

Reinforcement Learning for Autonomous Vehicle Control: Environment Representation, Reward Decomposition, and Adaptive Policies

Overview of Ph.D. Thesis by:
László Szőke



M Ű E G Y E T E M 1 7 8 2

Budapest University of Technology and Economics
Faculty of Transportation Engineering and Vehicle Engineering
Department of Control Transportation and Vehicle Systems

Supervisor:

Szilárd Aradi, PhD

Department of Control for
Transportation and Vehicle Systems
Budapest University of Technology and Economics

Submitted in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

June 2, 2025

1 Introduction

Autonomous vehicles (AVs) represent a transformative potential, promising safer and more comfortable transportation by mitigating human error and fatigue, and responding faster than human drivers in critical situations [1, 2].

Advancements in artificial intelligence (AI) and machine learning (ML) have accelerated developments in autonomous driving, enabling complex decision-making, scene understanding, and real-time adaptation [3, 4]. While traditional control methods, such as Proportional–Integral–Derivative (PID) and Linear Quadratic Regulator (LQR), remain effective in structured scenarios, they struggle with the diversity and unpredictability encountered in real-world driving [5, 6, 7]. Autonomous vehicle systems must perform accurately and robustly across numerous varying conditions, a challenge difficult to address solely through classical, rule-based algorithms [8, 9]. As such, learning-based methods have emerged as promising solutions, capable of generalizing and adapting to unforeseen situations without explicit rules [10].

This dissertation focuses on advancing autonomous vehicle motion control through reinforcement learning (RL) methodologies. Given the complexity and variability of real-world driving scenarios, traditional control strategies often fall short in adaptability and generalization. The dissertation addresses key challenges, including environment representation, reward design, handling imbalanced and conflicting objectives, adaptability across dynamic systems, and effective transferability of learned behaviors. It explores novel methodologies to enhance adaptability and generalization, particularly emphasizing reward design and the dynamic adaptability of learned behaviors.

Moreover, the work explores the broader implications of RL methodologies beyond autonomous driving, demonstrating their versatility and efficacy in related domains such as adaptive traffic signal control. By examining various learning methodologies, including Successor Features [11, 12], reward decomposition, and curriculum-based approaches [13], this work aims to provide insights and practical tools for future practitioners in the domain of intelligent transportation systems.

2 Problem definition

Autonomous vehicle control represents a significant technical and conceptual challenge in modern transportation systems, requiring novel solutions capable of handling real-world complexity and variability. The core issue this dissertation addresses is the development of robust, adaptable, and generalizable decision-making frameworks for autonomous vehicles, specifically utilizing reinforcement learning techniques [14, 7]. Traditional control methods, although reliable within fixed and predictable environments, typically lack the flexibility needed for complex scenarios encountered in autonomous driving, such as sudden lane changes, unpredictable pedestrian behavior, and dynamic traffic conditions [6, 7].

Thus, the primary problems investigated by this research include:

- Identifying and evaluating suitable RL algorithms tailored to diverse autonomous driving functionalities, ensuring they can handle varying complexity and unpredictable scenarios effectively.[15]
- Exploring and refining methodologies that improve RL agents' adaptability and generalization to unseen environments and scenarios, addressing a fundamental limitation of classical control solutions[13, 11, 16].
- Investigating critical aspects of reward design in RL, which significantly influence the quality and efficiency of training processes, and proposing practical solutions for addressing challenges such as reward imbalance, conflicting objectives, and degenerate behaviors [13, 11].
- Demonstrating how RL approaches, particularly those utilizing reward decomposition and curriculum learning, can dynamically adapt agent behaviors to changing preferences and operational requirements without retraining, enhancing the practical viability and flexibility of learned models [16, 13].
- Developing techniques to ensure robust control policies capable of managing system uncertainties and variations, thereby facilitating effective adaptation of behavior in control scenarios, despite differences between simulation environments and robot dynamic parameters [17].

This research's significance stems from its potential to contribute new insights and robust methodologies for integrating reinforcement learning into autonomous vehicle systems. By addressing these core problems, this work seeks to not only overcome existing limitations but also pave the way for further advancements in intelligent transportation, ultimately enhancing vehicle safety, efficiency, and adaptability.

3 Contributions of the Thesis

Thesis I.

I investigated whether applying image-based state representations for reinforcement learning agents helps scene understanding and yields higher rewards and success rates for highway traffic scenarios. I designed two agents: one with a minimal state space representation and a simple neural network structure. The second agent was designed with temporal and spatial capabilities using CNN and LSTM network components for enhanced scenario understanding. I tested my agents on a self-tailored test environment for a highway driving scenario using the Simulation of Urban Mobility software. I compared how agents' state space design and reward functions influence the performance. While the image-based representation is more memory-intensive, when combined with spatial and temporal processing

(LSTM), it effectively models traffic situations and yields better performance compared to the structural state space representation. As a result, I presented an efficiently functioning, learning-capable agent that uses temporal and spatial state representations and performs well in highway traffic scenarios.

Publications Related to the Thesis: [18, 19, 15]

Thesis II.

I hypothesized that implementing a successor features-based algorithm for reward decomposition can help policy learning and induce customizable behavior without retraining in autonomous vehicle highway control. I developed a reinforcement learning algorithm (DFRL) for the efficient behavior of autonomous vehicles on highways, utilizing the Successor Features method, which interprets multi-objective optimization tasks by decomposing them into elementary reward functions. The trained agent can adapt to changes in preferences in a parameterized manner without the need for retraining. Also, it is capable of a smooth and safe highway commute, with the ability to change preferences rapidly. I tested my method in an environment where an autonomous vehicle is operating in highway traffic. Experiments underline that the objectives of the agent could be altered, and it was capable of maintaining similar performance and episode completion rates in the new tasks without any retraining. I demonstrated the trained agent's various behavioral modes by modifying the reward components, emphasizing the system's flexibility. I proved my hypothesis and presented an agent that is capable of efficient and safe highway commutes while it adapts its behavior based on changing preferences.

Publications Related to the Thesis: [11, 12]

Thesis III.

I designed and trained a neural agent for traffic signal control at intersections. The agent is parameterizable based on multiple criteria, and it realizes efficient control by adjusting the duration of predefined phases of the intersection traffic lights based on the current preference setting. I applied the Successor Features method to the traffic signal control domain. I showed how the rewards and objective function can be defined, enabling adjustments to optimize the objective function based on changing preferences. I suggested a decomposition of the traffic signal control objective function into elementary reward functions, which enables the usage of the DFRL algorithm in this domain. I demonstrated the power of the algorithm through an example that considers aspects such as emissions, fuel consumption, delays, traffic queues, and the prioritization of certain vehicles. I tested my solution in a four-way traffic intersection, where the agent has to control the traffic lights to achieve

efficient traffic flow, minimal CO2 emissions, and minimal time loss for the participants. I highlighted how adaptive behavior can be beneficial in the case of changing preferences for intersection control without retraining or fine-tuning. I evaluated the agent using a prioritized vehicle that needed to exit the intersection as soon as possible. I presented that my algorithm outperforms both the classic and smart controllers in terms of all the monitored metrics, and I proved that using successor feature-based reward decomposition is beneficial in the Traffic Signal Control domain.

Publications Related to the Thesis: [16]

Thesis IV.

I developed a new reinforcement learning algorithm training method combining Successor Features and Curriculum Learning called PrefVeC. The method is applicable in the case of imbalanced reward functions and contradicting goals. I developed the ability to adaptively and automatically weight reward function components of varying magnitudes during the learning process. This method enables the gradual prioritization of critical rewards, ensuring their influence is reflected in the agent’s behavior. I proposed 2 types of the PrefVeC algorithm to ensure adaptivity and minimize domain knowledge necessity in case of applying my method. I tested my work in 3 different environments, where the safety awareness and acting capabilities were measured. I compared the proposed training approach with state-of-the-art algorithms. I concluded that my solution improves convergence for the learning agent, even with reward functions characterized by components of different orders of magnitude. The PrefVeC agents are superior because they avoid trivial solutions (not doing anything) and learn meaningful policies. I conducted ablation studies on the components of my proposed training method and discussed their contribution to the overall performance of the algorithm.

Publications Related to the Thesis: [13]

Thesis V.

I hypothesized that disentangling the physical parameters of the controlled system from the task can improve the adaptivity of RL agents. I developed a technique for managing dynamic parameter uncertainties in the case of varying scenarios by creating adaptive policies. The designed adaptive structure is capable of controlling systems with dynamic uncertainties using discrete control signals. It utilizes a neural encoder trained on historical state transition data to represent the dynamics of the current controlled system in a latent space. The additional information extracted from the historical transitions contributes to the agent’s performance. I proved the performance gains in the environment of a

differential-driven robot with a discrete action space, where I trained an agent that realizes smooth control by estimating the dynamic parameters of the robot. The environment can be randomized to propose challenging scenarios for the trained agent. The created agent can adapt using latent representations derived from historical transition data, thereby improving control quality in real-time applications. Using historical data during training, I encompass meaningful information that can aid the uncertainties raised by the new robot configurations. Thanks to my training methodology, the agent can solve tasks and control robots with different dynamics. I compared the agent to conventional methods, and I showed that both the control quality and final performance were better or on par with them. However, my solution is out of the box and is applicable to new robot setups without the need for retraining or extra fine-tuning.

Publications Related to the Thesis: [17]

4 Future Research

Although this dissertation has laid a comprehensive foundation for using reinforcement learning in autonomous vehicle systems, numerous opportunities for future research remain open. Building upon the insights and methodologies developed herein, several promising research directions can further enhance RL applications in safety-critical domains such as autonomous vehicles and intelligent traffic management.

Extended Application of Reward Decomposition Techniques This work successfully explored reward decomposition, primarily through linear compositions via successor features. Future research should investigate advanced techniques, such as non-linear reward decompositions or hierarchical successor features, to address more intricate and realistically complex decision-making scenarios.

Multi-Agent Coordination and Communication The focus of the dissertation was primarily on single-agent RL approaches. Extending the current work to multi-agent reinforcement learning (MARL) scenarios presents significant opportunities, especially in traffic control or fleet-based autonomous vehicle applications. Future work should explore coordinated multi-agent decision-making, cooperative perception systems, and the exchange of information between autonomous entities and infrastructure.

Real-World Deployment and Sim-to-Real Transfer While the proposed adaptation mechanisms addressed some challenges related to Sim-to-Real transfer, substantial work remains in bridging simulation-trained RL agents and real-world deployment. Future research could focus on robustifying transfer methods, utilizing domain randomization,

adaptive meta-learning, or hybrid approaches combining model-based methods and RL to reduce the performance gap between simulated and physical environments.

Integration of Safe Reinforcement Learning Safety-critical environments such as autonomous driving inherently require robust mechanisms to ensure fail-safe operations. Future research should explicitly focus on integrating SafeRL principles with existing successor feature-based methods, curriculum learning, and dynamic reward adjustment strategies presented in this work. Creating risk-aware policy optimization techniques, constraint-based RL frameworks, and formal verification methods could be interesting follow-up topics.

Scalable and Adaptive Curriculum Learning The curriculum learning methods introduced in this thesis were shown to address imbalanced rewards and conflicting objectives effectively. Expanding these concepts to more sophisticated curricula generation strategies presents a promising research direction.

In summary, the pathways outlined here aim to broaden and deepen the practical applications of reinforcement learning within autonomous systems. By pursuing these future research directions, the methods and insights developed in this dissertation can further shape autonomous vehicle technology, traffic management, and beyond, ultimately contributing toward safer, more adaptable, and more intelligent transportation systems.

References

- [1] James M. Anderson et al. *Autonomous Vehicle Technology: A Guide for Policymakers*. Santa Monica, CA: RAND Corporation, 2016. ISBN: 978-0-8330-9472-8. URL: https://www.rand.org/pubs/research_reports/RR443-2.html.
- [2] Steven E. Shladover. “Connected and Automated Vehicle Systems: Introduction and Overview”. In: *Journal of Intelligent Transportation Systems* 22.3 (2018), pp. 190–200. DOI: 10.1080/15472450.2017.1336053.
- [3] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. “Deep Learning”. In: *Nature* 521.7553 (2015), pp. 436–444. DOI: 10.1038/nature14539.
- [4] Szilárd Aradi. “Survey of Deep Reinforcement Learning for Motion Planning of Autonomous Vehicles”. In: *IEEE Transactions on Intelligent Transportation Systems* 23.2 (2022), pp. 740–759. DOI: 10.1109/TITS.2020.3024655.
- [5] Brian Paden, Michal Čáp, Sze Zheng Yong, Dmitry Yershov, and Emilio Frazzoli. “A Survey of Motion Planning and Control Techniques for Self-Driving Urban Vehicles”. In: *IEEE Transactions on Intelligent Vehicles* 1.1 (2016), pp. 33–55. DOI: 10.1109/TIV.2016.2578706.

- [6] Simon Ulbrich, Till Menzel, Andreas Reschka, Fabian Schuldt, and Markus Maurer. “Defining and Substantiating the Terms Scene, Situation, and Scenario for Automated Driving”. In: 2015, pp. 982–988. DOI: 10.1109/ITSC.2015.164.
- [7] Wilko Schwarting, Javier Alonso-Mora, and Daniela Rus. “Planning and Decision-Making for Autonomous Vehicles”. In: *Annual Review of Control, Robotics, and Autonomous Systems* 1.1 (2018), pp. 187–210. ISSN: 2573-5144. DOI: 10.1146/annurev-control-060117-105157.
- [8] Chris Urmson, Joshua Anhalt, Drew Bagnell, Christopher Baker, and Robert Bittner. “Autonomous Driving in Urban Environments: Boss and the Urban Challenge”. In: *Journal of Field Robotics* 25.8 (2008), pp. 425–466. ISSN: 1556-4959. DOI: 10.1002/rob.20255.
- [9] Michael Montemerlo, Jan Becker, Suhrid Bhat, Hendrik Dahlkamp, and Dmitri Dolgov. “Junior: The Stanford Entry in the Urban Challenge”. In: *Journal of Field Robotics* 25.9 (Sept. 2008), pp. 569–597. DOI: 10.1002/rob.20258.
- [10] Ardi Tampuu, Tambet Matiisen, Maksym Semikin, Dmytro Fishman, and Naveed Muhammad. “A Survey of End-to-End Driving: Architectures and Training Methods”. In: *IEEE Transactions on Neural Networks and Learning Systems* 33.4 (2022), pp. 1364–1384. DOI: 10.1109/TNNLS.2020.3043505.
- [11] Laszlo Szoke, Szilárd Aradi, Tamás Bécsi, and Péter Gáspár. “Skills to Drive: Successor Features for Autonomous Highway Pilot”. In: *IEEE Transactions on Intelligent Transportation Systems* (2022). DOI: 10.1109/TITS.2022.3150493.
- [14] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. 2nd. Adaptive Computation and Machine Learning series. MIT Press, 2018. ISBN: 9780262039246. URL: <https://books.google.de/books?id=uWVDwAAQBAJ>.

Own Publications

- [12] Laszlo Szoke, Szilárd Aradi, and Tamás Tettamanti. “Investigating Successor Features in the domain of Autonomous Vehicle Control”. In: 2021. DOI: 10.1016/j.trpro.2022.02.023.
- [13] Laszlo Szoke, Shahaf S. Shperberg, Jarrett Holtz, and Alessandro Allievi. “Adaptive Curriculum Learning With Successor Features for Imbalanced Compositional Reward Functions”. In: *IEEE Robotics and Automation Letters* 9.6 (2024), pp. 5174–5181. DOI: 10.1109/LRA.2024.3387134.
- [15] Laszlo Szoke, Szilárd Aradi, Tamás Bécsi, and Péter Gáspár. “Vehicle Control in Highway Traffic by Using Reinforcement Learning and Microscopic Traffic Simulation”. In: IEEE, 2020, pp. 21–26. DOI: 10.1109/SISY50555.2020.9217076.

- [16] Laszlo Szoke, Szilárd Aradi, and Tamás Bécsi. “Traffic Signal Control with Successor Feature-Based Deep Reinforcement Learning Agent”. In: *MDPI Electronics* 12.6 (Mar. 2023), p. 1442. DOI: [10.3390/electronics12061442](https://doi.org/10.3390/electronics12061442).
- [17] Laszlo Szoke, Peter Farkas, Szilard Aradi, and Tamas Becsi. “Adapting RL Control Policies to Changing Dynamics for Improved Robustness”. In: Cham: Springer Nature Switzerland, 2025, pp. 78–89. ISBN: 978-3-031-87620-2. DOI: https://doi.org/10.1007/978-3-031-87620-2_8.
- [18] Laszlo Szoke and Szilárd Aradi. “Autonóm ágens tanítása megerősítéses tanulás alkalmazásával autópályán való optimális közlekedéshez”. In: 2019. URL: https://mmaws.bme.hu/2019/pages/program/papers/Paper_16_Szoke_L_Aradi_Sz_IFFK_2019.pdf.
- [19] Laszlo Szoke, Szilárd Aradi, Tamás Bécsi, and Péter Gáspár. “Driving on Highway by Using Reinforcement Learning with CNN and LSTM Networks”. In: IEEE. 2020, pp. 121–126. DOI: [10.1109/INES49302.2020.9147185](https://doi.org/10.1109/INES49302.2020.9147185).