

Bi-Touch: Bimanual Tactile Manipulation with Sim-to-Real Deep Reinforcement Learning

Yijiong Lin, Alex Church, Max Yang, Haoran Li, John Lloyd, Dandan Zhang, Nathan F. Lepora

Abstract—Bimanual manipulation with tactile feedback will be key to human-level robot dexterity. However, this topic is less explored than single-arm settings, partly due to the availability of suitable hardware along with the complexity of designing effective controllers for tasks with relatively large state-action spaces. Here we introduce a dual-arm tactile robotic system (Bi-Touch) based on the Tactile Gym 2.0 setup that integrates two affordable industrial-level robot arms with low-cost high-resolution tactile sensors (TacTips). We present a suite of bimanual manipulation tasks tailored towards tactile feedback: bi-pushing, bi-reorienting and bi-gathering. To learn effective policies, we introduce appropriate reward functions for these tasks and propose a novel goal-update mechanism with deep reinforcement learning. We also apply these policies to real-world settings with a tactile sim-to-real approach. Our analysis highlights and addresses some challenges met during the sim-to-real application, e.g. the learned policy tended to squeeze an object in the bi-reorienting task due to the sim-to-real gap. Finally, we demonstrate the generalizability and robustness of this system by experimenting with different unseen objects with applied perturbations in the real world. Code and videos are available at <https://sites.google.com/view/bi-touch/>.

Index Terms—Force and Tactile Sensing; Reinforcement Learning

I. INTRODUCTION

BIMANUAL robotic manipulation is a useful and natural way of manipulating large, unwieldy or coupled objects due to the better manoeuvrability, flexibility and a larger workspace compared to single-arm settings [1]. Furthermore, the higher dimensionality of the dual-arm state-action space can enable more realistic tasks for real-world application, particularly when tactile sensing is leveraged to complement vision [2]. However, there are challenges in applying bimanual touch: (1) dual-arm systems often introduce more complexity in terms of system integration and controller design [3]; and (2) the high cost of existing dual-arm systems makes them less accessible to the research community. Moreover, while vision is commonly used as the primary sensing modality for bimanual manipulation, tactile sensing complements those aspects where vision is limited, such as enabling physically-interactive control with soft contacts and ensuring robustness in scenarios where visual occlusion may occur.

Manuscript received March 23, 2023; Revised June 22, 2023; Accepted July 6, 2023. This paper was recommended for publication by Editor Ashis Banerjee upon evaluation of the Associate Editor and Reviewers' comments. YL was supported by a UoB-CSC-joint scholarship. NL were supported by a Leadership Award from the Leverhulme Trust on 'A biomimetic forebrain for robot touch' (RL-2016-39). (Corresponding author: Yijiong Lin)

All authors are with the Department of Engineering Mathematics and Bristol Robotics Laboratory, University of Bristol, Bristol BS8 1UB, U.K. (email: {yijiong.lin, n.lepora}@bristol.ac.uk)

Digital Object Identifier (DOI): see top of this page.

Thus, tactile feedback is needed for precise, safe and reliable dual-arm manipulation, since careful regulation of the applied force on the local contact during robot-object interaction is needed for stable control while avoiding damage [3]. However, it remains an open challenge to design effective controllers for dual-arm robots using high-resolution tactile sensing to manipulate unknown objects in uncertain environments [4].

Recent advances in deep reinforcement learning (RL) for robotics indicate a plausible route to acquire sophisticated control policies for manipulation tasks with high-dimensional state-action space [5]–[7]. Nevertheless, progress in data-driven methods for bimanual robotics is challenging because of the sim-to-real gap, research reproducibility and inaccessible robotic hardware. Also, learning tactile dual-arm manipulation in the real world is difficult, such as preventing the arms from colliding, which is especially difficult at the start of training. Moreover, RL requires frequent manual resetting of tasks which is too laborious in the real world. Lastly, large-scale training in the real world risks damage to sensors. Therefore, the present study aims to advance recent progress with sim-to-real tactile deep RL methods applied to low-cost high-resolution tactile sensors with an affordable dual-arm manipulation system.

The main contributions of this work are as follows:

- 1) We adapt and extend the Tactile Gym 2.0 [8] to a low-cost dual-arm tactile robot setting with three new contact-rich bimanual tasks: bi-pushing, bi-reorienting, and bi-gathering.
- 2) We introduce appropriate reward functions for these tasks and show that deep RL reaches satisfactory performance using only proprioceptive and tactile feedback. To improve the robustness of the policies for real-world applications, we improve the sim-to-real transfer for bi-reorienting and propose a novel goal-update mechanism (GUM) for bi-gathering.
- 3) We demonstrate that the bimanual policies learned in simulation can be transferred well to the physical dual-arm robot. We further demonstrate the generalizability and robustness of the learned policies by testing the system on unseen objects.

To the best of our knowledge, this is the first framework for sim-to-real deep RL tailored to bimanual tactile manipulation.

II. RELATED WORK

A. Bimanual Robot System with Deep RL

Designing an effective controller for dual-arm manipulation is a long-standing challenge [9], [10], and the recent success of deep RL incentivizes robotics researchers to apply it to this problem [11], [12]. Previous work [13] successfully applied an RL policy learned in simulation to a real-world

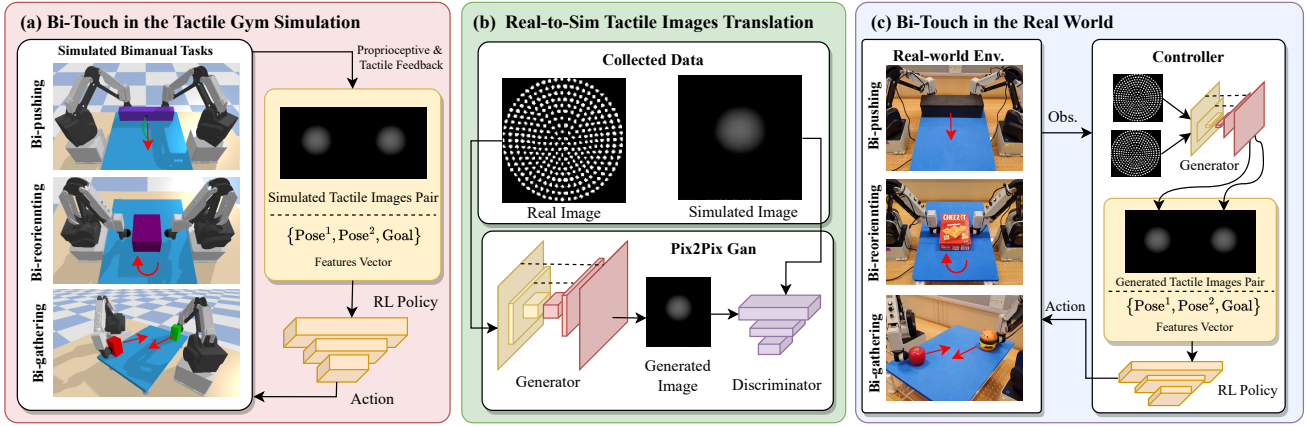


Figure 1. Overview of the proposed dual-arm tactile robotic system (Bi-Touch) with sim-to-real deep RL. a) Deep RL is applied to learn policies for three simulated bimanual tactile manipulation tasks (red arrows show desired displacements) using Tactile Gym. b) Real-to-sim tactile image generator learnt for the surface feature. c) The real-world evaluation feeds real tactile images through the generator into the RL policy concatenated with proprioceptive information.

task (connecting magnets) with a dual-arm platform, but it relied on a marker-based visual tracking system and did not show any generalization ability to unseen objects. Grannen et al. [14] proposed an RL framework to learn bimanual scooping policies for food acquisition and demonstrated its generalizability on various unseen food.

Since deep RL relies on large amounts of training, in recent years several bimanual manipulation simulators have been developed. Fan et al. developed a distributed RL framework called SURREAL for a set of dual-arm manipulation tasks [15]. Similarly, Zhu et al. [16] designed a suite of single-arm and dual-arm manipulation environments with Mujoco for algorithm development and evaluation. Chen et al. [17] developed a simulation benchmark for bimanual dexterous hands with a suite of bimanual manipulation tasks and tried solving these tasks with different RL methods. However, these benchmarks have not considered tactile sensing, which hinders their application to tasks that require direct, detailed information about the local contact for fine-grained manipulation.

B. Bimanual Robot System with Tactile Sensing

Tactile sensing enables dual-arm robots to perceive the local contact features with a granularity that cannot be achieved exclusively with vision. Nevertheless, this topic is not widely studied. Sommer et al. [18] leveraged a dual-arm tactile robot for object exploration with a Gaussian-Process-based filter and grasp-pose selection using a Gaussian mixture model; the method required human demonstrations for particular objects and only one arm had tactile sensing. Hogan et al. [19] developed dual-arm controllers with tactile palms for pusher-slider manipulation that can explicitly control the object's trajectory. A drawback of this approach was that it relied on a set of pre-designed motion skills and assumed full knowledge of the environment, hindering its ability to generalize to uncertain environments where unseen objects and unpredictable perturbations may be encountered. Here we aim to facilitate research on bimanual tactile robotics by developing a dual-arm tactile robot platform as a benchmark to design sophisticated controllers for complex bimanual tasks.

III. METHODS

A. Accessible Dual-arm Tactile Robotic System

1) *Desktop Dual-arm Platform*: To facilitate affordable automation and lower the entry barrier, we develop a low-cost dual-arm tactile robotic system while keeping high accuracy, which is comprised of two industry-capable desktop robotic arms (Dobot MG400) with vision-based tactile sensors mounted at the wrists as end-effectors (Fig. 1c). The proposed platform is developed with the Tactile Gym 2.0 [8] simulation (Fig. 1a) for deep RL-based policy training (see Sec. III-C).

Although the Dobot MG400 has a 4-DoF workspace with only Cartesian positions and rotations around the z -axis of the end-effector actuated, its accuracy is the same as larger industrial robot arms such as the UR5 at a considerably improved cost and convenience [8]. To maximally leverage the workspace of two robots while having one fixed configuration suitable for all the bimanual tactile robotic tasks considered here, we introduce a table (120 mm height) into the physical dual-arm platform to support objects, placed between the two arms (shown in blue in Fig. 1a,c). The two arm bases are set centrally below the board separated by a distance of 700 mm.

2) *High-resolution optical tactile sensing*: To endow the dual-arm robot with tactile sensing, we equip it with two low-cost, high-resolution biomimetic optical tactile sensors used in previous tactile robot research with the Dobot MG400 (the TacTip) [20] as the end-effectors. The sensor features an internal array of biomimetic markers on protruding pins inside the soft tactile skin, and the sensing principle of this tactile sensor is to capture marker-based movement that amplifies the skin deformation induced by physical contact against external stimuli. We refer to references [21], [22] for more details.

B. Sim-to-Real Deep RL Framework for Bimanual Tactile Robotic Manipulation

To apply the deep RL policies learned in simulation to the physical dual-arm tactile robotic system, we take a sim-to-real approach [7] consisting of three parts (shown in Fig. 1): 1) An online agent training in simulation (Fig. 1a), where

deep RL policies are learned in the Tactile Gym for three bimanual tactile robotic tasks (bi-pushing, bi-reorienting and bi-gathering) with observations comprising simulated tactile images and proprioceptive feedback. **2)** A real-to-sim domain adaption process where a translation model is learned to transfer real to simulated tactile images. **3)** A sim-to-real application with networks trained in the previous two parts, for transferring deep RL policies to the physical system.

To apply this approach to the Bi-Touch, we make several changes. First, we develop a simulated dual-arm tactile robot in the configuration described above (Sec. III-A) in Tactile Gym that is suited to the three bimanual robotic tasks. In the simulation learning phase, we concatenate two simulated tactile images (depicted in Fig. 1a) to be used as part of the observation of an RL agent. The simulated tactile images are captured by synthetic cameras embedded within simulated tactile sensors built from CAD models of the real sensors.

In the real-to-sim domain adaption phase, a real tactile image dataset paired with a simulated tactile image dataset is required for training. All three bimanual tasks considered here need to model a specific contact feature: the shape of a flat (or relatively flat) contact surface. We adopt the procedure proposed in [7] for dataset collection. However, a larger sensing space is necessary, as the bimanual tasks require the robot to learn more difficult control than for the previous single-arm tasks [7], [8]. Thus we collect surface-feature data with depth range [0.5, 8] mm and rotation range $[-30^\circ, 30^\circ]$, broader than in [7], [8]. The training dataset comprises 5000 tactile images and the validation dataset has 2000 tactile images collected both in simulation and with a desktop tactile robot on paired random contacts. Samples are labelled with the relative poses between the sensor and a known flat surface. Finally, we use an image-to-image translation Generative Adversarial Network (GAN) [23] to learn the real-to-sim tactile image translation, with hyperparameters from [7].

Since the RL policies take concatenated tactile images as part of the observation, in the sim-to-real application phase (Fig. 1c), we first transfer the real tactile images from both TacTips separately into simulated ones. These tactile images were then concatenated as observation input for the RL agent.

C. Bimanual Tactile Manipulation Tasks

In this study, we propose three bimanual tactile control tasks to benchmark the aforementioned dual-arm tactile system: bi-pushing, bi-gathering and bi-reorienting. The action space of the dual-arm robot in the bi-reorienting task is 6-dimensional comprising the x -, y -position and Rz -rotation angle of each robot's tool-centre-point (TCP), while the ones in the bi-pushing task (x and Rz) and the bi-gathering task (y and Rz) are 4-dimensional. Note that each end-effector (TacTip) is controlled and moves in its own TCP frame.

1) Bi-Pushing: An advantage of dual-arm robots over single-arm robots is that they can move relatively large and unwieldy objects. The goal of this bi-pushing task is to move a large-size object on a planar surface collaboratively with two robot arms with end-effectors (TacTips) to achieve a sequence of goals along a given trajectory. Each goal comprises the

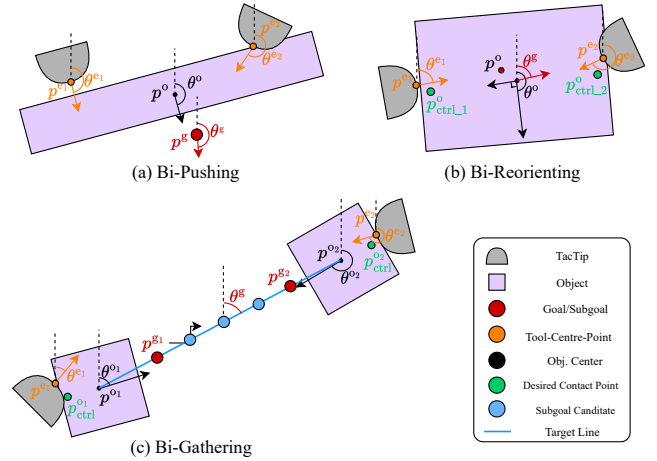


Figure 2. Illustration of the reward functions for all three proposed bimanual tasks: (a) bi-pushing, (b) bi-reorienting, and (c) bi-gathering. The notations in red are the goals (or subgoals) and those in orange are the proprioceptive information that is part of the observation.

target object's position $p^g(x, y)$ and orientation θ^g . We express the reward R_t^{BP} at time step t for the bi-pushing task as:

$$R_t^{\text{BP}} = -w_1 \|p_t^g - p_t^o\|_2 - w_2 S(\theta_t^g, \theta_t^o) - w_3 \sum_{i=1}^2 S(\theta_t^{e_i}, \theta_t^o), \quad (1)$$

where $w_j > 0$ ($j \in \{1, 2, 3\}$) are reward weights; p_t^g and θ_t^g are the position and orientation of the current goal respectively; p_t^o and θ_t^o are the current position and orientation of the object; $\theta_t^{e_i}$ is the current orientation of the TCP e_i for robot arm $i \in \{1, 2\}$; $S(\phi, \psi) = 1 - \cos(\phi - \psi)$ is the cosine distance between the angles ϕ and ψ . All notation is illustrated in Fig. 2a. We interpret the first and second terms of Eq. (1) as encouraging the robot to push the object towards the goal, while the third term is to encourage the robot to maintain the TacTips normal to the contact surface for stable pushing.

2) Bi-Reorienting: Reorienting an object with two arms is necessary when the object size exceeds the limit of what can be held by a gripper or a robot hand. The goal of this bi-reorienting task is for two robotic arms to reorient an object located at the workspace centre to a given target angle θ^g while keeping the object centre fixed in place. The dual-arm robot should reorient the object with gentle contact while keeping the end-effectors (TacTips) normal to the contact surface. We express the reward R_t^{BR} at step t for the bi-reorienting task as

$$R_t^{\text{BR}} = -w_1 \|p_0^o - p_t^o\|_2 - w_2 S(\theta_t^g, \theta_t^o) - w_3 \sum_{i=1}^2 S(\theta_t^{e_i}, (-1)^i(\pi/2 + \theta_t^o)) - w_4 \sum_{i=1}^2 \|p_{\text{ctrl}_i}^o - p_t^{e_i}\|_2 \quad (2)$$

where $w_j > 0$ ($j \in \{1, 2, 3, 4\}$) are reward weights; p_0^o , p_t^o and θ_t^o are the initial/current positions and orientation of the object respectively; $p_t^{e_i}$ and $\theta_t^{e_i}$ are the current position and orientation of each TCP e_i ($i \in \{1, 2\}$) respectively; $p_{\text{ctrl}_i}^o$ ($i \in \{1, 2\}$) are the desired contact points to either side of an object that are used to specify the desired positions of the TCPs when in contact with the object. All notation is illustrated in

Fig. 2b. The second term in Eq. (2) encourages the robot to reorient the object to the target angle θ^g while maintaining the object at its original position with the first term. The third term encourages the robot to maintain the TCP normal to the contact surface for stable reorienting. The final term encourages the dual-arm robot to keep both TCPs close to the desired contact points, which helps maintain the contact, especially the contact depth, of the TCP to avoid losing contact with the object.

However, a challenge we met during the sim-to-real application for this task is that a policy trained using Eq. (2) with the original simulated TacTip dynamics (stiffness, damping, etc) tended to squeeze the object when the goal angle is large in the real world, and this phenomenon became worse as the object length increased (see accompanying video). We attribute this phenomenon to a sim-to-real gap where the dual-arm robot tried to finish the task quickly for maximum rewards, but then learned to rotate the object by squeezing it in simulation to achieve this. However, in the real world, the TacTips may break if over-deformed and the objects are not always of high stiffness, so this learned policy is not suitable.

To solve this problem, we made several changes for policy learning in the simulation: 1) we tuned the simulated TacTip skin stiffness and damping coefficients to make it more elastic; 2) the dual-arm robot needed to maintain the object in the same pose for 10 time steps to achieve the goal orientation; 3) we increased the penalty coefficient if the dual-arm robot over-squeezes the object (large contact depth); 4) we trained the policy with a higher probability of larger goal angles. After conducting an ablation study for the above changes, we found that change 1) contributes most to the successful sim-to-real application, as the higher stiffness and lower damping make the TacTip skin more elastic against external contact, which enables the policy to be more sensitive to the contact depth as the tactile image are more responsive to depth in simulation.

3) *Bi-Gathering*: Gathering objects together is a common behaviour in our daily life, from tidying our desks to moving and sorting packages in warehouses. The goal of this bi-gathering task is for the dual-arm robot to gather two objects together by pushing them towards each other on a planar surface. Thus, each end-effector of the dual-arm robot has to push an object towards a dynamically changing goal (as the goal position of each object is the other object's current position), which makes it a harder exploration problem for RL as compared to single-arm pushing [8] or bi-pushing (above) that have static goals. The task is considered completed when the distance d between the two objects becomes closer $d < \epsilon$ than a distance threshold ϵ (related to the size of the objects). We express the reward R_t^{BG} at t of the bi-gathering task as

$$R_t^{\text{BG}} = -w_1 \|p_t^{\text{o}_1} - p_t^{\text{o}_2}\|_2 - w_2 \sum_{i=1}^2 S(\theta_t^{\text{o}_i}, \theta_t^{\text{o}_i}) - w_3 \sum_{i=1}^2 \|p_{\text{ctrl}}^{\text{o}_i} - p_t^{\text{o}_i}\|_2, \quad (3)$$

where $w_j > 0$ ($j = \{1, 2, 3\}$) are reward weights; $p_t^{\text{o}_i}$ and $\theta_t^{\text{o}_i}$ are the current position and orientation of the object o_i ; $p_t^{\text{o}_i}$ and $\theta_t^{\text{o}_i}$ are the current position and orientation of the TCP e_i ; $p_{\text{ctrl}}^{\text{o}_i}$ specifies the desired contact point position on

the object o_i for controlling the contact depth of the TCP e_i ($i \in \{1, 2\}$). All notation is illustrated in Fig. 2c. The first term of Eq. (3) pushes two objects closer while the second term tries to maintain the TCP normal to the contact surface of the object for stable pushing. The third term encourages the robot to keep the TCPs close to the contact points as much as possible to avoid losing contact.

To further explore the limit of the dual-arm tactile robot, we also introduce random perturbations to the objects during the gathering. Specifically, a random force is applied to an object's centre of mass at a random time step when training.

Although the most obvious formulation would be to place the goal of an object at another object, we found that this did not work well for a moving sparse goal both without and with perturbations, resulting in suboptimal performance as shown in Fig. 3c (green and purple plots). Thus, we propose a novel goal-update mechanism (GUM) that can generate subgoals to help the robot learn the task with auxiliary reward signals. Specifically, we generate n static points at equal interval distances along a target line as subgoals for the two objects (as shown by the blue line in Fig. 2c). We have experimented it with $n = [5, 10, 20]$ and found $n = 10$ performed slightly better than the others. The target line is updated every h time steps, and the subgoal location for each robot arm is set to the generated point nearest to the object that is closer to it. We express the reward $R_t^{\text{BG-GUM}}$ with GUM as

$$R_t^{\text{BG-GUM}} = R_t^{\text{BG}} - w_4 \sum_{i=1}^N \|p_t^{\text{g}_i} - p_t^{\text{o}_i}\|_2 - w_5 \sum_{i=1}^N S(\theta_t^{\text{o}_i}, (-1)^i \theta_t^c), \quad (4)$$

where $w_j > 0$ ($j = \{4, 5\}$) are reward weights; $p_t^{\text{g}_i}$ and θ_t^c represent the position of subgoal g_i for object o_i at time step t respectively. All notation is illustrated in Fig. 2c. The second term of Eq. (4) guides objects towards the nearest subgoals on the target line, providing denser auxiliary rewards for R_t^{BG} . The final term encourages object to be pushed along the target line direction. The hyperparameter h controls the target line update rate, which should update with appropriate frequency to provide sufficient guidance for the robot to move objects in the desired direction, but also avoid updating too often, which can cause the robot to fail to reach the subgoals due to insufficient time steps. We have experimented it with $h = [25, 50, 75, 100]$ and found $h = 75$ performed the best.

For the sim-to-real application in this task, since the object pose is unknown in real-world experiments, we instead select the current tool-centre-points (TCPs) for constructing the target line. However, during training in simulation, we found that training from scratch with a line between TCP positions failed to learn. This is because the initial random policy cannot maintain contact with objects and accordingly, this line cannot provide useful information for learning. To circumvent this, we devised a 2-step curriculum in the simulation where we first use objects' centres to construct the target line to train a policy from scratch, and then switch to TCPs for extra training so that the policy can be attuned to the real-world setting. We

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

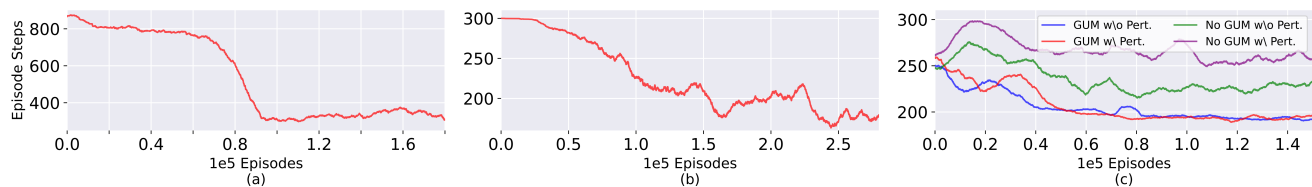


Figure 3. Task completion times during training for learned policies averaged over 10 trials (a) bi-pushing, (b) bi-reorienting, and (c) bi-gathering tasks. Note that less task completion time means better performance. Our proposed goal-update mechanism (GUM) in the bi-gathering task resulted in improved performance: the blue and red plots (with GUM) show reduced numbers of time steps required to achieve the task in one episode, under both perturbed and unperturbed conditions, compared to the green and purple plots (without GUM).

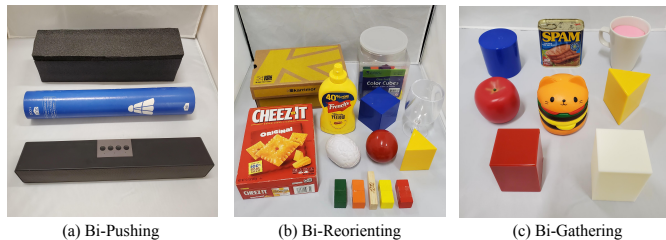


Figure 4. The objects used in the three proposed dual-arm manipulation tasks for real-world testing: a) a tripod box, a shuttles tube, and a loudspeaker for the bi-pushing task; b) a plastic cube, a cracker box (red), a shoe box (yellow), a cubes box (transparent), a mustard bottle, a goblet, a plastic ball, a soft brain toy, a plastic triangle prism, and a set of bricks for the bi-gathering task; c) a blue cylinder, a spam can, a ceramic mug, an apple, a foam toy, a triangle prism, and two cubes for the bi-gathering task. These objects are unseen during training and are chosen to vary in size, weight, shape and stiffness.

found that this simple curriculum was critical for successful training in simulation and sim-to-real application. During the curriculum, we also increased the probability of perturbation and its magnitude to further improve the policy’s robustness when applied in the real world.

IV. EXPERIMENTS AND RESULTS

A. Evaluations on the Bi-Touch in Simulation

An on-policy model-free deep-RL algorithm called Proximal Policy Optimization (PPO) [24] is used to train policies in simulation for all three bimanual tactile robotic tasks described above. Specifically, we use the Stable-Baselines-3 [25] implementation of PPO for learning the policies.

We obtained successfully trained policies in simulation for all three tasks (training curves in Figs 3a-c respectively). The bi-pushing is the easiest task to learn with a smooth learning

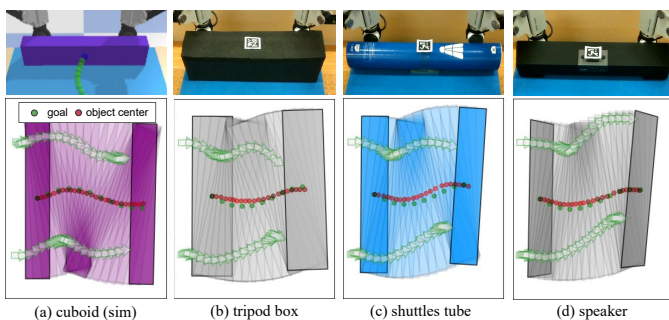


Figure 5. The objects’ trajectories plots (in red) of the bi-pushing task with (a) a cuboid (simulation), (b) a tripod box, (c) a shuttles tube and (d) a loudspeaker. The TacTips’ trajectories are indicated with green arrows.

curve and convergence at an early time step. The other two tasks have subtleties in learning that we describe below.

1) *Bi-Pushing*: In each episode, we selected a sequence of goals from a sampled linear or sinusoidal path as a desired object trajectory (parameterised by $y = kx$ and $y = a \sin(x/50)$ respectively, where $k \in [-0.28, 0.28]$, $a \in [-20, 20]$ and $x \in [-280, 50]$ mm). The simulated dual-arm tactile robot can learn to push a large object through a trajectory with a training curve shown in Fig. 3a. A successful example is shown in Fig. 5a. The accuracy from 20 simulated tests (10 for each type of trajectory) is 12.3 ± 4.8 mm (Table I, top row).

2) *Bi-Reorienting*: In each episode, the object length l is uniformly sampled from $[50, 200]$ mm and a goal angle θ^g is uniformly sampled from $[30^\circ, 90^\circ]$. Note that the goal angle is evenly divided into 10 subgoals as curriculum learning. The simulated dual-arm tactile robot can learn to reorient various lengths of objects to goal angles with the proposed reward design described in Sec. III-C, with a training curve shown in Fig. 3b. A successful example is shown in Fig. 6a. Note that the translation error is calculated by subtracting the final position of the object from the starting position $p_{t=0}^o$. The average translation and orientation errors from 10 simulated tests are 10.2 ± 4.8 mm and $3.4 \pm 1.8^\circ$ respectively (Table II).

3) *Bi-Gathering*: In each episode, the two object’s o^i ($i = \{1, 2\}$) initial positions $p_0^{o^i} = (x^i, y^i)$ are uniformly sampled from ranges $x^i \in (-1)^i [50, 200]$ mm, $y^i \in (-1)^i [0, 100]$ mm, respectively. The other hyperparameters are the termination distance ($\epsilon = 90$ mm; approximate object size), the number of points for subgoals ($n = 10$) and updating the subgoal target line every $h = 75$ steps. The results (without perturbations) demonstrate that the proposed GUM improves the performance of the simulated dual-arm tactile robot in achieving this task more robustly in shorter episode times (Fig. 3c, blue plot), as compared to the one without this mechanism (green plot). This indicates its ability to learn and adapt more effectively: even though the exact object poses are unknown to the robot, it learns to bring together the object locations with only tactile and proprioceptive feedback.

When perturbations are also considered, these are uniformly sampled from $[1, 5]$ N. In that situation, without GUM, the dual-arm tactile robot cannot learn to achieve this task, with longer episode times compared to the one without perturbations (Fig. 3c, green and purple plots respectively). With the use of GUM, the task is learned successfully even in the presence of perturbations, reaching a performance similar to that achieved without perturbations (Fig. 3c, red and blue plots respectively). A successful example with perturbations is shown in Fig. 7a. Upon testing the policy trained under

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

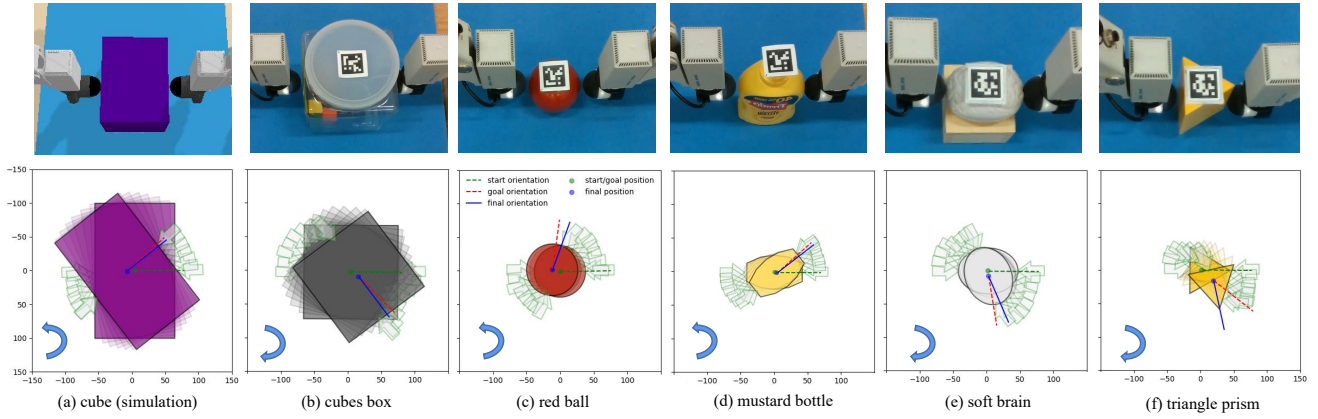


Figure 6. Object trajectories for the bi-reorienting task for the (a) cuboid (simulation), (b) cubes box, (c) plastic ball, (d) mustard bottle, (e) soft brain toy and (j) triangular prism. End-effector positions during the trajectories (green arrows) are shown, and the start, goal and final object orientations (lines).

Table I

Mean errors and standard deviations of the object trajectories from the ground truth for the bi-pushing task. The object size is also shown.

| Object | Accuracy | Size |
|----------------------------|-------------------|--------|
| Cuboid (Simulation) | 12.3 ± 4.8 mm | 400 mm |
| Box | 14.2 ± 6.4 mm | 351 mm |
| Tube | 16.6 ± 7.7 mm | 386 mm |
| Loudspeaker | 17.4 ± 8.1 mm | 377 mm |

perturbations with the goal-update mechanism, the success rates are 100% with different perturbation times in each of 5 simulated tests (Table III, top row).

B. The Performance of the Bi-Touch in Reality

1) *Bi-Pushing*: We set up the real bi-pushing task with the same configuration as the simulated one, testing it with three unseen objects (Fig. 4a) varying in weight and contact shape. The real dual-arm tactile robot can push a large object through a desired trajectory (example of sinusoidal trajectories in Fig. 5b-d). The accuracy from 20 real-world tests (10 for each type of trajectory) for the tripod box, the shuttle tube, and the loudspeaker are 14.2 ± 6.4 mm, 16.6 ± 7.7 mm, and 17.4 ± 8.1 mm respectively (Table I), compared to an overall distance travelled of 300-420 mm. The performances on all objects is similar despite notable differences in their contact shapes (e.g. flat, curved, and sloping surface), showing the generalization ability of the learned policy. Videos for these trajectories are provided in the supplementary materials.

Comparing the simulated and real trajectories of the end effectors (TacTips), while the box and tube behaved similarly (Fig. 5, three left plots), the loudspeaker travelled more distance (right-most plot). We attribute this behaviour to the larger frictional force due to the loudspeaker's heavier weight (879 g) compared with the box (244 g) and tube (127 g). The dual-arm robot needs to move its end-effector closer to the end of the loudspeaker to better counteract the larger frictional force and maintain the object's centre on the goal path. These behaviours show that the dual-arm robot learns to interact with objects of various weights using its tactile feedback, demonstrating the generalization ability and robustness of the learned policy.

2) *Bi-Reorienting*: The real bi-reorienting task is considered with the same configuration as the simulated task, testing with a set of new objects that vary in size, shape, weight, and

Table II

Mean errors and standard deviations of the reorientation angles and the positions of the objects from the ground truth for the bi-reorienting task.

| Object | Translation Err. | Orientation Err. | Size |
|-------------------------|-------------------|----------------------|--------|
| Cube (Sim.) | 10.2 ± 4.8 mm | $3.4 \pm 1.8^\circ$ | 100 mm |
| Plastic Cube | 12.5 ± 5.3 mm | $7.5 \pm 3.9^\circ$ | 100 mm |
| Shoe Box | 16.2 ± 6.8 mm | $11.5 \pm 5.2^\circ$ | 193 mm |
| Cracker Box | 13.9 ± 5.6 mm | $8.4 \pm 4.7^\circ$ | 172 mm |
| Cubes Box | 15.5 ± 5.5 mm | $9.2 \pm 5.6^\circ$ | 186 mm |
| Mustard Bottle | 13.6 ± 5.4 mm | $7.6 \pm 4.1^\circ$ | 122 mm |
| Goblet | 18.1 ± 6.1 mm | $12.2 \pm 5.8^\circ$ | 113 mm |
| Soft Brain Toy | 17.2 ± 6.7 mm | $9.7 \pm 5.3^\circ$ | 85 mm |
| Red Ball | 19.5 ± 7.0 mm | $13.4 \pm 6.5^\circ$ | 70 mm |
| Triangular Prism | - | - | 57 mm |

stiffness (Fig. 4b). We trained different policies for different directions. Note that the object position is defined as the top centre position where the ArUco marker is attached. The robot is considered to achieve the task when the object angle does not change more than 1° in 10 time steps for the final subgoal. The robot achieved this task with most of the selected objects with translation error from 12.5 ± 5.3 mm to 19.5 ± 7.0 mm, and orientation error from $7.5 \pm 3.9^\circ$ to $13.4 \pm 6.5^\circ$ (Table II), except the triangular prism where there was a problem with the sharp edge. Trajectories for a set of real-world objects are shown in Fig. 6 and videos are provided in the supplementary material.

Specifically, the real dual-arm robot can achieve higher accuracy with cube-shaped objects (plastic cube, shoe box, cracker box, mustard bottle, cubes box) compared to round-shaped objects (goblet, plastic ball, soft brain toy). This is because the policy is only trained with cube-shaped objects in simulation so the policy does not have knowledge of the object geometry. We also noticed that the larger the manipulated object is, the larger error will present. We attribute this to the increased difficulty for coordinating both arms to manipulate the larger objects as it requires the end effector to travel more distance during the task. Even so, it was able to successfully manipulate various unseen round objects of various stiffness, which demonstrates the learned policy's generalization ability to rounder objects. On the other hand, we attribute the difficulties on the prism to slippage caused by an unstable contact on the sharper edge, leading to an unrecoverable situation that was never experienced in training.

To further examine the generalization capability, we tried an additional experiment on a set of bricks. Specifically, the

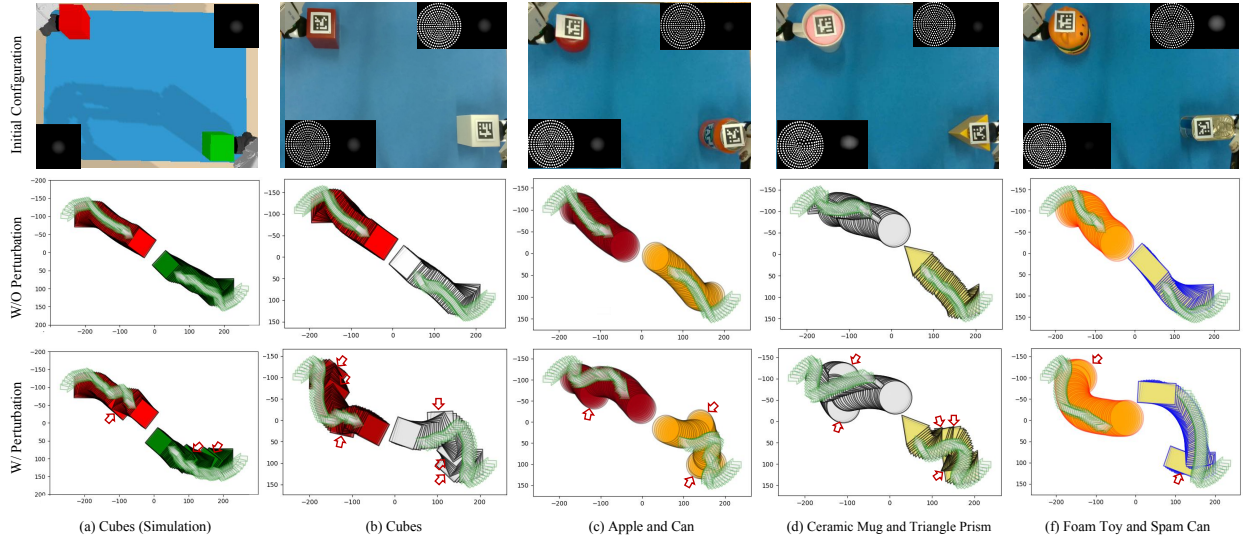


Figure 7. Bi-gathering on one simulated pair (a) and four real pairs of objects, corresponding to the (b) cubes, (c) plastic apple and can, (d) ceramic mug and triangular prism and (e) soft foam toy and spam can. The first row shows the initial configurations of the objects and the end effectors (TacTips), along with examples of real and simulated tactile images. The second and third rows show the end-effector (green arrows) and object trajectories. The red arrows in the third row show applied force perturbations. Note that ArUco markers are just for quantitative evaluation, as our policy does not use visual feedback.

bricks are loosely placed together initially, then the robot needs to reorient and gather them together by pushing and rotating the bricks it is contacting. The robot is able to successfully complete this task and, moreover, during the reorientation we can remove the middle brick to demonstrate the robustness to further perturbation. The robot can achieve this (see the video in supplementary material) with a 95% success rate in 20 tests. Overall, this demonstrates that the policy can lead to a new emergent behaviour not experienced during training.

3) *Bi-Gathering*: The real bi-gathering task is considered with the same configuration as in the simulation, with testing on a set of new objects (Fig. 4c) varying in shape, weight and stiffness. Here we applied the policy trained under perturbations in simulation with the target-line goal-update mechanism. We run 10 tests for each object pair with random force perturbations applied at random times. The robot is considered to achieve the task when the centre of the objects becomes closer than 7 cm (detected using ArUco markers) within 300 time steps. Note that object lengths are about 6 cm.

The tactile robot successfully completed the bi-gathering task without perturbation in all sets of 10 trials for each object. Regarding the effect of perturbations, the success rates of achieving the task are summarized in Table III. The robot completes the task at 100% success rate when the perturbation is applied twice or fewer times. The success rate decreases when the number of applied perturbations is increased for all pairs of objects, decreasing most for the irregular items (mug, triangular prism, foam toy and spam can). We show example trajectories of successful trials in Fig. 7.

Examining the results more closely, the dual-arm tactile robot can complete the task with objects of different weights and shapes to those during training. Comparing the foam toy to the cube in simulation (Figs. 7f, a), the TacTip is deformed much more lightly in reality than in simulation, due to the foam toy’s light weight and soft material. Even so, the robot can achieve the task with generated tactile images that are

Table III
Success rates for bi-gathering under different numbers of perturbations.

| Objects Pairs | No. Perturb. | | | | | |
|-------------------------------------|--------------|------|------|------|------|------|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| Cube & Cube (Simulation) | 100% | 100% | 100% | 100% | 100% | 100% |
| Cube & Cube | 100% | 100% | 100% | 100% | 90% | 90% |
| Apple & Can | 100% | 100% | 100% | 100% | 90% | 70% |
| Mug & Triangular Prism | 100% | 100% | 80% | 70% | 40% | 10% |
| Foam Toy & Spam Can | 100% | 100% | 70% | 50% | 20% | 10% |

notably different from those in training, both in terms of depression and in the contact shapes (e.g. the mug handle in Fig. 7d). These results demonstrate the strong generalization ability of the policy trained with our goal-update mechanism among different stiffness, shapes and object weights.

We observed that the failure cases are caused by the robot running out of workspace as it tried to push a perturbed object back together through a large turnaround. This happened mostly with the spam can and mug since they present additional challenges for pushing due to their larger weights. Even so, the dual-arm tactile robot still demonstrated a capability to push perturbed objects together (Fig. 7, third row).

V. DISCUSSION AND FUTURE WORK

In this paper, we developed a low-cost dual-arm tactile robot system called Bi-Touch for sim-to-real deep reinforcement learning based on Tactile Gym 2.0 [8]. The hardware includes two industry-capable desktop robot arms (Dobot MG400), each equipped with a low-cost high-resolution optical tactile sensor (TacTip) as end-effectors. We also designed a workspace configuration suited for three proposed bimanual tasks tailored towards tactile feedback and integrated into the Tactile Gym simulation methods and environments.

The performance of our low-cost sim-to-real deep RL dual-arm tactile robot system was evaluated in these three bimanual tasks in the real world. We introduced appropriate reward functions for these tasks in simulation, then investigated how

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

these policies apply to the real world. The experimental results show that the developed dual-arm tactile system is effective for all tasks on real objects unseen in the simulation learning.

For bi-gathering, we proposed a goal-update mechanism using a linear set of subgoals. These subgoals provide dense auxiliary rewards, enabling the dual-arm robot to learn from a high-dimensional state-action space. While this idea shares similarities with Hierarchical Visual Foresight (HVF) [26], we faced challenges in directly applying it to our tactile manipulation tasks due to the limited goal state information provided by tactile images. Instead, we simplified the problem by generating subgoals based on the current object positions in simulation or TCP positions in the real world. This mechanism, which preserves the distribution of initial environmental states, can be seen as a form of implicit curriculum learning [27]. It periodically generates goals that are easier to achieve compared to sparse goal settings where only object positions serve as goals. Unlike Hindsight Experience Replay [27], which is applicable only to off-policy RL methods, our method applies on-policy, such as PPO used in this study.

While here we demonstrate the feasibility of our work with 4-DoFs desktop robot arms and TacTips [21], the proposed framework should also work with different types of optical tactile sensors, e.g. the GelSight [28] and robots with more DoFs, as the Tactile Gym has demonstrated its scalability in both situations [7], [8]. A limitation of the present work is that shear deformation of the tactile sensor has not yet been considered in the simulation and may be needed for more complex tasks relying on the control of frictional forces. A future direction is to approximate the shear effect in simulation with reliable methods to further close the sim-to-real gap.

A future direction of this work is to apply our Bi-Touch framework and tactile robot platform to fine manipulation of held objects without a supporting table. To show the promise of this approach, we additionally developed a bimanual lifting (bi-lifting) task along with a reward function based on principles similar to the other three tasks: a) manipulating the object to achieve a goal state with tactile feedback, b) while maintaining stable contact with the object. Preliminary experiments for the bi-lifting task, both in simulation and the real world, demonstrate that this approach to dexterous manipulation can be effective, which we have included in a supplementary video that accompanies this paper.

REFERENCES

- [1] C. Smith, Y. Karayiannidis, L. Nalpantidis, X. Gratal, P. Qi, D. V. Dimarogonas, and D. Kragic. Dual arm manipulation—a survey. *Robotics and Autonomous systems*, 60(10):1340–1353, 2012.
- [2] J. Hansen, F. Hogan, D. Rivkin, D. Meger, M. Jenkin, and G. Dudek. Visuotactile-rl: Learning multimodal manipulation policies with deep reinforcement learning. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 8298–8304. IEEE, 2022.
- [3] D. Kruse, J. T. Wen, and R. J. Radke. A sensor-based dual-arm tele-robotic system. *IEEE Transactions on Automation Science and Engineering*, 12(1):4–18, 2015.
- [4] A. Sintov, S. Macenski, A. Borum, and T. Bretl. Motion planning for dual-arm manipulation of elastic rods. *IEEE Robotics and Automation Letters*, 5(4):6065–6072, 2020.
- [5] S. James, P. Wohlhart, M. Kalakrishnan, D. Kalashnikov, A. Irpan, J. Ibarz, S. Levine, R. Hadsell, and K. Bousmalis. Sim-to-real via sim-to-sim: Data-efficient robotic grasping via randomized-to-canonical adaptation networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12627–12637, 2019.
- [6] OpenAI: M. Andrychowicz, B. Baker, M. Chociej, R. Jozefowicz, B. McGrew, J. Pachocki, A. Petron, M. Plappert, G. Powell, A. Ray, et al. Learning dexterous in-hand manipulation. *The International Journal of Robotics Research*, 39(1):3–20, 2020.
- [7] A. Church, J. Lloyd, R. Hadsell, and N. Lepora. Tactile Sim-to-Real Policy Transfer via Real-to-Sim Image Translation. In *Proceedings of the 5th Conference on Robot Learning*, pages 1–9. PMLR, October 2021.
- [8] Y. Lin, J. Lloyd, A. Church, and N. Lepora. Tactile gym 2.0: Sim-to-real deep reinforcement learning for comparing low-cost high-resolution robot touch. volume 7 of *Proceedings of Machine Learning Research*, pages 10754–10761. IEEE, August 2022.
- [9] K. Zhang, M. Sharma, M. Veloso, and O. Kroemer. Leveraging multimodal haptic sensory data for robust cutting. In *2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*, pages 409–416. IEEE, 2019.
- [10] I. Garcia-Camacho, M. Lippi, M. C. Welle, H. Yin, R. Antonova, A. Varava, J. Borras, C. Torras, A. Marino, and G. Alenya. Benchmarking bimanual cloth manipulation. *IEEE Robotics and Automation Letters*, 5(2):1111–1118, 2020.
- [11] Z. Chiu, F. Richter, E. K. Funk, R. K. Orosco, and M. C. Yip. Bimanual regrabbing for suture needles using reinforcement learning for rapid motion planning. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 7737–7743, 2021.
- [12] F. Amadio, A. Colomé, and C. Torras. Exploiting symmetries in reinforcement learning of bimanual robotic tasks. *IEEE Robotics and Automation Letters*, 4(2):1838–1845, 2019.
- [13] S. Kataoka, S. K. S. Ghasemipour, D. Freeman, and I. Mordatch. Bi-manual manipulation and attachment via sim-to-real reinforcement learning. *arXiv preprint arXiv:2203.08277*, 2022.
- [14] J. Grannen, Y. Wu, S. Belkhal, and D. Sadigh. Learning bimanual scooping policies for food acquisition. *arXiv preprint:2211.14652*, 2022.
- [15] L. Fan, Y. Zhu, J. Zhu, Z. Liu, O. Zeng, A. Gupta, J. Creus-Costa, S. Savarese, and F. Li. Surreal: Open-source reinforcement learning framework and robot manipulation benchmark. In *Conference on Robot Learning*, pages 767–782. PMLR, 2018.
- [16] Y. Zhu, J. Wong, A. Mandekar, M. Roberto, J. Abhishek, N. Soroush, and Y. Zhu. robosuite: A modular simulation framework and benchmark for robot learning. In *arXiv preprint arXiv:2009.12293*, 2020.
- [17] Y. Chen, Y. Yang, T. Wu, S. Wang, X. Feng, J. Jiang, Z. Lu, S. M. McAleer, Dong H., and S. Zhu. Towards human-level bimanual dexterous manipulation with reinforcement learning. In *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2022.
- [18] N. Sommer, M. Li, and Aude Billard. Bimanual compliant tactile exploration for grasping unknown objects. In *2014 IEEE International Conference on Robotics and Automation*, pages 1–8. IEEE, 2014.
- [19] F. R. Hogan, J. Ballester, S. Dong, and A. Rodriguez. Tactile dexterity: Manipulation primitives with tactile feedback. In *2020 IEEE international conference on robotics and automation*, pages 1–7. IEEE, 2020.
- [20] N. F. Lepora, Y. Lin, B. Money-Coomes, and J. Lloyd. Digitac: A digitactip hybrid tactile sensor for comparing low-cost high-resolution robot touch. *IEEE Robotics and Automation Letters*, 7(4):9382–9388, 2022.
- [21] B. Ward-Cherrier, N. Pestell, L. Cramphorn, B. Winstone, M. E. Giannaccini, J. Rossiter, and N. Lepora. The TacTip Family: Soft Optical Tactile Sensors with 3D-Printed Biomimetic Morphologies. *Soft Robotics*, 5(2):216–227, April 2018.
- [22] N. Lepora. Soft Biomimetic Optical Tactile Sensing With the TacTip: A Review. *IEEE Sensors Journal*, 21(19):21131–21143, October 2021.
- [23] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1125–1134, July 2017.
- [24] J. Schulman, F. Wolski, A. Radford, and O. Klimov. Proximal policy optimization algorithms. *arXiv preprint:1707.06347*, 2017.
- [25] A. Raffin, A. Hill, M. Ernestus, A. Gleave, A. Kanervisto, and N. Dornmann. Stable baselines3, 2019.
- [26] S. Nair and C. Finn. Hierarchical foresight: Self-supervised learning of long-horizon tasks via visual subgoal generation. In *International Conference on Learning Representations*, 2020.
- [27] M. Andrychowicz, F. Wolski, A. Ray, J. Schneider, R. Fong, P. Welinder, B. McGrew, J. Tobin, P. Abbeel, and W. Zaremba. Hindsight experience replay. *Advances in neural information processing systems*, 30, 2017.
- [28] W. Yuan, S. Dong, and E. H. Adelson. Gelsight: High-resolution robot tactile sensors for estimating geometry and force. *Sensors*, 17(12):2762, November 2017.