УДК 004.8:004.94:629.113.018.7
DOI: 10.18523/2617-3808.2025.8.68-75

V. Beimuk, D. Kuzmenko

ENERGY CONSERVATION FOR AUTONOMOUS AGENTS USING REINFORCEMENT LEARNING

Reinforcement learning (RL) has shown strong potential in autonomous racing for its adaptability to complex and dynamic driving environments. However, most research prioritizes performance metrics such as speed and lap time. Limited consideration is given to improving energy efficiency, despite its increasing importance in sustainable autonomous systems. This work investigates the capacity of RL agents to develop multi-objective driving strategies that balance lap time and fuel consumption by incorporating a fuel usage penalty into the reward function. To simulate realistic uncertainty, fuel usage is excluded from the observation space, forcing the agent to infer fuel consumption indirectly. Experiments are conducted using the Soft Actor-Critic algorithm in a high-fidelity racing simulator, Assetto Corsa, across multiple configurations of vehicles and tracks.

We compare various penalty strengths against the non-penalized agent and evaluate fuel consumption, lap time, acceleration and braking profiles, gear usage, engine RPM, and steering behavior. Results show that mild to moderate penalties lead to significant fuel savings with minimal or no loss in lap time. Our findings highlight the viability of reward shaping for multi-objective optimization in autonomous racing and contribute to broader efforts in energy-aware RL for control tasks. Results and supplementary material are available on our project website.

Keywords: reinforcement learning, autonomous driving, energy efficiency, multi-objective optimization, Soft Actor-Critic, racing simulation.

Introduction

As autonomous driving technology advances, it has the potential to reshape mobility, offering benefits ranging from reduced traffic congestion to fewer accidents [5]. Yet, developing autonomous agents involves complex challenges in perception, planning, control, and decision-making in unpredictable environments [8].

Within this broader field, autonomous racing has emerged as an insightful research area [1]. Like traditional motorsport, it pushes systems to operate at their limits, making it a powerful testbed for high-performance, safe, and efficient algorithms [7]. RL has become a popular approach in this domain due to its ability to learn from high-dimensional inputs [3, 9]. However, most current RL applications in racing focus solely on maximizing speed or minimizing lap time [1]. Energy efficiency is rarely addressed, despite its growing societal and environmental impact.

Our work addresses this gap by investigating how penalizing fuel consumption in the reward function affects RL agent behavior. We aim to encourage the agent to balance speed and energy efficiency, forcing it to learn non-trivial trade-offs. Our key hypothesis is that shaping the reward function to penalize fuel use and incentivize speed leads to more energy-efficient strategies without significantly affecting lap time.

We evaluate agents across multiple vehicle-track combinations and penalty strengths. Results show that even mild penalties can lead to significant fuel savings with minimal performance loss, in some cases even outperforming baselines.

The results highlight the importance of reward design in multi-objective RL and contribute to the broader efforts in energy-aware autonomous systems. Supplementary material and additional results are available at https://nomadflamingo.github.io/assetto_corsa_gym/.

Related Work

Autonomous driving systems have traditionally followed a perception-planning-control pipeline, widely used in both industry and research [1]. More recently, end-to-end systems using RL have gained popularity due to their ability to learn complex behaviors through interaction with the environment [3]. RL has been successfully applied to tasks ranging from highway driving [11] to aggressive maneuvers like overtaking [10].

RL has also shown potential in reducing fuel consumption. For instance, Kim et al. [6] trained a neural network to predict the most fuel-efficient speeds based on road data and trip constraints. Yet, few studies examine how RL agents adapt when fuel usage is treated as a direct constraint rather than a prediction target. This leaves open questions regarding agent behavior under explicit fuel usage penalties.

Another limitation lies in the continuous reliance on simplified simulators that often lack realistic vehicle dynamics; this limits the generalizability of learned policies to real conditions. Additionally, most RL work assumes full observability, despite real-world agents operating without access to direct sensor information [3].

Our work addresses these gaps by applying the Soft Actor-Critic (SAC) algorithm in a high-fidelity simulator, compatible with the OpenAI Gym interface [2]. We introduce fuel efficiency objectives via reward shaping, while excluding fuel level from the observation space to simulate real-world constraints.

Methods

We use the AssettoCorsaGym interface developed by Remonda et al. [7], which integrates a high-fidelity racing simulator, Assetto Corsa, with the OpenAI Gym environment. Figure 1 shows an overview of the AssettoCorsaGym platform.

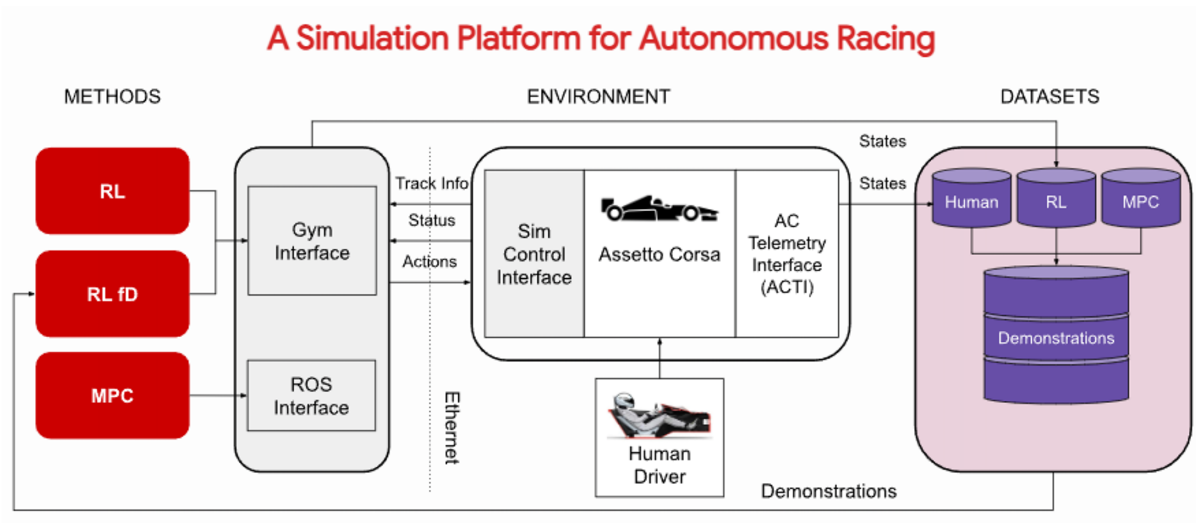

Figure 1. The architecture of the AssettoCorsaGym platform [8]

We used the SAC RL algorithm designed specifically for continuous control tasks [4]. It offers strong performance in autonomous racing benchmarks [7], as well as robustness in tasks that require balance between speed, control, and long-term planning [4, 7].

The setup included two tracks: Track A (Austria) and Track B (Monza) (Figure 2), and two vehicle models: a lightweight Formula 3 series car (F317) and a heavier GT3 series car (BMW Z4 GT3). These selections were made from the Assetto Corsa environment to align with the original AssettoCorsaGym dataset.

Track A offers a balanced layout for general driving evaluation, while Track B, containing tighter corner sequences, tests performance in more challenging conditions.

The two vehicles differ in dynamics. The F3 series car is a lightweight and high-downforce car that is nimble and agile, while the GT3 is a heavier, high-power vehicle that requires more cautious driving strategies, especially during sharp turns.

We extended the AssettoCorsaGym platform to include fuel usage data in the reward calculation. The reward function provided in AssettoCorsaGym is based on the car's velocity and penalizes deviation from the optimal driving line. It is computed as:

$$r = v \cdot (1 - a \cdot d)$$

where v is the car's current speed, d is the L2 distance from the optimal path (as determined by the simulator), and a is a penalty coefficient [7].

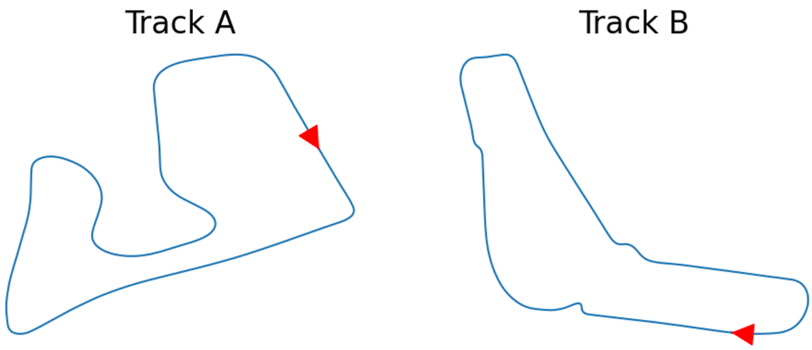

Figure 2. Two tracks chosen for training and evaluating the model in the Assetto Corsa simulator. Red triangles indicate the start positions and directions.

To encourage fuel efficientcy, we extended the reward function with a penalty for fuel consumption. The resulting reward function is defined as:

$$r = v \cdot (1 - a \cdot d) - b \cdot f$$

where f is the change in fuel since the last timestep, and b controls the strength of the penalty. Notably, fuel consumption data was excluded from the agent's observation space to encourage the agent to learn through implicit feedback.

We experimented with four values of the b coefficient – corresponding to approximate reward reductions of 2%, 5%, 10%, and 20%, relative to the original reward function formulation. All training runs were performed on an RTX 3060 laptop GPU. On average, it took approximately 48 hours to complete 500 training episodes.

Experiments

We examined the effect of fuel penalties on RL agent performance during training.

Figure 3 (Top) shows that low to moderate penalties (2-5%) often improved lap times during early training compared to the baseline, particularly with the GT3 car on Track A. However, higher penalties (20%) led to suboptimal policies that heavily prioritized fuel savings. On the more complex Track B, penalties above 2% consistently prevented agents from completing valid laps.

Figure 3 (Bottom) shows the evolution of the fuel consumption rates per lap during training. Across all setups, penalized agents consistently reduced fuel usage over time. This effect was strong on Track A for both vehicles. On Track B, fuel savings were less noticeable and limited by the lower magnitude of the penalty (2%), as higher penalties failed to converge.

Table 1 summarizes the best lap times and corresponding fuel consumption across all setups. On Track A, penalties led to fuel savings up to 0.4L per lap without major performance drops. On Track B, even mild penalties degraded performance and convergence. "DNF" values indicate failure to complete valid laps.

Table 1. **Comparison of best lap times and corresponding fuel consumption rates per lap during training across different setups**

Fuel Penalty	Random Seed	Best Lap Time (s) ↓	Fuel/Lap for Best Lap ↓
Track A, Car: GT3 series			
0%	0	94.97	1.88
	1	93.13	1.80
5%	0	96.59	1.88
	1	**92.57**	1.59
	2	95.09	1.47
10%	1	94.35	1.44
	2	93.80	1.43
20%	1	115.93	**1.01**
Track A, Car: F3 series			
0%	0	**83.62**	0.97
5%	0	83.86	**0.90**

Fuel Penalty	Random Seed	Best Lap Time (s) ↓	Fuel/Lap for Best Lap ↓
Track B, Car: GT3 series			
0%	0	**112.99**	2.45
2%	0	113.85	**2.39**
5%	0	DNF	DNF
10%	0	DNF	DNF

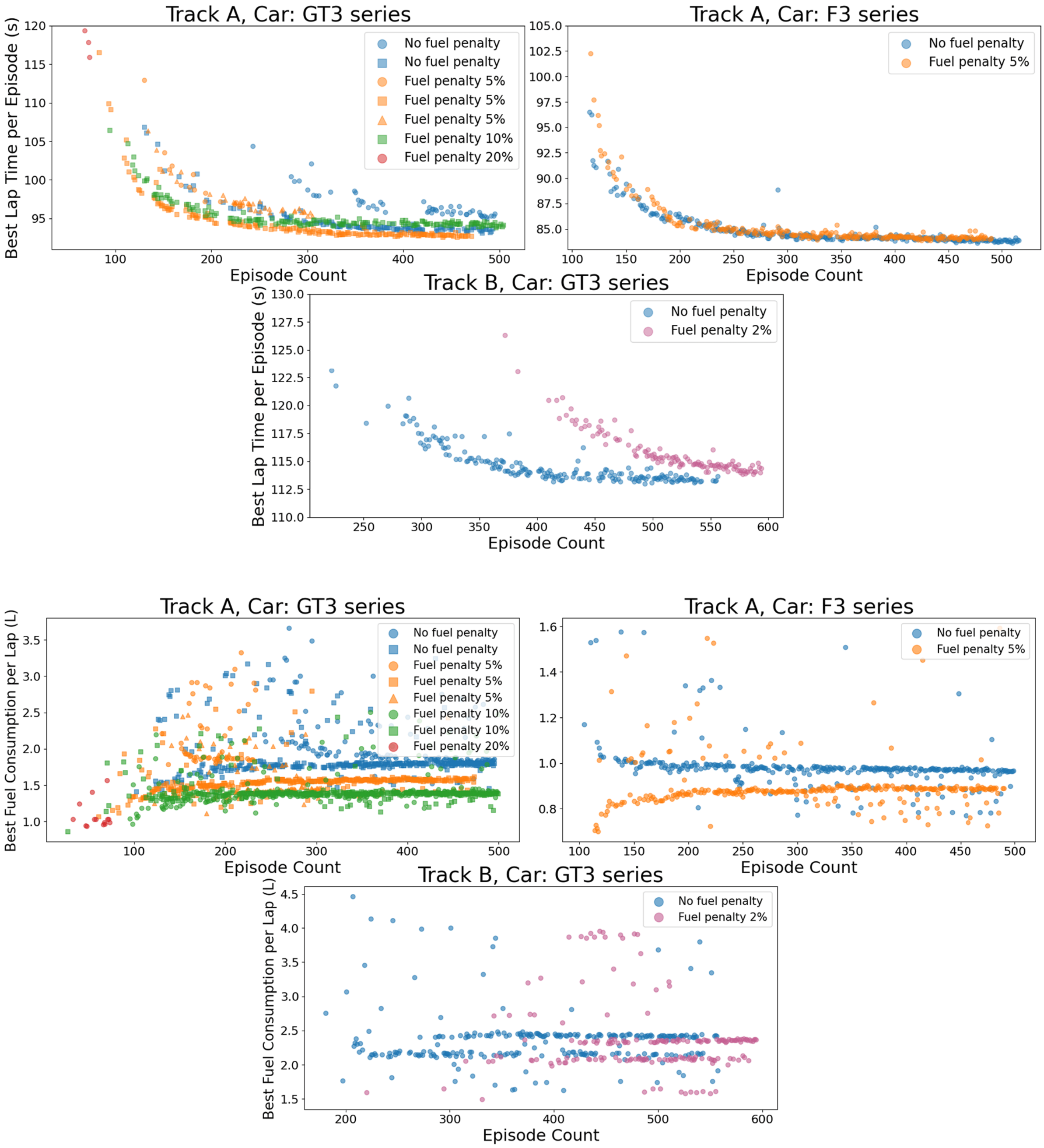

Figure 3. Top: Evolution of the best lap times per episode during training. Bottom: Best fuel consumption rates per lap. Different shapes represent different random seeds.

We compared the trained models against the baseline on Track A and Track B. On Track A, we tested models with 0% and 5% penalties. On Track B, models with 0% and 2% penalties were tested. Models were selected to have similar lap times for fair fuel efficiency comparison.

As per Table 2, the 5% penalty model on Track A used 0.19L less fuel per lap (10.7% savings) and was faster by 0.18 seconds (0.2%). On Track B, the 2% penalty model saved 0.064L (2.6%) but was slower by 0.84 seconds (0.7%). Training durations were similar between models.

Table 2. **Performance comparison of agents trained under different fuel penalties**

Fuel Penalty	**Best Lap Time (s) ↓**	**Mean Lap Time (s) ↓**	**FC for Best Lap (L) ↓**	**Mean FC Per Lap (L) ↓**	**Training Episodes**
Track A, Car: GT3 series					
0%	92.964	92.975	1.842	1.841	768
5%	**92.784**	**92.796**	**1.646**	**1.644**	474
Track B, Car: GT3 series					
0%	**113.198**	**113.206**	2.431	2.431	560
2%	114.04	114.203	**2.367**	**2.366**	595

Figure 4 shows spatial differences in fuel usage and lap time. Penalized models consumed less fuel, especially before major turns. On Track A, the penalty also led to faster lap times. On Track B, fuel savings came at the cost of slower lap times.

Figure 5 shows a comparison of driving behavior between the two models. On Track A, penalized agents reduced acceleration, braking, steering amplitude, and engine RPM – all factors responsible for the increased fuel consumption, directly or indirectly. On Track B, adaptations were weaker due to the lower penalty level: acceleration and RPM decreased, braking remained similar, and steering angle amplitude increased.

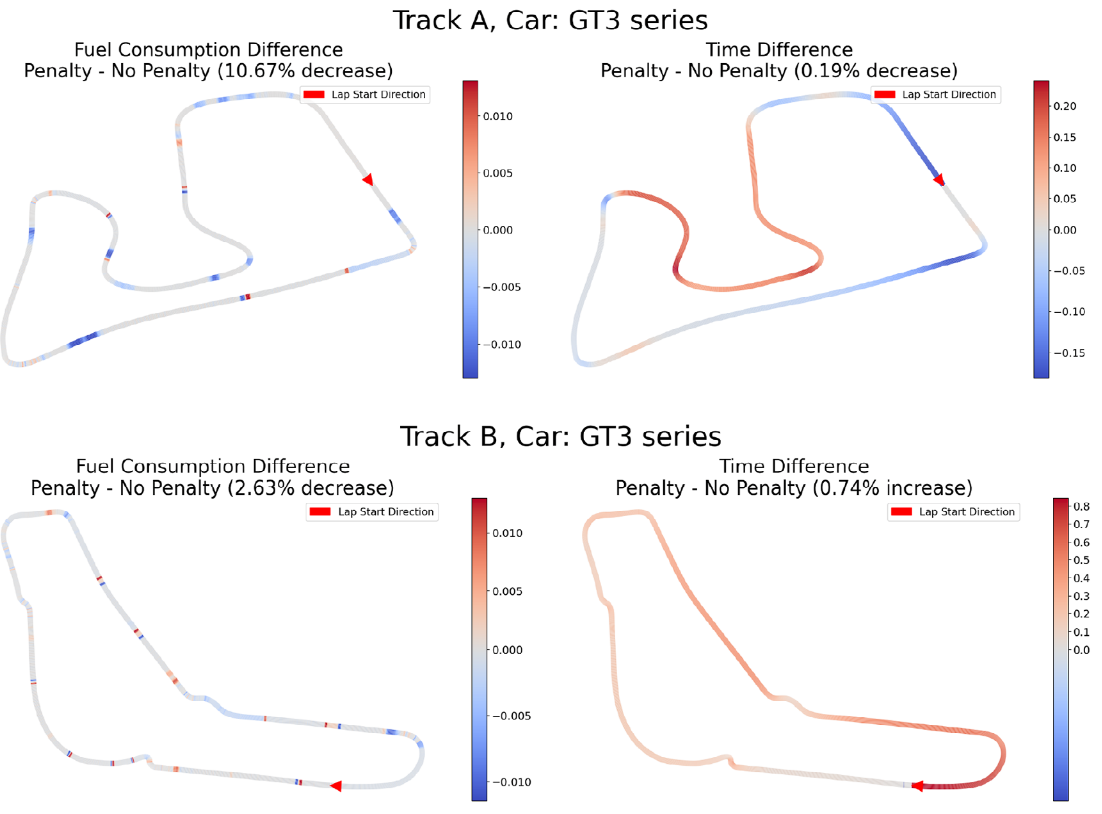

Figure 4. Spatial differences in fuel consumption and lap times between the two models on Tracks A and B. Percentage changes are computed as the difference in mean values between penalized and baseline laps. Blue regions indicate a decrease in the measured metric compared to the baseline, while red regions indicate an increase.

Results

Our experiments show that adding a fuel consumption penalty to the reward function leads to more fuel-efficient policies without significantly reducing lap time. On easier tracks, mild penalties (≤5%) often improved both fuel efficiency and speed, with faster early convergence during training. However, penalties above 10% led agents to prioritize fuel savings at the cost of increased lap time.

On the more challenging Track B, even a 2% penalty impaired performance, and higher penalties prevented agents from completing valid laps, confirming that penalty effectiveness depends on both track difficulty and the magnitude of the penalty.

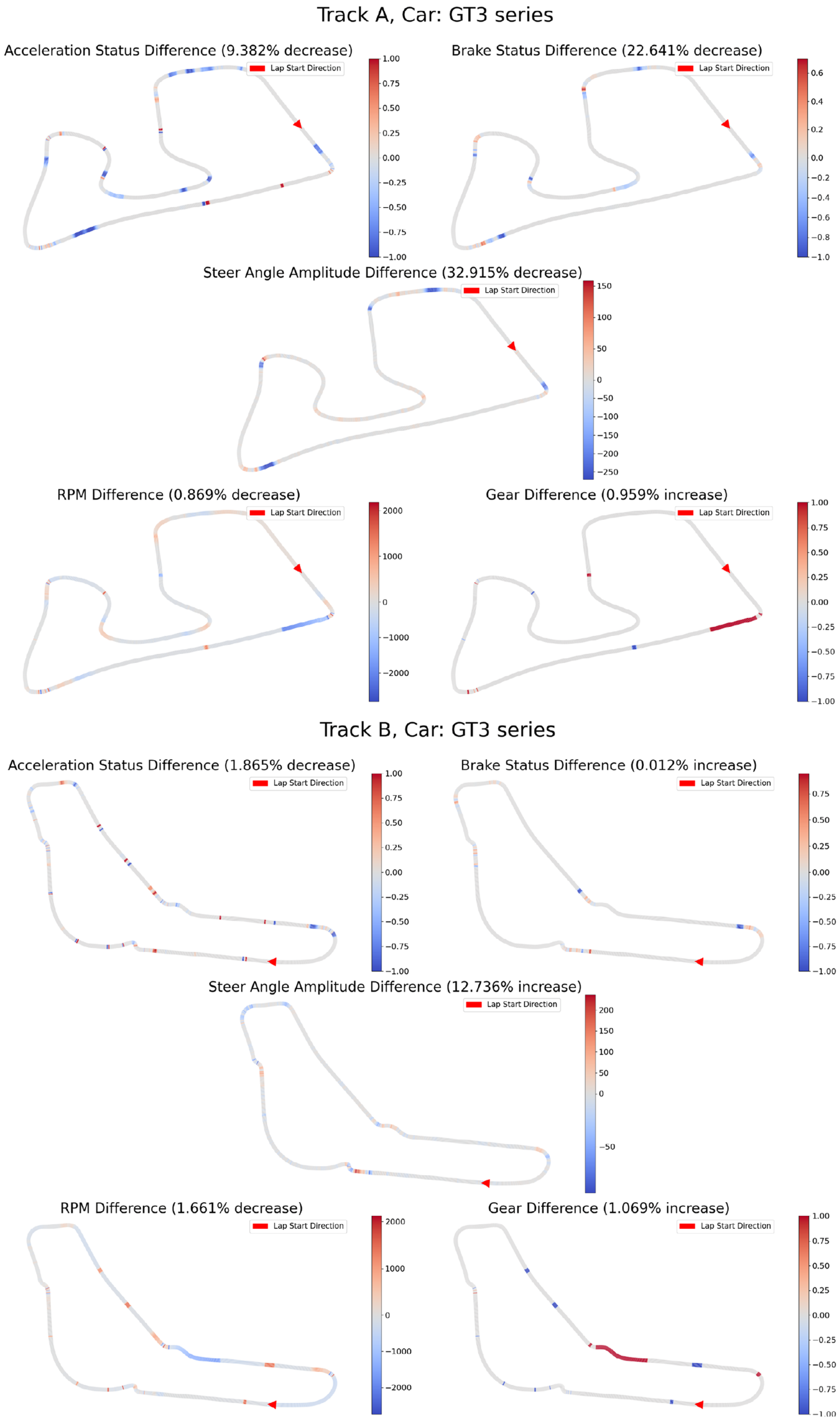

Figure 5. Spatial differences in driving behavior between the baseline and penalized agents

The agent learned key driving strategies to reduce fuel usage, mimicking real-world energy-saving strategies used in motorsport, such as maintaining momentum and staying in higher gears. Interestingly, it avoided early braking, likely suggesting that the agent prioritized maintaining speed over extra fuel savings.

Overall, the SAC algorithm effectively adapted to multi-objective rewards and learned fuel-efficient driving strategies under realistic constraints.

Discussions

Our work highlights the ability of RL agents to adopt fuel-efficient behaviors given appropriate reward shaping. However, several limitations remain that could be explored in future work.

First, experiments were limited to two vehicles and two tracks, raising concerns about generalizability. This is particularly relevant since strategies learned on the simpler Track A did not entirely transfer to Track B.

Second, the agent lacked manual gear-shifting control and could only influence gears indirectly via speed. This restricted fuel-saving techniques like short-shifting.

Third, the Assetto Corsa simulator runs only in real time, which considerably slows down training (~48 hours per 500 episodes). This limited our ability to test alternative reward designs or repeat training with different random seeds to ensure stability.

Notably, only one reward function design was explored. Future work could explore alternative formulations, like penalizing fuel-related factors directly or adding a short history of past fuel usage to the state space.

Finally, fuel usage was excluded from the observation space of the agent to simulate partial observability. While it did not prevent the agent from learning efficient strategies, future work could compare outcomes with and without partial observability.

It would also be valuable to test these methods with other RL algorithms beyond SAC, or with classical control frameworks such as MPC, LQR, or PID.

Conclusions

Our work explores the ability of RL agents in autonomous racing environments to adapt to multi-objective tasks that optimize both lap times and energy efficiency. We incorporated a fuel usage penalty into the reward function and demonstrated that low to moderate penalties lead to considerable fuel savings with minimal lap-time performance loss. The agents adapted by modifying their driving behaviors — reducing acceleration, managing engine RPM through gear changes, and increasing steering smoothness. However, these effects did not transfer to more complex tracks, where even small penalties impaired learning, highlighting the need for environment-specific penalty calibration.

Overall, our results suggest that RL can be effectively used to balance performance and energy efficiency. Future work could focus on generalizing these strategies across additional tracks and vehicles, and explore alternative reward function designs or observational conditions.

Acknowledgments

The Assetto Corsa simulation environment was used solely for the purposes of this research. We do not intend to use it in any public-facing or commercial context, nor in any activity involving public distribution or display.

The simulator is publicly available on the Steam platform. However, to use Assetto Corsa for research purposes, it is required to obtain the appropriate permissions from the Assetto Corsa Support Team.

Список літератури

1. Betz J. Autonomous Vehicles on the Edge: A Survey on Autonomous Vehicle Racing [Electronic resource] / J. Betz IEEE // Open Journal of Intelligent Transportation Systems. — 2022. — Vol. 3. — Pp. 458—488. — Mode of access: https://doi.org/10.1109/ojits.2022.3181510.
2. Brockman G. OpenAI Gym [Electronic resource] / G. Brockman et al. // ArXiv — 2016. — Mode of access: https://doi.org/10.48550/arXiv.1606.01540.
3. Elallid B. A Comprehensive Survey on the Application of Deep and Reinforcement Learning Approaches in Autonomous Driving [Electronic resource] / B. Elallid et al. // Journal of King Saud University — Computer and Information Sciences. — 2022. — Vol. 34, no. 9. — Pp. 7366–7390. — Mode of access: https://doi.org/10.1016/j.jksuci.2022.03.013.
4. Harrnoja T. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor [Electronic resource] / T. Haarnoja et al. // ArXiv — 2018. — Mode of access: https://doi.org/10.48550/arXiv.1801.01290.

5. Hu X. How Simulation Helps Autonomous Driving: A survey of Sim2real, Digital Twins, and Parallel Intelligence [Electronic resource] / X. Hu et al. // IEEE Transactions on Intelligent Vehicles. — 2024. — Vol. 9, no. 1. — Pp. 593–612. — Mode of access: https://doi.org/10.1109/TIV.2023.3312777.
6. Kim H. Autonomous Vehicle Fuel Economy Optimization with Deep Reinforcement Learning [Electronic resource] / H. Kim et al. // Electronics. — 2020. — Vol. 9, no. 11. — 3. 1911. — Mode of access: https://doi.org/10.3390/electronics9111911.
7. Remonda A. A Simulation Benchmark for Autonomous Racingwith Large-Scale Human Data [Electronic resource] / A. Remonda et al. // ArXiv. — 2024. — Mode of access: https://doi.org/10.48550/arXiv.2407.1668.
8. Revell J. Sim2real: Issues in transferring autonomous driving model from simulation to real world [Electronic resource] / J. Revell, D. Welch, J. Hereford // SoutheastCon 2022. — 2022. — Pp. 296–301. — Mode of access: https://doi.org/10.1109/SoutheastCon48659.2022.9764110.
9. Sutton R. Reinforcement Learning: An Introduction / R. S. Sutton, A. G. Barto // MIT Press, London, 1998. — Mode of access: https://books.google.ca/books?id=CAFR6IBF4xYC.
10. Song Y. Autonomous Overtaking in Gran Turismo Sport Using Curriculum Reinforcement Learning [Electronic resource] / Y. Song et al. // IEEE International Conference on Robotics and Automation (ICRA). — 2021. — Pp. 9403–9409. — Mode of access: https://doi.org/ 10.1109/ICRA48506.2021.9561049.
11. Tang X. Highway Decision-Making and Motion Planning for Autonomous Driving via Soft Actor-Critic [Electronic resource] / X. Tang et al. // IEEE Transactions on Vehicular Technology. — 2022. — Vol. 71, no. 5. — Pp. 4706–4717. – Mode of access: https://doi.org/10.1109/TVT.2022.3151651.

References

Betz, J., Zheng, H., Liniger, A., Rosolia, U., Karle, P., Behl, M., Krovi, V., & Mangharam, R. (2022). Autonomous Vehicles on the Edge: A Survey on Autonomous Vehicle Racing. *IEEE Open Journal of Intelligent Transportation Systems*, *3*, 458–488. https://doi.org/10.1109/OJITS.2022.3181510.

Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., & Zaremba, W. (2016). *OpenAI Gym* [Preprint]. arXiv. https://doi.org/10.48550/ARXIV.1606.01540.

Elallid, B. B., Benamar, N., Hafid, A. S., Rachidi, T., & Mrani, N. (2022). A Comprehensive Survey on the Application of Deep and Reinforcement Learning Approaches in Autonomous Driving. *Journal of King Saud University - Computer and Information Sciences*, *34* (9), 7366–7390. https://doi.org/10.1016/j.jksuci.2022.03.013.

Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018). *Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor* (Version 2) [Preprint]. arXiv. https://doi.org/10.48550/ARXIV.1801.01290.

Hu, X., Li, S., Huang, T., Tang, B., Huai, R., & Chen, L. (2024). How Simulation Helps Autonomous Driving: A Survey of Sim2real, Digital Twins, and Parallel Intelligence. *IEEE Transactions on Intelligent Vehicles*, *9* (1), 593–612. https://doi.org/10.1109/TIV.2023.3312777.

Kim, H., Pyeon, H., Park, J. S., Hwang, J. Y., & Lim, S. (2020). Autonomous Vehicle Fuel Economy Optimization with Deep Reinforcement Learning. *Electronics*, *9* (11), 1911. https://doi.org/10.3390/electronics9111911.

Remonda, A., Hansen, N., Raji, A., Musiu, N., Bertogna, M., Veas, E., & Wang, X. (2024). *A Simulation Benchmark for Autonomous Racing with Large-Scale Human Data* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2407.16680.

Revell, J., Welch, D., & Hereford, J. (2022). Sim2real: Issues in transferring autonomous driving model from simulation to real world. *SoutheastCon 2022*, 296–301. https://doi.org/10.1109/SoutheastCon48659.2022.9764110.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT Press. https://books.google.ca/books?id=CAFR6IBF4xYC.

Song, Y., Lin, H., Kaufmann, E., Durr, P., & Scaramuzza, D. (2021). Autonomous Overtaking in Gran Turismo Sport Using Curriculum Reinforcement Learning. *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 9403–9409. https://doi.org/10.1109/ICRA48506.2021.9561049.

Tang, X., Huang, B., Liu, T., & Lin, X. (2022). Highway Decision-Making and Motion Planning for Autonomous Driving via Soft Actor-Critic. *IEEE Transactions on Vehicular Technology*, *71* (5), 4706–4717. https://doi.org/10.1109/TVT.2022.3151651.

Беймук В. О., Кузьменко Д. О.

ЗБЕРЕЖЕННЯ ЕНЕРГІЇ ДЛЯ АВТОНОМНИХ АГЕНТІВ ІЗ ВИКОРИСТАННЯМ НАВЧАННЯ З ПІДКРІПЛЕННЯМ

Метою роботи є дослідження можливостей алгоритмів навчання з підкріпленням для формування стратегій автономного водіння з урахуванням компромісу між енергоефективністю та швидкістю.

Робота реалізована з використанням алгоритму Soft Actor-Critic у середовищі Assetto Corsa шляхом додавання штрафу за витрату пального у функцію винагороди. Досліджено вплив різних рівнів штрафу на витрати пального та швидкість руху. Також проаналізовано ключові поведінкові зміни, зокрема прискорення, оберти двигуна, передачі та амплітуди кермового кута.

Ключові слова: автономне водіння, навчання з підкріпленням, компроміс швидкість-ефективність, енергоефективність, симуляція перегонів.

Матеріал надійшов 24.06.2025