Proceedings of the 4th International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/2026.34236

Application and Analysis of Reinforcement Learning in Autonomous Racing

Jiarui Tang

Faculty of Applied Sciences, Macao Polytechnic University, Macao, China
p2211342@mpu.edu.mo

Abstract. In recent years, reinforcement learning has gradually become an important technical path for autonomous racing research because it can learn the long-term reward optimal policy through environmental interaction. At the same time, imitation learning, world model, offline training and safety constraints have also been introduced to make up for the shortcomings of pure reinforcement learning in sample efficiency and deployment safety. In this paper, the application of reinforcement learning in autonomous racing is reviewed. Firstly, the racing task is formalized as a Markov decision process. Secondly, the main paths of model-free reinforcement learning, model-driven reinforcement learning, imitation learning and reinforcement learning fusion, offline reinforcement learning and secure reinforcement learning were compared. Finally, combined with the application characteristics and key problems of the existing research, the main bottlenecks and future development trends of the current research are summarized. In general, the evolution of methods in the field of autonomous racing is not simply replacing traditional control with reinforcement learning, but gradually shifting to a fusion system of reinforcement learning + model knowledge + expert experience + safety constraints.

Keywords: Reinforcement learning, Autonomous racing, Autonomous driving, Safety control, Offline reinforcement learning

1. Introduction

With the development of artificial intelligence, robotic control, and intelligent transportation technologies, autonomous driving research is expanding from regular road scenarios to mission environments with higher dynamics and closer to vehicle limits. Autonomous racing is an important research direction formed in this context. Different from ordinary autonomous driving, which mainly emphasizes comfort, rule-following and safety redundancy, autonomous racing pays more attention to achieving optimal track tracking, fast decision-making and stable control under extreme conditions, thus becoming an important testing ground for testing the robustness, real-time performance and extreme handling ability of autonomous driving systems [1].

From the perspective of technological development, early autonomous racing systems mainly rely on classical control and model-based planning methods, such as Pure Pursuit, MPC, and race line optimization. These methods have the advantages of strong interpretability and mature engineering implementation. However, in high-speed, strongly nonlinear and complex interaction

© 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: 10.54254/2755-2721/2026.34236

scenarios, vehicle dynamics errors, environmental noise, and the uncertainty of opponent behavior prediction will significantly increase the modeling difficulty. In contrast, reinforcement learning can directly learn policies from state, action and reward relationships, which have natural advantages for complex nonlinear control and long-term reward optimization problems, and thus has been gradually introduced into the research of autonomous racing [2].

In recent years, with the development of visual perception, end-to-end control, shrinking racing car platforms, and high-fidelity simulation environments, researchers have begun to apply reinforcement learning to track tracking, corner control, overtaking games, and real vehicle deployment. Visual imitation of intensive study, for example, can be used to directly learn from the input image study [3] the car control strategy; And platforms such as F1TENTH provide the foundation for unified testing and validation of reinforcement learning in autonomous racing cars [4]. These studies show that reinforcement learning can not only be used as a general-purpose algorithmic tool, but also is gradually becoming a key component in the decision-making and control framework of autonomous racing.

However, the application of reinforcement learning in autonomous racing still faces obvious challenges. Firstly, online exploration on real racing platforms is costly and risky, and pure online training is often infeasible. Secondly, high-dimensional observation and continuous action control will significantly increase the sample demand and training instability. Thirdly, the strategy learned in the simulation environment is often affected by the dynamics deviation, sensor noise and environmental disturbance on the real vehicle, and it is difficult to transfer stably. Finally, multi-vehicle competition and high-speed overtaking further increase the complexity of safety control and game decision-making [5]. Therefore, the focus of current research has gradually shifted from "how to make the vehicle run faster" to "how to run faster stably under the premise of safety".

Based on the above background, the motivation of this paper is that autonomous racing has become one of the most representative high-dynamic application scenarios in reinforcement learning research. However, there is still a lack of unified review on the specific application methods of different reinforcement learning methods in autonomous racing cars. This article analyzes the typical usage of reinforcement learning in autonomous racing cars from an application perspective, rather than remaining at the level of algorithmic principles. The article covers three aspects. It explains the specific application scenarios of different reinforcement learning methods in autonomous racing cars. It also analyzes how model knowledge, expert experience, and offline data can help improve the sample efficiency and training stability in racing tasks. It summarizes how to balance speed performance and safety constraints in actual deployment. Through the above analysis, this article hopes to provide a clearer framework for the current research status and future development direction of enabling reinforcement learning for autonomous racing.

2. Task modeling and technical framework

From the perspective of reinforcement learning, autonomous racing can be represented as a Markov decision process:

$$M = (S, A, P, R, \gamma) \tag{1}$$

The state space S contains vehicle position, speed, heading, lateral deviation, track geometry and opponent information. Action space A generally includes steering Angle, throttle and brake; P denotes state transition; R is the reward function; And γ is the discount factor. The policy optimization objective can be written as follows.

DOI: 10.54254/2755-2721/2026.34236

$$J(\pi) = E_{\{\pi\}}[\sum_{\{t=0\}}^{\{T\}} \gamma^{t} r_{t}] \tag{2}$$

In autonomous racing, the reward function usually reflects not only speed gains but also track maintenance and safety constraints, which can be summarized as follows.

$$r_t = w_1 v_{\{progress\}} - w_2 e_{\{track\}} - w_3 e_{\{yaw\}} - w_4 C_{\{collision\}} \tag{3}$$

This indicates that the racing task is not a single pursuit of higher speed, but to achieve a balance between lap time, stability and safety. If safety constraints are further introduced, the problem can be formulated as follows.

$$maxJ(\pi), s.t. E[\sum \gamma^t c_t] \leq d \tag{4}$$

Where c_t represents the cost caused by boundary crossing, collision or dynamic constraint violation, and d is the safety budget. This formalism illustrates that autonomous racing is essentially a continuous control optimization problem with strong constraints.

In terms of system implementation, reinforcement learning autonomous racing usually consists of a closed loop of sensory input, state representation, policy network, value assessment/world model and vehicle execution module.

3. Classification and comparison of methods

3.1. Model-free reinforcement learning

Model-free reinforcement learning, which does not explicitly learn the environment dynamics but directly optimizes the policy or value function through interaction data, is currently one of the most common paths in autonomous racing research. In autonomous racing scenarios, such methods are often used for tasks such as single-vehicle track tracking, time racing, and continuous steering control, because these tasks can directly map the vehicle state to the control action, and form a clear reward feedback through indicators such as lap time, track retention, and stability [2].

In the case of PPO, the objective function can be written as follows.

$$L_{\{PPO\}}(\theta) = E_{t[min(r_{t(\theta)\hat{A}_t}, clip(r_{t(\theta)}, 1-\varepsilon, 1+\varepsilon)\hat{A}_t)]} \tag{5}$$

In autonomous racing cars, PPO is often used to learn the control strategy of the car in different corners and different speed conditions. Researchers usually take the vehicle's current speed, lateral deviation, heading Angle error and the local geometry information of the track as the state input, and take the steering Angle and throttle as the action output. Through a large number of simulation training, the car gradually learns when to slow down in advance, when to stay close to the ideal line, and when to quickly resume acceleration after the corner. In other words, the actual role of PPO in autonomous racing is not to optimize the strategy in the abstract, but to learn a set of track tracking control laws that can reduce the whole lap time [5].

SAC is also a common model-free reinforcement learning method in autonomous racing, which is characterized by the addition of an entropy regularization term to the objective function to enhance the exploration ability:

DOI: 10.54254/2755-2721/2026.34236

$$J(\pi) = \sum E[r(s,a) + \alpha H(\pi(\cdot|s))] \tag{6}$$

In autonomous racing, SAC is more suitable for dealing with continuous action control and robust decision making problems under complex track conditions. For example, when the car needs to maintain high control accuracy across different track curvatures and speed intervals, SAC can learn smoother acceleration, deceleration and steering behaviors through more adequate exploration, thereby improving control stability. In summary, the application of model-agnostic reinforcement learning in autonomous racing mainly lies in enabling the car to directly learn "when to turn, when to decelerate, and when to accelerate" through repeated interactions, thereby forming a more flexible race control strategy than fixed rules [2, 5].

The limitations of this approach in autonomous racing are also obvious. It heavily relies on a large amount of interaction data, and the real racing platform cannot perform high-risk trial-and-error operations as it does in a simulation environment. Its performance is highly dependent on the design of the reward function. If the reward design is unreasonable, the car may learn conservative but slow behaviors, or aggressive but unstable ones. The trained strategy may also be affected by changes in the environmental distribution and dynamic deviations when applied to the real world. Although model-agnostic reinforcement learning is suitable for exploring high-performance control strategies in autonomous racing, it is usually more suitable as the core method for simulation training and algorithm verification rather than undertaking all practical deployment tasks independently.

3.2. Model-driven reinforcement learning

By clearly understanding the dynamic characteristics of the learning environment or directly using existing vehicle models to assist in strategy training, model-driven reinforcement learning in autonomous racing focuses not merely on predicting future states but on using dynamic knowledge to help reinforcement learning reduce the cost of trial and error, improve training efficiency, and increase the feasibility of actions. This is particularly important in autonomous racing as the car task itself heavily relies on vehicle dynamic characteristics, and the cost of trial and error is very high at high speeds. If people completely rely on random exploration, the training cost will be extremely high [3].

In practical applications, researchers usually build an approximate vehicle dynamic model to predict the position of the vehicle in the next few moments under different steering and speed control conditions, and then use the predicted results to assist in the decision-making of reinforcement learning. For example, when a car is speeding towards a sharp turn, the model can predict that if it maintains high throttle, the car may understeer and deviate from the track. Therefore, the strategy will take measures to decelerate and adjust the route in advance. In this way, reinforcement learning training in autonomous racing no longer relies entirely on blind trial and error but can learn more efficient and safer control behaviors through prediction [3, 6].

Another typical role of model-driven methods in autonomous racing is to combine with the world model or potential space prediction to help the system update strategies with fewer actual interactions. Such applications are particularly suitable for reducing the scale of the racing platform or actual vehicles, as the number of actual experiments is limited, and the dynamics model can provide additional state evolution information for reinforcement learning, thereby reducing reliance on actual track data. However, these methods have limitations: if the learned model deviates from the actual dynamics, errors may accumulate during high-speed control, affecting the performance of the strategy. Therefore, in autonomous racing, model-based reinforcement learning is more suitable

DOI: 10.54254/2755-2721/2026.34236

as an auxiliary module to improve training efficiency and action feasibility rather than completely replacing actual interactive learning.

3.3. Imitation learning merges with reinforcement learning

In autonomous racing, the core role of the integration of imitation learning and reinforcement learning is to solve the problem of "randomly initialized strategies cannot drive and are not suitable for repeated trials". Compared with ordinary tasks, autonomous racing has the characteristics of high speed, high risk, and high-dimensional input. Therefore, if the strategy is explored from completely random actions, it will often hit walls, go out of bounds, or become unstable frequently in the early stage of training, and it is difficult to accumulate effective experience. Based on this practical problem, researchers usually use expert demonstrations or traditional controllers to generate demonstration data, so that the car can first learn the basic driving ability, and then further improve the performance through reinforcement learning [3].

This application is particularly common in vision-based autonomous racing tasks. Researchers can first collect demonstration data from an expert driver or an existing controller on the track, including the front view of the vehicle, the speed state, and the corresponding steering and throttle operations. Then, imitation learning is used to train an initial strategy, so that the vehicle has the basic ability to drive along the track stably. On this basis, reinforcement learning is introduced to let the car continue to optimize the turn entry point, turn acceleration time and line selection, so as to gradually exceed the original expert performance [3, 7]. This shows that in autonomous racing, the integration of imitation learning and reinforcement learning is not simply to splice the two methods together, but to first solve the problem of "can you drive", and then solve the problem of "can you drive faster".

The advantage of this approach is that it fits the engineering needs of autonomous racing very well. Because hazard exploration on real platforms is expensive, and expert data is often more readily available, leveraging expert experience to build a drivable initial strategy can significantly reduce training risk, shorten training time, and improve sample utilization efficiency. However, its shortcomings are also clear: if the quality of expert data is not high, the upper limit of the policy will be limited; Without reasonable constraints in subsequent reinforcement learning stages, the car may still deviate from the safe driving zone. Therefore, in autonomous racing, such methods often need to be combined with robustness training and safety mechanisms to achieve a balance between high performance and high reliability [7, 8].

3.4. Offline reinforcement learning

Offline reinforcement learning is only trained on a pre-collected dataset and does not interact with the environment in real time during training. For autonomous racing cars, this path is of great practical interest, because real racing platforms are expensive to train, with hardware wear, and risk, while test logs, race records, and expert demonstration data can be continuously accumulated. Therefore, the application of offline reinforcement learning in autonomous racing cars focuses on transforming the accumulated driving data into new training resources instead of online trial and error [9].

In practical applications, researchers usually compile offline datasets of vehicle states, driving actions, track results, and environmental feedback recorded in past tests or races, and then directly use these data to train reinforcement learning policies. For example, the system could learn from historical records that certain corners were driven too conservatively or with insufficient exit speed,

DOI: 10.54254/2755-2721/2026.34236

and then learn to use more aggressive throttle and steering control at those locations. This allows the car to continuously optimize the performance of the strategy based on the available data, without having to do a lot of dangerous testing again [9].

The advantages of offline reinforcement learning in autonomous racing are clear: it reduces the cost of real interaction and avoids risky exploration, making it particularly suitable as an important intermediate stage after simulation training but before real deployment. This method still faces a common problem of "unstable estimation of out-of-distribution behaviors". If the learned behavior by the strategy deviates from the region supported by the data, the value function may be overestimated incorrectly, which ultimately leads to a decline in deployment performance. Therefore, in autonomous racing cars, offline reinforcement learning usually needs to be combined with conservative estimation, behavior constraints, or distribution restrictions to ensure that the trained strategy does not deviate from the range supported by the existing data. The practical value of offline reinforcement learning in autonomous racing cars lies in enabling the system to iteratively use historical data safely, rather than relying entirely on new track experiments.

3.5. Safe reinforcement learning

In autonomous racing cars, safety is not an additional indicator but a prerequisite for training and deployment. The actual application goal of safe reinforcement learning is not merely to make the racing car "move more slowly", but to make it compete competitively while avoiding dangerous behaviors, achieving "fast but without excessive risk" strategy control [5]. This is obviously different from general reinforcement learning tasks, because in the racing scene, a boundary crossing, a collision or a high-risk instability can directly lead to training interruption or even hardware damage.

In autonomous racing, safety reinforcement learning is usually applied in three specific ways. The first is to add penalty terms such as boundary crossing, collision, too close distance and dynamic instability to the reward function, so that the policy automatically avoids high-risk behaviors during training. The second is to formulate the problem as an optimization task with constraints, which limits the cumulative risk cost while pursuing the speed performance. The third is to add a safety filtering module or feasible region supervisor in the execution layer. When the action output by the policy may cause a collision or run off the track, the system will correct the action in real time to ensure that the vehicle is still within the safety boundary [5, 10].

For example, in the multi-vehicle autonomous racing scenario, the reinforcement learning policy should not only consider how to complete overtaking as soon as possible, but also avoid collision with the opponent. Therefore, researchers can add safety constraints to the training: when the vehicle is too close to the opponent, the overtaking Angle is too extreme, or the predicted trajectory may intersect, the system will increase the penalty, or the action will be directly corrected by the safety filtering module. In this way, the car will not only pursue a temporary overtaking opportunity, but will learn to complete the overtaking at a more appropriate position and time [10]. This shows that the application of safety reinforcement learning in autonomous racing focuses on internalizing the safety boundary as part of policy learning, rather than adding an additional safety module after training.

In general, safe reinforcement learning is important not only because it reduces crashes, but also because it improves training sustainability and deployment confidence. For real autonomous racing systems, only policies learned within safety constraints are more likely to actually land. Therefore, more and more research tends to adopt a hierarchical structure of "high-level reinforcement learning

DOI: 10.54254/2755-2721/2026.34236

+ low-level safety control" to establish a stable balance between speed performance and safety constraints [4, 10, 11].

4. Challenges and development trends

Currently, the application of reinforcement learning in autonomous racing still faces four core challenges. First, sample efficiency is still the bottleneck. Autonomous racing is expensive to train, real data is scarce, and model-free reinforcement learning highly relies on a large number of interactions, so it is still necessary to rely on expert demonstrations, offline data, and world models to improve training efficiency in the future [6, 9]. Second, safety must be further endogenized. In high-speed environments, any unconstrained exploration can lead to severe consequences, so safety constraints should not just be an additional penalty in the reward function, but should be part of the training objective and system structure [5, 10]. Third, the ability of generalization and sim-to-real transfer is insufficient. Many strategies perform well on a single track, but their performance significantly decreases after changing tracks, weather, or vehicles, indicating that the robustness still needs to be strengthened, which is also the focus of some end-to-end reinforcement learning research [8]. Fourthly, the unified evaluation system still needs to be improved. At present, different studies vary greatly in input forms, reward functions, track difficulty and evaluation indicators, resulting in difficulties in horizontal comparison. However, the development of open platforms such as F1TENTH is providing an important foundation for unified benchmarks [4].

From the overall trend, the future research on autonomous racing will not be dominated by a single algorithm, but is more likely to move towards the integration route of "reinforcement learning + model knowledge + expert experience + safety constraints". On the one hand, visual reinforcement learning, lidar-driven strategies and on-board online learning will continue to promote the development of autonomous racing systems towards higher autonomy [7, 11, 12]. On the other hand, overtaking game, safety control and real deployment research will also make this field pay more attention to system-level executability and verifiability [10, 11]. Only on this basis, reinforcement learning can truly move from laboratory performance demonstration to replicable, deployable and extensible technical system.

5. Conclusion

In general, reinforcement learning has become an important method in the research of autonomous racing, but its value does not lie in simply replacing traditional control, but in providing new policy optimization capabilities for high-speed, strongly nonlinear, and strongly interactive scenarios. Through the comparison of model-free reinforcement learning, model-driven reinforcement learning, imitation learning fusion, offline reinforcement learning and safe reinforcement learning, it can be found that each method has its own emphasis. The former pays more attention to the performance limit, while the latter pays more attention to sample efficiency, training safety and engineering implementation.

Therefore, the future development direction of the autonomous racing field is not a single algorithm "winning", but a unified task framework that combines reinforcement learning with expert knowledge, prediction models, safety constraints, and system engineering depth. Only in this way, the research of reinforcement learning in autonomous racing can truly move from proof of concept to stable deployment, and provide methodological support for a wider range of autonomous driving limit control problems.

DOI: 10.54254/2755-2721/2026.34236

References

[1] Betz, J., Zheng, H., Liniger, A., et al. (2022). Autonomous vehicles on the edge: A survey on autonomous vehicle racing. IEEE Open Journal of Intelligent Transportation Systems, 3, 458–488.
[2] Kiran, B. R., Sobh, I., Talpaert, V., et al. (2021). Deep reinforcement learning for autonomous driving: A survey. IEEE Transactions on Intelligent Transportation Systems, 23(6), 4909–4926.
[3] Cai, P., Wang, H., Huang, H., et al. (2021). Vision-based autonomous car racing using deep imitative reinforcement learning. IEEE Robotics and Automation Letters, 6(4), 7262–7269.
[4] Evans, B. D., Trumpp, R., Caccamo, M., et al. (2024). Unifying F1TENTH autonomous racing: Survey, methods and benchmarks. arXiv. https: //doi.org/10.48550/arXiv.2402.18558
[5] Evans, B. D., Engelbrecht, H. A., & Jordaan, H. W. (2023). High-speed autonomous racing using trajectory-aided deep reinforcement learning. IEEE Robotics and Automation Letters, 8(9), 5353–5359.
[6] Leng, B., Zhang, W., Li, Z., et al. (2026). Expert knowledge-driven reinforcement learning for autonomous racing via trajectory guidance and dynamics constraints. arXiv. https: //doi.org/10.48550/arXiv.2603.05842
[7] Vasco, M., Seno, T., Kawamoto, K., et al. (2024). A super-human vision-based reinforcement learning agent for autonomous racing in gran turismo. arXiv. https: //doi.org/10.48550/arXiv.2406.12563
[8] Murdoch, A., Schoeman, J. C., & Jordaan, H. W. (2023). Partial end-to-end reinforcement learning for robustness against modelling error in autonomous racing. arXiv. https: //doi.org/10.48550/arXiv.2312.06406
[9] Koirala, P., & Fleming, C. (2024). F1TENTH autonomous racing with offline reinforcement learning methods. In 2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC) (pp. 3922–3929). IEEE.
[10] Steiner, E., van der Spuy, D., Zhou, F., et al. (2025). Accelerating real-world overtaking in F1TENTH racing employing reinforcement learning methods. In 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 4528–4534). IEEE.
[11] Hildisch, B., Ghignone, E., Baumann, N., et al. (2025). Drive fast, learn faster: On-board RL for high performance autonomous racing. arXiv. https: //doi.org/10.48550/arXiv.2505.07321
[12] Mammadov, M. (2023). End-to-end lidar-driven reinforcement learning for autonomous racing. arXiv. https: //doi.org/10.48550/arXiv.2309.00296