Computers and Artificial Intelligence；Vol.2, No.1; 2025
E-ISSN: 3006 -3728 P-ISSN：3006-371X
DOI: https://doi.org/10.70267/cai.25v2n1.0815
Published by: Zeus Press

The Current Development and Future Prospects of Autonomous Driving Driven by Artificial Intelligence

Zichen Fu*

Affiliated High School of Jiangxi Normal University International Department, Nanchang, Jiangxi 330013, China

**Corresponding author: Zichen Fu, E-mail: 3059421890@qq.com.*

Abstract

This paper explores the application and development of artificial intelligence in autonomous driving and analyses its current status, challenges, and future trends. Autonomous driving systems integrate multiple core technologies in vehicle perception and driving decision-making, achieving a leap from assisted driving to commercial deployment. Leveraging emerging methods such as machine learning, deep learning, and reinforcement learning, autonomous driving systems have significantly improved perception accuracy, decision-making capabilities, and environmental adaptability. However, current autonomous driving systems still face technical bottlenecks, including insufficient model generalizability and low training efficiency, while also encountering legal and societal challenges such as data privacy protection, accident liability determination, and algorithmic ethical biases. In the future, high-precision multimodal perception architectures, edge computing deployment solutions, and the construction of a vehicle–road collaborative ecosystem will be key breakthrough directions for enabling fully autonomous driving across all scenarios.

Keywords

autonomous driving, artificial intelligence, multimodal perception, reinforcement learning, intelligent transportation system

1. Introduction

As a core component of future intelligent transportation systems, autonomous driving technology has gradually attracted the attention of the public. Its development not only transforms traditional driving models but also has led to revolutionary changes in various fields, such as modern transportation, logistics, and urban management. Artificial intelligence (AI), as one of the key enablers of autonomous driving, provides advanced perception, decision-making, control, and optimization capabilities, offering crucial support for the advancement of autonomous driving.

By enhancing computational power and expanding algorithmic applications, AI technology has continuously driven innovation in autonomous driving systems. Across multiple stages of autonomous driving, AI algorithms play a pivotal role in pushing the technology toward higher levels of automation. Although existing research focuses primarily on individual technical aspects and lacks a systematic analysis of AI technology frameworks and industrial implementation, the challenges in technology, regulations, and ethics remain significant. Nevertheless, with ongoing research and continuous efforts from various sectors, achieving fully autonomous driving is gradually becoming a reality.

This paper innovatively discusses the topic from three dimensions: technological evolution, industrial bottlenecks, and societal impact. A review of the latest advancements in artificial intelligence within the field of autonomous driving aims to reveal key technological pathways and provide both theoretical foundations and practical insights for building safe and reliable autonomous driving systems.

2. Overview of Autonomous Driving Technology

2.1 Definition and Development of Autonomous Driving

Autonomous driving technology refers to an integrated intelligent system that enables a vehicle to autonomously complete environmental perception, behavioral decision-making, and vehicle control without requiring continuous intervention from a human driver. This is achieved through core technologies such as multisensor fusion, environmental perception, path planning, and decision control. According to the J3016 standard established by the International Society of Automotive Engineers (SAE), autonomous driving technology is classified into six levels (L0--L5), with L3 and above systems capable of performing all dynamic driving tasks under specified conditions. L5 represents the fully autonomous driving capability. The technical framework encompasses four key components: the environmental perception module (LiDAR, cameras, millimeter-wave radar, etc.), positioning and mapping module (high-precision maps, SLAM technology), decision planning module (behavior prediction, path planning), and vehicle control module (steering control execution system).

Research into autonomous driving can be traced back to the 1950s, when studies focused primarily on achieving basic driver assistance systems. In 1984, a research team from Karlsruhe University in Germany launched the SV-1 autonomous vehicle prototype, which uses LiDAR for environmental perception. In 1994, Carnegie Mellon University's Navlab project successfully developed the first autonomous vehicle capable of navigating independently. In 1995, Stanford University's Stanford Cart project demonstrated a vehicle that could autonomously avoid obstacles in dense environments. In 2004, the U.S. Defense Advanced Research Projects Agency (DARPA) held the first autonomous driving challenge. In 2005, the winning team of the DARPA Challenge successfully completed a full-length autonomous drive.

In the 2010s, Google introduced the first autonomous vehicle prototype, and Tesla launched an autopilot system with autonomous driving capabilities, pushing the commercialization of autonomous driving technology. In 2016, Uber began trial operations of autonomous taxis, further advancing the application of autonomous driving in the shared mobility sector. In 2020, Waymo became the first company to launch Level 4 autonomous taxis globally, marking the beginning of fully autonomous driving commercialization. Today, autonomous driving technology has been deployed worldwide and is gradually entering a broader commercialization phase.

2.2 Components of Autonomous Driving Systems

A complete autonomous driving system typically consists of four core components: the perception layer, decision layer, execution layer, and communication layer. These layers work together to achieve the various functions of autonomous driving.

The perception layer is the foundation of the autonomous driving system. It collects data from the surrounding environment through various sensors (such as LiDAR, millimeter-wave radar, cameras, and ultrasonic sensors) and uses computer vision technologies (such as semantic segmentation, object detection, and lane line recognition) and environmental modelling (such as high-precision maps and SLAM) to help the system understand complex spatial structures and vehicle positions. This provides data support for subsequent decision-making and execution.

The decision layer plans and makes decisions about the vehicle's actions on the basis of perception data. This layer involves multiple complex algorithms and models to complete three key tasks: (1) Path planning: Considering factors such as road type, traffic signals, and potential obstacles, it calculates the optimal driving route for the autonomous vehicle. (2) Behavior prediction: Predicting the acceleration, deceleration, and steering actions of surrounding vehicles on the basis of traffic conditions. (3) Real-time decision-making:

Decision algorithms such as deep reinforcement learning, model predictive control (MPC), and game theory are used to make immediate responses.

The execution layer translates decisions into specific actions to ensure that the vehicle follows the planned path. This layer consists of two key components: (1) Vehicle control system: This layer regulates the vehicle's acceleration, braking, and steering through control algorithms such as PID control and fuzzy control. (2) Actuators: Actuators such as the electric steering system and electronic braking system receive control commands and execute precise movements to achieve vehicle motion.

The communication layer ensures real-time information exchange and updates between the vehicle, other vehicles, road infrastructure, and cloud platforms. Vehicle-to-everything (V2X) communication technology enables the vehicle to exchange information with other vehicles, traffic lights, and road cameras, enabling information sharing and collaborative control. Cloud platforms store vehicle operation data, and real-time communication between the vehicle and the cloud platform helps update maps, optimize decision algorithms, and enable intervehicle collaboration.

3. Application of Artificial Intelligence in Autonomous Driving

AI technology has become an indispensable part of autonomous driving systems. It not only makes autonomous driving smarter and more autonomous but also improves the system's safety and efficiency.

3.1 Computer Vision and Deep Learning

Computer vision technology enables vehicles to understand the image information of the surrounding environment and is one of the key technologies in the perception layer. By using cameras, autonomous driving systems can capture real-time image data and recognize surrounding traffic signs, lane markings, pedestrians, other vehicles, and obstacles. Computer vision tasks typically include object detection, image classification, semantic segmentation, and depth estimation, all of which provide crucial perception data for the decision layer of autonomous driving.

The basic principle of computer vision is to extract useful feature information from images through image processing and analysis and then classify, recognize, and locate these features. Taking object detection as an example, the process begins with preprocessing the raw image, such as denoising, color space conversion, and image enhancement, to improve the accuracy of subsequent algorithms. Then, convolutional neural networks (CNNs), which represent the target objects in latent space, are used to extract feature vectors from the image. Finally, pretrained deep learning models are used to recognize and locate targets such as pedestrians, vehicles, traffic signs, and traffic signals in real-time images. Accurate recognition of these targets is the foundation for decision-making in autonomous driving systems.

Deep learning, particularly convolutional neural networks (CNNs), has become the main technology for solving computer vision tasks. The structure of a CNN typically includes convolutional layers, pooling layers, and fully connected layers. In the convolutional layer, the input image is convolved with multiple filters to extract different visual features, such as edges, textures, and corners. The pooling layer is used to reduce the dimensionality of the convolution results, retaining the most significant features, which reduces the computational load and prevents overfitting. Through multiple layers of convolution and pooling, features of the input image are progressively extracted, resulting in a set of feature vectors that represent the content of the image. The fully connected layer combines the features extracted from the convolution and pooling layers and outputs the final prediction result.

CNNs can automatically learn image features from a large amount of labelled training data without the need for manually designed feature extractors, which makes them perform well in tasks such as image classification and object detection. Tesla's full self-driving (FSD) system uses the HydraNet multitask architecture, with a backbone network based on an improved EfficientNet model. This model reduces the computational requirements by 75% through depthwise separable convolutions, achieving an end-to-end inference speed of 8 ms on the NVIDIA DRIVE platform. Baidu's Apollo system innovatively applied a 3D-CNN architecture in complex intersection scenarios. This network combines multiview camera data with

LiDAR point clouds into spatiotemporal voxels to construct a 360-degree 3D obstacle model in real time. In the Yizhuang testing area of Beijing, this system achieved a pedestrian recognition accuracy of 98.5% at intersection crossings, a 21% improvement over traditional 2D detection methods, effectively solving the problem of missed detections caused by vehicle occlusion.

3.2 Sensor Fusion and Multimodal Perception

Autonomous driving systems are typically equipped with multiple types of sensors, such as cameras, millimeter-wave radar (Radar), and LiDAR. Each sensor has its own specific advantages, such as cameras providing low-cost, rich scene information, millimeter-wave radar performing excellently in adverse weather conditions, and LiDAR generating high-precision 3D point cloud images. Considering the limitations of a single sensor in complex environments, autonomous driving systems usually employ sensor fusion techniques to integrate data from multiple sensors. This helps overcome the shortcomings of individual sensors, improves overall perception accuracy, and enhances the system's robustness and reliability in handling exceptional situations.

One challenge faced by multimodal data fusion is the issue of data alignment and consistency. Data from different modalities have different feature dimensions and representations, and their error distributions also vary, making it difficult to efficiently fuse data and improve data quality in practical applications. The Kalman filter is a commonly used data fusion algorithm. Its core idea is to combine prior state information of the system with measurement data, performing recursive prediction and correction to gradually optimize system state estimation. In multimodal data fusion, the Kalman filter is used to weight and combine data from different sensors, eliminating noise and errors from individual sensors.

High-definition maps (HD maps) are crucial for multimodal perception. HD maps typically contain precise geometric details of roads, lane markings, traffic signs, road surface conditions, intersection shapes, slopes, and other information. In complex environments such as cities or highways, autonomous driving systems combine GPSs, IMUs, and HD maps. The GPS provides the vehicle's approximate location, the IMU offers information about the vehicle's motion state, and the HD map provides detailed information about the road and environment. By integrating these data, the system can eliminate GPS signal errors and achieve centimeter-level precision in navigation.

With the development of deep learning, models such as convolutional neural networks (CNNs), deep neural networks (DNNs), and long short-term memory networks (LSTMs) have also been applied to multimodal data fusion. Through pretraining and fine-tuning, these models can accept raw data from different modalities as inputs and extract more useful features for analysis.

3.3 Reinforcement Learning and Decision Making

Reinforcement learning (RL) is a machine learning method that learns optimal strategies through interaction with the environment. Unlike traditional supervised learning, RL does not rely on labelled training data but instead learns how to make optimal decisions in specific situations through the interaction between the agent (in the case of autonomous driving, the vehicle) and the environment, using feedback mechanisms (rewards and penalties). The result of each decision influences future states, and the agent continuously adjusts its strategy on the basis of reward and penalty signals, eventually learning how to maximize cumulative rewards in a complex environment.

In autonomous driving systems, reinforcement learning is widely applied in the decision-making and planning stages. In uncertain and dynamic traffic environments, vehicle decisions rely not only on static road information and traffic signs but also on the behavior of other traffic participants, changes in road conditions, and real-time traffic flow. Reinforcement learning offers a flexible, adaptive approach to decision-making that can dynamically adjust on the basis of environmental changes.

The successful application of reinforcement learning in autonomous driving largely depends on deep learning models, particularly their ability to process high-dimensional input spaces such as images and point cloud data. By combining deep reinforcement learning (DRL), autonomous driving systems can extract features from large amounts of raw data and make decisions on the basis of these features. DRL introduces deep neural networks, such as deep Q networks (DQNs), to automatically learn complex state-action value functions, thereby handling more complex environments and tasks. The deep deterministic policy gradient (DDPG) is also a commonly used reinforcement learning algorithm. It combines the actor–critic architecture with the DQN and uses experience replay buffers and target networks to improve learning stability and efficiency. This method is particularly suitable for problems in continuous action spaces, such as adjusting vehicle speed or selecting steering angles, making it widely applied in autonomous driving.

Additionally, time series-based deep learning models, such as long short-term memory (LSTM) networks and gated recurrent units (GRUs), are applied to dynamic object behavior prediction tasks. LSTM and the GRU can capture long-term dependencies in time series data, analyse historical data and combine them with current states to predict the future behavior of other traffic participants. These predictions provide real-time decision support to autonomous driving systems, enabling the planning of safe paths during the decision-making phase and preventing conflicts with dynamic targets.

4. Development Status and Challenges

4.1 Visual Model and Detection Algorithm Optimization

The perception layer of autonomous driving systems extensively utilizes detection algorithms based on visual models. The detection accuracy and computational efficiency of these algorithms are two important indicators of system performance. Optimizing models and algorithms to improve accuracy and reduce latency is a key issue that recent research aims to address.

With the introduction of the attention mechanism (Vaswani et al., 2017), deep learning models based on the transformer architecture have demonstrated stronger feature extraction capabilities and versatility. Madake et al. (2024) incorporated the attention mechanism into deep neural networks, which, compared with pretrained models, improved the accuracy of detecting small objects such as traffic lights and pedestrians. Self-supervised learning enables models to learn features from unlabelled data, which is highly important in the autonomous driving domain, where data annotation is challenging. Xu et al. (2024) focused on LiDAR-camera 3D perception models and proposed a self-supervised deep learning framework that introduces a unified pretraining strategy for multimodal data, effectively enhancing the model's ability to recognize targets in the presence of unlabelled data. To improve the performance of the visual system in complex environments, related research has attempted to enhance model generalization and achieve multimodal data fusion. Jiang et al. (2024) proposed a complementary network, MCNet, that integrates RGB modality data and thermal modality data. This network achieves complementary and flexible adjustments between color, texture, and contour information, ensuring high detection accuracy even when one modality's information is difficult to perceive, thus improving the system's robustness in complex weather conditions. For visual perception in autonomous driving under nighttime or low-light conditions, generative adversarial networks (GANs) have become a feasible solution. One application of generative models is image enhancement, which can effectively restore the details of low-resolution images. Pham et al. (2020) proposed the deep retinal neural network DriveRetinex, inspired by the theory of retinal image capture. This network is divided into two subnetworks, one for color image decomposition and the other for light level enhancement, improving object detection accuracy under low light conditions.

Despite the significant progress in visual model research for autonomous driving, many challenges remain. For example, the method of using GANs for image enhancement has limitations when handling complex scenes. How to implement image enhancement in dynamic environments and improve real-time performance and stability remains a pressing issue. Autonomous driving systems may encounter unexpected events, and since these events occur infrequently in actual road scenarios, existing visual models still perform poorly when faced with such events, even with the application of data augmentation and self-supervised learning methods. Future research needs to find more effective ways to improve the model's adaptability. Real-time performance is a core requirement of autonomous driving systems, but deep learning models typically consume a large amount of computational resources, which makes it difficult to meet real-time demands in complex road conditions. How to effectively utilize lightweight neural networks and hardware acceleration methods to further reduce computational consumption is an area that warrants attention.

4.2 Acceleration of Reinforcement Learning Training

In dynamic and complex traffic scenarios, deep reinforcement learning (DRL) has become a core component of the decision-making layer in autonomous driving systems because of its powerful decision-making capabilities. In recent years, much research has focused on how to improve the training efficiency of reinforcement learning algorithms, reduce the search space during training, and ensure the safety of the training process.

One challenge in reinforcement learning model training lies in the need to use real-world data as input to simulate the agent's exploration process. The amount of input data and its compliance with real-world physical laws can significantly affect the reliability of the model's decisions. Wu et al. (2023) proposed a reinforcement learning method based on uncertainty-aware models, which builds an environment model with uncertainty assessment capabilities for virtual interaction. This method outperforms both model-free and model-based reinforcement learning in terms of learning efficiency and performance. Imitation learning is an effective method for accelerating training, typically involving simulating human driving behavior and using expert demonstration data. This greatly reduces the need for random exploration. Coelho et al. (2024) reported that existing algorithms rely on offline demonstrations and that the strategies learned may not be applicable in the latest scenarios. To address this, this paper introduces a policy network that outputs two standard deviations for exploration and training, effectively bridging the distribution gap between the demonstration environment and the real environment. Hu et al. (2024) proposed a safety reinforcement learning algorithm based on short-term constraints, aiming to enhance both short-term state safety during exploration and overall decision safety. The method adds safety violation costs to the training objective, significantly reducing the likelihood of the agent taking dangerous actions.

Although reinforcement learning holds great potential in autonomous driving decision-making, deep reinforcement learning models have not yet been widely applied in commercial autonomous driving systems. The fundamental reason is the difficulty in balancing training speed, decision quality, and computational efficiency. The decision quality of reinforcement learning relies on high-quality data, and constructing a training set that can simulate all possible situations in complex and dynamic real-world traffic environments remains challenging. On the other hand, even when simulated and demonstration data are used, the training costs of models can range from several hours to days. Some models that use complex frameworks (such as the actor-critics) often have response times in the range of hundreds of milliseconds or even seconds. Finding the optimal balance among these three factors—training speed, decision quality, and computational efficiency—is an ongoing challenge that needs continuous optimization and long-term attention.

5. Future Outlook

The future of autonomous driving is full of innovation and challenges. In the coming years, AI will play an even more profound role in multiple aspects of autonomous driving, pushing the technology toward higher levels of automation and intelligence.

5.1 High-Precision Perception and Intelligent Decision-Making

With the help of high-precision sensors and more advanced perception algorithms, future autonomous driving systems will have a more accurate understanding of the surrounding environment. The continuous optimization of visual models and detection algorithms, combined with further applications of transfer learning and self-supervised learning techniques, will improve the system's adaptability in adverse weather conditions, such as rain, night-time driving, or fog. This will enhance driving safety. Currently, the main challenges faced by reinforcement learning are training efficiency and real-time decision responsiveness. In the future, distributed training, multilevel training strategies, and ensemble learning may be employed to accelerate model training. By using virtual training environments for large-scale simulations and combining expert data with real-time data from the real world, autonomous driving systems can quickly learn and optimize driving strategies to handle unexpected events.

5.2 Efficient Computing Architecture and Edge Computing

Currently, most autonomous driving systems rely on cloud computing or high-performance onboard computing platforms to process data. While cloud computing provides immense processing power, it faces limitations in terms of real-time responsiveness and bandwidth. In the future, edge computing will play a crucial role in autonomous driving by processing data and making decisions locally within the vehicle. This reduces the dependence on remote servers, decreases the data transmission latency, and enables real-time responses. The collaboration between the cloud and edge computing will achieve the best balance in processing power and data storage. To meet the increasing demands of onboard computing, AI algorithms will focus more on computational efficiency, with lightweight and highly parallel neural network models becoming mainstream.

5.3 Intelligent Vehicle Coordination and Vehicle-to-Infrastructure Collaboration

The future intelligent transportation system will not operate as a single vehicle acting independently but as a multivehicle, vehicle-to-infrastructure (V2X) coordinated system. Advancements in V2X technology will enable real-time communication of road conditions between vehicles and infrastructure, facilitating high levels of intelligent collaboration. This will increase road traffic efficiency and reduce accidents, especially at complex intersections and highways.

5.4 Legal and Ethical Issues

AI's application in autonomous driving presents vast potential, but to achieve widespread commercialization, the development of relevant laws and regulations still needs to be improved. In the future, legal frameworks will likely evolve alongside the development of autonomous driving, providing clearer standards and guidelines for technology implementation. Additionally, ethical issues, such as the "trolley problem" (the issue of selective harm), remain a major concern in society. Finding a balance between technology and ethics to improve public acceptance will become a critical direction for future development.

6. Conclusion

This paper provides an in-depth exploration of the application and development of artificial intelligence (AI) in the field of autonomous driving, systematically analysing its innovative achievements and practical applications in perception, decision-making, and coordination. Through research on key technologies such as computer vision, sensor fusion, and reinforcement learning, this paper reveals how AI drives autonomous driving technology toward higher levels of automation. The findings indicate that AI technology has not only significantly improved the environmental perception accuracy and decision-making intelligence of autonomous driving systems but also provided effective support for solving dynamic problems in complex traffic scenarios.

However, there are still some shortcomings in the current research. For example, the robustness of deep learning models in handling extreme weather conditions and unforeseen events needs further improvement, and the training efficiency and real-time responsiveness of reinforcement learning algorithms have not been fully addressed. Additionally, the widespread application of autonomous driving technology still faces challenges related to legal regulations and ethical issues.

Future research should focus on further optimizing perception algorithms and decision models, enhancing the collaboration between edge computing and cloud computing, and advancing vehicle-to-infrastructure coordination technologies. At the same time, efforts should be made to strike a balance between technology and ethics and improve relevant laws and regulations to increase societal acceptance of autonomous driving technology.

References

Coelho, D., Oliveira, M., & Santos, V. (2024). RLfOLD: Reinforcement learning from online demonstrations in urban autonomous driving. *Proceedings of the AAAI Conference on Artificial Intelligence*, *38*(10), 11660-11668. https://doi.org/10.1609/AAAI.V38I10.29049

Hu, X., Chen, P., Wen, Y., Tang, B., & Chen, L. (2024). Long and short-term constraints driven safe reinforcement learning for autonomous driving. *arXiv preprint*, arXiv:2403.18209. https://doi.org/10.48550/arXiv.2403.18209

Jiang, X., Hou, Y., Tian, H., & Zhu, L. (2024). Mirror complementary transformer network for RGB-thermal salient object detection. *IET Computer Vision*, *18*(1), 15-32. https://doi.org/10.1049/cvi2.12221

Madake, J., Lokhande, T., Mali, A., Mahale, N., & Bhatlawande, S. (2024, 8-9 May 2024). *TransVOD: Transformer-based visual object detection for self-driving cars* [Paper presentation]. 2024 International Conference on Current Trends in Advanced Computing (ICCTAC), Bengaluru, India.

Pham, L. H., Tran, D. N. N., & Jeon, J. W. (2020, 1-3 Nov. 2020). *Low-light image enhancement for autonomous driving systems using DriveRetinex-Net* [Paper presentation]. 2020 IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia), Seoul, Korea.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention is all you need* [Paper presentation]. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

Wu, J., Huang, Z., & Lv, C. (2023). Uncertainty-aware model-based reinforcement learning: Methodology and application in autonomous driving. *IEEE Transactions on Intelligent Vehicles*, *8*(1), 194-203. https://doi.org/10.1109/TIV.2022.3185159

Xu, X., Zhang, T., Yang, J., Johnson-Roberson, M., & Huang, X. (2024). Self-supervised pre-training for transferable multi-modal perception. *arXiv e-prints*, arXiv: 2405.17942. https://doi.org/10.48550/arXiv.2405.17942

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

Not Applicable.

Copyrights