Proceedings of the 4th International Conference on Materials Chemistry and Environmental Engineering
DOI: 10.54254/2755-2721/85/20240948

A survey of reinforcement learning theories in autonomous vehicles

Shuwen Bai

School of Information Science and Engineering, East China University of Science and Technology, Shanghai, 200237, China

baishuwen@mail.ecust.edu.cn

Abstract. This survey explores the core theories, techniques, and core applications of reinforcement learning (RL) in the domain of autonomous vehicles. RL has emerged as a significant contributor to decision-making techniques in this field, enabling agents to make informed decisions based on actions and rewards derived from the environment. The ultimate objective of RL in autonomous driving is to maximize cumulative rewards over time. However, RL faces challenges related to interpretability and sample efficiency, particularly in complex driving scenarios. This survey extensively investigates the utilization of RL in decision-making and control, encompassing various scenarios and addressing the challenges encountered within RL-based autonomous vehicles. By emphasizing the design of effective reward functions, enhancing sample efficiency, and improving model interpretability, future advancements in reinforcement learning for autonomous vehicles can foster the development of more robust, efficient, and trustworthy autonomous systems. Moreover, this survey provides valuable insights into the limitations of RL techniques in autonomous driving decision-making, highlighting areas that require further research and development.

Keywords: reinforcement learning, autonomous vehicles, decision-making planning.

1. Introduction

Reinforcement learning (RL) has become a key method for autonomous decision-making and control with the quick development of autonomous driving technology. RL is rooted in the theoretical foundations of Markov Decision Processes (MDP), dynamic programming, and approximate value functions.

Traditionally, RL methods have primarily focused on offline training and decision-making in simulated environments. These methods optimize driving decisions by constructing value functions or policy functions. However, the rise of deep learning has propelled the prominence of Deep Reinforcement Learning (DRL), capable of handling complex and high-dimensional driving decision problems. DRL combines RL algorithms with deep neural networks, enabling end-to-end learning and decision-making directly from raw sensor data [1].

Deep RL algorithms have advanced quickly in the past few years, which has prompted the development of several DRL-based autonomous driving strategies. These include value-based methods like Deep Q-Network (DQN) and Double Deep Q-Network (DDQN), along with policy gradient techniques like Deterministic Policy Gradient (DPG) and Deep Deterministic Policy Gradient (DDPG).

DOI: 10.54254/2755-2721/85/20240948

Additionally, optimal control theory-based methods like Model Predictive Control (MPC) and Reinforcement Learning Model Predictive Control (RLMPC) have also been explored [2].

Reinforcement learning (RL) has become a significant contributor to the development of decision-making techniques for autonomous vehicles. It is a type of machine learning that enables agents to decide what to do based on actions and rewards received from the environment. By maximizing the total rewards gathered throughout its existence, the agent uses its understanding of the expected value connected to various state-action combinations to increase its eventual rewards.

Several studies using driving simulators like Carla and Torcs, as well as other simulation environments, have shown the superiority and flexibility of the RL algorithm in autonomous driving decision-making [3],[4],[5],[6]. RL facilitates data-driven decision-making by processing unstructured data and learning from high-dimensional perceptual information, enabling end-to-end solutions[7].

However, RL still confronts some challenges, such as low interpretability and sample efficiency [8], particularly in complex and uncertain interactive driving scenarios. Achieving fully autonomous decision-making remains a highly challenging task.

This survey aims to comprehensively explore the theories and applications of reinforcement learning in autonomous vehicles. Section II introduces the utilization of reinforcement learning, focusing on its application in decision-making and control. Section III presents various scenarios where RL is employed. Section IV discusses the challenges confronted by autonomous vehicles based on reinforcement learning.

2. Core Principles

The utilization of Reinforcement Learning (RL) in autonomous driving tasks encompasses various aspects, including environment perception, decision-making planning, and control execution. The subsequent sections will delve into the latter two aspects in more detail. Before exploring those topics, it is vital to examine the establishment of state space, action space, and reward systems within the context of autonomous vehicle challenges.

2.1. State Spaces, Action Spaces and Rewards

The proper design of state spaces, action spaces, and rewards is of utmost importance when applying reinforcement learning to autonomous driving tasks. In Figure 1, the agent checks the current state and makes a decision based on that observation by choosing an action. The environment then provides occasional rewards or punishments based on the agent's chosen action. This feedback from the environment helps the agent learn and improve its decision-making abilities over time. In their work, Kiran et al. provide a comprehensive description of these three key points commonly used in this context [9]. The state space features frequently employed in autonomous vehicles encompass fundamental attributes such as the vehicle's position, heading, and speed. These features also include detecting surrounding obstacles within the vehicle's sensor range. Moreover, it is customary to utilize a coordinate system centered on the autonomous vehicle, augmenting it with supplementary details such as lane information, path curvature, the vehicle's past and anticipated trajectories, longitudinal information, and other relevant factors. A widely adopted approach for presenting this comprehensive information is through a bird's-eye view perspective. However, the design of reward functions for reinforcement learning agents in autonomous vehicles remains an open question, with ongoing research efforts.

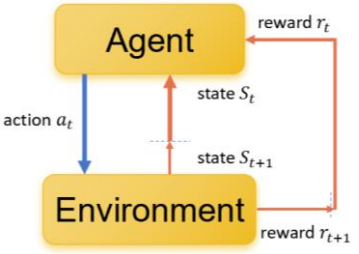

Figure 1. The basic reinforcement learning scenario.

DOI: 10.54254/2755-2721/85/20240948

2.2. Decision-making Planning Using RL

In the realm of autonomous vehicle behavioral decision-making, reinforcement learning has exhibited remarkable efficacy, particularly in highway scenarios and urban intersections. When considering the cutting-edge decision-making technologies for single-vehicle autonomous driving, Inverse Reinforcement Learning (IRL) and Hierarchical Reinforcement Learning (HRL) are widely adopted and currently regarded as relatively mature methodologies.

Imitation learning, which serves as the foundation for Inverse Reinforcement Learning (IRL), operates on the premise that decisions made by experts represent optimal or near-optimal courses of action. In this approach, the cumulative reward generated by the expert's behavior is assumed to be the highest. IRL leverages techniques such as maximum marginalization or probabilistic modeling to infer the reward function from observed expert behavior or current tactics. By doing so, IRL addresses challenges associated with significant reward errors, sparse rewards, and convergence issues. In the context of decision-making studies for autonomous vehicles, IRL frequently employs expert driver behavioral data to learn and infer the underlying rewards. Subsequently, in conjunction with the specific driving scenario, the reinforcement learning (RL) algorithm is executed in a forward manner, optimizing the inferred reward function-based driving behavior approach. This integration of IRL and RL enables autonomous vehicles to learn from expert demonstrations and refine their driving strategies, ultimately enhancing their overall performance in real-world driving scenarios. You et al. explored the driver's driving style as a basis for constructing a reward function that incorporates state-action pairs [10]. They employed a framework that combines Q-learning with the maximum entropy principle to determine the optimal driving strategy in a multi-lane environment. On a similar note, Liu et al. utilized Principal Component Analysis (PCA) to extract driving styles from expert prior knowledge [11]. Subsequently, they applied IRL) based on the maximum entropy principle to accomplish the lane-changing task, taking into account the specific driving style identified through PCA analysis.

Table 1. Different RL Formulation On The Motion Planning.

	Algorithm	State	Action	Reward
Chen et al. [3]	DDPG	front-view image	{discrete}	{+: road alignment, speed, distance to lane boundaries, overtake}
Du et al. [12]	DDPG	Vehicle relative states	{longitudinal acceleration}	{+: driving efficiency, comfort, energy efficiency}
Liu et al. [5]	SAC	bird-eye view image	{brake, throttle, discrete}	{+: efficiency, ride comfort, safety}
Chen et al. [6]	DDQN	bird-view images	{direction}	{+: speed, smoothness, -: collision, lane}
Nageshrao et al. [13]	DDQN	the ego car	{lane change}	{+: speed, lane position, safety}

Based on the driver's actual driving process as a hierarchical approach consisting of discrete and continuous levels, the Hierarchical Reinforcement Learning (HRL) algorithm establishes a framework that combines MDP and POMDP. In this framework, the upper-level decision-making is discretized, while the lower-level execution is continuous. Different local policies within the hierarchical system serve as independent sub-functions. For example, the overall strategy for highway driving can be divided into separate sub-tasks such as lane changing (left/right), lane keeping, and car-following. This approach simplifies the state space, effectively addresses the issue of dimensionality challenges that may arise in classical RL algorithms, and enhances overall decision-making performance. Chen et al. proposed an HRL framework based on the DDPG algorithm, which incorporates temporal and spatial attention mechanisms [3]. This integration enhances the structural capability of neural networks and improves lane-changing efficiency. For the lane-changing job, the system was built and assessed in the TORCS

DOI: 10.54254/2755-2721/85/20240948

simulator. Duan et al. focused on highway driving scenarios and applied the HRL approach to decompose the driving task into three options [4]. They employed an asynchronous parallel training method to learn both sub-policies and a main policy for each action. This approach allows for the effective learning of different strategies for each sub-action while maintaining a coherent overall driving strategy.

To provide further insights, Table 1 presents a compilation of the key points corresponding to several algorithms.

2.3. Control Approaches Using RL

The control approach founded on DRL avoids the requirement for explicitly developing a precise mathematical model, unlike alternative intelligent control systems. Instead, the control strategy is acquired through learning during system interactions, undergoing continual optimization and iteration. Du et al. conducted a noteworthy study where they utilized the DDPG algorithm to address the complex multi-objective optimization problem of controlling vehicle acceleration on uneven road surfaces [12]. The framework proposed by Bautista-Montesano et al. demonstrates notable efficacy in addressing the particular control problem at hand. The authors devised a system that integrates longitudinal control using RL and reactive Model Predictive Control (MPC) for longitudinal cruise control. This combined system effectively safeguards the vehicle from entering potentially hazardous situations [14].

Table 2 shows the strengths of each method which is mentioned above.

Table 2. The Strengths Of The Methods Mentioned In This Survey.

	Method	**Strengths**
You et al. [10]	Novel MaxEnt deep IRL algorithms specifically designed to tackle MDP problems without a predefined model	Consider the road shape that can let systems compare and contrast various cornering driving techniques
Liu et al. [11]	An approach for identifying driving styles that uses MFPCA; autonomous lane-change systems using an IRL methodology	Adaptable to various driving styles for multiple drivers
Chen et al. [3]	A H-DRL algorithm with an attention mechanism	learn multiple driving policies in a single model; improve performance with fewer examples
Duan et al. [4]	Self-driving automobiles make decisions using the H-RL approach without labeled driving data	Addresses the requirements for both low-level motion control and high-level maneuver selection in both lateral and longitudinal orientations
Du et al. [12]	A speed control model based on DRL	Improve the energy effectiveness and comfort of AVs on unpaved surfaces; in comparison with the MPC-based speed control model, gains are made in terms of computing efficiency, effectiveness of energy, and soothe while riding vertically
Bautista-Montesano et al. [14]	RL and MPC modules are made into a control system	Enhance the safety to RL based navigation system

DOI: 10.54254/2755-2721/85/20240948

3. Applications

This section focuses on two application scenarios that are based on RL including urban autonomous driving and highway autonomous driving.

3.1. RL for Urban Autonomous Driving

Owing to the complicated driving conditions, which include varying traffic conditions, interactions between many agents, and the need to balance efficiency, comfort, and safety, the successful application of self-driving vehicles in urban environments remains a significant problem [5].

The majority of methods first use the frontal image as the data source and train the strategy in an end-to-end manner. However, in urban driving contexts, the rich and high-dimensional visual elements present major learning challenges, leading to extensive sample complexity [6].

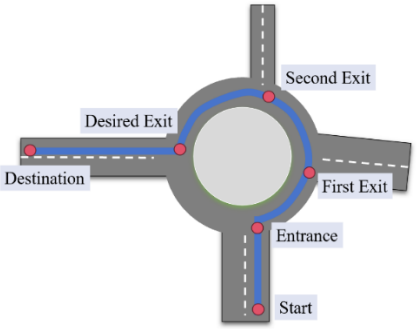

Figure 2. View of the roundabout task situation from above.

To mitigate this challenge, Chen et al. propose an alternative approach by devising a distinct source representation to lessen sample complexity [6]. To record low-dimensional hidden states, they use visual representation techniques, which makes the problem more tractable for reinforcement learning. Notably, their method exhibits promising performance in a challenging roundabout scenario, as illustrated in Figure 2. To address the issue of fuzzy rule or model definitions encountered in rule- or model-based methods, Liu et al. introduce an innovative framework that combines reinforcement learning with expert demonstration [5]. The goal of this framework is to obtain an action-control approach that is especially suited for city environments. This suggested technique also outperforms existing baselines with sample efficiency and overall efficacy in a situation including an urban roundabout with heavy traffic.

3.2. RL for Highway Autonomous Driving

Highway driving scenarios are distinct from urban scenarios due to factors such as higher speeds, smoother traffic flow, and optimized infrastructure for long-distance travel. The increased speeds on highways pose risks, including reduced reaction times and longer stopping distances. Nageshrao et al. address these risks by augmenting the Double Deep Q-Network (DDQN) learning agent with a safety check that considers short-term horizons [13]. This safety check modifies the initial action choice of the DDQN agent in critical scenarios if a safer alternative exists, enhancing learning efficiency by eliminating the need for repeated resets during training.

In the context of autonomous driving, addressing overtaking problems in highway driving scenarios is crucial for safety, traffic flow optimization, and effective decision-making. Liao et al. propose a driving strategy for autonomous vehicles on highways that is based on DRL [15]. Their decision-making strategy, adaptable to complex scenarios, utilizes a hierarchical control structure to manage ego and surrounding vehicle motions. The DDQN algorithm is employed in the decision-making process used on highways, and simulation experiments confirm the efficiency and safety of the overtaking policy based on DDQN for completing highway-related tasks.

DOI: 10.54254/2755-2721/85/20240948

4. Challenges

This section presents three challenges that may be areas for future endeavors. These challenges involve the incentive system's architecture, the DRL models' interpretability, and the low sample efficiency.

4.1. The Design of Reward Functions

The capacity of reinforcement learning to function without having explicit knowledge of the environment is one of its primary benefits. However, derivating whether an action is positive or negative in a given context can be challenging. Ad-hoc reward functions are commonly utilized, incorporating manually chosen coefficients to strike a balance between safety, efficiency, comfort, adherence to traffic regulations, and other factors. These designs often rely on brute-force methods, lacking both theoretical foundations and comprehensive research. It is important to note that different criteria for designing rewards can result in significantly distinct driving behaviors [9].

4.2. The Interpretability in DRL Models

Due to the intricate architecture of deep neural networks (DNNs), conventional deep reinforcement learning (DRL) approaches suffer from limited interpretability. The complexity of DNNs makes it infeasible to track and comprehend each parameter, rendering the implicit features and their management within the network uncertain. Consequently, DRL models have become black boxes, leaving specialists unable to discern how agents learn from their environment or why they make specific decisions. This lack of transparency poses a constraint on the implementation of DRL, as trust in an agent's actions diminishes without an explanation, especially when those actions contradict expectations [16].

4.3. Low Sample Efficiency

Sample efficiency poses a significant challenge in reinforcement learning, as the learning process requires a large number of samples to acquire an appropriate policy. This challenge becomes particularly apparent in scenarios where acquiring valuable expertise is costly or hazardous. When it comes to autonomous driving, postponed and sparse rewards, unbalanced distribution of information across a broad state space, and other factors further reduce sampling efficiency [9].

5. Conclusion

In conclusion, this survey has provided a comprehensive analysis of the fundamental techniques and applications of reinforcement learning in the field of autonomous vehicles. The discussion has encompassed recent research advancements in decision-making and control methods that leverage reinforcement learning in autonomous vehicle systems. Additionally, two specific application scenarios were presented to demonstrate the practical implementation of reinforcement learning in autonomous driving.

Moving forward, future efforts should prioritize addressing key challenges in reinforcement learning for autonomous vehicles. Firstly, there is a need to design effective and appropriate reward functions, considering the intricate trade-offs between various objectives such as safety, efficiency, comfort, and regulatory compliance. Moreover, improving sample efficiency is crucial, and innovative approaches should be explored to reduce the number of samples required for learning, enabling faster and more efficient policy acquisition. One possible approach is to modify the incentive structure by incorporating more frequent reward functions, which encourage the machine to learn quickly from fewer examples. Additionally, agents can enhance efficiency by initially learning an offline policy through imitation learning, leveraging expert-provided roll-outs. Subsequently, reinforcement learning can be employed to further improve performance through interaction with the environment. Furthermore, the challenge of enhancing interpretability in reinforcement learning models for autonomous vehicles requires further research. It is essential to develop methods that shed light on the decision-making process and provide transparent explanations. Explainable Artificial Intelligence (XAI) methodologies have emerged as a means to elucidate predictions made by complex models, particularly deep neural networks in image data analysis. While current efforts in Explainable Reinforcement Learning (XRL) are problem-specific

DOI: 10.54254/2755-2721/85/20240948

and unable to address practical challenges in reinforcement learning, ongoing research endeavors aim to explain predictions in models like Graph Neural Networks (GNNs) and clarify the reasoning behind Deep Reinforcement Learning (DRL) operations. These areas represent active fields of investigation [17].

By addressing these challenges, future advancements in reinforcement learning for autonomous vehicles will significantly contribute to the development of more robust, efficient, and trustworthy autonomous systems. The integration of well-designed reward functions, improved sample efficiency, and enhanced model interpretability will pave the way for safer and more reliable autonomous driving technologies in the coming years.

References

[1] F. Ye, S. Zhang, P. Wang, and C.-Y. Chan, "A survey of deep reinforcement learning algorithms for motion planning and control of Autonomous Vehicles," 2021 IEEE Intelligent Vehicles Symposium (IV), 2021. doi:10.1109/iv48863.2021.9575880

[2] S. Aradi, "Survey of deep reinforcement learning for motion planning of Autonomous Vehicles," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 2, pp. 740–759, 2022. doi:10.1109/tits.2020.3024655

[3] Y. Chen et al., "Attention-based hierarchical deep reinforcement learning for Lane change behaviors in autonomous driving," 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2019. doi:10.1109/iros40897.2019.8968565

[4] J. Duan, S. Eben Li, Y. Guan, Q. Sun, and B. Cheng, "Hierarchical reinforcement learning for self‐driving decision‐making without reliance on labelled driving data," IET Intelligent Transport Systems, vol. 14, no. 5, pp. 297–305, 2020. doi:10.1049/iet-its.2019.0317

[5] H. Liu, Z. Huang, J. Wu, and C. Lv, "Improved Deep Reinforcement Learning with expert demonstrations for urban autonomous driving," 2022 IEEE Intelligent Vehicles Symposium (IV), 2022. doi:10.1109/iv51971.2022.9827073

[6] J. Chen, B. Yuan, and M. Tomizuka, "Model-Free Deep Reinforcement Learning for Urban Autonomous Driving," 2019 IEEE Intelligent Transportation Systems Conference (ITSC), 2019. doi:10.1109/itsc.2019.8917306

[7] M. Hussonnois and J.-Y. Jun, "End-to-end autonomous driving using the ape-X algorithm in Carla simulation environment," 2022 Thirteenth International Conference on Ubiquitous and Future Networks (ICUFN), 2022. doi:10.1109/icufn55119.2022.9829674

[8] Z. Zhu and H. Zhao, "A survey of Deep RL and IL for autonomous driving policy learning," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 9, pp. 14043–14065, 2022. doi:10.1109/tits.2021.3134702

[9] B. R. Kiran et al., "Deep Reinforcement Learning for Autonomous Driving: A Survey," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 6, pp. 4909–4926, 2022. doi:10.1109/tits.2021.3054625

[10] C. You, J. Lu, D. Filev, and P. Tsiotras, "Advanced planning for autonomous vehicles using reinforcement learning and deep inverse reinforcement learning," Robotics and Autonomous Systems, vol. 114, pp. 1–18, 2019. doi:10.1016/j.robot.2019.01.003

[11] J. Liu, L. N. Boyle, and A. G. Banerjee, "An inverse reinforcement learning approach for customizing Automated Lane Change Systems," IEEE Transactions on Vehicular Technology, vol. 71, no. 9, pp. 9261–9271, 2022. doi:10.1109/tvt.2022.3179332

[12] Y. Du et al., "Comfortable and energy-efficient speed control of autonomous vehicles on rough pavements using deep reinforcement learning," Transportation Research Part C: Emerging Technologies, vol. 134, p. 103489, 2022. doi:10.1016/j.trc.2021.103489

[13] S. Nageshrao, H. E. Tseng, and D. Filev, "Autonomous Highway driving using Deep Reinforcement Learning," 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), 2019. doi:10.1109/smc.2019.8914621

DOI: 10.54254/2755-2721/85/20240948

[14] R. Bautista-Montesano, R. Galluzzi, K. Ruan, Y. Fu, and X. Di, "Autonomous Navigation at unsignalized intersections: A coupled reinforcement learning and model predictive control approach," Transportation Research Part C: Emerging Technologies, vol. 139, p. 103662, 2022. doi:10.1016/j.trc.2022.103662

[15] J. Liao et al., "Decision-making strategy on highway for autonomous vehicles using Deep Reinforcement Learning," IEEE Access, vol. 8, pp. 177804–177814, 2020. doi:10.1109/access.2020.3022755

[16] Y. Qing, S. Liu, J. Song, H. Wang, and M. Song, "A survey on Explainable Reinforcement Learning: Concepts, algorithms, challenges," arXiv.org, https://arxiv.org/abs/2211.06665 (accessed Oct. 8, 2023).

[17] S. Munikoti, D. Agarwal, L. Das, M. Halappanavar, and B. Natarajan, "Challenges and opportunities in deep reinforcement learning with Graph Neural Networks: A comprehensive review of algorithms and applications," IEEE Transactions on Neural Networks and Learning Systems, pp. 1–21, 2023. doi:10.1109/tnnls.2023.3283523