Robust Grasping via Human Adversary

Jiali Duan*, Qian Wang*, Lerrel Pinto, C.-C. Jay Kuo, Stefanos Nikolaidis
MCL & ICAROS Lab, University of Southern California
Carnegie Mellon University

Problem and Motivation

We formulate the problem as a *two-player game with incomplete information*, played by a human (H), and a robot (R).

Notations:

$s \in S$: state of the world
$s^+ \in S^+$: state after robotic action
$s^{++} \in S^{++}$: state after human action
$\mathcal{T}: S \times A^R \to \Pi(S^+)$: transition
$\mathcal{T}: S^+ \times A^H \to \Pi(S^{++})$: transition
$\pi^R: (s, a^R)$: robot action
$\pi^H: (s^+, a^H)$: human action

Rewards:

$r: (s, a^R, s^+, a^H, s^{++}) \to r$: reward
$r = R^R(s, a^R, s^+) - \alpha R^H(s^+, a^H, s^{++})$

Goal:

$\pi^R_* = argmax_{\pi^R} \mathbb{E}[r(s, a^R, a^H | \pi^H]$

An overview of our framework for a robot learning robust grasps by interacting with a human adversary.

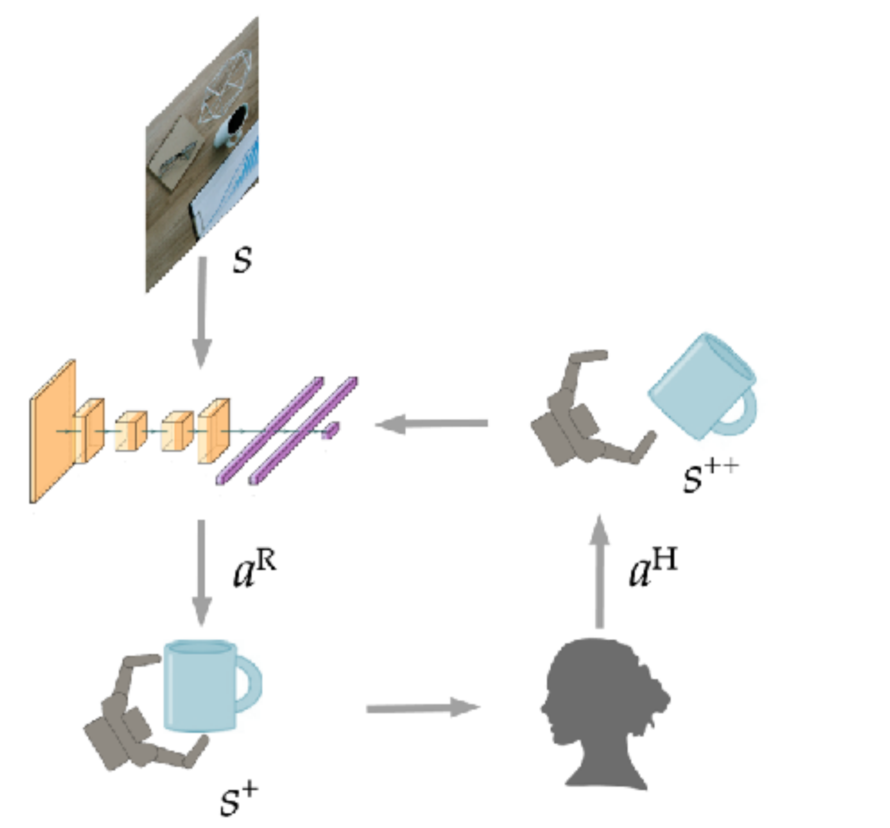

Fig.1: Overall training framework with human-in-the-loop

Existing works have explored cases where human acts as a supervisor that assists the robot. In reality, human observers tend to also act in an ***adversarial*** manner towards robotic systems. How can we leverage human adversarial actions to improve robustness of learned policies?

- **Pioneering Work on Human Adversarial Actions.** To the best of our knowledge, this is the first effort of robot learning with adversarial human users. In a manipulation task, we show that grasping success improves significantly when the robot trains with a human adversary as compared to training in a self-supervised manner.
- **Learning Robust Grasps.** By jointly training robot arm with human adversary, we show that it can lead to robust grasping solutions. We use self-supervised training and joint-training with simulated adversary as our baselines.
- **Comprehensive User Study.** We proposed two hypothesis and verified them by conducting a comprehensive user study involving 25 users. We plot success rates before/after applying human adversary as well as plots concerning action selected over time for different users.
- **Simulation Environment.** For the training, we developed a customized simulation environment based on Mujoco that allows a human user interacting with the physics engine.

A Framework for Robot Learning with Human Adversary

Pipeline: Robot Learning with Adversarial Human Actions

Algorithm 1 Learning with a Human Adversary

```
Initialize parameters W of robot's policy π^R
for batch = 1, B do
    for episode = 1, M do
        observe s
        sample action a^R ~ π^R_*(s)
        execute action a^R and observe s^+
        if s^+ is not terminal then
            observe human action a^H and state s^++
        observe r given by Eq. (1)
        record s, a^R, r
    update W based on recorded sequence
return W
```

Our Approach:

- **Adversarial Disturbance:** After the robot grasps an object successfully, the human can attempt to pull the object away from the robot's end-effector, by applying a force through our user-interactive interface, as shown in Fig.2.
- **Network Architecture:** We use a fully-connected ConvNet architecture similar to AlexNet as shown in Fig. 3. The network takes image as input and outputs grasping location and angle: (x_g, y_g, θ).
- **Network Training:** We initialized with a pertained model released by Pinto et al. The model was pretrained with different objects and patches. To train the model, we treat the reward r that the robot receives as a training target for the network. Specifically, we set $R^R(s, a^R, s^+) = 1$ if the robot succeeds and 0 if the robot fails. Similarly, $R^H(s^+, a^H, s^{++}) = 1$ if human succeeds and 0 if human fails. We then calculate cross-entropy loss between the network's prediction and the reward received and optimized with RMSProp.

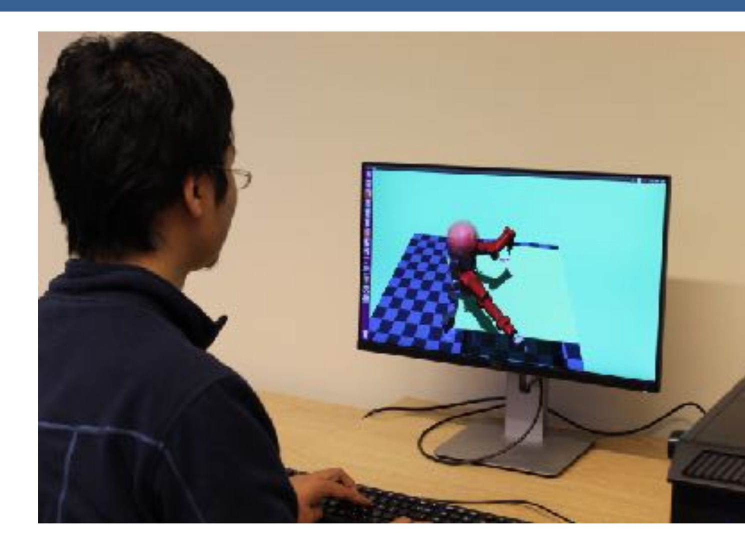

Fig.2: Participants interacted with our user-interface

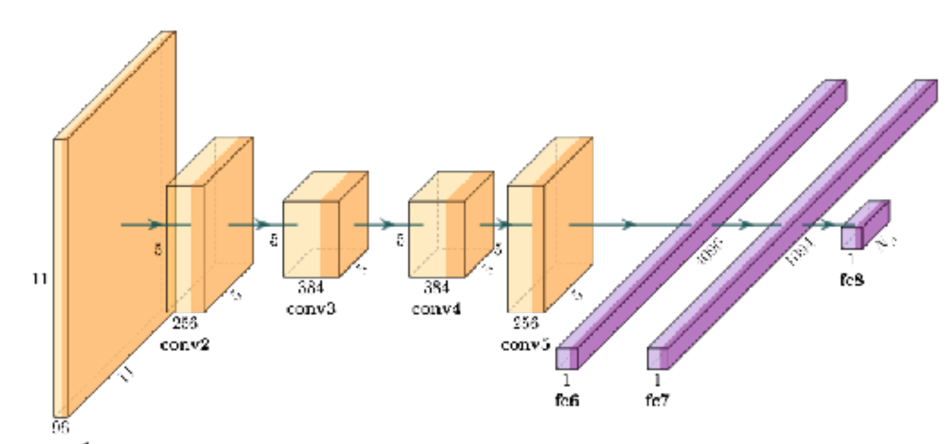

Fig.3: Neural network structure for grasping policy

Experimental Results: From Theory to Users

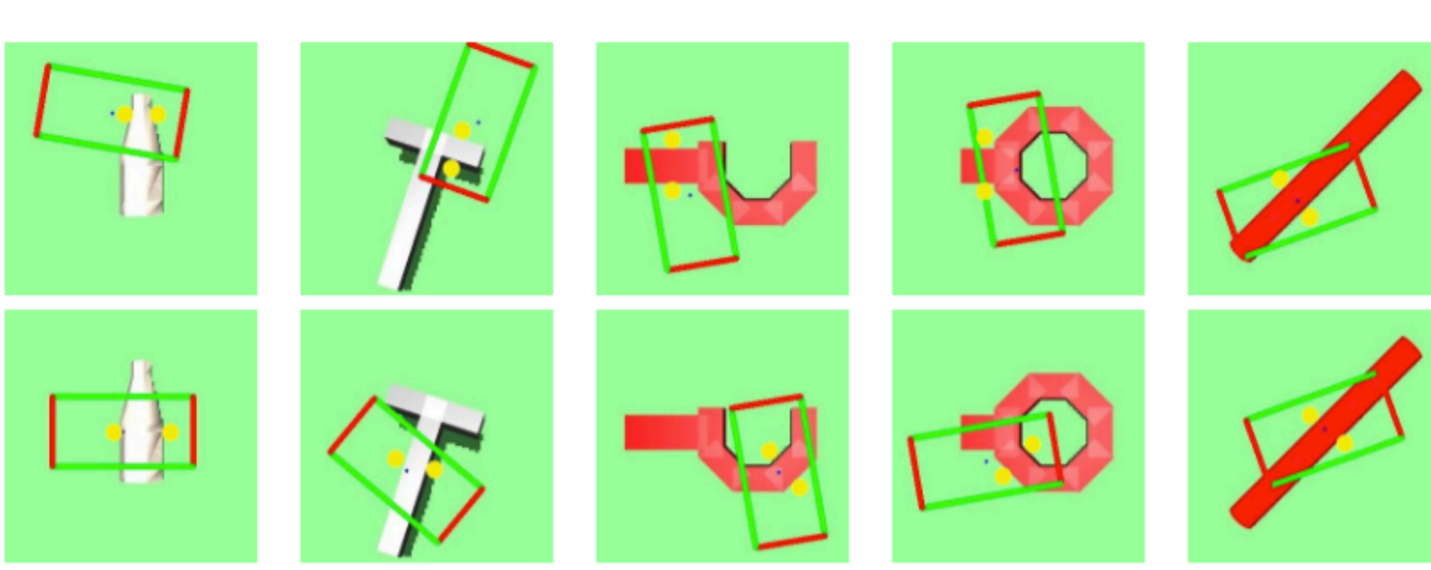

Fig.4: Selected grasp predictions before (top row) and after (bottom row) training with the human adversary.

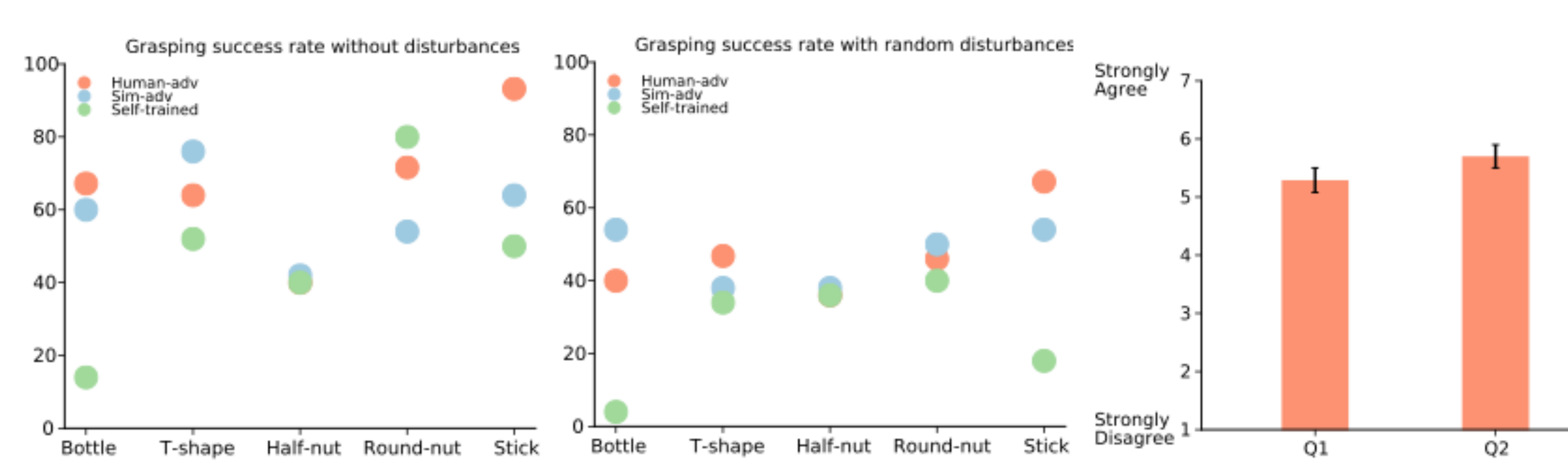

Fig.5: Success rate without/with random disturbances (Left two). Subjective metrics (Right)

- **Grasping Prediction:** In Fig.4, the red bars show the open gripper position and orientation, while the yellow circles show the grasping points when the gripper has closed.
- **Evaluation Metrics:** Fig.5 shows both the quantitative evaluation metric (Left two figures) as well as qualitative evaluation metric (Rightmost figure). A two-way multivariate ANOVA with object and framework as independent variables showed a statistically significant interaction effect for both measures: ($F(16,38)=3.07$, $p=0.002$, Wilks' $\Lambda = 0.19$). A Post-hoc Tukey tests with Bonferroni correction showed that success rates were significantly larger for the human adversary condition than the self-trained condition, both with ($p<0.001$) and without random disturbances ($p=0.001$). In subjective test, we asked user to evaluate if the robot learned throughout the study and if the performance of robot improved throughout the study.
- **Action Distribution:** Fig.7 shows the disturbances applied over time for different users. Observing the participants behaviors, we see that some participants *used their model of the environment to apply disturbances effectively.*

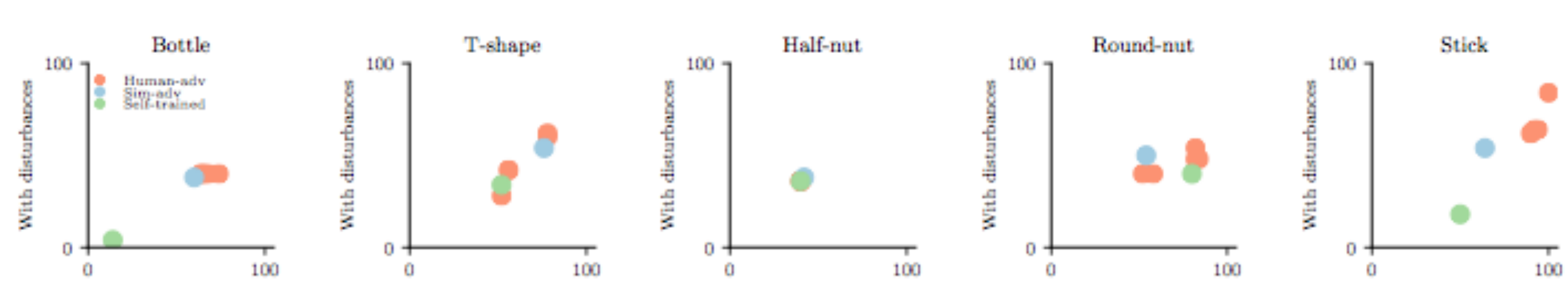

Fig.6: Success rates for each object with (y-axis) and without (x-axis) random disturbances for all participants

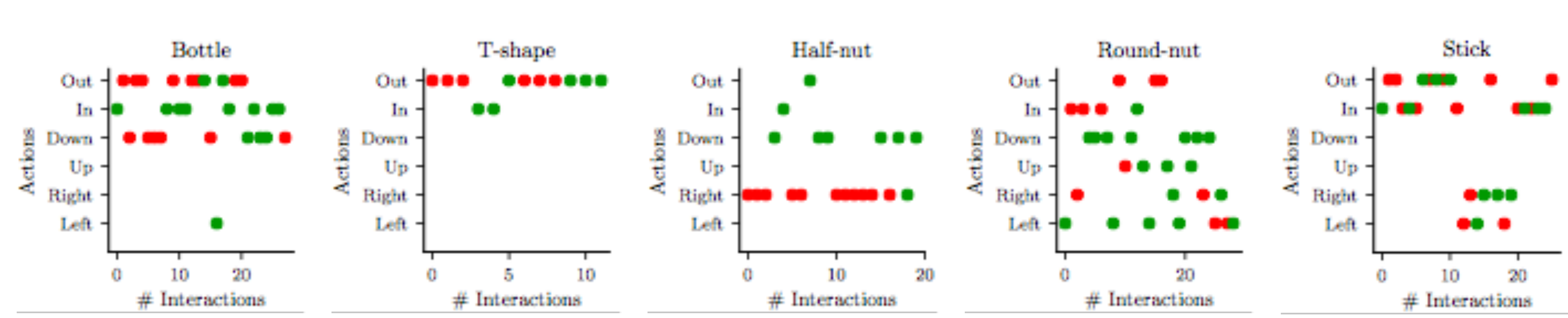

Fig.7: Actions applied by selected human adversaries over time. Red dot denotes successful grasping and green fails