

Sim-to-Real via Real-to-Sim: Improving real-world performance through real-world learning

PartIII project proposal

Ziyi Liu (zl413), Christ's College

Project Supervisor: Dr. Amanda Prorok, Jan Blumenkamp

Abstract

Data-driven approaches for multi-robot control problems have proven very promising with a variety of model architectures, including deep reinforcement learning, GNN, and Gaussian Processes. So far, simulation data has been the main sampling source, where typically a large amount of training data is required. The model trained from the simulation, when deployed onto the real world (sim-to-real transfer), can lead to a significant reality gap due to the limitations of simulation assumptions, real-world noises, and robot-robot interactions. Alternatively, we propose that a direct mapping of current states and action policies to next states can be learnt as the simulator, such that the real action policies for different tasks can be learnt with RL methods. We also propose a general framework for such real-world data-driven models on single- and multi-robot systems with an active and efficient data sampling mechanism.

1 Introduction, approach and outcomes

In a wide range of robotics tasks, gathering data and training policies and then transferring them into the real world(sim-to-real) has been a standard approach, and the reality gap [1] - the performance disparity between simulations and real world - as a result of this transfer has been dealt with many approaches, such as domain adaptation and domain randomization. However, instead of relying on simplified simulation assumptions or devising extremely complicated analytical simulations for training action policies, we can directly train an end-to-end neural network that maps the current state to control parameters in deployment. This data-driven approach has shown great potential in many tasks for multi-robot systems, but each trained model is isolated and only specific to one task with poor generalizability.

Another two main considerations for the data-driven approach are the large size of data required and the sensitivity of performance to the quality and coverage of the data we feed into it. One intuitive method would be collecting samples that have not been covered by our training data so far independent of specific tasks. However, it is not pragmatic to sample the whole action space. An alternative view is to leverage active learning methods for continuous data sampling that minimizes model uncertainty via active information acquisition [2, 3] by strategically picking the most informative samples. Therefore, one key objective of this project is to explore the adaptive sampling mechanism for better and more efficient data collection, potentially leading to a data collection library as a deliverable.

The main objective of the project is to learn a model that maps the current state and desired action to the next state with neural networks [4], which serves as the simulator(real-to-sim transfer); and this trained model will then be used for learning the actual action policy. This approach follows the same intuition from VR-Goggles real-to-sim transfer [5], and we expect much better action generations for the sim-to-real deployment. In the multi-robot case, GNN-based methods will also be considered for modelling robot-robot interactions [6]. We will first consider a single-robot system (RoboMaster)

with only robot-environment interactions, and then the problem can be extended to either a multi-RoboMaster or a single-drone system. Ideally, the trained model will be task-agnostic, such that specific tasks - such as static or dynamic targets, evader pursuit problems [7], narrow passage problem for a robot formation [6], and collision avoidance - would use a subset of dynamics that have been captured by the learnt model. There are also a few simulation environments to choose from (such as Pybullet [8] or Gazebo) for different tasks as a baseline. The final outcome for the real-to-sim adaptation approach will be compared to the simple simulation results on different tasks and systems as another success criterion.

There are four potential extensions to this project following the core problem described above: (1) Generalizing the adaptation techniques to multiple drones that accounts for significant robot-robot interactions (such as down-wash effects), (2) constructing a ready-to-use framework for different environments, formations, and tasks following the essential data collection procedures and experiments, (3) embedding physical knowledge priors into a Gaussian Process regression and comparing the hybrid modelling with our main framework [9] and (4) exploration of safe reinforcement learning approach, which concerns maximizing total reward within the safety constraints [10].

2 Workplan

An overview of the work plan broken into 2-week blocks is shown in the table on the next page. Some tasks divided into different chunks are not necessarily isolated, as they may take place simultaneously for different parts of the experiments.

The first two blocks are dedicated to literature review and early environment setups, as well as a full mathematical formulation of the core tasks. Before the start of Lent term, data collection pipelines and early results for different data collection schemes should be completed, which will then serve as the basis for all of the following real-world data-based experiments. The main focus for weeks 7-10 will be on single-robot systems training and evaluation, such that by the end of week 10, a preliminary set of full experiment results and benchmarking should be delivered.

The next 6 weeks will be primarily on the extension to multi-robot systems and adapting our current data collection and training framework to multi-agent problems, and by the end of week 16, full experiment results, analysis, and the expanded framework will be concluded.

More extensions mentioned in Introduction section will be explored in the next two chunks before the final dissertation writing in Easter, where some additional techniques and ideas will be tested, and one will be explored more in-depth depending on the preceding results. Note that one chunk is left blank as required in case of emergencies, but if things go as planned this will be used for extension exploration as well.

Weeks	Tasks	Notes
1-2	Literature Review, Code base Review	
3-4	Familiarize with Hardware, tools, and single-robot data collection plans	Formal work formulation
5-6	Set up data collection pipelines	Milestone: construction of a framework for data collection with different methods (random, hand-picked, active acquisition)
7-8	Single-robot experiment setup and training	
9-10	Simulation training and deployment for benchmarking	Early experiment results
11-12	Extension to multi-robot scenarios and new experiment designs	
13-14	Multi-robot task formulations, setup, and model designs(with GNN and DRL methods)	
15-16	Multi-robot training and evaluation	Experiment results and extended framework produced
17-18	Further extensions (Framework building, multi-drone system)	
19-20	Further extensions (safe RL considerations and physical priors)	Only one or two of the mentioned extensions will be explored in depth
21-22	Buffer weeks	In case of contingencies
23-24	Dissertation writing	
25-26	Dissertation writing and proofreading	

References

- [1] Nick Jakobi, Phil Husbands, and Inman Harvey. Noise and the reality gap: The use of simulation in evolutionary robotics. In Federico Morán, Alvaro Moreno, Juan Julián Merelo, and Pablo Chacón, editors, *Advances in Artificial Life*, pages 704–720, Berlin, Heidelberg, 1995. Springer Berlin Heidelberg.
- [2] Yiannis Kantaros and George J. Pappas. Scalable active information acquisition for multi-robot systems. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 7987–7993, 2021.
- [3] Gautam Salhotra, Christopher E. Denniston, David A. Caron, and Gaurav S. Sukhatme. Adaptive sampling using pomdps with domain-specific considerations. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2385–2391, 2021.
- [4] Anusha Nagabandi, Gregory Kahn, Ronald S. Fearing, and Sergey Levine. Neural network dynamics for model-based deep reinforcement learning with model-free fine-tuning, 2017.
- [5] Jingwei Zhang, Lei Tai, Peng Yun, Yufeng Xiong, Ming Liu, Joschka Boedecker, and Wolfram Burgard. Vr-goggles for robots: Real-to-sim domain adaptation for visual control, 2019.
- [6] Jan Blumenkamp, Steven Morad, Jennifer Gielis, Qingbiao Li, and Amanda Prorok. A framework for real-world multi-robot systems running decentralized gnn-based policies, 2021.
- [7] Sumeet Batra, Zhehui Huang, Aleksei Petrenko, Tushar Kumar, Artem Molchanov, and Gaurav Sukhatme. Decentralized control of quadrotor swarms with end-to-end deep reinforcement learning, 2021.
- [8] Jacopo Panerati, Hehui Zheng, SiQi Zhou, James Xu, Amanda Prorok, and Angela P. Schoellig. Learning to fly – a gym environment with pybullet physics for reinforcement learning of multi-agent quadcopter control, 2021.
- [9] Lucas Rath, Andreas René Geist, and Sebastian Trimpe. Using physics knowledge for learning rigid-body forward dynamics with gaussian process force priors. In *5th Annual Conference on Robot Learning*, 2021.
- [10] Javier García and Fernando Fernández. A comprehensive survey on safe reinforcement learning. *Journal of Machine Learning Research*, 16(1):1437–1480, 2015.