ICRA2026 Late Breaking Results Poster presented at
2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Heterogeneous Skill Learning for Asynchronous Multi-Robot Relay Pushing in Complex Environments

Hui Zhi*, David Navarro-Alarcon
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hong Kong

***Abstract*—This paper presents a heterogeneous skill learning framework for asynchronous multi robot relay pushing in complex and cluttered environments. To support cooperative relay transportation, we construct a skill library comprising room robot pushing, corridor helper pushing, and standby behaviors. We further propose a geometry aware pushing strategy that enables contact rich manipulation without relying on external force sensors. For the room robot, curriculum learning is adopted to decompose training into an approach to parcel phase and a parcel to target pushing phase, thereby improving training stability and task progression. For long horizon transportation in constrained corridors, an affordance network is introduced to model the local feasibility of pushing actions, providing structured guidance that improves policy learning efficiency. The overall framework combines Soft Actor Critic with Dijkstra based reachability maps to coordinate the"Corridor Helper Pushing" skills. Experimental results demonstrate high success rates across progressive curriculum lessons, suggesting that the proposed framework provides an effective skill primitive for cooperative multi robot transportation.**

***Index Terms*—Multi robot collaboration, nonprehensile manipulation, relay pushing, deep reinforcement learning, affordance learning.**

I. INTRODUCTION

Nonprehensile manipulation, particularly object pushing, offers significant operational versatility and maintains mechanical simplicity; however, it introduces formidable nonlinear dynamical challenges when operating within constrained workspaces. In large scale maze environments, multirobot systems frequently encounter coordination and collision difficulties, especially when navigating restricted areas. To address these complexities, we investigate a Relay Pushing paradigm that decomposes intricate transport tasks into a sequence of manageable subtasks. The successful implementation of such a relay strategy necessitates a robust and reliable underlying pushing controller. Existing methodologies often rely on direction analysis [1] which frequently imposes restrictive curvature constraints on the resulting trajectories. While Model Predictive Control (MPC) is a common alternative for handling such highly constrained optimization problems, its inherent computational intensity often renders it ineffective for dynamic tasks requiring responses in real time. Consequently, this work leverages Soft Actor Critic (SAC)[2] to develop a highly adaptable and reactive controller specifically designed for pushing maneuvers in complex environments.

The main contributions are summarized as follows:

1) A geometry aware pushing method is developed to achieve contact rich manipulation without requiring force sensors.
2) An affordance network is introduced to model local physical pushing feasibility, providing structured guidance that improves policy learning for long horizon parcel transportation tasks.
3) The learned pushing behavior provides a foundational skill primitive for multi robot cooperative relay parcel transportation.

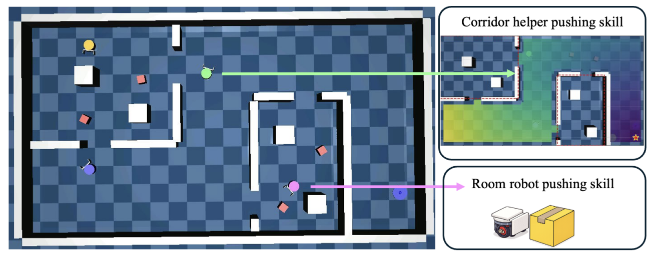

Fig. 1. Overview of the proposed heterogeneous skill learning framework for asynchronous multi-robot relay pushing, illustrating the interaction between the room pushing and corridor helper pushing modules for parcel transportation in complex environments.

II. METHODOLOGY

The proposed approach focuses on learning specialized behavioral primitives for nonprehensile manipulation. We develop three distinct skills—Room Pushing, Corridor Helper, and Standby—to manage the varying geometric constraints of complex environment.

A. Room Pushing Skill

The Room Pushing Skill is designed for object transportation in relatively open workspaces and is learned using the SAC framework. The observation space $\mathcal{O}_{room}$ combines the robot state, the relative state of the parcel, and the relative state of the target, thereby providing a comprehensive description of the manipulation process. In addition, an occupancy grid is incorporated to capture nearby obstacles and support collision avoidance. The reward function r_{room} is formulated to encourage task progress while preserving stable and smooth motion:

$$r_{room} = R_{prog} + R_{reach} + R_{cont} - (P_{coll} + P_{smooth}) \quad (1)$$

where R_{prog} denotes the progress of the parcel toward the target, R_{reach} rewards the robot for approaching the parcel, R_{cont} is a contact bonus, P_{coll} penalizes collisions, and P_{smooth} penalizes abrupt control changes. To improve training stability and promote reliable convergence, this skill is learned through a three-stage curriculum. In Lesson 1, the robot is initialized in a pushing-ready configuration near the parcel. In Lesson 2, the robot starts farther away, requiring an additional approach phase before contact. In Lesson 3, both the robot and the parcel are randomly initialized within the room, resulting in the most challenging setting and encouraging robust generalization.

B. Corridor Helper Pushing Skill

To facilitate navigation within constrained passages, we introduce the Corridor Helper Pushing Skill, which utilizes an Affordance guided SAC policy. This skill is specifically designed to overcome the challenges of long horizon tasks in restricted geometries, where sparse rewards and complex contact dynamics often hinder policy convergence. We characterize local physical pushing feasibility through an Affordance Net f_ψ that maps local spatial perceptions to discrete action priors. The network is trained using trajectory episodes harvested from both heuristic and room pushing policies, with samples labeled based on the success of future rollouts over K steps.

The Affordance Net performs a binary classification to identify the geometric feasibility of a potential push. The training objective is defined by the cross entropy loss:

$$\mathcal{L}(\psi) = -\sum_{t}\sum_{i\in\mathcal{C}} y_{t,i}\log(f_\psi(o_t)_i) \tag{2}$$

where the output space $\mathcal{C} = \{\text{Effective Contact, Reposition}\}$ represents the core tactical decisions. An *Effective Contact* label indicates that the current robot pose is conducive to object progress, while a *Reposition* label signals the need for the agent to adjust its contact point to avoid kinematic singularities or stuck states.

The observation space for this skill is augmented as $\mathcal{O}_t^{aug} = \{s_t, \nabla\mathcal{D}(p_t), p_{push}, \mathcal{A}_t\}$, incorporating the move direction derived from Dijkstra map gradients $\nabla\mathcal{D}(p_t)$, candidate push position p_{push} and the resulting affordance map $\mathcal{A}_t$. By integrating these local spatial primitives, the system achieves geometry aware manipulation without requiring expensive force sensors in complex corridor layouts. This structured guidance allows the SAC agent to maintain stable control over long horizons, significantly improving transport success rates compared to baseline end to end policies.

C. Standby Skill

The Standby Skill serves as a foundational behavioral primitive for multirobot coordination within the relay framework. This skill is optimized to maintain the robot in a state of readiness while the object is actively manipulated by another agent. The primary objectives are to minimize spatial interference with the active pushing trajectory and to ensure the agent remains within a reachable proximity to the designated handover zone. By treating standby behavior as a learned skill, the system dynamically manages agent positioning using the shared Dijkstra reachability map, thereby facilitating seamless transitions between individual manipulation behaviors.

III. Simulation Results

The simulation results validate the performance of the learned behavioral primitives within the MuJoCo environment. As illustrated in Fig. 1(a), the Room Pushing Skill successfully converges to a success rate between 0.8 and 0.9 across all three curriculum lessons.

The Corridor Helper Pushing Skill evaluation in Fig. 1(b) highlights the critical impact of affordance guidance on learning performance. In Lesson 1, the policy utilizing the Affordance Net demonstrates significantly higher sample efficiency than the baseline. This advantage is further amplified in Lesson 2, where the affordance guided agent achieves a success rate near 0.8, while the baseline without affordance guidance fails to surpass 0.5. These findings confirm that characterizing local physical pushing feasibility is essential for solving long horizon tasks in constrained geometries.

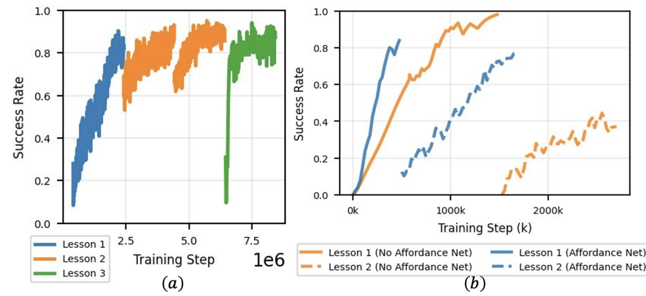

Fig. 2. The learning performance of room robot push skill (a) and corridor helper push skill (b).

IV. Conclusion

This paper presented a heterogeneous skill learning framework for asynchronous multi-robot relay pushing in complex environments. By decomposing the task into Room Pushing, Corridor Helper Pushing, and Standby skills, the proposed method enables effective manipulation and coordination under different geometric constraints. In particular, curriculum-based SAC supports robust pushing in open spaces, while affordance-guided learning improves long-horizon performance in narrow corridors.

Simulation results demonstrate reliable convergence and strong success rates across increasingly challenging scenarios. Overall, the proposed framework provides an effective solution for cooperative relay transportation without external force sensing. Future work will focus on learned multi-agent scheduling and real-world robotic validation.

References

[1] F. Bertoncelli, M. Selvaggio, F. Ruggiero, and L. Sabattini, "Task oriented contact optimization for pushing manipulation with mobile robots," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 1639–1646.

[2] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *International conference on machine learning*. Pmlr, 2018, pp. 1861–1870.