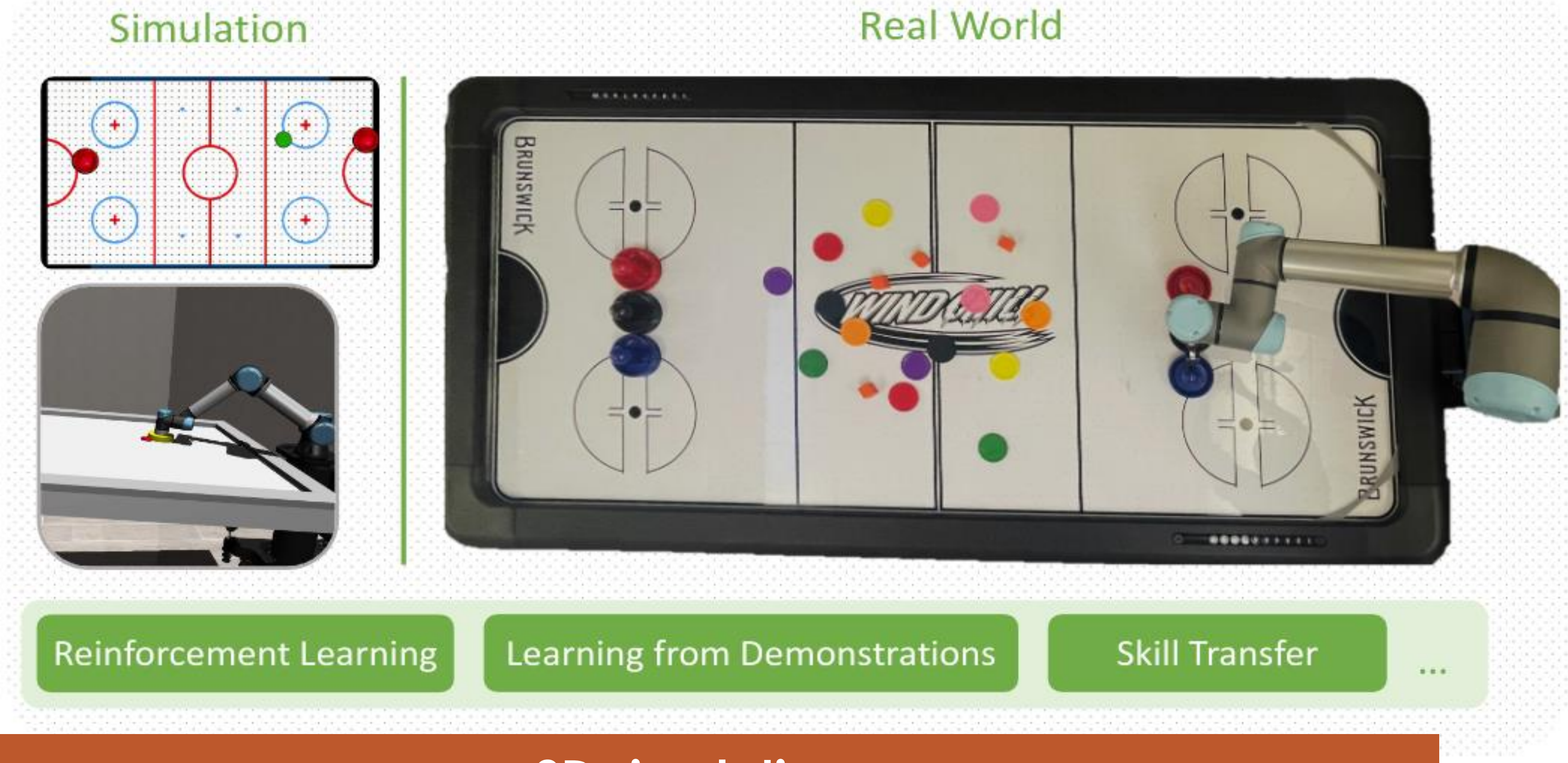
Robot Air Hockey: A Manipulation Testbed for Robot Learning with Reinforcement Learning

Caleb Chuck, Carl Qi, Michael J. Munje, Shuozhe Li, Max Rudolph, Chang Shi, Siddhant Agarwal, Harshit Sikchi, Abhinav Peri, Sarthak Dayal, Evan Kuo, Kavan Mehta, Anthony Wang, Peter Stone*, Amy Zhang, Scott Niekum**

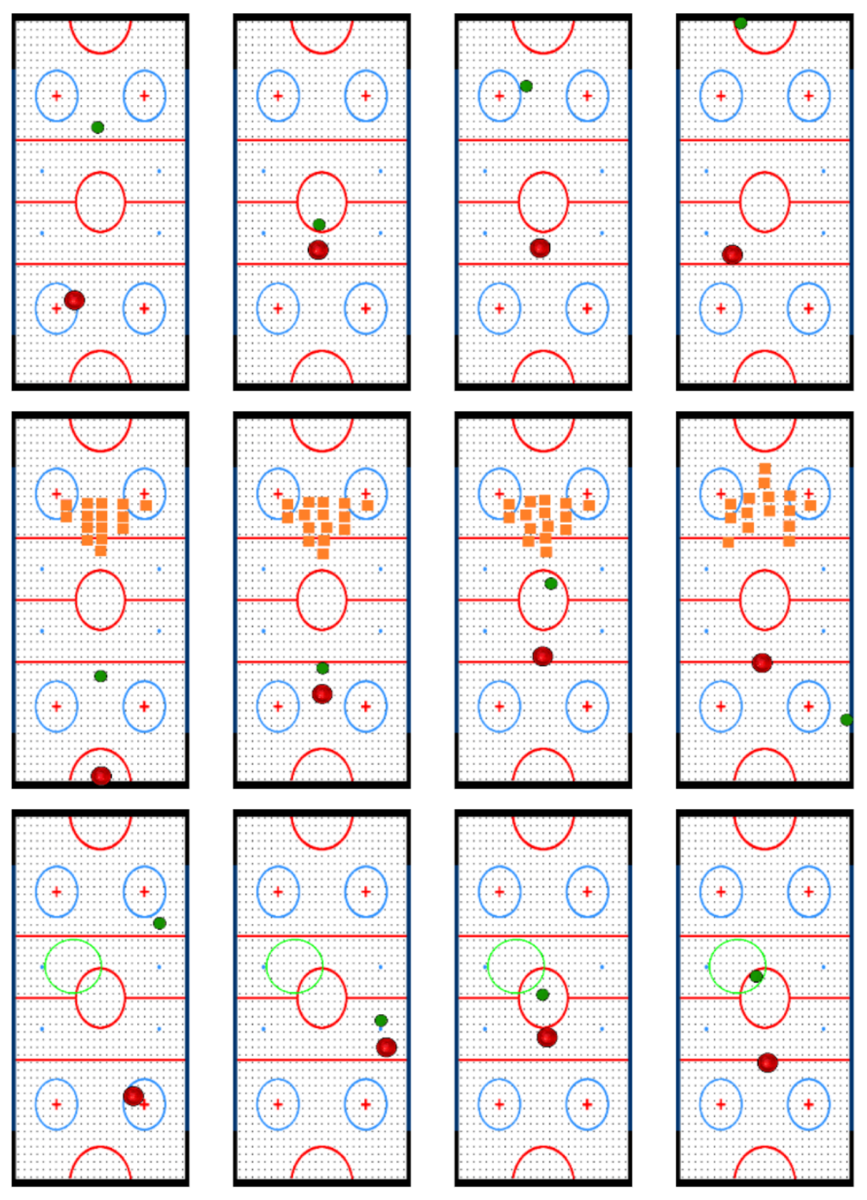
Code: email contact Contact: calebc@cs.utexas.edu Paper:

Overview



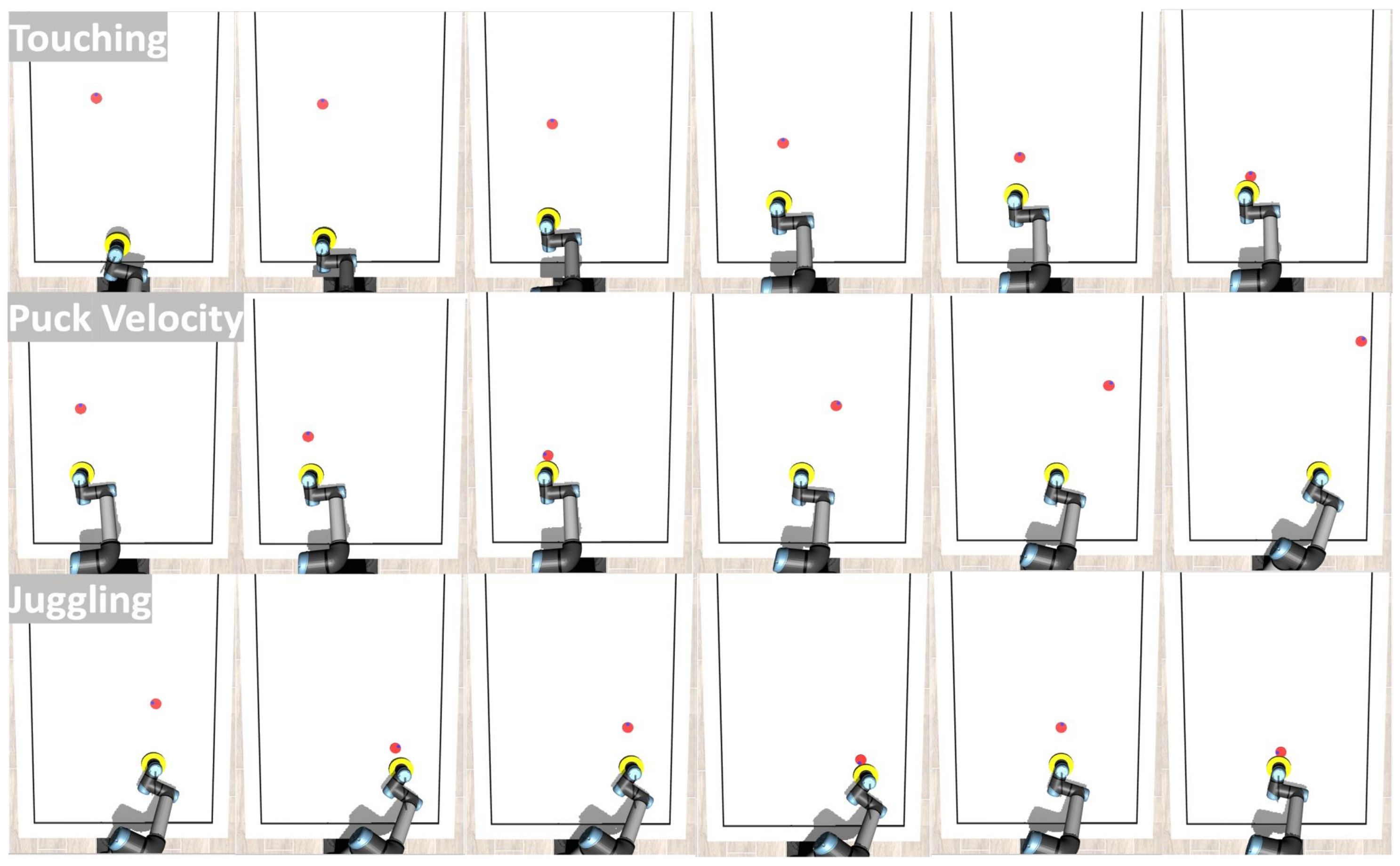
Abstract: Reinforcement Learning is a promising tool for learning complex policies even in fast-moving and object-interactive domains where human teleoperation or hard-coded policies might fail. To effectively reflect this challenging category of tasks, we introduce a dynamic, interactive RL testbed based on robot air hockey. By augmenting air hockey with a large family of tasks ranging from easy tasks like reaching, to challenging ones like pushing a block by hitting it with a puck, as well as goal-based and human-interactive tasks, our testbed allows a varied assessment of RL capabilities. The robot air hockey testbed also supports sim-to-real transfer with three domains: two simulators of increasing fidelity and a real robot system. Using a dataset of demonstration data gathered through two teleoperation systems: a virtualized control environment, and human shadowing, we assess the testbed with behavior cloning, offline RL, and RL from scratch.

2D simulation



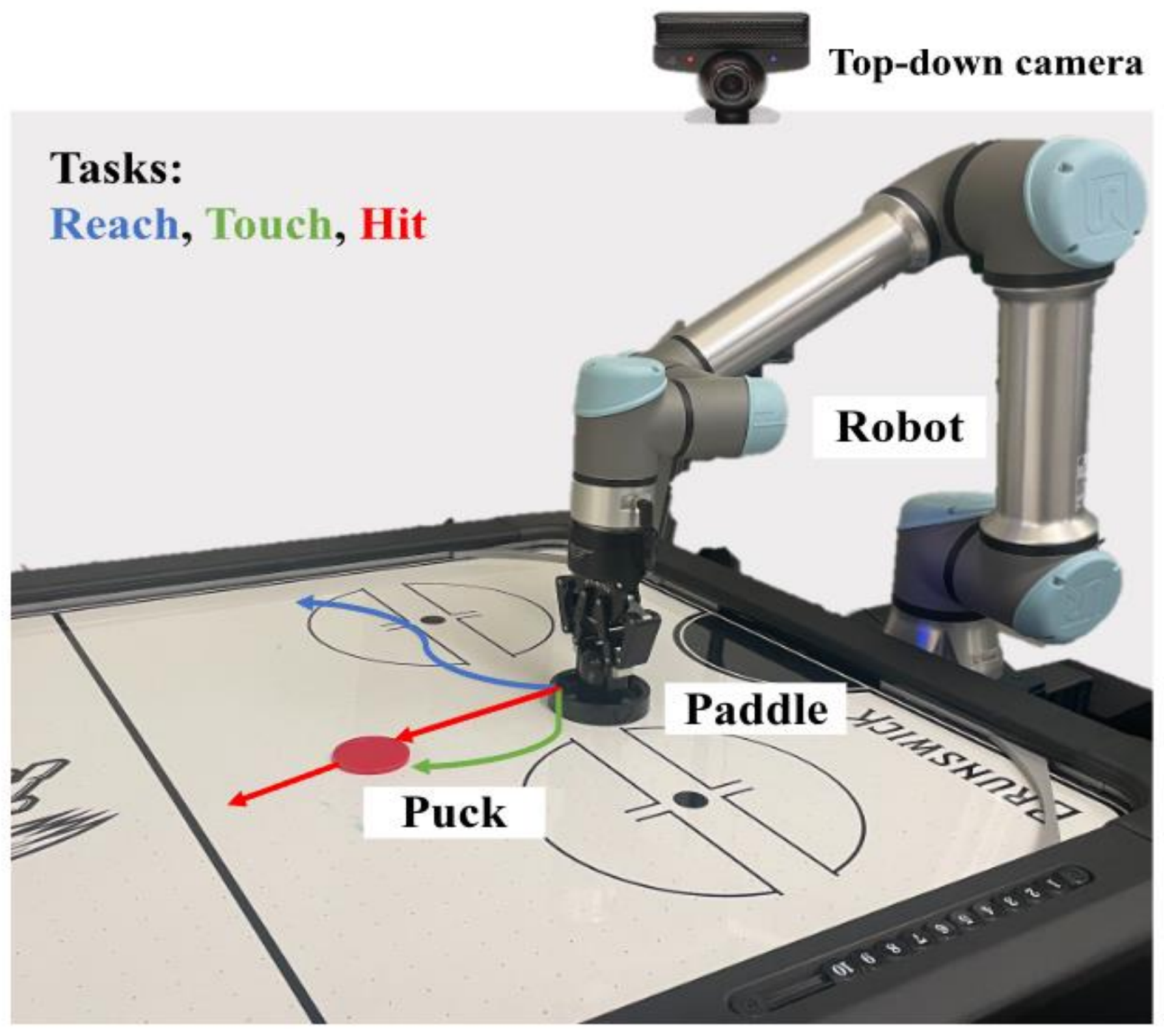
Our 2D simulation environment uses the Python implementation of Box2D as a physics simulation backend. We can assess a wide range of tasks with shaped rewards and penalize actions empirically unrealistic realistic for a robot arm. This environment has many changeable world parameters such as paddle mass, puck mass, dampening, friction, gravity, starting puck velocities, and many more parameters.

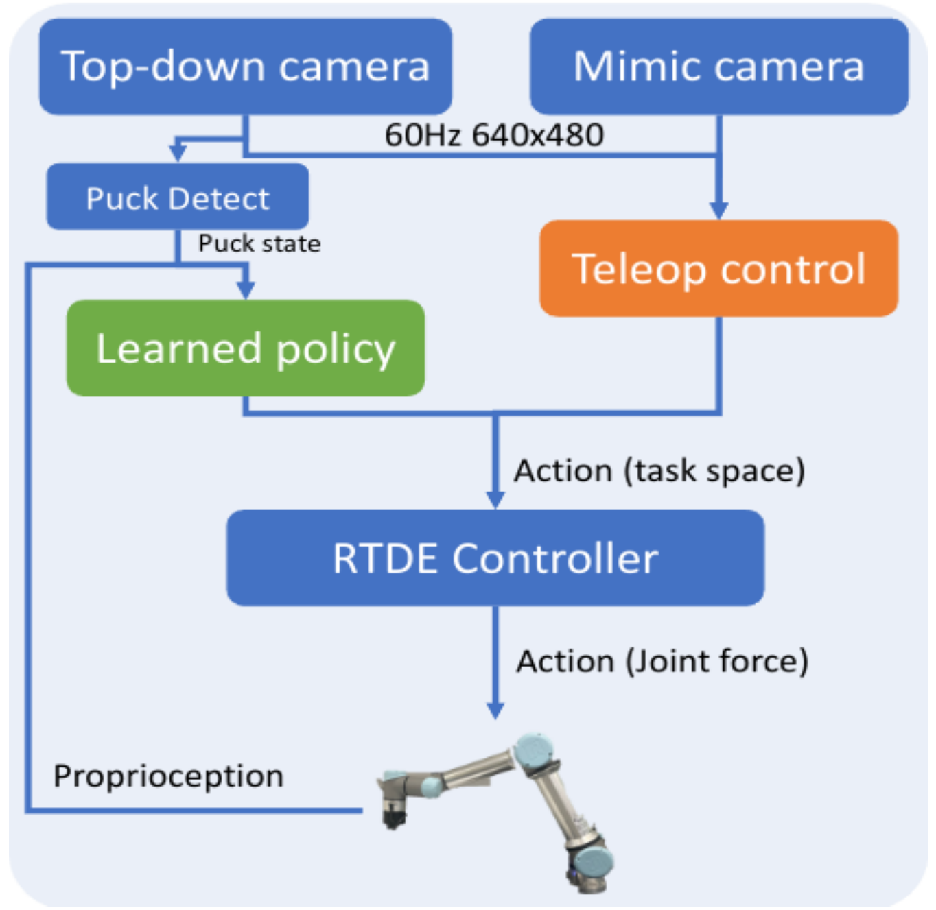
3D simulation



Our 3D simulation environment is a custom Robosuite setting which builds on top of a MuJoCo as a simulation backend. For control, we modify the operation space controller to maintain stable contact with the table. A second simulation environment characteristically different from 2D opens an avenue for assessing transfer through sim-to-sim.

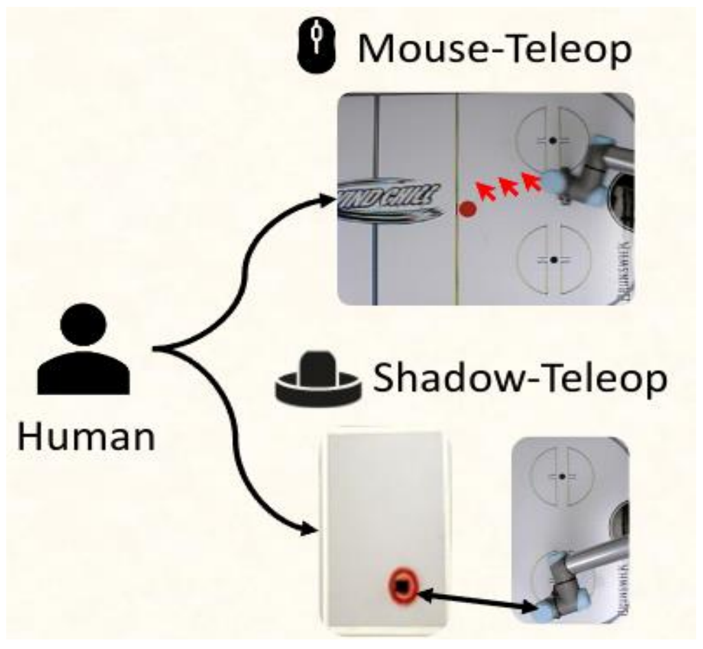
Real Environment

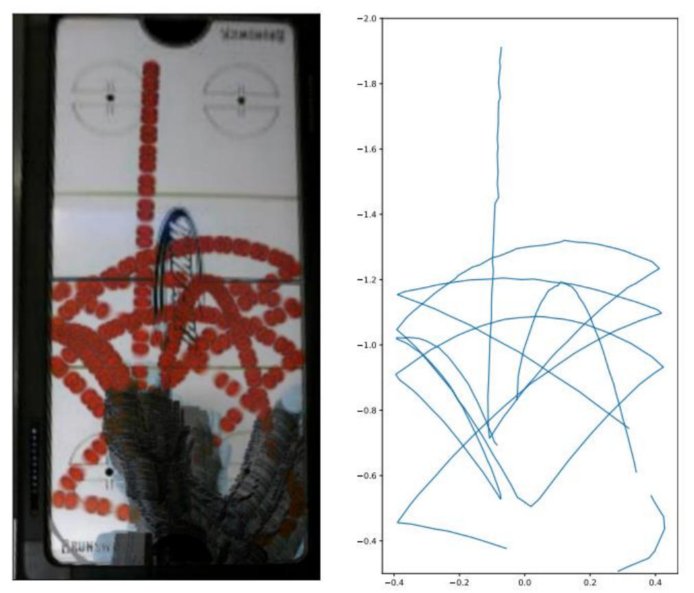




Unlike the simulators, the real world must contend with occlusions, control frequency limits and UR5 emergency stopping when taking strong hard actions. Furthermore, the friction, collision and other dynamics are unknown. Directly transferring a policy from the simulators to the real robot is thus not possible, opening the possibility for future work in sim-to-real transfer.

On the real robot, we provide two teleoperation methods: mouse and human teleoperation to generate a dataset of 800 trajectories gathered from 8 participants of varying skill. Data can be easily generated because of efficient replacing The juggling task require fine grain control that is challenging for human demonstrators—while able to hit the puck at least once, can struggle to achieve multiple hits. Human demonstrates are only able to juggle the puck (perform 4 or more consecutive hits) 30% of the time.





Evaluation

Environment	Method	Robot Air Hockey Tasks									
		Reach	Reach V.	Touch	Strike	Strike Crowd	Juggle	Puck V.	Block	Hit Goal	Hit Goal V.
Box2D [17]	BC	0.9	0.8	**1.0**	0.7	0.3	0.3	0.7	0.0	0.1	0.0
	IQL	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	0.0	0.4	0.0
	RL	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	0.0	**0.9**	0.0
Robosuite [18]	BC	0.9	0.8	0.8	-	-	0.6	0.6	-	0.1	-
	IQL	0.9	0.9	0.8	-	-	0.7	0.8	-	0.1	-
	RL	**1.0**	**1.0**	**1.0**	-	-	**0.9**	**0.9**	-	**0.2**	-
Real World	BC	0.9	**0.1**	0.3	-	-	-	0.1	-	-	-
	IQL	**1.0**	0.0	0.6	-	-	-	0.3	-	-	-
	Human	**1.0**	0.0	**1.0**	-	-	**0.3**	**1.0**	-	-	-

Overall, online RL performs the best among the baselines in simulation, showing that online interactions are crucial for solving our dynamic tasks. In the real world, all of our baselines fall short to human performance, leaving room for potential improvements for future work. Directly applying RL to the real world is infeasible both in sample efficiency, and random jitter is ineffective on a robot arm. However, offline RL notably outperforms behavior cloning, which suggests that dynamic, interactive tasks benefit from a reward signal to learn fine-grained control behaviors like hitting a moving puck.

SCALAR LAB Safe, Confident, and Aligned Learning + Robotics
MIDI lab
LARG Learning Agents Research Group The University of Texas at Austin
TEXAS The University of Texas at Austin
TEXAS Robotics
University of Massachusetts Amherst
Sony AI

Affiliations:

- UT Austin, * Also Sony AI, ** UMass Amherst