

Variable-Friction In-Hand Manipulation for Arbitrary Objects via Diffusion-Based Imitation Learning

Qiyang Yan¹, Zihan Ding², Xin Zhou¹, and Adam J. Spiers¹

Abstract—Dexterous in-hand manipulation (IHM) for arbitrary objects is challenging due to the rich and subtle contact process. Variable-friction manipulation is an alternative approach to dexterity, previously demonstrating robust and versatile 2D IHM capabilities with only two single-joint fingers. However, the hard-coded manipulation methods for variable friction hands are restricted to regular polygon objects and limited target poses, as well as requiring the policy to be tailored for each object. This paper proposes an end-to-end learning-based manipulation method to achieve *arbitrary object* manipulation for *any target pose* on real hardware, with minimal engineering efforts and data collection. The method features a diffusion policy-based imitation learning method with co-training from simulation and a small amount of real-world data. With the proposed framework, arbitrary objects including polygons and non-polygons can be precisely manipulated to reach arbitrary goal poses within 2 hours of training on an A100 GPU and only 1 hour of real-world data collection. The precision is higher than previous customized object-specific policies, achieving an average success rate of 71.3% with average pose error being 2.676 mm and 1.902°. Code and videos can be found at: <https://sites.google.com/view/vf-ihm-il/home>.

I. INTRODUCTION

Humans effortlessly perform in-hand manipulation (IHM) for everyday tasks, such as reorienting and repositioning arbitrary complex objects like pens or keys without regrasping [1]. These actions involve controlled transitions and re-orientations that are crucial but challenging for robots. While impressive IHM tasks have been performed by dexterous hands through learning-based methods [2], [3], [4], [5], those platforms tend to suffer from mechanical complexity. This complexity translates to extremely high hardware and maintenance costs as well as control complexity.

In comparison, the 2-DOF variable-friction (VF) hand of [6] can achieve high object-dexterity despite its simple morphology. Inspired by the ability of human fingertips to selectively slide and grip objects, the VF-hand can achieve the same by actively switching its finger surface between high and low friction states. This allows objects to be slid or rolled along the finger surfaces.

Several manipulation-planning strategies have been developed for VF grippers [7], [8]. The current state-of-the-art is a hybrid of planning and visual servoing [7] via an offline A* pathfinding algorithm to determine the most efficient trajectory through the combination of hard-coded

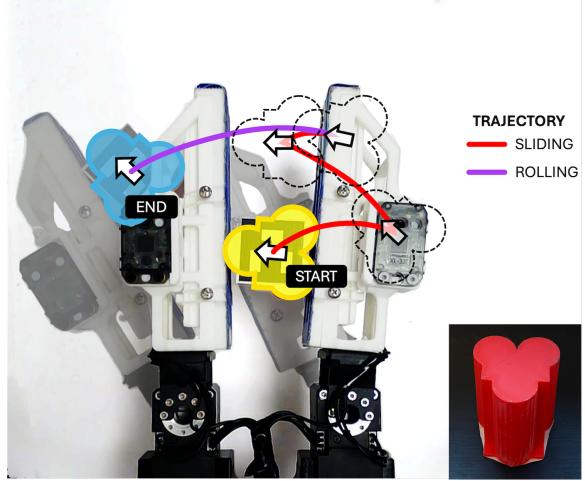


Fig. 1. The manipulation trajectory of a complex Three-Cylinder object using the learned control policy. An Aruco marker obscures the body of the object, which is why the shape has been superimposed.

model-based manipulation ‘primitives’. While this technique demonstrates the potential of variable friction hands, it also suffers from several limitations: (a) The model-based controller restricts manipulation to simple polygonal objects, such as cube and hexagonal prisms. (b) Rotational primitives only allow rolling an object from one flat surface to another. (c) Target poses (orientation and location) are restricted to positions where a flat surface of the object is in contact with a finger.

We address these shortcomings via a new learning-based approach for VF-manipulation. An end-to-end data-efficient learning framework has been developed, based on imitation learning with diffusion policy. This enables the VF gripper to manipulate complex object shapes from arbitrary initial poses to arbitrary target poses in a continuous manipulation range. While our work focuses on the VF gripper, we believe this methodology is well suited to other uncommon gripper morphologies with non-holonomic object motion constraints.

In addition, we determined that imitation learning with diffusion policy [9] can effectively learn high-precision policies, even with demonstrations generated via an imprecise RL policy optimized for manipulation smoothness. Finally, we have demonstrated that co-training the diffusion policy with a mix of real and simulated data increases both task success rate and resource efficiency.

II. RELATED WORK

Reinforcement learning (RL) learns optimal policies through interactions, and has shown promising results for in-hand manipulation (IHM) in recent years. Some work

*Research supported by Imperial College London internal funds

¹Manipulation and Touch Lab, Department of Electrical and Electronic Engineering, Imperial College London, UK
a.spiers@imperial.ac.uk

²Electrical and Computer Engineering Department, Princeton University, US.

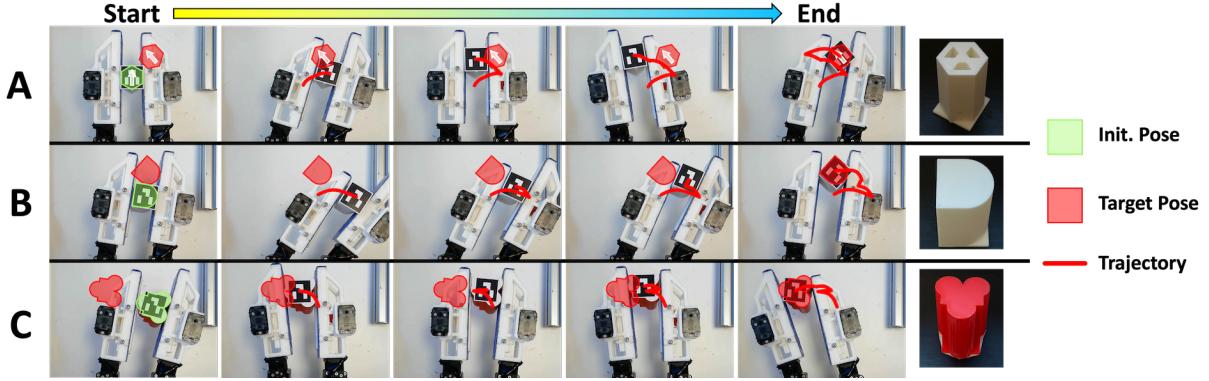


Fig. 2. In-hand manipulation of complex objects using the variable-friction hand, trained through our diffusion-based co-train imitation learning pipeline.

utilizing model-based approaches with explicit environment representations [10], [11], [12], while model-free approaches learn policies directly [13], [2], [14]. IHM RL policy training is commonly performed in simulation [15], as collecting enough interaction samples in the real world is often too resource intensive. However, training in simulation still requires a large amount of computing power (usually performed on CPU / GPU clusters) and training time due to IHM's task complexity. Imitation Learning (IL) is a frequently-used method to reduce the required training effort for learned IHM policies [16] with growing popularity for IHM applications [17]. In addition, the behavior cloning in IL does not require carefully engineered simulation environments (no reward function is needed and it is more forgiving regarding action & observation space). Robot hands trained via this method learn from expert demonstrations, often generated via observing human object interactions [18], [19], [4], [20], teleoperation systems [21], [22], RL policies [3], or hand-crafted control strategies [23].

Recently, IL with the diffusion policy [9], [24] (which generates manipulation behaviour from demonstrations via a denoising diffusion process) has shown to be very effective. General tasks such as object-transportation or reorientation [9], and even complex domestic tasks such as cooking or chair rearranging [25] have been successfully tested with this approach. In the context of IHM, [5] presented a complex hand with 4 fingers, each consisting of a delta mechanism, which used the diffusion policy to perform basic tasks such as cap twisting and syringe pushing.

III. PROBLEM FORMULATION

A. Imitation Learning with Diffusion Policy

The manipulation policy is formulated as a Denoising Diffusion Probabilistic Model (DDPM) [26], which effectively captures the multi-modal distribution through a denoising process. For matching data distribution with probability $p(x_0)$, the forward diffusion process iteratively adds Gaussian noise ε to the data sample such that at step k , we have $x_k = \sqrt{\bar{\alpha}_k}x_0 + \sqrt{1 - \bar{\alpha}_k}\varepsilon$, $1 \leq k \leq K$ with noise $\varepsilon \sim \mathcal{N}(0, \mathbf{I})$ and $\bar{\alpha}_k = \prod_{i=1}^k \alpha_i$ being a predetermined noise schedule. The

reverse denoising process follows:

$$x_{k-1} = \frac{1}{\sqrt{\alpha_k}}x_k - \frac{1 - \alpha_k}{\sqrt{1 - \bar{\alpha}_k}\sqrt{\alpha_k}}\varepsilon_\theta + \mathcal{N}(0, \sigma_k^2 \mathbf{I}) \quad (1)$$

as the posterior distribution of Gaussian following the Bayes' rule, with variance $\sigma_k = \frac{(1 - \alpha_k)(1 - \bar{\alpha}_{k-1})}{1 - \bar{\alpha}_k}$. ε_θ is the predicted noise by a model parameterized by θ , which can be optimized with the following diffusion loss:

$$\mathcal{L} = \mathbb{E}_{x_0 \sim \mathcal{D}, \varepsilon \sim \mathcal{N}(0, \mathbf{I}), k \in \{1, \dots, K\}} (\varepsilon - \varepsilon_\theta(\sqrt{\bar{\alpha}_k}x_0 + \sqrt{1 - \bar{\alpha}_k}\varepsilon, k))^2 \quad (2)$$

The diffusion policy use DDPM to approximate the conditional distribution $p(a|o)$, i.e., predicting action a conditioned on observation o . For imitation learning, the dataset $\mathcal{D} = \{(o, a)\}$ is pre-collected with a given behavior policy π^b , and the current diffusion policy $\pi_\theta(a|o)$ is optimized to minimize the distribution divergence from the behavior policy by Eq. (2).

B. Reinforcement Learning

For RL, we define a Markov decision process $(\mathcal{O}, \mathcal{A}, R, \mathcal{T}, \gamma)$, where \mathcal{O} is the observation space, \mathcal{A} is the action space, $R(o, a) : \mathcal{O} \times \mathcal{A} \rightarrow \mathbb{R}$ is the reward function, $\mathcal{T}(o'|o, a) : \mathcal{O} \times \mathcal{A} \rightarrow \text{Pr}(\mathcal{O})$ is the stochastic transition function, and $\gamma \in [0, 1]$ is the discount factor for value estimation. The policy in RL is optimized by maximizing its discounted cumulative reward: $\mathbb{E}_\pi[\sum_{t=0}^{\infty} \gamma^t r(o_t, a_t)]$.

IV. SYSTEM OVERVIEW

A. Hardware Setup

1) *Variable Friction Hand*: The VF hand in this work builds on the design of [6] and is shown in Fig. 3. This hand is 3D printed in PLA material. Two Dynamixel XM-430 servo motors actuate the two rotary fingers, while two lower-cost Dynamixel XL-320 servos switch between low- and high-friction finger surfaces via a pulley-cam mechanism described in [6]. Objects are manipulated via a ‘push-pull’ control scheme: the pulling finger is in position control mode and ‘pulls away’ from the object, whilst the pushing finger is in current control mode, pushing the object firmly against the other finger with a constant torque. During this motion, objects slide along the finger surface if the finger is in low-friction mode (Fig. 3B); if in high-friction mode,

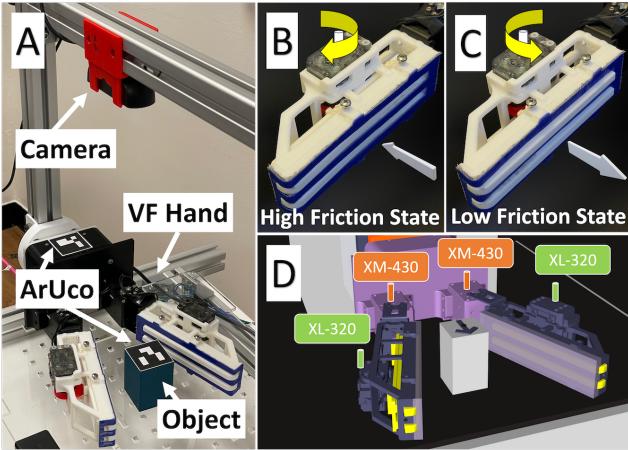


Fig. 3. A: The Variable Friction hand mounted on the training & testing rig. B: When both fingers are in high-friction mode, objects pivot. C: When one of the fingers is in low-friction mode, objects slide along the finger. D: Rendering of the hand and object in the MuJoCo Simulation Environment.

objects typically pivot around the contact point (Fig. 3C). This behaviour is demonstrated in the accompanying video.

The variable friction hand is mounted on an aluminium strut frame shown in Fig. 3A. An optional transparent acrylic board was utilized during experiments such that all ArUco markers at the same height, ensuring visual tracking reliability. Our experiments verify the policy performances with and without object support, as demonstrated on the website. A generic 1080p webcam is mounted on the aluminium strut above the hand and is used for tracking object poses. An AruCo marker [27] is placed on each manipulated object, and its pose (X and Y location & Z orientation) is inferred relative to a reference AruCo marker placed on the hand's base. The visual tracking system is calibrated and implemented following [27]. The tracking system was evaluated to have a mean absolute error of approximately 1 mm with a standard deviation of 0.75 mm.

B. Simulation Environment

The simulation environment, built in the MuJoCo engine, features the 2-DoF Variable Friction Hand and objects for manipulation and visualization (see Fig. 3D).

To simulate the friction pad's properties, MuJoCo's soft contact model is hardened by adjusting solver parameters (solref, solimp), using an elliptical friction cone, Newton solver, and a high impratio. The Noslip solver (optional) ensures no slip, matching the real gripper's behavior. The MultiCCD flag is enabled for more stable contact with meshes at an additional computational cost.

To minimize the sim-to-real gap, we set the physical, geometric, and dynamics parameters through system identification (SI) [28] by comparing the simulation and real trajectories as in [29]. Domain randomization [30], [29] has been applied to parameters in the policy training and data collection process, with their ranges identified during SI.

V. SIM-REAL CO-TRAINING FRAMEWORK

Our initial experiments testified the effectiveness of RL with careful reward engineering and domain randomization

for a single regular polygon (a cube). However, as displayed in Fig 1, the successful trajectory for variable friction in-hand manipulation does not follow a minimal straight-line distance principle. Hence, the reward function had to be designed specifically for each regular polygon object according to the individual contact surface and interior angle for precise IHM. Such reward shaping is infeasible for non-polygons.

For efficiently manipulating arbitrary objects towards any specified target pose, we propose the imitation learning approach outlined in Fig. 4. To capture the multi-modality in the behaviour policy dataset, we adopt the diffusion model for policy representation following [9]. The benefits of our IL pipeline over the baseline RL approach are manifold: (1) Our straightforward co-training method combines real and simulated demonstrations and effectively bridges the sim-to-real gap. (2) The heavy reward engineering needed for each object under the baseline RL-only approach is avoided. Through *hindsight goal relabeling* [31], any smooth trajectory in either simulation or the real world can be relabeled with the actual final pose to compose the demonstration dataset. (3) The restrictions on the object shape and target pose are removed. (4) The IL policy training is significantly more computationally efficient than RL-only policy training (2 hours vs. 15 hours in our experiments).

A. Manipulation Policy Formulation

The IL and RL manipulation policies $\pi(a|o) : \mathcal{O} \rightarrow \text{Pr}(\mathcal{A})$ share a common formulation, which is detailed as follows.

a) Observation space: The observation space for RL $o_t \in \mathcal{O} \subset \mathbb{R}^{25}$ consists of the joint state, object state and goal information. For IL, the velocity information in the observation space is zero-padded due to the discrepancy of simulation and real-world velocity.

b) Hybrid action space: The hybrid action space $\mathcal{A} = \mathcal{A}_{\text{continuous}} \times \mathcal{A}_{\text{discrete}}$ is designed to allow the VF gripper to perform robust manipulation by switching among discrete friction modes, reducing exploration by focusing on task execution rather than finger coordination. This allows manipulation without support beneath the object even in the presence of sliding.

The discrete action space $\mathcal{A}_{\text{discrete}}$ specifies the high-level operating mode, specifically in which direction the fingers rotate and which friction modes are used (outlined in Sec. IV-A.1)). This results in 6 discrete actions (see Table I). The friction mode switching and finger movements are executed alternatively in each action step to avoid object slipping and motion instability.

The continuous action space $\mathcal{A}_{\text{continuous}}$ specifies the relative action a_t between the active finger's current joint angle (q_t^{active}) and its target joint angle (q_t^{target}):

$$q_t^{\text{target}} = q_t^{\text{active}} + a_t$$

The action range is constrained to within $[0, 18.9]$ degrees to limit the finger movement velocity, as a higher a_t at each time step would cause the position-controlled leading finger to accelerate excessively. Note that a_t is never negative, as

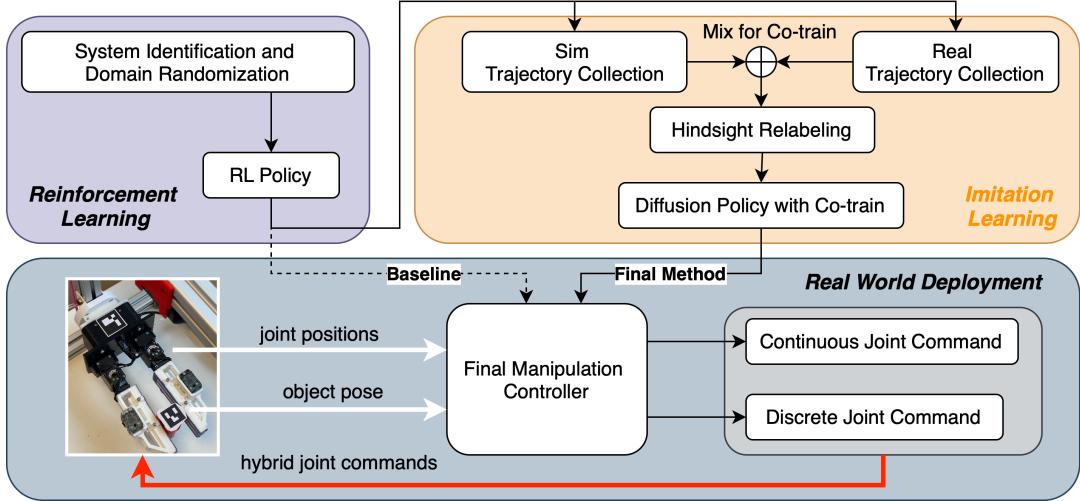


Fig. 4. Training Framework of our Co-Train IL method. The IL policy is represented as a diffusion model, and trained with a mix of simulation and real data. The RL policy is used to generate demonstrations for IL, but also used as a baseline during performance analysis.

a_{discrete}	Action Type	L	R
0	Slide up on right finger	PC+HF	TC+LF
1	Slide down on right finger	TC+HF	PC+LF
2	Slide up on left finger	TC+LF	PC+HF
3	Slide down on left finger	PC+LF	TC+HF
4	Rotate clockwise	TC+HF	PC+HF
5	Rotate anticlockwise	PC+HF	TC+HF

TABLE I

DISCRETE ACTION SPACE L: LEFT FINGER. **R:** RIGHT FINGER. **PC:** POSITION-CONTROLLED. **TC:** TORQUE-CONTROLLED. **HF:** HIGH-FRICTION. **LF:** LOW-FRICTION

the leading finger is per definition moving in one direction (the leading finger becomes the following finger once the movement direction is reversed). The control frequency for the above system is 2.5 Hz.

B. Smooth Manipulation via Reinforcement Learning

The hybrid action space is handled by a multi-head model architecture for both continuous and discrete prediction. As a key component in RL policy training, the reward function for a precise RL policy required heavy manual design for each object, which was feasible for only regular polygonal objects. Thus we opted for a reward function to obtain a RL policy capable of smoothly manipulating unseen objects with loosened precision requirements, significantly reducing training time.

1) *Reward function:* The task is successful if the Euclidean distance \bar{d} and angular $\bar{\theta}$ difference between the object's current and the goal pose is both smaller than the threshold value $\bar{d}, \bar{\theta}$, i.e. $\Delta d \leq \bar{d}, \Delta \theta \leq \bar{\theta}$. This comprises the sparse success reward. However, only with sparse reward the policy learning can be less efficient due to lack of dense reward signals. Considering the gripper's special side-to-side movement pattern (Fig. 1), we choose to also use the angular coordinates (r, θ) apart from the Euclidean coordinates for the object pose. Specifically, we use $\Delta r_t = (|r_1 - r_{g1}| + |r_2 - r_{g2}|)/2$ as the dense reward function, where r_1 and r_2 are the distances from the two finger bases to the object's

current position, while r_{g1} and r_{g2} are the same distances but for the goal position. We also penalized the agent for manipulating the object out of the legal operation range with a sparse penalty term. The complete reward function is mathematically expressed as:

$$\begin{aligned} r &= c_1 \mathbb{1}(|\Delta \theta_t| < \bar{\theta} \wedge |\Delta d_t| < \bar{d}) && \text{sparse success reward} \\ &\quad - c_2 |\Delta r_t| && \text{dense task reward} \\ &\quad - c_3 \mathbb{1}(\text{Object out of legal range}) && \text{sparse penalty} \end{aligned}$$

where $c_1, c_2, c_3 > 0$ are weighting coefficients, and $\mathbb{1}$ is an indicator function that identifies whether the condition within the bracket is satisfied.

2) *Training Details:* The RL algorithm applied in our experiments is TD3 [32] with Hindsight Experience Replay [31], with parallel environments (44 CPU cores) for data sampling. The same RL training setting is applied to both our IL approach and the baseline RL policy for comparable results. The exploration policy is trained on the Cube object only, taking roughly 3.5 hours; while the baseline RL is trained for at least 15 hours to converge to a high success rate with the same level of precision for each object. The episodes are terminated if the object does not reach success within performing 10 action steps. This led to efficient and smooth (but imprecise) manipulation policies for arbitrary unseen objects. This smoothness-optimized policy is then used to generate expert demonstrations described further in the next subsection.

C. Collecting Demonstrations with Hindsight Relabeling

The demonstration dataset is collected with the smoothness-optimized RL policy in both simulation and reality. We adopt *hindsight goal relabeling* [31], [33] to collect demonstration trajectories. The final state of a sampled trajectory is relabeled as the goal state and treated as an expert demonstration. The major benefit of this method is that it does not require the RL policy to be optimal for arbitrary objects and specified targets, as long

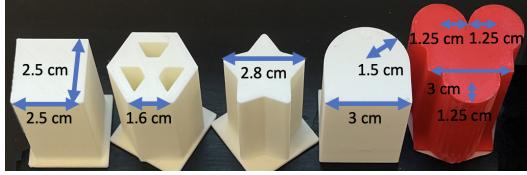


Fig. 5. The objects used in the experiments. From left to right: Cube, Hexagon (regular polygons), Star (irregular polygon), Cube Cylinder and Three-Cylinder (non-polygons)

as the trajectory is smooth within the task domain. The behavior policy is not restricted to a RL policy.

D. Sim-Real Co-Trained Diffusion Imitation Learning

For the next stage, we use behaviour cloning with a diffusion policy [9] to generate object-specific precise manipulation policies. We introduce the sim-real co-training framework for training the diffusion policy, using demonstrations generated via the above hindsight goal relabeling approach. For each object, we collect 100 real demos $\mathcal{D}_{\text{real}}$ (1 h/object), and 10,000 simulation demos \mathcal{D}_{sim} (30 min/object).

The training objective for the diffusion policy is defined as:

$$\min_{\theta} \mathbb{E}_{(o^i, a^i) \sim \mathcal{D}_{\text{sim}} \cup \mathcal{D}_{\text{real}}} [\mathcal{L}_{\theta}(o^i, a^i)] \quad (3)$$

$$= \min_{\theta} \mathbb{E}_{\substack{(o^i, a^i) \sim \mathcal{D}_{\text{sim}} \cup \mathcal{D}_{\text{real}} \\ \epsilon \sim \mathcal{N}(0, I), k \in \{1, \dots, K\}}} (\epsilon - \epsilon_{\theta}(\sqrt{\bar{\alpha}_k} a^i + \sqrt{1 - \bar{\alpha}_k} \epsilon, o^i, k))^2 \quad (4)$$

where o^i is the observation consisting of object goal pose, joints and object state, $a^i \in \mathbb{R}^2$ aligns with the action space defined for RL, and \mathcal{L} is the diffusion loss following Eq. (2). ϵ_{θ} is θ -parameterized diffusion policy for noise prediction. We sample with equal probability from the sim data \mathcal{D}_{sim} and real data $\mathcal{D}_{\text{real}}$ with a batch size of 1024. Observations are normalized based on the statistics of the real data distribution.

VI. REAL-WORLD EXPERIMENTS

A. Experiment setting

As shown in Fig. 5, we designed 5 objects of varying shapes, including regular and irregular polygons, as well as non-polygons. The objects are 3D-printed with the same geometries as their simulation counterparts, but surface properties like friction may differ. The expert demonstrations were collected via the same exploration RL policy trained on the Cube, but final precise IL policies are object-specific.

Performance was evaluated on the real-world platform outlined in Sec. IV-A. We average performance across 3 random seeds with 10 trials each and state standard deviations. Three performance measures were used: success rate of objects ending within 5 mm and 5.73 degrees (0.1 radians) from targets with a maximum of 10 action steps; positional error (center-distance to target) for successful episodes; and rotation error for successful episodes.

B. IL Performance Results

1) Main Results: As shown in Table. II, our method successfully manipulates the five test objects to arbitrary goal poses sampled within range with a success rate of 71.3% and an average positional error of 2.676 mm and a rotational error of 1.902 degrees for successful attempts. Notably, the error levels remain consistent across objects of varying complexity, which we attribute to system error. Fig. 7 visualizes these success rates tested on the real-world VF Hand.

TABLE II
FINAL CO-TRAINED IL METHOD (10,000 SIM AND 100 REAL DEMOS)
TESTED ON THE VF HAND

Object	Success Rate (%) In Reality	Position Error (mm)	Rotation Error (degree)
Cube	76.7 ± 20.8	2.550 ± 0.910	1.260 ± 0.998
Hexagon	76.7 ± 12.5	2.415 ± 1.195	2.105 ± 1.259
Star	73.3 ± 12.5	2.847 ± 1.218	2.153 ± 1.301
Cube Cylinder	70.0 ± 10.0	2.997 ± 1.295	1.432 ± 1.216
Three-Cylinder	60.0 ± 26.5	2.573 ± 1.171	2.561 ± 2.022
Average	71.3 ± 16.5	2.676 ± 1.158	1.902 ± 1.359

We observed that the agent can execute effective trajectories toward unseen target poses sampled in distribution at test time. It achieves the goal pose with at most 4 friction changes in most trials, leading to fast task completion with additional benefits of reduced hardware wear and tear.

2) IL vs. Baseline RL Policy: As an RL-only policy following Sec. V-B required heavy reward engineering for each object, we only conducted this comparison on the Cube. To achieve high success rates, the RL baseline takes 15 hours of training; while the exploration RL policy for our IL method takes only 3.5 hours. From Table. III, we observe that our final IL method with co-train achieves a higher success rate with similar levels of errors for successful episodes, but requires drastically less training time (36.7%). To manipulate unseen objects, only 2 extra hours of IL are needed as the exploration RL policy is shared.

TABLE III
PERFORMANCE COMPARISON OF REWARD-ENGINEERED RL AND OUR
FINAL IL FOR CUBE MANIPULATION ON THE REAL HAND

Method	Success Rate	Training Duration	Real Error (Pos./Rot.)
Baseline RL	66.7%	15h	(2.094 ± 0.863) mm / (3.032 ± 1.814)°
Our Method	76.7%	3.5h(RL) + 2h(IL)	(3.336 ± 1.211) mm / (2.163 ± 1.645)°

3) IL vs. Model-based Policy: Previous work for variable friction IHM tasks adopts the A* planner method [7], where manipulation primitives are hard-coded to achieve only face-to-face rolling for restricted target pose reaching, therefore A* is infeasible for arbitrary target poses. Our IL approach does not have such constraints. In Table IV, we compare the two methods under two settings: for 4 pre-specified goal poses (same as in [7]) and for arbitrary poses. Across the Cube and Hexagon, our method demonstrates a lower positional error compared to A* planner, for even arbitrary goal poses.

TABLE IV
POSITIONAL ERROR (MM) OF THE A* METHOD AND OUR IL METHOD
ON THE REAL HAND

Object	4 Different Achievable Goals		Arbitrary Goal Poses	
	Our Method	A*	Our Method	A*
Cube	3.055 ± 1.635	3.800 ± 0.900	3.336 ± 1.211	-
Hexagon	2.719 ± 1.506	3.800 ± 0.500	2.415 ± 1.195	-

4) *Failure Cases*: On occasions where compounded errors lead to target positions being missed, the gripper is observed to perform an additional sliding action followed by a rotation, often managing to eventually reach the target within the required 10 total action steps. Another type of failure is observed for three-cylinders, where the gripper intended to slide the object toward the gripper base but resulted in object rolling. This type of failure is due to the VF hand's fixed base width: the angles between fingers cannot be independently controlled, leading to non-ideally angled forces that cause rolling behaviour even under low frictions. Such failure can be avoided by introducing an actuated prismatic palm with more control over finger angles such as in [34].

C. The Effects of Co-Training

1) Ablation Study – Determining Real-Data Efficiency:

Real data collection usually requires intensive human involvement which is expected to be minimized in the sim-real co-training framework in the pipeline (Sec. V-D). By keeping the simulation data fixed as 10,000 trajectories and varying the amount of real data demonstrations, we investigate the real data efficiency in the co-training framework. The amount of real data is in $[0, 100, 200, 300]$ trajectories, with a real-world collection speed of 100 trajectories per hour on the real IHM system. The object used for this experiment is the Cube Cylinder. The results averaged across three random seeds are shown in Fig. 6. This indicates that the success rate increases along with the usage of more real data, but plateaus after 200 trajectories, with highest success rate 93.3% for co-training and 73.3% for real data only. 100 trajectories are selected as the default amount in our main experiments in Sec. VI-B and VI-C.2, due to a good balance of a low labor-intensive real data collection effort and high success rate of 70% (a more than 2-fold improvement from 33.3%). The simulation data also provides significant performance boosts across different real data amounts. This testifies the

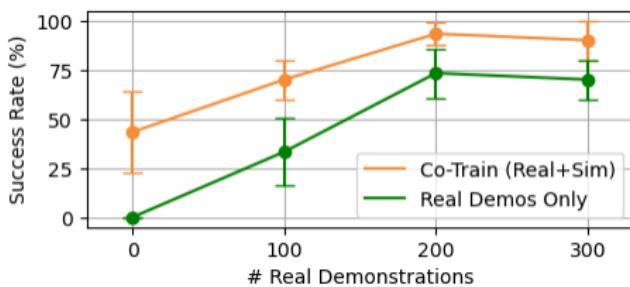


Fig. 6. Real robot success rate for using different amounts of real-world demonstrations for object Cube Cylinder.

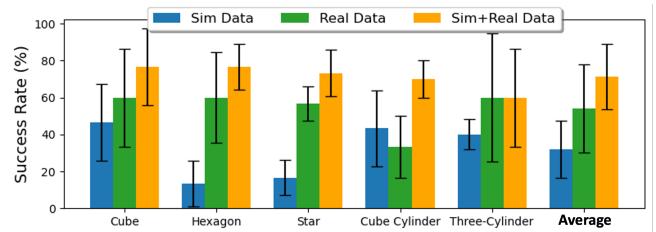


Fig. 7. Co-train compared to sim/real demonstrations only, tested on the real-world VF hand.

effectiveness of the proposed co-training framework with a mix of simulation and real data.

2) *Co-Training vs. Real/Sim Demos Only*: The results in Fig. 7 demonstrate the effectiveness of co-training which leverages both simulation and real-world data. The success rate increases significantly for most objects when co-training is applied, with the Three-Cylinder object being an exception, whose shape was too challenging for the gripper's mechanical design to handle as explained in Sec.VI-B.4. On average across five objects, the success rate improved from 32.0% (simulation only) and 54.0% (real data only) to 71.3% when using both. Additionally, the standard deviation is reduced for most tasks, indicating greater stability and consistency in performance.

3) *Co-Training vs. Fine-tuning*: We also made a comparison with fine-tuning using the cube cylinder by first training the diffusion policy on the simulation demonstrations and then fine-tuning it with the real demonstrations. The results for this experiment use 200 episodes real data. Fine-tuning achieved a success rate of 76.7%, with a lower success rate and higher standard deviation compared to co-training. Our conjecture is that fine-tuning can lead to overfitting on few real data easily thus losing its generalization on new poses. These findings confirm that co-training effectively enhances task success while improving reliability across different object manipulations.

TABLE V
COMPARISON OF CO-TRAIN AND FINE-TUNE ON THE REAL HAND

	Real Data Only	Co-train	Finetune
Success Rate (%)	73.3 ± 17.3	93.3 ± 5.8	76.7 ± 12.5

VII. CONCLUSION

In summary, our method unlocks the VF gripper's IHM capabilities. We trained an automated demonstration generation agent through reinforcement learning, allowing effortless demonstration collection in both simulation and the real world. By deploying diffusion policies co-trained with simulation data, we addressed the scarcity of real-world data and reduced the simulation-to-real-world discrepancy, allowing learning to manipulate arbitrary objects for arbitrary start and goal pose with 2-hours training and 1 hour real-world data collection. Future work includes a general policy for arbitrary objects, through mixing training demonstrations across sample objects.

REFERENCES

- [1] I. M. Bullock and A. M. Dollar, "Classifying human manipulation behavior," in *2011 IEEE international conference on rehabilitation robotics*. IEEE, 2011, pp. 1–6.
- [2] O. 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*, vol. 39, no. 1, pp. 3–20, 2020.
- [3] T. Chen, M. Tippur, S. Wu, V. Kumar, E. Adelson, and P. Agrawal, "Visual dexterity: In-hand reorientation of novel and complex object shapes," *Science Robotics*, vol. 8, no. 84, p. eadc9244, 2023.
- [4] Y. Qin, H. Su, and X. Wang, "From one hand to multiple hands: Imitation learning for dexterous manipulation from single-camera teleoperation," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 10 873–10 881, 2022.
- [5] Z. Si, K. L. Zhang, Z. Temel, and O. Kroemer, "Tilde: Teleoperation for dexterous in-hand manipulation learning with a delta hand," *arXiv preprint arXiv:2405.18804*, 2024.
- [6] A. J. Spiers, B. Calli, and A. M. Dollar, "Variable-friction finger surfaces to enable within-hand manipulation via gripping and sliding," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 4116–4123, 2018.
- [7] G. Narayanan, J. A. Raj, A. Gandhi, A. A. Gupte, A. J. Spiers, and B. Calli, "Within-hand manipulation planning and control for variable friction hands," in *Experimental Robotics: The 17th International Symposium*. Springer, 2021, pp. 600–610.
- [8] A. Sahin, A. J. Spiers, and B. Calli, "Region-based planning for 3d within-hand-manipulation via variable friction robot fingers and extrinsic contacts," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 6549–6555.
- [9] C. Chi, S. Feng, Y. Du, Z. Xu, E. Cousineau, B. Burchfiel, and S. Song, "Diffusion policy: Visuomotor policy learning via action diffusion," *arXiv preprint arXiv:2303.04137*, 2023.
- [10] V. Kumar, E. Todorov, and S. Levine, "Optimal control with learned local models: Application to dexterous manipulation," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 378–383.
- [11] A. Nagabandi, K. Konolige, S. Levine, and V. Kumar, "Deep dynamics models for learning dexterous manipulation," *CoRR*, vol. abs/1909.11652, 2019. [Online]. Available: <http://arxiv.org/abs/1909.11652>
- [12] H. Han, G. Paul, and T. Matsubara, "Model-based reinforcement learning approach for deformable linear object manipulation," 08 2017, pp. 750–755.
- [13] T. Chen, J. Xu, and P. Agrawal, "A system for general in-hand object re-orientation," *CoRR*, vol. abs/2111.03043, 2021. [Online]. Available: <https://arxiv.org/abs/2111.03043>
- [14] OpenAI, I. Akkaya, M. Andrychowicz, M. Chociej, M. Litwin, B. McGrew, A. Petron, A. Paine, M. Plappert, G. Powell, R. Ribas, J. Schneider, N. Tezak, J. Tworek, P. Welinder, L. Weng, Q. Yuan, W. Zaremba, and L. Zhang, "Solving rubik's cube with a robot hand," *CoRR*, vol. abs/1910.07113, 2019. [Online]. Available: <http://arxiv.org/abs/1910.07113>
- [15] Z. Ding, Y. Chen, A. Z. Ren, S. S. Gu, Q. Wang, H. Dong, and C. Jin, "Learning a universal human prior for dexterous manipulation from human preference," *arXiv preprint arXiv:2304.04602*, 2023.
- [16] M. Zare, P. M. Kebria, A. Khosravi, and S. Nahavandi, "A survey of imitation learning: Algorithms, recent developments, and challenges," 2023. [Online]. Available: <https://arxiv.org/abs/2309.02473>
- [17] A. I. Weinberg, A. Shirizly, O. Azulay, and A. Sintov, "Survey of learning approaches for robotic in-hand manipulation," *arXiv preprint arXiv:2401.07915*, 2024.
- [18] A. Rajeswaran, V. Kumar, A. Gupta, J. Schulman, E. Todorov, and S. Levine, "Learning complex dexterous manipulation with deep reinforcement learning and demonstrations," *CoRR*, vol. abs/1709.10087, 2017. [Online]. Available: <http://arxiv.org/abs/1709.10087>
- [19] H. Zhu, A. Gupta, A. Rajeswaran, S. Levine, and V. Kumar, "Dexterous manipulation with deep reinforcement learning: Efficient, general, and low-cost," *CoRR*, vol. abs/1810.06045, 2018. [Online]. Available: <http://arxiv.org/abs/1810.06045>
- [20] S. P. Arunachalam, S. Silwal, B. Evans, and L. Pinto, "Dexterous imitation made easy: A learning-based framework for efficient dexterous manipulation," in *2023 ieee international conference on robotics and automation (icra)*. IEEE, 2023, pp. 5954–5961.
- [21] T. Z. Zhao, V. Kumar, S. Levine, and C. Finn, "Learning fine-grained bimanual manipulation with low-cost hardware," *arXiv preprint arXiv:2304.13705*, 2023.
- [22] R. Ding, Y. Qin, J. Zhu, C. Jia, S. Yang, R. Yang, X. Qi, and X. Wang, "Bunny-visionpro: Real-time bimanual dexterous teleoperation for imitation learning," 2024. [Online]. Available: <https://arxiv.org/abs/2407.03162>
- [23] S. Yuan, L. Shao, C. L. Yako, A. Gruebele, and J. K. Salisbury, "Design and control of roller grasper v2 for in-hand manipulation," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 9151–9158.
- [24] Z. Ding and C. Jin, "Consistency models as a rich and efficient policy class for reinforcement learning," *arXiv preprint arXiv:2309.16984*, 2023.
- [25] Z. Fu, T. Z. Zhao, and C. Finn, "Mobile aloha: Learning bimanual mobile manipulation with low-cost whole-body teleoperation," 2024. [Online]. Available: <https://arxiv.org/abs/2401.02117>
- [26] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.
- [27] F. J. Romero-Ramirez, R. Muñoz-Salinas, and R. Medina-Carnicer, "Speeded up detection of squared fiducial markers," *Image and vision Computing*, vol. 76, pp. 38–47, 2018.
- [28] P. Abbeel and A. Y. Ng, "Exploration and apprenticeship learning in reinforcement learning," in *Proceedings of the 22nd international conference on Machine learning*, 2005, pp. 1–8.
- [29] E. Valassakis, Z. Ding, and E. Johns, "Crossing the gap: A deep dive into zero-shot sim-to-real transfer for dynamics," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 5372–5379.
- [30] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," in *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2017, pp. 23–30.
- [31] M. Andrychowicz, F. Wolski, A. Ray, J. Schneider, R. Fong, P. Welinder, B. McGrew, J. Tobin, O. Pieter Abbeel, and W. Zaremba, "Hindsight experience replay," *Advances in neural information processing systems*, vol. 30, 2017.
- [32] S. Fujimoto, H. Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," in *International conference on machine learning*. PMLR, 2018, pp. 1587–1596.
- [33] O. M. Team, D. Ghosh, H. Walke, K. Pertsch, K. Black, O. Mees, S. Dasari, J. Hejna, T. Kreiman, C. Xu *et al.*, "Octo: An open-source generalist robot policy," *arXiv preprint arXiv:2405.12213*, 2024.
- [34] X. Zhou and A. J. Spiers, "E-troll: Tactile sensing and classification via a simple robotic gripper for extended rolling manipulations," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 11 826–11 833.