

Reinforcement Learning for Autonomous Driving: Efficiency and Comfort Optimization

Anrui Wang

March 9, 2025

Abstract

Keywords: Reinforcement Learning, Autonomous Driving, Multi-Objective Optimization, Efficiency, Comfort

1 Introduction

1.1 Background and Current Research Status

Autonomous driving has become a major application domain for deep reinforcement learning (RL) in recent years. Modern RL algorithms can learn complex driving policies in high-dimensional environments such as realistic traffic simulators¹. A variety of driving tasks have been tackled with RL, including lane keeping, car-following (adaptive cruise control), intersection handling, and overtaking. For example, Hossain et al. (2023) used Deep Q-Networks (DQN) in the CARLA simulator to train an agent to drive at high speed while avoiding collisions¹. RL agents have even begun to match or exceed human driving performance in certain controlled tasks².

Early and foundational work demonstrated the feasibility of RL for continuous vehicle control. The Deep Deterministic Policy Gradient (DDPG) algorithm, for instance, was applied to autonomous driving scenarios, showing that an RL agent could learn to steer and accelerate a car in a racing simulator with human-like skill³. Since then, numerous studies have employed deep RL methods for single-objective goals in driving. Common objectives include safety (collision avoidance) and efficiency (minimizing travel time or fuel). For example, RL controllers have been trained to minimize fuel consumption – a PPO-based agent achieved a 17% improvement in fuel economy

over a human baseline while maintaining comparable travel time⁴. Likewise, RL policies have been optimized for safety by heavily penalizing collisions or infractions, yielding agents that can drive for long durations without accidents. These single-objective successes validate RL as a powerful tool for autonomous vehicle decision-making⁵.

However, driving is inherently a multi-faceted problem involving trade-offs among objectives like time efficiency, energy efficiency, safety, and passenger comfort. Traditional single-objective approaches (e.g. focusing solely on speed or fuel efficiency) often ignore secondary criteria like ride comfort. Recently, researchers have started incorporating multiple objectives into the RL framework, recognizing that a truly human-like autonomous driver must be not only efficient but also safe and comfortable. The next sections review the state-of-the-art in multi-objective RL for autonomous driving, with an emphasis on jointly optimizing efficiency (time/fuel) and comfort (ride smoothness).

1.2 Research Gaps and Challenges

1.3 Our Contributions

2 Related Work

2.1 Deep RL Algorithms

A range of deep RL algorithms have been applied to autonomous driving, each with its own strengths. As comprehensively documented by Kiran et al.⁶ in their survey, value-based methods like DQN and Double DQN work with discrete action spaces (e.g., discrete throttle/brake levels or high-level decisions) and have been used for lane-following and simple driving tasks¹. For instance, a DQN agent can learn to keep a car in its lane and avoid obstacles by choosing from a few steering angles. Pioneering work by Sallab et al.⁷ demonstrated the feasibility of using end-to-end deep RL to perform lane-keeping assist from raw pixels, showing that neural networks could learn steering control policies without manual rule-based controllers. However, discrete actions make it hard to capture the fine control needed for comfort (smooth braking, etc.), so value-based RL is often combined with action discretization.

Policy gradient and actor-critic methods are more popular for continuous control. DDPG (Deep Deterministic Policy Gradient) is an off-policy actor-critic algorithm that outputs continuous steering or acceleration commands. It has seen wide use in driving scenarios that require smooth control, such as

car-following. Zhu et al.³ employed DDPG for adaptive cruise control, noting it directly optimizes continuous acceleration and can incorporate multiple reward terms for safety, efficiency, and comfort. Their evaluation using real traffic data showed the RL agent maintained safer time headways and smaller acceleration changes, indicating improved comfort and safety simultaneously compared to conventional controllers.

Proximal Policy Optimization (PPO), an on-policy method, has also been applied to driving for its stability in training. In an eco-driving study, a PPO agent with an LSTM network was used to optimize a speed profile, successfully reducing fuel consumption while respecting travel time⁴. Soft Actor-Critic (SAC), which maximizes a trade-off of reward and entropy, has been tested in driving simulations as well; its sample-efficient learning and robust convergence make it suitable for high-dimensional tasks (e.g., controlling vehicles from raw sensor inputs).

Recent advances include hybrid action space architectures that combine discrete and continuous control. Jin et al.² proposed a hybrid action-based reinforcement learning approach for multi-objective autonomous driving that handles both discrete decisions (lane changes) and continuous control (steering, acceleration) simultaneously. Their framework demonstrated improved compatibility between efficiency and safety objectives compared to traditional methods.

In general, actor-critic algorithms (DDPG, PPO, SAC) are preferred for handling the continuous and smooth action space required by comfortable driving, as they allow for the fine-grained control necessary for passenger comfort. The DDPG algorithm, introduced by Lillicrap et al.⁸, has been particularly influential in autonomous driving applications due to its ability to handle continuous action spaces effectively.

2.2 Model-based and Hybrid RL

To overcome sample inefficiency and incorporate physical knowledge, some works explore model-based RL. One example is the Safe Model-Based Off-Policy RL (SMORL) proposed by Zhu⁴, which integrated a learned vehicle dynamics model and safety checker into the training loop. This approach addressed issues of reward engineering and constraint violations by guiding the agent with model predictions.

Another line of research combines RL with classical planners or controllers. Bautista-Montesano et al. cited in⁵ coupled RL with a Model Predictive Control (MPC) module for navigating intersections, using RL for high-level decisions and MPC for low-level control. Such hybrid systems leverage the strengths of model-based control (smoothness, stability) and

learning (adaptability to complex scenarios).

Multi-objective reinforcement learning (MORL) has emerged as a promising approach for balancing competing objectives in autonomous driving. Li and Czarnecki⁹ introduced a thresholded lexicographic Q-learning method for urban driving that first optimizes higher-priority goals (safety and rule compliance) and only then optimizes comfort and progress. Their multi-objective DQN agent learned to navigate multi-lane roads and intersections while yielding, changing lanes, and respecting traffic rules, without the need to hand-tune a single reward trade-off. Yang et al.¹⁰ proposed a generalized MORL algorithm capable of adapting policies to different objective preferences, effectively approximating the Pareto frontier of optimal driving behaviors. This allows for more flexible policy deployment based on varying situational priorities.

More recently, Zhang et al.¹¹ applied a lexicographic actor-critic to urban driving, demonstrating improved performance by giving strict priority to safety-critical criteria before optimizing ride comfort and efficiency. Similarly, He and Lv¹² focused specifically on energy efficiency as a key objective alongside safety and progress, developing a multi-objective RL approach that significantly improved energy conservation without compromising safety.

Overall, while pure model-free RL dominates current autonomous driving research, these model-based and hybrid enhancements are increasingly important for multi-objective tasks that balance efficiency, safety, and comfort.

2.3 Autonomous Driving Frameworks and Architectures

RL can be applied at different levels of the driving stack. Some approaches train end-to-end policies that map sensor inputs (e.g., camera images or LiDAR) directly to steering/throttle. End-to-end RL is powerful but faces high-dimensional observations and less interpretability⁵. A milestone example is "Learning to Drive in a Day" by Kendall et al.¹³, which achieved the first real-world end-to-end driving via deep RL. Starting from scratch, their agent learned a lane-keeping policy from monocular camera input by interacting with a real car, proving that with careful reward design and safety constraints, an RL policy can be trained on hardware in a short time frame.

Other approaches use a mid-level representation of the state (such as vehicle speeds, distances to other cars, etc.) and let the RL agent make decisions like gap selection or acceleration, which are then executed by lower-level controllers. This can simplify the learning problem and improve sample efficiency⁵. Xu et al.¹⁴ presented an RL approach for highway driving decisions,

modeling it as an MDP with separate rewards for safety, smoothness, and speed. They used a multi-objective policy iteration method to learn when to change lanes or overtake, and notably deployed the learned policy on a real vehicle, confirming that the RL agent could handle highway merges and overtakes in real traffic.

For example, a common setup for highway driving is to give the RL agent a state representation including the relative positions and velocities of surrounding vehicles (often termed an observation space), and let it choose a continuous acceleration or lane-change action. This kind of formulation was used by Jin et al.² and others to handle multi-lane highway scenarios with RL, often with a high-level decision followed by a low-level PID controller for execution.

Selvaraj et al.¹⁵ proposed a comprehensive framework that explicitly addresses the triple objective of efficiency, safety, and comfort in autonomous driving. Their approach used a deep reinforcement learning agent that balances these competing objectives through careful reward design and an integrated co-simulation platform (CoMoVe) that combines traffic simulation (SUMO), network communication (ns-3), and vehicle dynamics (MATLAB/Simulink).

Beyond single-agent scenarios, there is growing interest in multi-agent RL for driving, such as multiple autonomous vehicles interacting or an autonomous vehicle in traffic with human drivers.

Multi-agent settings introduce additional objectives like cooperative efficiency (e.g., managing intersections or merging) and fairness, which complicate reward design further. Surveys like Bai et al.⁵ and Kiran et al.⁶ document how RL has been applied across these contexts and note that despite progress, issues like sample complexity, interpretability, and multi-objective trade-offs are still open problems.

2.4 Simulators and Testing Environments

Because deploying unproven RL policies on real cars is unsafe, researchers rely on high-fidelity simulators to train and evaluate autonomous driving RL agents. CARLA, introduced by Dosovitskiy et al.¹⁶, is an open-source urban driving simulator that has become a de facto standard environment for RL-based autonomous driving research. CARLA provides realistic city scenes, traffic scenarios, and sensor streams (camera, LiDAR), enabling training of deep RL policies in complex urban conditions. Many RL studies use CARLA to test policies in tasks like lane following, obstacle avoidance, and navigation under different weather conditions. For example, after training a DQN agent, Hossain¹ validated it in CARLA’s town simulation to ensure it maintains lane

position and avoids collisions at high speed.

MetaDrive is another platform introduced by Li et al.¹⁷ to facilitate RL research in diverse driving scenarios. It can generate an infinite variety of roads and traffic situations through procedural generation, supporting both single-agent and multi-agent driving tasks. MetaDrive has been used to benchmark generalization (training on some maps and testing on unseen maps) and safe exploration for RL algorithms. Li et al. demonstrated that training on a wide diversity of road layouts and traffic situations leads to improved zero-shot performance on unseen roads, highlighting the importance of diversity for robust policy learning.

SUMO (Simulation of Urban MObility), documented by Lopez et al.¹⁸, is often employed for traffic-level simulations, where the RL agent may control one vehicle among many. SUMO provides a microscopic traffic model, making it useful for studying how an RL-controlled vehicle interacts with flow or for multi-vehicle coordination (e.g., platooning).

Researchers sometimes integrate multiple simulators to capture different aspects of driving. Selvaraj et al.¹⁵ developed a co-simulation framework called CoMoVe that links SUMO (for traffic), a network simulator (ns-3 for V2X communication), and a vehicle dynamics model in MATLAB/Simulink. This allows training an RL agent under realistic vehicle and communication conditions.

In a typical reinforcement learning framework for driving, the agent consists of a learned policy and the RL algorithm that updates it. At each time step t , the agent observes the current state S_t (e.g., positions and velocities of vehicles) and chooses an action A_t (e.g., acceleration or steering). The action is applied to the environment (simulator), which then provides the agent with a reward R_t and the next state. This feedback loop repeats as the agent learns a policy that maximizes cumulative reward.

3 Methodology

4 Experiments and Results

5 Discussion and Future Work

6 Conclusion

References

- [1] Jumman Hossain. Autonomous driving with deep reinforcement learning in carla simulation. *arXiv preprint arXiv:2306.11217*, 2023. URL <https://arxiv.org/abs/2306.11217>.
- [2] Guizhe Jin, Zhuoren Li, Bo Leng, Wei Han, Lu Xiong, and Chen Sun. Hybrid action based reinforcement learning for multi-objective compatible autonomous driving. *arXiv preprint*, arXiv:2501.08096v1, Jan 2025. URL <https://arxiv.org/abs/2501.08096>.
- [3] Meixin Zhu, Yinhai Wang, Ziyuan Pu, Jingyun Hu, Xuesong Wang, and Ruimin Ke. Safe, efficient, and comfortable velocity control based on reinforcement learning for autonomous driving. *Transportation Research Part C: Emerging Technologies*, 117:102662, 2020. URL https://meixinzhu.github.io/pdf/Safe_efficien_Zhu.pdf.
- [4] Zhaoxuan Zhu. *Reinforcement Learning in Eco-driving for Connected and Automated Vehicles*. PhD thesis, Ohio State University, 2021. URL https://rave.ohiolink.edu/etdc/view?acc_num=osu1638536531293253.
- [5] S. Bai. A survey of reinforcement learning theories in autonomous vehicles. *Applied and Computational Engineering*, 85:187–194, 2024. URL https://www.researchgate.net/publication/382758563_A_survey_of_reinforcement_learning_theories_in_autonomous_vehicles.
- [6] B. R. Kiran, Ibrahim Sobh, Victor Talpaert, Patrick Mannion, Ahmad A. Al Sallab, Senthil Yogamani, and Patrick Pérez. Deep reinforcement learning for autonomous driving: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 23(6):4909–4926, June 2022. doi: 10.1109/TITS.2021.3054625.

- [7] Ahmad EL Al-Sallab, Mahmoud Abdou, Etienne Perot, and Senthil Yogamani. End-to-end deep reinforcement learning for lane keeping assist. In *NIPS 2016 Workshop on Machine Learning for Intelligent Transportation Systems*, 2016.
- [8] Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. *arXiv preprint*, arXiv:1612.04340, 2016. URL <https://arxiv.org/pdf/1509.02971>.
- [9] Changliu Li and Krzysztof Czarnecki. Urban driving with multi-objective deep reinforcement learning. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, pages 359–367, 2019.
- [10] Runzhe Yang, Xingyuan Sun, and Karthik Narasimhan. A generalized algorithm for multi-objective reinforcement learning and policy adaptation. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/4a46fbfca3f1465a27b210f4bdf6ab3-Paper.pdf.
- [11] Hengrui Zhang, Youfang Lin, Sheng Han, and Kai Lv. Lexicographic actor-critic deep rl for urban autonomous driving. *IEEE Transactions on Vehicular Technology*, 72(4):4308–4319, 2023. doi: 10.1109/TVT.2022.3226579.
- [12] Xiangkun He and Chen Lv. Towards energy-efficient autonomous driving: A multi-objective reinforcement learning approach. *IEEE/CAA Journal of Automatica Sinica*, 10(JAS-2022-1495):1329, 2023. ISSN 2329-9266. doi: 10.1109/JAS.2023.123378. URL <https://www.ieee-jas.net/en/article/doi/10.1109/JAS.2023.123378>.
- [13] Alex Kendall, Jeffrey Hawke, David Janz, Przemyslaw Mazur, Daniele Reda, John-Mark Allen, Vinh-Dieu Lam, Alex Bewley, and Amar Shah. Learning to drive in a day. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 8248–8254, 2019. doi: 10.1109/ICRA.2019.8793742.
- [14] Xin Xu, Lei Zuo, Xin Li, Lilin Qian, Junkai Ren, and Zhenping Sun. A reinforcement learning approach to autonomous decision making of intelligent vehicles on highways. *IEEE Transactions on Sys-*

- tems, Man, and Cybernetics: Systems*, 50(10):3884–3897, 2020. doi: 10.1109/TSMC.2018.2870983.
- [15] Dinesh Cyril Selvaraj, Shailesh Hegde, Nicola Amati, Francesco Deflorio, and Carla Fabiana Chiasserini. A deep reinforcement learning approach for efficient, safe and comfortable driving. *Applied Sciences*, 13(9), 2023. ISSN 2076-3417. doi: 10.3390/app13095272. URL <https://www.mdpi.com/2076-3417/13/9/5272>.
 - [16] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. CARLA: An open urban driving simulator. In Sergey Levine, Vincent Vanhoucke, and Ken Goldberg, editors, *Proceedings of the 1st Annual Conference on Robot Learning*, volume 78 of *Proceedings of Machine Learning Research*, pages 1–16. PMLR, 13–15 Nov 2017. URL <https://proceedings.mlr.press/v78/dosovitskiy17a.html>.
 - [17] Quanyi Li, Zhenghao Peng, Lan Feng, Qihang Zhang, Zhenghai Xue, and Bolei Zhou. Metadrive: Composing diverse driving scenarios for generalizable reinforcement learning, 2022. URL <https://arxiv.org/abs/2109.12674>.
 - [18] Pablo Alvarez Lopez, Michael Behrisch, Laura Bieker-Walz, Jakob Erdmann, Yun-Pang Flötteröd, Robert Hilbrich, Leonhard Lücken, Johannes Rummel, Peter Wagner, and Evamarie Wiessner. Microscopic traffic simulation using sumo. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 2575–2582, 2018. doi: 10.1109/ITSC.2018.8569938.