Robust Grasping via Human Adversary

Jiali Duan*, Qian Wang*, Lerrel Pinto, C.-C. Jay Kuo and Stefanos Nikolaidis

***Abstract*—Much work in robotics has focused on "human-in-the-loop" learning techniques that improve the efficiency of the learning process. However, these algorithms have made the strong assumption of a *cooperating* human supervisor that assists the robot. In reality, human observers tend to also act in an *adversarial* manner towards deployed robotic systems. We show that this can in fact improve the robustness of the learned models by proposing a physical framework that leverages perturbations applied by a human adversary, guiding the robot towards more robust models. 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.**

***Index Terms*—human-robot interaction, human adversary, simulated adversary, self-supervised training**

I. Introduction

We focus on end-to-end learning of robust manipulation grasps that can withstand perturbations using input images from an on-board camera.

Our main motivation is *how can we leverage human adversarial behaviors to improve robustness of the learned policies, given human domain knowledge?* While there has been a rich amount of human-in-the-loop learning, to the best of our knowledge this is the first effort of robot learning with *adversarial* human users. We propose a "human-adversarial" framework where a robotic arm collects data for a manipulation task, and a human user attempts to make the robot learner fail. For instance, if the learner grasps an object, the human can apply forces or torques to remove it from the robot. Contrary to a robot adversary in previous work [1], the human may already have domain knowledge about the best way to attempt the grasp, by observing the grasp orientation and their prior knowledge of the object's geometry and physics.

II. Overview of Framework

The overall pipeline is shown in Fig. 1. We define s to be state of the world. A robot and a human are taking turns in actions. A robot action results in a stochastic transition to new state $s^+ \in S^+$. The human then acts based on a stochastic policy, also unknown to the robot, $\pi^{\mathrm{H}} : (s^+, a^{\mathrm{H}})$. After the human and the robot's actions, the robot observes the final state s^{++} and receives a reward signal $r : (s, a^{\mathrm{R}}, s^+, a^{\mathrm{H}}, s^{++}) \mapsto r$.

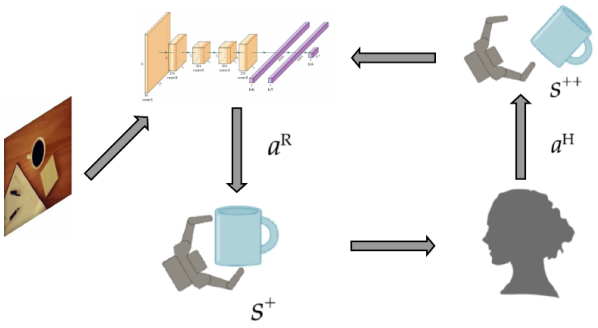

Fig. 1. An overview of our framework for a robot learning robust grasps by interacting with a human adversary.

Specifically, we formulate r as a linear combination of two terms: the reward that the robot would receive in the absence of an adversary, and the penalty induced by the human action:

$$r = R^{\mathrm{R}}(s, a^{\mathrm{R}}, s^+) - \alpha R^{\mathrm{H}}(s^+, a^{\mathrm{H}}, s^{++}) \tag{1}$$

The goal of the system is to develop a policy $\pi^{\mathrm{R}} : s \mapsto a_t^{\mathrm{R}}$ that maximizes this reward.

$$\pi_*^{\mathrm{R}} = \underset{\pi^{\mathrm{R}}}{\operatorname{argmax}}\ \mathbb{E}\left[r(s, a^{\mathrm{R}}, a^{\mathrm{H}})|\pi^{\mathrm{H}}\right] \tag{2}$$

Through this maximization, the robot implicitly attempts to minimize the reward of the human adversary.

III. Experiments

We conducted a user study, with 25 participants (5 users for each object) interacting with the robot in our customized physical engine[1] (Fig. 2 (Middle)) built upon mujoco [2]. The purpose of our study is to test whether the robustness of the robot's grasps can improve when interacting with a human adversary. We are also interested to explore how 5 objects (with varying difficulty and geometry) affects the adversarial strategies of the users, as well as how users perceive robot's performance [2]. We verify the following hypotheses:

H1. Robot trained with the human adversary will perform better than the robot trained in a self-supervised manner.

H2. Robot trained with the human adversary will perform better than the robot trained with a simulated adversary.

Objective metrics. Fig. 4 shows the mean success rates for different objects. We have two dependent variables, the success

* Duan and Wang contributed equally to the work.

Duan and Kuo are with the Department of Electrical and Computer Engineering, University of Southern California, Los Angeles 90089, USA. (e-mail: jialidua@usc.edu, cckuo@sipi.usc.edu).

Wang and Nikolaidis are with the Department of Computer Science, University of Southern California, Los Angeles 90089, USA. (e-mail: {wang215, nikolaid}@usc.edu).

Pinto is with the Robotics Institute, Carnegie Mellon University, Pittsburgh 15213, USA. (e-mail: lerrelp@cs.cmu.edu).

[1]Code available at: https://github.com/davidsonic/Interactive-mujoco_py

[2]The anonymized log files of the human adversarial actions are publicly available at https://github.com/davidsonic/human_adversarial_grasping_data

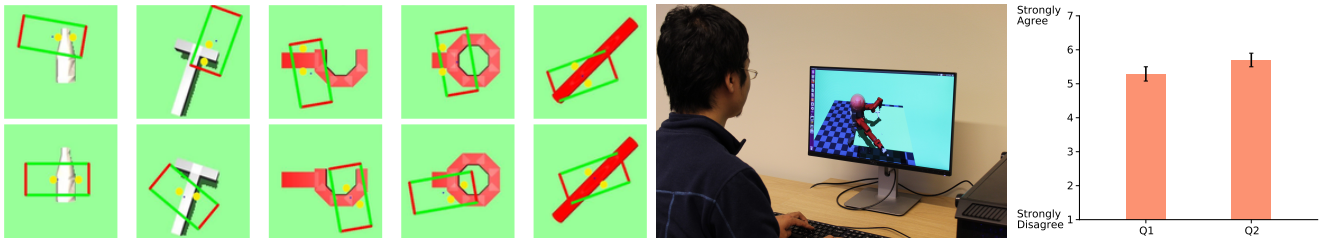

Fig. 2. **Left:** Selected grasp predictions before (top row) and after (bottom row) training with the human adversary. The red bars show the open gripper position and orientation, while the yellow circles show the grasping points when the gripper has closed. **Middle:** Participants interacted with a simulated Baxter robot in the customized Mujoco simulation environment. **Right:** Subjective metric from user study (q1: The robot learned throughout the study; q2: The performance of the robot improved throughout the study).

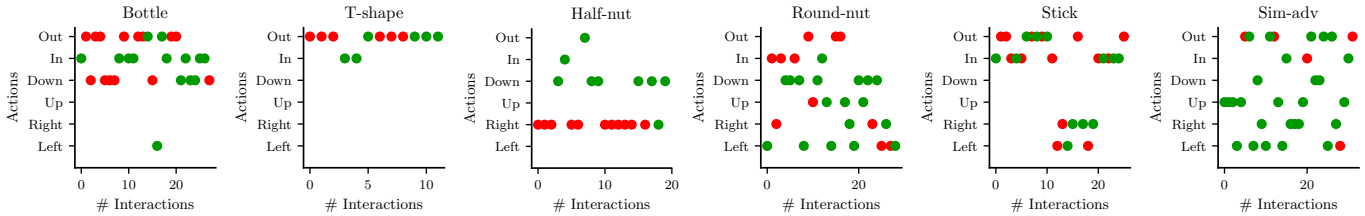

Fig. 3. Actions applied by selected human adversaries (first 5) over time. We plot in green adversarial actions that the robot succeeds in resisting, and in red actions that result in the human 'snatching' the object. The last plot compares simulated adversary for stick object.

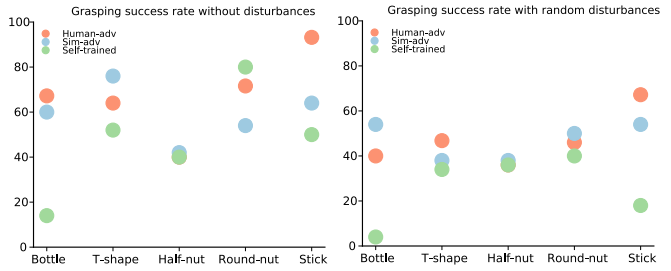

Fig. 4. Success rates between 3 different frameworks (human adversary/simulated adversary/self-supervised training) for all five participants (w/wo disturbances).

rate of robot grasping an object in the testing phase in the absence of any perturbations, and the success rate with random perturbations being applied. A two-way multivariate ANOVA with object and framework as independent variables showed a statistically significant interaction effect for both dependent measures: $(F(16, 38) = 3.07, p = 0.002,$ Wilks' $\Lambda = 0.19)$. In line with **H1**, 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 $(p = 0.001)$.

The success rate averaged over all human adversaries was higher for three out of five objects. The reason is it was easy for the self-trained policy to pick up these objects without a robust grasp while the network trained with the human adversary rejected these unstable grasps, and learned quickly robust grasps for these objects. In contrast, round nut and half-nut objects could be grasped robustly at the curved areas of the object. The self-trained network thus got "lucky" finding these grasps, and the difference was negligible. In summary, *Training with a human adversary is particularly beneficial for objects that have few robust grasp candidates that the network needs to search for.*

Fig. 3 shows the disturbances applied over time for different users. The first 3 plots show that human adversarial quickly explored successful perturbations by exploring object geometry, which is indicated by the red dots. Gradually, the robot learned a more robust grasping policy, which resulted in the user failing to snatch the object (green dots). The 4th one shows the user adapted their strategy as well when the robot learned to withstand adversarial actions. The 5th and 6th plot compare the user with the simulated adversary for the same object (stick). We observe that the simulated adversary explores different perturbations that are unsuccessful in snatching the object, which translates to worse performance for that object in the testing phase.

IV. Conclusion

Our work shows that human-in-the-loop adversarial learning can be leveraged to improve the robustness of robotic grasping, because humans can understand stability and robustness better than learned adversaries. We conclude our analysis with reporting the users' subjective responses (Fig. 2 (right)). A Cronbach's $\alpha = 0.86$ showed good internal consistency. Participants generally agreed that the robot learned throughout the study, and that its performance improved.

References

[1] L. Pinto, J. Davidson, and A. Gupta, "Supervision via competition: Robot adversaries for learning tasks," in *Robotics and Automation (ICRA), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1601–1608.

[2] E. Todorov, T. Erez, and Y. Tassa, "Mujoco: A physics engine for model-based control," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012, pp. 5026–5033.