

# 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

The integration of deep reinforcement learning (RL) into autonomous driving has emerged as a significant research domain in recent years. Contemporary RL algorithms demonstrate remarkable capabilities in learning complex driving policies within high-dimensional environmental contexts, particularly in sophisticated traffic simulation platforms. The application spectrum of RL in vehicular autonomy encompasses diverse operational tasks, including lane maintenance, adaptive cruise control, intersection navigation, and overtaking maneuvers. Notably, in certain controlled scenarios, RL-based systems have demonstrated performance metrics comparable to or exceeding human driving capabilities<sup>1</sup>.

Foundational research established the feasibility of applying RL to continuous vehicle control systems. The Deep Deterministic Policy Gradient (DDPG) algorithm, when implemented in autonomous driving contexts, demonstrated that RL agents could acquire human-comparable proficiency in vehicle steering and acceleration within racing simulation environments<sup>2</sup>. Subsequently, numerous investigations have employed deep RL methodologies to optimize singular objectives in driving tasks. Prevalent objectives

include safety parameters (collision avoidance) and efficiency metrics (minimization of travel duration or fuel consumption). Exemplifying this approach, RL controllers optimized for fuel efficiency have achieved notable improvements—specifically, a Proximal Policy Optimization (PPO)-based agent demonstrated a 17% enhancement in fuel economy relative to human baseline performance while maintaining comparable transit times<sup>3</sup>. Similarly, RL policies optimized for safety through substantial penalization of collisions or regulatory infractions have yielded agents capable of extended accident-free operation periods. These single-objective accomplishments substantiate RL as an efficacious tool for autonomous vehicle decision-making processes<sup>4</sup>.

Nevertheless, driving inherently constitutes a multidimensional problem domain involving complex trade-offs among various objectives including temporal efficiency, energy conservation, operational safety, and passenger comfort. Traditional single-objective approaches—such as those exclusively prioritizing velocity or fuel efficiency—frequently neglect secondary criteria like ride quality. Contemporary research has begun incorporating multiple objectives within the RL framework, acknowledging that authentically human-like autonomous driving must optimize not only efficiency parameters but also safety and comfort metrics. The subsequent sections review the state-of-the-art developments in multi-objective RL for autonomous driving, with particular emphasis on the simultaneous optimization of efficiency (temporal/fuel) and comfort (ride smoothness) parameters.

## **1.2 Research Gaps and Challenges**

## **1.3 Our Contributions**

# **2 Related Work**

## **2.1 Deep RL Algorithms**

A comprehensive spectrum of deep RL algorithms has been applied to autonomous driving challenges, each presenting distinct advantages. As extensively documented by Kiran et al.<sup>5</sup> in their systematic review, value-based methodologies such as DQN and Double DQN—operating within discrete action spaces (e.g., discrete throttle/brake gradations or high-level decision sets)—have been successfully implemented in lane-following and fundamental driving tasks<sup>6</sup>. For instance, DQN agents demonstrate capability in maintaining lane positioning and obstacle avoidance through selection from a finite set of steering angle options. Pioneering research by Sallab et al.<sup>7</sup> established the viability of end-to-end deep RL for lane-keeping assistance

derived from raw visual input, demonstrating that neural network architectures could acquire steering control policies without reliance on manual rule-based controller systems. However, discrete action spaces present limitations in capturing the nuanced control necessary for comfort optimization (e.g., smooth braking transitions), necessitating the combination of value-based RL with action discretization techniques.

Policy gradient and actor-critic methodologies demonstrate greater applicability for continuous control scenarios. The Deep Deterministic Policy Gradient (DDPG) algorithm—an off-policy actor-critic approach generating continuous steering or acceleration commands—has found widespread application in driving scenarios necessitating smooth control implementations, such as car-following behaviors. Zhu et al.<sup>2</sup> employed DDPG for adaptive cruise control, noting its direct optimization of continuous acceleration parameters and capacity to incorporate multiple reward components addressing safety, efficiency, and comfort considerations. Their empirical evaluation using authentic traffic data demonstrated that the RL agent maintained enhanced safety time headways and reduced acceleration variations, simultaneously indicating improvements in comfort and safety metrics compared to conventional control systems.

Proximal Policy Optimization (PPO), an on-policy methodology, has also been implemented in driving scenarios due to its training stability characteristics. In an eco-driving investigation, a PPO agent incorporating Long Short-Term Memory (LSTM) network architecture was utilized to optimize velocity profiles, successfully reducing fuel consumption while adhering to temporal constraints<sup>3</sup>. Soft Actor-Critic (SAC), which maximizes the trade-off between reward acquisition and entropy maintenance, has undergone testing in driving simulations; its sample-efficient learning paradigm and robust convergence properties render it appropriate for high-dimensional tasks (e.g., vehicle control derived from raw sensor inputs).

Recent advancements include hybrid action space architectures that integrate discrete and continuous control mechanisms. Jin et al.<sup>1</sup> proposed a hybrid action-based reinforcement learning framework for multi-objective autonomous driving that simultaneously manages discrete decisions (lane transitions) and continuous control parameters (steering, acceleration). Their framework exhibited enhanced compatibility between efficiency and safety objectives compared to conventional methodologies.

Generally, actor-critic algorithms (DDPG, PPO, SAC) are preferentially selected for managing the continuous and smooth action space required for comfortable driving experiences, as they facilitate the fine-grained control essential for passenger comfort optimization. The DDPG algorithm, introduced by Lillicrap et al.<sup>8</sup>, has been particularly influential in autonomous

driving applications due to its effective handling of continuous action spaces.

## 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 Zhu<sup>3</sup>, 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 in<sup>4</sup> 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 Czarnecki<sup>9</sup> 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.<sup>10</sup> 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.<sup>11</sup> 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 Lv<sup>12</sup> 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 interpretability<sup>4</sup>. A milestone example is "Learning to Drive in a Day" by Kendall et al.<sup>13</sup>, 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 efficiency<sup>4</sup>. Xu et al.<sup>14</sup> 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.<sup>1</sup> 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.<sup>15</sup> 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 compli-

cate reward design further. Surveys like Bai et al.<sup>4</sup> and Kiran et al.<sup>5</sup> 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.<sup>16</sup>, 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, Hossain<sup>6</sup> 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.<sup>17</sup> 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.<sup>18</sup>, 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.<sup>15</sup> 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] 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>.
- [2] 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\\_efficient\\_Zhu.pdf](https://meixinzhu.github.io/pdf/Safe_efficient_Zhu.pdf).
- [3] 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](https://rave.ohiolink.edu/etdc/view?acc_num=osu1638536531293253).
- [4] 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](https://www.researchgate.net/publication/382758563_A_survey_of_reinforcement_learning_theories_in_autonomous_vehicles).
- [5] 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.
- [6] Jumman Hossain. Autonomous driving with deep reinforcement learning in carla simulation. *arXiv preprint arXiv:2306.11217*, 2023. URL <https://arxiv.org/abs/2306.11217>.
  - [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/4a46fbfca3f1465a27b210f4bdfef6ab3-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2019/file/4a46fbfca3f1465a27b210f4bdfef6ab3-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 Systems, Man, and Cybernetics: Systems*, 50(10):3884–3897, 2020. doi: 10.1109/TSMC.2018.2870983.
  - [15] Dinesh Cyril Selvaraj, Shailesh Hegde, Nicola Amati, Francesco Defflorio, 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.