

# Pick-and-place Manipulation Across Grippers Without Retraining: A Learning-optimization Diffusion Policy Approach

Xiangtong Yao<sup>1\*</sup>, Yirui Zhou<sup>1\*</sup>, Yuan Meng<sup>1</sup>, Liangyu Dong<sup>1</sup>, Lin Hong<sup>2</sup>,  
Zitao Zhang<sup>1</sup>, Zhenshan Bing<sup>1,3†</sup>, Kai Huang<sup>4</sup>, Fuchun Sun<sup>5</sup>, Alois Knoll<sup>1</sup>

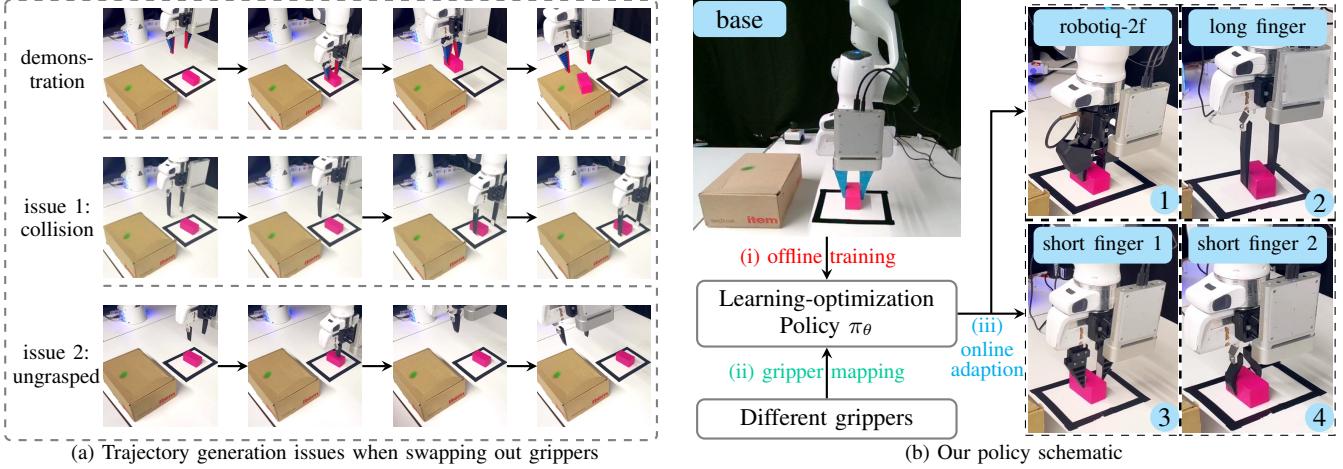


Fig. 1: (a) Different issues of the trajectory generation of the vanilla diffusion policy across grippers without the policy retaining. (b) Our policy aims to transfer pick-and-place skills across grippers without policy retraining while ensuring task completion, involving (i) offline training the policy  $\pi_\theta$  with data collected via a base gripper, (ii) gripper-aware mapping, and (iii) online optimization adaption to enforce the generated trajectories to fit different unseen grippers.

**Abstract**— Current robotic pick-and-place policies typically require consistent gripper configurations across training and inference. This constraint imposes high retraining or fine-tuning costs, especially for imitation learning-based approaches, when adapting to new end-effectors. To mitigate this issue, we present a diffusion-based policy with a hybrid learning-optimization framework, enabling zero-shot adaptation to novel grippers without additional data collection for retraining policy. During training, the policy learns manipulation primitives from demonstrations collected using a base gripper. At inference, a diffusion-based optimization strategy dynamically enforces kinematic and safety constraints, ensuring that generated trajectories align with the physical properties of unseen grippers. This is achieved through a constrained denoising procedure that adapts trajectories to gripper-specific parameters (e.g., tool-center-point offsets, jaw widths) while preserving collision avoidance and task feasibility. We validate our method on a Franka Panda robot across six gripper configurations, including 3D-printed fingertip, flexible silicone gripper, and Robotiq 2F-85 gripper. Our approach achieves a 93.3% average task success rate across grippers (vs. 23.3–26.7% for diffusion policy baselines), supporting tool-center-point variations of 16–23.5 cm and jaw widths of 7.5–11.5 cm. The results demonstrate that constrained diffusion enables robust cross-gripper manipulation while maintaining the sample efficiency of imitation learning,

eliminating the need for gripper-specific retraining. Video and code are available at <https://github.com/yaoxt3/GADP>.

## I. INTRODUCTION

Robotic pick-and-place operations form the cornerstone of automated manipulation, which is essential for activities from household to industrial scenarios. A variety of methods have been proposed to endow robots with pick-and-place capabilities, including reinforcement learning [1]–[3], imitation learning [4], and motion planning [5]. While these methods have demonstrated the ability to manipulate multiple objects (known or unknown) [6], [7], operate in dynamic environments [5], or perform one-shot or few-shot tasks [8], [9], they typically assume fixed end-effector morphology throughout training and deployment. This assumption restricts the transfer of pick-and-place skills across grippers and compromises overall generalizability.

Recent research mitigates this limitation through multi-embodiment learning strategies. For example, HPT trains a transformer-based policy with large-scale heterogeneous robot datasets covering diverse embodiments and tasks [10]. Wang et al. [11] employ latent space alignment to transfer pick-and-place skills between different robot manipulators. Despite promising results, these methods often require substantial embodiment-specific data [10] or policy fine-tuning to accommodate new gripper configurations [10], [11], leading to additional adaptation costs and limiting

<sup>1</sup> Technical University of Munich, Munich, Germany

<sup>2</sup> Hong Kong University of Science and Technology, China

<sup>3</sup> Nanjing University, Nanjing, China

<sup>4</sup> Sun Yat-sen University, Guangzhou, China

<sup>5</sup> Tsinghua University, Beijing, China

\* Equal contribution

†Corresponding author: Zhenshan Bing zhenshan.bing@tum.de

deployment flexibility. This challenge is particularly pronounced for behavior cloning algorithms, such as diffusion policies [12], [13]. In online developments, switching from a base gripper to one with a different morphology can shift the end-effector’s tool-center-point (TCP), causing collisions or missed grasps if a pretrained diffusion policy is naively applied to generate trajectories, as shown in Fig. 1a.

We bridge this gap through an integration of diffusion policy [12], [13] with a learning-optimization hybrid strategy. This strategy allows the trained policy to fit new grippers without policy retraining or fine-tuning, as shown in Fig. 1b. The policy first learns pick-and-place manipulation primitives following diffusion policy, where the training demonstration data is collected via a base gripper. During the policy’s online inference, the base gripper is replaced with a new one. To ensure the generative trajectories are valid for new gripper configurations, we first incorporate the gripper-specific dimensional offsets into the policy inputs, and recast the traditional diffusion sampling process as a constrained-optimization process. The optimization introduces mathematical encoding of task and safety constraints (e.g., object grasping, avoiding collisions) to the denoising process, utilizing a quadratic programming optimization method to incrementally refine an initially noisy trajectory to conform to unseen gripper geometries. Crucially, this approach does not require policy retraining or fine-tuning, thereby preserving flexibility for real-world deployment with minimal overhead.

We validate our approach with a Franka Panda manipulator equipped with various grippers for pick-and-place tasks—including 3D-printed fingertips of different heights, flexible fingertips, and the Robotiq 2F gripper—achieving robust performance without policy retraining. Key contributions are summarized as follows.

- A learning-optimization strategy is introduced into the diffusion policy, dynamically optimizing the generation of trajectories to counteract the TCP offset problem caused by switching grippers, thus ensuring task completion. This ensures the trained policy can solve pick-and-place tasks with new grippers without retaining and fine-tuning operations.
- A general formulation of the task and safety constraints is designed for the optimization process. Moreover, a cumulative optimization is utilized to ensure the temporal consistency of the generated trajectory.
- Real-world experiments across different grippers show our method achieves 93.3% and 70.0% success rates for seen and unseen objects, respectively—outperforming baseline diffusion policies (23.3% and 20.0%).

## II. RELATED WORK

Transferring robot manipulation knowledge between different embodiments can improve the policy’s flexibility and generalizability for adapting to new tasks or hardware settings [10], including the knowledge transitions between different grippers [14] or robot manipulators [10], [11], [15]. For example, Zakka et al. [14] combine the temporal

cycle-consistency method with imitation learning to learn an invariant feature space for different embodiments, empowering robot manipulators (target domain) to perform the tasks by imitating video demonstrations of human experts (source domain). While the feature correspondence between the source and target domains has to be re-trained if a new source or target domain is introduced. Wang et al. [11] mitigate this limitation by learning a feature latent space for the source (a 7DoF manipulator) and target (a 6DoF manipulator) domains alongside the encoder and decoder for each domain. Then, they employ adversarial training and cycle consistency for latent alignment, which aligns the target encoder-decoder to that of the source domain, respectively. Similarly, in [15], the authors employ latent alignment for source and target domains, achieving knowledge transferring between heterogeneous manipulators. By leveraging large-scale heterogeneous robot data, Wang et al. [10] propose a heterogeneous pre-trained transformer (HPT) that can perform different tasks across heterogeneous robot manipulators.

However, the manipulation knowledge transition of these approaches relies on transfer learning or domain adaptation/alignment [10], [11], [15], where introducing new hardware configuration (e.g., grippers) to the policy require fine-tuning policy [10] or retraining auxiliary networks (e.g., encoder and decoder) [11], [15], limiting the flexibility of the policy. Our method circumvents these fine-tuning and retraining operations by integrating a learning-optimization hybrid strategy into diffusion policy [12], necessitating the introduction of task-oriented constraints for new hardware configurations and employing optimization during the policy online inference phase to ensure completion of the task.

Diffusion models have shown promise for solving decision-making tasks, including motion planning [16], [17], imitation learning [12], [13], [18], and reinforcement learning [19], [20]. When solving decision-making tasks through generative models, the policy needs not only to accomplish the task goal but also to satisfy certain task constraints, such as collision avoidance. Some works add a residual loss to the training objectives if the task constraints are consistent during the policy training and inference [17], [21]. A more flexible method is training diffusion models via classifier-free guidance [22], introducing conditioning variables that represent constraints into the policy training, such as physical constraints for guiding human motion generation [23]. However, the model conditioning encourages the generated samples to adhere to task constraints rather than strict guarantee constraints [24]. Alternative post-processing methods draw constraints into the last denoising stage of the sample-generating process and obtain samples that satisfy the constraints by solving an optimization problem [25], [26]. As the optimization problem does not consider the unknown data likelihood, post-processing may result in samples significantly deviating from the data distribution [27]. To mitigate this issue, an iterative projection is integrated into the denoising process, confining each generated sample to the constraints. Some works employ it for sequential

decision-making [24], which is time-consuming. Or model-based trajectory control [27], falling short of seamlessly adapting to a new model since introducing a new gripper changes the robot dynamics. Our approach utilizes projection alongside optimization techniques in the denoising process but mitigates the time consumption and makes the policy seamlessly adaptable to new grippers.

### III. BACKGROUND ON DIFFUSION POLICY

Denoising diffusion probabilistic models (DDPMs) [28] are generative models that iteratively refine noisy data samples to match a target distribution. They employ two complementary Markovian processes: a forward diffusion process that incrementally adds Gaussian noise to an original data sample  $x^0$  over  $k$  steps, and a reverse denoising process that removes the accumulated noise step by step. Building on this foundation, Diffusion Policy [12] adapts DDPMs for robotic manipulation. In this setting, a policy  $\pi_\theta(A_t | S_t)$  is trained via behavior cloning to generate action sequences conditioned on visual and robot state observations. The model injects noise into the action samples during training and learns to predict this noise over multiple denoising steps, thereby guiding the policy to produce coherent actions at inference. Specifically, at test time, the reverse denoising process is iterated  $K$  times to obtain a “clean” action sequence  $A_t^0$ . Concretely, assume a horizon  $T_a$  and a state sequence  $S_t = \{s_{t-T_a+1}, s_{t-T_a+2}, \dots, s_t\}$ . The policy generates an action sequence  $A_t = \{a_t, a_{t+1}, \dots, a_{t+T_a-1}\}$  based on the learned diffusion mechanism. After executing  $A_t$  on the robot, new sensor readings yield a subsequent state sequence  $S_{t+T_a}$ , from which the policy infers the next action sequence. By iterating this diffusion-based inference, the robot continuously adapts its actions to changing environment states, enabling robust long-horizon control.

## IV. METHODOLOGY

### A. Problem Statement

We formalize the challenge of gripper-agnostic manipulation via diffusion policies through three core components:

- **Gripper Configuration:** Let  $\mathbb{G} \subset \mathbb{R}^{d_g}$  denote the space of two-finger gripper parameters encoding morphology, maximum width  $w^{\max}$ , and tool-center-point (TCP)  $z$ , as shown in Fig. 2.
- **Observation Domain:**  $\mathcal{O} = \mathcal{S}_{\text{sce}} \times \mathcal{S}_{\text{rob}}$  where  $\mathcal{S}_{\text{sce}}$  represents scene observations (3D point clouds) and  $\mathcal{S}_{\text{rob}} = SE(3) \times [0, g]$  the robot state (end-effector pose  $\mathbf{x}_{\text{ee}} \in SE(3)$ , gripper width  $g$ )<sup>1</sup>.  $\mathbf{x}_{\text{ee}}$  is gripper-agnostic and reading from the robot.  $g$  is specific for grippers.
- **Action Space:**  $\mathcal{A} \subset \mathbb{R}^{d_a}$  containing end-effector displacements  $\Delta \mathbf{x}_{\text{ee}}$  and gripper commands.

The policy  $\pi_\theta$  is trained on demonstrations  $\mathcal{D} = \{\tau^{(i)}\}_{i=1}^N$  collected with a reference gripper  $\mathbb{G}_0 \in \mathbb{G}$ . Each trajectory  $\tau = \{(\mathbf{o}_t, \mathbf{a}_t)\}_{t=0}^T$  satisfies:  $\mathbf{o}_t = (\mathcal{S}_{\text{sce}}^0, \mathbf{x}_{\text{ee}}^0, g_t^0)$  and  $\mathbf{a}_t \sim \pi_{\text{expert}}(\cdot | \mathbf{o}_t)$ , where superscript 0 indicates  $\mathbb{G}_0$  parameters.

<sup>1</sup>This paper focuses on two-finger grippers problems in  $SE(3)$  space.  $SE(6)$  space and three-finger grippers implementation is future work.



Fig. 2: Gripper variants of different morphologies. (Unit: cm)

During deployment with novel gripper  $\mathbb{G}_i \neq \mathbb{G}_0$ , the observation-action distribution shifts due to (1) visual/kinematic differences  $\mathcal{O}_i \neq \mathcal{O}_0$ , and (2) policy mismatch  $p_\theta(\mathcal{A}|\mathcal{O}_i) \neq p_\theta(\mathcal{A}|\mathcal{O}_0)$ . This manifests as trajectory divergence:

$$\|\tau_{1:T}^{\mathbb{G}_i} - \tau_{1:T}^{\mathbb{G}_0}\|_{\mathcal{W}} > \delta_{\text{tol}}, \quad (1)$$

where  $\mathcal{W}$  is the task-specific metric space and  $\delta_{\text{tol}}$  the success threshold, e.g., objects cannot be grasped with shorter grippers, and collisions can result from using longer grippers.

To mitigate these issues, we develop policy  $\pi_\theta^*$  that maintains task performance under gripper variation, combining (1) gripper-invariant knowledge learning and (2) morphology-aware trajectory optimization to achieve  $\pi_\theta^*$  without policy retraining. The overview framework is shown in Fig. 3.

### B. Learning Gripper-agnostic Grasping Knowledge

Visuomotor policies, like Diffusion policy and 3D Diffusion Policy [12], [13], depend on visual observations to generate robot trajectories. However, swapping out different grippers during the online execution can alter visual observations (both RGB and point cloud inputs), as shown in Fig. 2. Such alterations lead to out-of-distribution trajectory generations, reducing the task success rate [29]. To mitigate this limitation, we introduce a gripper-invariant *grasping probability map*  $\mathcal{G}_{\text{prob}}$  as an additional observation component, which captures object-centric grasp affordances that are independent of end-effector geometry, thereby guiding the policy to focus on relevant object features rather than gripper-specific visual patterns. By decoupling object-related cues from the gripper’s appearance,  $\mathcal{G}_{\text{prob}}$  enhances the policy’s robustness to variations in gripper morphology, maintaining stable task performance across different grippers.

We adopt the Generative Grasping CNN (GG-CNN) [30] for  $\mathcal{G}_{\text{prob}}$  synthesis from depth images. GG-CNN is pre-trained on the Cornell Grasping Dataset [31], which contains 885 RGB-D images with annotated parallel-jaw grasps across 240 objects. However, real-world pick-and-place manipulations introduce two key challenges: (1) the hand-eye camera moving with the robot, causing scale variations in object pixels, and (2) lighting changes disturb depth sensor readings. These factors degrade GG-CNN’s output stability, i.e.,  $\mathcal{G}_{\text{prob}}$ , and destabilize policy training and inference performance.

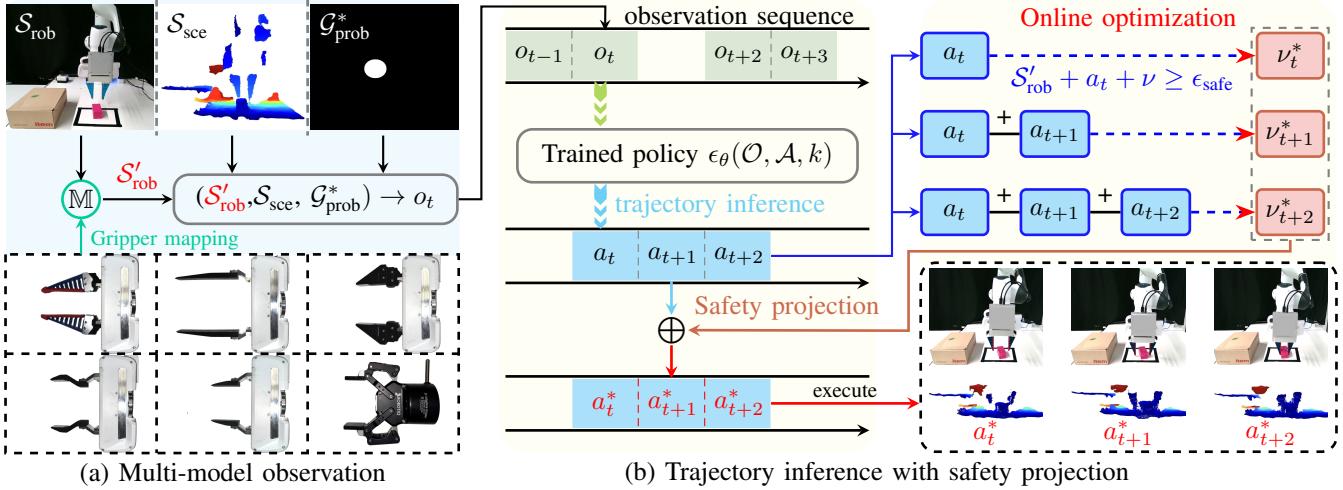


Fig. 3: Overview of policy. (a) The multi-modal observation consists of robot pose  $S'_{\text{rob}}$ , scene point clouds  $S_{\text{sce}}$ , and grasping probability map  $G^*_{\text{prob}}$ . Gripper morphological variations are encoded into  $S'_{\text{rob}}$  via gripper mapping. (b) Safety-Constrained trajectory projection via online optimization process, enforcing the executive trajectory to satisfy task and safety constraints.

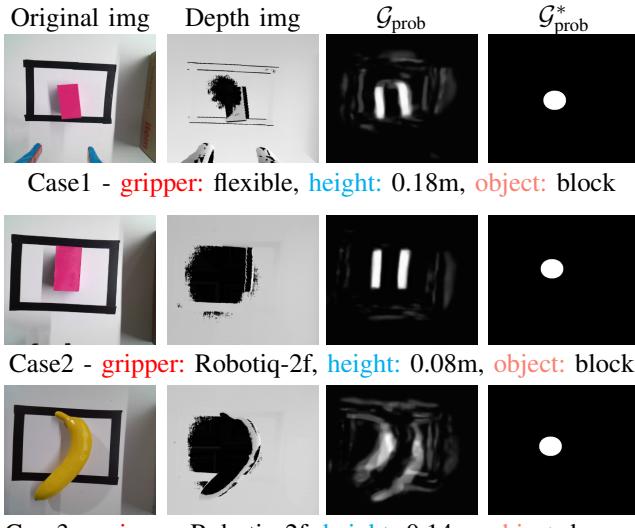


Fig. 4: The gripper-agnostic grasping knowledge. Dynamic changes in robot pose and gripper variants cause changes in visual observations, including RGB-D images and object grasp probabilities,  $\mathcal{G}^*_{\text{prob}}$  provides stable visual information.

To address this issue, our solution involves: (1) threshold filtering: discard pixels with  $\mathcal{G}_{\text{prob}} < 0.7$ , (2) centroid computation:  $\mathbb{O} = \frac{1}{N} \sum_{i=1}^N (u_i, v_i)$  for remaining pixels, and (3) region masking: generate  $\mathcal{G}^*_{\text{prob}}$  through circular masking ( $r = 30$  pixels) about  $\mathbb{O}$ . The map  $\mathcal{G}^*_{\text{prob}}$  satisfies:

$$\mathcal{G}^*_{\text{prob}}(u, v) = \begin{cases} 1 & , \text{if } \|(u, v) - \mathbb{O}\|_2 \leq 30 \\ 0 & , \text{otherwise} \end{cases} \quad (2)$$

This spatial filtering maintains grasp affordance information while eliminating outlier predictions caused by sensor noise, as shown in Fig.4. The policy observation consists of:

$$\mathcal{O}^* = \mathcal{G}^*_{\text{prob}} \times \mathcal{S}_{\text{sce}} \times \mathcal{S}_{\text{rot}}, \quad (3)$$

### C. Optimizing Trajectory Generation

The training scheme of our policy  $\pi_\theta$  is consistent with that of Diffusion Policy (DP) [12], i.e. DDPM, with observations  $\mathcal{O}^*$  and MSE training loss. During inference,  $\pi_\theta$  employs Denoising Diffusion Implicit Models (DDIM) [32] sampling with 10 denoising iterations. However, our policy introduces a strategy to enforce the generative trajectory to fit different grippers, comprising gripper-geometry aware mapping and safety-constrained trajectory projection.

**Gripper mapping:** Gripper morphological variations induce the end-effector's pose discrepancies during identical object manipulation, primarily along: (1) vertical axis ( $z$ ): tool-center-point offset, and (2) gripper state ( $g$ ): grasping width differences. The discrepancies causes inconsistent actions predicted by  $\pi_\theta$  across grippers.

Let  $\mathbb{G}_{(0)}$  denote the reference gripper used during training, and  $\mathbb{G}_{(i)}$  represent a novel gripper of category  $i$ . We define differentiable mapping functions  $\text{IM}_h(\cdot)$  and  $\text{IM}_g(\cdot)$  that project  $\mathbb{G}_{(i)}$  parameters to the  $\mathbb{G}_{(0)}$  basis:

$$\begin{aligned} z'_{(i)} &= \text{IM}_h(z_{(i)}) = z_{(i)} + \Delta h_{(i)}, \\ g'_{(i)} &= \text{IM}_g(g_{(i)}) = \alpha_{(i)} g_{(i)}, \end{aligned} \quad (4)$$

where  $z_{(i)}$  is the measured height of end-effector equipped with  $\mathbb{G}_{(i)}$ ,  $\Delta h_{(i)} = z_{(0)} - z_{(i)}$  is the height offset from reference gripper,  $g_{(i)} \in [g_{\min}, g_{\max}]$  is the real-time grasping width,  $\alpha_{(i)} = g_{(0)}^{\max}/g_{(i)}^{\max}$  scales widths. The mapping parameters  $\{\Delta h_{(i)}, \alpha_{(i)}\}$  are obtained through offline calibration with mechanical measurement of gripper dimensions.

This transformation preserves the policy's internal representation while adapting to physical gripper properties, enabling zero-shot generalization to novel end-effectors. Current implementation focuses on translational pose adaptation, and the rotational compensation remains future work. During policy execution, the transformed pose  $S'_{\text{rob}} = (x, y, z'_{(i)}, g'_{(i)})$  is fed to  $\pi_\theta$  instead of  $S_{\text{rob}}$  in (3), maintaining observation-space consistency across grippers.

**Safety-Constrained Trajectory Projection** While gripper mapping aligns geometric parameters, visual perception differences from gripper morphology can still induce unsafe trajectory variations. To guarantee constraint satisfaction, we integrate a projection layer into the DDIM denoising process [32]. The modified reverse diffusion step becomes:

$$\mathbf{a}_t^{k-1} = \text{Proj}_{\mathcal{C}}(\mu_k(\mathbf{a}_t^k, \epsilon_\theta(\mathbf{a}_t^k, \mathbf{o}_t, k))), \quad (5)$$

where  $\text{Proj}_{\mathcal{C}}(\cdot)$  enforces safety constraints  $\mathcal{C}$  through the following two steps.

a) **Constraint-aware denoising:** For efficiency, projection activates only in the final denoising steps ( $k \leq 2$ ). At each step  $k$ , we solve a quadratic program problem:

$$\begin{aligned} \nu_t^{k*} &= \arg \min_{\nu_t^k} \|\nu_t^k\|_2^2 \\ \text{s.t. } \mathcal{S}'_{\text{rob}}(z)_t + \Phi(\mathbf{a}_t^k) + \nu_t^k &\geq \epsilon_{\text{safe}} \end{aligned} \quad (6)$$

where  $\Phi(\cdot)$  maps latent actions to Cartesian displacement, which is denormalization in our case,  $\epsilon_{\text{safe}} = 0.01$  m (safety margin), and  $\nu_t^k$  is the minimal corrective offset.

b) **Temporal consistency enforcement:** To maintain safety over the policy's  $T_a$ -step action horizon ( $j \in [0, T_a - 1]$ ), we extend (6) with cumulative constraints:

$$\mathcal{S}'_{\text{rob}}(z)_t + \sum_{r=0}^j \Phi(\mathbf{a}_{t+r}^k) + \nu_{t+j}^k \geq \epsilon_{\text{safe}} \quad (7)$$

The projected actions  $\mathbf{a}_t^{k*} = \Phi^{-1}[\Phi(\mathbf{a}_t^k) + \nu_t^{k*}]$  guarantee:

$$\mathbb{P}\left(\bigcap_{j=0}^{T_a} \{\mathcal{S}'_{\text{rob}}(z)_{t+j} \geq \epsilon_{\text{safe}}\}\right) = 1, \quad (8)$$

indicating the cumulative trajectory is always safe, with an example of  $\mathcal{S}'_{\text{rob}}(z)_{t+1} = \mathcal{S}'_{\text{rob}}(z)_t + \Phi(\mathbf{a}_t^{k*})$ .

## V. EXPERIMENTS

To evaluate our proposed method, we designed experiments to (i) evaluate the stability of grasping probability maps under different object and robot pose conditions, (ii) validate different policy's generalization for pick-and-place tasks across diverse two-finger grippers without retraining the policy, and (iii) assess the generated trajectory's safety and task completion issues across different policies.

### A. Experiment Setup

The experimental setup employs a Franka Emika Panda robotic arm equipped with interchangeable end-effectors, including a native parallel gripper augmented with custom 3D-printed fingertips and a Robotiq 2F-85 adaptive gripper, as shown in Fig. 2. Real-time control is managed via a workstation running Ubuntu 20.04 with a real-time kernel to ensure deterministic low-latency operation, while Diffusion Policy framework is trained and executed on another workstation with a NVIDIA RTX-4090 GPU and an AMD 7950x CPU. Perception integrates an egocentric Azure Kinect DK camera and a fixed Intel RealSense D455 camera. Sensor data synchronizations are facilitated through ROS noetic.

---

### Algorithm 1 Gripper-Aware Trajectory Generation

---

**Require:** Novel gripper  $\mathbb{G}_i$ , Observation  $\mathbf{o}_t$ , Safety margin  $\epsilon_{\text{safe}}$ , trained noise predicted  $\epsilon_\theta$

**Ensure:** Safe trajectory  $\tau = \{\mathbf{a}_{t:t+T_a-1}\}$

// Online Inference:

- 1:  $\mathcal{S}'_{\text{rob}} \leftarrow \text{IM}_h(z_{(i)}) = z_{(i)} + \Delta h_{(i)}$  // Gripper Mapping;
- 2:  $\mathbf{g}'_{(i)} \leftarrow \text{IM}_g(g_{(i)}) = \alpha_{(i)} g_{(i)}$
- 3:  $\tilde{\mathbf{o}}_t \leftarrow (\mathcal{G}_{\text{prob}}^*, \mathcal{S}_{\text{sce}}, \mathcal{S}'_{\text{rob}})$
- 4:  $\mathbf{a}_t^K \sim \mathcal{N}(0, \mathbf{I})$  // Diffusion Process:
- 5: **repeat**
- 6:      $k \leftarrow K - 1$ , and  $K \leftarrow K - 1$
- 7:      $\mathbf{a}_t^k \leftarrow \mathcal{N}(\mu_k(\mathbf{a}_t^{k+1}, \epsilon_\theta(\mathbf{a}_t^{k+1}, \tilde{\mathbf{o}}_t, k + 1)), 0)$
- 8:     **if**  $k \leq 1$ : // Safety Projection:
- 9:         **for**  $j \leftarrow 0$  **to**  $T_a - 1$  **do**
- 10:              $\nu_{t+j}^{k*} \leftarrow \arg \min_{\nu} \|\nu_{t+j}^k\|_2^2$
- 11:             s.t.  $\mathcal{S}'_{\text{rob}}(z)_t + \sum_{r=0}^j \Phi(\mathbf{a}_{t+r}^k) + \nu_{t+j}^k \geq \epsilon_{\text{safe}}$
- 12:          $\mathbf{a}_{t:t+T_a-1}^k \leftarrow \Phi^{-1}[\Phi(\mathbf{a}_{t:t+T_a-1}^k) + \nu_{t:t+T_a-1}^{k*}]$
- 13:     **until**  $K = 0$
- 14:      $\tau \leftarrow \text{Decode}(\mathbf{a}_{t:t+T_a-1}^0)$

---

### B. Data Collection and Training

1) **Data Collection:** Our dataset comprises 59 expert demonstration trajectories of pick-and-place tasks with the gripper  $\mathbb{G}_0$  in Fig. 2. Multimodal observations are captured through three synchronized modalities: (1) Proprioceptive states: including the 6-DoF end-effector pose and continuous gripper width measurements obtained via the Franky API<sup>2</sup>; (2) 3D point clouds are acquired via the D455 camera and are processed with farthest point sampling. To estimate the gripper-agnostic *grasping probability map*  $\mathcal{G}_{\text{prob}}^*$ , we adapt the grasping probability representation from [33] to process the Kinect camera's depth and RGB images; (3) Kinesthetic demonstration signals: recorded via a joystick controller operating at 2Hz, encoding relative Cartesian displacements alongside a one-hot encoded gripper action (open/closed).

2) **Training:** We select the Simple DP3 implementation [13] as our policy's backbone for efficiency. To integrate the *grasping probability map*  $\mathcal{G}_{\text{prob}}^*$  into the policy framework, we adapt Simple DP3 to Simple DP3 MM (Multi-Modal Simple Diffusion Policy 3D). This architecture concatenates three encoded inputs: Proprioceptive states and point cloud features processed identically to Simple DP3;  $\mathcal{G}_{\text{prob}}^*$  encoded through a ResNet-18 network. The model retains the same pruned UNet backbone as Simple DP3. Training follows the original Simple DP3 procedure but extends the duration to 8,000 epochs to accommodate the additional modality.

### C. Real-world Experiment

1) **Evaluation of Grasping Probability Map:** We evaluate the stability of grasping probability maps under three perturbation scenarios, including graspable object variation, gripper morphology, and viewpoint changes (i.g. end-effector

<sup>2</sup><https://github.com/TimSchneider42/franky>

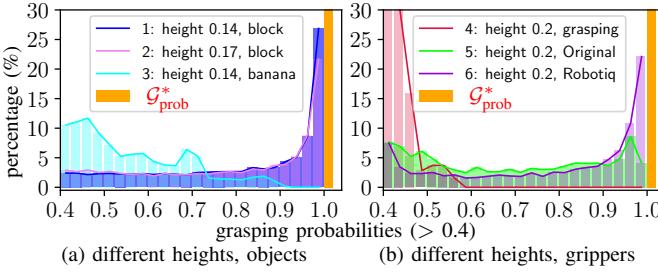


Fig. 5: Grasping probabilities map under different conditions.

TABLE I: KL divergence between distributions in Fig. 5.

	$\mathcal{G}_{\text{prob}}^{(2)}$	$\mathcal{G}_{\text{prob}}^{(3)}$	$\mathcal{G}_{\text{prob}}^{(4)}$	$\mathcal{G}_{\text{prob}}^{(5)}$	$\mathcal{G}_{\text{prob}}^{(6)}$
$\text{KL}(\mathcal{G}_{\text{prob}}^{(i)}, \mathcal{G}_{\text{prob}}^{(1)})$	0.009	0.929	0.276	2.271	0.052
$\text{KL}(\mathcal{G}_{\text{prob}}^{*(i)}, \mathcal{G}_{\text{prob}}^{*(1)})$	0.000	0.000	0.000	0.000	0.000

heights). We define stability as the extent to which the probability distribution of these maps remains consistent across the different scenarios. If the variance in distribution is large, the resulting shift can adversely affect the policy’s trajectory generation from training to inference. Fig. 4 visualizes raw GG-CNN outputs  $\mathcal{G}_{\text{prob}}$  versus our processed maps  $\mathcal{G}_{\text{prob}}^*$  across these conditions. Fig. 5 presents the difference between  $\mathcal{G}_{\text{prob}}$  and  $\mathcal{G}_{\text{prob}}^*$  through normalized probability histograms, here only shows high-probability parts ( $> 0.4$ ). The baseline  $\mathcal{G}_{\text{prob}}$  demonstrates multimodal distributions versus our unimodal concentration. Moreover, Tab. I quantifies the distribution difference across conditions via KL divergence, indicating  $\mathcal{G}_{\text{prob}}^*$  provides stable grasping information against perturbations. The stabilized grasp cues enable the policy to generate spatially consistent approach trajectories despite sensory perturbations, which is critical for reliable manipulation across grippers.

## 2) Policy Performance Comparison across Grippers:

(a) **Setup:** Our policy and baselines are evaluated in different pick-and-place task setting, including utilizing different grippers and manipulating different objects. Notably, all training demonstrations included only one type of gripper ( $\mathbb{G}_0$ ) and graspable object (a pink block).

(b) **Success Rate Definition:** Picking the object from the tabletop and placing it on a box is a successful trial. All other outcomes are deemed failures, including failure to grasp the object or dropping it during manipulation.

## (c) Baseline and Ablation Study Setting:

- Diffusion Policy (DP) [12]: Implements a baseline framework for temporal action sequence generation through diffusion modeling. DP receives two viewpoint RGB images from the Realsense and Kinect cameras.
- Diffusion Policy 3D (DP 3D) [13]: Extends the Diffusion Policy paradigm by incorporating 3D spatial reasoning. It receives point clouds from the Realsense camera as visual observations.
- Ours w/o Projection: The trajectory optimization strategy in IV-C is not activated in the online infer-

TABLE II: Pick-and-place task across grippers.

Method	Seen object: Block (unit: %)						
	$\mathbb{G}_0$	$\mathbb{G}_1$	$\mathbb{G}_2$	$\mathbb{G}_3$	$\mathbb{G}_4$	$\mathbb{G}_5$	$\mu$
Diffusion Policy	20.0	0.0	60.0	40.0	0	40.0	26.7
Diffusion Policy 3D	20.0	0.0	60.0	60.0	0.0	0.0	23.3
DP + Projection	100.0	-	-	-	-	-	-
DP 3D + Projection	80.0	-	-	-	-	-	-
Ours w/o Projection	100.0	0.0	60.0	40.0	0.0	0.0	33.3
Ours w/o $\mathcal{G}_{\text{prob}}^*$	80.0	20.0	40.0	80.0	100.0	0.0	53.3
Ours	<b>100.0</b>	<b>80.0</b>	<b>100.0</b>	<b>80.0</b>	<b>100.0</b>	<b>100.0</b>	<b>93.3</b>

Method	Unseen object: Banana (unit: %)						
	$\mathbb{G}_0$	$\mathbb{G}_1$	$\mathbb{G}_2$	$\mathbb{G}_3$	$\mathbb{G}_4$	$\mathbb{G}_5$	$\mu$
Diffusion Policy	20.0	0.0	40.0	<b>60.0</b>	40.0	20.0	30.0
Diffusion Policy 3D	20.0	0.0	40.0	40.0	20.0	0.0	20.0
Ours w/o Projection	80.0	0.0	40.0	40.0	0.0	20.0	30.0
Ours w/o $\mathcal{G}_{\text{prob}}^*$	60.0	0.0	40.0	<b>60.0</b>	40.0	0.0	33.3
Ours	<b>80.0</b>	<b>60.0</b>	<b>100.0</b>	<b>60.0</b>	<b>60.0</b>	<b>60.0</b>	<b>70.0</b>

- The training data is collected with the base gripper  $\mathbb{G}_0$  and a pink block.

- The initial pose and position of the object are identical across tests.

-  $\mathbb{G}_i$  indicates different grippers reported in Fig. 2. w/o: without this module. - Every method is validated 5 times for each gripper, totaling 30 evaluations for calculating the average success rate  $\mu$  across different grippers.

ence phase, including gripper mapping and safety-constrained trajectory projection.

- Ours w/o  $\mathcal{G}_{\text{prob}}^*$ : Utilizing RGB images from the Kinect camera to train policy, instead of gripper-agnostic grasping knowledge  $\mathcal{G}_{\text{prob}}^*$ .

(e) **Quantitative results:** Our method demonstrates generalization across diverse gripper geometries relative to baselines. As shown in Table II, our method achieves an average success rate of **93.3%** on all gripper configurations for the seen graspable object (block). In contrast, for the original gripper  $\mathbb{G}_0$ , both DP and DP 3D exhibit notably lower success rates (e.g., DP succeeds only 20%, while DP 3D fails entirely), primarily due to their imprecise motion generation, resulting in collisions with the table. When transitioning to shorter grippers ( $\mathbb{G}_2$  and  $\mathbb{G}_3$ ), DP and DP 3D can complete the task, as the reduced gripper length diminishes the likelihood of collisions. Notably, our policy without the Projection (“Ours w/o Projection”) is capable of successfully completing the task for the original gripper  $\mathbb{G}_0$ ; however, it fails for grippers that deviate more significantly from  $\mathbb{G}_0$  (specifically,  $\mathbb{G}_1$ ,  $\mathbb{G}_4$ , and  $\mathbb{G}_5$ ). Furthermore, for the ablation setting with “Ours w/o  $\mathcal{G}_{\text{prob}}^*$ ”, the performance on  $\mathbb{G}_0$ ,  $\mathbb{G}_3$ , and  $\mathbb{G}_4$  is comparable to, or slightly worse than, that of our full policy. However, when evaluated on grippers that appear different from  $\mathbb{G}_0$  from the perspective of the Kinect camera (namely,  $\mathbb{G}_1$  and  $\mathbb{G}_5$ ), its performance degrades (the success rate ranging from 0 to 20%). These results underscore the critical roles of both the Projection and gripper-agnostic grasping knowledge  $\mathcal{G}_{\text{prob}}^*$  in ensuring robust generalization across varying gripper geometries. Our method demonstrates robust generalization to unseen object geometries compared to baseline approaches, as evidenced by experiments with objects such as bananas. Our method achieves an average success rate of 70%, outperforming baseline policies. While DP and DP 3D occasionally succeed (30% and 20%, respectively) when the unseen object is placed identically to training demonstrations, their reliance on position-specific motion primitives limits adaptability. For

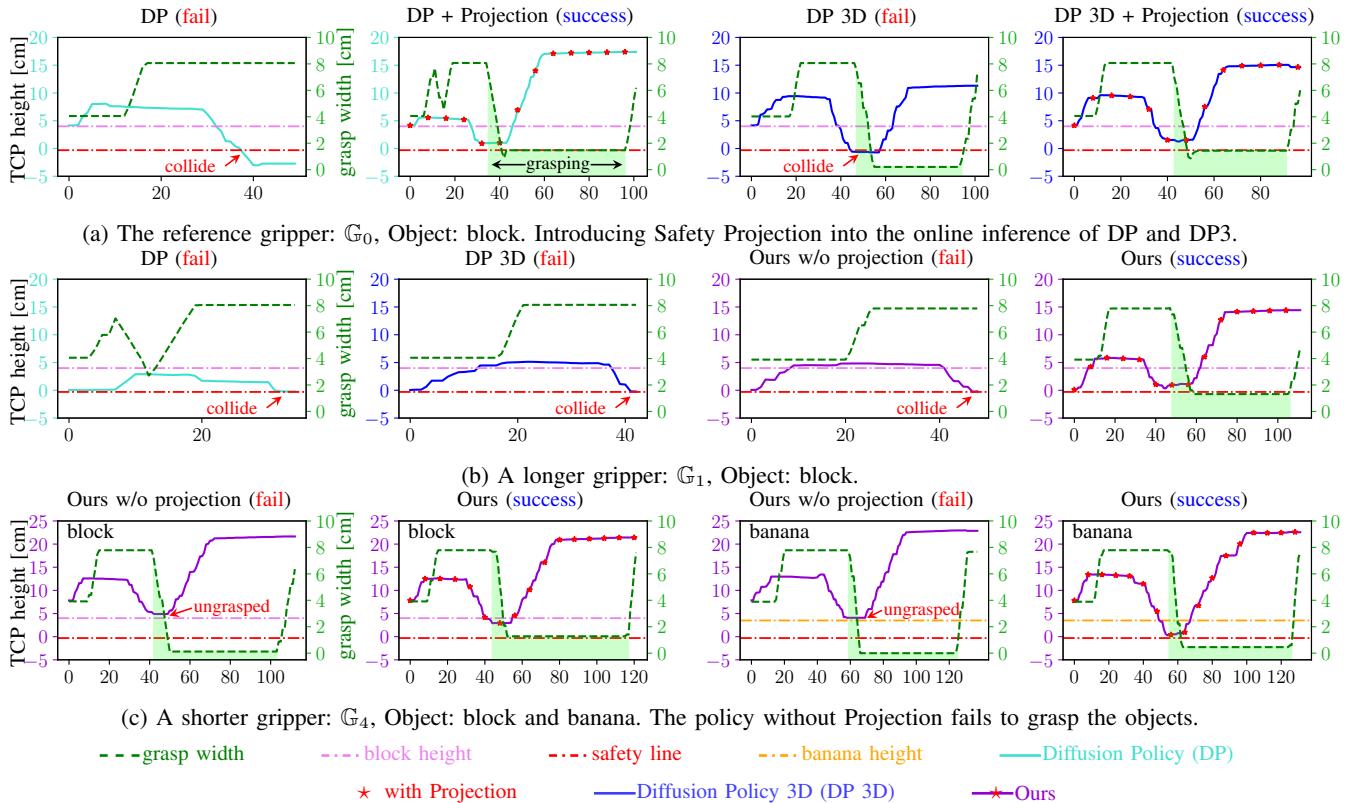


Fig. 6: Safety analysis of different methods under various conditions. Tool-center-point (TCP) height less than the **safety line** indicates that the gripper collided with the table, causing the task to fail. If the TCP height is greater than the height of the object, it indicates a failure to grasp the object (**ungrasped**).

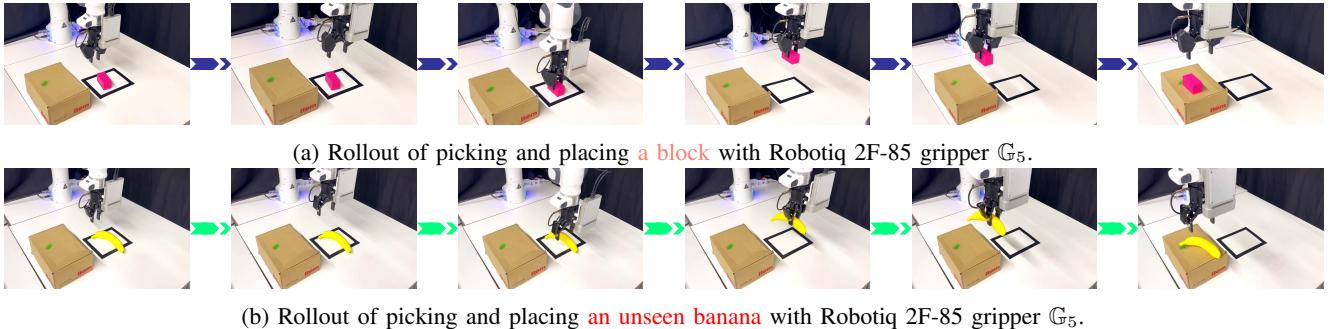


Fig. 7: The rollout of our method for solving the pick-and-place task with unseen gripper and objects.

instance, DP’s rigid trajectory generation fails to adjust to novel object shapes, often resulting in misaligned grasps (e.g., sliding off the curved surface of a banana).

When object positions are randomized, both DP and DP 3D collapse entirely (success rate is 0). Notably, our projection module plays a pivotal role in handling unseen geometries. Compared to the ablation without gripper-agnostic grasping knowledge “Ours w/o  $G_{\text{prob}}^*$ ”, our policy exhibits a better understanding of grasping unseen objects by leveraging gripper-agnostic grasping knowledge. Similarly, our policy without the Projection module (“Ours w/o Projection”) is still able to locate the grasping point. However, it fails for grippers that deviate more significantly from  $G_0$  (specifically,  $G_2$ ,  $G_4$ , and  $G_5$ ). These results highlight

the limitations of baseline methods in disentangling object geometry from positional priors. This capability is critical for real-world applications where object and experiment setup diversity are inherent.

**3) Safety Analysis:** We analyze how **gripper mapping** and **safety-constrained trajectory projection** ensure safe trajectory generation and successful task completion across different grippers. As shown in Fig. 6, for Gripper  $G_0$ , as mentioned in quantitative results, DP and DP 3D often result in collisions. After introducing the **safety-constrained trajectory projection** to DP and DP 3D on the inference phase, the generated motions are projected within the defined safety constraints, enabling both policies to successfully complete the task without collisions, as shown in Fig. 6a. When testing

on gripper  $\mathbb{G}_1$ , which is longer than gripper  $\mathbb{G}_0$  with TCP's shift +3.5cm, as shown in Fig. 6b, DP, DP 3D, and our policy without the projection module consistently lead to table collisions. In contrast, our policy ensures collision-free execution by correctly adjusting the motion to the extended gripper length. For the shorter gripper  $\mathbb{G}_4$ , our policy without **gripper mapping** fails to adapt the motion to the shorter gripper compared to  $\mathbb{G}_0$ , where TCP's shift -4cm, resulting in a grasping point that is too high above the object. However, our full policy successfully generates adaptive motions that account for the shorter gripper length, ensuring accurate and stable object grasping. Fig. 7 displays our method's rollout of pick-and-place tasks with different grippers.

## VI. CONCLUSIONS AND OUTLOOK

This paper proposes a diffusion-based policy for transferring pick-and-place knowledge across different grippers. This knowledge transition does not require retraining or fine-tuning the policy with the new gripper's configuration. Instead, it only needs to introduce the configuration in the policy inference phase and make the generated trajectories satisfy safety constraints, ensuring the successful completion of pick-and-place tasks. To enhance the flexibility of integrating different grippers into our method, the gripper's configurations can be described with free-form language instructions during the policy inference phase.

## REFERENCES

- [1] X. Yao, Z. Bing, G. Zhuang, K. Chen, H. Zhou, K. Huang, and A. Knoll, "Learning from symmetry: Meta-reinforcement learning with symmetrical behaviors and language instructions," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023, pp. 5574–5581.
- [2] Z. Bing, A. Koch, X. Yao, K. Huang, and A. Knoll, "Meta-reinforcement learning via language instructions," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 5985–5991.
- [3] Y. Meng, Z. Bing, X. Yao, K. Chen, K. Huang, Y. Gao, F. Sun, and A. Knoll, "Preserving and combining knowledge in robotic lifelong reinforcement learning," *Nature Machine Intelligence*, pp. 1–14, 2025.
- [4] H. Zhou, Z. Bing, X. Yao, X. Su, C. Yang, K. Huang, and A. Knoll, "Language-conditioned imitation learning with base skill priors under unstructured data," *IEEE Robotics and Automation Letters*, vol. 9, no. 11, pp. 9805–9812, 2024.
- [5] T. Migimatsu and J. Bohg, "Object-centric task and motion planning in dynamic environments," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 844–851, 2020.
- [6] C. Mitash, R. Shome, B. Wen, A. Boualiaris, and K. Bekris, "Task-driven perception and manipulation for constrained placement of unknown objects," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5605–5612, 2020.
- [7] A. Zeng, P. Florence, J. Tompson, S. Welker, J. Chien, M. Attarian, T. Armstrong, I. Krasin, D. Duong, V. Sindhwan, *et al.*, "Transporter networks: Rearranging the visual world for robotic manipulation," in *Conference on Robot Learning*. PMLR, 2021, pp. 726–747.
- [8] F. Di Felice, S. D'Avella, A. Remus, P. Tripicchio, and C. A. Avizzano, "One-shot imitation learning with graph neural networks for pick-and-place manipulation tasks," *IEEE Robotics and Automation Letters*, 2023.
- [9] J. Deng, H. Zhang, J. Hu, X. Zhang, and Y. Wang, "Class incremental robotic pick-and-place via incremental few-shot object detection," *IEEE Robotics and Automation Letters*, vol. 8, no. 9, 2023.
- [10] L. Wang, X. Chen, J. Zhao, and K. He, "Scaling proprioceptive-visual learning with heterogeneous pre-trained transformers," in *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.
- [11] T. Wang, D. Bhatt, X. Wang, and N. Atanasov, "Cross-embodiment robot manipulation skill transfer using latent space alignment," *arXiv preprint arXiv:2406.01968*, 2024.
- [12] C. Chi, Z. Xu, S. Feng, E. Cousineau, Y. Du, B. Burchfiel, R. Tedrake, and S. Song, "Diffusion policy: Visuomotor policy learning via action diffusion," *The International Journal of Robotics Research*, p. 02783649241273668, 2023.
- [13] 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.
- [14] K. Zakka, A. Zeng, P. Florence, J. Tompson, J. Bohg, and D. Dwibedi, "Xirl: Cross-embodiment inverse reinforcement learning," in *Conference on Robot Learning*. PMLR, 2022, pp. 537–546.
- [15] G. Tatiya, J. Francis, and J. Sinapov, "Transferring implicit knowledge of non-visual object properties across heterogeneous robot morphologies," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 11315–11321.
- [16] J. Carvalho, A. T. Le, M. Baierl, D. Koert, and J. Peters, "Motion planning diffusion: Learning and planning of robot motions with diffusion models," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023, pp. 1916–1923.
- [17] K. Mizuta and K. Leung, "Cobl-diffusion: Diffusion-based conditional robot planning in dynamic environments using control barrier and lyapunov functions," *arXiv preprint arXiv:2406.05309*, 2024.
- [18] M. Reuss, M. Li, X. Jia, and R. Lioutikov, "Goal conditioned imitation learning using score-based diffusion policies," in *Robotics: Science and Systems*, 2023.
- [19] Z. Liang, Y. Mu, M. Ding, F. Ni, M. Tomizuka, and P. Luo, "Adaptddiffuser: Diffusion models as adaptive self-evolving planners," in *International Conference on Machine Learning*. PMLR, 2023, pp. 20725–20745.
- [20] S. Venkatraman, S. Khaitan, R. T. Akella, J. Dolan, J. Schneider, and G. Berseth, "Reasoning with latent diffusion in offline reinforcement learning," in *The Twelfth International Conference on Learning Representations*, 2024.
- [21] J.-H. Basteck, W. Sun, and D. M. Kochmann, "Physics-informed diffusion models," *arXiv preprint arXiv:2403.14404*, 2024.
- [22] J. Ho and T. Salimans, "Classifier-free diffusion guidance," *arXiv preprint arXiv:2207.12598*, 2022.
- [23] Y. Yuan, J. Song, U. Iqbal, A. Vahdat, and J. Kautz, "Physdiff: Physics-guided human motion diffusion model," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 16010–16021.
- [24] J. K. Christopher, S. Baek, and F. Fioretto, "Constrained synthesis with projected diffusion models," in *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.
- [25] F. Mazé and F. Ahmed, "Diffusion models beat gans on topology optimization," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, no. 8, 2023, pp. 9108–9116.
- [26] G. Giannone, A. Srivastava, O. Winther, and F. Ahmed, "Aligning optimization trajectories with diffusion models for constrained design generation," *Advances in Neural Information Processing Systems*, vol. 36, pp. 51830–51861, 2023.
- [27] R. Römer, A. von Rohr, and A. P. Schoellig, "Diffusion predictive control with constraints," *arXiv preprint arXiv:2412.09342*, 2024.
- [28] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.
- [29] J. Liu, Z. Shen, Y. He, X. Zhang, R. Xu, H. Yu, and P. Cui, "Towards out-of-distribution generalization: A survey," *arXiv preprint arXiv:2108.13624*, 2023.
- [30] D. Morrison, P. Corke, and J. Leitner, "Learning robust, real-time, reactive robotic grasping," *The International journal of robotics research*, vol. 39, no. 2-3, pp. 183–201, 2020.
- [31] I. Lenz, H. Lee, and A. Saxena, "Deep learning for detecting robotic grasps," *The International Journal of Robotics Research*, vol. 34, no. 4-5, pp. 705–724, 2015.
- [32] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," in *International Conference on Learning Representations*, 2021.
- [33] X. Wang and Q. Xu, "Transferring grasping across grippers: Learning-optimization hybrid framework for generalized planar grasp generation," *IEEE Transactions on Robotics*, vol. 40, pp. 3388–3405, 2024.