

# Learning Cross-hand Policies of High-DOF Reaching and Grasping

Qijin She<sup>1</sup>, Shishun Zhang<sup>1</sup>, Yunfan Ye<sup>3</sup>, Ruizhen Hu<sup>2</sup>, and Kai Xu<sup>1\*</sup>

<sup>1</sup> National University of Defense Technology, China

<sup>2</sup> Shenzhen University, China

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

**Abstract.** Reaching-and-grasping is a fundamental skill for robotic manipulation, but existing methods usually train models on a specific gripper and cannot be reused on another gripper. In this paper, we propose a novel method that can learn a unified policy model that can be easily transferred to different dexterous grippers. Our method consists of two stages: a gripper-agnostic policy model that predicts the displacements of pre-defined key points on the gripper, and a gripper-specific adaptation model that translates these displacements into adjustments for controlling the grippers' joints. The gripper state and interactions with objects are captured at the finger level using robust geometric representations, integrated with a transformer-based network to address variations in gripper morphology and geometry. In the experiments, we evaluate our method on several dexterous grippers and diverse objects, and the result shows that our method significantly outperforms the baseline methods. Pioneering the transfer of grasp policies across dexterous grippers, our method effectively demonstrates its potential for learning generalizable and transferable manipulation skills for various robotic hands.

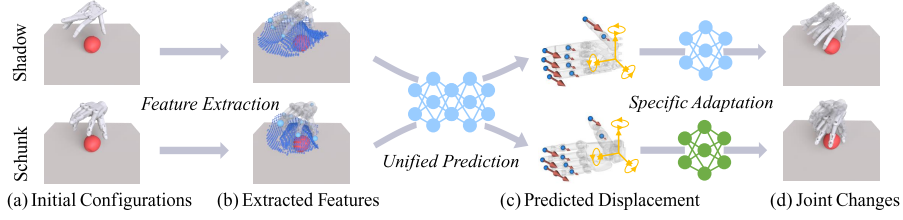
**Keywords:** Generalizable Dexterous Grasping · Policy Transfer

## 1 Introduction

Reaching-and-grasping problem is a fundamental and crucial challenge in robotics and computer graphics. Its core objective involves moving a gripper to approach an object and grasping it, where the prediction of the gripper's pose plays an important role. In this research field, traditional methods usually separate the synthesis of grasp poses from the planning of the reaching process, leading to additional time costs in replanning and a lack of robustness. To solve this problem, learning-based methods typically train models to predict the gripper's next pose frame by frame in real time, and have achieved significant success with both two-finger grippers [16] and dexterous grippers [26, 27]. However, the generalization capability of learning-based methods is limited. While many previous works have improved the models' generalization across objects [10, 25], generalization across grippers, particularly for dexterous grippers, remains largely unexplored.

---

\* Corresponding author. Email: kevin.kai.xu@gmail.com



**Fig. 1:** The overview of one step of our proposed framework. Given the context of the scene and the configuration of the gripper, our method initially extracts gripper-agnostic features. These features are uniformly sent to the policy model to predict gripper-agnostic point displacements, which are forwarded to the adaptation models of various grippers for precise gripper control.

Enabling the policy model to transfer across grippers could eliminate the need for expensive data collection and the time-consuming process of training individual models for each gripper.

Prior research has aimed to create a broad grasp generalization across various grippers [22, 23, 40, 47]. However, these efforts mainly revolve around the synthesis of final grasp poses, which implies that the target pose cannot be adjusted during execution. In this study, we aim to develop a universal High-DOF grasping policy that can be trained on a single gripper and subsequently transferred to other grippers with minimal effort. Our principal hypothesis posits that there are commonalities in grasping skills among grippers despite of their difference in geometries and morphologies. It is the representation, serving as the input and the output of the policy that restricts the generalization ability of grasping skills across grippers. Consequently, we propose that the main challenge lies in identifying a gripper-agnostic geometric representation that can mitigate the influence of the two distinct factors: (i) Gripper morphology: the policy model need understand the gripper’s state and predict state changes, which are typically represented in joint space. Due to distinct morphologies, the joint spaces of different grippers can vary significantly; (ii) Gripper geometry: the policy model should also receive spatial information about the gripper and the scene, which are typically represented as images or point clouds, to comprehend their spatial relationship. However, these representations may lead the policy model to overfit on gripper’s geometry, potentially impacting the model’s ability to generalize.

Inspired by animation systems such as IK Rig [9] where animators manipulate key points on characters to alter their poses, we advocate the use of key points on and shared among dexterous grippers (e.g. fingertip). We use these key point positions as the state and their displacements as the action, to overcome the morphological differences of the grippers. Thus, the policy can be regarded as “dragging” these key points to adjust the gripper’s poses. This pose adjustment can be accomplished through an adaptation model that converts key point displacements into gripper joint changes, simultaneously avoiding self-collisions.

Furthermore, to reduce the interference of gripper geometries on the policy’s generalization, we introduce the Interaction Bisector Surface (IBS) [49] to characterize the spatial interaction between the gripper and the object. The IBS, computed as the Voronoi diagram between two geometric objects, has proved successful in enhancing the policy’s robustness to object geometries [41] in the grasping task. We believe that IBS should have similar robustness to gripper geometries. Through our experiments, we discovered that incorporating the IBS as an additional state representation can significantly assist policies in transferring between different grippers. To facilitate the policy’s transfer to dexterous grippers with different fingers, we implement the policy network based on the transformer network [44] for utilizing the attention mechanism to integrate information from these two representations and extract relations among fingers.

In the experimental section, we undertake an extensive series of performance verification and ablation studies. The findings indicate that, when compared to benchmark methods, our approach consistently produces grasps with higher quality for diverse objects. Moreover, our method can adapt to different grippers with minimal performance degradation, showcasing its generalization. The main contributions of this work can be summarized as:

- We formulate gripper-agnostic state and action representations for the grasping policy, enabling seamless transferability across different grippers;
- We introduce an innovative policy network designed to improve the extraction of relation among fingers and representations.
- We put forward a two-stage hierarchical framework that separates the prediction of unified grasps from the control of specific grippers.

## 2 Related Work

### 2.1 Static Generalizable Grasp Synthesis

Grasp synthesis has been extensively analyzed from various angles. Existing grasp synthesis techniques can be broadly categorized into analytical and data-driven methods. Analytical approaches employ sampling methods [28] or optimization techniques [43] to search for gripper poses that ensure physical stability. By optimizing grasps for each object and gripper on a case-by-case basis, such methods naturally generalize across different grippers but run slowly.

Data-driven methods train models that can directly predict grasp poses based on object features. To make these methods generalizable for different grippers, previous studies have proposed numerous cross-gripper grasp representations for prediction. Unigrasp [40] and EfficientGrasp [22] predict contact points on objects for grasping, and then employ inverse kinematics and reinforcement learning, respectively, to obtain grasps from contact points for a specific gripper. Following a distinct approach, AdaGrasp [47] focuses on providing the pre-grasp poses for grasping and then achieves stable grasps by closing the grippers. GenDexGrasp [23] generates the contact map of objects, subsequently minimizing the discrepancy between the actual contact map and the predicted ones, to

obtain grasps for a specific gripper with a similar approach used in [3]. Neural-Grasp [19] takes a unique route by learning an implicit field defined by distances to the gripper and to the object. It generates grasps by retrieving its nearest neighbors in the latent space from a grasp database.

Nevertheless, the representations employed in the aforementioned works are specifically crafted to depict the final grasp poses and may not be well-suited for capturing the entirety of the reaching-and-grasping process.

## 2.2 Kinematic Motion Retargeting

Motion retargeting refers to the process of transferring motions from one embodied entity to another. There are primarily two types of motion retargeting techniques. We summarized as learning-based and heuristic methods. Learning-based methods frame motion retargeting as a sequence-to-sequence generation problem [1, 2, 45]. Heuristic methods employ correspondences such as matched joints [30] or key points [21] between two robots to calculate the configurations of the target robot. Within the realm of robotic grasping, some methods have achieved real-time motion retargeting [13, 36, 42], facilitating the transfer of motions from a human hand to a dexterous gripper. These methods are used for teleoperation and robot trajectory collection.

Indeed, while motions generated by motion retargeting methods may visually resemble the original motions, replicating them in a dynamic environment becomes challenging due to unforeseen errors. Hence, the retargeted motions are often employed primarily as additional data [34, 35] or constraints [27, 31, 32] to train robust policies.

## 2.3 Dynamic Policy Transfer

In the pursuit of generating stable motions for new robots without the necessity for retraining, some studies have transferred policies directly to new robots in locomotion tasks, instead of transferring motions. The core concept underlying these studies is the integration of the agent’s morphological information into the policy learning process. This morphological information can be encoded as a latent embedding upon which the policy conditions [7, 39]. Given that robot morphologies are often depicted as graphs, the utilization of graph-structured neural networks [38] presents an effective approach to explicitly encode the connection of different components into the policy [15, 29, 46]. Furthermore, transformer networks [44] have also provided additional options to implicitly extract relationships among various components [11, 14, 20].

However, these locomotion works are mainly designed to achieve generalization on robot morphologies without considering the impact of robot appearance and geometries, which cannot be ignored in contact-rich tasks such as grasping task. Some studies in manipulation utilize wrist-mounted cameras [48] or image inpainting [6] to mitigate the impact of the robot’s appearance on policy generalization. Nonetheless, their approaches rely on the assumption of two-finger grippers and present difficulties in adapting to the dexterous grasping task.

### 3 Methods

Our approach is predicated on a typical grasping scenario where an object is situated on a table and a gripper is assigned the task of reaching and grasping it under the guidance of the policy. Given the point clouds of the scene segmented into the foreground object and background, as well as the configurations of the gripper, the model is expected to output joint changes of the gripper. These changes are used to control the gripper’s motion through PD controllers. The model also has to discern the task’s completion when high-quality grasps form. The environment is constructed based on PyBullet [8].

Figure 1 illustrates the process by which our method controls various grippers in a single step. Our approach features a two-stage hierarchical model framework, distinguishing the prediction of high-level grasp motions from the specific gripper control. The framework consists of a unified policy model and a gripper-specific adaptation model. The policy model, which is shared among different grippers, ingests the gripper-object interaction and key points. The model predicts the key point displacements, subsequently translated into gripper configuration changes through the gripper-specific adaptation model. These three processes will be elaborated upon in Sections 3.1, 3.2 and 3.3.

#### 3.1 Gripper-agnostic Feature Extraction

The crux of realizing a generalized grasp policy lies in identifying a gripper-agnostic representation that encapsulates common gripper traits while retaining grasp-related task information. We combine the semantic key points and the interaction bisector surface (IBS) [41] which is the set of points equidistant to the scene and the gripper, to represent our policy model’s state.

Semantic key points establish a natural correspondence across different hands. We select the fingertip point  $p_k^1$  and middle phalanx point  $p_k^0$  on each finger  $k$ , as well as the gripper’s root  $p_0$  on the palm. The middle phalanx points offer bending information of fingers, facilitating precise finger control. Semantic key point coordinates are calculated via forward kinematics and defined in the gripper’s local coordinate system with its root to be the origin. We incorporate an additional 3-D vector  $r$  to represent the gripper’s rotation relative to the world coordinate system, defined as the initial frame of the object pose. Thus, the full semantic key point input  $s_{key} = [r, p_0, p_1^0, p_1^1, \dots, p_K^0, p_K^1]$  comprises  $6(K + 1)$  dimensions where  $K$  is the number of fingers.

The Interaction Bisector Surface (IBS) captures the spatial interaction between the gripper and the object. Computing the exact IBS is computationally intensive due to the need for Voronoi diagram extraction. To maintain computational efficiency, we implement the approach proposed by [41] to compute sampled IBS points  $\{p_{ibs}\}$  near the gripper. We define a sphere centered at the palm with a radius  $r = 18cm$  and voxelize its bounding box into  $v^3$  cells, where  $v = 20$ . For each cell, we compute the distance from its center to both the gripper and the scene, and consider its center as IBS points if the difference between these distances is below a threshold. To enhance efficiency and precision, cells

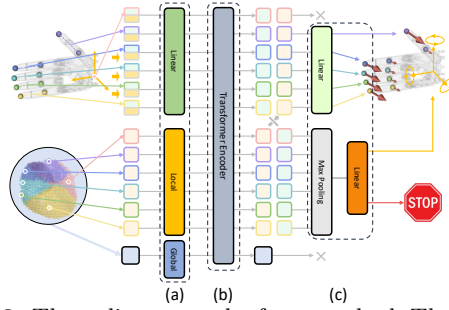
distant from the potential IBS are excluded from the computation, and the positions of sampled IBS points are refined for closer proximity to the exact IBS surface. An IBS illustration is available in Figure 1 (b).

The sampled IBS points undergo downsampling to a fixed number  $n = 4096$  for network input. It’s noteworthy that the origin of these points is established at the gripper’s palm center. Each IBS point  $p_b$ , with its closest points on the scene and the gripper denoted as  $p_s$  and  $p_g$ , respectively, is associated with the following features: (i) coordinate  $c = (x, y, z) \in R^3$ ; (ii) distance to the scene  $d_s \in R$ ; (iii) distance to the gripper  $d_g \in R$ ; (iv) indicator of whether  $p_s$  is located on the foreground object  $b_s \in \{0, 1\}$ ; (v) one-hot indicator of the gripper component that  $p_g$  belongs to  $c_g \in \{0, 1\}^{k+1}$ ; (vi) indicator of which side of the gripper  $p_g$  is located on  $a_g \in [-1, 1]$ . Here,  $a_g = n_g \cdot d_{up}$  is the dot product of the normal direction  $n_g$  of point  $p_g$  on gripper in rest pose and the upright direction  $d_{up}$  perpendicular to the palm and pointing outwards.

### 3.2 Unified Policy Model

The policy action is also configured to be gripper-agnostic. The policy model ingests IBS points  $\{p_b\}$  and semantic key points  $s_{key}$ , predicting the displacements  $\Delta p_k^i$  of each point  $p_k$ , the global translation change  $\Delta p$ , the global rotation change  $\Delta r$ , as well as a special stop value  $a_s$  used to determine task termination in cases of more than two contacts between the object and fingers.

We introduce a novel transformer-based network for extracting the relationship between fingers, semantic key points, and IBS. Figure 2 provides an overview of our policy model structure. The model comprises three components: (i) Finger-wise encoders for semantic points and IBS that separately translate them into a latent feature space; (ii) A transformer module that executes attention over semantic key points and IBS features; (iii) Finger-wise and global feature aggregation for finger-wise actions and global action.



**Fig. 2:** The policy network of our method. The network includes three components : (a) finger-wise feature encoder; (b) a transform encoder fusing information among fingers and representations; (c) finger-wise and global.

The hand can be segmented into  $K + 1$  components. We categorize semantic points and IBS points according to their corresponding component and encode them separately using MLP and PointNet [33]. For each semantic key point group, we append the global rotation, resulting in a  $9D$  vector.

Each IBS group is processed with a local PointNet network to obtain local features. Additionally, a global PointNet computes the global features of the entire IBS. The tokens undergo a transformer-based encoder incorporating  $N$  self-attention layers. This attention module plays a crucial role in fostering spatial reasoning between key points and IBS defined in different coordinate systems. It effectively integrates information from other components for each hand component. This enables individual finger action prediction, enhancing model generalizability for grippers with varying finger numbers.

The procedure involves concatenating the semantic key point embedding with the IBS embedding for each finger. Subsequently, this combined vector is projected into an output that signifies the desired semantic point displacements of the fingers within the gripper’s local coordinate system. The palm point’s embedding vectors don’t predict displacements but are passed into two separate MaxPooling layers along with other components’ embeddings to predict actions impacting all fingers, i.e., the global translation and rotation change of the gripper in the world coordinate system, as well as the stop value.

### 3.3 Specific Adaptation Model

Upon determining desired key point displacements, we need to translate them into specific joint changes for gripper control, while avoiding finger self-collision. Though this conversion can be solved with optimization methods, simultaneous optimization of these two objectives is time-consuming and unsuitable for real-time gripper control in our task. Hence, we employ a lightweight neural network, the adaptation model, to directly predict the gripper’s joint changes.

The adaptation model takes the current gripper configuration, current gripper semantic points, and desired semantic key point displacements  $\{\Delta p_k^i\}$  as inputs. It then produces the corresponding joint changes that effectuate the semantic key point displacements. The input semantic key point positions  $s'_{key} = s_{key} \setminus \{r\}$  exclude the global rotation. The gripper configuration input  $s_j = \{j\} \in R^C$  is defined as the actuated joint angles of the gripper where  $C$  is the gripper’s degrees of freedom. The model outputs the changes in gripper’s joints  $\{\Delta j\} \in R^C$ . The adaptation network consists of only several MLP layers and can be efficiently trained.

### 3.4 Network Training

The model training in our work comprises two stages: joint training and transfer training. Joint training aims to develop a generalizable policy model. During this stage, the policy model and the adaptation model are trained simultaneously, each with its independent loss function and no gradients are exchanged between them. Contrastingly, transfer training is intended to adapt the pre-trained policy model to a new gripper. During this stage, the policy model remains fixed while a new specific adaptation model is trained from scratch. We conduct 800k updates for both the policy and adaptation models during joint training, and 50k updates for the new adaptation model during the transfer process.

The Unified Policy Model is trained in a reinforcement learning manner. We adopt the reward function and training strategy provided by [41]. Our reward function comprises two components: (i) a task reward for successful and stable final grasping; (ii) a reaching reward for preventing collision between the gripper and the scene. We adopt the Soft Actor-Critic [12] algorithm to train our policy. Additional data is used to accelerate the efficiency, which is generated by the interpolation between static grasps and their pre-grasp poses and between the pre-grasp poses and the gripper’s initial poses. We use grasps from the work [24].

We train the Specific Adaptation Model with a self-supervised cycle loss. Given the predicted semantic key point displacements  $\Delta p_k^i$ , current joint angles  $\{j\}$ , and current semantic key point positions  $\{p_k^i\}$ , the Specific Adaptation Model outputs desired joint changes  $\{\Delta j\}$ . we utilize a differentiable forward kinematic function to compute the expected semantic key point positions after joint movement. and construct MSE Loss based on the difference between predicted key points and expected key points.  $\theta$  is the parameters of the network

$$L_{point}(\theta) = \frac{1}{2} \sum_{k=1}^K \sum_{i=0}^1 (e_k^i - p_k^i - \Delta p_k^i)^2, e_k^i = FK_k^i(j + \Delta j) \quad (1)$$

where  $FK$  is a differentiable forward kinematic function. To avoid self-collision, i.e. the collision among different gripper components, we incorporate a self-collision loss adapted from [24] to the original loss. This loss is written as:

$$L(\theta) = L_{point}(\theta) + \omega L_{self}(\theta), \quad (2)$$

where  $\omega = 1$  is the parameter to balance different loss terms. The self-collision loss  $L_{self}$  is defined as:

$$L_{self}(\theta) = \sum_{m=1}^L \sum_{n=1}^N \max(D(p_n(j + \Delta j)), H_m(j + \Delta j)), 0), \quad (3)$$

where  $L$  is the number of gripper links,  $N$  is the number of points  $p_j(j + \Delta j)$  sampled from each link,  $H_i(j + \Delta j)$  is the convex hull of each link, and  $D$  is the signed distance from a point to a convex hull. More details about the self-collision loss  $L_{self}$  can be found in this work [24].

## 4 Experiments

### 4.1 Experiment Setting

We adapt several dexterous grippers for training and testing in our environment. These grippers include Shadow Dexterous Hand, Schunk SVH Hand, Rutgers Hand, Allegro Hand, and actuated virtual human hand adapted from Mano [37]. Figure 3 displays these hands in their rest poses. We select graspable objects from [24] which consists of 500 objects collected from four object datasets. We use the objects from KIT Dataset [18] and GD Dataset [17] as training objects,

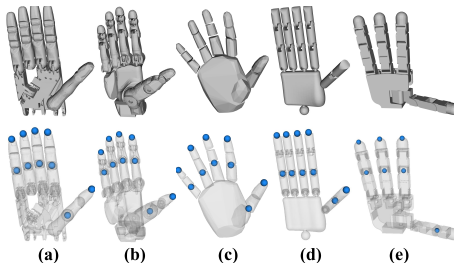
	Features enabled					Shadow(Origin)		Schunk		Mano		Rutgers		Allegro	
	UNI	OCM	GCM	IBS	TR	$SR$	$Q_1$	$SR$	$Q_1$	$SR$	$Q_1$	$SR$	$Q_1$	$SR$	$Q_1$
Single				✓		<b>72.2%</b>	<b>0.177</b>	-	-	-	-	-	-	-	-
UNI+IBS	✓			✓		68.0%	0.168	54.6 %	0.166	61.2%	<b>0.169</b>	42.6%	0.143	-	-
UNI+OCM	✓	✓				50.1%	0.152	41.2%	0.164	45.5%	0.166	38.9%	0.145	-	-
UNI+GCM	✓		✓			64.0%	0.159	45.5%	0.148	50.4%	0.138	41.4%	0.132	-	-
Ours	✓			✓	✓	71.3%	0.169	<b>65.3 %</b>	<b>0.167</b>	<b>65.2%</b>	0.167	<b>54.8%</b>	<b>0.147</b>	<b>55.0%</b>	<b>0.133</b>

**Table 1:** Ablation Results of our method and its degraded variants using our YCB object dataset. “-” means the method cannot be adapted to the gripper due to its limitations.

and objects from YCB Dataset [5] as test objects. We further expand our test set with objects from ContactPose Dataset [4]. We remove the object with volume  $v > 1.5dm^3$  and reduce similar objects in the dataset. In our experiment, we used Shadow Hand for policy training and other hands for adaptation tests. To generate the initial poses of grippers in the episodes, for each object, we use its center to create a sphere with a radius  $r = 20cm$ . We sampled points on the upper hemisphere as the origin of the local coordinate system of the gripper and rotated the gripper to make its palm face the object center and its thumb point upwards. In the test process, we set a fixed set of initial poses for each object.

We utilize two metrics to assess policy performance. The first is the success rate of the final grasp. Our success criteria are adapted from the AdaGraps [47]. A grasp is deemed successful if, upon moving the gripper upwards by  $0.6m$ , the object’s center of mass ascends at least  $0.2m$ . Besides, we incorporate the generalized  $Q_1$  [24] for an additional grasp stability evaluation. We compute its average success rate (denoted “ $SR$ ”) among all objects and initial positions and the average generalized  $Q_1$  values (denoted “ $Q_1$ ”) of all successful grasps.

To validate the benefits of using IBS in representation for policy adaptation, we implement two types of contact map representation inspired by [3, 23] as alternatives. The object contact map representation (OCM) is the feature points sampled from the object’s surface. The point features include its coordinate with the object center as its origin, its distance to the gripper, the indicator of whether it is located in the foreground, and the one-hot indicator of the gripper component to which its closest point on the gripper belongs. The gripper contact map representation (GCM) is the feature points sampled from the surface of the gripper. It has features similar to the “OCM”. The point features include its coordinates with the gripper root as the origin, its distance to the scene, an



**Fig. 3:** The hands and their semantic key points. (a) Shadow; (b) Schunk; (c) Mano; (d) Rutgers; (e) Allegro.

indicator of whether its nearest point in the scene is located on the foreground object, and a one-hot indicator of the gripper component to which it belongs. These alternatives maintain the same point number as IBS.

#### 4.2 Ablation Studies on the Policy Model

We evaluate models under diverse configurations using our YCB object dataset. Table 1 presents the results of these variants and our method. In our ablation, we assess the impact of the following five features on the model’s performance. “Single” means an end-to-end model with state and action defined in the gripper’s joint space. “UNI” refers to using our unified policy model and hierarchical framework. “OCM” and “GCM” refer to the use of contact map representation or gripper map representation, respectively. “IBS” stands for the use of the IBS representation. “Ours” means using our transformer-based policy model instead of a naive policy model that directly concatenates features.

**Effectiveness of two-stage hierarchical framework** In our hierarchical framework, the unified policy model “Uni+IBS” demonstrates its capability to control various grippers, achieving comparable performance on these hands to the end-to-end specific model “Single”. This is achieved despite its performance on the original hand being marginally lower.

**Importance of IBS** The “Uni+OCM” model, which uses feature points sampled from the object surface, underperforms in the original hand due to overfitting observed on training objects. Despite this, it exhibits commendable transfer performance across new grippers, albeit with a slight performance drop compared to the original hand. On the contrary, the “Uni+GCM” model, which employs feature points from the gripper surface as state, performs well on the original hand but struggles significantly on new hands. This indicates that its generalizability is limited by the original grippers’ geometry. The “Uni+IBS” model, using an IBS representation that balances object and gripper geometry, efficiently mitigates the limitations of the preceding models. It outperforms both “Uni+OCM” and “Uni+GCM” across all hands, suggesting that IBS is a viable representation for generalizable grasps encompassing various objects and grippers.

**Importance of transformer-based policy** By replacing the simple feature concatenation model with our transformer-based model, the learned policy, “Ours”, not only surpasses “Uni+IBS” in transfer performance across all hands but also successfully adapts to the four-fingered Allegro Hand. This indicates that our policy design not only capitalizes on the input information more effectively but also boasts a flexible structure that accommodates morphological variations in the number of fingers.

#### 4.3 Ablation Studies on the Adaptation Module

We perform an additional experiment that justifies our learning-based adaptation with two variants of optimization-based adaptation methods, affirming

	Time Cost Per Frame (ms)				Grasp Performance		Collision	
	Feature Extraction	Unified Prediction	Adaptation	Total	$SR$	$Q_1$	Collision Percentage	Collision Loss
OB-IK	53.3	6.0	6.0	65.3	61.3%	0.146	2.0%	0.65
OB-IK+SC	90.3	6.6	182.7	279.6	64.0%	0.151	0.0%	0.0
LB-IK+SC	53.0	6.0	0.4	59.4	62.0%	0.170	0.2%	0.09

**Table 2:** The computational efficiency and performance of our unified policy under different adaptation approaches using our YCB object dataset. The experiment is conducted on the Schuck Hand using a subset of initial poses.

our design choice. The optimization method is derived from [34]. “Optimization-based Adaptation for Inverse Kinematics only (OB-IK)”: this method employs an optimization-based approach to translate predicted key point displacements into joint changes. “Optimization-based Adaptation for Inverse Kinematics and self-collision avoidance (OB-IK+SC)”: the optimizer settings mirror OB-IK, with the addition of a self-collision loss to the optimization function. “Learning-based Adaptation for IK and self-collision avoidance (LB-IK+SC)”: our approach leverages a neural network to directly predict joint changes, incorporating inverse kinematics and self-collision as loss during the training phase.

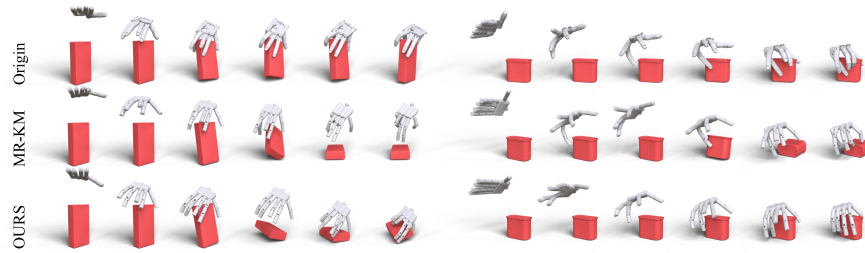
We evaluate the policy with these modules on the Schunk Hand and our YCB object set in a subset of test initial episodes as shown in Table 2. Evaluation metrics comprise the average consumption of each module in our framework,  $SR$ ,  $Q_1$  as well as the percentage of collision frames throughout the process and the average self-collision value when such an event occurs, to illustrate the impact of self-collision avoidance loss. The experimental results reveal that when self-collision is not accounted for, the optimization-based adaptation approach can satisfy real-time requirements, albeit with potential finger collisions. Including self-collision in the optimization function significantly escalates the algorithm’s time cost, despite completely averting collision, and this results in a running time of approximately 280ms per frame, which is difficult to meet the requirements of real-time dynamic grasping tasks. Moreover, due to limited overall resources, the increased overhead of the optimization algorithm encroaches upon the computing resources of IBS, consequently slowing down IBS computation. In contrast, the learning-based method not only achieves the lowest time overhead but also prevents self-collision during the grasping process.

#### 4.4 Comparison to Baselines

We have implemented four baseline methods for comparison. These baselines transfer the policy trained on the specific joint space of the Shadow Hand using different strategies. Motion Retargeting (MR) involves the offline generation of motion trajectories using a policy, which are then transferred to another hand. The converted trajectories are then replayed in the simulator with PD controllers. Policy Transfer (PT) is an online process where the configuration of a new hand is mapped to the training hand, and the policy is invoked to predict the subsequent configuration for the training hand, which is then mapped back

Method	Schunk		Mano		Rutgers	
	$SR$	$Q_1$	$SR$	$Q_1$	$SR$	$Q_1$
MR-JM	21.8%	0.086	6.4%	0.091	10.9%	0.069
MR-KM	35.9%	0.136	33.8%	0.124	33.7%	0.126
PT-JM	40.1%	0.139	48.9 %	0.151	26.9%	0.123
PT-KM	35.6%	0.132	30.9 %	0.132	31.6%	0.141
OURS	<b>61.8%</b>	<b>0.167</b>	<b>63.4 %</b>	<b>0.165</b>	<b>52.5%</b>	<b>0.143</b>

**Table 3:** Comparison results with baseline methods using both the YCB and Contact-Pose object datasets.



**Fig. 4:** The visual comparisons to motion retargeting method. Our method can adapt to the pose change of the object.

to the new hand to control it. This hand configuration mapping can be executed using either keypoint matching (KM) or joint matching (JM). Combining different generation and mapping methods, we have developed four baseline methods.

Table 3 presents results on two object sets and three test grippers, demonstrating that our method significantly outperforms all the baseline approaches. Notably, “MR-KM” surpasses “MR-JM” in both success rate and  $Q_1$ , which indicates that matching key points can aid in finding a more precise correspondence between the poses of two grippers. The online conversion process of “PT-JM” allows the policy to re-plan the pose, even when the conversion lacks accuracy. Consequently, it performs better than its offline version “MR-JM”, and even exceeds “MR-KM” on Schunk Hand and Mano Hand. Figure 4 showcases example sequences of “MR-KM” and “OURS”, highlighting the advantages of replanning. However, the expected performance improvement from “MR-KM” to “PT-KM” does not materialize. We attribute this primarily to the morphologies’ differences of the grippers. The slight displacements of key points between frames on the source gripper do not trigger corresponding displacements of the key points on the target gripper during optimization. This leads to the target gripper frequently sticking at a specific pose. In our method, we forecast the desired displacements of the key points on the target hand directly. This enhances the control precision over each key point, thereby facilitating the policy’s adaptability across a variety of grippers.

Method	Shadow		Schunk		Mano		Rutgers		Allegro	
	$SR$	$Q_1$	$SR$	$Q_1$	$SR$	$Q_1$	$SR$	$Q_1$	$SR$	$Q_1$
Complete	<b>71.3%</b>	0.169	<b>65.3%</b>	0.167	<b>65.2%</b>	0.167	54.8%	0.147	<b>55.0%</b>	0.133
Diff Init	66.8%	<b>0.189</b>	59.2%	0.203	59.9%	<b>0.196</b>	54.8%	<b>0.167</b>	47.1%	<b>0.159</b>
Fix Init	67.3%	0.182	58.6%	<b>0.207</b>	62.4%	0.184	<b>57.1%</b>	0.166	46.2%	0.155

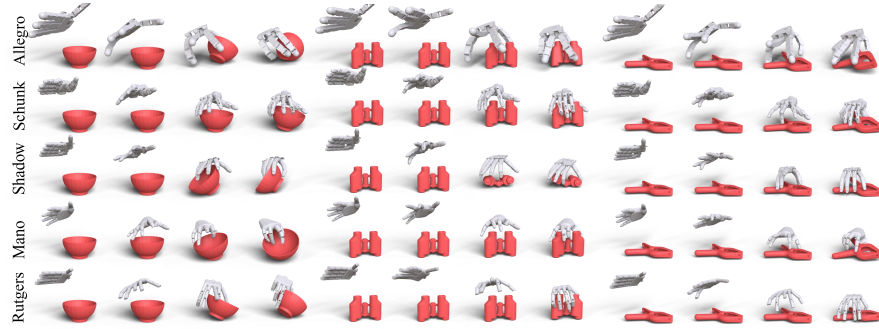
**Table 4:** Results of the original setting and two partial observation test settings on the YCB object dataset.

#### 4.5 Verification of the Partial Observation

To verify whether our method can work in the more realistic partial observation setting that is closer to the real-world scenarios, we take the incomplete point clouds captured by a depth camera as the raw input. The camera is positioned in a fixed location under the world coordinate system, gazing upon the object from an oblique upper angle. We set up two test settings with different initialization strategies for gripper poses: In “Different gripper initialization (Diff Init)”, the gripper starts from the initial pose around the object. The set of initial positions is equivalent to the test set of initial positions mentioned in Section 4.1. In “Fixed gripper initialization (Fix Init)”, the initial poses of the hand is fixed on the side of the camera, and the palm surface is parallel to the optical axis of the camera.

We test our method on these two settings as well as the original setting where the point cloud is complete. We compute the average success rate and the average generalized  $Q_1$  of all successful grasps on the YCB object set for comparison. The results are shown in Table 4. We find our method can adapt to partial observation without much performance drop in success rates. There are two reasons why our method generalizes well to the partial observation. First, we use IBS as a dynamic state representation. The shape of IBS remains robust even when the point cloud of the gripper and object is incomplete. For example, when the gripper bends to grasp the object and the camera can only partially capture the object’s point cloud through the gaps between fingers, the complete interaction surface can still be computed using this incomplete object point cloud. Secondly, our method involves learning a policy model. Even if partial observations lead to incorrect movements of the fingers, the model can adjust the gripper’s behavior to rectify the error based on the new observation. The visual example can be found in our supplementary video.

When comparing the performances of different dexterous hands, we observe a more significant performance drop in the Allegro Hand. This is due to its larger volume, leading to more severe occlusion between the hand and the object compared to other hands. In contrast, the Rutgers Hand, which has a relatively smaller size, experiences the smallest performance loss when provided with partial observations as input. Additionally, we observed an interesting result. In cases of incomplete point clouds, the average  $Q_1$  of successful grasps using our method shows a noticeable increase. This can be attributed to the task termination requirement in our environment, which mandates the finger number contacted with



**Fig. 5:** The visual results of our method on different grippers (up to down) and objects (left to right). For each case, we show the initial configurations of grippers and two sampled frames during the reaching process with the final grasping.

the objects. Consequently, in scenarios with incomplete point clouds, fulfilling the task termination require the contacts captured by the camera. It is possible that other fingers make contact with parts of the object not visible in the captured point cloud. Since the generalized Q1 value is closely linked to the number of contacts between the fingers and the object, grasps under partial observations result in a higher Q1 value.

#### 4.6 Visualization of Grasping Processes

Figure 5 illustrates the visual results of our method, showcasing the control of grippers to grasp various objects with smooth motions. The complete reaching processes and additional examples can be found in our supplementary video.

### 5 Conclusion

To the best of our knowledge, our method represents the first attempt to transfer grasp policies across various dexterous grippers. Our approach adopts a two-stage hierarchical model framework, which separates the prediction of grasp key points' motions from the control of specific grippers. This, along with the use of a unified gripper-agnostic state and action designs, as well as a novel policy network, enables our method to be easily adaptable to various dexterous grippers while achieving exceptional performance.

However, our method still has some limitations. First, our method cannot transfer the policy trained on an anthropomorphic hand (e.g. Shadow Hand) to non-anthropomorphic hands (e.g. Robotiq 3F) because of large differences in topology. We believe learning a more flexible correspondence across grippers would make our method more generalizable, and would like to leave this for future exploration. Moreover, in our experiments, we standardized the physical parameters for various dexterous hands, such as mass and joint friction. Adapting the algorithm to account for diverse physical parameters across different hands is essential for the successful implementation of our method in real-world settings.

## Acknowledgements

We thank the anonymous reviewers for their valuable comments. We also thanks Min Liu for helpful discussions. This work was supported in by the NSFC (62325211, 62132021, 62322207), the Major Program of Xiangjiang Laboratory (23XJ01009), Guangdong Natural Science Foundation (2021B1515020085), and Shenzhen Science and Technology Program (RCYX20210609103121030).

## References

1. Aberman, K., Li, P., Lischinski, D., Sorkine-Hornung, O., Cohen-Or, D., Chen, B.: Skeleton-aware networks for deep motion retargeting. *ACM Transactions on Graphics (TOG)* **39**(4), 62–1 (2020)
2. Aberman, K., Wu, R., Lischinski, D., Chen, B., Cohen-Or, D.: Learning character-agnostic motion for motion retargeting in 2d. *ACM Transactions on Graphics (TOG)* **38**(4), 1–14 (2019)
3. Brahmbhatt, S., Handa, A., Hays, J., Fox, D.: Contactgrasp: Functional multi-finger grasp synthesis from contact. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 2386–2393. IEEE (2019)
4. Brahmbhatt, S., Tang, C., Twigg, C.D., Kemp, C.C., Hays, J.: Contactpose: A dataset of grasps with object contact and hand pose. In: *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII* 16. pp. 361–378. Springer (2020)
5. Calli, B., Singh, A., Bruce, J., Walsman, A., Konolige, K., Srinivasa, S., Abbeel, P., Dollar, A.M.: Yale-cmu-berkeley dataset for robotic manipulation research. *The International Journal of Robotics Research* **36**(3), 261–268 (2017)
6. Chen, L.Y., Hari, K., Dharmarajan, K., Xu, C., Vuong, Q., Goldberg, K.: Mirage: Cross-embodiment zero-shot policy transfer with cross-painting. *arXiv preprint arXiv:2402.19249* (2024)
7. Chen, T., Murali, A., Gupta, A.: Hardware conditioned policies for multi-robot transfer learning. *Advances in Neural Information Processing Systems* **31** (2018)
8. Coumans, E., Bai, Y.: Pybullet, a python module for physics simulation for games, robotics and machine learning. <http://pybullet.org> (2016–2021)
9. Engine, U.: Ik rig in unreal engine. Tech. rep. (2021), <https://docs.unrealengine.com/5.0/en-US/ik-rig-in-unreal-engine/>, accessed: June 8, 2023
10. Fang, H.S., Wang, C., Fang, H., Gou, M., Liu, J., Yan, H., Liu, W., Xie, Y., Lu, C.: Anygrasp: Robust and efficient grasp perception in spatial and temporal domains. *IEEE Transactions on Robotics* (2023)
11. Gupta, A., Fan, L., Ganguli, S., Fei-Fei, L.: Metamorph: Learning universal controllers with transformers. In: *International Conference on Learning Representations* (2022)
12. Haarnoja, T., Zhou, A., Abbeel, P., Levine, S.: Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In: *International conference on machine learning*. pp. 1861–1870. PMLR (2018)
13. Handa, A., Van Wyk, K., Yang, W., Liang, J., Chao, Y.W., Wan, Q., Birchfield, S., Ratliff, N., Fox, D.: Dexpilot: Vision-based teleoperation of dexterous robotic hand-arm system. In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. pp. 9164–9170. IEEE (2020)

14. Hong, S., Yoon, D., Kim, K.E.: Structure-aware transformer policy for inhomogeneous multi-task reinforcement learning. In: International Conference on Learning Representations (2021)
15. Huang, W., Mordatch, I., Pathak, D.: One policy to control them all: Shared modular policies for agent-agnostic control. In: International Conference on Machine Learning. pp. 4455–4464. PMLR (2020)
16. Kalashnikov, D., Irpan, A., Pastor, P., Ibarz, J., Herzog, A., Jang, E., Quillen, D., Holly, E., Kalakrishnan, M., Vanhoucke, V., et al.: Qt-opt: Scalable deep reinforcement learning for vision-based robotic manipulation. *Robotics: Science and Systems (RSS)* (2018)
17. Kappler, D., Bohg, J., Schaal, S.: Leveraging big data for grasp planning. In: 2015 IEEE International Conference on Robotics and Automation (ICRA). pp. 4304–4311. IEEE (2015)
18. Kasper, A., Xue, Z., Dillmann, R.: The kit object models database: An object model database for object recognition, localization and manipulation in service robotics. *The International Journal of Robotics Research* **31**(8), 927–934 (2012)
19. Khargonkar, N., Song, N., Xu, Z., Prabhakaran, B., Xiang, Y.: Neuralgrasps: Learning implicit representations for grasps of multiple robotic hands. In: Conference on Robot Learning. pp. 516–526. PMLR (2023)
20. Kurin, V., Igl, M., Rocktäschel, T., Boehmer, W., Whiteson, S.: My body is a cage: the role of morphology in graph-based incompatible control. In: ICLR 2021 (2021)
21. Lee, J., Shin, S.Y.: A hierarchical approach to interactive motion editing for human-like figures. In: Proceedings of the 26th annual conference on Computer graphics and interactive techniques. pp. 39–48 (1999)
22. Li, K., Baron, N., Zhang, X., Rojas, N.: Efficientgrasp: A unified data-efficient learning to grasp method for multi-fingered robot hands. *IEEE Robotics and Automation Letters* **7**(4), 8619–8626 (2022)
23. Li, P., Liu, T., Li, Y., Geng, Y., Zhu, Y., Yang, Y., Huang, S.: Gendexgrasp: Generalizable dexterous grasping. 2022 International Conference on Robotics and Automation (ICRA) (2022)
24. Liu, M., Pan, Z., Xu, K., Ganguly, K., Manocha, D.: Deep differentiable grasp planner for high-dof grippers. In: Proceedings of the Robotics: Science and Systems 2020 (2020)
25. Mahler, J., Matl, M., Liu, X., Li, A., Gealy, D., Goldberg, K.: Dex-net 3.0: Computing robust vacuum suction grasp targets in point clouds using a new analytic model and deep learning. In: 2018 IEEE International Conference on robotics and automation (ICRA). pp. 5620–5627. IEEE (2018)
26. Mandikal, P., Grauman, K.: Learning dexterous grasping with object-centric visual affordances. In: 2021 IEEE international conference on robotics and automation (ICRA). pp. 6169–6176. IEEE (2021)
27. Mandikal, P., Grauman, K.: Dexvip: Learning dexterous grasping with human hand pose priors from video. In: Conference on Robot Learning. pp. 651–661. PMLR (2022)
28. Miller, A.T., Allen, P.K.: Graspit! a versatile simulator for robotic grasping. *IEEE Robotics & Automation Magazine* **11**(4), 110–122 (2004)
29. Pathak, D., Lu, C., Darrell, T., Isola, P., Efros, A.A.: Learning to control self-assembling morphologies: a study of generalization via modularity. *Advances in Neural Information Processing Systems* **32** (2019)
30. Penco, L., Clément, B., Modugno, V., Hoffman, E.M., Nava, G., Pucci, D., Tsagarakis, N.G., Mouret, J.B., Ivaldi, S.: Robust real-time whole-body motion

- retargeting from human to humanoid. In: 2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids). pp. 425–432. IEEE (2018)
31. Peng, X.B., Coumans, E., Zhang, T., Lee, T.W., Tan, J., Levine, S.: Learning agile robotic locomotion skills by imitating animals. *Proceedings of the Robotics: Science and Systems* (2020)
  32. Peng, X.B., Kanazawa, A., Malik, J., Abbeel, P., Levine, S.: Sfv: Reinforcement learning of physical skills from videos. *ACM Transactions On Graphics (TOG)* **37**(6), 1–14 (2018)
  33. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 652–660 (2017)
  34. Qin, Y., Su, H., Wang, X.: From one hand to multiple hands: Imitation learning for dexterous manipulation from single-camera teleoperation. *IEEE Robotics and Automation Letters* **7**(4), 10873–10881 (2022)
  35. Qin, Y., Wu, Y.H., Liu, S., Jiang, H., Yang, R., Fu, Y., Wang, X.: Dexmv: Imitation learning for dexterous manipulation from human videos. In: *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXIX*. pp. 570–587. Springer (2022)
  36. Qin, Y., Yang, W., Huang, B., Van Wyk, K., Su, H., Wang, X., Chao, Y.W., Fox, D.: Anyteleop: A general vision-based dexterous robot arm-hand teleoperation system. In: *Robotics: Science and Systems* (2023)
  37. Romero, J., Tzionas, D., Black, M.J.: Embodied hands: Modeling and capturing hands and bodies together. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)* **36**(6) (2017)
  38. Scarselli, F., Gori, M., Tsoi, A.C., Hagenbuchner, M., Monfardini, G.: The graph neural network model. *IEEE transactions on neural networks* **20**(1), 61–80 (2008)
  39. Schaff, C., Yunis, D., Chakrabarti, A., Walter, M.R.: Jointly learning to construct and control agents using deep reinforcement learning. In: *2019 International Conference on Robotics and Automation (ICRA)*. pp. 9798–9805. IEEE (2019)
  40. Shao, L., Ferreira, F., Jorda, M., Nambiar, V., Luo, J., Solowjow, E., Ojea, J.A., Khatib, O., Bohg, J.: Unigrasp: Learning a unified model to grasp with multi-fingered robotic hands. *IEEE Robotics and Automation Letters* **5**(2), 2286–2293 (2020)
  41. She, Q., Hu, R., Xu, J., Liu, M., Xu, K., Huang, H.: Learning high-dof reaching-and-grasping via dynamic representation of gripper-object interaction. *ACM Transactions on Graphics (SIGGRAPH 2022)* **41**(4) (2022)
  42. Sivakumar, A., Shaw, K., Pathak, D.: Robotic telekinesis: learning a robotic hand imitator by watching humans on youtube. *Robotics: Science and Systems (RSS)* (2022)
  43. Turpin, D., Wang, L., Heiden, E., Chen, Y.C., Macklin, M., Tsogkas, S., Dickinson, S., Garg, A.: Grasp’d: Differentiable contact-rich grasp synthesis for multi-fingered hands. In: *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VI*. pp. 201–221. Springer (2022)
  44. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. *Advances in neural information processing systems* **30** (2017)
  45. Villegas, R., Yang, J., Ceylan, D., Lee, H.: Neural kinematic networks for unsupervised motion retargeting. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 8639–8648 (2018)

- 46. Wang, T., Liao, R., Ba, J., Fidler, S.: Nervenet: Learning structured policy with graph neural networks. In: Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada. vol. 30 (2018)
- 47. Xu, Z., Qi, B., Agrawal, S., Song, S.: Adagrasp: Learning an adaptive gripper-aware grasping policy. In: 2021 IEEE International Conference on Robotics and Automation (ICRA). pp. 4620–4626. IEEE (2021)
- 48. Yang, J., Sadigh, D., Finn, C.: Polybot: Training one policy across robots while embracing variability. arXiv preprint arXiv:2307.03719 (2023)
- 49. Zhao, X., Wang, H., Komura, T.: Indexing 3d scenes using the interaction bisector surface. ACM Transactions on Graphics (TOG) **33**(3), 1–14 (2014)