Learning Mobile Manipulation Skills via Autonomous Exploration

Russell Mendonca Deepak Pathak

Carnegie Mellon University

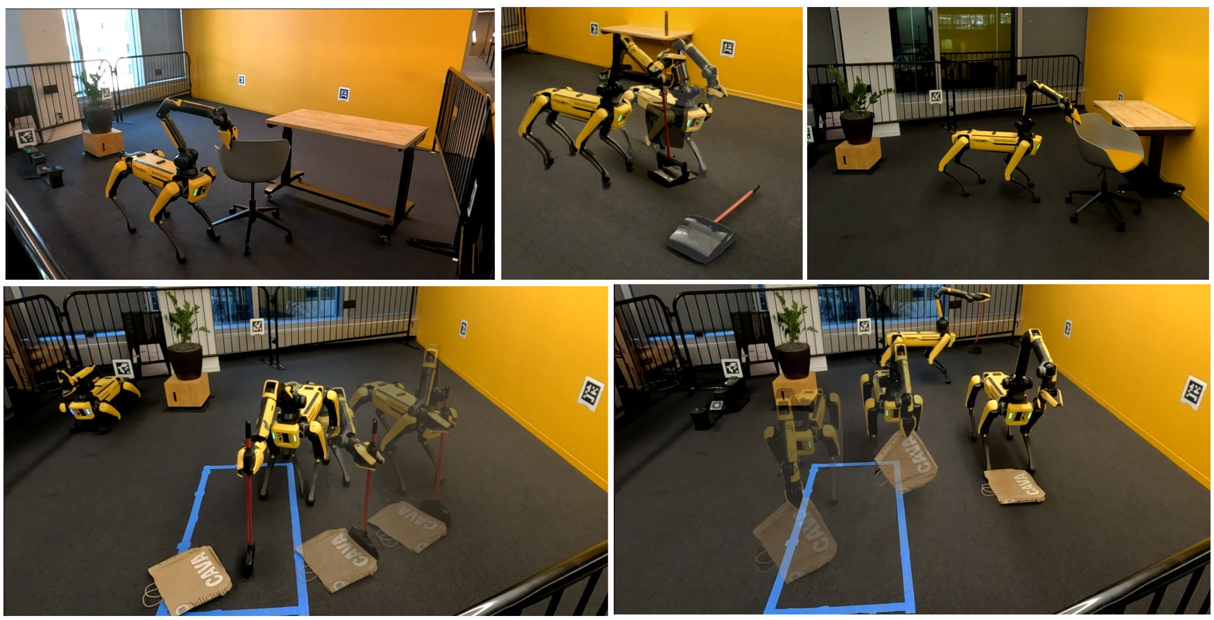

Figure 1. **Continual Autonomous Learning**: We enable a legged mobile manipulator to learn a variety of tasks such as moving chairs (top, left and right), righting a dustpan (top, middle), and sweeping (bottom) via practice in the real world with minimal human intervention.

Abstract

*To build generalist robots capable of executing a wide array of tasks across diverse environments, robots must be endowed with the ability to engage directly with the real world to acquire and refine skills without extensive instrumentation or human supervision. This work presents a fully autonomous real-world reinforcement learning framework for mobile manipulation that can both independently gather data and refine policies through accumulated experience in the real world. It has several key components: 1) automated data collection strategies by guiding the robot's exploration toward object interactions, 2) using goal cycles for real world RL such that the robot changes goals once it has made sufficient progress, where the different goals serve as resets for one another, 3) efficient control by leveraging basic task knowledge present in behavior priors in conjunction with policy learning and 4) formulating generic rewards that combine human-interpretable semantic information with low-level, fine-grained state information. We demonstrate our approach on Boston Dynamics Spot robots in continually improving performance on a set of four challenging mobile manipulation tasks and show that this enables competent policy learning, obtaining an average success rate of 80% across tasks, a **3-4**× improvement over existing approaches.*

1. Introduction

As robots transition from the structured confines of fully mapped industrial settings into the dynamic and unstructured realm of our daily lives, there is an increasing need to build generalist systems capable of executing a wide array of tasks across diverse environments. While visuomotor policies trained with reinforcement learning (RL) have demonstrated significant potential to bring robots into open-world environments[9–11], in practice, they first require training in simulation [1–3, 7, 15, 17]. However, it is challenging and not scalable to build simulations that capture the unbounded diversity of real-life tasks, especially involving complex manipulation. What if we instead adopt a strategy where learning occurs through direct engagement with the real world, without extensive environmental instrumentation or human supervision during the training process? We address multiple challenges for such a system.

Challenge 1: Automated collection of useful data: Consider a complex, high-dimensional system like a legged mobile manipulator operating in open spaces where undirected actions often do not affect any meaningful change in the environment. The first challenge in building an effective real-world learning system is in autonomous, task-relevant data collection because good robot autonomy does not imply the resulting data has a useful learning signal. For example, we would like to avoid the robot simply waving its arm in the air without interacting with objects if its goal is to acquire manipulation skills. While such a system could, in theory, learn sophisticated mobile manipulation strategies given enough data, we propose using off-the-shelf visual models to design automated strategies that make learning in the real world feasible by guiding the robot's exploration toward object interactions.

Challenge 2: How to ensure diverse practice? The second challenge is how to allow the robot to purposely practice achieving goals from diverse initial states without human resetting. Once the robot is close to its goal, it does not get to practice the task from states that are further from the goal. For instance, consider a robot tasked to move furniture. The robot may learn to move a piece of furniture to its target location; however, now that the furniture is very close to the goal, continuing to practice the task from this starting state will not yield further benefits. Instead, if the environment state could be reset back to the initial state distribution, the robot could practice repeating its success. In the absence of such resets, how can we enable autonomous robots to return to the harder initial state distribution for practicing tasks? The approach we use is to set up '*goal-cycles*' [5, 6, 8], where we switch the goal once the robot has made sufficient progress on the previous one, or spent a budget of a fixed interval of trajectories attempting it. Hence, the goals serve as resets for one another, and this multi-goal learning setup ensures that the robot does not stagnate in a limited region of the state space near any particular goal.

Challenge 3: Efficient control in the real world: Even with a favorable initial state distribution, policy learning poses a daunting challenge due to large observation and action spaces. This challenge is especially severe in the case of legged mobile manipulation, where the robot needs to move and simultaneously maintain contact with objects and retain control. Our approach expedites learning control policies by leveraging basic task knowledge present in behavior priors. These priors can take the form of planners with a simplified incomplete model or automated procedurally generated behaviors. It is important to note that while these priors bootstrap learning and help provide a signal for learning, particularly in the early stages, the priors might not be very competent at performing the task, owing to their simplicity. In our experiments, the average success rate of the prior is just 20% across tasks but as low as 5% for the challenging task of sweeping. In contrast, our learning-based approach enables an average success rate of 80%, a $4\times$ improvement. Hence, the priors are not a substitute for learning controllers but rather serve to structure exploration.

Challenge 4: Defining rewards in the real world: For the system to benefit from the previously described structure and get better at performing tasks, it must evaluate the relative benefit of different actions by receiving reward feedback from the environment. Providing reward supervision in the real world often requires physical instrumentation in the form of specialized sensors [13, 16] or needs humans in the loop [4, 12, 14]. Furthermore, the ability of these robots to keep collecting data and learning to improve is bottlenecked by how expensive or difficult it is to scale these approaches. In this work, we seek a *flexible* way for humans to specify objectives for *arbitrary* tasks. To this end, we devise a generic reward modeling recipe that combines human-interpretable, semantic information, i.e., text-based detection and segmentation models, along with low-level, fine-grained state information, i.e., vision and depth-based observations for object estimation. Despite yielding noisy estimates, we find the resulting reward is sufficient to allow the robot to learn challenging tasks.

The main contribution of this work is a general approach for continuously learning mobile manipulation skills directly in the real world with autonomous RL. The main components of our approach involve: (1) task-relevant autonomy for collecting data with useful learning signal, (2) efficient control by integrating priors with learning policies, and (3) flexible reward specification combining high-level visual-text semantics with low-level depth observations. Our approach enables a Boston Dynamics Spot robot to continually improve in performance on a set of 4 challenging mobile manipulation tasks, including moving a chair to a goal with the table in the corner or center of the playpen, picking up and vertically balancing a long-handled dustpan, and sweeping a paper bag to a target region. Our experiments show that our approach gets an average evaluation success rate of about 80% across tasks, which is a **$4\times$ improvement** over using either RL or the prior individually.

References

[1] Ilge Akkaya, Marcin Andrychowicz, Maciek Chociej, Mateusz Litwin, Bob McGrew, Arthur Petron, Alex Paino, Matthias Plappert, Glenn Powell, Raphael Ribas, et al. Solving Rubik's Cube with a Robot Hand. *arXiv preprint arXiv:1910.07113*, 2019. 2

[2] Xuxin Cheng, Kexin Shi, Ananye Agarwal, and Deepak Pathak. Extreme Parkour with Legged Robots. *arXiv preprint arXiv:2309.14341*, 2023.

[3] Mark Cutler, Thomas J Walsh, and Jonathan P How. Reinforcement Learning with Multi-Fidelity Simulators. In *ICRA*, pages 3888–3895. IEEE, 2014. 2

[4] Justin Fu, Avi Singh, Dibya Ghosh, Larry Yang, and Sergey Levine. Variational Inverse Control with Events: A General Framework for Data-Driven Reward Definition. In *NeurIPS*, 2018. 2

[5] Abhishek Gupta, Justin Yu, Tony Z Zhao, Vikash Kumar, Aaron Rovinsky, Kelvin Xu, Thomas Devlin, and Sergey Levine. Reset-Free Reinforcement Learning via Multi-Task Learning: Learning Dexterous Manipulation Behaviors without Human Intervention. In *ICRA*, pages 6664–6671. IEEE, 2021. 2

[6] Abhishek Gupta, Corey Lynch, Brandon Kinman, Garrett Peake, Sergey Levine, and Karol Hausman. Demonstration-Bootstrapped Autonomous Practicing via Multi-Task Reinforcement Learning. In *ICRA*, pages 5020–5026. IEEE, 2023. 2

[7] Tuomas Haarnoja, Ben Moran, Guy Lever, Sandy H Huang, Dhruva Tirumala, Markus Wulfmeier, Jan Humplik, Saran Tunyasuvunakool, Noah Y Siegel, Roland Hafner, et al. Learning Agile Soccer Skills for a Bipedal Robot with Deep Reinforcement Learning. *arXiv preprint arXiv:2304.13653*, 2023. 2

[8] W. Han, S. Levine, and P. Abbeel. Learning Compound Multi-Step Controllers under Unknown Dynamics. In *IROS*, 2015. 2

[9] Dmitry Kalashnikov, Alex Irpan, Peter Pastor, Julian Ibarz, Alexander Herzog, Eric Jang, Deirdre Quillen, Ethan Holly, Mrinal Kalakrishnan, Vincent Vanhoucke, et al. QT-Opt: Scalable Deep Reinforcement Learning for Vision-Based Robotic Manipulation. *arXiv preprint arXiv:1806.10293*, 2018. 2

[10] Dmitry Kalashnikov, Jacob Varley, Yevgen Chebotar, Benjamin Swanson, Rico Jonschkowski, Chelsea Finn, Sergey Levine, and Karol Hausman. MT-Opt: Continuous Multi-Task Robotic Reinforcement Learning at Scale. *arXiv preprint arXiv:2104.08212*, 2021.

[11] Sergey Levine, Chelsea Finn, Trevor Darrell, and Pieter Abbeel. End-to-End Training of Deep Visuomotor Policies. *JMLR*, 2016. 2

[12] Huihan Liu, Soroush Nasiriany, Lance Zhang, Zhiyao Bao, and Yuke Zhu. Robot Learning on the Job: Human-in-the-Loop Autonomy and Learning During Deployment. *arXiv preprint arXiv:2211.08416*, 2022. 2

[13] C. Schenck and D. Fox. Visual Closed-Loop Control for Pouring Liquids. In *ICRA*, 2017. 2

[14] Avi Singh, Larry Yang, Kristian Hartikainen, Chelsea Finn, and Sergey Levine. End-to-End Robotic Reinforcement Learning without Reward Engineering. In *RSS*, 2019. 2

[15] Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World. In *IROS*, pages 23–30. IEEE, 2017. 2

[16] A. Yahya, A. Li, M. Kalakrishnan, Y. Chebotar, and S. Levine. Collective Robot Reinforcement Learning with Distributed Asynchronous Guided Policy Search. In *IROS*, 2017. 2

[17] Ruihan Yang, Yejin Kim, Aniruddha Kembhavi, Xiaolong Wang, and Kiana Ehsani. Harmonic Mobile Manipulation. *arXiv preprint arXiv:2312.06639*, 2023. 2