Learning with Less Effort: Efficient Training and Generalization in (Multi-)Robot Systems

Peihong Yu
University of Maryland
College Park, United States
peihong@umd.edu

ABSTRACT

The growing demand for automation has sparked interest in multi-robot systems that can handle complex tasks through collaboration. While these systems offer advantages in speed, coverage, and capability compared to single robots, getting multiple robots to learn and coordinate effectively remains challenging — training robots requires extensive effort on data and computation, and learned policies often struggle to generalize beyond training conditions. My research addresses two fundamental challenges in multi-robot learning: reducing the training effort required and improving generalization to reduce policy retraining. First, we propose methods to make training more efficient — using human-drawn sketches rather than teleoperated demonstrations for manipulation tasks, and utilizing individual robot demonstrations instead of joint multi-robot ones for learning collaborative behaviors. Second, we develop techniques to help learned policies adapt to new scenarios without retraining — introducing frameworks that maintain coordination under different observation conditions and enable effective information sharing across varying initial states. My work aims to create more practical and adaptable multi-robot systems that can be efficiently trained and deployed across diverse real-world settings.

KEYWORDS

Multi-Agent Reinforcement Learning, Learn from Demonstration, Generalization, Manipulation

ACM Reference Format:
Peihong Yu. 2025. Learning with Less Effort: Efficient Training and Generalization in (Multi-)Robot Systems. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 3 pages.

1 INTRODUCTION

Multi-robot systems are becoming essential in applications from warehouse automation to search and rescue, offering advantages in speed, coverage, and capability compared to single robots [7, 12, 17]. However, getting multiple robots to learn and coordinate effectively remains a fundamental challenge. While multi-agent reinforcement learning (MARL) [23] provides a framework for training robot teams, it faces several key difficulties limiting practical application. The first major challenge is the substantial training effort required for effective coordination. As the number of robots increases, the state-action space grows exponentially, making exploration extremely difficult — robots must learn not only their individual tasks but also how to coordinate with teammates. To address this challenge, current approaches often rely on collecting demonstrations where multiple human experts control the robots simultaneously to show desired coordination patterns [14, 15]. However, such joint demonstrations are time-consuming to collect and must be recollected whenever the team composition changes. The second major challenge persists even after successful training — learned behaviors often fail to generalize beyond training conditions. Changes in observation conditions, initial states, or team configurations [16] can severely degrade performance. This brittleness forces robots to either undergo constant retraining or operate only in highly controlled environments, limiting their real-world applicability.

My research addresses these fundamental challenges in multi-robot learning through two main directions: (1) Reducing initial training effort by developing methods that leverage more accessible forms of human guidance — using simple 2D sketches for robot manipulation and individual demonstrations for team behaviors; (2) Improving generalization to reduce retraining effort by developing frameworks that enable learned policies to maintain coordination across varying conditions. These advances would significantly impact real-world robotics applications. Lower training effort means multi-robot systems become accessible to a broader range of users and applications, not just specialized settings with abundant resources. Better generalization capabilities allow robots to adapt to changing conditions without requiring constant retraining, making them more practical for dynamic real-world environments.

2 LESS TRAINING EFFORT

Learning from Demonstration (LfD) [1, 5, 10] has emerged as a key approach for efficient robot training, allowing robots to learn directly from expert behaviors rather than requiring extensive trial-and-error exploration. However, demonstration collection remains a major bottleneck. For manipulation tasks, they typically rely on kinesthetic teaching [4, 9] or teleoperation [13] requiring specialized hardware and expertise. For multi-robot coordination, they often need multiple experts controlling robots simultaneously [14, 15], which scales poorly as more robots are added. These requirements create significant barriers to deploying robot systems in practice.

For robot manipulation tasks, we observe that humans naturally communicate motion ideas through simple 2D sketches - like drawing a path to show how to navigate or demonstrate a desired motion pattern. Previous works have explored leveraging sketches in robotics in different ways. For example, RT-Trajectory [3] uses sketches to condition policies in imitation learning, while Zhi et al. [24] proposed diagrammatic teaching that directly fits and executes

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

trajectories from sketches. However, these approaches are limited to replicating the provided sketches and require new sketches for each task execution. We developed Sketch-To-Skill to leverage sketches more broadly in reinforcement learning. The framework first learns to map 2D sketches to 3D trajectories through a pretrained generator. These trajectories enable autonomous collection of initial demonstrations through open-loop servoing. We then utilize these sketch-generated demonstrations in two ways: to pretrain an initial policy through behavior cloning and to refine this policy through reinforcement learning with guided exploration. We evaluate our approach on six manipulation tasks in MetaWorld [22] including tasks that require precise gripper control like BoxClose and CoffeePush. Despite using only basic sketches as input and no explicit gripper information, this approach achieves ~96% of the performance of policies trained with teleoperated demonstrations while exceeding pure reinforcement learning performance by ~170%. We further validate our approach on physical UR3e robot hardware on several real-world task, achieving an ~80% success rate in randomized settings.

In multi-robot coordination tasks, collecting joint demonstrations becomes increasingly challenging as team size grows. Recent works like MAGAIL [14] and DM2 [15] have explored using demonstrations to guide multi-agent learning, but they can only converge when using joint demonstrations from co-trained policies, which naturally provide compatible behaviors. When demonstrations come from mixed sources, the behaviors can conflict, preventing successful learning. Additionally, when team configurations change or new agents are introduced, demonstrations must be recollected. We developed PegMARL [20] to address these challenges through personalized demonstrations — demonstrations that show individual agents performing their tasks independently rather than as part of a team. This approach offers natural scalability: agents of the same type can share demonstrations regardless of team size, and new demonstrations are only needed when introducing new types of agents. Our framework leverages these personalized demonstrations through two discriminators: a personalized behavior discriminator that provides positive incentives for actions that align with demonstrations and negative incentives for divergent ones, and a personalized transition discriminator that adjusts these incentive weights based on whether actions lead to desired state changes similar to those observed in demonstrations. Together, these enable robots to learn effective coordination strategies even though the demonstrations don't explicitly show cooperative behaviors. Our experiments demonstrate PegMARL's effectiveness across different scenarios. In gridworld environments, PegMARL with personalized demonstrations shows faster convergence than baselines and better scaling with increasing agent numbers. We further validate our approach in StarCraft [11] multi-agent scenarios, where PegMARL converges effectively even with joint demonstrations from mixed sources, showing its ability to bootstrap from and improve upon provided demonstrations.

3 LESS POLICY RETRAINING EFFORT

Even after successful training, learned multi-robot policies often fail to generalize beyond training conditions. Changes in observation conditions, initial states, or team configurations can severely degrade performance. Current approaches typically require complete retraining for each new scenario or extensive fine-tuning, making deployment impractical for real-world applications where conditions frequently vary.

Drawing inspiration from how humans adapt their communication based on what information others need, we developed TACTIC [21], a Task-Agnostic Contrastive pre-Training framework for Inter-agent Communication. Our approach enables agents to maintain coordination even when sight ranges during execution differ significantly from those during training. TACTIC consists of two key stages: offline contrastive pretraining and online policy integration. In the pretraining stage, we train two key communication modules — a message generator and a message-observation integrator — using contrastive learning [6]. The objective aligns the integration of local observations and messages with the full egocentric state for each agent, enabling agents to effectively "see" beyond their limited sight ranges through communication. During online policy learning, these pretrained communication modules are frozen and incorporated into agents' policy learning, enabling dynamic communication adaptation while preserving the learned task-agnostic properties. We evaluate TACTIC in the SMACv2 [2] benchmark across different scenarios and sight ranges. Unlike baseline methods that struggle to generalize across sight ranges, TACTIC maintains consistent performance by learning to communicate task-relevant information effectively - for example, achieving stable win rates across varying sight ranges in combat scenarios where traditional methods' performance drops significantly.

In multi-agent systems, agents are also prone to failure when faced with shifts in state distribution. To address this, we developed an approach based on Common Operating Picture (COP) integration [19]. Each agent is equipped with the capability to integrate its observations, actions, and received messages into a COP that is dynamically updated to reflect the environment and mission. This process takes into account both current observations and historical information. Rather than directly communicating local observations which can overwhelm communication channels, agents share processed summaries that help maintain a consistent global understanding. Our results in StarCraft2 [11] show that COP-based training produces robust policies that significantly outperform state-of-the-art MARL methods under out-of-distribution initial states.

4 CONCLUSION AND FUTURE RESEARCH

This research develops methods to make multi-robot learning more practical by reducing training effort and improving generalization. We have shown that sketches and personalized demonstrations can effectively replace more complex training data, while frameworks for information sharing with a focus on global awareness enable better generalization across varying conditions.

Our future work will explore zero-shot generalization [8, 18] in multi-robot systems. As large language models exhibit strong ability to understand and generate human instructions, we aim to investigate how these models can help robots better understand team dynamics and adapt to new partners. The goal is to create truly adaptable robot teams that can coordinate effectively with new teammates without requiring additional training, making deployment more flexible across different configurations and scenarios.

REFERENCES

[1] Jie Chen and Wenjun Xu. 2022. Policy gradient from demonstration and curiosity. *IEEE Transactions on Cybernetics* (2022).

[2] Benjamin Ellis, Jonathan Cook, Skander Moalla, Mikayel Samvelyan, Mingfei Sun, Anuj Mahajan, Jakob Foerster, and Shimon Whiteson. 2024. Smacv2: An improved benchmark for cooperative multi-agent reinforcement learning. *Advances in Neural Information Processing Systems* 36 (2024).

[3] Jiayuan Gu, Sean Kirmani, Paul Wohlhart, Yao Lu, Montserrat Gonzalez Arenas, Kanishka Rao, Wenhao Yu, Chuyuan Fu, Keerthana Gopalakrishnan, Zhuo Xu, et al. 2023. Rt-trajectory: Robotic task generalization via hindsight trajectory sketches. *arXiv preprint arXiv:2311.01977* (2023).

[4] Auke Jan Ijspeert, Jun Nakanishi, Heiko Hoffmann, Peter Pastor, and Stefan Schaal. 2013. Dynamical Movement Primitives: Learning Attractor Models for Motor Behaviors. *Neural Computation* 25, 2 (Feb. 2013), 328–373. https://doi.org/10.1162/neco_a_00393

[5] Bingyi Kang, Zequn Jie, and Jiashi Feng. 2018. Policy optimization with demonstrations. In *International conference on machine learning*. PMLR, 2469–2478.

[6] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. 2020. Supervised Contrastive Learning. In *Advances in Neural Information Processing Systems*, Vol. 33. 18661–18673.

[7] Tianxu Li, Kun Zhu, Nguyen Cong Luong, Dusit Niyato, Qihui Wu, Yang Zhang, and Bing Chen. 2022. Applications of multi-agent reinforcement learning in future internet: A comprehensive survey. *IEEE Communications Surveys & Tutorials* 24, 2 (2022), 1240–1279.

[8] Reuth Mirsky, Ignacio Carlucho, Arrasy Rahman, Elliot Fosong, William Macke, Mohan Sridharan, Peter Stone, and Stefano V Albrecht. 2022. A survey of ad hoc teamwork research. In *European conference on multi-agent systems*. Springer, 275–293.

[9] Alexandros Paraschos, Elmar Rueckert, Jan Peters, and Gerhard Neumann. 2015. Model-free Probabilistic Movement Primitives for physical interaction. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. https://doi.org/10.1109/iros.2015.7353771

[10] Desik Rengarajan, Gargi Vaidya, Akshay Sarvesh, Dileep Kalathil, and Srinivas Shakkottai. 2022. Reinforcement Learning with Sparse Rewards using Guidance from Offline Demonstration. In *International Conference on Learning Representations*. https://openreview.net/forum?id=YJ1WzgMVsMt

[11] Mikayel Samvelyan, Tabish Rashid, Christian Schroeder De Witt, Gregory Farquhar, Nantas Nardelli, Tim GJ Rudner, Chia-Man Hung, Philip HS Torr, Jakob Foerster, and Shimon Whiteson. 2019. The starcraft multi-agent challenge. *arXiv preprint arXiv:1902.04043* (2019).

[12] Shai Shalev-Shwartz, Shaked Shammah, and Amnon Shashua. 2016. Safe, multi-agent, reinforcement learning for autonomous driving. *arXiv preprint arXiv:1610.03295* (2016).

[13] Weiyong Si, Ning Wang, and Chenguang Yang. 2021. A review on manipulation skill acquisition through teleoperation-based learning from demonstration. *Cognitive Computation and Systems* 3, 1 (2021), 1–16.

[14] Jiaming Song, Hongyu Ren, Dorsa Sadigh, and Stefano Ermon. 2018. Multi-agent generative adversarial imitation learning. *Advances in neural information processing systems* 31 (2018).

[15] Caroline Wang, Ishan Durugkar, Elad Liebman, and Peter Stone. 2023. DM2: Decentralized Multi-Agent Reinforcement Learning via Distribution Matching. (2023).

[16] Weixun Wang, Tianpei Yang, Yong Liu, Jianye Hao, Xiaotian Hao, Yujing Hu, Yingfeng Chen, Changjie Fan, and Yang Gao. 2020. From few to more: Large-scale dynamic multiagent curriculum learning. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34. 7293–7300.

[17] Erfu Yang and Dongbing Gu. 2004. *Multiagent reinforcement learning for multi-robot systems: A survey.* Technical Report. tech. rep.

[18] Lebin Yu, Yunbo Qiu, Quanming Yao, Xudong Zhang, and Jian Wang. 2023. Improving zero-shot coordination performance based on policy similarity. In *Proceedings of the International Conference on Automated Planning and Scheduling*, Vol. 33. 438–442.

[19] Peihong Yu, Bhoram Lee, Aswin Raghavan, Supun Samarasekera, Pratap Tokekar, and James Zachary Hare. 2023. Enhancing Multi-Agent Coordination through Common Operating Picture Integration. In *First Workshop on Out-of-Distribution Generalization in Robotics at CoRL 2023*. https://openreview.net/forum?id=fADcJl0B0P

[20] Peihong Yu, Manav Mishra, Alec Koppel, Carl Busart, Priya Narayan, Dinesh Manocha, Amrit Singh Bedi, and Pratap Tokekar. 2025. Beyond Joint Demonstrations: Personalized Expert Guidance for Efficient Multi-Agent Reinforcement Learning. *Transactions on Machine Learning Research* (2025). https://openreview.net/forum?id=kzPNHQ8ByY

[21] Peihong Yu, Manav Mishra, Syed Zaidi, and Pratap Tokekar. 2024. TACTIC: Task-Agnostic Contrastive pre-Training for Inter-Agent Communication. In *Multi-Agent reinforcement Learning for Transportation Autonomy*. https://openreview.net/forum?id=CMYiApcudp

[22] Tianhe Yu, Deirdre Quillen, Zhanpeng He, Ryan Julian, Karol Hausman, Chelsea Finn, and Sergey Levine. 2019. Meta-World: A Benchmark and Evaluation for Multi-Task and Meta Reinforcement Learning. In *Conference on Robot Learning (CoRL)*. arXiv:1910.10897 [cs.LG] https://arxiv.org/abs/1910.10897

[23] Kaiqing Zhang, Zhuoran Yang, and Tamer Başar. 2021. Multi-agent reinforcement learning: A selective overview of theories and algorithms. *Handbook of Reinforcement Learning and Control* (2021), 321–384.

[24] Weiming Zhi, Tianyi Zhang, and Matthew Johnson-Roberson. 2023. Learning from demonstration via probabilistic diagrammatic teaching. *arXiv preprint arXiv:2309.03835* (2023).