Efficient Robot Learning from Diverse Data

Chang Shi

The University of Texas at Austin
Austin, TX, United States
changshi@utexas.edu

Abstract

The lack of large-scale clean data for learning has been a challenge that significantly hinders robots from developing superior level of autonomous intelligence. This urges the necessity to utilize diverse data in a more efficient way. This work approaches the challenge from four perspectives: efficient learning from expert demonstration, efficient dynamics modeling from in-the-wild videos, efficient learning from heuristics guidance, and adjustment for efficient deployment. We provide an overview of preliminary results in each area and outline proposed research on extracting controllable representation from data, aiming at efficient cross-embodiment learning, as well as learning from multi fidelity data.

Introduction

Robotics has long been suffering from a scarcity of large-scale, clean, and standardized datasets, especially when compared to fields like computer vision or natural language processing. Real-world robot data is expensive to collect, noisy due to hardware and environmental variability, and often fragmented across labs with inconsistent formats and task definitions. Despite the notable efforts include RoboNet (Dasari et al. 2019), BridgeData (Ebert et al. 2021), and more recent Open-X Embodiment dataset (O'Neill et al. 2024) that aggregates demonstrations from multiple robot platforms have begun addressing this gap, it is still beyond feasibility to build scalable and clean datasets for every robot embodiment and task domain. Therefore, it is worth viewing the challenge from an orthogonal perspective, and tackle through the lens of efficient learning from diverse data, that is to say, maximizing the utility of limited accessible data according to the characteristics of each data source, and well tuning the learned policy according to the deployment constraints. I believe that these areas are far from extensive examination.

In short, this work tries to address the pivotal question:

How can robot agents efficiently learn policies from diverse data and efficiently deploy those policies for real-time decision making?

With this motivation in mind, this work explore the following four dimensions:

1. **Efficient Learning from Expert Demonstration.** Expert demonstrations for robots usually require human teleopration, which is time-consuming. To maximize the utility of those precious data, this work will develop a learning scheme that imitates not only the robot trajectories from the data, but also all kinds of available priors such as skills, execution stages etc.

2. **Efficient Dynamics Modeling from In-the-Wild Videos.** While in-the-wild videos do not come with paired action labels as expert demonstrations, their large data magnitude provide abundant dynamics information. This work would like to focus on efficient dynamics modeling in complex scenes through factorization, with a goal of cross-embodiment generalization.

3. **Efficient Learning from Heuristics Guidance.** Unlike demonstrations with limited distribution, heuristics is a compact format of prior knowledge that can offer universal guidance in extensive state coverage, when demonstrations are not available. It equips the agent with capability to learn from imagined planning. This work investigate how planning using heuritics can guide the online policy learning towards better sample efficiency.

4. **Adjustment for Efficient Deployment.** Successful deployment of the learned policy on real robot systems while satisfying the constraints of real-time execution is the last but not least piece of complete robot learning. To avoid robot from waiting for policy inference, we design inference time acceleration method and introduce test-time prior injection.

Preliminary Results

In this section, we give an overview of some preliminary results along each dimension of the main question.

We introduce a transformer-based imitation learning model (Shi, Zheng, and Fey 2022), which predicts both robot trajectories and gestures of the teleoprators. We conducted experiments on an expert dataset of a bimanual surgical robot and show that the transformer-based architecture can better capture long-term dependencies, and thus predict both future motion of the robot and intent of the teleoprators more accurately. Further, we propose to use the learned prediction model over expert demonstrations to infer a real-time reference for robot manipulation, and offer a haptic guidance for the human teleoprators (Shi et al. 2023). Results show that the operation time to finish a task and the user perception of difficulty significantly decrease, with the assistance of our model. These two works explored novel ways to learn from expert demonstrations (Dimension 1).

To deal with noisy, complex, action-free video data, we introduce FLAM (Shi et al. 2025), a factored world model that infers latent actions through joint learning of inverse and forward dynamics models. Through factorization, we are able to capture the dynamics more precisely in multi-entity scenarios. Then the latent actions can be used to do robot policy learning, given a few-shot latent action decoder. The latent actions can also be used to generate more dynamics-aligned data which can be used for robot training. Results show this framework is well suitable for learning robot manipulation that involves interaction with multiple objects. This contribution addresses the key question of how to take advantage of in-the-wild videos for robot learning when true action labels are not available (Dimensions 2) .

In the conditions where even wild videos are scarse, heuritics can be a remedy data source to support robot learning for autonomy. We introduce TP-PPO (Shi and Zhang 2025b), which uses Monte Carlo tree-search based lookahead planning to guide PPO policy training. We conducted a study on 3D bin packing problem where no demonstrations but only heuristics for stacking stability are available. The results show that heuritics-guided learning is able to achieve much denser packing than learning from scratch. This work offers an insight about how to utilize heuritics source of data to assist efficient robot learning (Dimension 3) when no demonstration data accessible.

Considering the real-time decision making demand of robots, we introduce FastDP (Shi and Zhang 2025a), a low-latency diffusion policy method that achieves competitive robot task performance while significantly reducing inference delay through an efficient state space model backbone and test-time prior trajectory knowledge injection. Experimental findings indicate that adjustment for efficient deployment (Dimension 4) could significantly conductive to robotics tasks that real-time decision-making.

Proposed Research

Endogenous Representation Pretraining. While preliminary results show that FLAM (from dimension 2) can disentangle the factors for more accurate dynamics modeling, it does not yet have the capability to distinguish controllable factors from background noise factors, i.e., the world model learns independent dynamics models for each factor but not able to tell which factor is controlled by the agent. In this case, to learn from the action-free video and the latent actions, robot would have to aggregate back the information from each factors, and learn an latent action decoder.

A natural way to think of for efficient learning is to extract controllable representations through extra supervision from a task classifier, so that the robot only needs to learn from the effective control related information, getting rid of the dynamics of unrelated factors. After learning the policy over controllable representations, we can then utilize the model for cross-embodiment learning, where a robot can learn from demonstration videos on robots with different embodiment, or even from human video, with less complicated depth estimation, joint alignment that classic methods need.

Learning from Multi-Fidelity Data. So far, we have been dealing with efficient robot learning from multiple perspectives, however, we assume the diverse data origins from sources with high fidelity, which is not true in the noisy real-world environment robots operate. And given the different underlying physical solvers that different simulators use, the data gathered from different simulation environments usually do not share the same level of fidelity, especially for contact-rich robot tasks. We propose to use large amount of lower-fidelity but low-cost data from GPU-accelerated simulations to help reducing the learning variance over small amount of higher-fidelity but high-cost data from slower simulations. We hope this can largely reduce the online data collection time and lead to more efficient robot learning.

References

Dasari, S.; Ebert, F.; Tian, S.; Nair, S.; Bucher, B.; Schmeckpeper, K.; Singh, S.; Levine, S.; and Finn, C. 2019. Robonet: Large-scale multi-robot learning. *arXiv preprint arXiv:1910.11215*.

Ebert, F.; Yang, Y.; Schmeckpeper, K.; Bucher, B.; Georgakis, G.; Daniilidis, K.; Finn, C.; and Levine, S. 2021. Bridge data: Boosting generalization of robotic skills with cross-domain datasets. *arXiv preprint arXiv:2109.13396*.

O'Neill, A.; Rehman, A.; Maddukuri, A.; Gupta, A.; Padalkar, A.; Lee, A.; Pooley, A.; Gupta, A.; Mandlekar, A.; Jain, A.; et al. 2024. Open x-embodiment: Robotic learning datasets and rt-x models: Open x-embodiment collaboration 0. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 6892–6903. IEEE.

Shi, C.; Madera, J.; Boyea, H.; and Fey, A. M. 2023. Haptic guidance using a transformer-based surgeon-side trajectory prediction algorithm for robot-assisted surgical training. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1942–1949. IEEE.

Shi, C.; Wang, Z.; Hu, J.; Martín-Martín, R.; and Stone, P. 2025. FLAM: Scaling Latent Action Models with Factorization. In *NeurIPS 2025 Workshop on Embodied World Models for Decision Making*.

Shi, C.; and Zhang, A. 2025a. FastDP: Deployable Diffusion Policy for Fast Inference Speed. In *RLC 2025 Workshop on Practical Insights into Reinforcement Learning for Real Systems*.

Shi, C.; and Zhang, A. 2025b. TP-PPO:Non-Myopic 3D Bin Packing with Tree Search Guided Online Planning. Submitted to 2026 International Conference on Automated Planning and Scheduling (ICAPS).

Shi, C.; Zheng, Y.; and Fey, A. M. 2022. Recognition and prediction of surgical gestures and trajectories using transformer models in robot-assisted surgery. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 8017–8024. IEEE.