Leveraging Sim-and-Real Co-Training for Autonomous Dexterous Manipulation

Jared Berry[1]; Toby Buckley[1]; Kevin Lynch[1]
[1]Northwestern University, Evanston, IL

HAND — HUMAN AUGMENTATION VIA DEXTERITY

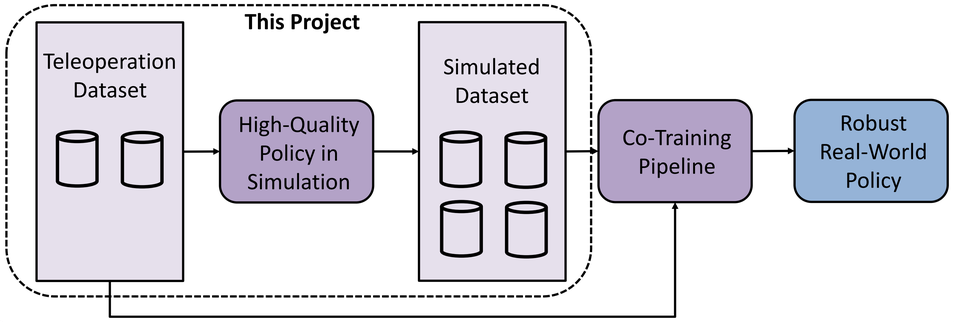
Northwestern McCORMICK SCHOOL OF ENGINEERING

Background

Goals

Train an **autonomous** agent to perform **dexterous manipulation** tasks with the DexNex hardware setup.

Use **Sim-and-Real Co-Training** [1] to train a **more robust** policy that performs well on a **varied set** of real-world manipulation tasks.



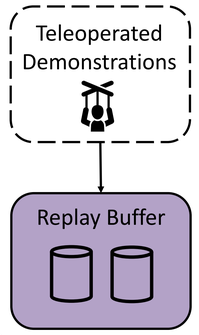
Model Framework

A. Specifications

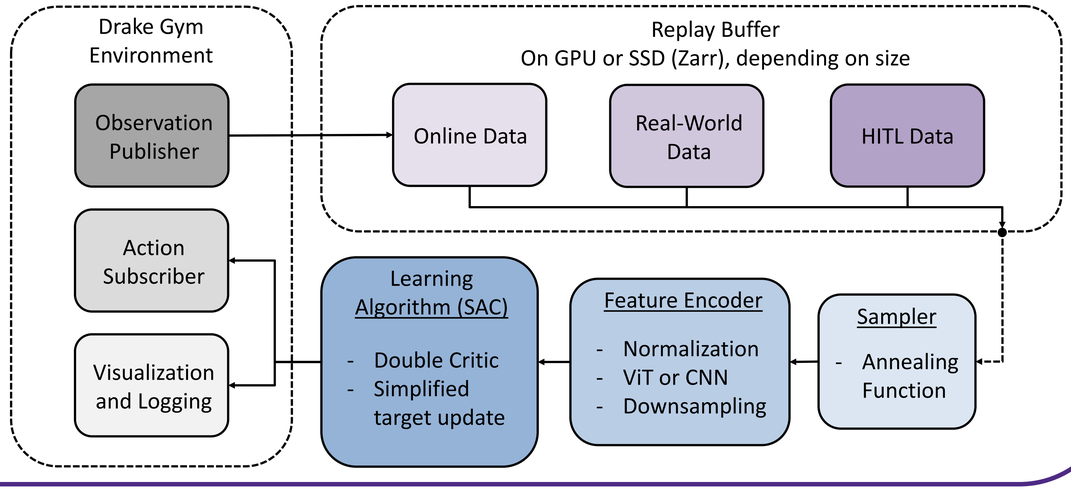
Algorithm Type: SAC [2]

Simulator: Drake [3]

B. Pre-training



C. Training



Pipeline Testing

Sparse Reward Function

$$r = r_f(p_5) \qquad r_f(p_5) = \begin{cases} R, & \text{if } p_5 \text{ in right box} \\ 0, & \text{otherwise} \end{cases}$$

Algorithm 1 Soft Actor-Critic (SAC)

Initialize parameter vectors ψ, θ_1, θ_2, ϕ
for each iteration **do**
  for each environment step **do**
    $a_t \sim \pi_\phi(a_t \mid s_t)$
    $s_{t+1} \sim p(s_{t+1} \mid s_t, a_t)$
    $D \leftarrow D \cup \{(s_t, a_t, r(s_t, a_t), s_{t+1})\}$
  end for
  for each gradient step **do**
    $\theta_i \leftarrow \theta_i - \lambda_Q \nabla_{\theta_i} J_Q(\theta_i)$, **for** $i \in \{1, 2\}$
    $\phi \leftarrow \phi - \lambda_\pi \nabla_\phi \hat{J}_\pi(\phi)$
    $Q_{\bar{\psi}_i} \leftarrow \tau Q_{\psi_i} + (1 - \tau) Q_{\bar{\psi}_i}$, **for** $i \in \{1, 2\}$
  end for
end for

Figure 1: SAC Algorithm

Shaped Reward Function

$$r = -\lambda_1 \|\mathbf{p_2} - \mathbf{p_1}\| - \lambda_2 \|\mathbf{p_4} - \mathbf{p_3}\| - \lambda_3 \|\mathbf{p_6} - \mathbf{p_5}\| + r_f(p_5)$$

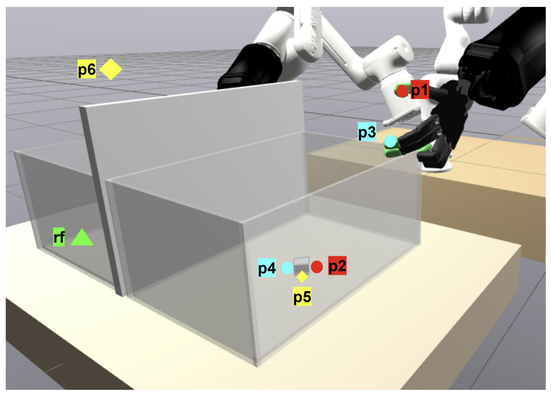


Figure 2: Reward Function Diagram

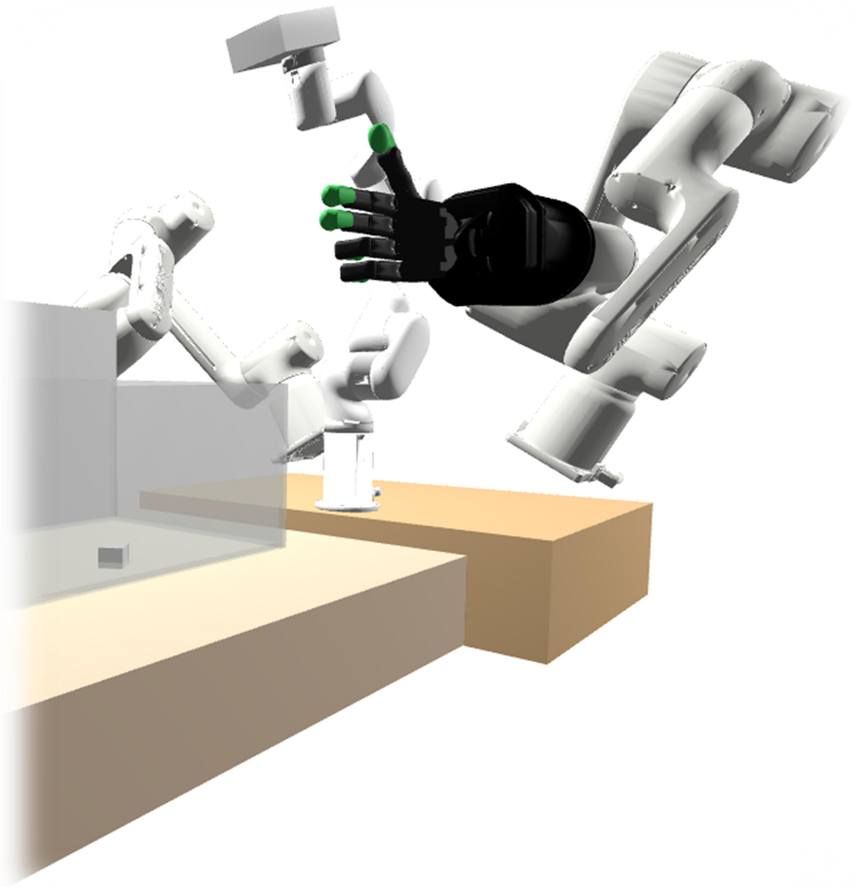


Figure 3: DexNex avatar visualized in Drake

Conclusions and Learning

- Difficulty learning with sparse reward
- Critic instability
- Computational bottlenecks
- Convergence to local minima

Future Work

- Creation of HITL demonstrations
- Domain randomization
- Fix computational bottlenecks
- Simulated dataset generation
- Integration with Co-Training pipeline

Contact Information:
Jared Berry jaredberry2026@u.northwestern.edu

Acknowledgements: This work was funded by the National Science Foundation Research Experience for Undergraduates Program, and supported by the HAND ERC and Northwestern University.

References:

1. Wei *et al* (2025). Empirical Analysis of Sim-and-Real Cotraining of Diffusion Policies for Planar Pushing from Pixels. *IROS 2025*
2. Haarnoja *et al* (2018). Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. *ICML 2018*
3. Tedrake, Russ (2025). Robotic Manipulation: Perception, Planning, and Control. *Course Notes, MIT*