2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

Robot Learning in Mixed Adversarial and Collaborative Settings

Seung Hee Yoon and Stefanos Nikolaidis
Department of Computer Science
The University of Southern California
Los Angeles, CA 90089
yoon@usc.edu, nikolaid@usc.edu

***Abstract*— Previous work has shown that interacting with a human adversary can significantly improve the efficiency of the learning process in robot grasping. However, people are not consistent in applying adversarial forces; instead they may alternate between acting antagonistically with the robot or helping the robot achieve its tasks. We propose a physical framework for robot learning in a mixed adversarial/collaborative setting, where a second agent may act as a collaborator or as an antagonist, unbeknownst to the robot. The framework leverages prior estimates of the reward function to infer whether the actions of the second agent are collaborative or adversarial. Integrating the inference in an adversarial learning algorithm can significantly improve the robustness of learned grasps in a manipulation task.**

I. Introduction

We address the problem of end-to-end learning robust manipulation grasps in robotics. For instance, we would like a robotic arm to learn how to grasp robustly a brush, to perform a hair brushing task to a user using an on-board camera.

This is challenging and it requires a large number of samples. For instance, to achieve high accuracy by learning in a self-supervised manner, researchers have performed thousands of grasping attempts [1], which involved up to 14 manipulators over the course of two months [2]. Training in a self-supervised manner requires a significant amount of resources. Importantly, the feedback to the system is a binary signal of grasping success or failure, rendering the system unable to distinguish between stable or unstable grasps.

Recent work has leveraged robotic and human adversaries to improve sample efficiency and grasping stability. Specifically, researchers [3] showed that a robotic adversary that attempt to learn how to "snatch" objects away from the learner can improve sample efficiency. Our previous work [4] has shown that human adversaries have proven to be particularly good in enabling the robot to learn stable grasps since they have an intuitive notion of grasp stability and robustness. They interacted with the robot through a user interface, removing objects from the robot's gripper by selecting a direction that the force would be applied. This guided the learning algorithm towards more stable grasps.

However, people did not act consistently in an adversarial manner. Our previous user studies have shown that people occasionally applied forces that moved the object towards more *stable* grasps, despite being instructed to act adversarially. In a post-experimental interview, one participant stated that "it seems some perturbations were challenging; so after some time I didn't apply that perturbation again." This indicates that the participant did not follow our instructions and wanted to assist the robot instead.

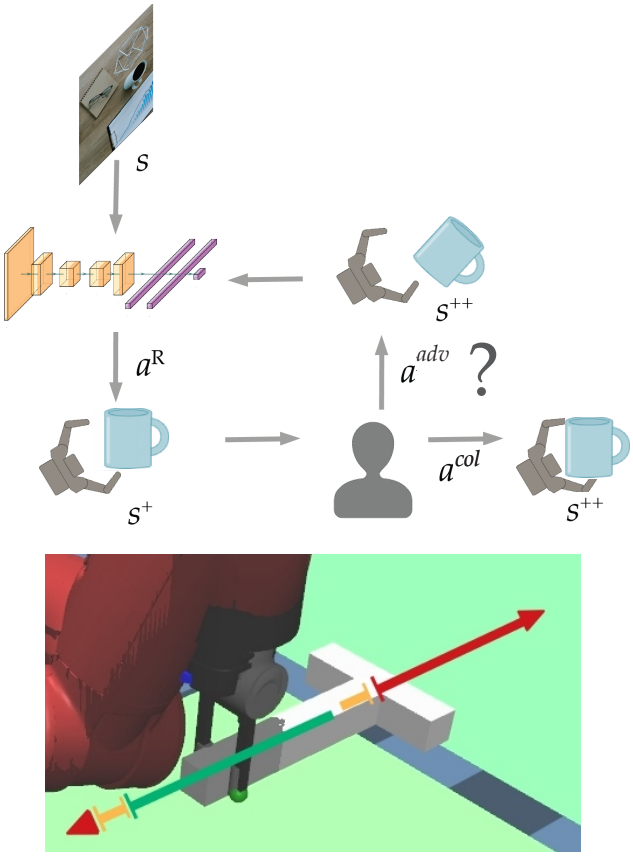

Fig. 1: In a mixed cooperative setting, an agent may act adversarially or collaboratively, unbeknownst to the robot. By evaluating the object's configuration before an external force is applied, the robot can identify adversarial forces and use them to learn robust grasps.

This indicates that robots should not rely on the strong assumption of a consistently adversarial agent. Instead, they should distinguish between adversarial and collaborative actions, and use this information in their learning.

This work explores how the robot can learn in a mixed adversarial/collaborative setting, when the second agent acts sometimes adversarially and sometimes collaboratively, without the robot knowing the agent's role at a given time.

Specifically, we focus on the grasping problem and we address the following research question:

> How can we enable effective learning of robust grasps in the presence of both adversarial and collaborative forces?

Our key insight is:

> Adversarial forces will reduce grasp quality, which can be estimated quite accurately with a prior learning experience, while collaborative forces will improve it.

We propose a generalized framework, where a robotic arm collects data for a grasping task, interacting with a second agent applying disturbances to the grasped object. When the second agent applies a force, the system evaluates the configuration of the object in the robot's gripper before and after the force has been applied, using the prior estimate of grasping quality, as if the robot were to grasp the object from these configurations. If the grasping quality has improved, the system classifies this as a collaborative force and ignores it. If the grasping quality has failed, or the object is no longer grasped, the system classifies it as an adversarial force and updates the learned policy, guiding the robot towards more robust grasps.

We implement the framework in a virtual environment, where we apply randomly generated forces on different objects grasped by a robotic arm. We show that, compared to treating all forces as adversarial, the robot can learn significantly more robust grasping policies.

While there are limitations on the implementation in the virtual environment, we view this as an exciting step towards enabling robot learning in mixed adversarial and collaborative settings.

II. Related Work

We focus on enabling the robot to grasp objects robustly, with a small number of interactions. Previous work [5], [6] on grasping ranges from physics-based modeling [7]–[10] to data-driven techniques [1], [2]. Deep learning combined with self-supervision [2], [3], [11]–[13] analytical techniques [14]–[16] and domain adaptation [17], [18] has been particularly effective in this domain. In previous work [19], a robot learner interacts with a robotic adversary, which attempts to either snatch the objects away from the robot's gripper or shake the gripper to remove the object. Our own work [4] has shown that a human adversary can be particularly effective in removing objects away from the robot in a virtual environment and therefore guiding the robot towards more robust grasps. Both these works assume that the second agent acts consistently in an antagonistic manner.

A significant amount of research [20]–[26] has focused on learning through positive or negative feedback given by a human supervisor who observes the agent's learning process. There has been significant work on how to best interpret the human input, which sometimes is meant to encourage future behaviour rather than reward or punish previous actions. In our work, the agent applies forces directly to the robot's gripper, which can be either collaborative or adversarial.

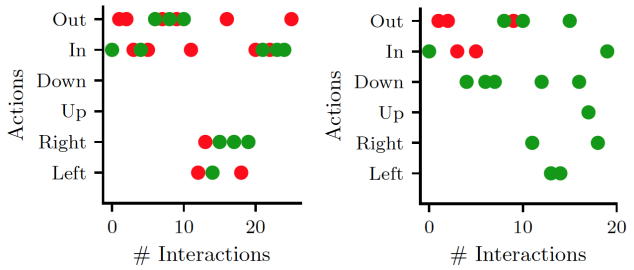

Fig. 2: Difference between training with a human acting consistently as an adversary and a human that starts assisting the robot in the later parts of the interaction in our previous user study [4]. The red dots indicate human success in snatching the object, while the green dots indicate robot success in withstanding the human perturbation.

Our work also bears resemblances to generative adversarial methods [27]–[29] which can increase classification performance. We use estimates of the grasping configuration to estimate whether the input to the discriminative model is adversarial or not, before adopting the model.

III. Categorizing External Forces

Motivation

Our previous work has shown that applying adversarial forces can improve the learning of robust grasps. However, in our user study people did not always apply adversarial forces, even when they were instructed to do so.

Fig. 2 shows the external forces applied by a participant in the study who acted consistently adversarially, contrasted with the forces applied by a participant that started acting adversarially but then assisted the robot. The robot's performance during testing was significantly worse for the second participant since the learning was guided towards more robust grasps, misinterpreting the human actions as adversarial.

But how can the robot infer whether a force is adversarial or collaborative?

Clearly, if a certain force successfully removes an object from the robot's gripper, no matter in what direction, or how strong it is, this should be an "adversarial force." The robot should interpret this as feedback that the grasp is unstable and update the policy towards more stable grasps.

However, when an object still remains in the gripper after the applied force, the situation is more complicated. If the agent applying the force is assumed to be acting adversarially, this would imply that the grasp is robust and the robot should receive a signal reinforcing the current grasp.

On the other hand, if the force was collaborative, then interpreting it as adversarial might guide the learning towards unstable grasps.

Therefore, one has to distinguish between different types of forces. We propose doing this by observing the effect of the force on the object's configuration.

Categorization Factors

A collaborative force may differ from an adversarial one based on a number of factors, such as direction, magnitude and duration. One would also need to factor the presence of

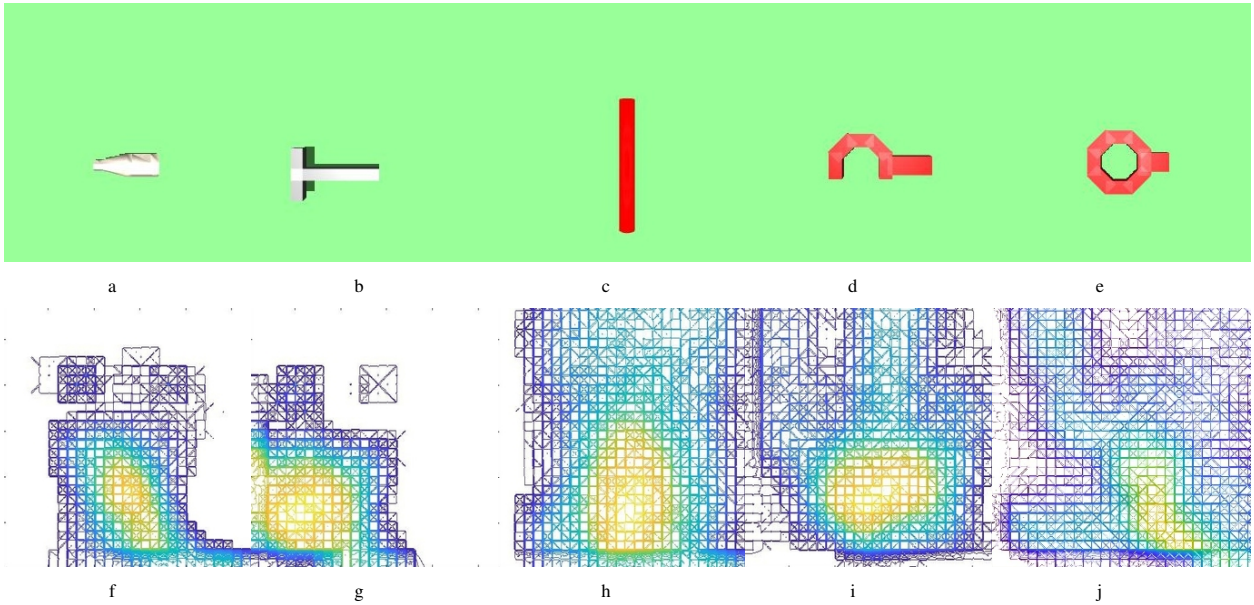

Fig. 3: Objects used in our experiments and ground truth scores of grasp robustness for grasps centred at the corresponding pixels in the image. The scores are specific to a given object position and orientation. Yellow colour indicates more robust grasps.

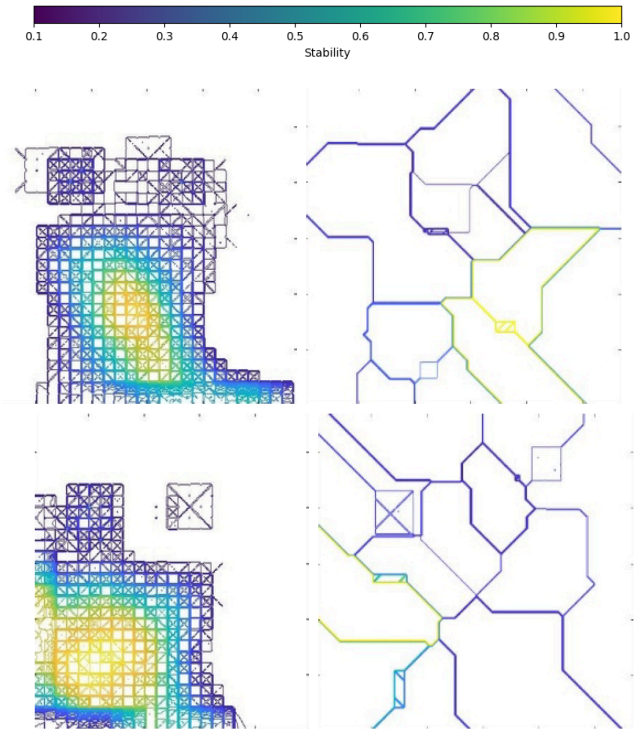

Fig. 4: Comparison of the ground truth of robustness when grasping every 10 pixels on the grid (left) and the sampled predicted reward on the same grid (right) for the bottle(top) and T-shape object (bottom).

nearby obstacles, that an adversary could use to remove the object.

This work focuses on two main factors: (1) the force magnitude, and (2) the direction of the force. Magnitude is important since, regardless of the direction, if it is strong enough to remove the object, then it should be classified as "adversarial." If it does not remove the object, then it should be classified as (ineffective) adversarial or collaborative based on its direction.

Grasping Robustness

We stated that we should classify a force as collaborative or ineffective adversarial based on the effects on the object, specifically on whether it moves the object from a "robust" grasp to a "less robust" grasp.

The first step is to define grasp robustness; it is the likelihood of withstanding external perturbation of various different magnitudes and directions. We can, therefore, map a given grasping configuration to a scalar metric of robustness by applying random forces in a range of directions and registering a scalar reward based on the effect of the forces.

To provide intuition on how grasping robustness can be computed, we performed simulations where the robot attempts repeatedly grasp centred at different points in an image of 285x285. For every 10 pixels in the image, the robot attempts 50 grasping attempts at random grasping orientations. The simulation then applies a force randomly in one of four directions (left-right, inwards-outwards). We define a robustness score as follows: (0: Failed to grip, 0.5: Succeed to grip and failed to stand against a force, 1: Succeed to grip and stand against a force). Then, we filled the corresponding position in the grid with the average rewards of 50 trials depending on the effect of the forces, and we interpolated the values filling up every pixel. Fig. 3 shows the result of the experiment for five different objects. The same objects were used in our previous work [4]. The yellow areas indicate higher scores. While there are errors from the sampling process, we see that the scores provide intuition on the more stable parts of the grasp, for instance, they are higher towards the centre of the bottle, or the left-most part of the T-shape.

This data can not only inform grasp quality but also classification of the different forces, *based on how the object moves inside the robot's gripper*; if a force moves the object towards an area of a higher score, that would likely be a collaborative force. The opposite would hold for an adversarial force. While this is implemented in simulation, in the real world such forces would be applied by people interacting with the robot.

Of course, during grasping the robot does not have access to these grids, which are specific to given object orientation. We provide these grids only for intuition. However, the robot can estimate the grasping robustness score using a partially trained convolutional neural network.

Estimating Grasping Robustness

The key insight behind this work is that we can approximate the ground-truth robustness score, described above, using a partially trained neural network. We use the same ConvNet architecture with previous work [1], modelled on AlexNet [30]. The output of the network is scaled to $(0, 1)$ using a sigmoidal response function.

We initialized the network with a pre-trained model by Pinto et al. [1]. The model was pre-trained with completely different objects and patches. We then train the model for 25 iterations, using the adversarial learning framework from our previous work [4], treating all forces as adversarial. This gives a reasonable estimate of the quality of each grasping configuration. While the framework will perform better for previously seen objects, it will iteratively update a reward map for new objects as well.

We then compare the reward *predicted* by the network with the ground-truth computed with extensive sampling. Fig. 4 shows the result of sampling 128 points in the grid and interpolating between the points for the bottle and the T-shape objects. While the reward map is simple, it has a comparable structure to the ground truth obtained by extensive sampling.

This shows that we can use the very same network to classify forces to different types, before integrating them to the learning process. In turn, this will lead to better predictions by the networks, that will iteratively make for more accurate classification. We formally specify the problem and the algorithm in Section IV.

IV. Problem Statement

We formulate the problem as a *two-player game with incomplete information* [31], played by a robot (R) and a second agent ($\mathcal{A}$). We define $s \in S$ to be the *state* of the world. A robot and the second agent are taking turns in actions. A robot action results in a stochastic transition to new state $s^+ \in S$, based on some unknown transition function $\mathcal{T} : S \times A^{\mathrm{R}} \to \Pi(S)$.

We assume that the second agent acts based on a stochastic policy, also unknown to the robot, so that $\pi^{\mathcal{A}} : (s^+, a^{\mathcal{A}})$.

The policy may either invoke an adversarial action, or a collaborative action $A^{\mathcal{A}} : A^{adv} \cup A^{col}$.

After the second agent's and the robot's actions, the robot observes the final state $s^{++} \in S$ and receives a reward signal $r : (s, a^{\mathrm{R}}, s^+, a^{\mathcal{A}}, s^{++}) \mapsto r$.

In a collaborative setting, both agents receive the same reward

$$r = R^{\mathrm{R}}(s, a^{\mathrm{R}}, s^+) \equiv R^{\mathcal{A}}(s^+, a^{col}, s^{++}) \tag{1}$$

In an adversarial setting, the robot attempts to maximize r, while the agent wishes to minimize it. 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 action taken by the adversary:

$$r = R^{\mathrm{R}}(s, a^{\mathrm{R}}, s^+) - \alpha R^{\mathcal{A}}(s^+, a^{adv}, s^{++}) \tag{2}$$

In Eq. (2), α controls the tradeoff between grasping objects from the ground and withstanding disturbances.

The robot does not observe the reward of the second agent, therefore it does not directly observe whether the second agent intended to act adversarially or collaboratively.

Our goal is to improve robustness to adversarial actions:

$$\pi^{\mathrm{R}}_* = \underset{\pi^{\mathrm{R}}}{\operatorname{argmax}}\ \mathbb{E}\left[r(s, a^{\mathrm{R}}, a^{adv}) | \pi^{\mathcal{A}}\right] \tag{3}$$

Through this maximization, the robot implicitly attempts to minimize the reward of the second agent, when the agent acts adversarially.

V. Approach

Algorithm

We parameterize the robot's policy π^{R} with a set of parameters W, represented by a convolutional neural network [4].

The robot observes state representation s and executes action a^{R} (Algorithm 1.) It then observes a new state s^+, and waits for the secondary agent to act, unless the result is terminal state, e.g., the robot failed to grasp the object. After the second agent takes an action, the robot observes the final state s^{++}. At this stage, it has to decide whether the action was adversarial or collaborative; the parameterized robot's policy is updated only when the adversarial force was applied.

If s^{++} is terminal, e.g., the object was removed from the robot's gripper, the given external force was adversarial.

If not, it compares the predicted reward that the robot would receive before (state s^+) and after (state s^{++}) the second agent acted.

In the grasping setting, we let $a^{\mathrm{R}+}_{s^+}$ be the grasping configuration, if *the robot had attempted to grasp the object in the same configuration as the one currently holding the object:* $a^{\mathrm{R}+}_{s^+} \equiv (x^+_g, y^+_g, \theta^+_g)$, where x^+_g, y^+_g and θ^+_g the grasp position and orientation at state s^+. To compute the predicted reward, the robot uses the state-action pairs $(s^+, a^{\mathrm{R}+}_{s^+})$,$(s^{++}, a^{\mathrm{R}+}_{s^+})$ and the network parameters W.

If the predicted reward in s^{++} is higher than in s^+, then the robot interprets the result as that the given action was collaborative. Formally, we define a set of collaborative forces $A^{Col} \subset A^{\mathcal{A}}$ as: $A^{Col} = \{a^{col} | a^{col} \in A^{\mathcal{A}} \wedge R_{s^{++}} - R_{s^+} \geq 0\}$ where R_s indicates predicted reward based on state s . We define the adversarial forces as $A^{Adv} = A^{\mathcal{A}} \setminus A^{Col}$.

If the given force is decided as *adversarial*, the robot computes the reward r based on Eq. (2), records the triplet

Algorithm 1 Learning in a Mixed Adversarial and Collaborative Setting

Initialize parameters W of robot's policy π_s^{R}
for batch $= 1, B$ **do**
  for episode $= 1, M$ **do**
    observe s
    sample action $a^{\mathrm{R}} \sim \pi_*^{\mathrm{R}}$
    execute action a^{R} and observe s^+
    if s^+ is not terminal **then**
      observe the state s^{++} based on $a^{\mathcal{A}}$
      if observed s^{++} is not terminal **then**
        predict reward $R_{s^{++}}$.
        predict reward R_{s^+}.
        if $R_{s^{++}} \geq R_{s^+}$ **then**
          label $a^{\mathcal{A}}$ as *"Collaborative"*
        else
          label $a^{\mathcal{A}}$ as *"Adversarial"*
      else
        label $a^{\mathcal{A}}$ as *"Adversarial"*
      if $a^{\mathcal{A}} \in A^{col}$ **then**
        continue
    observe r given by Eq. (2)
    record s, a^{R}, r
  update W based on recorded sequence
return W

(s, a^{R}, r) and updates W. A new world state is then sampled randomly.

Initialization

We initialize the parameters W by optimizing only for $R^{\mathrm{R}}(s, a^{\mathrm{R}}, s+)$, that is for the reward in the absence of the adversary. This allows the robot to choose actions that have a high probability of grasp success, which in turn enables the external force to be applied in response. We then refine the network by training it briefly with random external forces using the adversarial learning framework from previous work [4]. This provides good initial values of the predicted reward that we can use to classify forces.

Grasping Prediction

Following the previous work [1], [4], we formulate grasping prediction as a classification problem. Given a 2D input image I, taken by a camera with a top-down view, we sample N_g image patches. We then discretize the space of grasp angles to N_a different angles. We use the patches as input to a convolutional neural network, which predicts the probability of success for every grasping angle with the grasp location being the centre of the patch. The output of the ConvNet is a N_a-dimensional vector giving the likelihood of each angle. This results in a $N_g \times N_a$ grasp probability matrix. The policy then chooses the best patch and angle to execute the grasp. The robot's policy thus uses as input the image I, and as output the grasp location (x_g, y_g), which is the centre of the sampled patch, and the grasping angle θ_g: $\pi^{\mathrm{R}} : I \mapsto (x_g, y_g, \theta_g)$.

External Force

After a successful grasp, we sample uniformly the direction (Dir) and the magnitude (F) of an external force, so that: $Dir \sim DU\{1, 2, 3, 4\}$, and $F \sim U(F_{min}, F_{max})$. We set four directions for simplicity: Left, Right, Inward, and Outward. From a top-camera view, the directions can be considered as North, South, East, and West.

Categorizing Forces

After an external force is applied and the object is still held by the robot, the robot reasons over whether the force is adversarial or collaborative, by sampling an action from π^{R} at the states s^+ (before the force was applied) and s^{++} (after the force was applied). We use the network to predict the reward for each action and label the external force as collaborative if the reward has increased and adversarial otherwise. In practice, we use for robustness a *voting* scheme, where we sample a set of actions $\pi_{s^+}^{\mathrm{R}} \mapsto ((x_g^+, y_g^+) + \epsilon, \theta_g^+)$ $\epsilon \sim \mathcal{N}(0, \Sigma)$ and classify the force as collaborative or adversarial based on the majority of the votes.

We ignore collaborative samples, since otherwise the robot may learn to grasp the objects in an unstable manner, "expecting" a collaborative force. We also note that s^+ and s^{++} are observed using a top-view camera, identically to the initial state s. While lifting an object changes the scale of the object view, the CNN model is robust over the changing scale and able to recognize the features of an object similarly before and after lifting an object.

Reward System

Our reward system follows Eq. (2) and we use it as training target. We set $R^{\mathrm{R}}(s, a^{\mathrm{R}}, s^+) = 1$ if the robot succeeds to grasp the object from the ground and $R^{\mathrm{R}}(s, a^{\mathrm{R}}, s^+) = -1$ otherwise. We set $R^{\mathcal{A}}(s^+, a^{adv}, s^{++}) = 1$ if the adversary manages to snatch the object and 0 if the force fails.

Therefore, based on Eq. (2) and $\alpha = 1$, the signal received by the robot is:

$$r = \begin{cases} -1 & \text{failed grasp} \\ 1 & \text{successful grasp and the adversary fails} \\ 0 & \text{successful grasp and the adversary succeeds} \end{cases} \tag{4}$$

VI. EXPERIMENTS

Our experiments aim to answer the following question: in a mixed adversarial/collaborative setting, can we improve the performance of the grasping policy by identifying and ignoring collaborative samples? We trained and tested models on simulated Mujoco environment having trained two types of models: without force classification as in the previous work [4] and with force classification.

Definitions

We name "pre-success" a case where the robot successfully grasps an object and lifts it. "Post-success" is when the robot successfully lifts the object and withstands external forces. Finally, 'robust rate' refers to the number of post-successes

Type	Bottle	T shape	Bar	Half Nut	Round Nut
Baseline framework [4]	69.59	32.16	28.2	83.27	84.96
Finetuned Baseline framework	**80.36**	**82.18**	**83.38**	**86.24**	**91.86**

TABLE I: Pre-success rate comparison between the framework from previous work and the finetuned baseline framework.

Object Type	Bottle		T Shape		Bar		Half Nut		Round Nut	
Force Type	Adv	Col	Adv	Col	Adv	Col	Adv	Col	Adv	Col
Detected / Total (Fraction)	10/14	19/25	20/23	24/33	31/38	24/29	21/24	19/27	15/23	39/44
Detected / Total (Decimal)	0.714	0.760	0.870	0.727	0.861	0.828	0.876	0.704	0.652	0.886

TABLE II: Force classification performance.

Test #	Bottle	T shape	Bar	Half Nut	Round Nut
1	57 57	75 60	79 79	81 76	78 20
2	52 52	71 66	82 81	85 81	68 19
3	48 52	71 66	79 79	90 75	70 30
4	40 53	68 63	74 79	93 78	77 20
5	50 48	73 68	86 76	88 85	72 30
Average	49.4 **52.4**	**71.6** 64.6	**80.0** 78.8	**87.4** 79	**73.0** 23.8

TABLE III: Number of post-successes for 98 test iterations with (left) and without (right) force classification.

Test #	Bottle	T shape	Bar	Half Nut	Round Nut
1	0.904 0.740	0.815 0.789	0.929 0.929	0.952 0.904	0.896 0.215
2	0.800 0.722	0.816 0.804	0.891 0.952	0.965 0.920	0.829 0.220
3	0.827 0.590	0.865 0.767	0.929 0.963	0.989 0.872	0.886 0.315
4	0.727 0.670	0.809 0.759	0.913 0.950	0.989 0.906	0.846 0.217
5	0.781 0.539	0.829 0.809	0.966 0.930	0.942 0.977	0.888 0.326
Average	**0.808** 0.652	**0.827** 0.786	0.926 **0.945**	**0.967** 0.916	**0.869** 0.259

TABLE IV: Robust rates for each object among 98 test iterations with (left) and without(right) force classification.

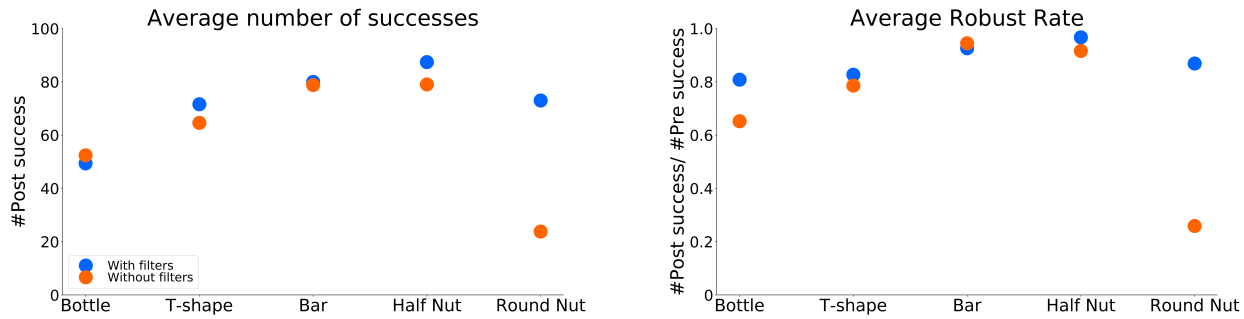

Fig. 5: Average success rates (left) and robust rates (right) without force classification (red) and with force classification (blue).

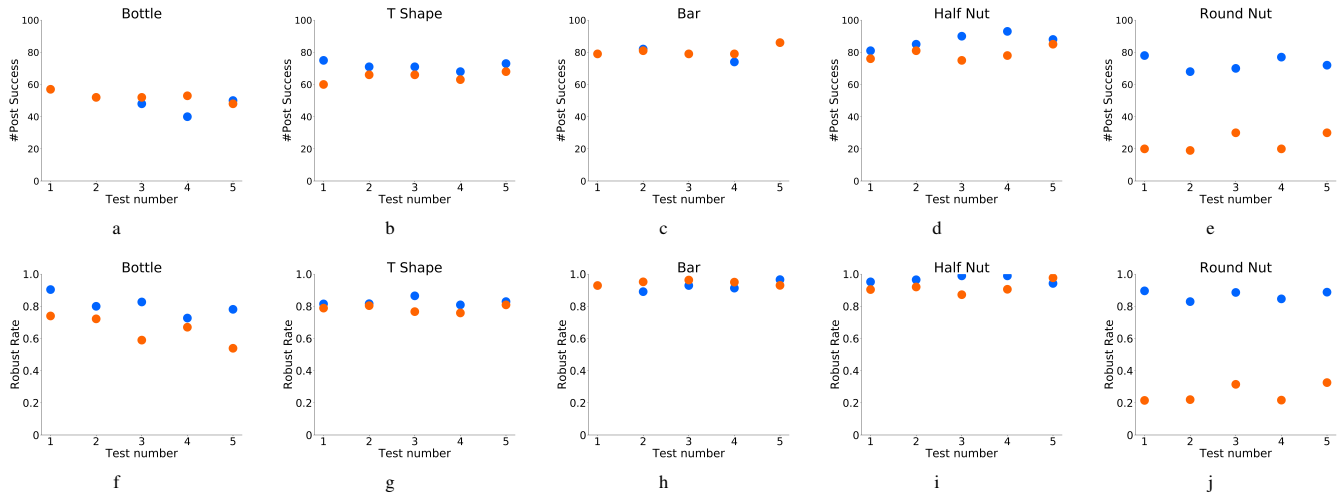

Fig. 6: Post-success scores (top) and robust rates (bottom) with force classification (blue) and without force classification (red).

divided by the number of pre-successes, which indicates how good the robot is in withstanding disturbances after it has grasped the object.

Baseline Model

We extensively finetuned the baseline model to achieve the best possible performance for a fair comparison. We first changed the reward range from $r \in [0, 1]$ in previous work [4] to $[-1, 1]$ in Eq. , since we observed that penalizing failed grasping attempts resulted in higher pre-success rate accuracy. Additionally, we introduced an exploration rule, where if the robot failed to grasp the object from the ground, the robot would randomly sample a grasping position and orientation at the next time step, rather than picking the best one. Table I shows the improvement in the baseline model after fine-tuning, compared to the adversarial learning model of the previous work [4] after 100 iterations in the 5 different objects shown on Fig. 3.

Force Classification Performance

Before testing our framework, we want to assess whether the system can accurately classify external forces. We focus on ineffective adversarial forces, that is on adversarial forces that fail to remove the object from the robot's gripper. An experimenter first applied an external force through a user interface in Mujoco [4] and labelled the force as collaborative or adversarial. We then compared the human labels with the output of the classification system. Table II shows the result in 100 iterations. We see that the system achieved high performance in most of the objects.

Manipulated Variables

We manipulated (1) the robot's learning framework and (2) the objects that the robot interacted with. We had two conditions for the first independent variable: (a) the robot executing Algorithm 1, classifying forces as adversarial or collaborative, and (b) the robot executing the baseline adversarial framework from previous work [4], where it treats all forces as adversarial.

We had five different objects (Fig. 3), identical to those of previous work [4]. The objects are of varying grasping difficulty and geometry. We used 100 iterations of training for each object. We did not classify forces in the first 25 iterations, in order to get a good prior to the predicted rewards.

Dependent Measures

We run 5 tests per object and learning algorithm, each having 98 episodes, after the end of the training. During testing, we did not perform any parameter updates and we picked the best actions based on the learned policies. We computed the number of post-successes and robust rates.

Hypothesis

We hypothesize that the robot trained with force classification will perform better than the robot trained without it. We expect that treating collaborative forces as adversarial will reinforce grasps that are not robust. Based on the force classification performance results in Fig. II, we expect the proposed framework to distinguish and disregard collaborative forces.

Analysis

Table III and IV show the post-success numbers and robust rates of the two algorithms in each of the five objects. A two way multivariate ANOVA with object and learning algorithm as independent variables showed a significant interaction effect between objects and learning algorithm on both dependent measures ($p < 0.001$). In line with our hypothesis, Post-hoc Tukey tests showed that post-success attempts and robust rates were significantly higher for the proposed framework, compared to the baseline framework ($p < 0.001$).

We note that the post-hoc analysis should be viewed with caution, because of the significant interaction effect. To interpret these results, we plot the post-success numbers and robust rates for each object in Fig. 5, as well as for each test trial in Fig. 6.

We observe that the average post-success score is higher in four out of the five objects. In the object 'Bottle', we observe a seemingly contradictory result; the proposed framework has a lower post-success score, but a higher robust rate. The reason is that the bottle is a hard object to grasp in the first place; ignoring the collaborative forces leads to a smaller number of training samples for the learning algorithm, therefore a lower probability of picking up the object successfully. On the other hand, when the robot of the proposed framework does grasp the object, it uses more robust configurations, which explains the difference in the robust rate.

On the other hand, at the round-nut, T-shape and Half nut objects the robot can succeed in grasping them from a range of different configurations, while there are significant differences in the how robust these configurations are. For instance, grasping the round-part from the centre of the circle is significantly more robust to disturbances than grasping it from the rightmost part. Therefore, for these objects, the framework achieved better performance.

VII. Conclusion

Limitations

Our work is limited in many ways. To estimate the effect of a force we assume a top-down camera view of the object; in practice, occlusions may affect the image input to the system, and it would be interesting to combine our approach with active exploration techniques [32]–[34]. In addition, the applied forces are sampled from a predefined set of directions and magnitude. In real-world settings, we expect people to interact with robots in richer, more nuanced ways [35], and identifying adversarial actions will require higher-level reasoning over their beliefs, desires and intentions [36].

Implications

Even in adversarial settings, people do not consistently act as antagonists; sometimes they challenge each other, other times they will give a helping hand. The same holds for human-robot interactions, and being able to distinguish

between collaborative and adversarial actions is essential for effective robot learning. We are excited to explore further how the robot can learn robust behaviours in mixed cooperative / adversarial settings in manipulation, navigation and social interactions.

REFERENCES

[1] L. Pinto and A. Gupta, "Supersizing self-supervision: Learning to grasp from 50k tries and 700 robot hours," in *Robotics and Automation (ICRA), 2016 IEEE International Conference on*. IEEE, 2016, pp. 3406–3413.

[2] S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz, and D. Quillen, "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection," *The International Journal of Robotics Research*, vol. 37, no. 4-5, pp. 421–436, 2018.

[3] L. Pinto and A. Gupta, "Learning to push by grasping: Using multiple tasks for effective learning," in *Robotics and Automation (ICRA), 2017 IEEE International Conference on*. IEEE, 2017, pp. 2161–2168.

[4] J. Duan, Q. Wang, L. Pinto, C.-C. J. Kuo, S. Nikolaidis *et al.*, "Robot learning via human adversarial games," *arXiv preprint arXiv:1903.00636*, 2019.

[5] J. Bohg, A. Morales, T. Asfour, and D. Kragic, "Data-driven grasp synthesis—a survey," *IEEE Transactions on Robotics*, vol. 30, no. 2, pp. 289–309, 2014.

[6] A. Bicchi and V. Kumar, "Robotic grasping and contact: A review," in *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065)*, vol. 1. IEEE, 2000, pp. 348–353.

[7] J. Mahler, F. T. Pokorny, B. Hou, M. Roderick, M. Laskey, M. Aubry, K. Kohlhoff, T. Kröger, J. Kuffner, and K. Goldberg, "Dex-net 1.0: A cloud-based network of 3d objects for robust grasp planning using a multi-armed bandit model with correlated rewards," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 1957–1964.

[8] D. Berenson, R. Diankov, K. Nishiwaki, S. Kagami, and J. Kuffner, "Grasp planning in complex scenes," in *2007 7th IEEE-RAS International Conference on Humanoid Robots*. IEEE, 2007, pp. 42–48.

[9] D. Berenson and S. S. Srinivasa, "Grasp synthesis in cluttered environments for dexterous hands," in *Humanoids 2008-8th IEEE-RAS International Conference on Humanoid Robots*. IEEE, 2008, pp. 189–196.

[10] E. Rimon and J. Burdick, *The Mechanics of Robot Grasping*. Cambridge University Press, 2019.

[11] I. Lenz, H. Lee, and A. Saxena, "Deep learning for detecting robotic grasps," *The International Journal of Robotics Research*, vol. 34, no. 4-5, pp. 705–724, 2015.

[12] S. Levine, C. Finn, T. Darrell, and P. Abbeel, "End-to-end training of deep visuomotor policies," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 1334–1373, 2016.

[13] P. Agrawal, A. V. Nair, P. Abbeel, J. Malik, and S. Levine, "Learning to poke by poking: Experiential learning of intuitive physics," in *Advances in Neural Information Processing Systems*, 2016, pp. 5074–5082.

[14] J. Mahler, J. Liang, S. Niyaz, M. Laskey, R. Doan, X. Liu, J. A. Ojea, and K. Goldberg, "Dex-net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics," *arXiv preprint arXiv:1703.09312*, 2017.

[15] J. Mahler, M. Matl, V. Satish, M. Danielczuk, B. DeRose, S. McKinley, and K. Goldberg, "Learning ambidextrous robot grasping policies," *Science Robotics*, vol. 4, no. 26, p. eaau4984, 2019.

[16] V. Satish, J. Mahler, and K. Goldberg, "On-policy dataset synthesis for learning robot grasping policies using fully convolutional deep networks," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1357–1364, 2019.

[17] K. Bousmalis, A. Irpan, P. Wohlhart, Y. Bai, M. Kelcey, M. Kalakrishnan, L. Downs, J. Ibarz, P. Pastor, K. Konolige *et al.*, "Using simulation and domain adaptation to improve efficiency of deep robotic grasping," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 4243–4250.

[18] K. Fang, Y. Bai, S. Hinterstoisser, S. Savarese, and M. Kalakrishnan, "Multi-task domain adaptation for deep learning of instance grasping from simulation," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 3516–3523.

[19] 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.

[20] W. B. Knox and P. Stone, "Reinforcement learning from simultaneous human and mdp reward," in *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*. International Foundation for Autonomous Agents and Multiagent Systems, 2012, pp. 475–482.

[21] ——, "Combining manual feedback with subsequent mdp reward signals for reinforcement learning," in *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*. International Foundation for Autonomous Agents and Multiagent Systems, 2010, pp. 5–12.

[22] R. Loftin, B. Peng, J. MacGlashan, M. L. Littman, M. E. Taylor, J. Huang, and D. L. Roberts, "Learning behaviors via human-delivered discrete feedback: modeling implicit feedback strategies to speed up learning," *Autonomous agents and multi-agent systems*, vol. 30, no. 1, pp. 30–59, 2016.

[23] S. Griffith, K. Subramanian, J. Scholz, C. L. Isbell, and A. L. Thomaz, "Policy shaping: Integrating human feedback with reinforcement learning," in *Advances in neural information processing systems*, 2013, pp. 2625–2633.

[24] W. B. Knox and P. Stone, "Tamer: Training an agent manually via evaluative reinforcement," in *2008 7th IEEE International Conference on Development and Learning*. IEEE, 2008, pp. 292–297.

[25] D. Arumugam, J. K. Lee, S. Saskin, and M. L. Littman, "Deep reinforcement learning from policy-dependent human feedback," *arXiv preprint arXiv:1902.04257*, 2019.

[26] J. MacGlashan, M. K. Ho, R. Loftin, B. Peng, G. Wang, D. L. Roberts, M. E. Taylor, and M. L. Littman, "Interactive learning from policy-dependent human feedback," in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 2017, pp. 2285–2294.

[27] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[28] V. Dumoulin, I. Belghazi, B. Poole, O. Mastropietro, A. Lamb, M. Arjovsky, and A. Courville, "Adversarially learned inference," *arXiv preprint arXiv:1606.00704*, 2016.

[29] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, 2014.

[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[31] R. Lavi, "Algorithmic game theory," *Computationally-efficient approximate mechanisms*, pp. 301–330, 2007.

[32] H. Liu, Y. Yuan, Y. Deng, X. Guo, Y. Wei, K. Lu, B. Fang, D. Guo, and F. Sun, "Active affordance exploration for robot grasping," in *International Conference on Intelligent Robotics and Applications*. Springer, 2019, pp. 426–438.

[33] G. Kahn, P. Sujan, S. Patil, S. Bopardikar, J. Ryde, K. Goldberg, and P. Abbeel, "Active exploration using trajectory optimization for robotic grasping in the presence of occlusions," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 4783–4790.

[34] J. Aleotti, D. L. Rizzini, and S. Caselli, "Perception and grasping of object parts from active robot exploration," *Journal of Intelligent & Robotic Systems*, vol. 76, no. 3-4, pp. 401–425, 2014.

[35] D. Brščić, H. Kidokoro, Y. Suehiro, and T. Kanda, "Escaping from children's abuse of social robots," in *Proceedings of the tenth annual acm/ieee international conference on human-robot interaction*, 2015, pp. 59–66.

[36] A. S. Rao, M. P. Georgeff *et al.*, "Bdi agents: from theory to practice." in *ICMAS*, vol. 95, 1995, pp. 312–319.