Proceedings of the 3rd International Conference on Mechatronics and Smart Systems
DOI: 10.54254/2755-2721/2025.22931

Multi-objective Optimization in Autonomous Driving Based on Reinforcement Learning

Chenyang Li

Software College of Northeastern University, Shenyang, China
19213200658@163.com

Abstract: With autonomous driving technology playing an increasingly important role in intelligent transportation systems, how to improve ride comfort while ensuring safety has become an urgent challenge. This paper proposes a multi-objective optimization approach for autonomous driving based on reinforcement learning. By designing a multi-objective reward function that integrates rewards based on position, speed, direction, and acceleration, the method aims to balance driving efficiency and ride comfort. The Proximal Policy Optimization (PPO) algorithm is employed for training on the high-fidelity simulation platform MetaDrive, and experiments in multiple scenarios verify the effectiveness of the proposed approach. The experimental results show that as the comfort penalty coefficient in the reward function changes, the success rates for left turns, straight driving, and right turns exhibit a non-linear trend of first increasing then decreasing, with the best performance achieved when the parameter value is 0.001. This fully demonstrates the critical impact of parameter selection on the performance of autonomous driving strategies. It provides an optimization solution for reinforcement learning-based autonomous driving decision-making that balances safety and ride comfort, and offers a reference for subsequent related research.

Keywords: autonomous driving, multi-objective optimization, reinforcement learning, ride comfort, PPO

1. Introduction

1.1. Background and motivation

With the continuous increase in global vehicle ownership and the rapid development of artificial intelligence technology and the automotive industry, autonomous driving technology has gradually matured and become an important component of intelligent transportation. Autonomous driving systems integrate advanced AI algorithms, various high-precision sensors, built-in high-precision maps, and high-performance control systems to create a comprehensive perception network that overcomes the limitations of traditional human perception, enabling vehicles to autonomously plan routes and drive safely. As a result, not only has the traffic accident rate been significantly reduced, but traffic efficiency and resource utilization have also been markedly improved. In the context of ever-increasing demands for travel quality and intelligence, autonomous driving has undoubtedly become a landmark technology in the transportation field, paving the way toward intelligent and modernized mobility.

DOI: 10.54254/2755-2721/2025.22931

1.2. Advantages of reinforcement learning-based autonomous driving

Traditional supervised learning relies on vast amounts of labeled data, which is expensive to collect and may not cover complex scenarios. Reinforcement learning, on the other hand, uses simulation environments to mimic real driving situations, enabling agents to autonomously learn and optimize strategies while reducing dependence on labeled data. For example, SenseTime's R-UniAD system leverages reinforcement learning to lower the required data scale[1]. Additionally, when combined with deep neural networks, reinforcement learning can process multimodal information, optimize avoidance strategies, adapt to unexpected situations, and overcome the limitations of traditional rule-based systems.

1.3. Two main objectives of autonomous driving: comfort and safety

In practical applications, autonomous driving systems primarily pursue two main objectives:

Comfort: Vehicles should maintain smooth and stable driving, minimizing sudden acceleration, hard braking, and sharp turns to enhance the passenger experience.

Safety: Vehicles must effectively avoid collisions, strictly adhere to traffic rules, and have the ability to respond quickly to potential risks in emergencies to ensure the safety of all personnel.

There is a trade-off between them: Pursuing comfort by reducing speed can impact traffic efficiency; Prioritizing safety can lead to overly conservative driving, reducing road capacity. Thus, designing a reward function that balances both is crucial.

1.4. Existing research progress and research gap

In recent years, deep reinforcement learning algorithms—such as Deep Q-Networks and policy gradient methods—have achieved significant progress in the decision-making and control aspects of autonomous driving. They have demonstrated strong learning and adaptability in tasks such as lane keeping, adaptive cruise control, and obstacle avoidance. However, most studies have primarily focused on single-objective optimization, such as enhancing driving efficiency or solely pursuing safety. Research on multi-objective balancing, especially on how to simultaneously consider ride comfort and safety, remains insufficient. This research gap provides a clear entry point and motivation for this paper.

1.5. Main contributions of this paper

To address the aforementioned issues, the main contributions of this paper include:

Multi-Objective Reward Function Design:

This paper proposes a method for designing a multi-objective reward function that organically integrates rewards based on position, speed, direction, and acceleration.

Training Based on the PPO Algorithm:

The paper employs the Proximal Policy Optimization (PPO) algorithm for training, leveraging its stable updates and fast convergence. Through systematic parameter tuning, a high decision success rate is achieved.

Experimental Validation:

Extensive experiments are conducted in a high-fidelity simulation environment. Comparative analysis demonstrates that the proposed method significantly improves the stability, responsiveness, and overall performance of the autonomous driving system under complex road conditions, while simultaneously balancing ride comfort and driving safety.

DOI: 10.54254/2755-2721/2025.22931

2. Related work

2.1. Challenges in designing reinforcement learning reward functions

In autonomous driving tasks, the design of the reward function directly determines the agent's learning efficiency and final performance. However, in practice, there are multiple challenges.

2.1.1. Multi-objective balancing

Autonomous driving tasks often require balancing multiple objectives simultaneously, such as driving efficiency, ride comfort, and driving safety. Traditional reward functions tend to focus solely on a single objective—for example, optimizing driving speed or safety—while neglecting the importance of comfort [2].

In some experiments, a reward design that solely pursues high speed can lead to frequent sudden braking, thereby diminishing the passenger experience; whereas overemphasizing safety may cause the vehicle to act overly conservatively, affecting overall traffic efficiency.

To address this issue, recent studies have attempted to employ weighted summation, decomposition strategies, or Pareto optimization methods to balance different objectives, but determining the appropriate weights and strategies remains an open problem [3].

2.1.2. Reward signal sparsity

In real-world driving scenarios, the agent often only receives rewards when completing certain key actions—such as successfully navigating an intersection or maintaining a safe following distance—which leads to sparse reward signals. When the vehicle travels in a long, uninterrupted straight line without significant changes, the agent finds it difficult to obtain timely feedback from the environment. This delay can hinder the convergence of the policy and increase the difficulty of learning.

To alleviate this issue, researchers have proposed methods such as reward shaping, intrinsic motivation, and hierarchical reinforcement learning, aiming to provide more frequent feedback signals and thus accelerate the learning process [4].

2.2. Progress in reinforcement learning-based autonomous driving research

In recent years, reinforcement learning has made considerable progress in the field of autonomous driving, although some limitations still remain.

2.2.1. End-to-end vs. modular design

Early autonomous driving research predominantly employed end-to-end deep learning methods, directly mapping raw sensor data to control commands. However, due to issues with interpretability and generalization, recent studies have gradually shifted towards a modular design, where perception, decision-making, and control are separated and optimized individually [5].

Some systems based on end-to-end design perform poorly when encountering entirely new scenarios, whereas modular design allows for the independent optimization of each submodule, thereby enhancing the overall system's robustness and safety [6].

DOI: 10.54254/2755-2721/2025.22931

2.2.2. Applications of reinforcement learning algorithms

Various reinforcement learning algorithms—such as DQN, policy gradient, Actor-Critic, PPO, and SAC—have been widely applied to tasks like lane keeping, adaptive cruise control, and obstacle avoidance[7].

For example, studies using DQN have achieved certain success in lane keeping tasks; however, at complex intersections, agents often exhibit sudden braking behavior due to simplistic reward design. In contrast, another study using policy gradient methods for adaptive cruise control achieved smooth speed adjustments through carefully designed reward functions, although there is still room for improvement in safety metrics [8].

These findings indicate that while existing algorithms perform well on individual tasks, they still exhibit clear limitations in multi-objective optimization.

2.2.3. Experimental cases and comprehensive evaluation

Some studies have attempted to combine reinforcement learning with other learning paradigms—such as imitation learning and supervised learning—in order to fully leverage expert data and accelerate the convergence process[8].

For example, a hybrid approach utilized imitation learning to provide an initial policy for reinforcement learning, followed by reward shaping for further optimization. This method achieved relatively strong performance in lane changing and obstacle avoidance tasks. However, it still lacks a systematic solution for balancing comfort and safety[9].

In addition, some experimental results have shown that, in high-fidelity simulation environments, integrating multi-objective reward design can significantly improve task success rates and system stability. Nevertheless, adaptability and robustness in real-world environments still require further validation.

3. Methodology

3.1. Reward function design

To balance both driving comfort and safety, we design a comprehensive reward function consisting of the following components:

Position Reward Function:

This component first determines whether the vehicle is on the reference lane to establish the correct driving direction. Rewards or penalties are then assigned based on the vehicle's lateral deviation from the lane center. If the vehicle deviates from the center, the reward decreases proportionally; if the vehicle is in a wrong-way driving state, a significant penalty is applied[10].

Forward Movement Reward Function:

A positive reward is given based on the vehicle’s longitudinal progress along the lane, encouraging steady forward motion.

Speed Reward Function:

Rewards are calculated based on the ratio between the vehicle’s current speed and the maximum allowable speed. This encourages the vehicle to maintain a reasonable speed, thus balancing safety with traffic efficiency.

Acceleration Reward Function:

A redesigned acceleration reward function is used to constrain abrupt changes in acceleration. When acceleration exceeds the safe range, a penalty is applied. This promotes smooth acceleration and deceleration, thereby enhancing ride comfort.

Integrated Weighting Strategy:

DOI: 10.54254/2755-2721/2025.22931

All reward components are combined using a weighted fusion strategy. The weights are fine-tuned through extensive simulation experiments to ensure that the reward function reflects driving efficiency, comfort, and safety, guiding the agent to learn an optimal driving policy.

This multi-objective reward design enables the reinforcement learning algorithm to focus not only on driving speed but also on safety and passenger experience, thereby making it better suited for complex traffic environments.

3.2. Training method

Algorithm Selection: The PPO algorithm is adopted due to its advantages of stable updates and fast convergence. The experiment is implemented in Python and debugged using PyCharm.

Simulation Environment Construction: The MetaDrive platform is used to provide various scenarios. SubprocVecEnv is employed to achieve multi-process parallel simulation, accelerating data collection and policy updates.

Parameter Tuning: Based on the Stable Baselines3 library, experimental configurations with log recording, hyperparameter tuning, and environment parallelization are used to ensure high training efficiency and scientifically sound parameter tuning.

Packages Used: metadrive, stable_baselines3, numpy, torch, random, matplotlib, IPython.display, os, json, datetime, etc.

Through the effective integration of the above methods and packages, this work ensures high efficiency in simulation training and scientific rigor in parameter tuning, providing a solid experimental foundation for multi-objective optimization in autonomous driving strategies.

4. Results

4.1. Hyperparameter settings

In our experiments, the following key hyperparameters were used:

Table 1: Hyperparametre

Parameter Name	Description	Value
random_seed	Random seed	0
wrong_way_penalty	Wrong direction penalty	50.0
lateral_reward_weight	Weight for lateral reward	1
acceleration_reward_coefficient	Acceleration penalty coeff.	0.0005
a_x_max	Max x-acceleration	3
a_y_max	Max y-acceleration	1
map	Map layout used in the environment	"X"

These hyperparameters were carefully selected and tuned to ensure the effectiveness and reproducibility of the training process in a diverse and dynamic driving simulation environment.

4.2. Comfort function

To measure passenger comfort in autonomous driving, this paper builds a reward function based on ISO 2631 and ISO 22133, focusing on acceleration and its rate of change. This reflects passengers' physiological perception and provides an optimization objective for reinforcement learning algorithms[10].

DOI: 10.54254/2755-2721/2025.22931

4.2.1. Acceleration modeling

Passengers' perception of acceleration during vehicle motion exhibits nonlinear characteristics: low-magnitude acceleration has minimal impact on comfort, while discomfort increases sharply once the acceleration exceeds a certain threshold. Based on this observation, this study adopts an exponential decay model to quantify the impact of acceleration on comfort:

$$S1 = ie^{-\alpha(|a\_x|+|a\_y|)} \tag{1}$$

Variable Definitions:

a_x and a_y represent the vehicle's acceleration in the x and y directions, respectively.

α is the decay coefficient, which characterizes the passenger's sensitivity to acceleration. Based on preliminary experiments and literature comparison, a recommended range for α is between 0.3 and 0.5.

A higher α value causes the comfort score to drop rapidly under higher accelerations, thus imposing stricter penalties.

4.2.2. Jerk (rate of change of acceleration) modeling

Passengers are generally more sensitive to the rate of change of acceleration (jerk) than to acceleration itself, as frequent or abrupt changes in acceleration tend to cause stronger discomfort. Therefore, this study also adopts an exponential model to describe the impact of jerk on comfort:

$$S2 = e^{-\beta(|j\_x|+|j\_y|)} \tag{2}$$

Variable Definitions:

j_x and j_y represent the vehicle's jerk in the x and y directions, respectively. Their calculation formulas are:

$$j\_x = a\_x\_current - a\_x\_\Delta t \tag{3}$$

$$j\_y = a\_y\_current - a\_y\_\Delta \tag{4}$$

β is the coefficient that controls the influence of jerk. Based on analysis of passenger experience and experimental data, the recommended range for β is 0.6 to 1.0.

A larger β value causes the reward function to give stronger negative feedback for sudden changes in acceleration, encouraging the policy to avoid abrupt variations.

4.2.3. Construction of the overall comfort function

To simultaneously reflect the different impacts of acceleration and jerk (rate of change of acceleration) on passenger comfort, the final comfort score is computed using a weighted sum:

$$S = w\_1 \times S1 + w\_2 \times S2 \tag{5}$$

Weight Settings:

The weights w_1 and w_2 satisfy the condition w_1 + w_2 = 1.

Comprehensive analysis and experimental results indicate that jerk has a more significant impact on passenger comfort. Therefore, w_1 is set to 0.4 (acceleration component), and w_2 is set to 0.6 (jerk component).

This configuration balances the need for smooth driving while ensuring that the optimization process is not overly influenced by a single factor, thereby enhancing the robustness of the overall driving policy.

DOI: 10.54254/2755-2721/2025.22931

4.2.4. Hyper-parameter tuning and experimental validation

In the actual training process, sensitivity analysis of different parameter combinations is conducted through a large number of pre-experiments to validate the model's ability to portray the comfort experience and ensure that the reward function can effectively guide the strategy search. Specific parameters were set as follows: α was set to 0.4. With this setting, the vehicle will quickly obtain a lower comfort score when facing a larger acceleration, thus prompting the reinforcement learning algorithm to prioritize the smooth acceleration strategy and reduce passenger discomfort.

β was set to 0.8, which well balanced the penalty for acceleration change rates. This allowed the system to maintain safety while dealing with sudden road changes or emergency avoidance, and also ensured passenger comfort as much as possible.

Through comparison experiments with several alternatives, it is finally determined that the above parameters can function stably in different road scenarios, effectively improving the smoothness of the vehicle driving process and passenger comfort without affecting the realization of the safety objectives.

4.3. Experimental results

By adjusting the penalty coefficient for acceleration in the reward function, the impact on the autonomous driving strategy's performance was verified. The experiment focused on the success rates of left turns, straight driving, and right turns, and evaluated driving smoothness and comfort through the average acceleration during turns. The table below presents the experimental data for different parameter configurations:

Table 2: Experimental results with different parameters

Parameter Value	Left Turn Success Rate	Straight Success Rate	Right Turn Success Rate	Left Turn a average	Straight a average	Right Turn a average
0	0.19	0.58	0.81	5.36	5.16	5.86
0.001	0.77	0.83	0.90	4.83	4.81	5.64
0.002	0.14	0.40	0.66	6.06	5.63	6.43
0.0005	0.18	0.55	0.77	5.63	4.93	5.64
0.0015	0.72	0.60	0.83	4.88	4.93	5.97

4.3.1. Convergence analysis

We trained the model until convergence, and the convergence curve is shown in the figure 1.

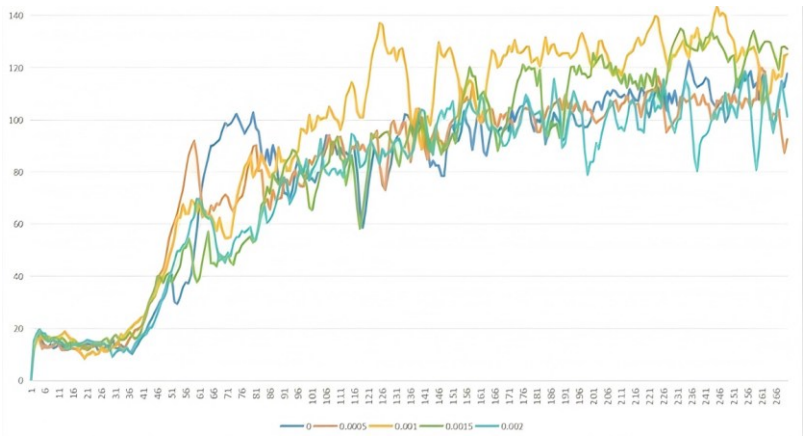


Figure 1: Convergence curve

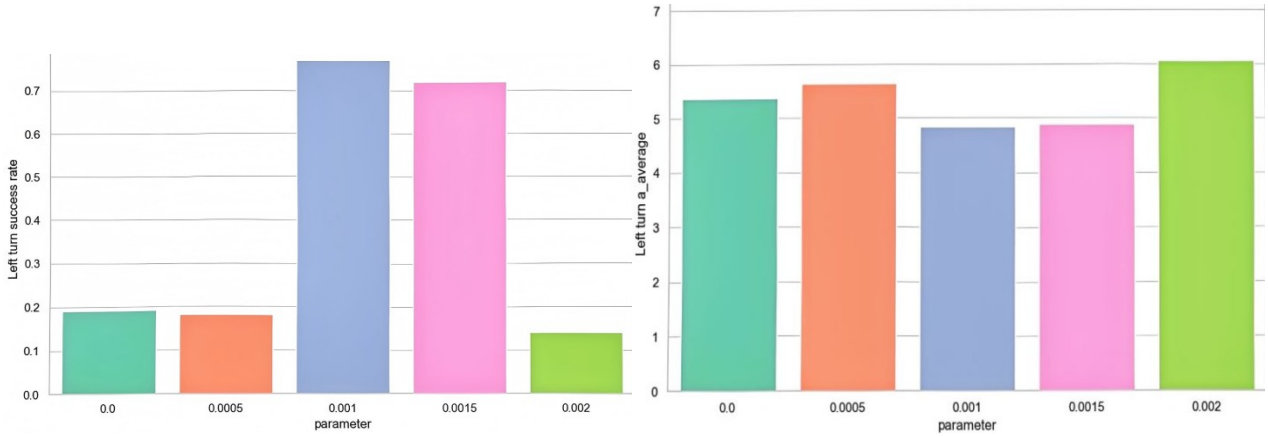
DOI: 10.54254/2755-2721/2025.22931

4.3.2. Success rate analysis

Left turn task:

When the parameter is set to 0.001, the left turn success rate reaches 0.77, which is a significant improvement compared to the cases with parameters of 0 (only 0.19) and 0.002 (only 0.14). Although the left-turn power was also higher (0.72) with a parameter of 0.0015, the overall combined performance was slightly worse than 0.001.

The experiment result is shown in figure 2.



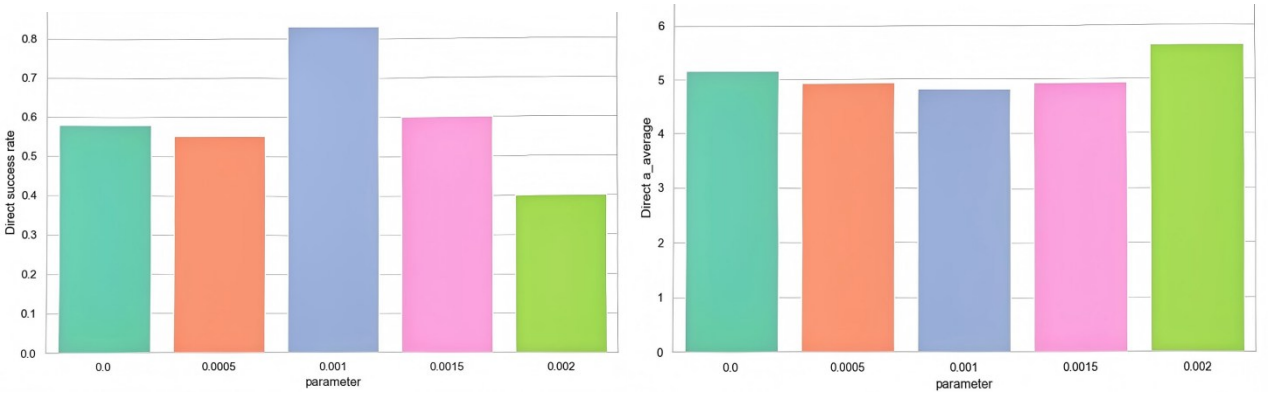
(a) (b)

Figure 2: The success rate and a_average in left turn

Straight ahead vs. right turn task:

In the straight ahead task, parameter 0.001 yields the highest success rate (0.83), while in the right turn task, it also performs the best with a success rate of 0.90. This suggests that an appropriate acceleration penalty (e.g., 0.001) in the parameter tuning can better balance the decision-making requirements in different driving scenarios, and improve the robustness and accuracy of the strategy.

The experiment result is shown in figure 3.



(a) (b)

Figure 3: The success rate and a_average in straight

4.3.3. Analysis of average acceleration

Ride comfort:

As an important indicator of passenger comfort, lower values of average acceleration usually indicate a smoother vehicle ride. The experimental results show that when the parameter is 0.001, the

DOI: 10.54254/2755-2721/2025.22931

a_average values obtained in the left-turn, straight and right-turn scenarios are 4.83, 4.81 and 5.64, respectively, which are lower than the other parameter settings (e.g., 0.002 corresponds to 6.06, 5.63, and 6.43), which indicates that the acceleration of the vehicle at this time is smoother, which is more conducive to improving the ride comfort. comfort.

4.4. Synthesized discussion

The synthesized experimental data shows that the setting of the comfort penalty coefficient in the reward function has a significant effect on the autopilot strategy.

Too low a penalty coefficient (parameter 0 or 0.0005): the intelligence lacks sufficient smoothness constraints, resulting in a low success rate of turning and high acceleration, and is prone to rapid acceleration or deceleration.

Excessive penalty coefficient (0.002): excessive penalties can make the strategy too conservative, which reduces the turn success rate, while the vehicle may be underpowered in some cases.

Moderate value (0.001): performs best in balancing driving efficiency, cornering success and ride comfort. At this time, not only the success rate of all types of turning tasks is significantly improved, but also the acceleration of the vehicle is kept at a low level, fully reflecting the requirements of smooth driving.

Overall, by comparing the experimental results under different parameters, this paper verifies the effectiveness of the multi-objective reward design. The strategy trained based on the PPO algorithm can better balance the safety and comfort objectives when the parameter is 0.001, thus realizing high success rate and excellent ride experience in complex driving scenarios.

5. Conclusion

5.1. Summary

In this paper, a multi-objective optimization method based on reinforcement learning is proposed for autonomous driving. By designing a multi-objective reward function that integrates position, speed, direction and acceleration rewards, and using the Proximal Policy Optimization (PPO) algorithm to train on a high-fidelity simulation platform, a better balance between safety and ride comfort is achieved. The experimental results show that under the influence of different acceleration penalty coefficients, the success rate of each turning task shows a tendency of increasing and then decreasing, in which the overall performance reaches a better level when the penalty coefficient takes the value of 0.001.

5.2. Problems addressed

This paper mainly solves the problem of how to simultaneously balance driving safety and ride comfort in automatic driving decision-making. By comprehensively designing the multi-objective reward function, the behavior of the vehicle in the case of rapid acceleration and deceleration is effectively constrained, so as to improve the success rate of turning while reducing the change of vehicle acceleration and improving the ride experience.

5.3. Future prospects

Reward function improvement: Explore a more adaptive reward mechanism, dynamically adjust the reward weights according to different driving scenarios, and further improve the flexibility and stability of the strategy.

DOI: 10.54254/2755-2721/2025.22931

Algorithm structure optimization: Consider introducing hierarchical reinforcement learning or hybrid learning methods to improve the performance of traditional algorithms in complex decision-making and increase training efficiency.

Training strategy adjustment: optimize the training strategy and parameter adjustment methods for different simulation environments and task requirements, in order to enhance the generalization ability and robustness of the model.

References

[1] *Wang, X. (2025). A Study on Autonomous Driving Control Strategy Considering Passenger Comfort - China Road Traffic Safety Network. China Road Traffic Safety Network. Retrieved from https://www.163.com/dy/article/JP181JVE05118O92.html[^39^].*

[2] *Abouelazm, A., Michel, J., & Zöllner, J. M. (2024). A Review of Reward Functions for Reinforcement Learning in the Context of Autonomous Driving. arXiv preprint arXiv:2405.01440.*

[3] *Wang, H., & Chan, C. Y. (2021). Multi-objective optimization for autonomous driving strategy based on soft actor–critic algorithm. Journal of Autonomous Intelligence, 3(1), 1-10.*

[4] *Li, S. E. (2023). Reinforcement Learning for Sequential Decision and Optimal Control. Springer.*

[5] *Chen, D., & Huang, X. (2024). End-to-end Autonomous Driving: Challenges and Frontiers. IEEE Transactions on Pattern Analysis and Machine Intelligence.*

[6] *Pranav, S. C., & Singh, P. (2023). Recent Advancements in End-to-End Autonomous Driving using Deep Learning: A Survey. arXiv preprint arXiv:2307.04370. arXiv*

[7] *Liu, H., Huang, Z., Wu, J., & Lv, C. (2021). Improved Deep Reinforcement Learning with Expert Demonstrations for Urban Autonomous Driving. arXiv preprint arXiv:2102.09243.*

[8] *Lu, Y., Fu, J., Tucker, G., et al. (2022). Imitation Is Not Enough: Robustifying Imitation with Reinforcement Learning for Challenging Driving Scenarios. arXiv preprint arXiv:2212.11419.*

[9] *Booher, J., Rohanimanesh, K., Xu, J., et al. (2024). CIMRL: Combining Imitation and Reinforcement Learning for Safe Autonomous Driving. arXiv preprint arXiv:2406.08878.*

[10] *Yixu He, Yang Liu, Lan Yang, Xiaobo Qu. Exploring the design of reward functions in deep reinforcement learning-based vehicle velocity control algorithms. Transportation Letters, 2024, 16(10): 1338-1352. ISSN 1942-7867. https://doi.org/10.1080/19427867.2024.2305018.*