AugInsert: Learning Robust Visual-Force Policies via Data Augmentation for Object Assembly Tasks

Adam Imdieke (presenting)

Ryan Diaz[1], Adam Imdieke[1], Vivek Veeriah[2], Karthik Desingh[1]

[1]Department of Computer Science and Engineering, University of Minnesota, [2]Google DeepMind

A ***multisensory*** *peg-in-hole policy learning framework should be robust to* ***previously unseen task variations***

Introduction:

A robot deployed in an unstructured environment such as a household should be capable of handling unanticipated changes to the task environment. To this end, we develop a policy learning framework and a simulation environment in MuJoCo [1] for a contact-rich peg-in-hole task that takes in RGB and force-torque (F/T) readings as input. We perform an extensive evaluation of this framework on several types of task variations.

Objectives:

1. To evaluate which types of peg-in-hole task variations pose the greatest generalization challenges for a multisensory vision + F/T framework.
2. To explore a method to improve task variation robustness through multisensory data augmentation.

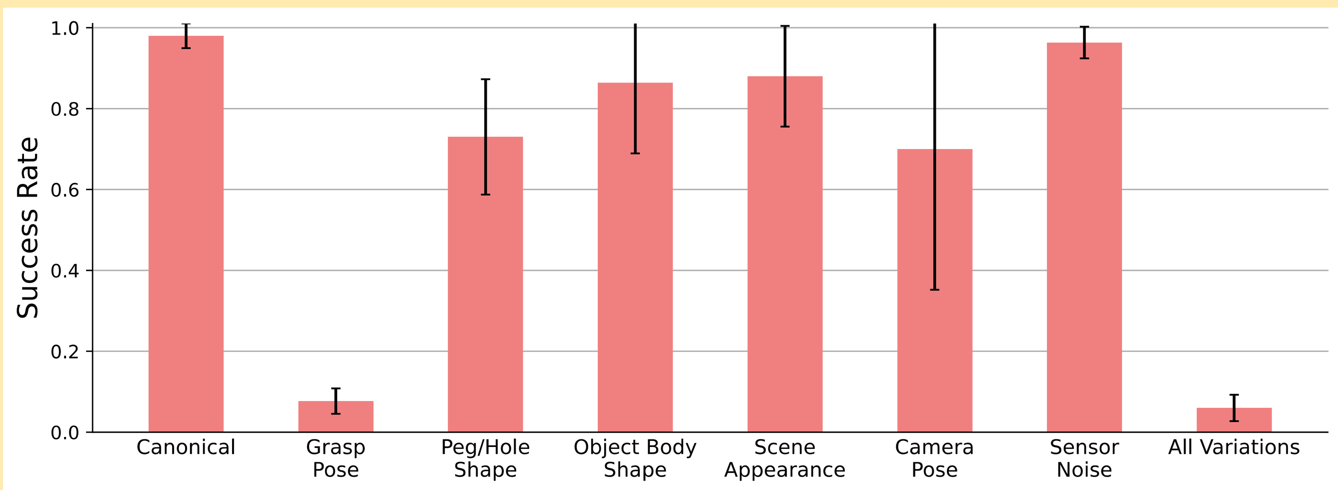
Learning and Evaluation Framework:

Our pipeline is composed of four main steps:

1. **Expert Demonstration Collection:** A human expert provides EEF trajectories through teleoperation.
2. **Data Augmentation:** Expert trajectories are replayed in environments with a subset of task variations applied.
3. **Policy Learning:** The observation encoder [2] and policy network is trained with a behavior cloning objective [3].
4. **Evaluation:** The trained policy is evaluated in an environment with task variations unseen during training.

Results:

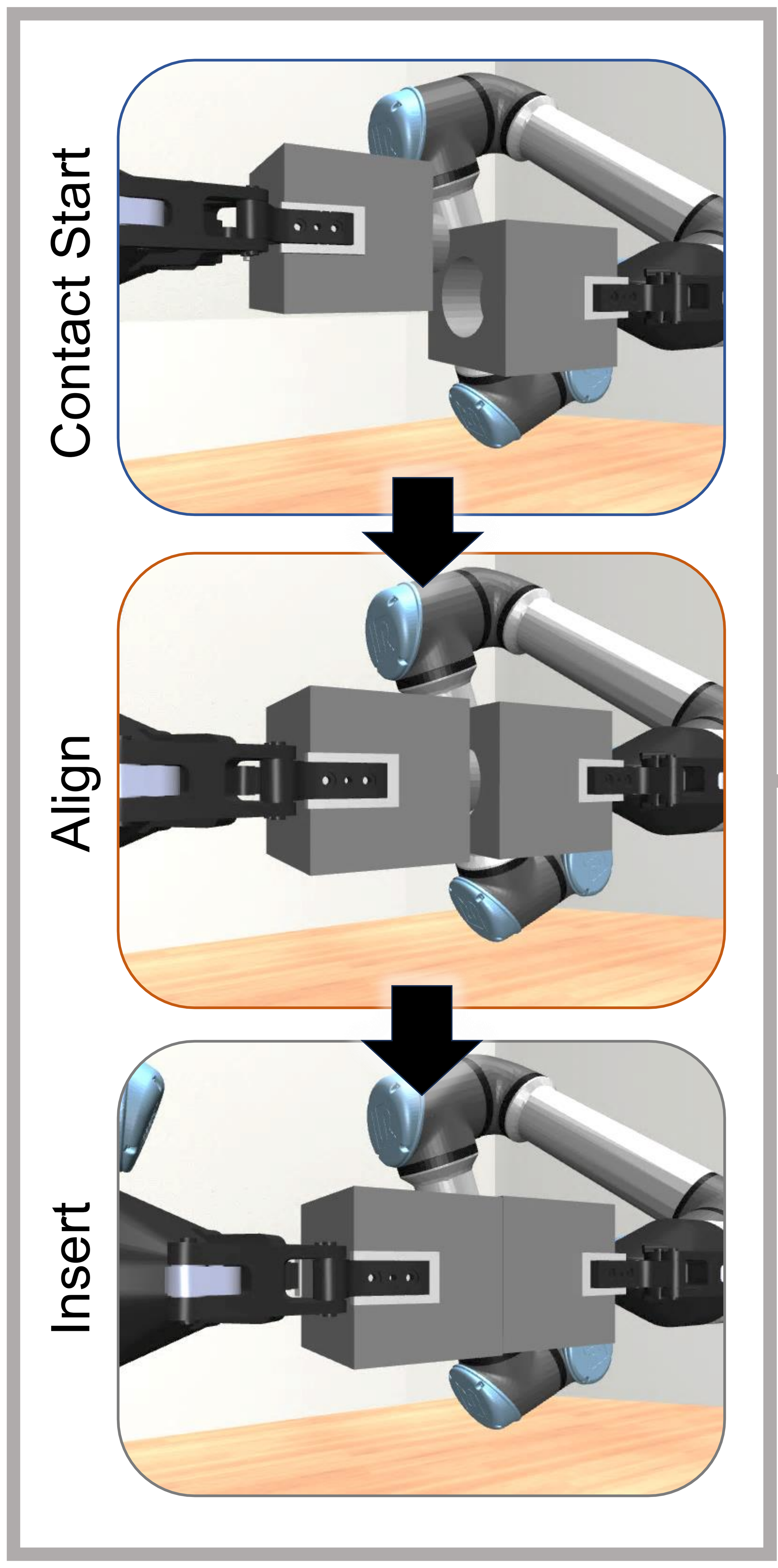
- The physical *Grasp Pose* task variation presents the largest generalization challenge for our model.



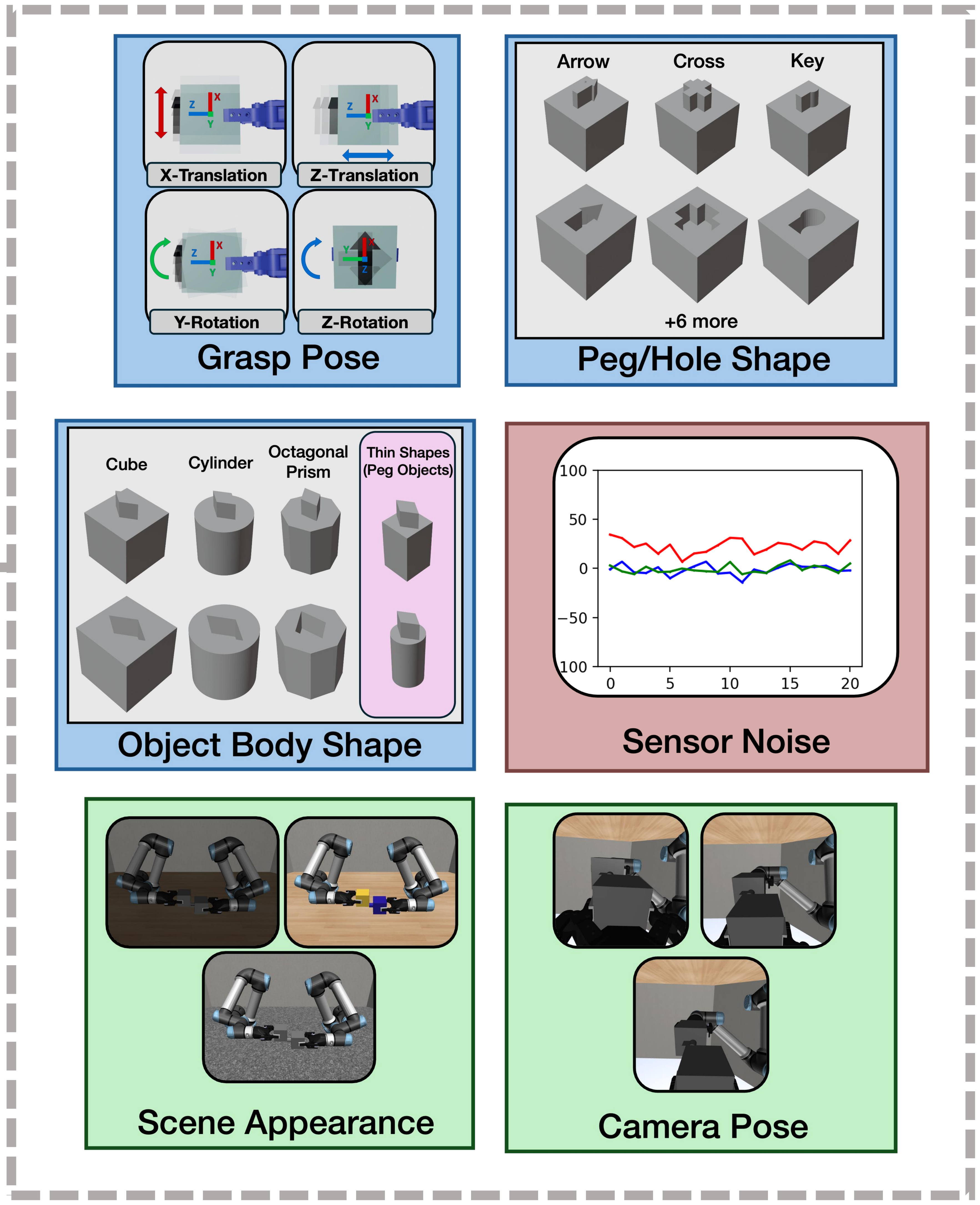
- Data augmentation with physical variations improved performance on new instances of those variations.
- Tactile (F/T) information is shown to be the most important for our task, while visual (RGB) information is the least important.

References:

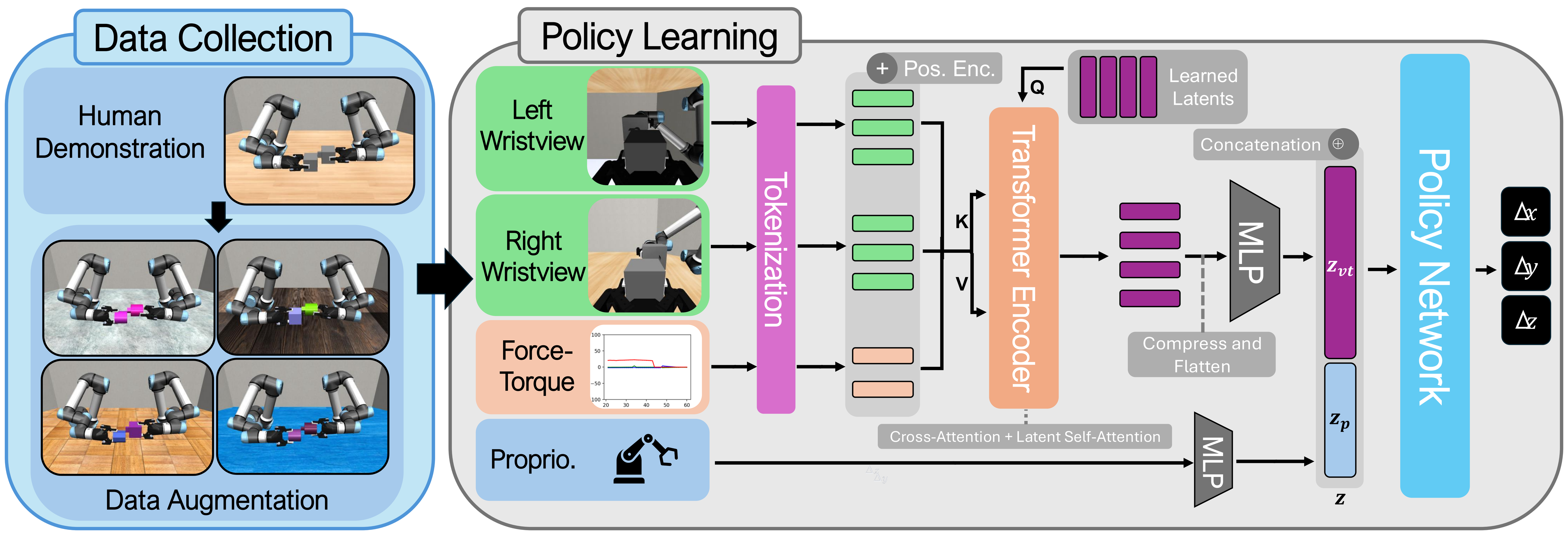
[1] E. Todorov, T. Erez, and Y. Tassa, "Mujoco: A physics engine for model-based control," in *2012 IEEE/RSJ International Conference on Intellligent Robots and Systems*, 2012, pp. 5026-5033.
[2] A. Jaegle, S. Boregeaud, J-B. Alayrac, C. Doersch, C. Ionescu, D. Ding, S. Koppula, D. Zoran, A. Brock, E. Shelhamer *et al.*, "Perceiver io: A general architecture for structured inputs and outputs," in *International Conference on Learning Representations*, 2022.
[3] A. Mandlekar, D. Xu, J. Wong, S. Nasiriany, C. Wang, R. Kulkarni, L. Fei-Fei, S. Savarese, Y. Zhu, and R. Martín-Martín, "What matters in learning from offline human demonstrations for robot manipulation," in *Conference on Robot Learning*. PMLR, 2022, pp. 1678-1690.



Task Rollout



Task Variations



Learning and Evaluation Framework

ICRA IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION ATLANTA 2025

Computer Science & Engineering

Robotics: Perception & Manipulation Lab