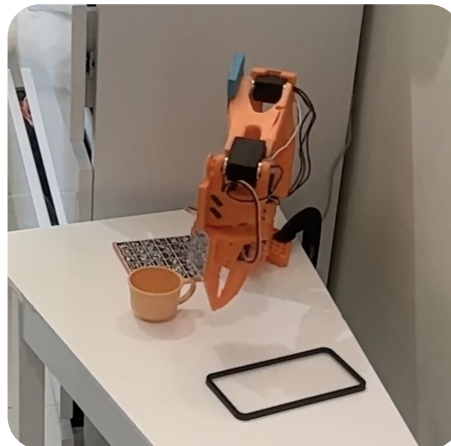
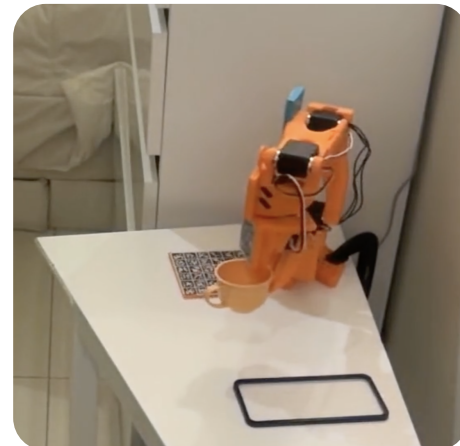


Train: 10 Demos x 4 Viewpoints

Baseline (EEF Regression) RAPS (Ours)



Success: 1/10



Success: 9/10

Moving Hand-Held Camera Mid-Inference (Success)



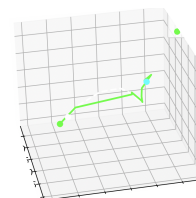
How: Pixel-Aligned Robot Actions Generalize Well



DINOv2



Pred. 2D+Height Traj.
(+ rotation, gripper)



3D Lifted
Trajectory



Robot Actions
via IK