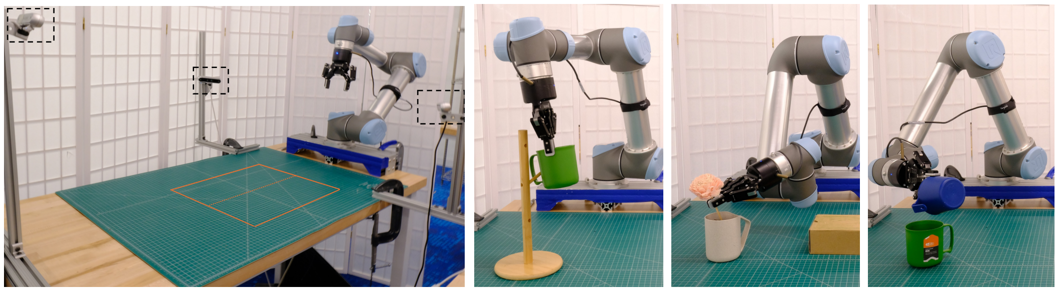
6 Appendix

6.1 Real Robot Experiment



(a) Workspace Settings *(b)* Mug-Tree *(c)* Plug-Flower *(d)* Pour-Ball

Figure 7. Settings and tasks of real-world experiments.

Task	# demos	# pick completions	# place completions	# completions / # trials	success rate
Mug-Tree	10	15/15 (100%)	12/15 (80.0%)	12 /15	80.0%
Plug-Flower	10	15/15 (100%)	14/15 (93.3%)	14/15	93.3%
Pour-Ball	10	14/15 (93.3%)	14/14 (100%)	14/15	93.3%

Table 3. Performance on real-world experiments.

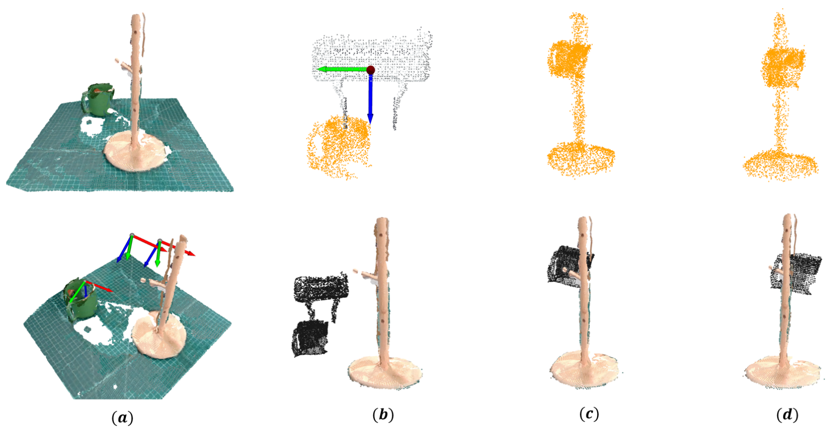


Figure 8. Action inference on *Mug-Tree* with real-sensor data: (a) the observed real-sensor point cloud and the inferred pick, preplace and place action from IMAGINATION POLICY, (b) pick generation, (c) preplace generation, and (d) place generation. The top row shows the generated points with orange color and the bottom row demonstrates the configurations of pick, preplace, and place with the calculated rigid transformations. Please note that we used the point cloud from Franka-Emika Panda gripper to train the model and evaluated it with the Robotiq-85 gripper.

We validated IMAGINATION POLICY on a physical robot. We trained a multi-task agent from scratch on 3 tasks using a total of just 30 demonstrations. There was no use of the simulated data or prrtaining in this experiment - all demonstrations were performed on the real robot.

Settings. The experiment was performed on a UR5 robot with a Robotiq-85 end effector, as shown in Figure 7a. The workspace was a $48\text{cm} \times 48\text{cm}$ region on a table. There were three RealSense 455 cameras mounted pointing toward the workspace. We split the workspace into two parts to place the object and the placement. The segmented point cloud can be directly obtained by cropping the workspace accordingly. To collect the demonstration, we released the UR5 brakes to push the arm physically and record data of the form (initial observation, pick pose, preplace pose, place pose). The combined point cloud P_{ab} was constructed with segmented points and the poses.[2] During testing, we used MoveIt as our path planner to execute the action sequentially.

[2] Please note that we used the point cloud from Franka-Emika Panda gripper to train the pick model.

Tasks. We evaluate IMAGINATION POLICY on three pick and place tasks, as shown in Figure 7bcd. ***Mug-Tree:*** The robot needs to pick up the mug and place it on the mug holder. ***Plug-Flower:*** This task consists of picking up the flower and plugging it into the mug. ***Pouring-Ball:*** The agent is asked to grasp the small blue cup and pour the ball into the big green cup.

Results. We collected 10 human demonstrations of each task. Our model was trained for 200k SGD steps with the same settings as the simulated experiments. We evaluated 15 unseen configurations of each task. The results are reported in Table 3. Visualizations of the captured observation and the generated actions are shown in Figure 8. Videos can be found in supplementary materials. Our failures are mainly caused by the distortion of observations and motion planning errors. For example, the handle of the green mug in *Mug-Tree* task might disappear due to the sensor noise and calibration, which would result in a place failure.

6.2 Detailed Results on RLbench task

We report the results of our method and baselines on RLbench tasks with ± 1.98 std error in Table 4.

Model	# demos	phone-on-base	stack-wine	put-plate	put-roll	plug-charger	insert-knife
IMAGINATION POLICY (ours)	1	4.00 ± 4.52	2.67 ± 2.61	1.33 ± 2.61	2.78 ± 2.72	0	0
IMAGINATION POLICY (ours)	5	78.67 ± 10.45	**97.33 ± 2.61**	0	1.39 ± 2.71	24.00 ± 1.57	38.67 ± 2.61
IMAGINATION POLICY (ours)	10	**90.67** ± 2.61	**97.33 ± 2.61**	34.67 ± 10.45	**23.61 ± 5.44**	**26.67 ± 13.82**	**42.67 ± 9.42**
RVT [2]	10	56.00 ± 4.52	18.67 ± 2.61	**53.33 ± 6.91**	0	0	8.00 ± 4.52
PerAct [1]	10	66.67 ± 11.39	5.33 ± 2.62	12.00 ± 4.52	0	0	0
Diffusor 3D [3]	10	29.33 ± 5.22	26.67 ± 14.55	12.00 ± 0	0	0	0
Discrete Expert		98.67	100	74.6	56	72	90.6

Table 4. Detailed performance comparisons on RL benchmark. Success rate (%) on 25 tests v.s. the number of demonstration episodes (1, 5, 10) used in training. Results are averaged over 3 runs. Even with only 5 demos, our method can outperform existing baselines by a significant margin.