GEORGE BROWN PRESS
JOURNALS, BOOKS & PROCEEDINGS

Journal of Computer, Signal, and System Research

Article

Research on Autonomous Driving Path Planning Technology Based on Reinforcement Learning

Linfeng Hao [1,*]

[1] Robert Morris University, Moon Township, Pennsylvania, 15108, USA
* Correspondence: Linfeng Hao, Robert Morris University, Moon Township, Pennsylvania, 15108, USA

Abstract: Path planning is one of the key technical challenges in the field of autonomous driving. Thanks to the rapid advancement of reinforcement learning (RL) technology in the field of artificial intelligence, research on autonomous driving path planning based on RL is increasingly receiving attention. This study explores the use of reinforcement learning to implement autonomous driving path planning technology, analyzing the needs of multiple aspects such as environmental perception, model construction, obstacle avoidance, and path length optimization. A practical application scheme of reinforcement learning for path planning in different autonomous driving scenarios has been proposed. By comparing reinforcement learning algorithms such as DQN, A3C, PPO, etc., the adaptability and optimization ability of these algorithms in handling complex environments were explained, and the strategy of using multi-agent reinforcement learning for path planning was discussed.

Keywords: reinforcement learning; autonomous driving; path planning; Deep Q-Network; A3C algorithm

Received: 29 March 2025
Revised: 04 April 2025
Accepted: 16 April 2025
Published: 20 April 2025

1. Introduction

With the rapid advancement of computers, the Internet of Things, artificial intelligence algorithms, and automation technology, autonomous driving technology is also continuously upgrading and iterating, gradually becoming well-known to the public. For example, Mercedes Benz's self-driving car Bertha, Tesla, Baidu's autonomous driving car Apollo, and so on. Path planning technology has undergone a long period of exploration and is becoming mature in the field of robotics. The path planning strategy in the field of autonomous driving largely follows the relevant algorithms of robotics technology. However, due to the slow speed of traditional robots and the difference in operational environments between robots and autonomous vehicles, conventional robotic path planning technologies cannot meet the requirements of autonomous vehicles under complex and rapidly changing road conditions, especially for timely avoidance of dynamic obstacles. Path planning technology based on reinforcement learning can continuously improve decision-making algorithms and enhance the artificial intelligence level and adaptability of path planning through the interaction between vehicles and their surrounding environment.

2. Characteristics of Autonomous Driving

Autonomous driving technology relies on the vehicle's sensing devices, judgment programs, and control components to achieve the function of independently completing driving tasks without the need for driver participation. Its characteristics lie in the level of intelligence, system integration, and extremely high degree of automation. This technology utilizes numerous sensors (such as LiDAR, cameras, millimeter wave radar, etc.) to monitor the surrounding environment, as shown in Figure 1, the sensors synchronously capture road conditions and real-time dynamic information, enabling a comprehensive understanding of the surroundings. The autonomous driving system possesses excellent computational and decision-making abilities. With the help of deep learning and reinforcement learning technology, it can make real-time and accurate judgments in complex environments and plan the best driving route. In addition, the system can automatically complete driving actions such as acceleration, braking, and steering through advanced control units. Autonomous driving technology also helps alleviate driver fatigue, optimize traffic flow, and promote energy conservation and emission reduction [1].

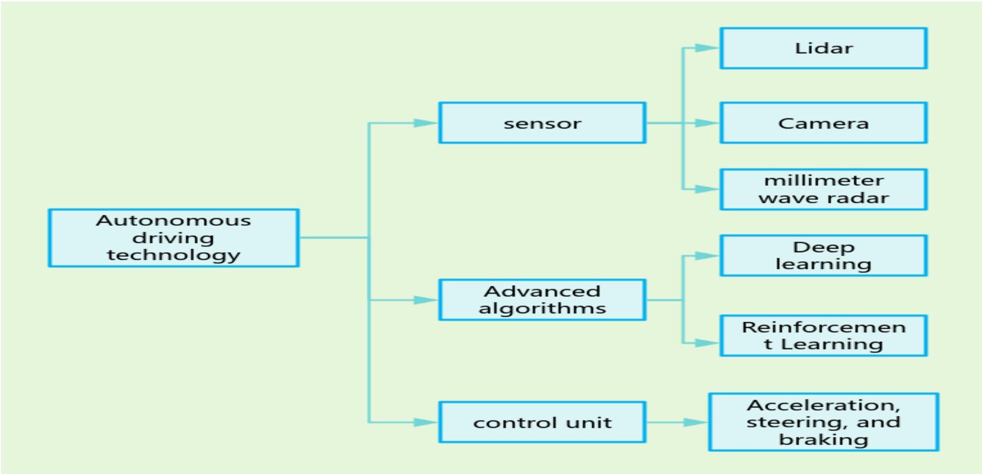

Figure 1. Characteristics of Autonomous Driving.

3. Requirements for Path Planning Based on Reinforcement Learning

3.1. Environmental Perception and Modeling Requirements

In the field of autonomous driving technology, environmental perception and modeling are the core, providing key data support and decision-making references for path planning. Autonomous vehicles rely on various sensors to obtain information about the surrounding environment in real time. These devices are responsible for collecting data on road conditions, obstacles, traffic signals, and pedestrians, and constructing a continuously changing and updated dynamic environmental model. The system needs to use these perceived data to accurately interpret the current road structure and predict traffic conditions and potential risks. Path planning under reinforcement learning requires the system to quickly optimize decision plans based on environmental perception data. Unlike traditional path planning algorithms that rely on fixed rules and models, reinforcement learning based algorithms can continuously improve strategies through interaction with the environment [2].

In terms of environment perception, the auto drive system needs to have the ability to build a detailed and comprehensive environment model, covering many details such as road layout, obstacle location, traffic light status and so on. Reinforcement learning can assist systems in optimizing their perception algorithms through repeated experimentation and feedback, gradually enhancing their understanding and processing skills for complex environments. Environmental information continues to change, such as traffic flow and pedestrian activity patterns showing strong dynamic characteristics. While perceiving these changes, the auto drive system updates the environment model in real time, and strengthens learning to assist the system in making flexible path adjustments to environmental changes through incentive mechanisms.

3.2. Collision Risk Avoidance Strategy Requirements

The ability to avoid collisions is a core aspect of autonomous driving route planning, directly determining the safety of vehicles. The autonomous driving system must promptly analyze potential risks in the surrounding environment and take appropriate measures based on the collected data. Considering the complexity of traffic conditions, avoiding collisions not only means bypassing fixed obstacles, but also involves dynamic prediction and response to moving objects such as other vehicles, pedestrians, or animals. The advantage of reinforcement learning in this field lies in its ability to continuously learn and optimize collision avoidance strategies through interaction with the environment. The path planning algorithm of reinforcement learning can guide the system to gradually learn how to adjust driving strategies to avoid collisions in different traffic scenarios through incentive mechanisms.

A well-trained system can implement avoidance strategies promptly upon detecting collision threats, such as changing the driving trajectory, slowing down the driving speed, or applying brakes. The advantage of reinforcement learning technology is its high ability to self-adjust and cope with changing and complex traffic conditions. As the training progresses, the system will be able to automatically update its collision avoidance strategies based on various obstacle characteristics, traffic lights, and the action patterns of other participants on the road.

3.3. Path Length Optimization Requirements

In the path design of autonomous driving, optimizing the length of the travel path is a key task. At the same time, it is also important to ensure driving safety and passenger comfort, while minimizing mileage and time consumption as much as possible. When planning the travel route from the starting point to the destination, the auto drive system should select a feasible path, and be committed to improving the driving efficiency through path optimization, reducing energy consumption and time costs [3]. In this process, the reinforcement learning algorithm learns the optimal path selection strategy through continuous interaction with the external environment, helping the auto drive system gradually determine the best path from the starting point to the end point based on the mechanism of trial and feedback under the changeable and complex road conditions. In specific application environments, the route of vehicle travel is constrained by many factors, including but not limited to traffic flow, road layout, and changes in traffic signals. Traditional route planning techniques are difficult to achieve real-time adjustments in the face of a series of complex and ever-changing factors. By utilizing reinforcement learning techniques and strategies, the system can gradually improve its action strategy through continuous interaction with the external environment, achieve optimal routes, and make flexible adjustments based on real-time traffic conditions. For example, when the system detects congestion on a certain road section, reinforcement learning algorithms can guide vehicles to choose an alternative, shorter path to avoid the time loss caused by congestion [4].

3.4. Requirements for Switching between Different Driving Scenarios

The auto drive system must be able to switch smoothly in various driving scenarios. All kinds of driving scenarios have unique requirements for the formulation of travel routes and driving strategies, which requires that the auto drive system can flexibly adjust the strategies according to the changes in the surrounding environment. The role of reinforcement learning technology in switching between multiple scenarios is extremely critical. Its powerful adaptability allows the system to identify the optimal action strategy in each scenario through continuous learning and feedback, achieving smooth transitions and optimizing driving efficiency. Table 1 presentation of the switching requirements for different scenarios, specifically demonstrating the requirements for route planning in various scenarios.

Table 1. Requirements for Different Driving Scenarios.

Scene type	Main challenges	Reinforcement learning strategy requirements	Key optimization objectives
Urban Rd	Busy traffic, pedestrians, traffic signals	Dynamic decision-making, real-time feedback, and multi strategy adjustment	Safety, traffic efficiency, emergency response
Expressway	Constant speed driving and overtaking	Path smoothness, vehicle speed optimization, traffic density analysis	High speed driving safety, energy-saving efficiency
Intersection	Traffic signal lights, collision risk	Traffic signal prediction and collision risk avoidance	Signal light control, traffic priority
Narrow alleyway	Narrow space and speed restrictions	Precise control, local path planning	Vehicle speed control, space utilization

Observing Table 1, it can be seen that autonomous driving technology faces diverse challenges in various driving environments [5]. The application of reinforcement learning technology in such environments relies on real-time sensing, building environmental models, and continuous improvement of strategies, effectively overcoming these challenges and ensuring smooth transition and optimal selection of the system between different driving modes. Through continuous strategy adjustment and optimization, reinforcement learning algorithms can adapt to various environments, flexibly switch driving modes, and achieve high efficiency and safety in autonomous driving route planning.

4. Reinforcement Learning Based Autonomous Driving Path Planning Technology

4.1. Autonomous Driving Path Planning Technology Based on Deep Q-Network (DQN)

Deep Q-Network (DQN), a combination of deep learning and reinforcement learning, has been widely applied in autonomous driving path planning. This technology relies on deep neural networks to approximate the Q-value function, helping agents make reasonable choices in complex environments. In autonomous driving path planning, the core task of DQN is to learn a strategy that enables vehicles to select the optimal path in a given road environment, minimize travel time and path length, and ensure safety. In the practical application scenario of DQN, the task of path planning is constructed as a Markov decision framework. Within this framework, autonomous vehicles continuously optimize their actions (i.e. path selection) by interacting with the environment to maximize cumulative rewards. The core strategy of DQN is to use neural networks to simulate Q-functions, enabling autonomous vehicles to predict the value of taking a certain action in a given state. The definition of the Q-value function is:

$$Q(s,a) = E[R_t | S_t = s, A_t = a] \tag{1}$$

In formula (1), $Q(s, a)$ represents the expected return obtained by taking action a in state s; R_t is the total sum of future rewards; S_t and A_t are the states and actions at time t. Through neural networks, DQN predicts Q values by learning an approximation function $Q(s,a;\theta)$, where θ is a parameter of the network. The training objective of the network is to minimize the error between the predicted Q value and the actual Q value, which is updated through the following objective function:

$$L(\theta) = E[(y - Q\hat{}(s,a;\theta))^2] \tag{2}$$

In formula (2), $y = r + \gamma max_a \hat{Q}(s',a';\theta^-)$ for the target Q value, γ is the discount factor, and θ^- is the parameter of the target network. Through such training, DQN can master the skills of selecting the best route and action plan under complex traffic conditions, and ensure that the auto drive system can make efficient and safe judgments during driving [6].

3.2. Autonomous Driving Path Planning Based on A3C Algorithm

Asynchronous Advantage Actor Critic (A3C) algorithm is a deep learning technique that integrates reinforcement learning, value functions, and policy gradient methods. It demonstrates efficient performance in handling path planning tasks in complex environments. This algorithm adopts a multi-threaded asynchronous update strategy, which enhances the robustness of the training process and accelerates the convergence process. In the field of autonomous driving, the A3C algorithm can effectively learn and develop the optimal driving path that adapts to various scenarios based on different traffic conditions. The A3C algorithm draws on two unique network models, one of which is the decision network, responsible for selecting direct actions. The second is the value network, which is responsible for estimating the value of taking a certain action in a given state. Thanks to this structure, A3C is able to simultaneously optimize the strategy and value function, achieving more stable and efficient learning performance. A3C aims to pursue the maximization of expected returns by optimizing the advantage function to reduce strategy errors. In the field of path planning, the A3C algorithm can determine which driving route to choose through learning, achieving cumulative maximization of future rewards. The following is the core update formula of A3C algorithm:

$$L = E[\log \pi_\theta(a_t|s_t)\,\delta_t] \tag{3}$$

In formula (3), $\pi_\theta(a_t|s_t)$ is the probability of the policy network selecting action a_t in state s_t. $\delta_t = r_t + \gamma V(s_t + 1) - V(s_t)$ it is an advantage function that represents the advantage of a certain action relative to its state value. $V(st)$ is the predicted value of the value network in state s_t. Gamma is a discount factor used to balance the importance of immediate rewards and future rewards. In the path design of autonomous driving, the A3C algorithm utilizes multi-threaded concurrent processing to continuously optimize the strategy and valuation function, and determine the optimal travel trajectory for various driving conditions. The system relies on feedback mechanisms to optimize route length, while also taking into account factors such as driving safety and riding experience, achieving flexible and real-time route adjustments.

4.3. Multi Agent Path Planning Based on Reinforcement Learning

Multi-Agent Path Planning (MAPP) in the field of autonomous driving focuses on how to handle conflicts in path selection and optimize overall routes in multi vehicle sharing scenarios. Reinforcement learning has become a powerful tool for solving such problems due to its adaptability to complex decision spaces [7]. Adopting distributed reinforcement learning strategies, such as multi-agent deep deterministic policy gradients, can promote collaboration and avoid conflicts between vehicles. In this model, each autonomous vehicle acts as an independent agent, adjusting its strategy based on global data and local perception and local perception, aims to complete the selection of the best path in a changing environment. In order to evaluate the performance of the strategy, simulation tests were conducted to compare the differences in optimization effects between reinforcement learning and traditional path planning algorithms.

The experimental data shown in Table 2 demonstrates that multi-agent path planning using reinforcement learning techniques outperforms traditional algorithms in terms of overall planning time, total path distance, and success rate. Especially in complex traffic conditions, the collision probability is significantly reduced, demonstrating the advantages of reinforcement learning in efficiency and stability. Thanks to the integration of reinforcement learning technology, auto drive system can achieve more efficient collaboration and path optimization among multi-agent, which plays a key role in promoting the progress of intelligent transportation system.

Table 2. Data Analysis Table.

Algorithm	Total planning time (seconds)	Total length of the path (meters)	Collision rate (%)	Success rate (%)
A *algorithm	15.2	1050	8.4	91.6
Dijkstra algorithm	20.4	1100	10.2	89.8
MADDPG algorithm (reinforcement learning)	12.8	980	2.3	97.8

4.4. Path Planning Based on PPO (Near End Policy Optimization) Algorithm

The Proximal Policy Optimization (PPO) algorithm is a state-of-the-art reinforcement learning method, which has been widely used in autonomous driving path planning due to its excellent efficiency and simple implementation process. The PPO algorithm prevents the problem of policy failure that may occur in traditional optimization methods by introducing a clipping mechanism to constrain policy updates, accelerates the convergence process, and ensures the reliability of finding the global optimal solution. In path planning tasks, the PPO algorithm guides vehicles to learn the optimal path in a dynamic environment by defining a reward function R_t. The reward function integrates factors such as driving efficiency, safety, and energy consumption, and is formulated as follows:

$$R_t = \alpha \cdot V_{eff} - \beta \cdot C_{col} - \gamma \cdot E_{cons} \qquad (4)$$

In formula (4), V_{eff} represents the efficiency of path planning (such as average speed or shortest path length). C_{col} represents collision risk, calculated based on vehicle spacing and speed difference. E_{cons} represents the energy consumption of a vehicle during operation, α, β, and γ are weight parameters used to balance the influence of different factors. The PPO algorithm was applied to the scenario of autonomous driving path planning through simulation experiments. Table 3 are the simulation experiment results.

Table 3. Simulation Experiment.

Indicator	Average path length (meters)	Average speed (m/s)	Collision rate (%)	Energy consumption (kilojoule)
DDPG algorithm	1120	9.3	5.8	92.4
SAC algorithm	1105	9.7	4.1	89.3
PPO algorithm (this study)	1085	10.2	2.7	85.6

From the experimental data in Table 3, it can be seen that the navigation strategy using PPO algorithm performs the best in multiple key indicators such as total route length, driving speed, and driving safety. The PPO algorithm can flexibly adjust the scale of policy updates according to actual situations, effectively mitigating policy fluctuations induced by environmental dynamics, especially in complex traffic conditions where its outstanding performance is more prominent. The above mathematical formulas and experimental results have confirmed the enormous application prospects of PPO algorithm in the field of automatic navigation, laying a solid foundation for transitioning the technology from theory to practical deployment.

5. Conclusion

Reinforcement learning has shown great potential and research significance in path planning in the field of autonomous driving. This technology enables the auto drive system to make decisions in real time under dynamic traffic conditions, optimize the path selection, reduce the possibility of traffic accidents, and enhance the overall efficiency of driving. Despite notable advancements, challenges remain in handling highly dynamic traffic scenarios and ensuring effective coordination among multiple autonomous agents. Future research should focus on improving the intelligence and adaptability of algorithms, ensuring that the system has stronger robustness in complex and changing contexts. With the progress of computing technology and the continuous improvement of algorithms, the path planning technology relying on reinforcement learning is expected to play a more critical role in the actual auto drive system, helping to improve the intelligent transportation system.

References

1. R. Zhang, J. Yang, Y. Liang, S. Lu, Y. Dong, B. Yang, and L. Zhang, "Navigation for autonomous vehicles via fast-stable and smooth reinforcement learning," *Sci. China Technol. Sci.*, vol. 67, no. 2, pp. 423–434, 2024, doi: 10.1007/s11431-023-2483-x.
2. J. J. Wu, D. F. Song, X. M. Zhang, C. S. Duan, and D. P. Yang, "Multi-objective reinforcement learning-based energy management for fuel cell vehicles considering lifecycle costs," *Int. J. Hydrogen Energy*, vol. 48, no. 95, pp. 37385–37401, 2023, doi: 10.1016/j.ijhydene.2023.06.145.
3. P. Yan, K. Yu, X. Chao, and Z. Chen, "An online reinforcement learning approach to charging and order-dispatching optimization for an e-hailing electric vehicle fleet," *Eur. J. Oper. Res.*, vol. 310, no. 3, pp. 1218–1233, 2023, doi: 10.1016/j.ejor.2023.03.039.
4. H. Guo, Y. Zhang, L. Chen, and A. A. Khan, "Research on vehicle detection based on improved YOLOv8 network," *arXiv preprint arXiv:2501.00300*, 2024, doi: 10.48550/arXiv.2501.00300.
5. C. Wang, R. Liu, A. Tang, Z. Zhang, and P. Liu, "A reinforcement learning-based energy management strategy for a battery–ultracapacitor electric vehicle considering temperature effects," *Int. J. Circ. Theory Appl.*, vol. 51, no. 10, pp. 4690–4710, 2023, doi: 10.1002/cta.3656.
6. V. K. T. Mantripragada and R. K. Kumar, "Deep reinforcement learning-based antilock braking algorithm," *Vehicle Syst. Dyn.*, vol. 61, no. 5, pp. 1410–1431, 2023, doi: 10.1080/00423114.2022.2084119.
7. S. Ergün, "A study on multi-agent reinforcement learning for autonomous distribution vehicles," *Iran J. Comput. Sci.*, vol. 6, no. 4, pp. 297–305, 2023, doi: 10.1007/s42044-023-00140-1.