EasyChair Preprint

№ 13294

Evaluating the Sim-to-Real Transferability of End-to-End Control Policies for Autonomous Vehicles Operating on Deformable Terrains

Huzaifa Unjhawala, Zhenhao Zhou, Ishaan Mahajan,
Harry Zhang, Alexis Ruiz, Radu Serban and Dan Negrut

May 15, 2024

Evaluating the Sim-to-Real Transferability of End-to-End Control Policies for Autonomous Vehicles Operating on Deformable Terrains

Huzaifa Unjhawala, Zhenhao Zhou, Ishaan Mahajan, Harry Zhang, Alexis Ruiz, Radu Serban, Dan Negrut

Mechanical Engineering
University of Wisconsin-Madison
Madison, United States
{unjhawala,zzhou292,imahajan,hzhang699,aruiz26,serban,negrut}@wisc.edu

Abstract

We report on the transfer of an end-to-end control policy synthesized in simulation to a real-world setting. The policy guides a 1/6th scale vehicle, named *ART-B*, to a target location while navigating around obstacles with the aid of a 2D Lidar and GPS sensor. We utilize Gym-Chrono [1], a Reinforcement Learning (RL) environment based on the Project Chrono simulator, and the Open AI Gymnasium framework to synthesize this control policy trained using the Proximal Policy Optimization (PPO) algorithm. The approach involves training three versions of the policy: one for guiding *ART-B* across flat-rigid terrain, another for hilly-rigid terrain, and a third for hilly-deformable terrain. Subsequently, each policy will be tested in a real-world scenario with deformable terrain to answer the underlying research question – Does training an end-to-end control policy in a simulated setting with deformable terrain enhance its effectiveness in real-world applications?

1 Introduction

Using simulation in robotics is attractive due to its cost-effectiveness, safety, and expediency. This is particularly relevant when using Reinforcement Learning (RL), which demands millions of robot-environment interactions, typically feasible only in simulation. However, bridging the sim-to-real gap remains a challenge, especially in unstructured or off-road terrains [2]. Recent studies have demonstrated successful policy transfers from simulation to reality for legged quadruped robots in challenging terrains using privileged learning and adaptive terrain curricula [3]. Our study aims to extend this research to wheeled robots operating on deformable terrain. Specifically, gauging the importance of accounting for terrain deformation in the sim-to-real transfer is a main motivation of this work.

2 Method and Preliminary Results

Training Environment: The training environment is a 60×60 m patch of terrain on which obstacles are randomly placed. The vehicle's initial position is picked randomly in a 30 m diameter circle; the goal is placed on the opposite side of the same circle. In polar coordinates, given α the angle of the vehicle initial position, the angle of the goal will be $\alpha + \beta$, with β randomly picked in $[\frac{\pi}{2}, \frac{3\pi}{2}]$. *ART-B*, which hosts a 2D LiDAR and a GPS simulated using Chrono::Sensor, is tasked with navigating to within 5 m of the goal using this sensor suite. The terrain varies: it can be rigid-flat; rigid-hilly with a height map generated using smooth Perlin Noise; or deformable-hilly, modeled using the Soil Contact Model (SCM) [4]. Figure 1 shows *ART-B* performing obstacle avoidance on hilly-deformable terrain.

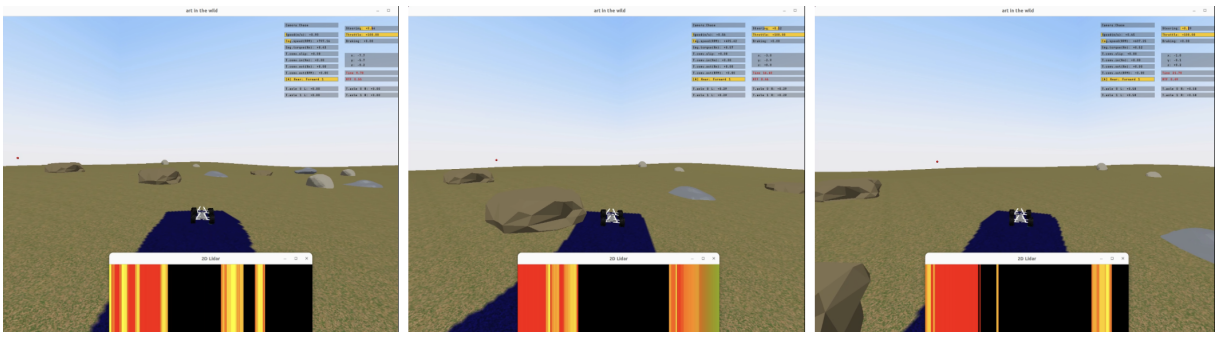

Figure 1: ART-B performing an obstacle avoidance maneuver on hilly-deformable terrain.

RL Algorithm: We use Proximal Policy Optimization (PPO), an on-policy RL method using separate actor and critic Neural Networks (NNs) [5]. A LiDAR and GPS preprocessing NN is also trained alongside the Actor and Critic networks. Inputs include LiDAR depth (clipped to $[0, 30]$ and of size 180×1)

and a $\mathbb{R}^{4\times1}$ vector of the vehicle's relative position, heading, and speed. The NN outputs a vector $\in \mathbb{R}^{2\times1}$ which represents the normalized throttle and steering commands. Rewards are given for goal-directed movement, with penalties for collisions, boundary breaches, and failing to reach the goal in 40 seconds.

Training Method: We use Curriculum Learning while progressively increasing environmental difficulty. Training starts on rigid-flat terrain with obstacles following a normal distribution $\mathcal{N}(3,3)$, capped at positive values. After every 10 NN updates (16 simulations each), the model is assessed for success rate; i.e., percentage of simulations where the vehicle reaches the goal. Surpassing a 70% success rate triggers an increase of 3 in the mean obstacle count, capping at a mean of 15. Post 200 updates on flat terrain, we save the model, switch to hilly terrain, and continue for another 200 updates. The process is repeated for deformable terrain.

Preliminary Findings: Figure 2 displays the moving averages of the success, failure, and time-out rates for each policy, calculated every 16 simulations relative to the number of neural network (NN) updates. Blue vertical dashed lines indicate the updates where the number of obstacles was raised. Upon increasing the mean of the number of obstacles to 15, there is a decrease in the success rate, but it recovers and stabilizes around 75%. When the terrain changes from rigid-flat to rigid-hilly, represented by the green dashed line, there is a significant decline in performance with the flat terrain policy; nevertheless, the success rate rebounds to approximately 75% following around 200 updates. Training commences on deformable terrain at the first yellow dashed line, initially restricting the max-min height to 0.5 m. Following this, at the second yellow line, the terrain's max-min height is elevated to 1 m. It is worth noting that the policy continues to perform well, preserving a success rate of roughly 75%. Yet, as soon as the terrain's max-min height is increased to 1.5 m, the policy's efficiency drops, failing to recover until the completion of the training process. An interesting observation is the surge in time-out rate, potentially implying that the vehicle finds it challenging to navigate steep terrains within the given time due to reduced speed caused by wheel slippage. An in-depth report of these findings is reserved for the final presentation. Additionally, we intend to validate these observations via actual-world testing.

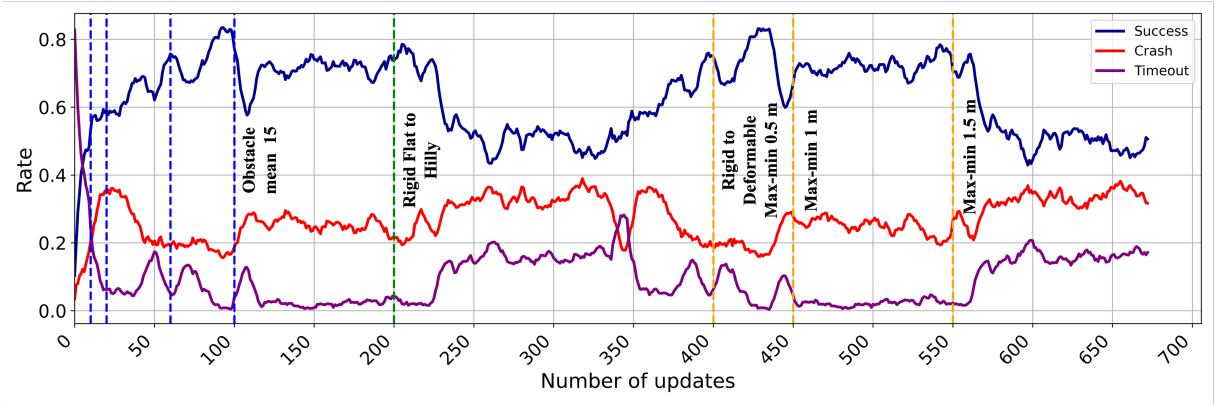

Figure 2: Moving averages of the policy success, failure, and time-out rates. Blue dashed lines indicate the updates where the number of obstacles was raised. The green dashed line represents the transition from rigid-flat to rigid-hilly terrain with terrain height varying between -0.5 and 0.5 m. The yellow dashed lines indicate the transition from rigid-hilly to deformable terrain with varying terrain heights.

References

[1] S. Benatti, A. Young, A. Elmquist, J. Taves, R. Serban, D. Mangoni, A. Tasora, and D. Negrut, "Pychrono and gym-chrono: A deep reinforcement learning framework leveraging multibody dynamics to control autonomous vehicles and robots," in *Advances in Nonlinear Dynamics*. Springer, 2022, pp. 573–584.

[2] W. Zhao, J. P. Queralta, and T. Westerlund, "Sim-to-real transfer in deep reinforcement learning for robotics: a survey," in *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2020, pp. 737–744.

[3] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning quadrupedal locomotion over challenging terrain," *Science Robotics*, vol. 5, no. 47, p. eabc5986, 2020. [Online]. Available: https://www.science.org/doi/abs/10.1126/scirobotics.abc5986

[4] R. Serban, J. Taves, and Z. Zhou, "Real time simulation of ground vehicles on deformable terrain," in *Proceedings of ASME IDETC/CIE 2022*, St. Louis, MO, 2022.

[5] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *CoRR*, vol. abs/1707.06347, 2017. [Online]. Available: http://arxiv.org/abs/1707.06347