

Planning-Guided Diffusion Policy Learning for Contact-Rich Bimanual Object Reorientation

Xuanlin Li^{1,2}, Tong Zhao¹, Bo Ai^{1,2}, Xinghao Zhu¹, Jiuguang Wang¹, Tao Pang¹, Kuan Fang^{1,3}

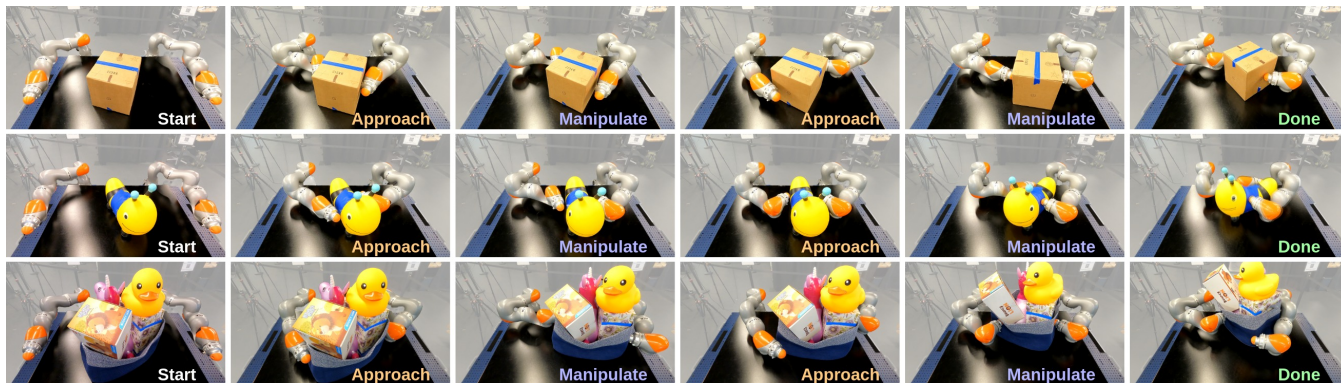


Fig. 1: Contact-rich bimanual manipulation of objects with diverse geometries and physical properties performed by our method. Each row illustrates a trajectory where the robot executes multiple phases of contact (*approach* and *manipulate*) to reorient the object to a target SE(2) pose. We train a single conditional diffusion policy to control the robot in a closed-loop manner given the observed point clouds.

Abstract—Contact-rich bimanual manipulation involves precise coordination of two arms to change object states through strategically selected contacts and motions. Due to the inherent complexity of these tasks, acquiring sufficient demonstration data and training policies that generalize to unseen scenarios remains a largely unresolved challenge. Building on recent advances in planning through contacts, we introduce Planning-Guided Diffusion Policy Learning (LIDE), an approach that effectively learns to solve contact-rich bimanual manipulation tasks by leveraging model-based motion planners to generate demonstration data in high-fidelity physics simulation. Through efficient planning in randomized environments, our approach generates large-scale and high-quality synthetic motion trajectories for tasks involving diverse objects and transformations. We then train a task-conditioned diffusion policy via behavior cloning using these demonstrations. To reduce the sim-to-real gap, we propose a set of designs in feature extraction, action prediction, and data augmentation that enable learning robust prediction of smooth action sequences and generalization to unseen scenarios. Through experiments in both simulation and the real world, we demonstrate that our approach can enable a bimanual robotic system to effectively manipulate objects of diverse geometries, dimensions, and physical properties.

I. INTRODUCTION

From warehouse logistics to home services, a broad range of essential robotic applications relies on manipulation involving multi-contact interactions between objects and the manipulators. For example, as shown in Fig. 1, the task is to control two robotic arms to manipulate different objects to a specified target pose. These objects are often bulky and heavy, making them not directly graspable by the end-effectors. To rearrange and reorient the object, the two arms must hold the object robustly through contacts at multiple links, and then reorient the object over possibly long horizons that take multiple contact phases to reach the goal. Due to such inherent complexity, solving contact-rich bimanual manipulation for diverse and complex objects remains an open challenge.

To tackle this challenge, recent advances in model-based planning methods utilizing smoothed contact models have begun to demonstrate effectiveness [1–4]. However, such motion planners need complete knowledge of object states and environment geometry, thereby limiting their ability to operate in novel environments where objects exhibit diverse geometries and physical properties. Additionally, their computational overhead prevents them from generating trajectories online, which can be a critical limitation, especially in dynamic environments that require real-time adaptation. These limitations highlight the need for approaches that can effectively and robustly perform contact-rich bimanual manipulation over diverse objects.

Alternatively, an increasing number of works have aimed to acquire robust contact-rich manipulation skills by learning from trajectory data [5–9]. In spite of the promising progress, key challenges remain, particularly for bimanual manipulation with complex embodiments. First, training robust, generalizable policies for complex visuomotor skills usually requires large scale, high-quality trajectory data such as expert demonstrations. However, it is particularly difficult and costly to collect demonstrations for complex systems and tasks like contact-rich bimanual manipulations using traditional approaches such as teleoperation in the real world [10–12]. Second, for approaches that learn policies using simulation data, the reality gap in perception and dynamics poses notable challenges for effective policy deployment and generalization. Such challenges also become more significant as the complexity of the task increases.

In this work, we propose a method that learns contact-rich bimanual manipulation, specifically reorientation of bulky and heavy objects, by addressing the aforementioned challenges. Our approach, named PLanning-GuIded Diffusion Policy LEarning (LIDE), is built upon recent advances in diffusion policies [9] and model-based motion planning [3] to boost scalability for both data and model. Using a contact-

¹RAI Institute. ²University of California, San Diego. ³Cornell University.

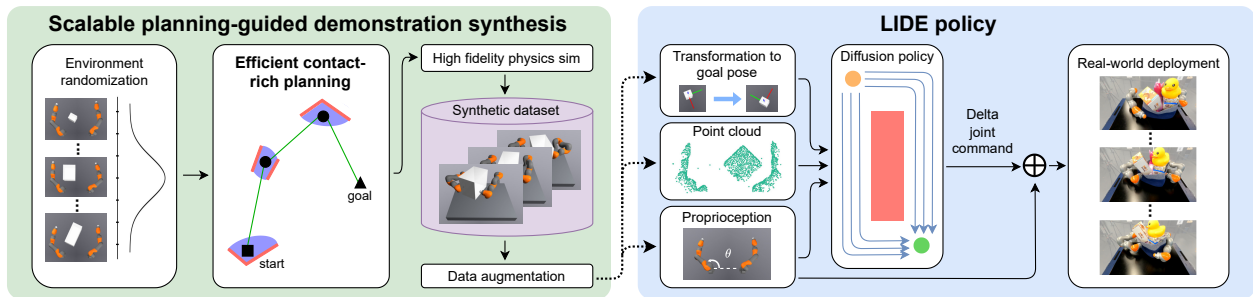


Fig. 2: Overview of PLanning-GuIded Diffusion Policy LEarning (LIDE) for contact-rich bimanual object manipulation. **Left:** Scalable synthetic demonstration generation using efficient motion planning in physical simulator (Sec. III-B). **Right:** Our LIDE policy is a task-conditioned point cloud diffusion policy trained through behavior cloning. We further introduce essential design options that significantly enhance our policy’s ability to transfer to the real world and generalize to unseen scenarios (Sec. III-C).

implicit trajectory optimization solver [4], which derives its efficiency from a smoothed linear approximation of local robot-object contact dynamics, we generate massive and high-quality demonstration data for contact-rich bimanual manipulation in physics simulation. Compared to prior long-horizon motion planning methods that exhibit high computational costs, this planner significantly improves data generation efficiency with minimal impact on trajectory quality. Additionally, we implement our planner to greedily approach the goal object states, further speeding up synthetic data generation. To learn generalizable visuomotor skills for contact-rich and bimanual manipulation, we build on recent advancements in diffusion policy [8, 13] that effectively capture multimodal action distributions. In contrast to [8, 13], which learns a specific set of parameters for each task, we design a task-conditioned diffusion policy network that controls the robot to manipulate the object to *arbitrary* target poses specified by the users. To reduce the reality gap between the simulation and the real-world system, we introduce a set of design choices in feature extraction, action prediction, and data augmentation, which significantly improve our policy’s performance in sim-to-real transfer. We evaluate LIDE in a series of simulated and real-world contact-rich bimanual manipulation tasks involving both in-distribution and out-of-distribution (OOD) objects. Through detailed analysis, we demonstrate that our approach can robustly accomplish contact-rich bimanual manipulation in unseen scenarios.

II. RELATED WORK

Contact-rich and bimanual manipulation have been widely studied, with key challenges arising from non-smooth contact dynamics and navigating non-convex cost landscapes [3]. Previous work in planning has formulated contact-rich manipulation as a Mixed-Integer Program (MIP), using integer variables to represent contact modes [14–16]. While MIP-based methods offer optimality, they scale poorly for tasks like full-arm bimanual manipulation due to the exponential growth in modes with increasing contact pairs and task horizons. Recent planning approaches have improved efficiency by sacrificing optimality for speed through analytic smoothing [3, 4, 17], but planners rely on full knowledge of object states and geometry, limiting their generalization to diverse real-world objects. To address generalization, another line of prior work has approached contact-rich manipulation through behavior cloning and reinforcement learning [5–8, 18, 19]. While

they have shown promising results, few have focused on generalizable bimanual contact-rich manipulation due to its inherent complexity. A recent work [20] has made progress in this area, but it only uses two end-effectors with single contact points to perform manipulation rather than the full bimanual arms, and it lacks visual feedback. This limits its ability to manipulate bulky or lengthy objects. In this work, we propose a method that demonstrates generalizable contact-rich bimanual manipulation over objects with diverse dimensions, geometries, and physical properties, utilizing visual feedback and without assuming privileged information (e.g., object shapes) to be known.

Visuomotor policy learning for manipulation. Modern machine learning approaches have enabled robots to learn visuomotor policies for a variety of tasks like grasping, pushing, rotating, and insertion [21–29]. Compared to these primitive skills, contact-rich bimanual manipulation introduces critical challenges to policy learning due to the complex dynamics and multimodal distribution of behaviors. With recent advances in deep generative models [30–32], diffusion policies [9, 33–36] have been introduced as a new class of policy networks that are capable of capturing high-dimensional multi-modal action distributions. While these methods have shown promise, they are often applied to specific tasks with fixed goals. In this work, we design a diffusion policy network for contact-rich bimanual manipulation given point clouds.

Planning-guided data synthesis. While traditional motion planning systems require complete knowledge of scenes and objects, limiting their real-world deployment, they are highly valuable for synthetic data generation for learning-based methods. Prior imitation learning work has used planning data to train effective, generalizable policies for collision avoidance [37, 38], manipulation [39–41], locomotion [42, 43], and autonomous driving [44]. Model-based planning has also been widely used to improve sample efficiency and performance in reinforcement learning [45–50]. However, using planning to generate high-quality demonstrations for contact-rich bimanual manipulation scalably remains under-explored due to the challenges posed by complex contact dynamics and planning efficiency. In this work, we build on recent advances in trajectory optimization and motion planning [3, 4] to effectively and scalably generate demonstrations for contact-rich bimanual manipulation.

III. METHOD

LIDE centers around two key focuses: generating diverse and high-quality training data for contact-rich bimanual manipulation tasks and learning visuomotor policies that can generalize to unseen environments and task specifications.

In this section, we will first describe the problem formulation for learning contact-rich bimanual manipulation. Next, we will propose a data synthesis pipeline that uses an efficient contact-rich planner. Finally, we will devise a conditional diffusion policy that generates action sequences given observed point clouds and task specifications.

A. Problem Formulation

We consider the problem of controlling a bimanual robotic system to change the pose of objects. Such objects can be bulky and heavy, and thus cannot be directly grasped by the end-effectors. To accomplish this task, the two robot arms need to strategically approach the object to make contact and then reorient the object to the target pose. For complex objects and challenging target poses, this process might take multiple rounds of approaching and manipulation due to the limitation of the configuration space of the robot. To enable generalization to unseen objects, we do not assume the shape and initial pose of the object to be known. Instead, the robot only receives visual observations of the environment.

Formally, we define the problem with the environment state space \mathcal{S} , the observation space \mathcal{O} , the action space \mathcal{A} , the task space \mathcal{C} , and the time horizon H . In each episode, the robot starts at an initial environment state $s_0 \in \mathcal{S}$ and is commanded to change the environment according to a task specification $c \in \mathcal{C}$. At each time step $t = 1, \dots, H$, the robot receives the observation $o_t \in \mathcal{O}$ and takes an action $a_t \in \mathcal{A}$. To solve this problem, we aim to learn a single policy $\pi_\theta(a|o, c)$ with a trainable set of parameters θ . Due to the complexity of the dynamics and control in bimanual manipulation tasks, directly collecting massive and high-quality data in the real world is hard. Instead, we aim to generate *synthetic* demonstration data \mathcal{D} to train the policy π . As no real-world training data is collected and used, we would need to design the policy network and learning algorithm such that the policy trained on \mathcal{D} is transferrable to the real world at deployment.

In this work, we specifically consider a table-top environment involving two 7-DoF robotic arms without end-effectors, as shown in Fig. 1. The policy does not have access to the true environment state s_t but receives uncolored point clouds from a static camera and the proprioceptive joint states of the robot as the observation o_t . Such point cloud observations with invariance to lighting conditions, textures, and colors have been shown to facilitate sim-to-real transfer [51–53]. The target object to manipulate is supported by the table surface, and the task is defined by a transformation of the object pose in $\mathbb{SE}(2)$. At each time step, we compute the task specification c_t as the transformation between the current object pose and the target pose as part of the inputs to the policy π .

B. Demonstration Synthesis via Efficient Planning

To scalably generate demonstrations for our bimanual object reorientation task, we use motion planning that lever-

ages privileged object and robot states in simulation to produce trajectories. We build upon the planning-through-contact framework in [3] and incorporate the very recent advances in [4] in our planning pipeline, which proposes a smoothed linear approximation of local robot-object contact dynamics that significantly improves planning efficiency.

Our planner’s input include (i) the initial state of the system, $s_0 := (q_0^a, q_0^u)$, where q_0^a is the robot joint angles and q_0^u the object pose; and (ii) the goal object pose q_{goal}^u . An action a consists of commanded joint angles for both arms. The planner generates a sequence of actions $T := (a_0, a_1, \dots)$ that takes the object from the initial pose to the goal pose. Compared with the sampling-based planner in [3], we greedily approach the goal to encourage consistency in the demonstrations, which prior work has shown to be more effective for behavioral cloning [54]. The greedy approach did not significantly impact the planner’s success rate in practice. The resulting planner, summarized in Algorithm 1, consists of the following components:

- A *contact sampler* that generates robot joint configurations q_{grasp}^a where the robot arms make contact with the object in a way that facilitates manipulation (i.e., a “grasp”). In our implementation, we use inverse kinematics to generate grasps where the robot’s distal links can pinch the object.
- A *collision-free planner* using bidirectional RRT [55] with shortcutting to plan a collision-free trajectory from the current robot joint configuration q^a to the next grasp q_{grasp}^a .
- A *contact planner* which, given a current robot configuration q^a that is already grasping the object and the object’s current configuration q^u , greedily moves the object towards the goal configuration q_{goal}^u as much as possible while ensuring that the robot does not exceed joint limits.

Algorithm 1: Demonstration Synthesis via Planning

```

1 Input:  $q_0^a, q_0^u, q_{\text{goal}}^u$ 
2 Output: Action trajectory  $T$ 
3  $q^a, q^u, T \leftarrow q_0^a, q_0^u, \text{list}()$ 
4 while  $q^u \neq q_{\text{goal}}^u$  do
5    $q_{\text{grasp}}^a \leftarrow \text{SAMPLECONTACT}(q^u)$ 
6   while  $q^a \neq q_{\text{grasp}}^a$  do
7      $q^a, a \leftarrow \text{PLANCOLLISIONFREE}(q^a, q_{\text{grasp}}^a)$ 
8      $T.\text{extend}(a)$ 
9   while  $q^a$  not at joint limit and  $q^u \neq q_{\text{goal}}^u$  do
10     $q^a, q^u, a \leftarrow \text{PLANCONTACT}(q^a, q^u, q_{\text{goal}}^u)$ 
11     $T.\text{extend}(a)$ 
12 return  $T$ 

```

Contact Planner Details. Much prior work has focused on trajectory optimization through contact [1, 2, 16, 56, 57], but their high computational costs due to long-horizon trajectory optimization and the exponential number of contact modes prohibit efficient trajectory generation. To reduce these costs, we adopt a *single step* variant of trajectory optimization and solve it through the approach in [4]. In short, we use a linear approximation f_{local} of the local contact dynamics to solve the optimization problem with the following objective:

$$\min_{q_+^u, a} (q_+^u - q_{\text{goal}}^u)^T \mathbf{Q} (q_+^u - q_{\text{goal}}^u) + (a - q^a)^T \mathbf{R} (a - q^a)$$

Here, $q_+^u = f_{\text{local}}(q^u, q^a, a)$ denotes the object’s approximate configuration after the robot action a is taken, and \mathbf{Q} and \mathbf{R} are user-specified cost matrices.

Filtered Behavior Cloning. As the collected trajectories are

often suboptimal due to the approximity of the contact dynamics and the stochasticity of the RRT in the planner, directly performing behavior cloning on all trajectories leads to poor policy performance. To address this, we extract high-quality demonstrations by filtering trajectories via rollouts in a high-fidelity simulator [58] to verify their accuracy. The trajectories in which the object doesn’t reach the goal and the suboptimal trajectories that take too long to reach the goal are discarded in this process. Finally, we rebalance the trajectories to be uniformly distributed across objects and render uncolored point clouds for policy training. We save the resultant subset of demonstration trajectories as the dataset \mathcal{D} to train the policy.

C. Diffusion Policy Learning from Synthetic Demonstrations

Given the synthetic demonstrations \mathcal{D} , we train a diffusion policy $\pi_\theta(a|s, c)$ for contact-rich bimanual manipulation via behavior cloning. To enable π_θ to predict suitable manipulation actions for unseen scenarios and effectively transfer the learned knowledge to the real world, we introduce a set of designs to existing diffusion policy learning methods [9] in feature extraction, task representation, and action prediction.

To extract the geometric information of the environment from the noisy point cloud observations, we design a feature extraction backbone for π_θ that can facilitate generalization to unseen objects based on [9] and [59]. To enable generalization to unseen environments, we clip the point clouds within the robot’s workspace and remove the irrelevant background objects in each frame. Additionally, to address the reality gap and sensor noise encountered in the real world, we introduce a *Flying Point Augmentation* approach, where we randomly add large Gaussian noise to the points with a small probability (e.g., 0.5%). We find that this approach significantly improves our policy’s real-world performance while requiring minimal implementation effort.

While [9] and [59] aim to learn a predefined set of tasks using a different set of parameters for each task, we train a *single* policy that takes the task specification c as an additional input. As described in Sec. III-A, instead of assuming the object shape is known and directly specifying a target object pose, we implicitly specify the target pose using the initial visual observation o_0 along with the delta transformation c_0 from the initial object pose in o_0 to the target object pose. In each subsequent time step t , we recompute the transformation c_t from the current pose to the target pose given the current observation o_t . Without knowing the object shape and pre-defining an object frame, we propose to obtain c_t by segmenting the target object in o_0 using an open-vocabulary segmentation algorithm [60, 61], selecting keypoints using the farthest point sampling within the segmentation, and then tracking the keypoints in the 3D space through real-time object tracking [62, 63].

We design the policy’s prediction head to robustly generate smooth and feasible motion trajectories. Following [9], our policy predicts an action sequence of T_a steps. We use a larger $T_a = 20$ at test time to improve performance and train our policy with $T_a = 64$. Additionally, prior work like [9] usually directly predicts the absolute end-effector poses or joint angles as actions. However, we observe that



Fig. 3: Objects used for real-world evaluation: **(Left)** Boxes whose dimensions and physical properties are within our policy’s training distribution; **(Right)** Out-of-distribution (OOD) containers, some of which are made with soft materials like rubber (top right, bottom left) and fabric (bottom right). For half of OOD evaluations, we evaluate our policy using empty containers; for the other half, we evaluate using containers overfilled with miscellaneous objects.

this can result in poor generalization to unseen objects and non-smooth trajectories in the real world. To address this, we instead re-design the prediction head to predict the *delta* joint position actions $a_{t+1:t+T_a} = \{q_i - q_t\}_{i=t+1}^{t+T_a}$ with q_t being the current joint positions. Compared to absolute joint actions, the delta joint actions are much more consistent in scales and shifts across training trajectories, resulting in significantly better real-world policy performance.

IV. EXPERIMENTS

We design our experiments to answer the following questions: **(1)** Does LIDE effectively perform contact-rich bimanual object manipulation on objects whose geometries, dimensions, and physical properties are within the distribution of training demonstrations? **(2)** How well does LIDE generalize to objects with challenging geometries and physical properties? **(3)** How important are the design choices in LIDE to the performance?

A. Environments and Tasks

We perform bimanual object reorientation on a flat table using two 7-DoF KUKA LBR iiwa arms in both the DRAKE [58] simulator and the real world. We use a Realsense D455 camera mounted 2 meters above the table to capture real-world visual observations, and we actuate three joints per arm following [64]. In simulation, we synthesize planner demonstrations following our approach in Sec. III-B and then retain 12,000 successful trajectories for policy training (Sec. III-C), filtered from roughly 17,000 successful ones to ensure demonstration quality. During both demonstration generation and policy evaluation, we randomize object assets along with their initial and goal poses. An episode is successful if the final object pose is within 10 cm and 0.2 rad of the goal. The demonstration generation process takes about two days on a 96-CPU machine.

For demonstration generation and policy training, we construct 2,000 box primitives with randomized physical parameters: widths in the range 0.15–0.6 m, depths 0.16–0.6 m, heights 0.13–0.6 m, masses 0.3–2.5 kg, and friction coefficients 0.5–1.0. We additionally include cylinders with radii 0.1–0.4 m and heights 0.2–0.8 m. Although the policy is trained exclusively on these primitive shapes, we demonstrate generalization to objects with out-of-distribution geometries and physical properties during testing, such as irregular containers and plush toys.

B. Experiment Design

To comprehensively assess LIDE’s performance in contact-rich bimanual object reorientation, we evaluate it on

Task (In-Dist Eval)	Planner (Sim)	Policy (Sim)	Policy (Real)
Fixed 45° Rotation	0.337	0.740	0.800
Random Rotation (Easy)	0.227	0.610	0.600
Random Rotation (Medium)	0.141	0.410	0.360
Random Rotation (Hard)	0.099	0.180	0.200
Random Rotation (Overall)	0.156	0.400	0.387

TABLE I: Bimanual object reorientation success rates in both simulation and the real world for in-distribution objects ($N = 1000$ and 25 for the sim and real experiments, respectively). We report results for our point cloud policy and also include the planner results as a reference. The planner requires access to full state information of the box, infeasible for real-world deployment. For learning-based baseline (vanilla point cloud diffusion policy), see the last row of Tab. III and the first column of Tab. IV.

Task (Real OOD Eval)	Empty Containers	Overfilled Containers	Overall
Fixed 45° Rotation	0.688	0.625	0.657
Random Rotation	0.250	0.313	0.282

TABLE II: Success rates of real-world policy evaluation using out-of-distribution containers (which are visualized in Fig. 3 - Right). We evaluate our policy using both empty and overfilled containers.

objects with geometries, dimensions, and physical properties within the distribution of demonstration trajectories (i.e., “*In-Distribution Evaluation*”). Additionally, we put LIDE to the test and evaluate its ability to generalize to challenging, unseen objects outside of training distribution (“*Out-of-Distribution (OOD) Evaluation*”). Despite the perception and control challenges posed by the diverse shapes and physical properties of OOD real world objects, we hope that our policy can still effectively infer object boundaries from our processed point clouds where irrelevant backgrounds are removed, from which it can determine the necessary robot joint movements to approach and manipulate the object.

Furthermore, to enable a fine-grained analysis of LIDE’s success and limitations, we design our experiments to encompass increasing levels of task difficulty:

- **Fixed Rotation:** Let $\Delta\theta$ represent the orientation difference between the initial and goal object poses. In our “Fixed Rotation” setting, we set $\Delta\theta$ to a constant value across all demonstration and evaluation episodes (in our experiments, $\Delta\theta = 45^\circ$). Additionally, we initialize object orientations in multiples of $\frac{\pi}{2}$ (for boxes, their edges are aligned parallel to the table edges). Although simpler than the “Random Rotation” setting below, this setup is already challenging as it requires the policy to generalize to diverse objects with randomized positions.
- **Random Rotation:** Both $\Delta\theta$ and the initial object orientation are fully randomized. We further subdivide $\Delta\theta$ into 3 levels of difficulties (visualized in Fig. 4): *Easy* ($|\Delta\theta| \leq 45^\circ$); *Medium* ($45 < |\Delta\theta| \leq 90^\circ$); *Hard* ($90^\circ < |\Delta\theta| \leq 150^\circ$)¹. As $|\Delta\theta|$ increases, it becomes more likely that the robot needs to perform multiple rounds of approaching and manipulation to reach the object to the goal orientation due to the limitation of robot’s configuration space.

We generate demonstration sets for both the “Fixed Rotation” and “Random Rotation” tasks and train a diffusion policy on each. For the “Random Rotation” task, we evaluate a single policy across all three levels of task difficulty.

¹We set $|\Delta\theta| \leq 150^\circ$ as it is very challenging to generate successful demonstrations for $|\Delta\theta| > 150^\circ$ with our current planner.

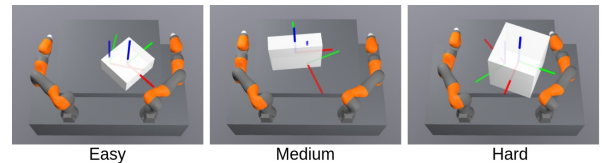


Fig. 4: Visualization of our bimanual box reorientation environment in the Drake simulator. The coordinate frame on the box represents its current pose, while the coordinate frame on the table indicates the goal pose. We illustrate 3 environment variations with randomized boxes and increasing difficulty levels: easy ($|\Delta\theta| \leq 45^\circ$), medium ($45 < |\Delta\theta| \leq 90^\circ$), and hard ($90^\circ < |\Delta\theta| \leq 150^\circ$).

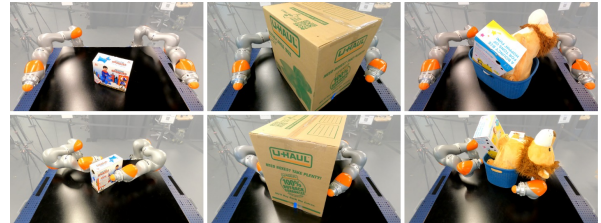


Fig. 5: Examples of policy evaluation on in-distribution (left, middle) and out-of-distribution (right) objects. The first row shows the initial state, and the second row shows the final state.

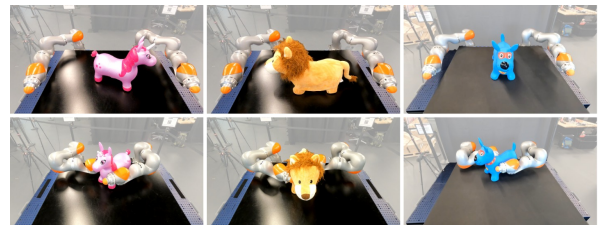


Fig. 6: Visualization of policy evaluation under a challenging OOD scenario using irregularly shaped toys. The success rate is 52% on the fixed 45° rotation task and 28% on the random rotation task.

C. In-Distribution Evaluation

We perform in-distribution evaluation of LIDE in both simulation and the real world. In simulation, we randomly sample boxes whose dimensions and physical properties are within our demonstration’s distribution. In the real world, we randomly choose a set of everyday boxes from the same distribution, as shown in Fig. 3. We conduct 100 policy evaluation trials in simulation and 25 trials in the real world. Additionally, we include our motion planner’s success rates in simulation as a reference. Note that the planner runs slower than our policy and requires access to complete knowledge of object geometry, which is infeasible for real-world evaluation. Due to the highly stochastic nature of our planner, we evaluate it over a larger number of 600 trials.

We present our results in Tab. I, and we include qualitative visualizations of policy rollouts in the first row of Fig. 1 and the left and middle columns of Fig. 5. Overall, LIDE demonstrates strong performance on in-distribution objects in both simulation and the real world. In simulation, the learned policy outperforms the planner by learning only from filtered, high-quality demonstrations, whereas the planner may produce suboptimal actions during online search. In real-world experiments, the policy exhibits minimal degradation from the sim-to-real gap and occasionally reports higher success rates, potentially statistical variations from the relatively limited number of trials. However, there is still significant room for improvement in the highly challenging setting

where the policy must manipulate the object towards a distant target orientation over a long horizon. Later in Sec. IV-F, we will analyze these failure cases in detail and suggest directions for future improvements.

D. Out-of-Distribution (OOD) Evaluation

We evaluate LIDE’s ability to generalize to out-of-distribution objects in the real world using a set of containers shown in Fig. 3-Right. These containers differ from the rigid boxes in both geometry and physical properties significantly: some have curved edges or a wider top than the bottom, while others are deformable, made from soft materials like rubber or fabric. We conduct 16 evaluation trials with the containers empty and another 16 trials with them *overfilled* with miscellaneous objects. As illustrated in the third row of Fig. 1 and the right column of Fig. 5, overfilled containers introduce irregular object boundaries and object weights heavier than those during training, adding further challenges for our policy. Results are shown in Tab. II. LIDE demonstrates robustness and generalization to OOD shapes and object properties, with a slight performance drop compared to in-distribution evaluation in Tab. I. Notably, the success rates are similar across both empty and overfilled containers, despite their distinct geometries and weight, showcasing LIDE’s robustness in handling diverse and challenging scenarios.

Additionally, we evaluate LIDE on inflatable toys as shown in Fig. 6 and the second row of Fig. 1. These toys present even larger distribution shifts from the training objects, featuring legged bases instead of the rectangular bottoms seen in boxes and containers. Additionally, their higher center of mass makes them prone to tipping if the two robot arms do not act with precise coordination. LIDE achieves 52% for the fixed rotation task and 28% for the random rotation task, further demonstrating its potential to adapt and generalize to challenging objects.

E. Ablations

In this section, we analyze the importance of different design choices to LIDE’s performance. In Tab. III, we ablate the delta robot joint action prediction and the flying point augmentation approaches we introduced in Sec. III-C to improve LIDE’s sim-to-real transfer. While these techniques show minimal impact in simulation, they significantly enhance policy performance in real-world evaluations. Omitting either approach results in a notable drop in real-world performance. In Tab. IV, we ablate the number of action prediction steps T_a for our diffusion policy evaluation. We find that a larger $T_a = 20$, compared to prior work [9] with $T_a = 8$, results in smoother trajectories and better performance. However, further increasing T_a can degrade performance due to the lack of real-time feedback, which limits the policy’s ability to make timely adaptations during object manipulation. In Tab. V, we study how the number of planner-generated demonstrations impacts policy performance. We find that thousands of demonstrations are needed for good policy performance on the easier “Fixed Rotation” task, with significantly more required for the challenging “Random Rotation” task. We also find that the performance doesn’t plateau as we scale up the number of demonstrations,

Delta Action	Flying Point Aug	Success (Sim, In-Dist)	Success (Real, In-Dist)
✓	✓	0.740	0.800
✓	✗	0.750	0.320
✗	✓	0.780	0.520
✗	✗	0.750	0.000

TABLE III: Ablation of different point cloud diffusion policy design choices on our fixed 45° clockwise object rotation task (100 trials in sim and 25 trials in real; minor evaluation noise may occur).

Task (Sim, In-Dist Eval)	$T_a = 8$	$T_a = 20$	$T_a = 40$	$T_a = 64$
Fixed 45° Rotation	0.440	0.740	0.760	0.770
Random Rotation	0.270	0.400	0.340	0.200

TABLE IV: Ablation on the number of action prediction steps T_a for diffusion policy evaluation. We use $T_a = 64$ for training.

Task (Sim, In-Dist Eval)	500 Demos	2500 Demos	7500 Demos	12000 Demos
Fixed 45° Rotation	0.330	0.690	0.740	0.800
Random Rotation	0.030	0.170	0.330	0.400

TABLE V: Ablation on the number of filtered successful expert planner demonstrations for diffusion policy learning.

suggesting that further scaling up demonstration generation will continue to improve policy effectiveness.

F. Failure Mode Analysis

While our method demonstrates effective generalizable bimanual contact-rich manipulation, there is still room for improvement on highly challenging scenarios requiring manipulating objects to distant target orientations. Among 52% of failure cases in our “Random Rotation-Hard” experiments, the robot is stuck in a poor joint configuration that prevents further object rotation, as illustrated in the wrapped figure. Additionally, object slippage occurs in 20% of cases due to unstable contact, and in 16% of cases, the robot exceeds torque limits by squeezing the object too hard. To address these failure scenarios, generating demonstrations from more diverse initial states, including near-failure situations, and incorporating tactile perception [20, 29] would be promising avenues for future work.



V. CONCLUSION

We presented LIDE, a planning-guided diffusion policy learning method for contact-rich bimanual manipulation. Leveraging the recent advances in efficient model-based planning through contact, we generate large-scale, high-quality demonstration trajectories in simulation. We then perform visuomotor imitation learning using a task-conditioned point cloud diffusion policy, and we propose designs in feature extraction, action prediction, and data augmentation that enable generalization to unseen scenarios and sim-to-real transfer. Evaluations in both simulation and the real world demonstrate our policy’s effectiveness in manipulating objects with diverse geometries, dimensions, and physical properties. In this work, we synthesized data using only primitive shapes. Incorporating objects from large-scale datasets could further enhance generalization. As our planner and policy architecture are largely task-agnostic, they can be applied to more complex domains, including dexterous and mobile manipulation, which are promising directions for future research.

REFERENCES

- [1] Y. Tassa, T. Erez, and E. Todorov, "Synthesis and stabilization of complex behaviors through online trajectory optimization," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012, pp. 4906–4913.
- [2] M. Posa, C. Cantu, and R. Tedrake, "A direct method for trajectory optimization of rigid bodies through contact," *The International Journal of Robotics Research*, vol. 33, no. 1, pp. 69–81, 2014.
- [3] T. Pang, H. T. Suh, L. Yang, and R. Tedrake, "Global planning for contact-rich manipulation via local smoothing of quasi-dynamic contact models," *IEEE Transactions on Robotics*, 2023.
- [4] H. Suh, T. Pang, T. Zhao, and R. Tedrake, "Dexterous contact-rich manipulation via the contact trust region," *arXiv preprint arXiv:2505.02291*, 2025.
- [5] 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.
- [6] 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.
- [7] A. Handa, A. Allshire, V. Makoviychuk, A. Petrenko, R. Singh, J. Liu, D. Makoviichuk, K. Van Wyk, A. Zhurkevich, B. Sundaralingam *et al.*, "Dextreme: Transfer of agile in-hand manipulation from simulation to reality," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 5977–5984.
- [8] Y. Qin, B. Huang, Z.-H. Yin, H. Su, and X. Wang, "Dexpoint: Generalizable point cloud reinforcement learning for sim-to-real dexterous manipulation," in *Conference on Robot Learning*. PMLR, 2023, pp. 594–605.
- [9] C. Chi, S. Feng, Y. Du, Z. Xu, E. Cousineau, B. Burchfiel, and S. Song, "Diffusion policy: Visuomotor policy learning via action diffusion," *The International Journal of Robotics Research*, 2023.
- [10] T. Z. Zhao, V. Kumar, S. Levine, and C. Finn, "Learning fine-grained bimanual manipulation with low-cost hardware," in *Robotics: Science and Systems*, 2023.
- [11] Z. Fu, T. Z. Zhao, and C. Finn, "Mobile aloha: Learning bimanual mobile manipulation with low-cost whole-body teleoperation," *arXiv preprint arXiv:2401.02117*, 2024.
- [12] C. Chi, Z. Xu, C. Pan, E. Cousineau, B. Burchfiel, S. Feng, R. Tedrake, and S. Song, "Universal manipulation interface: In-the-wild robot teaching without in-the-wild robots," *arXiv preprint arXiv:2402.10329*, 2024.
- [13] H. Qi, B. Yi, S. Suresh, M. Lambeta, Y. Ma, R. Calandra, and J. Malik, "General in-hand object rotation with vision and touch," in *Conference on Robot Learning*. PMLR, 2023, pp. 2549–2564.
- [14] T. Marcucci, R. Deits, M. Gabbicini, A. Bicchi, and R. Tedrake, "Approximate hybrid model predictive control for multi-contact push recovery in complex environments," in *2017 IEEE-RAS 17th international conference on humanoid robotics (Humanoids)*. IEEE, 2017, pp. 31–38.
- [15] F. R. Hogan and A. Rodriguez, "Feedback control of the pusher-slider system: A story of hybrid and underactuated contact dynamics," in *Algorithmic Foundations of Robotics XII: Proceedings of the Twelfth Workshop on the Algorithmic Foundations of Robotics*. Springer, 2020, pp. 800–815.
- [16] A. Aydinoglu, A. Wei, W.-C. Huang, and M. Posa, "Consensus complementarity control for multicontact mpc," *IEEE Transactions on Robotics*, vol. 40, pp. 3879–3896, 2024.
- [17] T. A. Howell, S. L. Cleac'h, J. Brüdigam, J. Z. Kolter, M. Schwager, and Z. Manchester, "Dojo: A differentiable physics engine for robotics," *arXiv preprint arXiv:2203.00806*, 2022.
- [18] D. Bertsekas, *Reinforcement learning and optimal control*. Athena Scientific, 2019, vol. 1.
- [19] J. Grannen, Y. Wu, B. Vu, and D. Sadigh, "Stabilize to act: Learning to coordinate for bimanual manipulation," in *7th Annual Conference on Robot Learning*, 2023.
- [20] Y. Lin, A. Church, M. Yang, H. Li, J. Lloyd, D. Zhang, and N. F. Lepora, "Bi-touch: Bimanual tactile manipulation with sim-to-real deep reinforcement learning," *IEEE Robotics and Automation Letters*, vol. 8, no. 9, pp. 5472–5479, 2023.
- [21] A. Mandlekar, Y. Zhu, A. Garg, J. Booher, M. Spero, A. Tung, J. Gao, J. Emmons, A. Gupta, E. Orbay *et al.*, "Roboturk: A crowdsourcing platform for robotic skill learning through imitation," in *Conference on Robot Learning*. PMLR, 2018, pp. 879–893.
- [22] A. Mandlekar, J. Booher, M. Spero, A. Tung, A. Gupta, Y. Zhu, A. Garg, S. Savarese, and L. Fei-Fei, "Scaling robot supervision to hundreds of hours with RoboTurk: Robotic manipulation dataset through human reasoning and dexterity," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 1048–1055.
- [23] C. Eppner, A. Mousavian, and D. Fox, "Acronym: A large-scale grasp dataset based on simulation," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 6222–6227.
- [24] F. Ebert, Y. Yang, K. Schmeckpeper, B. Bucher, G. Georgakis, K. Daniilidis, C. Finn, and S. Levine, "Bridge data: Boosting generalization of robotic skills with cross-domain datasets," in *Robotics: Science and Systems (RSS) XVIII*, 2022.
- [25] X. Zhu, Y. Zhou, Y. Fan, L. Sun, J. Chen, and M. Tomizuka, "Learn to grasp with less supervision: A data-efficient maximum likelihood grasp sampling loss," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 721–727.
- [26] M. Liu, X. Li, Z. Ling, Y. Li, and H. Su, "Frame mining: a free lunch for learning robotic manipulation from 3d point clouds," in *6th Annual Conference on Robot Learning*, 2022. [Online]. Available: <https://openreview.net/forum?id=d-JYso87y6s>
- [27] X. Zhang, S. Jin, C. Wang, X. Zhu, and M. Tomizuka, "Learning insertion primitives with discrete-continuous hybrid action space for robotic assembly tasks," in *2022 International conference on robotics and automation (ICRA)*. IEEE, 2022, pp. 9881–9887.
- [28] D. Seita, Y. Wang, S. Shetty, E. Li, Z. Erickson, and D. Held, "ToolFlowNet: Robotic Manipulation with Tools via Predicting Tool Flow from Point Clouds," in *Conference on Robot Learning (CoRL)*, 2022.
- [29] B. Ai, S. Tian, H. Shi, Y. Wang, C. Tan, Y. Li, and J. Wu, "Robopack: Learning tactile-informed dynamics models for dense packing," *Robotics: Science and Systems (RSS)*, 2024. [Online]. Available: <https://arxiv.org/abs/2407.01418>
- [30] A. Ajay, Y. Du, A. Gupta, J. Tenenbaum, T. Jaakkola, and P. Agrawal, "Is conditional generative modeling all you need for decision-making?" *arXiv preprint arXiv:2211.15657*, 2022.
- [31] L. Chen, K. Lu, A. Rajeswaran, K. Lee, A. Grover, M. Laskin, P. Abbeel, A. Srinivas, and I. Mordatch, "Decision transformer: Reinforcement learning via sequence modeling," *Advances in neural information processing systems*, vol. 34, pp. 15 084–15 097, 2021.
- [32] T. Tian, H. Li, B. Ai, X. Yuan, Z. Huang, and H. Su, "Diffusion dynamics models with generative state estimation for cloth manipulation," in *CoRL*, 2025. [Online]. Available: <https://arxiv.org/abs/2503.11999>
- [33] M. Janner, Y. Du, J. B. Tenenbaum, and S. Levine, "Planning with diffusion for flexible behavior synthesis," *arXiv preprint arXiv:2205.09991*, 2022.
- [34] P. Hansen-Estruch, I. Kostrikov, M. Janner, J. G. Kuba, and S. Levine, "Idql: Implicit q-learning as an actor-critic method with diffusion policies," 2023. [Online]. Available: <https://arxiv.org/abs/2304.10573>
- [35] M. Reuss, M. Li, X. Jia, and R. Lioutikov, "Goal-conditioned imitation learning using score-based diffusion policies," 2023. [Online]. Available: <https://arxiv.org/abs/2304.02532>
- [36] T. Pearce, T. Rashid, A. Kanervisto, D. Bignell, M. Sun, R. Georgescu, S. V. Macua, S. Z. Tan, I. Momennejad, K. Hofmann, and S. Devlin, "Imitating human behaviour with diffusion models," 2023. [Online]. Available: <https://arxiv.org/abs/2301.10677>
- [37] A. Fishman, A. Murali, C. Eppner, B. Boots, and D. Fox, "Motion policy networks," in *Conference on Robot Learning*. PMLR, 2023, pp. 967–977.
- [38] M. Dalal, J. Yang, R. Mendonca, Y. Khaky, R. Salakhutdinov, and D. Pathak, "Neural mp: A generalist neural motion planner," *arXiv preprint arXiv:2409.05864*, 2024.
- [39] M. Dalal, A. Mandlekar, C. Garrett, A. Handa, R. Salakhutdinov, and D. Fox, "Imitating Task and Motion Planning with Visuomotor Transformers," in *Conference on Robot Learning*, 2023.
- [40] T. Tosun, E. Mitchell, B. Eisner, J. Huh, B. Lee, D. Lee, V. Isler, H. S. Seung, and D. Lee, "Pixels to plans: Learning non-prehensile manipulation by imitating a planner," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 7431–7438.
- [41] J. Yamada, Y. Lee, G. Salhotra, K. Pertsch, M. Pflueger, G. Sukhatme, J. Lim, and P. Englert, "Motion planner augmented reinforcement learning for robot manipulation in obstructed environments," in *Conference on Robot Learning*. PMLR, 2021, pp. 589–603.
- [42] J. Carius, F. Farshidian, and M. Hutter, "Mpc-net: A first principles guided policy search," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2897–2904, 2020.
- [43] B. Abbatematteo, E. Rosen, S. Tellex, and G. Konidaris, "Bootstrap-

- ping motor skill learning with motion planning,” in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 4926–4933.
- [44] Y. Pan, C.-A. Cheng, K. Saigol, K. Lee, X. Yan, E. A. Theodorou, and B. Boots, “Agile autonomous driving using end-to-end deep imitation learning,” *Robotics: Science and Systems XIV*, 2017. [Online]. Available: <https://api.semanticscholar.org/CorpusID:53873353>
- [45] T. Jurgenson and A. Tamar, “Harnessing reinforcement learning for neural motion planning,” in *Robotics: Science and Systems*, 2019.
- [46] H. Ha, J. Xu, and S. Song, “Learning a decentralized multi-arm motion planner,” in *Conference on Robot Learning*, 2021.
- [47] S. Levine and V. Koltun, “Guided policy search,” in *International conference on machine learning*. PMLR, 2013, pp. 1–9.
- [48] M. Zhang, J. Barreiros, and A. O. Onol, “Plan-guided reinforcement learning for whole-body manipulation,” *arXiv preprint arXiv:2310.12263*, 2023.
- [49] F. Jenelten, J. He, F. Farshidian, and M. Hutter, “Dtc: Deep tracking control,” *Science Robotics*, vol. 9, no. 86, p. eadh5401, 2024.
- [50] J. Brüdigam, A.-A. Abbas, M. Sorokin, K. Fang, B. Hung, M. Guru, S. Sosnowski, J. Wang, S. Hirche, and S. L. Cleac’h, “Jacta: A versatile planner for learning dexterous and whole-body manipulation,” in *Conference on Robot Learning*, 2024.
- [51] Y. Qin, B. Huang, Z.-H. Yin, H. Su, and X. Wang, “Dexpoint: Generalizable point cloud reinforcement learning for sim-to-real dexterous manipulation,” *Conference on Robot Learning (CoRL)*, 2022.
- [52] B. Ai, Z. Wu, and D. Hsu, “Invariance is key to generalization: Examining the role of representation in sim-to-real transfer for visual navigation,” in *Experimental Robotics - The 18th International Symposium, ISER 2023, Melia, Chiang Mai, Thailand, 26-30 November 2023*, ser. Springer Proceedings in Advanced Robotics, M. H. A. Jr. and O. Khatib, Eds., vol. 30. Springer, 2023, pp. 69–80. [Online]. Available: https://doi.org/10.1007/978-3-031-63596-0_7
- [53] Y. Jiang, C. Wang, R. Zhang, J. Wu, and L. Fei-Fei, “Transic: Sim-to-real policy transfer by learning from online correction,” *arXiv preprint arXiv:2405.10315*, 2024.
- [54] H. Zhu, T. Zhao, X. Ni, J. Wang, K. Fang, L. Righetti, and T. Pang, “Should we learn contact-rich manipulation policies from sampling-based planners?” *IEEE Robotics and Automation Letters*, vol. 10, no. 6, pp. 6248–6255, 2025.
- [55] S. M. LaValle, *Planning algorithms*. Cambridge university press, 2006.
- [56] S. Le Cleac’h, T. A. Howell, S. Yang, C.-Y. Lee, J. Zhang, A. Bishop, M. Schwager, and Z. Manchester, “Fast contact-implicit model predictive control,” *IEEE Transactions on Robotics*, vol. 40, pp. 1617–1629, 2024.
- [57] H. J. T. Suh, T. Pang, and R. Tedrake, “Bundled gradients through contact via randomized smoothing,” *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4000–4007, 2022.
- [58] R. Tedrake and the Drake Development Team, “Drake: Model-based design and verification for robotics,” 2019. [Online]. Available: <https://drake.mit.edu>
- [59] Y. Ze, G. Zhang, K. Zhang, C. Hu, M. Wang, and H. Xu, “3d diffusion policy: Generalizable visuomotor policy learning via simple 3d representations,” in *Proceedings of Robotics: Science and Systems (RSS)*, 2024.
- [60] S. Liu, Z. Zeng, T. Ren, F. Li, H. Zhang, J. Yang, C. Li, J. Yang, H. Su, J. Zhu *et al.*, “Grounding dino: Marrying dino with grounded pre-training for open-set object detection,” *European Conference on Computer Vision*, 2024.
- [61] H. Cai, J. Li, M. Hu, C. Gan, and S. Han, “Efficientvit: Lightweight multi-scale attention for high-resolution dense prediction,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 17 302–17 313.
- [62] C. Doersch, Y. Yang, M. Vecerik, D. Gokay, A. Gupta, Y. Aytar, J. Carreira, and A. Zisserman, “Tapir: Tracking any point with per-frame initialization and temporal refinement,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 10 061–10 072.
- [63] N. Karaev, I. Rocco, B. Graham, N. Neverova, A. Vedaldi, and C. Rupprecht, “Cotracker: It is better to track together,” *arXiv:2307.07635*, 2023.
- [64] H. Zhu, T. Pang, X. Ni, T. Zhao, K. Fang, L. Righetti, and J. Wang, “Learning goal-conditioned diffusion policy for contact-rich bimanual manipulation through planning-guided data synthesis,” in *ICRA 2024 Workshop on Future Roadmap for Manipulation Skills*, 2023.