

One Hand to Rule Them All: Canonical Representations for Unified Dexterous Manipulation

Zhenyu Wei Yunchao Yao Mingyu Ding
University of North Carolina at Chapel Hill



Fig. 1: We introduce a canonical hand representation that unifies diverse dexterous hands into a shared parameter space and canonical URDF format, serving as a condition for cross-embodiment policy learning. It enables dexterous grasping and zero-shot generalization to novel hand morphologies, highlighting its potential for a wide range of dexterous manipulation tasks.

Abstract—Dexterous manipulation policies today largely assume fixed hand designs, severely restricting their generalization to new embodiments with varied kinematic and structural layouts. To overcome this limitation, we introduce a parameterized canonical representation that unifies a broad spectrum of dexterous hand architectures. It comprises a unified parameter space and a canonical URDF format, offering three key advantages. 1) The parameter space captures essential morphological and kinematic variations for effective conditioning in learning algorithms. 2) A structured latent manifold can be learned over our space, where interpolations between embodiments yield smooth and physically meaningful morphology transitions. 3) The canonical URDF standardizes the action space while preserving dynamic and functional properties of the original URDFs, enabling efficient and reliable cross-embodiment policy learning.

We validate these advantages through extensive analysis and experiments, including grasp policy replay, VAE latent encoding, and cross-embodiment zero-shot transfer. Specifically, we train a VAE on the unified representation to obtain a compact, semantically rich latent embedding, and develop a grasping policy conditioned on the canonical representation that generalizes across dexterous hands. We demonstrate, through simulation and real-world tasks on unseen morphologies (e.g., 81.9% zero-shot success rate on 3-finger LEAP Hand), that our framework unifies both the representational and action spaces of structurally diverse hands, providing a scalable foundation for cross-hand learning toward universal dexterous manipulation. Project Page: <https://zhenyuwei2003.github.io/OHRA/>

I. INTRODUCTION

Dexterous robotic hands with high degrees of freedom (DoF) and anthropomorphic designs offer advanced flexi-

bility and control beyond the capabilities of simpler parallel or underactuated grippers [7, 4], enabling robots to interact with diverse objects in dynamic and unstructured environments. Recent advances in both learning-based and model-based approaches have driven impressive progress in object grasping [30, 13, 29, 41, 26, 23], dynamic in-hand manipulation [6, 5, 28, 2, 37], and tool use for goal-directed behaviors [1, 14, 12, 40, 20, 33, 9]. However, the specific DoF requirements and structural layouts vary substantially across tasks and robotic platforms. As a result, existing methods generally remain tailored to specific robotic hands, limiting their generalization and reusability across embodiments.

Variations in morphology, DoF, and kinematic layouts make policies trained on one hand design difficult to transfer to another [33]. This lack of interoperability increases the cost of data collection and prevents leveraging heterogeneous datasets across platforms. Recent studies have begun exploring cross-embodiment manipulation: $\mathcal{D}(\mathcal{R}, \mathcal{O})$ Grasp [30] transfers grasping skills across hands by modeling hand-object interactions, but remains task-specific; DexUMI [33] uses the human hand as a universal interface, yet its reliance on human-like kinematics and teleoperation setups limits scalability; and particle-based dynamics learning methods [10] model both hands and objects as particle systems, enabling transfer for deformable manipulation but struggling with structurally distinct hands and broader manipulation skills. Despite these efforts, a central challenge remains: *How to establish a unified*

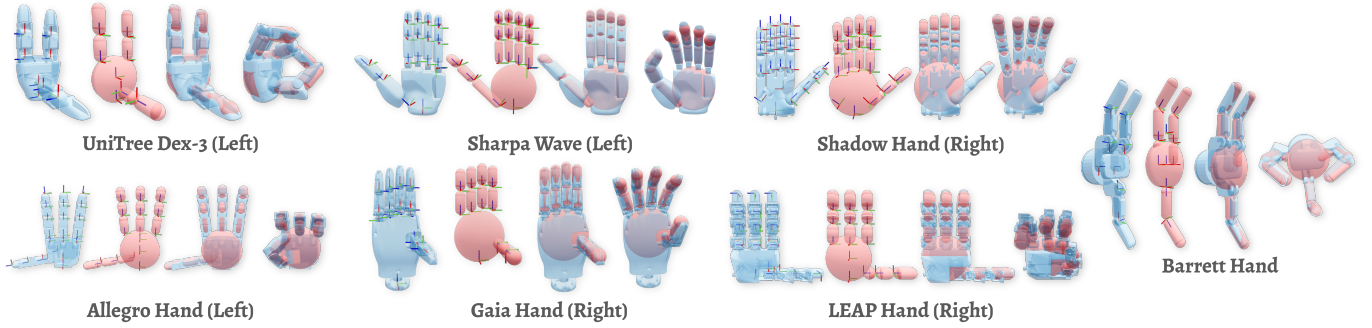


Fig. 2: Comparison of canonical and original URDFs across seven dexterous hands with different finger numbers and handedness. For each hand (from left to right): canonical URDF, original URDF, overlay of initial poses, and overlay of grasp poses, showing close morphological and kinematic consistency between the canonical and original models.

representation and action space that policies can generalize across different hand embodiments and tasks?

Addressing this question requires solving two fundamental challenges: (i) obtaining a compact yet expressive representation of hand morphology that can serve as a conditioning input for learning-based models, and (ii) establishing a unified action space that allows policies to function seamlessly across hands with varying DoFs and kinematic structures. Although the URDF format encodes a complete specification of hand geometry and kinematics, its hierarchical and heterogeneous nature makes it ill-suited for direct use in neural networks for learning-based approaches. This motivates us to introduce a parameterized canonical representation, which encodes hand structure in a learning-friendly format while standardizing the action space across diverse dexterous hands.

Building on this canonical representation, we encode the geometric and kinematic properties of dexterous hands into a fixed set of parameters defined under a canonical URDF format. This format can approximate a wide range of robotic hand designs while maintaining a unified action space, where inactive joints are treated as dummy variables. Through an automated pipeline, existing hand URDFs are converted into this canonical format and corresponding parameter representation, yielding two unified spaces: (i) a parameter space that compactly captures hand morphology, and (ii) an action space that standardizes joint movements across embodiments. This unified formulation enables joint policy training across diverse hand embodiments, bridging their structural and dimensional discrepancies.

We validate our framework through three key experiments. First, training a variational autoencoder (VAE) on the canonical parameter space reveals a structured latent manifold for hand geometry and kinematics, where interpolation between different embodiments produces smooth, physically meaningful morphology transitions. Second, grasp policy replay and in-hand reorientation tasks show that the canonical URDF preserves the kinematic and functional characteristics of the original hands, achieving performance comparable to their native URDFs. Finally, conditioning on hand morphology representations, we train a cross-embodiment dexterous grasp-

ing policy that generalizes across diverse hand designs. We first demonstrate effective policy transfer across three distinct robotic hands, where the shared policy outperforms per-hand baselines. We further scale this approach by training on near one hundred LEAP Hand variants, enabling zero-shot generalization to unseen hand morphologies and robust grasping performance in both simulation and real-world experiments, without additional fine-tuning.

Our main contributions are summarized as follows:

- 1) We propose a canonical representation for dexterous hands that standardizes diverse morphologies and kinematic structures into a unified parameterized format, enabling consistent and learning-friendly structural encoding across hands.
- 2) Extensive experiments including morphology latent interpolation, grasp policy replay, in-hand reorientation, and unseen cross-hand grasping demonstrate that the canonical format not only faithfully preserves functional behavior but also provides a unified action space for zero-shot and effective policy transfer across embodiments.
- 3) For the first time, we establish an interpretable, interpolable, and scalable representation foundation that enables joint policy training for cross-embodiment dexterous manipulation, paving the way for unified large-scale and morphology-aware learning.

II. RELATED WORKS

Dexterous Manipulation. High-DoF robotic hands provide the articulation needed for fine-grained contact control and multi-stage manipulation, supporting tasks from stable grasping to dynamic in-hand reconfiguration and tool-mediated interactions [11, 42, 5, 30, 13, 27, 32, 14]. These capabilities have been extensively explored through analytic models and data-driven learning, producing strong performance on individual embodiments. In particular, reinforcement learning and imitation learning have yielded high-quality grasping and manipulation policies when confined to a fixed dexterous hand [38, 6, 39, 40]. However, because such policies are optimized for the specific kinematics, actuation, and workspace of a single hand, they become tightly coupled to that embodiment [34, 26]. This embodiment-specific specialization limits

transfer to other hands and prevents the reuse of demonstrations across heterogeneous hardware, leaving progress fragmented across isolated hand designs rather than advancing toward methods that generalize to new embodiments.

Cross-Embodiment Policy Learning. Recent work has increasingly explored how to share manipulation abilities across robotic hands with distinct morphologies [30, 13, 33, 10, 31, 8, 3, 35, 16]. Much of this progress has centered on grasping. One line of work focuses on intermediate grasp representations that abstract away embodiment-specific kinematics. These methods leverage representations such as interaction-centric fields, contact patterns, or PCA-based pose synergies to enable grasp transfer across different robotic hands, but remain largely confined to grasp synthesis and do not naturally extend to sequential manipulation skills [30, 13, 18]. Beyond grasping, some approaches target more general cross-embodiment behaviors through higher-level or embodiment-agnostic interfaces. Human-centric representations treat the human hand as a universal manipulation prior but typically assume human-like kinematics and require specialized hardware mappings [33]. Particle-based dynamics learning provides an alternative by representing hands and objects as particle systems, yet is mainly applicable to structurally similar hands and constrained manipulation tasks [10]. Despite these efforts, a unified cross-embodiment paradigm capable of supporting general manipulation across heterogeneous robotic hands is still lacking.

III. CANONICAL HAND REPRESENTATION DESIGN

Overview. To enable learning policies that generalize across dexterous hands with different morphologies, we propose a parameterized canonical hand representation, which serves as the foundation for cross-embodiment manipulation. The goal is to express diverse robotic hands within a unified structural and kinematic framework that can be efficiently processed by learning-based models.

We begin by motivating the need for a canonical representation (Sec. III-A), highlighting the limitations of existing URDF formats and the challenges of defining a consistent and learnable description for heterogeneous dexterous hands. We then present the design of the canonical URDF (Sec. III-B), which captures shared human-inspired kinematic structure while enforcing consistent coordinate conventions. Next, we define the canonical parameter set that encodes key morphological and kinematic properties in a compact and interpretable form (Sec. III-C). We further describe the automatic parsing process that converts arbitrary hand URDFs into this canonical parameterization and generates standardized URDF models (Sec. III-D). Finally, we establish a unified action space that aligns control dimensions across hands with different degrees of freedom (Sec. III-E), enabling a single policy to act consistently over diverse embodiments.

A. Motivation

To develop a task-agnostic policy that generalizes across dexterous hands with diverse embodiments, the learning framework must incorporate a representation of the current

TABLE I: Comparison of representative dexterous robotic hands and their kinematic configurations. F, A, and R denote flexion, abduction/adduction, and axial rotation. DoF values in parentheses include passive or mechanically coupled joints.

Name	DoF	Joint Types				
		Thumb	Index	Middle	Ring	Little
Shadow Hand	22	RAAFF	AFFF	AFFF	AFFF	FAFFF
Sharpa Wave	22	AFAFF	AFFF	AFFF	AFFF	FAFFF
WUJI Hand	20	FAFF	AFFF	AFFF	AFFF	AFFF
DexHand 021	12(19)	AFFF	AFFF	FFF	AFFF	AFFF
Orca Hand	16	FAFF	AFF	AFF	AFF	AFF
Gaia Hand	15	AFF	AFF	AFF	AFF	AFF
XHAND1	12	AFF	AFF	FF	FF	FF
Inspire	6(12)	AFFF	FF	FF	FF	FF
Allegro	16	ARFF	RFFF	RFFF	-	RFFF
LEAP Hand	16	ARFF	AFFF	AFFF	-	AFFF
Barrett	8	FF	AFF	AFF	-	-
Dex3	7	RFF	FF	FF	-	-

hand as part of the model input. Equally important is a consistent action space across hands, enabling a single policy to operate seamlessly over different morphologies.

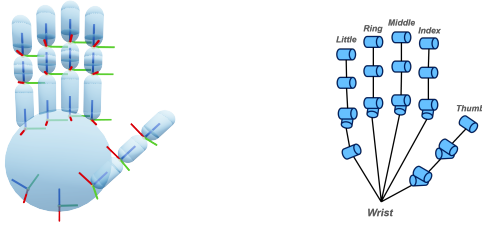
A natural approach is to describe each hand using general 3D representations, such as point clouds or signed distance fields. While these capture detailed geometry, they only reflect static structure and ignore the kinematic and dynamic properties crucial for manipulation. In contrast, the Unified Robot Description Format (URDF) encodes comprehensive information about a hand’s morphology and motion capabilities, including joint topology, link dimensions, and actuation limits. However, URDFs are typically handcrafted, vary across platforms, and are tree-structured and heterogeneous, making them difficult to use directly in learning-based pipelines.

To overcome these limitations, we introduce a canonical URDF representation that unifies the structural and kinematic description of dexterous hands. Each hand is represented as a compact set of parameters capturing key geometric and motion attributes, from which a standard URDF can be automatically generated. This canonical form preserves both structural and functional consistency while providing a format suitable for neural network input. It enables policies to be conditioned on hand morphology and executed in a shared action space, laying the foundation for scalable cross-hand generalization.

B. Canonical URDF Design

The canonical URDF unifies the structural representation of diverse dexterous hands by capturing shared kinematic principles in a standardized yet expressive format. The main challenge is selecting essential morphological parameters while remaining compatible with heterogeneous joint configurations and mechanical layouts. Preserving all native URDF details introduces numerous coordinate-dependent variables and inconsistent representations, while an overly compact model would constrain expressiveness. The canonical formulation therefore strikes a balance between representational richness, generality, and physical realizability.

1) *Kinematic Analysis of Existing Dexterous Hands:* To guide the canonical design, we analyzed a representative set



(a) Mesh and frame visualization (b) Kinetic skeleton diagram

Fig. 3: Structure of the canonical URDF. A right-hand configuration is shown for clarity, but the representation is applicable to both left- and right-handed hands.

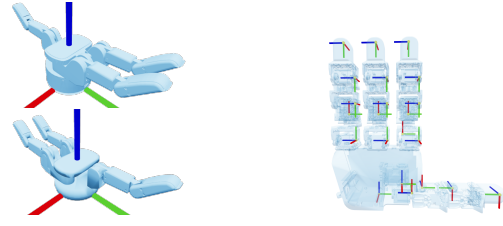
of commercial and open-source dexterous hands, focusing on kinematic organization and degrees of freedom (DoF) (Table I). Despite mechanical differences, most designs share a human-inspired layout with recurring motion patterns. The thumb typically has two distal flexion joints and, when present, a separate abduction–adduction joint, while proximal thumb joints exhibit greater variability due to mechanical design. Remaining fingers generally have an abduction–adduction proximal joint followed by a flexion chain dominating grasping, with some higher-DoF hands adding extra flexion for the little finger to approximate human-like coupled motion.

Guided by these observations, we define a canonical URDF supporting up to five fingers and 22 DoF (Fig. 3), capturing the shared human-like topology. To maintain geometric consistency across different link meshes, all links are represented as capsule primitives, which also reflect the cylindrical shape of most hands’ collision meshes. This abstraction reduces geometric complexity while preserving the essential kinematic relationships required for dexterous manipulation.

2) *Unified Coordinate Conventions:* Dexterous hand URDFs vary in global and local coordinate frames. Even identical hardware can have different global axes (Fig. 4a) or incompatible local joint orientations (Fig. 4b), causing unnecessary transformations and ambiguity. Our canonical URDF enforces a unified convention aligned with human-hand kinematics (Fig. 3a): the palm normal along $+x$, the thumb (right hand) along $+y$, and other fingers along $+z$, with local axes x for abduction–adduction, y for flexion–extension, and z for axial rotation when applicable. This standardized convention ensures consistent, interpretable kinematics and enables parameterization that captures key morphological and kinematic variations for cross-hand policy learning.

C. Parameter Definition

Building on the canonical URDF design, we define a compact parameter set for dexterous hands that captures key morphological and kinematic information. The canonical representation comprises 82 parameters, while a more comprehensive URDF template with 173 parameters is provided to support specialized hand designs. Full details and design rationale are given in the Supplementary Material. By consolidating geometric and kinematic attributes into this structured representation, the canonical model captures the most salient



(a) Global frame inconsistency (b) Local frame ambiguity

Fig. 4: Coordinate frame inconsistencies in URDFs. (a) Global orientations vary across sources, (b) local joint frames use inconsistent axis definitions, leading to kinematic ambiguity.

variations across embodiments and facilitates efficient learning and transfer among diverse dexterous hands.

D. URDF Parsing and Generation

To support a wide range of articulated hand designs, we develop an automatic framework for URDF parsing and generation. The framework extracts canonical parameters from arbitrary hand URDFs and reconstructs morphology and kinematics in the canonical space from minimal manual inputs, while also enabling full URDF generation from these parameters. Generation is implemented using the Jinja2 dynamic templating language [17], allowing conditional inclusion of elements and automatic adaptation to hands with varying numbers of fingers or links. Together, these capabilities provide consistent bidirectional conversion between diverse robot models and the unified representation (Sec. III-B). Full details of the parsing and generation procedures are provided in the Supplementary Material.

E. Action Space Unification

With the canonical URDF, all dexterous hands share a fixed 22-DoF structure defining a unified control and observation space. Hands with fewer active DoF have the corresponding joints set inactive and the associated links removed, maintaining a consistent joint and link representation across embodiments. Using the joint-to-joint mapping from parsing, we enable bidirectional conversion between original and canonical joint vectors with consistent indexing and sign conventions. This unified formulation standardizes joint dimensionality and ordering, providing a coherent action space for scalable cross-hand policy learning and transfer.

IV. CANONICAL REPRESENTATION APPLICATIONS

We evaluate the proposed canonical representation through four complementary studies. First, we examine whether the parameter space introduced in Sec. III-C induces a continuous and interpretable latent manifold (Sec. IV-A) using a variational autoencoder trained on diverse sampled hand morphologies. Second, we assess the physical fidelity of the canonical URDF by comparing in-hand object reorientation performance against the original hand models (Sec. IV-B), verifying that the canonical formulation preserves key kinematic and control properties. Third, leveraging the unified structure and action representation, we train a single grasping policy transferable

across multiple dexterous hands (Sec. IV-C), demonstrating scalability and cross-embodiment generalization. Finally, we investigate the zero-shot grasping capability enabled by hand conditioning using LEAP Hand and its variants (Sec. IV-D), assessing generalization to unseen hand morphologies.

A. Learning a Morphology Latent Space

To evaluate whether the canonical parameterization defines a compact and structured embedding, we train a variational autoencoder (VAE) to map hand morphology parameters into a 16-dimensional latent space. The latent representation captures essential geometric and kinematic variations across dexterous hands, forming a smooth manifold that supports interpolation and generalization.

A dataset of 65,536 synthetic hand configurations is generated by sampling each canonical parameter within physically feasible ranges. Continuous morphology parameters (palm/finger radii, link lengths) are sampled from bounded intervals, finger-origin translations preserve plausible proportions, joint axes are encoded as one-hot vectors over six canonical directions ($\pm x, \pm y, \pm z$), and joint availability is encoded as binary indicators over the 22 canonical DoFs. The VAE reconstructs these parameters using type-specific losses with KL regularization. Full model architecture, training objectives, and hyperparameters are provided in the Supplementary Material.

B. In-hand Reorientation

We evaluate the physical fidelity of the canonical representation using an in-hand object rotation task, where reinforcement learning agents are trained to rotate an object about a predefined axis. Separate policies are learned for the original dexterous hands and their canonical counterparts. Experiments are conducted in the IsaacGym simulator [15] using the Shadow Hand [21] and the LEAP Hand [24], following prior in-hand rotation setups [19, 24]. Observations include joint positions and velocities, PD controller targets, and the object’s pose and velocities. The reward promotes high object angular velocity while penalizing object drift, excessive hand motion, and torque usage. Policies are trained with Proximal Policy Optimization (PPO) [22] using MLP-based agents. Full training details are provided in the Supplementary Material.

C. Cross-Embodiment Dexterous Grasping

To evaluate the transferability of control across different robotic hands, we train a cross-embodiment grasp pose prediction model that operates entirely within the canonical representation. A grasp pose is defined by the wrist pose (T, R) relative to the object frame, parameterized by translation and rotation, together with the hand joint configuration θ .

As shown in Fig. 6, grasp generation follows a two-stage formulation inspired by Zhang et al. [41]. In the first stage, a diffusion-based conditional generator predicts the distribution of wrist translations around an object, conditioned on a grasp feature f_g and an explicitly provided wrist rotation R . Providing R separately decouples orientation from translation,

allowing direct control over grasp direction and supporting diverse floating grasps. Training data spans multiple object orientations, enabling the model to generalize to arbitrary directions while supporting orientation-constrained sampling at test time. The second stage uses a lightweight MLP to predict the corresponding hand joint configuration θ conditioned on the sampled wrist pose (T, R) and f_g , implementing a deterministic mapping.

The grasp feature f_g integrates both object and hand morphology information. Object features are extracted from the input point cloud using the $\mathcal{D}(\mathcal{R}, \mathcal{O})$ point-based encoder, while hand features are obtained from the frozen VAE latent embedding introduced in Sec. IV-A. This allows the model to generalize grasp predictions across hands with different morphologies.

The training objective combines diffusion-based translation prediction with deterministic joint regression, both optimized using a Smooth- L_1 loss:

$$\mathcal{L} = \text{SmoothL1}(\hat{T}, T) + \text{SmoothL1}(\hat{\theta}, \theta). \quad (1)$$

D. LEAP Hand Zero-Shot Generalization

To further demonstrate the zero-shot generalization capability enabled by hand conditioning, and considering both the limited availability of existing dexterous hand designs and the feasibility of real-world experiments, we adopt the open-source, modular LEAP Hand hardware. By systematically varying the presence of individual links for each finger, we construct $4^4 = 256$ LEAP Hand variants, denoted as `leap_xyzw`, where $x, y, z, w \in \{0, 1, 2, 3\}$ correspond to the number of links of the thumb, index, middle, and little fingers, respectively. For example, `leap_3333` corresponds to the original LEAP Hand design.

Under our canonical hand representation, different LEAP Hand variants can be generated by simply modifying the corresponding morphology parameters, allowing efficient and scalable instantiation of a large number of hand designs. Grasp data for each variant are generated using Lightning Grasp [36], where a dedicated configuration file is specified for each hand morphology. The generated grasps are further filtered using the data filtering strategy from $\mathcal{D}(\mathcal{R}, \mathcal{O})$ Grasp [30] to ensure physical plausibility. See the Supplementary Material for more details.

To balance practical grasp feasibility and the evaluation of zero-shot generalization, we construct the grasping dataset using 66 LEAP Hand variants satisfying $x + y + z + w \geq 8$. The grasping policy is trained following the same procedure as in Sec. IV-C, with the hand condition directly specified by the canonical morphology parameters. Zero-shot generalization is then evaluated on LEAP Hand variants with $x + y + z + w < 8$, as well as additional variants whose grasp data are entirely excluded during training, allowing us to assess generalization to unseen hand morphologies without further fine-tuning.

V. EXPERIMENTS

We evaluate the effectiveness of the proposed canonical URDF representation and the applications described in

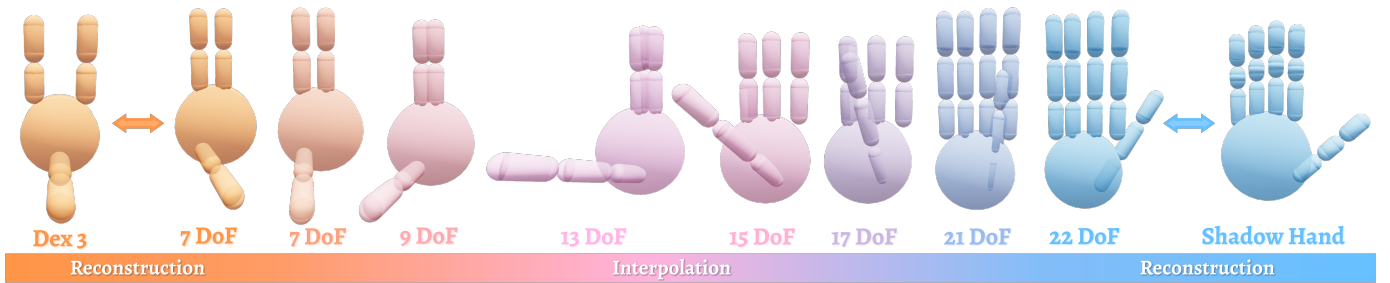


Fig. 5: Visualization of latent-space interpolation between two dexterous hands. Canonical URDFs are shown at the ends, with decoded reconstructions and interpolated morphologies in between, demonstrating smooth transitions in DoF, finger arrangement, and overall geometry.

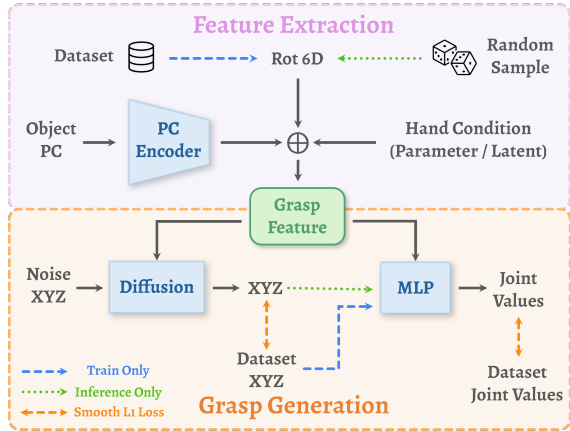


Fig. 6: Two-stage cross-embodiment grasp generation pipeline.

Sec. IV. Our experiments focus on four key aspects: (1) the quality and continuity of the learned morphology latent space, (2) the fidelity and dynamics preserved when motions are expressed using the canonical hand model, (3) the generalization of a single unified grasping policy across dexterous hands with diverse kinematic structures, and (4) the zero-shot capability enabled by hand conditioning. We further deploy the zero-shot grasping policy on physical hardware, demonstrating that models trained in simulation with the canonical model maintain strong performance in real-world grasping tasks.

A. VAE Latent Space Visualization

We first examine whether the variational autoencoder (VAE) introduced in Sec. IV-A learns a structured, continuous embedding of dexterous-hand morphology. To assess this, we visualize latent-space interpolations between two hands with distinct geometric and kinematic designs, such as a compact three-finger gripper and a high-DoF anthropomorphic hand.

Given latent features z_a and z_b for the two canonical encodings, we interpolate using $z(\alpha) = (1 - \alpha)z_a + \alpha z_b$ for $\alpha \in [0, 1]$, decode each feature into canonical URDF parameters, and generate the corresponding hand models. As shown in Figure 5, the interpolated structures evolve smoothly, with gradual variation in palm size, finger count and spacing, thumb placement, and DoF, indicating that the VAE captures a continuous morphological representation.

These results suggest that the latent space preserves intrinsic

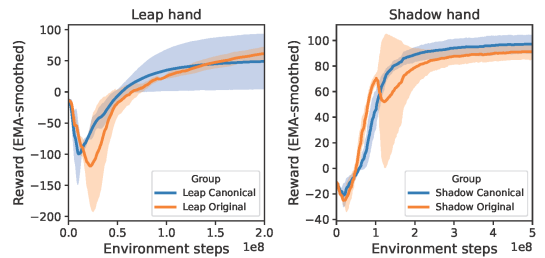


Fig. 7: Smoothed log of training reward over simulation steps during training. The solid lines mark the mean value, and the shaded area consists of the standard deviation.

TABLE II: Comparison of in-hand reorientation policies trained with the canonical URDF vs. the original URDF.

Policy	Steps-to-Fall $\uparrow$	Cumulative Rotation $\uparrow$
Shadow (Original)	369.66	9.09
Shadow (Canonical)	390.62	10.92
LEAP (Original)	397.62	5.63
LEAP (Canonical)	326.98	6.31

structural relationships among diverse hand designs, which is critical for downstream tasks such as morphology-conditioned grasping and cross-hand policy transfer.

B. Canonical Hand Fidelity

1) *In-Hand Reorientation*: We evaluate the physical fidelity of the canonical hand representation by comparing its performance to that of the original hand models in high-dynamic in-hand manipulation tasks, assessing whether essential kinematic and control properties are preserved.

Metrics. (1) *Steps-to-Fall*: the number of simulation steps the object remains stably grasped; and (2) *Cumulative Rotation*: the total rotation about the positive z-axis in radians, indicating how effectively the policy drives rotation. These metrics capture both stability and rotation, allowing direct comparison between canonical and original hands.

Results. Table II reports the average performance, showing that canonical representations achieve comparable Steps-to-Fall and Cumulative Rotation. Fig. 7 further indicates that the learning progress and convergence patterns are similar for both representations. These results demonstrate that the

TABLE III: Comparison of grasp success rates when transferring across canonical and original URDFs.

Method	Success Rate (%)		
	Allegro	Barrett	ShadowHand
Ours (Canonical)	84.20	88.10	62.90
Ours (Original)	71.60 (-12.60)	88.70 (+0.60)	62.60 (-0.30)
D(R,O) (Original)	92.30	87.30	83.00
D(R,O) (Canonical)	92.38 (+0.08)	87.34 (+0.04)	78.63 (-4.37)

TABLE IV: Grasp performance comparison.

Method	Success Rate (%) $\uparrow$			Time (sec.) $\downarrow$
	Allegro	Barrett	ShadowHand	
DFC	76.2	86.3	58.8	>1800
GenDexGrasp	51.0	67.0	54.2	19.71
D(R,O) Grasp	92.3	87.3	83.0	0.65
Ours	<u>84.2</u>	88.1	<u>62.9</u>	0.13

canonical parameterization preserves essential manipulation dynamics and can serve as a practical interface for downstream applications such as reinforcement learning.

2) *Grasp Policy Transfer via Action Mapping*: Using the bidirectional mappings introduced in Sec. III-E, we replay grasp policies between the canonical and original URDF spaces. The policy trained in canonical space (Ours, Sec. V-C) is mapped to the original URDF, while the policy trained on the original URDF ($\mathcal{D}(\mathcal{R}, \mathcal{O})$) is mapped to the canonical space. As shown in Table III, transfers in both directions achieve closely matched success rates, indicating that the canonical representation preserves the action semantics required for stable execution. The main discrepancy occurs with the Allegro Hand, due to the omission of its axial-rotation joint in the canonical URDF, creating a minor structural mismatch. Overall, these results demonstrate that the mapping is robust and that the canonical action space provides a coherent interface for executing policies across diverse hand designs.

C. Cross-Embodiment Grasping Performance

We build upon the filtered GenDexGrasp [13] dataset provided by $\mathcal{D}(\mathcal{R}, \mathcal{O})$ Grasp [30], which contains 24,764 valid grasps from three dexterous hands: Allegro, Barrett, and Shadow Hand. These hands vary substantially in geometry and kinematics, spanning 8–22 DoFs and three-, four-, and five-finger topologies. All grasps are converted into our canonical URDF, providing a consistent action space across embodiments and enabling evaluation of its generality and applicability across different hands.

We then train a single unified grasp-generation model on the canonical URDF introduced in Sec. IV-C, evaluating its performance on 10 unseen objects against state-of-the-art baselines and comparing unified training with embodiment-specific training. The evaluation metric is provided in the Supplementary Material.

Comparison with State-of-the-Art Methods. As shown in Table IV, our lightweight model achieves grasp success rates comparable to more complex pipelines, without requiring optimization-based refinement. Inference uses a 10-step DDIM

TABLE V: Comparison of grasp success rates. “Specific” indicates that each embodiment is trained independently, while “Unified” denotes joint training across all embodiments.

Method	Success Rate (%)		
	Allegro	Barrett	ShadowHand
Specific	82.1	87.6	55.4
Unified	84.2	88.1	62.9

sampler [25] and runs in only 0.13 s, making it the most efficient among the evaluated methods.

This comparison is intended to evaluate the canonical URDF as a downstream action space rather than to introduce a new grasping algorithm. The results show that even a simple model trained in this representation produces high-quality grasps across diverse hands, demonstrating that the canonical parameterization is expressive, coherent, and supports robust cross-hand grasp generation without reliance on hand-specific architectures or heavy engineering.

Unified Training vs. Embodiment-Specific Training. The unified model consistently outperforms embodiment-specific models (Table V), indicating that the canonical URDF enables effective policy sharing across morphologies. By learning in a shared action space, hands with distinct kinematics benefit from each other’s data, highlighting the representation’s capacity to support cross-embodiment generalization.

D. Zero-Shot Grasping Performance

We evaluate the zero-shot generalization of the canonical grasping policy using a set of 10 objects. Following the procedure described in Sec. IV-D, we generate and filter grasps for various LEAP Hand variants, limiting each hand-object pair to a maximum of 200 grasps and applying selection criteria on the hand variants. The resulting dataset contains 69,917 grasps. For zero-shot evaluation, the training data excludes grasps corresponding to the validated LEAP Hand variants.

We first train a full model on the complete dataset to assess cross-embodiment performance on seen hands. We then train three additional models, each excluding data from a specific LEAP Hand variant (leap_3033, leap_3303, and leap_3330) for zero-shot testing. As shown in Table VII, the models conditioned on hand morphology achieve performance on unseen hands comparable to that on seen hands, demonstrating strong zero-shot transfer capability across unobserved hand designs.

We further evaluate zero-shot generalization on a set of LEAP Hand variants with $x + y + z + w < 8$, which differ substantially from those in the training dataset, making the task more challenging. To isolate the effect of hand morphology on success rates, we generated and filtered grasp data specifically for these variants, training models only on their corresponding data. The resulting grasp data numbers and success rates are summarized in Table VIII.

As shown, the zero-shot models outperform the variant-specific models on leap_0312, leap_2203, and leap_3103. Performance is slightly lower on leap_0303

TABLE VI: Real-World grasp success rates.

Model	Success Rate										
	Apple	Band Aid	Coke	Cube	Football	Mayo	Orange	Pear	Sheep	Soccer	Average
leap_3333 (trained)	8/10	7/10	9/10	7/10	10/10	6/10	8/10	9/10	10/10	9/10	83/100
leap_3033 (trained)	8/10	8/10	2/10	6/10	9/10	6/10	7/10	9/10	10/10	10/10	75/100
leap_3033 (zero-shot)	8/10	10/10	5/10	5/10	7/10	2/10	9/10	7/10	9/10	9/10	71/100
leap_3303 (trained)	7/10	8/10	5/10	3/10	9/10	4/10	9/10	7/10	9/10	9/10	70/100
leap_3303 (zero-shot)	9/10	6/10	4/10	5/10	9/10	5/10	8/10	6/10	9/10	10/10	71/100



(a) Object list (b) Zero-shot (left four) and trained (right four) grasping results with different LEAP Hand variants.

Fig. 8: Real-world grasping objects and results.

TABLE VII: Comparison of grasp success rates. Underlined values indicate zero-shot evaluation.

Model	Success Rate (%)		
	leap_3033	leap_3303	leap_3330
All Data	76.1	85.4	43.3
No leap_3033 Data	<u>67.8</u>	83.4	31.5
No leap_3303 Data	81.5	<u>81.9</u>	46.9
No leap_3330 Data	74.7	81.6	<u>36.3</u>

TABLE VIII: Grasp success rates and dataset number for selected LEAP Hand variants.

Model	Success Rate (%)				
	0303	0312	2203	3030	3103
All Data	46.9	13.3	65.1	36.2	46.6
Specific Data	75.1	12.1	33.9	55.4	18.5
Data Num	37249	4368	2458	37217	2124

TABLE IX: Grasp results for leap_3033 across hand conditions.

Condition	Success Rate (%)	
	All Data	Zero-Shot
leap_3303	85.4	81.6
leap_3033	33.9	12.8
leap_3330	20.5	2.4
leap_3333	85.1	71.9

and leap_3030, likely because these two variants have only two fingers, resembling a gripper, and their grasp patterns differ markedly from those of dexterous hands in the training set. In addition, these variants generated a large number of grasps ($\approx 30k$) in a single round, which benefits the variant-specific models. Nevertheless, the zero-shot models still demonstrate strong generalization to these challenging morphologies, highlighting the effectiveness of the canonical hand conditioning and unified action space.

To further validate the impact of hand conditioning, we evaluated the model using incorrect hand conditions on other LEAP Hand variants. As shown in Table IX, we tested the variant leap_3033. Overall, applying an incorrect hand condition substantially reduces grasp success rates. Using the original LEAP Hand parameters (leap_3333) in the “All Data” trained model results in only a minor drop, as the model can partially overfit to the included variant. In contrast, in the zero-shot setting, success rates drop significantly, highlighting the critical role of hand conditioning.

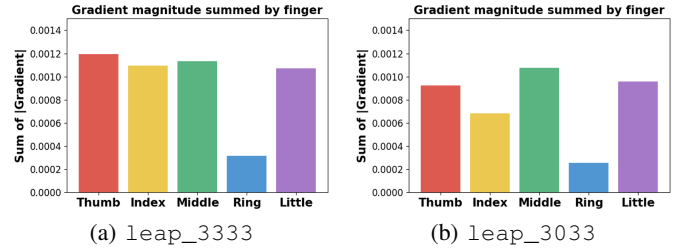


Fig. 9: Gradient magnitude visualization.

We also visualize the backward gradients with respect to the canonical parameters (Fig. 9). The gradient for the ring finger remains consistently low, as this finger is absent in the LEAP Hand. In variant leap_3033, where the index finger is removed, its gradient drops markedly, indicating that the model has learned to focus on the functional fingers essential for successful grasps.

E. Real-World Experiment

To validate the sim-to-real transfer and practical applicability of our approach, we deploy the LEAP Hand [24] grasping policies on a Franka Research 3 robotic arm. Object observations are captured using an Intel RealSense L515 depth camera. The evaluation is conducted on a set of 10 diverse objects, as illustrated in Fig. 8a, and across several LEAP Hand variants (Fig. 8b). Details of real-world experiments are in the Supplementary Material.

We evaluate both models trained on the canonical dataset and zero-shot models that have never seen the target hand variants. As summarized in Table VI, the trained models achieve high grasp success rates, demonstrating that the canonical hand representation preserves the essential dynamics and physical fidelity of the original hands, and that sim-to-real transfer is reliable. Importantly, the zero-shot models achieve success rates close to those of the trained models, highlighting strong generalization capabilities and the effectiveness of the hand condition in guiding grasping across unseen morphologies. The multi-panel illustrations in Fig. 8b further demonstrate that

even for diverse and unusual LEAP Hand variants, the policy executes stable and robust grasps, highlighting the strength and effectiveness of the canonical hand conditioning.

VI. CONCLUSION AND DISCUSSION

We introduced a canonical representation that maps heterogeneous dexterous hands into a shared parameter and action space, enabling scalable cross-embodiment learning. Its continuous morphology parameterization enables explicit hand-aware conditioning, and the unified action space facilitates data and policy sharing across platforms. Together, these properties enable embodiment-agnostic manipulation policies that generalize in a zero-shot manner to unseen hands, establishing the canonical URDF as a practical foundation for scalable cross-embodiment manipulation. Beyond dexterous hand grasping, our insights can extend to broader robotic embodiments such as humanoid robots and to more diverse manipulation tasks, potentially benefiting the embodied AI and robotics communities.

REFERENCES

- [1] Ananye Agarwal, Shagun Uppal, Kenneth Shaw, and Deepak Pathak. Dexterous functional grasping. *arXiv preprint arXiv:2312.02975*, 2023.
- [2] OpenAI: Marcin Andrychowicz, Bowen Baker, Maciek Chociej, Rafal Jozefowicz, Bob McGrew, Jakub Pachocki, Arthur Petron, Matthias Plappert, Glenn Powell, Alex Ray, et al. Learning dexterous in-hand manipulation. *The International Journal of Robotics Research*, 39(1):3–20, 2020.
- [3] Erik Bauer, Elvis Nava, and Robert K Katzschmann. Latent action diffusion for cross-embodiment manipulation. *arXiv preprint arXiv:2506.14608*, 2025.
- [4] A. Bicchi. Hands for dexterous manipulation and robust grasping: a difficult road toward simplicity. *IEEE Transactions on Robotics and Automation*, 16(6):652–662, 2000. doi: 10.1109/70.897777.
- [5] Tao Chen, Jie Xu, and Pulkit Agrawal. A system for general in-hand object re-orientation. In *Conference on Robot Learning*, pages 297–307. PMLR, 2022.
- [6] Tao Chen, Megha Tippur, Siyang Wu, Vikash Kumar, Edward Adelson, and Pulkit Agrawal. Visual dexterity: In-hand reorientation of novel and complex object shapes. *Science Robotics*, 8(84):eadc9244, 2023. doi: 10.1126/scirobotics.adc9244. URL <https://www.science.org/doi/abs/10.1126/scirobotics.adc9244>.
- [7] M.R. Cutkosky. On grasp choice, grasp models, and the design of hands for manufacturing tasks. *IEEE Transactions on Robotics and Automation*, 5(3):269–279, 1989. doi: 10.1109/70.34763.
- [8] Xin Fei, Zhixuan Xu, Huaicong Fang, Tianrui Zhang, and Lin Shao. T(r, o) grasp: Efficient graph diffusion of robot-object spatial transformation for cross-embodiment dexterous grasping. *arXiv preprint arXiv:2510.12724*, 2025.
- [9] Jingxiang Guo, Jiayu Luo, Zhenyu Wei, Yiwen Hou, Zhixuan Xu, Xiaoyi Lin, Chongkai Gao, and Lin Shao. Telepreview: A user-friendly teleoperation system with virtual arm assistance for enhanced effectiveness. *arXiv preprint arXiv:2412.13548*, 2024.
- [10] Zihao He, Bo Ai, Tongzhou Mu, Yulin Liu, Weikang Wan, Jiawei Fu, Yilun Du, Henrik I Christensen, and Hao Su. Scaling cross-embodiment world models for dexterous manipulation. *arXiv preprint arXiv:2511.01177*, 2025.
- [11] Binghao Huang, Yuanpei Chen, Tianyu Wang, Yuzhe Qin, Yaodong Yang, Nikolay Atanasov, and Xiaolong Wang. Dynamic handover: Throw and catch with bi-manual hands. *arXiv preprint arXiv:2309.05655*, 2023.
- [12] Kun Lei, Huanyu Li, Dongjie Yu, Zhenyu Wei, Lingxiao Guo, Zhennan Jiang, Ziyu Wang, Shiyu Liang, and Huazhe Xu. RL-100: Performant robotic manipulation with real-world reinforcement learning. *arXiv preprint arXiv:2510.14830*, 2025.
- [13] Puhao Li, Tengyu Liu, Yuyang Li, Yixin Zhu, Yaodong Yang, and Siyuan Huang. Gendexgrasp: Generalizable dexterous grasping. *arXiv preprint arXiv:2210.00722*, 2022.
- [14] Zhixuan Liang, Yao Mu, Yixiao Wang, Tianxing Chen, Wenqi Shao, Wei Zhan, Masayoshi Tomizuka, Ping Luo, and Mingyu Ding. Dexhanddiff: Interaction-aware diffusion planning for adaptive dexterous manipulation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 1745–1755, 2025.
- [15] Viktor Makoviychuk, Lukasz Wawrzyniak, Yunrong Guo, Michelle Lu, Kier Storey, Miles Macklin, David Hoeller, Nikita Rudin, Arthur Allshire, Ankur Handa, et al. Isaac gym: High performance gpu-based physics simulation for robot learning. *arXiv preprint arXiv:2108.10470*, 2021.
- [16] Austin Patel and Shuran Song. Get-zero: Graph embodiment transformer for zero-shot embodiment generalization. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 14262–14269. IEEE, 2025.
- [17] Pallets Projects. Jinja documentation. <https://jinja.palletsprojects.com/en/stable/>, 2022.
- [18] En Yen Puang, Federico Ceola, Giulia Pasquale, and Lorenzo Natale. Pchands: Pca-based hand pose synergy representation on manipulators with n -dof. In *2025 IEEE-RAS 24th International Conference on Humanoid Robots (Humanoids)*, pages 475–482. IEEE, 2025.
- [19] Haozhi Qi, Ashish Kumar, Roberto Calandra, Yi Ma, and Jitendra Malik. In-hand object rotation via rapid motor adaptation. In *Conference on Robot Learning*, pages 1722–1732. PMLR, 2023.
- [20] Aravind Rajeswaran, Vikash Kumar, Abhishek Gupta, Giulia Vezzani, John Schulman, Emanuel Todorov, and Sergey Levine. Learning complex dexterous manipulation with deep reinforcement learning and demonstrations. *arXiv preprint arXiv:1709.10087*, 2017.

- [21] Shadow Robot. Dexterous hand series. <https://shadowrobot.com/dexterous-hand-series/>.
- [22] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.
- [23] Lin Shao, Fabio Ferreira, Mikael Jorda, Varun Nambiar, Jianlan Luo, Eugen Solowjow, Juan Aparicio Ojea, Oussama Khatib, and Jeannette Bohg. Unigrasp: Learning a unified model to grasp with multifingered robotic hands. *IEEE Robotics and Automation Letters*, 5(2):2286–2293, 2020.
- [24] Kenneth Shaw, Ananye Agarwal, and Deepak Pathak. Leap hand: Low-cost, efficient, and anthropomorphic hand for robot learning. *arXiv preprint arXiv:2309.06440*, 2023.
- [25] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.
- [26] Weikang Wan, Haoran Geng, Yun Liu, Zikang Shan, Yaodong Yang, Li Yi, and He Wang. Unidexgrasp++: Improving dexterous grasping policy learning via geometry-aware curriculum and iterative generalist-specialist learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3891–3902, 2023.
- [27] Jun Wang, Yuzhe Qin, Kaiming Kuang, Yigit Korkmaz, Akhilan Gurumoorthy, Hao Su, and Xiaolong Wang. Cyberdemo: Augmenting simulated human demonstration for real-world dexterous manipulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17952–17963, 2024.
- [28] Jun Wang, Ying Yuan, Haichuan Che, Haozhi Qi, Yi Ma, Jitendra Malik, and Xiaolong Wang. Lessons from learning to spin “pens”. *arXiv preprint arXiv:2407.18902*, 2024.
- [29] Ruicheng Wang, Jialiang Zhang, Jiayi Chen, Yinzen Xu, Puhao Li, Tengyu Liu, and He Wang. Dexgraspnet: A large-scale robotic dexterous grasp dataset for general objects based on simulation. *arXiv preprint arXiv:2210.02697*, 2022.
- [30] Zhenyu Wei, Zhixuan Xu, Jingxiang Guo, Yiwen Hou, Chongkai Gao, Zhehao Cai, Jiayu Luo, and Lin Shao. $\mathcal{D}(\mathcal{R}, \mathcal{O})$ grasp: A unified representation of robot and object interaction for cross-embodiment dexterous grasping. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4982–4988, 2025. doi: 10.1109/ICRA55743.2025.11127754.
- [31] Zhiyuan Wu, Rolandos Alexandros Potamias, Xuyang Zhang, Zhongqun Zhang, Jiankang Deng, and Shan Luo. Cedex: Cross-embodiment dexterous grasp generation at scale from human-like contact representations. *arXiv preprint arXiv:2509.24661*, 2025.
- [32] Lixin Xu, Zhixuan Liu, Zhewei Gui, Jingxiang Guo, Zeyu Jiang, Zhixuan Xu, Chongkai Gao, and Lin Shao. Dexsingrasp: Learning a unified policy for dexterous object singulation and grasping in cluttered environments. *arXiv preprint arXiv:2504.04516*, 2025.
- [33] Mengda Xu, Han Zhang, Yifan Hou, Zhenjia Xu, Linxi Fan, Manuela Veloso, and Shuran Song. Dexumi: Using human hand as the universal manipulation interface for dexterous manipulation. *arXiv preprint arXiv:2505.21864*, 2025.
- [34] Yinzen Xu, Weikang Wan, Jialiang Zhang, Haoran Liu, Zikang Shan, Hao Shen, Ruicheng Wang, Haoran Geng, Yijia Weng, Jiayi Chen, et al. Unidexgrasp: Universal robotic dexterous grasping via learning diverse proposal generation and goal-conditioned policy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4737–4746, 2023.
- [35] Zhixuan Xu, Chongkai Gao, Zhixuan Liu, Gang Yang, Chenrui Tie, Haozhuo Zheng, Haoyu Zhou, Weikun Peng, Debang Wang, Tianrun Hu, et al. Manifoundation model for general-purpose robotic manipulation of contact synthesis with arbitrary objects and robots. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 10905–10912. IEEE, 2024.
- [36] Zhao-Heng Yin and Pieter Abbeel. Lightning grasp: High performance procedural grasp synthesis with contact fields. *arXiv preprint arXiv:2511.07418*, 2025.
- [37] Zhao-Heng Yin, Binghao Huang, Yuzhe Qin, Qifeng Chen, and Xiaolong Wang. Rotating without seeing: Towards in-hand dexterity through touch. *arXiv preprint arXiv:2303.10880*, 2023.
- [38] Ying Yuan, Haichuan Che, Yuzhe Qin, Binghao Huang, Zhao-Heng Yin, Kang-Won Lee, Yi Wu, Soo-Chul Lim, and Xiaolong Wang. Robot synesthesia: In-hand manipulation with visuotactile sensing. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6558–6565. IEEE, 2024.
- [39] Zhecheng Yuan, Tianming Wei, Shuiqi Cheng, Gu Zhang, Yuanpei Chen, and Huazhe Xu. Learning to manipulate anywhere: A visual generalizable framework for reinforcement learning. *arXiv preprint arXiv:2407.15815*, 2024.
- [40] Yanjie Ze, Gu Zhang, Kangning Zhang, Chenyuan Hu, Muhan Wang, and Huazhe Xu. 3d diffusion policy: Generalizable visuomotor policy learning via simple 3d representations. *arXiv preprint arXiv:2403.03954*, 2024.
- [41] Jialiang Zhang, Haoran Liu, Danshi Li, XinQiang Yu, Haoran Geng, Yufei Ding, Jiayi Chen, and He Wang. Dexgraspnet 2.0: Learning generative dexterous grasping in large-scale synthetic cluttered scenes. In *8th Annual Conference on Robot Learning*, 2024.
- [42] Yuanhang Zhang, Tianhai Liang, Zhenyang Chen, Yanjie Ze, and Huazhe Xu. Catch it! learning to catch in flight with mobile dexterous hands. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 14385–14391. IEEE, 2025.