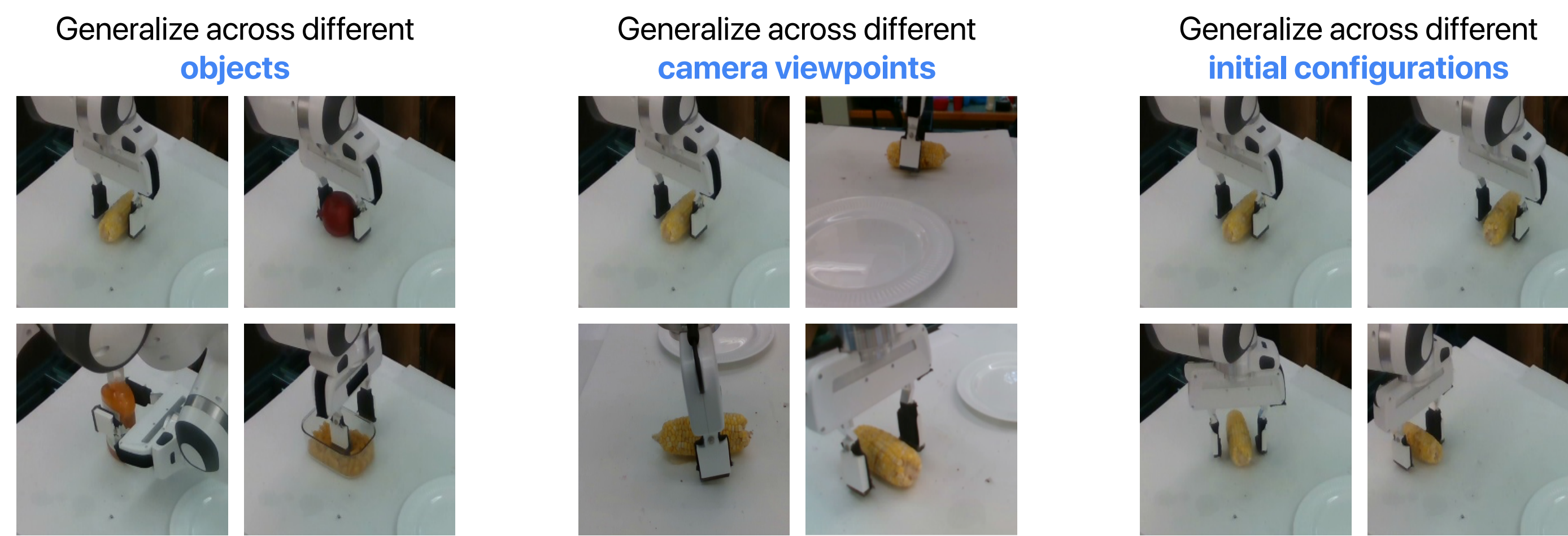


Visually-Grounded Library of Behaviors for Manipulating Diverse Objects across Diverse Configurations and Views

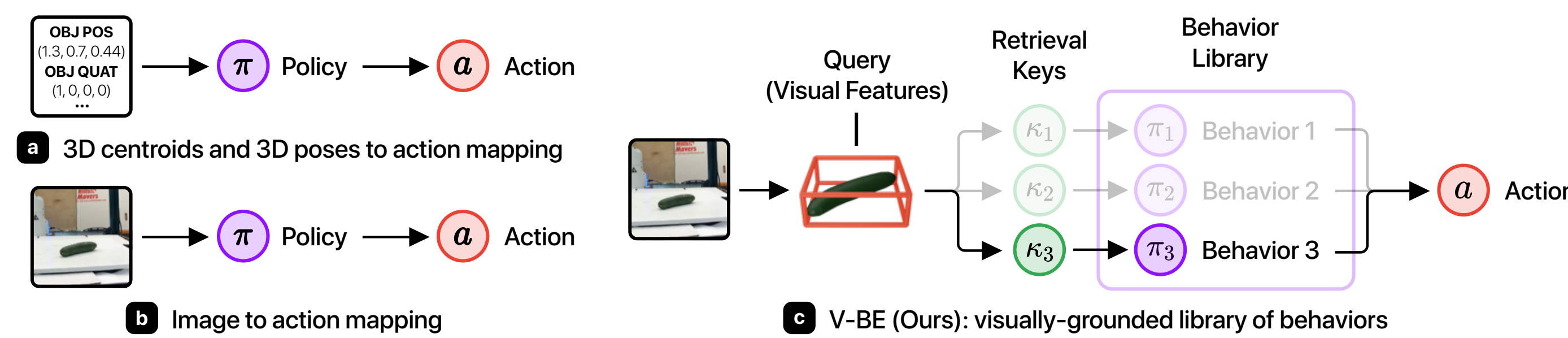
Jingyun Yang*, Hsiao-Yu Tung*, Yunchu Zhang*, Gaurav Pathak, Ashwini Pokle, Christopher G Atkeson, Katerina Fragkiadaki
 Project Website: <https://yjy0625.github.io/projects/v-be/>

Problem Setup



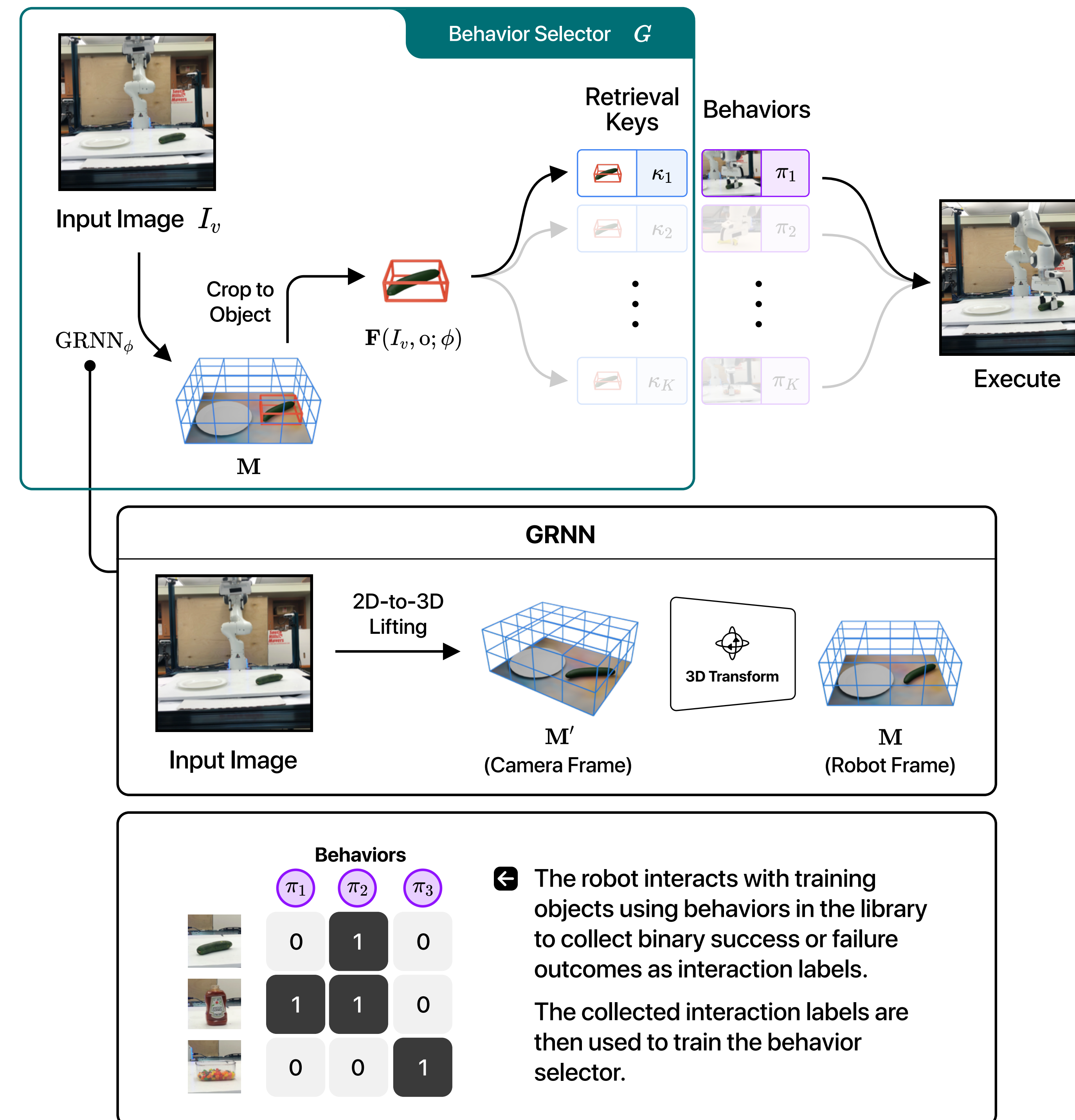
- We propose a method for **manipulating diverse objects across a wide range of initial and goal configurations and camera placements.**

Key Idea



- Previous works use state-to-action or image-to-action mappings as policy representations.
- Our proposed framework decomposes a policy into a **behavior selection module** and a **library of behaviors** to select from.
- This decomposition enables the selection module to work on a different state representation from each behavior in the library of behaviors.
- The selection module operates in a semantically-rich visual feature space, while the behaviors operate in an abstract object state space that facilitate efficient policy learning.

Method

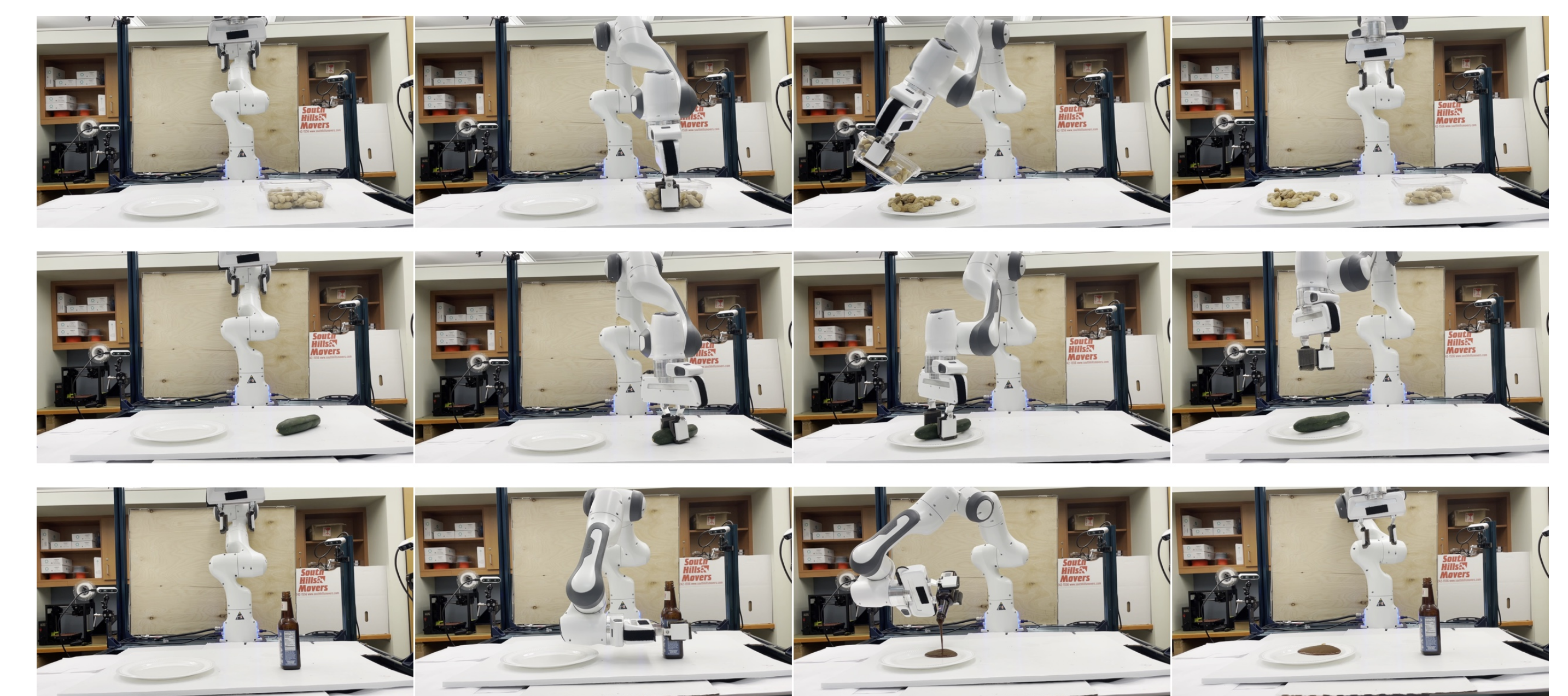
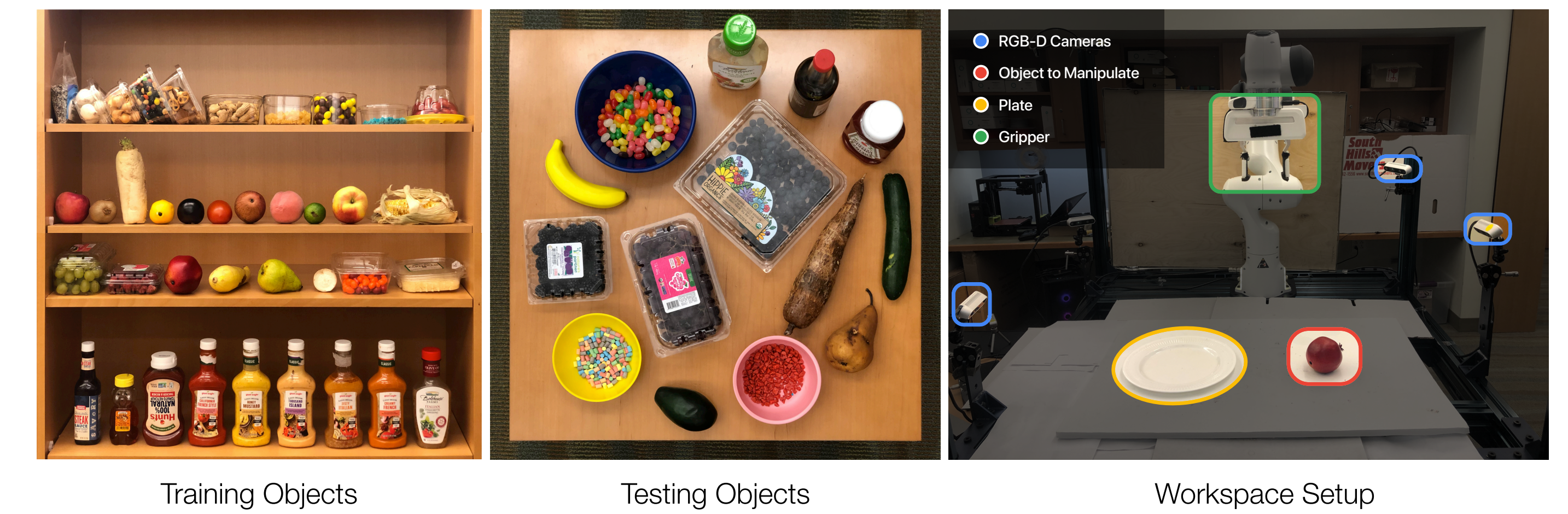


- The behavior selector is trained with a combination of view-prediction loss, occupancy loss, and affordance loss.
- Our behavior library consists of learned skills in pushing and scripted primitive skills in grasping and transporting.

Simulation Experiment Results

	Single Behavior			Library of Behaviors	Ablation study on the selector's visual feature representation	
	Abstract 3D	Abstract 3D + Image	Contextual 3D	V-BEs (Ours)	V-BEs w/ 2D features	V-BEs w/o Fine-tuning on Interaction Labels
grasping	0.30	0.35	0.20	0.78	0.46	0.31
pushing	0.83	0.70	0.10	0.88	0.81	0.46

Real Robot Experiment Results



- Our model achieves **88.6%** success rate on the real robot transporting task, while the *V-BEs with 2D features* baseline achieves **38.0%** success rate.