Learning Robot Manipulation from Cross-Morphology Demonstration

Gautam Salhotra*, I-Chun Arthur Liu*, Gaurav S. Sukhatme

ROBOTIC EMBEDDED SYSTEMS LABORATORY — USC University of Southern California — CoRL

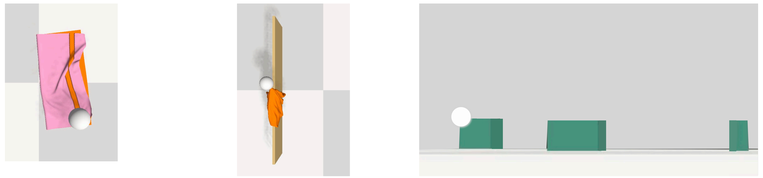
uscresl.github.io/mail

Motivation

Prior Learning from Demonstrations (LfD) cannot handle large morphological mismatches between expert and student.

Morphological Adaptation in Imitation Learning (MAIL) is a cross-morphology LfD method that can handle such mismatches, even with suboptimal teacher demonstrations.

Environments



ClothFold — Fold a cloth in half

DryCloth — Pick & hang dry a cloth

ThreeBoxes — Rearrange boxes along a line

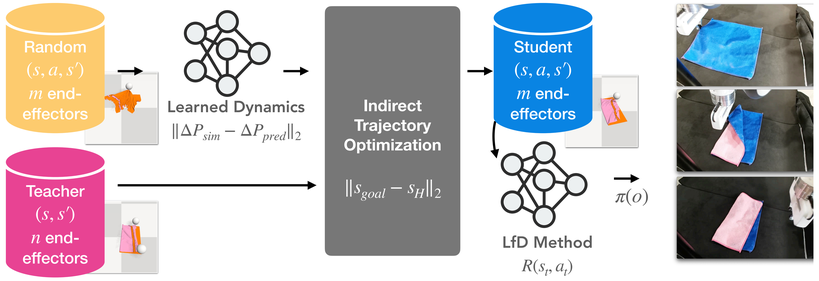
⚪ = end-effector

State s: xyz of each particle that is used to represent object.

Action a: pick (xyz) and place (xyz).

Approach

Train Dynamics → **Traj Opt** → **Train π** → **Rollout**



Generalizes across task variations and object properties. Extends from state-based demonstrations to observation-based policy.

Learned Dynamics

- CNN-LSTM-based forward dynamics model or Transformers.
- Objective: Minimize error in particle displacements $\|\Delta P_{sim} - \Delta P_{pred}\|_2$.

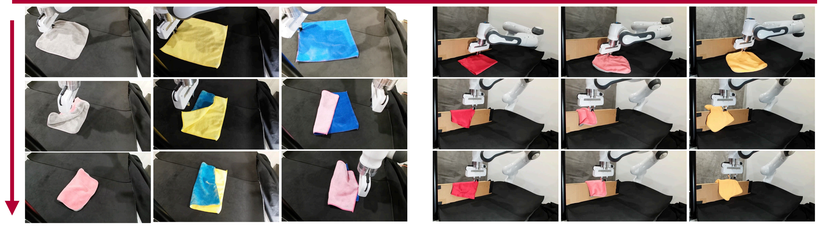
Trajectory Optimization

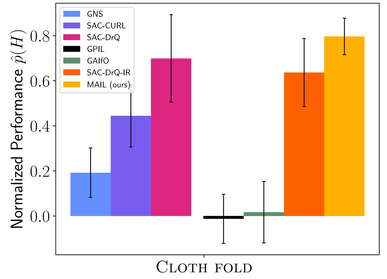
- Cross-Entropy Method (CEM) with the learned model for dynamics.
- Objective function: Match the object's goal state in the demonstration $\|s_{goal} - s_H\|_2$.

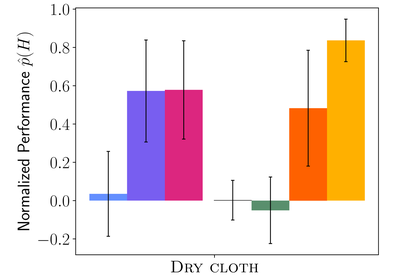
LfD Method

- Deformable Manipulation from Demonstrations (DMfD) [1], an off-policy actor-critic method for RL + LfD that balances exploration with exploitation.
  - Uses advantage-weighted samples in replay buffer, to encourage policy to stay close to stored demo states, similar to AWAC [2].
  - Entropy regularization to explore online, similar to SAC [3].
- You can choose any LfD or RL + LfD method.

Results

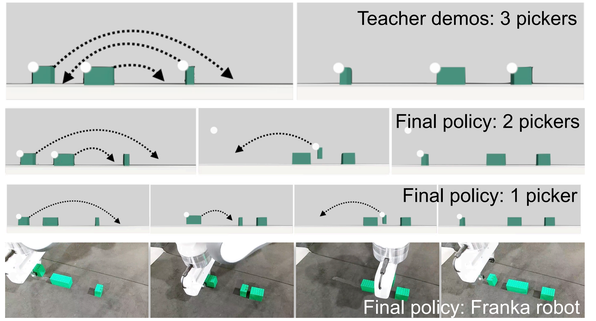






Generalization

MAIL can learn a policy for agents with M end-effectors from teachers with N end-effectors, shown in this rearrangement task.



[1] G. Salhotra, I.-C. A. Liu, M. Dominguez-Kuhne, and G. S. Sukhatme. Learning deformable object manipulation from expert demonstrations. IEEE Robotics and Automation Letters, 2022.
[2] X. B. Peng, A. Kumar, G. Zhang, and S. Levine, Advantage-weighted regression: Simple and scalable off-policy reinforcement learning. arXiv preprint arXiv:1910.00177, 2019
[3] T. Harmoja, A. Zhou, P. Abbeel, and S. Levine. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. International Conference on Machine Learning, 2018.