Proceedings of CONF-MLA 2025 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2025.LD28603

Progress of Deep Reinforcement Learning in Autonomous Driving in the Past Three Years

Fujia Yu

School of Economics and Trade, Statistics major, Henan University of Animal Husbandry and Economy, Zhengzhou, China
fujiayu@stepbystep.freeqiye.com

Abstract. Many challenges such as environmental complexity, decision-making security, and algorithm generalization ability remain as the main problems faced by autonomous driving technology. Current research focuses on multimodal perception, end-to-end control systems, and reinforcement learning frameworks, but still has problems such as insufficient handling of long-tail scenarios, weak interpretability of black-box decisions, and high training costs. This paper's research can be deepened in three directions: integrating large language models (LLMs) with visual foundation models (VLMs) to enhance the scene understanding and few-shot generalization ability of end-to-end systems through semantic reasoning; developing hybrid learning frameworks that combine imitation learning and model-based reinforcement learning (such as the DIRL method) to reduce the demand for high-risk interactions; and building high-fidelity simulation environments to generate dynamic scenes using multimodal trajectory prompts and optimize the robustness of algorithms in extreme conditions. This paper can solve the black-box problem of autonomous driving through decision transparency enhanced by LLMs; lightweight models and hybrid training strategies can significantly reduce computational costs; the simulation supplementation of long-tail scenarios by world models will promote the implementation of safety standards and provide technical support for the commercialization of fully autonomous driving.

Keywords: Deep reinforcement learning, Autonomous driving systems, Large language models

1. Introduction

In recent years, autonomous driving technology, as a core area of the integration of artificial intelligence and transportation systems, has undergone a critical transformation from theoretical exploration to practical application. Its core objective is to achieve autonomous and safe driving in complex environments through an integrated system of perception, decision-making, and control. However, this goal faces multiple challenges: environmental dynamics (such as pedestrian and vehicle interactions), sensor heterogeneity (multi-modal data fusion), decision-making reliability (generalization ability in extreme scenarios), and system transparency (end-to-end "black box" explainability).

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.LD28603

Traditional autonomous driving systems often adopt a modular architecture (such as separation of perception, planning, and control), which, although interpretable, is prone to performance degradation due to error accumulation [1,2]. To break through this limitation, end-to-end learning has gradually become a research focus: it directly maps raw sensor data to control instructions through deep neural networks [3,4], significantly enhancing system response speed and adaptability. For instance, the DAVE-2 model achieved 90% autonomous control within a limited area [3], while deep reinforcement learning (DRL) has approached human-level performance in games and racing control [5,6], demonstrating the potential of the end-to-end framework.

However, end-to-end systems rely on massive high-quality data and efficient training paradigms. Multi-modal datasets (such as nuScenes) provide a foundation for complex scene modeling by integrating 360° perception from cameras, radars, and lidars [7]; simulation platforms (such as CARLA) support safe and controllable algorithm verification [1]. Meanwhile, reinforcement learning (RL) guides agents to learn autonomously through reward functions [8,9], but it faces issues such as low sample efficiency, sparse rewards, and safety risks [2,6].

Recently, the rise of large language models (LLMs) and vision-language models (VLMs) has injected new impetus into autonomous driving: LLMs can enhance system reasoning and few-shot learning capabilities [3,10], while VLMs can improve three-dimensional environmental understanding [11]. Virtual world models (such as DriVerse) generate high-fidelity driving scene videos through trajectory prediction and motion alignment techniques, filling the data gap for extreme cases [12]. Additionally, hybrid optimization strategies (such as imitation reinforcement learning DIRL) combine the advantages of imitation learning (IL) and model-driven reinforcement learning, stabilizing policy robustness while reducing real-world interactions [6].

This paper aims to systematically review the evolution of autonomous driving control technologies, mainly focusing on end-to-end learning, reinforcement learning paradigm innovations, multi-modal large model integration, and simulation verification systems, and analyze current challenges and future directions, providing theoretical support for technology implementation.

2. The application of end-to-end learning in autonomous driving

One of the most notable trends in the field of autonomous driving is the shift from traditional modular systems to an end-to-end learning model, demonstrating the powerful capabilities of deep learning algorithms, especially convolutional neural networks (CNNs), which can directly map raw sensory data, such as images from cameras, to control commands using DAVE-2 [3,7,8]. As shown in Figure 1:

DOI: 10.54254/2755-2721/2025.LD28603

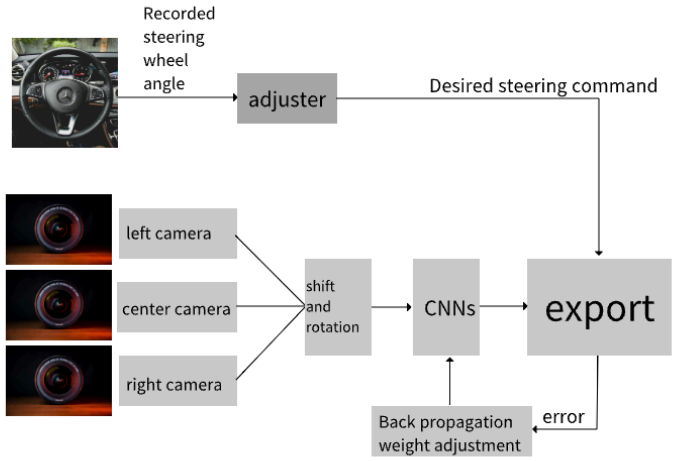

Figure 1. Training the neural network [3]

This method significantly reduces the need for modular systems and enables vehicles to learn driving behaviors with minimal human intervention in complex environments. For instance, the DAVE-2 system uses CNNs trained with driving data to learn steering actions, achieving up to 100% autonomy under certain conditions [3]. Similarly, using reinforcement learning (RL) can also enable autonomous learning of lane following tasks with minimal training data, demonstrating the potential for real-time learning and adaptation [8].

Traditional modular systems achieve this by decomposing autonomous driving tasks into subsystems such as perception, planning, and control. These systems typically operate based on pre-defined rules for each task, which may lead to cumulative and repetitive calculations between modules, resulting in a waste of computing power [2]. In contrast, end-to-end systems (such as NVIDIA's DAVE-2 and the systems explored by Wayve based on reinforcement learning) learn the entire control process within a unified large-scale complex model [4,8], thereby reducing the possibility of errors (repetitive calculations) between modules to improve efficiency. However, in practical applications, traditional modular systems still dominate because they have strong stability and transparency, especially suitable for critical tasks such as path planning [2].

3. Innovative advancements in the field of autonomous driving through deep reinforcement learning

Deep Reinforcement Learning (DRL) demonstrates revolutionary potential in the field of autonomous driving. Its core innovation lies in breaking away from the reliance on precise environmental modeling and preset rules, and shifting to a more general, adaptive, and efficient learning paradigm. Represented by the research of the Wayve team titled "Learning to Drive in a Day" [8], DRL was first proven to be capable of achieving end-to-end policy learning on real vehicles. Its breakthrough lies in:

In the first instance, Minimal Reward Design. Utilizing only the driving distance before the intervention of the safety driver as a sparse reward signal, without the need for complex lane centerline annotations or environmental maps.

DOI: 10.54254/2755-2721/2025.LD28603

In terms of the second factor, Efficiency and Authenticity. Based on the DDPG algorithm, the model only requires a small number of training rounds (approximately 30 minutes) to learn lane following from random initialization parameters, and all learning and exploration are completed on real vehicles, relying solely on monocular camera images, vehicle speed, and steering wheel angle as state inputs.

The paper "learning to drive in a day" was then read again by author, the author found that Task-Based Training Architecture still needed to be improved. This innovation proves that deep reinforcement learning is a data-based framework that does not require a model. It can directly learn driving strategies from the data input by sensors. This approach offers a new path for autonomous driving that distinguishes it from traditional modularization (perception-planning-control) or traditional imitation learning.

In addition to integration, from deep learning to enhancement, and the transition to end-to-end learning, there is innovation in autonomous driving. However, another important innovation area is the deep integration of advanced training and traditional sound control theory. This integration is primarily aimed at increasing the interference resistance and stability of the system. The value of this integration is proved by a study called "Neural Network Approach Super-Twisting Sliding Mode Control" by Kim et al. [13]:

To begin with this process, after years of research and exploration, neural networks can enhance robustness. When estimating the disturbances in the vehicle dynamics model (such as tire nonlinearity and parameter uncertainty), the radial basis function neural network (RBFNN) can perform real-time estimation of the vehicle. For example, under low adhesion road conditions ($\mu = 0.6$), RBFNN can effectively capture system changes.

In terms of the second factor, the suppression of controlling chattering. The estimated output of the RBFNN is integrated into the super-torque sliding mode control (STSMC). Compared with the traditional sliding mode control (SMC), STSMC can reduce high-frequency chattering. In fact, the precise disturbance estimation provided by the RBFNN can further reduce the amplitude of the control input, ultimately improving the smoothness of the driving. Moreover, in the simulations of the dual-lane change (DLC) and rapid path tracking (RPT) scenarios, the hybrid method (NN-STSMC) outperforms the traditional SMC and STSMC in terms of control input smoothness and path tracking accuracy.

With regard to the third aspect, the Theoretical Guarantee. Through Lyapunov function strict derivation, the control law and neural network weight update rules are ensured to maintain the stability of the closed-loop system.

These technological advancements have clarified several key development directions. Addressing the issue of efficient learning under sparse reward conditions is of vital importance in this context; take Wayve as an example, where a weakening of reward signals occurs during the later stages of policy optimization—a scenario that precisely falls into the category of sparse rewards, thus underscoring the practical necessity of resolving this problem. Another key focus lies in the integration of more prior knowledge—encompassing semantic information and depth information—and the core goal of this integration lies in improving the quality of state representation and learning efficiency, which in turn lays a more solid knowledge foundation for technical optimization. Notably, the in-depth integration of reinforcement learning with other methods (including Model Predictive Control (MPC) and adaptive control) will also serve as a critical pathway to achieving more complex and safer autonomous driving. Specifically, there are two main approaches to this integration: one involves utilizing reinforcement learning to optimize controller parameters, while the other employs reinforcement learning to provide high-level decision support. Both approaches

DOI: 10.54254/2755-2721/2025.LD28603

play a significant role in boosting the performance of autonomous driving systems, further highlighting the value of such integration.

Wayve's "Learn to Drive in One Day" and NN-STSMC's "Neural Network-Assisted Robust Control" respectively represent two complementary innovative paths. The former refers to data-driven end-to-end learning, while the latter stands for model-driven hybrid enhanced control. It is these two innovative paths that jointly drive the development of autonomous driving technology towards greater intelligence and robustness.

4. Simulation and virtual training environment technology research and development enhance autonomous driving's strengths

When the author was seeking connections among the papers, simulation platforms like CARLA and DriVerse played a unique role in training autonomous driving systems. These platforms simulate complex urban and rural traffic scenarios and provide rich datasets for training models [3,12]. For instance, CARLA offers flexible simulation settings for evaluating different autonomous driving strategies, including modular pipelines and end-to-end models using supervised and reinforcement learning methods [3]. The DriVerse model further enhances this by generating high-quality driving simulations from simple inputs (such as images and future trajectories) to achieve this goal [12].

Moreover, the simulation technology based on NVIDIA's end-to-end learning framework and DriVerse world model demonstrates significant advantages in reducing the cost and difficulty of obtaining autonomous driving data. Its core value is reflected in the following three aspects:

1. Minimal data-driven scene generation capability

Traditional bottleneck: Real-world road testing requires coverage of diverse scenarios (such as extreme weather and dangerous conditions), with data collection costs as high as $1-2 per mile (Waymo report), and the probability of obtaining long-tail scenarios (such as pedestrians crossing at night during heavy rain) is extremely low.

Technical breakthrough:

DriVerse only requires a single frame image + future trajectory to generate high-fidelity videos (FID 18.2, FVD 95.2), with geometric trajectory alignment error reduced by 55% compared to Vista, directly replacing expensive multi-sensor road testing data [12].

NVIDIA framework proves: A CNN trained with 100 hours of human driving videos can generalize to complex scenarios such as rural roads without lane markings and parking lots, significantly reducing annotation requirements [3].

2. Zero-risk coverage of high-risk scenarios

Safety redundancy design: The modular test interface built into CARLA supports custom accident scenarios (such as reversing vehicles, sudden braking), and optimizes responses through reinforcement learning training strategies (such as DDPG) [3].

Physical consistency guarantee: DriVerse's LMA module (Latent Motion Alignment) constrains the trajectories of dynamic objects through motion perception loss functions, eliminating the distortion phenomena of traditional generation models such as "vehicle teleportation", ensuring the physical credibility of virtual collision tests [12].

3. Closed-loop simulation acceleration algorithm iteration

End-to-end training revolution: The NVIDIA system achieves 98% road autonomy rate with 72 hours of real vehicle data at 30Hz real-time inference (DRIVE PX hardware), proving that the simulation environment can compress trial-and-error costs by a thousandfold [3].

In terms of adaptive scene generation, the Dynamic Window Generation (DWG) strategy adopted by DriVerse is capable of automatically switching key frames—this switching is triggered when Vt

DOI: 10.54254/2755-2721/2025.LD28603

< 0.6. It also supports the generation of long-term sequences with a duration of 30 seconds, achieving an FVD (Fréchet Video Distance) score of 281.4, which in turn provides a continuous decision-making context for planning algorithms [12].

These simulation environments play a role in alleviating the limitations inherent in real-vehicle testing. For instance, real-vehicle testing often struggles to gather sufficiently diverse data to cover all potential driving scenarios. Moreover, they enable safe testing of high-risk scenarios, such as emergency braking—tasks that would pose substantial dangers in real-world settings [3]. Beyond that, these environments can also lead to significant savings in human resources, material inputs, and financial costs.

5. The comprehensive application of large language models in autonomous driving

Recent studies have explored the possibility of integrating large language models (LLMs) into autonomous driving systems—an integration intended not only to enhance the decision-making capabilities of such systems but also to provide explanations for their actions. It is when LLMs are combined with visual models that autonomous driving systems gain a key improvement: not only can they significantly boost their understanding of open environments, but beyond that, they can address two critical issues in traditional autonomous driving solutions—on the one hand, the error accumulation that plagues traditional modular autonomous driving systems, and on the other, the inherent "black box" problem in end-to-end autonomous driving systems. Thanks to these improvements, vehicles are able to quickly adapt to unfamiliar scenarios, and to achieve this rapid adaptation, vehicles rely on few-shot learning; additionally, this also enables vehicles to make wise decisions when facing unfamiliar scenarios. This cross-application (of LLMs and autonomous driving systems) brings multiple functional supports: for instance, it allows LLMs to assist autonomous driving systems in handling complex reasoning tasks, and it also enables LLMs to provide real-time explanations of the decision logic used by autonomous driving systems—an example here is that LLMs can explain why a vehicle needs to avoid a vehicle behind it when changing lanes. Furthermore, this cross-application helps LLMs support autonomous driving systems in processing multimodal traffic information, with a specific case being that LLMs can assist in combining camera images with radar data, and the reason for combining these two types of data is to determine the intentions of pedestrians. When compared to traditional autonomous driving methods, this approach (integrating LLMs) shows clear advantages: one notable advantage is that it offers greater adaptability for autonomous driving, and another advantage is that it provides greater interpretability for the autonomous driving process.

From a safety perspective, people can analyze the application of relevant technologies: large language models (LLMs) contribute to enhancing the reliability of autonomous driving systems, while also helping to boost users' trust in such systems—specifically, LLMs achieve this by delivering transparent decision-making processes for autonomous driving systems. At the same time, techniques like super-twisting spiral sliding mode control (STSMC) serve distinct functions: that is, STSMC is capable of addressing sensor noise in autonomous driving systems and, in the meantime, managing road obstacles encountered during the vehicle's operation. When LLMs and STSMC are integrated with each other, they generate synergistic effects: firstly, this integration not only ensures stable path tracking for the vehicle but also renders the vehicle’s responses to unforeseen scenarios more auditable, thereby fully leveraging the complementary advantages of the two technologies in safeguarding the safety of autonomous driving.

That said, the application of LLMs in autonomous driving is not devoid of potential risks. One such risk lies in the model "hallucinations" of LLMs, which could prompt the autonomous driving

DOI: 10.54254/2755-2721/2025.LD28603

system to make misassessments of the driving environment; a second risk involves inherent model biases within LLMs, which might induce unfair decisions in the autonomous driving process. Additionally, there exists a risk of LLM-associated privacy breaches, a risk rooted in LLMs' handling of user data or environmental data in the autonomous driving system. Beyond these concerns, there are further challenges to autonomous driving system security: one such challenge is the "black box" characteristic of deep learning employed in LLMs, while another revolves around adversarial assaults aimed at LLMs—for instance, attackers could tamper with visual input in order to deceive LLMs, all of which present challenges to the security of the autonomous driving system.

6. Conclusion

When the author explored the recent research on the application of deep reinforcement learning in the field of autonomous driving, they found that the safety of autonomous driving remains a key challenge, especially in terms of the error handling capabilities of the control system and the ability to deal with unpredictable environments. To achieve safe and reliable control of vehicles, there is a need for powerful control algorithms that can adapt to disturbances (such as sensor noise or road obstacles, etc.). The Super-Twisting Sliding Mode Control (STSMC) technology is designed to handle disturbances while maintaining stable path tracking, thereby ensuring that vehicles can operate effectively in real-world conditions.

Overall, the development of autonomous driving systems is a highly collaborative effort involving multiple disciplines. It integrates deep learning, reinforcement learning, simulated environments, and advanced control technologies. End-to-end deep learning models, especially those using reinforcement learning, bring great potential for extensible and adaptable driving systems. However, issues such as sample efficiency, safety, and generalization ability to new environments still exist, which require continuous progress in the theory and practice of autonomous vehicle technology and further development in interdisciplinary and multi-disciplinary integration. Although the path to safe and reassuring autonomous driving technology models is long and challenging, it requires the collaborative efforts of technical talents in this field.

This article offers a review and summary of deep reinforcement learning research in the autonomous driving domain. Drawing on the substantial body of work from prior researchers, numerous forerunners have yielded invaluable contributions in theoretical framework development. Additionally, they have advanced efforts in empirical inquiry. Such work has laid the groundwork for the development of this review. This paper extend our sincere and profound thanks to these forerunners.

References

[1] Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., ... & Zieba, K. (2016). End to end learning for self-driving cars. arXiv preprint arXiv: 1604.07316.

[2] Yang, Z., Jia, X., Li, H., & Yan, J. (2023). LLM4Drive: A survey of large language models for autonomous driving. arXiv preprint arXiv: 2311.01043.

[3] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Hassabis, D. (2015). Human-level control through deep reinforcement learning. Nature, 518(7540), 529–533.

[4] Cai, P., Wang, H., Huang, H., Liu, Y., & Liu, M. (2021). Vision-based autonomous car racing using deep imitative reinforcement learning. IEEE Robotics and Automation Letters, 6(4), 7262–7269.

[5] Caesar, H., Bankiti, V., Lang, A. H., Vora, S., Liong, V. E., Xu, Q., ... & Beijbom, O. (2020). nuScenes: A multimodal dataset for autonomous driving. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 11621–11631).

DOI: 10.54254/2755-2721/2025.LD28603

[6] Kendall, A., Hawke, J., Janz, D., Mazur, P., Reda, D., Allen, J., ... & Shah, A. (2019). Learning to drive in a day. In 2019 International Conference on Robotics and Automation (ICRA) (pp. 8248–8254). IEEE.
[7] Pan, F., & Bao, H. (2021). Research progress on autonomous driving control technology based on reinforcement learning. Journal of Image and Graphics, (1).
[8] Chen, X., Peng, M., Tiu, P., Wu, Y., Chen, J., Zhu, M., & Zheng, X. (2024). GenFollower: Enhancing car-following prediction with large language models. IEEE Transactions on Intelligent Vehicles.
[9] Mengjie, W., Huiping, Z., Jian, L., Wenxiu, S., & Song, Z. (2025). Research on driving scenario technology based on multimodal large language model optimization. arXiv preprint arXiv: 2506.02014.
[10] Li, X., Wu, C., Yang, Z., Xu, Z., Liang, D., Zhang, Y., ... & Wang, J. (2025). DriVerse: Navigation world model for driving simulation via multimodal trajectory prompting and motion alignment. arXiv preprint arXiv: 2504.18576.
[11] Kim, H., & Kee, S. C. (2023). Neural network approach super-twisting sliding mode control for path-tracking of autonomous vehicles. Electronics, 12(17), 3635.