Data-Driven Simulation for Training High-Performance Polices

Grzegorz Czechmanowski Jan Wegrzynowski Piotr Kicki Krzysztof Walas
Poznań University of Technology, Poznań, Poland
name.surname@put.poznan.pl

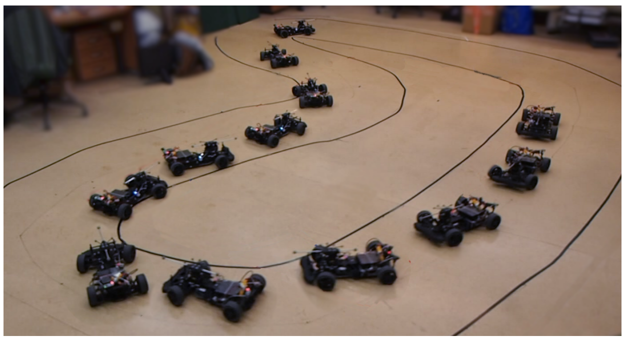

Figure 1. Our racing policy, trained entirely in simulation, deployed zero-shot on an F1TENTH car in the real world.

Abstract

Introduction. Autonomous racing is challenging control task — minimizing lap time requires operating at the limits of vehicle handling while strictly avoiding crashes. This makes simulation a compelling setting for learning racing policies, where high-risk exploration is safe and scalable, provided the simulator is sufficiently accurate to enable transfer to real-world. We therefore investigate the critical components of a real→sim→real pipeline: learning dynamics models from real-world data, training the policy in simulation, and transferring it zero-shot to a real F1TENTH car. We demonstrate the effectiveness of this pipeline by outperforming both a nonlinear Model Predictive Control (MPC) baseline and an expert human driver.

1. Differentiable vehicle simulator. We build a lightweight yet accurate, differentiable vehicle model by coupling a single-track (bicycle) dynamics model with Pacejka MF6.1 tyre forces [1] and first-order actuator dynamics. System identification is formulated as a multi-step prediction problem and solved via gradient-based optimization (backpropagation through time) on 36 min of motion-capture data.

2. Learning transferable policies. The multilayer perceptron policy is trained with PPO [2] for 120M steps in 400 parallel environments. We apply episode-reset domain randomization to bridge the sim–to–real gap. An ablation

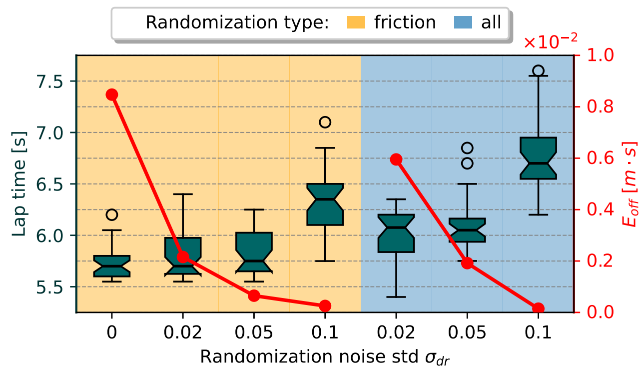

Figure 2. Lap times and E_{off} across randomization settings (friction-only vs. all single-track parameters and friction). Evaluated on an F1TENTH car over 30 laps per configuration.

over randomization magnitude and type shows a trade-off: stronger randomization and more randomized parameters lengthen lap times while reducing track-boundary violations E_{off} by inducing conservative driving (Fig. 2).

3. Additional design insights. (i) *Actuator modeling matters*: removing the first-order actuator model slows real laps by 0.7 s and increases boundary violations. (ii) *Action space matters*: commanding wheel-speed *accelerations* (rather than speeds setpoints) mitigates throttle spikes and enables reliable sim to real transfer of the policy.

4. Real-world evaluation. We evaluated the policy on a 17m L-shaped indoor track (Fig. 1), comparing against a state-of-the-art MPC [3] and an expert human driver. The RL policy achieved the fastest single-lap time, 5.561s, and the fastest time over 20-laps 115.36s. It surpassed MPC and the human by 0.078s and 0.286s on the fastest lap, and by 0.72s and 3.42s over the 20-laps. To the best of our knowledge, this is the first time that the RL policy has exceeded expert human performance in RC car racing.

Conclusion. The study showcases how a *learnable, differentiable simulator*, paired with targeted randomization, can close the Real to Sim to Real loop for dynamic control tasks.

This work was accepted for IEEE/RSJ IROS 2025
https://grzegorzczput.github.io/rl-racing/

Acknowledgments

This research was partially funded by PUT internal grant 0214/SBAD/0252. Work of Piotr Kicki was supported by the Foundation for Polish Science (FNP).

References

[1] A. J.C. Schmeitz I. J.M. Besselink and H. B. Pacejka. An improved Magic Formula/Swift tyre model that can handle inflation pressure changes. *Vehicle System Dynamics*, 48(sup1): 337–352, 2010. 1

[2] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal Policy Optimization Algorithms, 2017. arXiv:1707.06347 [cs]. 1

[3] Sirish Srinivasan, Sebastian Nicolas Giles, and Alexander Liniger. A Holistic Motion Planning and Control Solution to Challenge a Professional Racecar Driver. *IEEE Robotics and Automation Letters*, 6(4):7854–7860, 2021. 1