shows that deep neural networks can directly learn from forward-looking camera images and turn to control signals, laying a representative foundation for end-to-end autonomous driving research [12].

In recent years, end-to-end autonomous driving research has continued to expand, and some works have begun to integrate multimodal input, attention mechanisms, imitation learning, and reinforcement learning into a unified framework to improve the model's adaptability in complex traffic environments. Compared with traditional modular systems, end-to-end methods have theoretical advantages in overall

optimization, but they also face problems such as insufficient interpretability, high training data requirements, difficulty in handling long-tail scenarios, and complex safety verification [11][13].

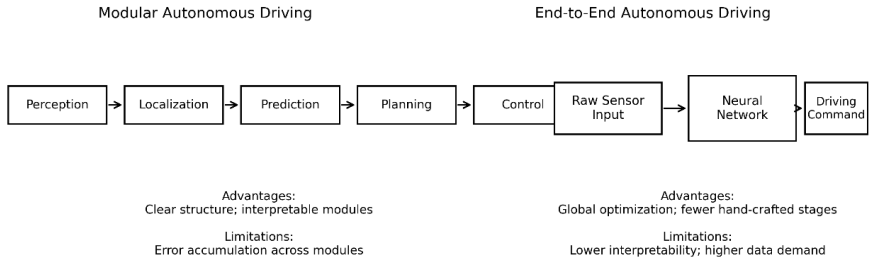

Fig 3. Comparison between modular autonomous driving and end-to-end autonomous driving.

6. The Role of Deep Reinforcement Learning in Autonomous Driving

Deep reinforcement learning learns decision-making strategies by maximizing cumulative rewards during continuous interaction with the environment, and has therefore received widespread attention in high-level behavioral decision-making, path selection, and interactive driving scenarios in autonomous driving.

Compared to supervised learning, reinforcement learning does not require human experts to label each state, but instead continuously optimizes behavioral strategies through reward mechanisms. This gives it potential advantages in complex traffic scenarios, such as autonomous lane changing, merging, intersection games, overtaking, and congestion environment decision-making [13]. However, the application of reinforcement learning in autonomous driving still faces several challenges, including high training sample costs, high risk of trial and error on real roads, insufficient policy stability, and difficulty in explicit modeling of safety constraints. Therefore, at the current stage, reinforcement learning is more often trained in simulation environments and combined with imitation learning, rule systems, or modular planning methods.

7. Current Challenges

Despite the significant advancements in deep learning technology, its practical application still faces numerous challenges.

First, there is a contradiction between model complexity and real-time performance. High-precision perception and fusion models typically have large parameter scales, placing high demands on computing resources and power consumption. However, the limited computing power and energy consumption budget of in-vehicle platforms constrain the deployment of high-performance models [14][15].

Second, autonomous driving is highly dependent on training data, but extreme weather, long-tail risk events, and rare interaction scenarios are sparsely distributed in real-world data, easily leading to insufficient generalization ability of models in actual roads. Even if a model performs well on public datasets, it may still fail to recognize objects, exhibit biased behavior predictions, and make decision-making errors in real-world complex environments.

Third, deep learning models generally suffer from insufficient interpretability. As a high-safety scenario, autonomous driving requires models to provide sufficiently transparent decision-making basis for safety verification, regulatory adaptation, and liability determination. However, most deep networks still exhibit significant "black box" characteristics, posing difficulties for engineering deployment and supervision [2][13].

Finally, autonomous driving systems are also constrained by issues such as regulations and ethics, safety redundancy, and the migration from simulation to real vehicles. How to achieve a balance between algorithm performance, system complexity, and commercial viability while ensuring safety remains a crucial research topic.

8. Future Development Trends

To address the aforementioned issues, the future development of deep learning in autonomous driving can be broadly summarized in the following directions:

First, lightweight models and efficient deployment will continue to receive significant attention. Model compression, knowledge distillation, sparsification, and efficient inference techniques for in-vehicle platforms will become crucial means to improve real-time performance and reduce power consumption.

Second, multimodal fusion and unified representation learning will be further deepened. Future autonomous driving systems will not only rely on multi-sensor fusion but will also emphasize the construction of a unified feature space across modalities to reduce information fragmentation between different modules. A unified framework based on bird's-eye view representation, spatiotemporal Transformer, and multi-task learning is expected to continue to develop.

Third, the integration of end-to-end and modular methods may become a more realistic technical route. While pure end-to-end methods possess overall optimization advantages, they currently still have shortcomings in terms of safety, interpretability, and verifiability. Therefore, hybrid systems combining explicit intermediate representations, rule constraints, and data-driven learning are more likely to emerge in the future.

Fourth, the importance of simulation platforms and closed-loop testing systems will be further enhanced. To reduce the risks of real-world road testing and improve coverage of long-tail scenarios and hazardous events, researchers will increasingly rely on high-fidelity simulation environments, data synthesis, and hybrid virtual-real training strategies.

Overall, deep learning research in autonomous driving is gradually moving from optimizing single perception models to a new stage of multimodal fusion, unified representation, closed-loop decision-making, and system-wide collaborative optimization.

9. Conclusion

Deep learning has become one of the core driving forces behind the development of autonomous driving technology. At the environmental perception level, CNNs and their derivative models have significantly improved object detection and semantic segmentation capabilities; at the prediction and planning level, temporal models such as LSTM have enhanced trajectory prediction and behavior modeling capabilities, while Transformer has further improved long-term dependency modeling and multimodal fusion levels; at the system architecture level, end-to-end learning provides new technical approaches for the overall optimization of autonomous driving.

Meanwhile, deep learning in autonomous driving still faces practical challenges such as limited computing power, data scarcity, insufficient interpretability, and difficulties in safety verification. Future research should continue to advance in areas such as efficient model design, unified multimodal representation, closed-loop decision optimization, simulation testing systems, and secure and reliable mechanisms. It is foreseeable that with the synergistic development of algorithms, hardware, and systems engineering, deep learning will play an even more important role in the field of autonomous driving.

References

[1] S. Grigorescu, B. Trasnea, T. Cocias, and G. Macesanu, "A Survey of Deep Learning Techniques for Autonomous Driving," *Journal of Field Robotics*, vol. 37, no. 3, pp. 362–386, 2020.

[2] E. Yurtsever, J. Lambert, A. Carballo, and K. Takeda, "A Survey of Autonomous Driving: Common Practices and Emerging Technologies," *IEEE Access*, vol. 8, pp. 58443–58469, 2020.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[4] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *International Conference on Learning Representations*, 2015.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[6] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 91–99.

[7] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7263–7271.

[8] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 801–818.

[9] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[10] A. Vaswani *et al*., "Attention Is All You Need," in *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.

[11] P. S. Chib and P. Singh, "Recent Advancements in End-to-End Autonomous Driving using Deep Learning: A Survey," *arXiv preprint* arXiv:2307.04370, 2023.

[12] M. Bojarski *et al*., "End to End Learning for Self-Driving Cars," *arXiv preprint* arXiv:1604.07316, 2016.

[13] B. R. Kiran, I. Sobh, V. Talpaert, P. Mannion, A. A. Al Sallab, S. Yogamani, and P. Pérez, "Deep Reinforcement Learning for Autonomous Driving: A Survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 6, pp. 4909–4926, 2022.

[14] Z. Li, W. Wang, H. Li, *et al*., "BEVFormer: Learning Bird's-Eye-View Representation from Multi-Camera Images via Spatiotemporal Transformers," in *Proceedings of the European Conference on Computer Vision*, 2022.

[15] X. Bai, Z. Hu, X. Zhu, Q. Huang, Y. Chen, H. Fu, and C. Tai, "TransFusion: Robust LiDAR-Camera Fusion for 3D Object Detection with Transformers," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 1803–1812.