TEXAS Robotics

RPL UT Robot Perception & Learning Lab

LARG Learning Agents Research Group The University of Texas at Austin

GROOT saves you from retraining visuomotor policies whenever you change backgrounds, move cameras, or use new objects!

GROOT: Learning Generalizable Manipulation Policies with Object-Centric 3D Representations

Yifeng Zhu, Zhenyu Jiang, Peter Stone, Yuke Zhu

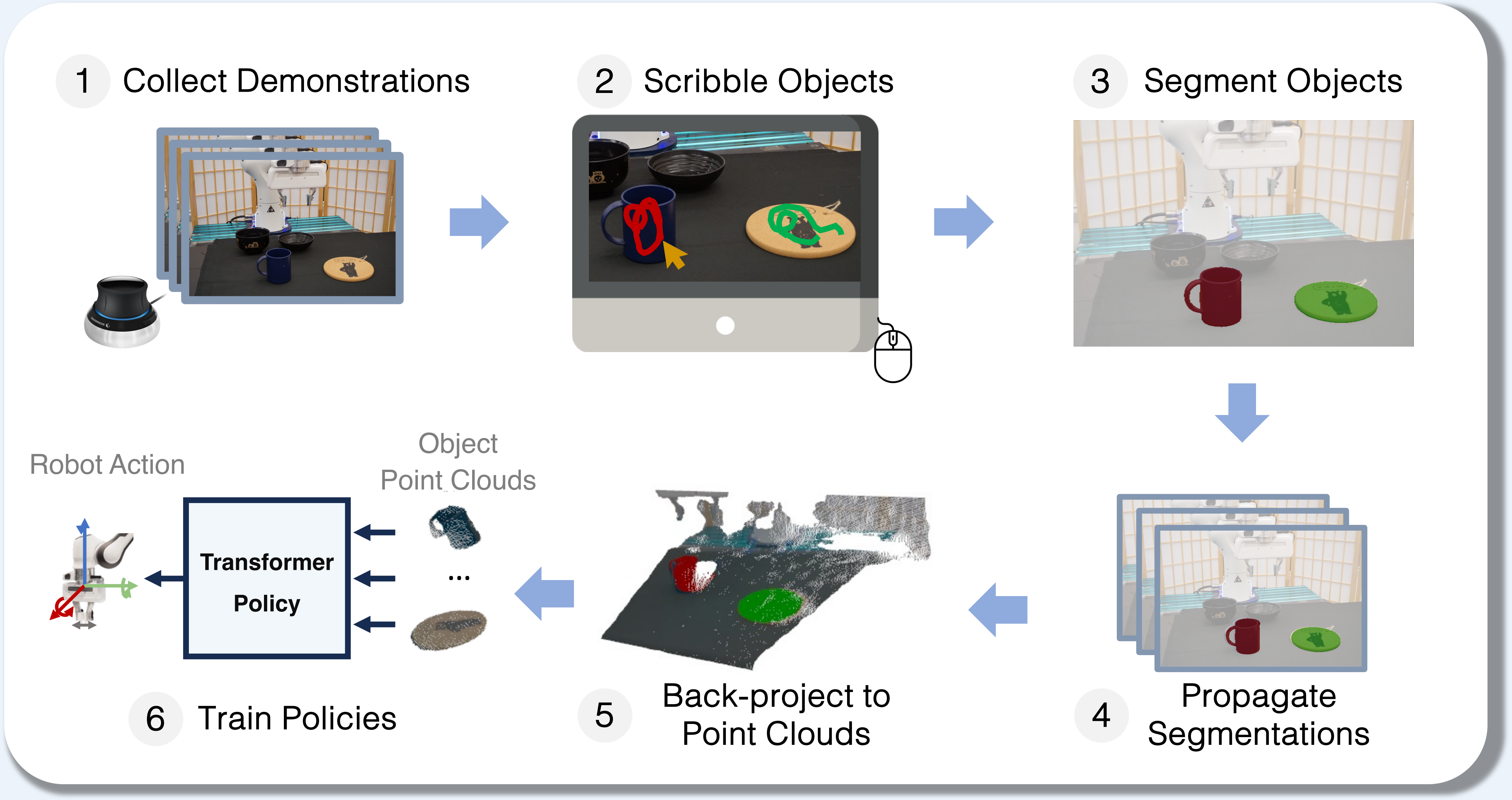
Motivation

- Prior imitation learning methods fall short in generalization beyond the training conditions
- We want to avoid the time-consuming data re-collection and model re-training in every new setting

Insights

- Object-centric:
  - Exploit compositional structure of visual scenes in objects and entities
  - Attend to task-relevant objects, minimize visual distractions
- 3D-aware:
  - Lift the spatial reasoning from the 2D plane to a unified reference frame of 3D coordinates
- New object generalization:
  - Use the open-vocabulary visual understanding of vision foundation models

GROOT Overview



Generalization of GROOT

Various Visual Backgrounds

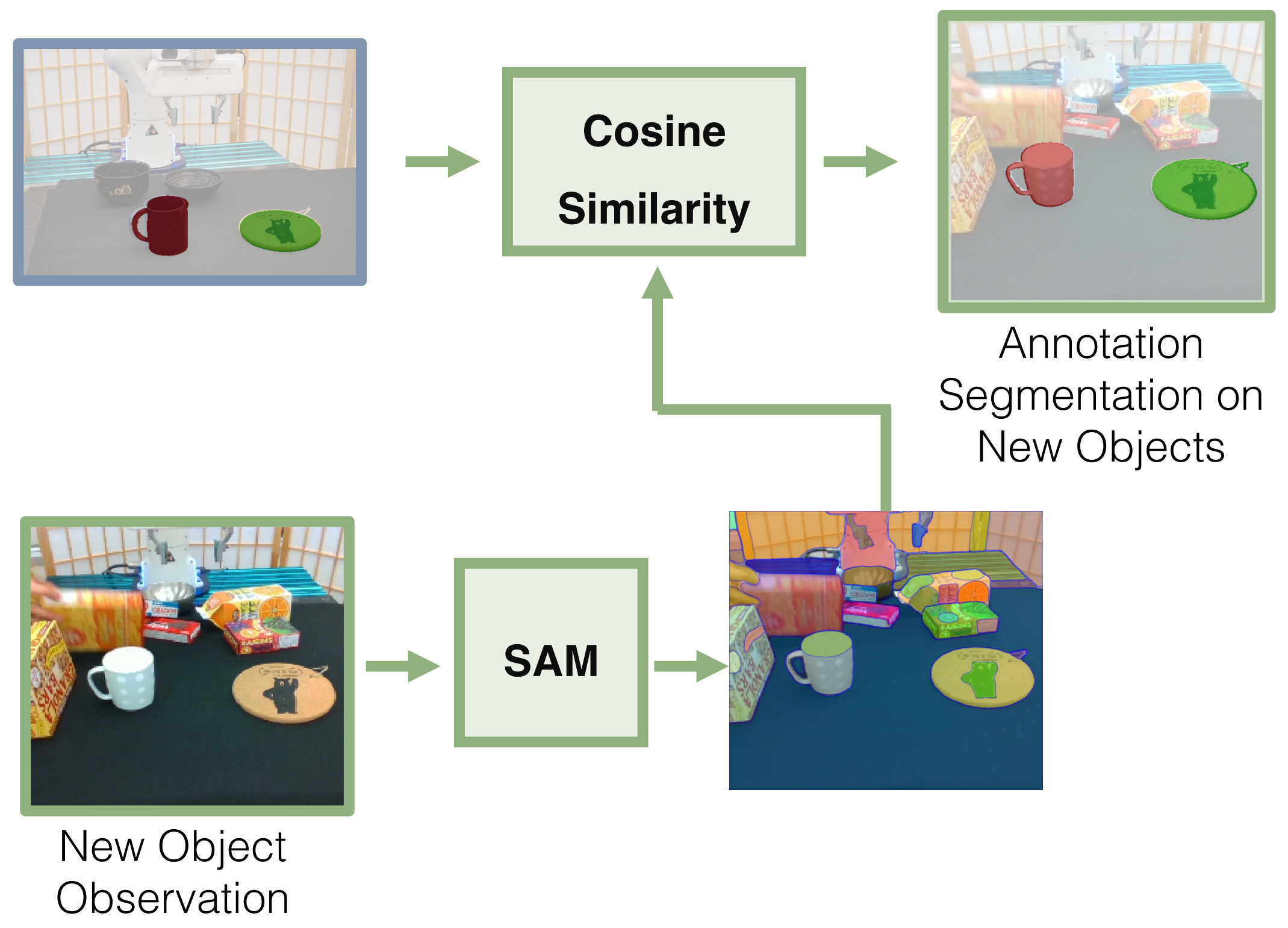
Different Camera Angles

New Object Instances

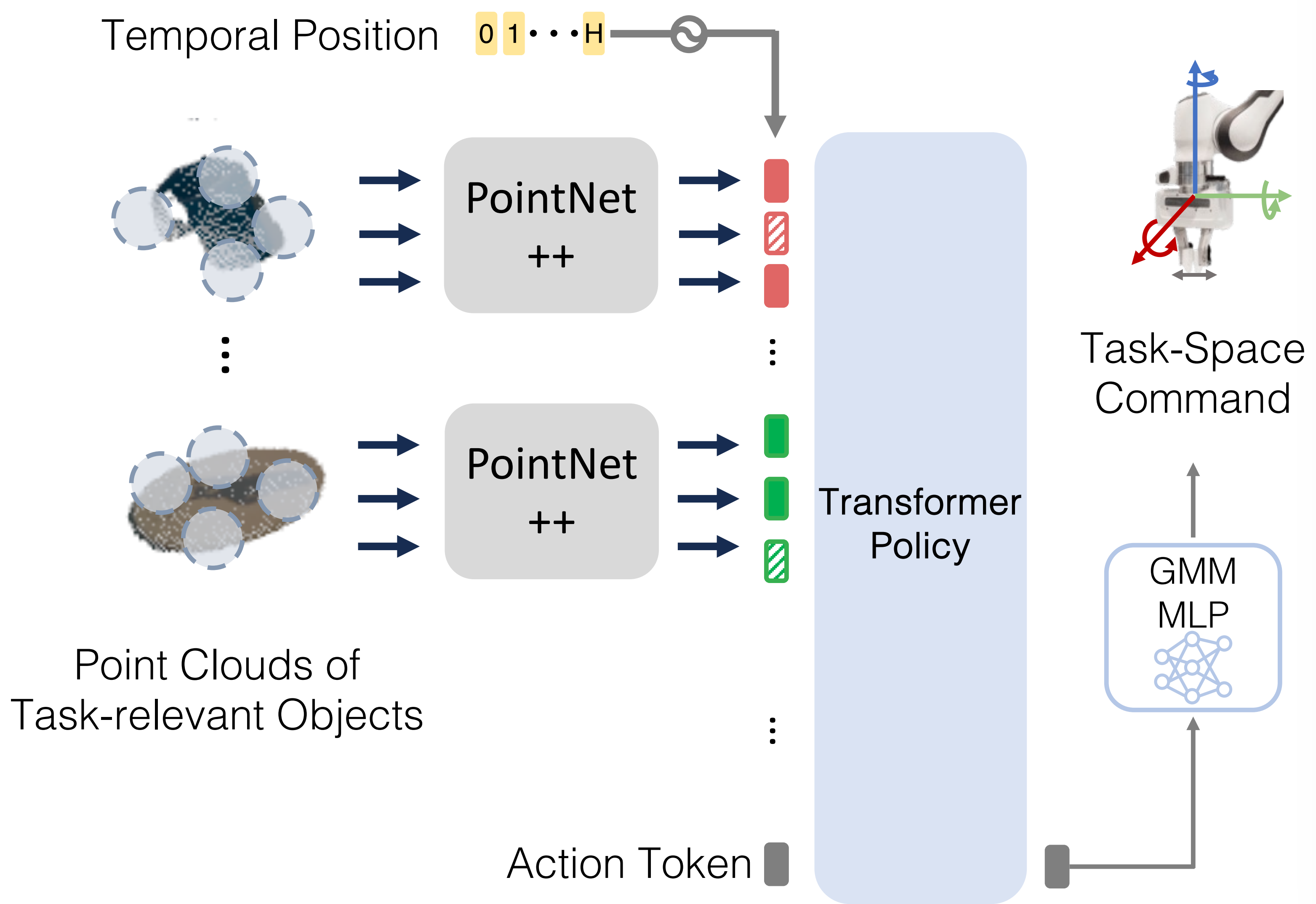
Segmentation Correspondence Model

*Only used during deployment

- SCM helps identify new object
- The new annotation segmentation then propagates to new objects



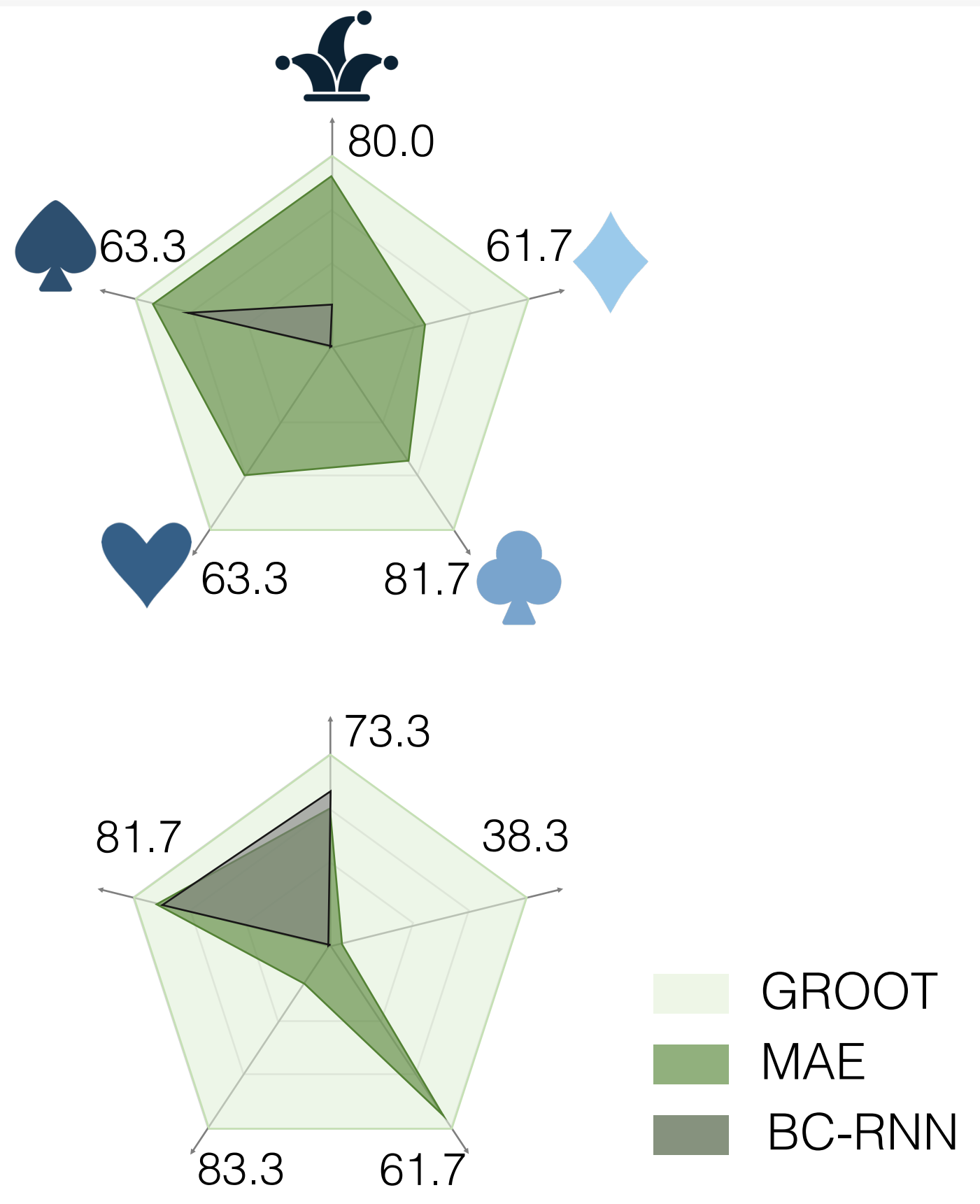
Transformer Policy



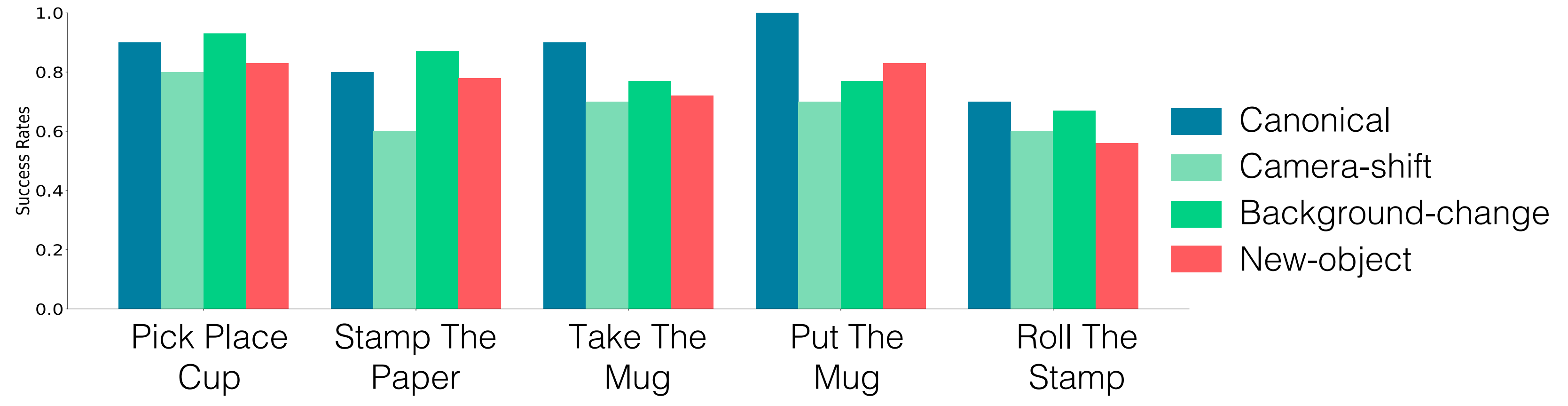
Experiments

A Simulation

Put the moka pot on the stove
Reposition the yellow and white mug



B Real Robot



C New Object Generalization