2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

CEDex: Cross-Embodiment Dexterous Grasp Generation at Scale from Human-like Contact Representations

Zhiyuan Wu[1], Rolandos Alexandros Potamias[2], Xuyang Zhang[1], Zhongqun Zhang[3], Jiankang Deng[2], Shan Luo[1]

***Abstract*— Cross-embodiment dexterous grasp synthesis refers to adaptively generating and optimizing grasps for various robotic hands with different morphologies. This capability is crucial for achieving versatile robotic manipulation in diverse environments and requires substantial amounts of reliable and diverse grasp data for effective model training and robust generalization. However, existing approaches either rely on physics-based optimization that lacks human-like kinematic understanding or require extensive manual data collection processes that are limited to anthropomorphic structures. In this paper, we propose CEDex, a novel cross-embodiment dexterous grasp synthesis method at scale that bridges human grasping kinematics and robot kinematics by aligning robot kinematic models with generated human-like contact representations. Given an object's point cloud and an arbitrary robotic hand model, CEDex first generates human-like contact representations using a Conditional Variational Auto-encoder pretrained on human contact data. It then performs kinematic human contact alignment through topological merging to consolidate multiple human hand parts into unified robot components, followed by a signed distance field-based grasp optimization with physics-aware constraints. Using CEDex, we construct the largest cross-embodiment grasp dataset to date, comprising 500K objects across four gripper types with 20M total grasps. Extensive experiments show that CEDex outperforms state-of-the-art approaches and our dataset benefits cross-embodiment grasp learning with high-quality diverse grasps. Project Page: https://georgewuzy.github.io/cedex-website/**

I. Introduction

Humans, by nature, possess remarkable dexterous grasping capabilities that can generate feasible and reliable grasps for given objects while seamlessly adapting to various constraints and generalizing across different finger configurations, *e.g.*, three or four-fingered grasps [1]. In robotic manipulation systems, such a capability of performing versatile grasping is equally important, as grasping is fundamental for complex downstream tasks [2]. However, most existing robotic grasping methods are tailored to specific end-effectors, which hampers their generalization. When faced with new robotic hands featuring novel morphologies, these approaches necessitate costly data collection and time-consuming retraining for each embodiment, which limits their practical deployment and scalability [3]. This limitation indicates the urgent need for a unified model capable of representing and generating grasps across various robotic embodiments. Such a capability, which enables the adaptive generation and optimization of grasps for arbitrary robotic hands, is referred to as **cross-embodiment** grasp synthesis [4]. While recent efforts [4]–[6] have demonstrated the potential of using unified representations to learn cross-embodiment dexterous grasping, training these models requires substantial amounts of reliable and diverse data. The data need to include grasps that satisfy both physical constraints and human-like kinematic plausibility, quantified through metrics such as force closure stability and dynamic simulation success rates. Additionally, these learning-based methods often face data imbalance, with certain grasp orientations being overrepresented while others remain underexplored. Therefore, it is crucial to develop a robust cross-embodiment grasp data synthesis strategy at scale.

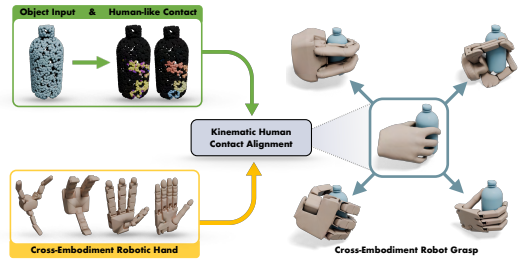

Fig. 1. Given an object point cloud and the generated human-like contact representations, our proposed CEDex method can generate stable and diverse dexterous grasps across various robotic hand embodiments at scale by integrating both human-like kinematics and physics-aware constraints.

Existing cross-embodiment grasp synthesis methods for large-scale data generation primarily rely on grasp optimization, utilizing physical constraints such as force closure to ensure grasp feasibility [3], [7]. However, these physics-based approaches focus solely on static equilibrium and neglect human grasping kinematics, failing to consider the dynamic nature of grasping process, where the hand must plausibly approach and engage with the object. This oversight leads to low success rates in practical scenarios. To incorporate dynamic kinematic considerations, some approaches have attempted to learn from human demonstrations [8], [9], but this process requires significant manual effort for data collection, leading to increased costs. Additionally, these methods often

This work was supported in part by the EPSRC projects "ViTac: Visual-Tactile Synergy for Handling Flexible Materials" (EP/T033517/2) and "TacDiff: Designing Tactile-based Robots via Differentiable Simulations".

[1]Zhiyuan Wu, Xuyang Zhang, and Shan Luo are with Department of Engineering, King's College London, Strand, London, WC2R 2LS, United Kingdom, {zhiyuan.1.wu, xuyang.zhang, shan.luo}@kcl.ac.uk.

[2]Rolandos Alexandros Potamias and Jiankang Deng are with Imperial College London, London, SW7 2AZ, United Kingdom, r.potamias@imperial.ac.uk, jiankangdeng@gmail.com.

[3]Zhongqun Zhang is with College of Software, Nankai University, Tianjin, 300350, China, zhongqunzhang@outlook.com.

necessitate calibrating gripper joints and remapping them to corresponding human hand joints, limiting their effectiveness primarily to anthropomorphic structures [2]. As a result, non-anthropomorphic robotic hands, such as the three-fingered Barrett and Robotiq-3F, still face significant challenges.

Recent advancements in human grasp synthesis have shown that model-based grasp optimization can effectively simulate how a human hand approaches and grasps an object [1], [10], [11]. These methods leverage contact representations to model grasping at the object-centric level and utilize learning-based models for generation, a process that is significantly more efficient than collecting and retargeting from human grasp demonstrations. Inspired by this, we propose CEDex, a novel **C**ross-**E**mbodiment **Dex**terous grasp synthesis method at scale that aligns robot kinematic models with generated human-like contact representations during the grasping process, as illustrated in Fig. 1. Specifically, given a point cloud of an object and the kinematic model of a robotic hand, CEDex consists of two key components: human contact generation and kinematic human contact alignment. Once a human-like contact is generated from a Conditional Variational Auto-encoder (CVAE) model pretrained on human contact data, the kinematic human contact alignment component conducts topological merging to integrate multiple human hand parts into cohesive robot components, aligned with the target robot's kinematic configuration. Subsequently, a signed distance function (SDF)-based grasp optimization with physics-aware constraints is employed to produce robust and diverse grasps that reflect human-like kinematic understanding. Using CEDex, we constructed a large-scale cross-embodiment dexterous grasp dataset, which contains 500K objects and four types of grippers, with 20M grasps in total. Extensive experiments have demonstrated the effectiveness and superiority of our proposed method and dataset over state-of-the-art (SoTA) approaches.

Our main contributions can be summarized as follows:

- We propose CEDex, a novel cross-embodiment dexterous grasp synthesis method at scale that aligns robot kinematic models and generated human-like contact representations to enable effective grasp optimization across diverse robotic hands.
- We construct the largest cross-embodiment grasp dataset to the best of our knowledge, with 500K objects, four types of grippers, and 20M grasps in total.
- Extensive experiments demonstrate that our proposed CEDex outperforms SoTA cross-embodiment grasp synthesis approaches. Furthermore, our constructed large-scale dataset provides high-quality diverse grasps that significantly benefit cross-embodiment grasp learning.

II. Related Works

A. Robotic Dexterous Grasping

Dexterous grasping acts as an essential element for various complex, human-like manipulation tasks. In recent years, data-driven approaches have emerged as a promising direction for dexterous grasping. Shao *et al.* has pioneered UniGrasp [12] for multi-fingered robotic hand grasping by learning generalizable contact point representations. Subsequent works such as GraspTTA [11] and ContactGen [1] employ contact maps as intermediate representations to effectively bridge hand-object geometry and synthesize diverse grasps for human hands through optimization-based approaches. The UniDexGrasp series [13], [14] further introduce universal policies capable of handling thousands of object instances via sophisticated curriculum learning and teacher-student distillation frameworks. Building upon this trend, recent research has shifted from single-embodiment, object-agnostic approaches to both hand-agnostic and object-agnostic methodologies, marking a significant advancement toward truly universal grasping systems. The GeoMatch series [5], [6] and the recent DRO-Grasp [4] achieve cross-embodiment dexterous grasp synthesis by learning geometric correspondences between diverse hand morphologies and object geometries. This development highlights the critical importance of robust and diverse grasp data, as well as the need for grasp synthesis methods at scale.

B. Dexterous Grasp Datasets

Generating dexterous grasp data is foundational for data-driven dexterous grasping. Direct capture of human grasp demonstrations, as exemplified by works including GRAB [15] and RealDex [9], provides the most behaviorally faithful and accurate supervision by retargeting human grasps to generate reliable grasp data. However, this reliance on manual capture and retargeting constrains dataset scale and object diversity. To reduce collection cost and expand coverage, other datasets synthesize grasps in simulator like GraspIt! [16] and Isaac Gym [17]. For instance, DDG [18] synthesizes dexterous grasp poses in GraspIt! [16], and DexGraspNet [19] generates grasps via differentiable force-closure with Isaac Gym physics validation. Recently, works such as GraspXL [20] and DexGrasp Anything [21] further push scalable generation by synthesizing objective-conditioned grasp motions and large-scale dexterous grasps over extensive object sets, *e.g.*, Objaverse [22]. Moving beyond single anthropomorphic settings, the community is pivoting toward cross-embodiment and multi-object generalization of dexterous grasping. DFC [7] first proposed to synthesizes grasps for arbitrary hand morphologies with a differentiable force-closure estimator. Based on it, MultiDex [3] offers a multi-hand dataset to support cross-hand generalization. Recently, Multi-GraspLLM [23] leverages LLM-driven annotations to directly generate final grasp poses. However, current approaches still lack human-like kinematic understanding of grasp dynamics and remain limited at scale (both object diversity and grasp counts), underscoring the need for more robust, diverse, large-scale datasets with rich contact and kinematically informed representations across embodiments.

III. Generating Cross-Embodiment Grasp from Human Contact

As illustrated in Fig. 2, CEDex takes 1) the point cloud $\boldsymbol{O} \in \mathbb{R}^{N\times 3}$ of an object, where N represents point number,

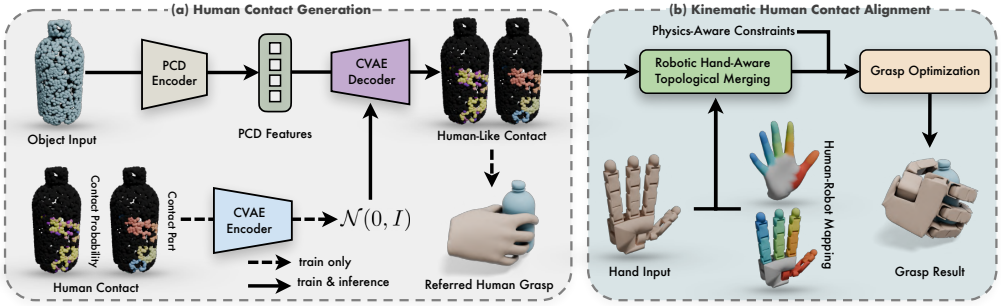

Fig. 2. The cross-embodiment dexterous grasp synthesis pipeline of our CEDex. Given a point cloud of an object and a robotic hand as input, CEDex first generates human-like contact representations using a CVAE model pretrained on human contact data. The kinematic human contact alignment component then performs topological merging to consolidate multiple human hand parts into unified robot components according to the robot's kinematic configuration, followed by a SDF-based grasp optimization with physics-aware constraints to generate robust and diverse grasps with human-like kinematic understanding.

and 2) the kinematic model of an arbitrary robotic hand as input, and generates a physically stable grasp configuration for the given robotic hand. The architecture consists of two key components: a) a human contact generation part that generates reliable human-like contact representations using a CVAE model pretrained on human contact data, and b) a kinematic human contact alignment part that performs topological merging to consolidate multiple human hand parts into unified robot components according to the target robot's kinematic configuration, followed by signed distance field-based grasp optimization with physics-aware constraints to generate robust and diverse grasps with human-like kinematics. The details of each component are elaborated below.

A. Human Contact Generation

We first generate human-like contact representations using a pretrained CVAE model. We employ a MANO hand [24] to represent human hand and divide it into $B = 16$ parts, referring to [1]. The object-centric human contact representations $[\boldsymbol{C}^h, \boldsymbol{P}^h]$ consist of a contact map $\boldsymbol{C}^h \in \mathbb{R}^{N\times 1}$ and a part map $\boldsymbol{P}^h \in \mathbb{R}^{N\times B}$. Each contact value $\boldsymbol{c}_k \in [0, 1]$ in $\boldsymbol{C}^h$ represents the contact probability of point $\boldsymbol{o}_k$ in $\boldsymbol{O}$. The definition and computation of $\boldsymbol{C}^h$ follows ContactOpt [10], where a virtual capsule is placed at each object point $\boldsymbol{o}_k$, and $\boldsymbol{c}_k$ is set to 1 if any point in the human hand point cloud lies inside the capsule and otherwise smoothly decays with distance. The one-hot part map $\boldsymbol{P}^h$ represents which part of the hand touches the object, indicating the hand part label in $\{1, \ldots, B\}$ in contact with each object point. The direction map from [1] is not considered, as the substantial shape differences between human fingers and robotic gripper fingers result in significantly different contact directions, making direct human-to-robot transfer ineffective. We employ CVAE to model the conditional probabilities $p(\boldsymbol{C}^h, \boldsymbol{P}^h|\boldsymbol{O})$ as:

$$p(\boldsymbol{C}^h, \boldsymbol{P}^h|\boldsymbol{O}) = p(\boldsymbol{P}^h|\boldsymbol{C}^h, \boldsymbol{O})p(\boldsymbol{C}^h|\boldsymbol{O}), \tag{1}$$

where $\boldsymbol{C}^h$ is conditioned on object $\boldsymbol{O}$ and $\boldsymbol{P}^h$ is additionally conditioned on $\boldsymbol{C}^h$. We use two Point Cloud [25] decoders $\mathcal{D}_c$ and $\mathcal{D}_p$ to predict $\boldsymbol{C}^h$ and $\boldsymbol{P}^h$ as:

$$\boldsymbol{C}^h = \mathcal{D}_c(\boldsymbol{z}_c, \boldsymbol{F}^o), \tag{2}$$

and

$$\boldsymbol{P}^h = \mathcal{D}_p(\boldsymbol{z}_p, \boldsymbol{C}^h, \boldsymbol{F}^o), \tag{3}$$

where $\boldsymbol{F}^o$ represents feature maps extracted from $\boldsymbol{O}$ via PointNet++ [26], and $\boldsymbol{z}_c \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ and $\boldsymbol{z}_p \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ represent latent codes sampled from Gaussian distributions. We pretrain our human contact generation model with:

$$\mathcal{L}^{recon} = |\boldsymbol{C}^h - \hat{\boldsymbol{C}}^h| + \lambda_p \mathcal{L}_{CE}(\boldsymbol{P}^h, \hat{\boldsymbol{P}}^h) + \lambda_{KL}\mathcal{L}^{KL}, \tag{4}$$

where $|\boldsymbol{C}^h - \hat{\boldsymbol{C}}^h|$ is the L1 loss for contact maps, $\mathcal{L}_{CE}(\boldsymbol{P}^h, \hat{\boldsymbol{P}}^h)$ is the cross-entropy loss for part maps, and a KL regularization loss $\mathcal{L}^{KL}$ [27]. The model is pretrained on the GRAB [15] and YCB Affordance [28] datasets.

B. Kinematic Human Contact Alignment

Directly transferring human-like contact representations to robotic hands poses significant challenges due to fundamental structural differences between human hands and robotic grippers. These differences manifest in two key aspects: 1) varying finger numbers across different robotic embodiments, and 2) distinct joint configurations within a single finger. Such morphological disparities make direct keypoint remapping or grasp optimization with raw human-like contact representations unavailable for most robotic hands. However, humans demonstrate remarkable adaptability in performing stable grasps using different finger combinations, where multiple fingers can collectively fulfill the mechanical role of a single finger across different configurations. This observation inspires our approach to perform topological merging of contact representations at the object level. We adaptively consolidate multiple human hand parts, such as the ring and pinky fingers, into fewer robotic components, such as a single gripper finger, based on the target robot's kinematic configuration. This strategy facilitates effective kinematic alignment between human grasping mechanics and the capabilities of robotic hands, enabling robots to perform human-like kinematics in grasping.

We predefine a kinematic human-robot mapping to align human hand parts with robotic hand components according to their configurations, as shown in Fig. 3. Since human hands have more complex part structures than robotic hands, this mapping requires merging multiple human parts into

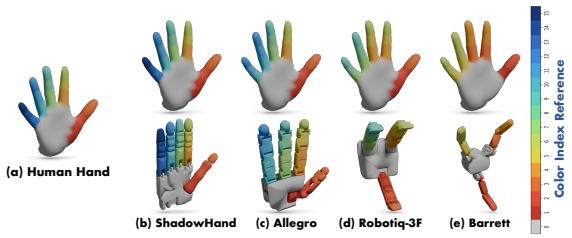

Fig. 3. Human-robot mapping to align human hand parts with robotic hand components. (a) Original human hand (b) Shadow hand (c) Allegro (d) Robotiq-3F (e) Barrett. Color index reference for visualization is provided.

single robot parts. To implement this predefined mapping, we develop a geometric-based topological merging approach that consolidates contact representations from multiple human parts into unified robot contact representations at the object level. Given human-like contact representations $[\boldsymbol{C}^h, \boldsymbol{P}^h]$ with contact map $\boldsymbol{C}^h \in \mathbb{R}^{N\times 1}$ and part map $\boldsymbol{P}^h \in \mathbb{R}^{N\times B}$, our goal is to generate robot-specific contact representations $[\boldsymbol{C}^r, \boldsymbol{P}^r]$, where $\boldsymbol{C}^r \in \mathbb{R}^{N\times 1}$ is the robot contact map and $\boldsymbol{P}^r \in \mathbb{R}^{N\times B'}$ is the robot part map with B' robot components. When human parts $\{b_i, b_j\} \subset \{1, ..., B\}$ need to be merged into a single robot component $b_m \in \{1, ..., B'\}$, we first extract the two contact parts $\boldsymbol{P}^i = \boldsymbol{P}^h[:,\ i]$ and $\boldsymbol{P}^j = \boldsymbol{P}^h[:,\ j]$. We then obtain the corresponding part-specific contact maps through element-wise multiplication as:

$$\boldsymbol{C}^i = \boldsymbol{C}^h \odot \boldsymbol{P}^i, \quad \boldsymbol{C}^j = \boldsymbol{C}^h \odot \boldsymbol{P}^j, \tag{5}$$

where $\boldsymbol{C}^i, \boldsymbol{C}^j \in \mathbb{R}^{N\times 1}$ represent the contact values for human parts i and j respectively, and $\odot$ denotes element-wise multiplication. For each point $\boldsymbol{o}_x$ with its contact value $\boldsymbol{c}_x^i > 0$, the projection direction is designed to consolidate the spatially separated contacts from two human parts into a unified region that can be effectively accessed by a single robot component, where the direction towards the centroid between the two parts ensures optimal geometric coverage for the merged contact. We compute the projecting direction as:

$$\boldsymbol{v}_x = \frac{\frac{\overrightarrow{\boldsymbol{M}^o\boldsymbol{o}_x}}{|\overrightarrow{\boldsymbol{M}^o\boldsymbol{o}_x}|} + \frac{\overrightarrow{\boldsymbol{M}^o\boldsymbol{M}^j}}{|\overrightarrow{\boldsymbol{M}^o\boldsymbol{M}^j}|}}{\left\| \frac{\overrightarrow{\boldsymbol{M}^o\boldsymbol{o}_x}}{|\overrightarrow{\boldsymbol{M}^o\boldsymbol{o}_x}|} + \frac{\overrightarrow{\boldsymbol{M}^o\boldsymbol{M}^j}}{|\overrightarrow{\boldsymbol{M}^o\boldsymbol{M}^j}|} \right\|}, \tag{6}$$

where $\boldsymbol{M}^o$ represents the object mass centroid and $\boldsymbol{M}^i$ and $\boldsymbol{M}^j$ represent the mass centroids of $\boldsymbol{C}^i$ and $\boldsymbol{C}^j$, computed as:

$$\boldsymbol{M}^o = \frac{1}{N}\sum_{k=1}^{N} \boldsymbol{o}_k, \tag{7}$$

$$\boldsymbol{M}^{i,j} = \frac{\sum_{k=1}^{N} \boldsymbol{c}_k^{i,j} \cdot \boldsymbol{o}_k}{\sum_{k=1}^{N} \boldsymbol{c}_k^{i,j}}, \tag{8}$$

where $\boldsymbol{c}_k^i$ and $\boldsymbol{c}_k^j$ represent the contact value of k-th point in $\boldsymbol{C}^i$ and $\boldsymbol{C}^j$, and $\boldsymbol{o}_k$ represent the k-th point in the object point cloud $\boldsymbol{O}$. The remapped position $\boldsymbol{o}_x{}'$ is then determined as the nearest point in $\boldsymbol{O}$ to the ray from $\boldsymbol{M}^o$ in direction $\boldsymbol{v}_x$. By applying the same remapping process to all points in human part b_j towards b_i, we obtain the symmetric remapped contact pair. Next, given the two remapped contact pairs $[\boldsymbol{C}^{i'}, \boldsymbol{P}^{i'}]$ and $[\boldsymbol{C}^{j'}, \boldsymbol{P}^{j'}]$ with $\boldsymbol{C}^{i'}, \boldsymbol{C}^{j'}, \boldsymbol{P}^{i'}, \boldsymbol{P}^{j'} \in \mathbb{R}^{N\times 1}$, the merged contact $[\boldsymbol{C}^m, \boldsymbol{P}^m]$ is obtained by taking the union of the two remapped contact pairs as:

$$\boldsymbol{c}_k^m = \begin{cases} \boldsymbol{c}_k^{i'} + \boldsymbol{c}_k^{j'} & \text{if } \boldsymbol{p}_k^{i'} > 0 \text{ and } \boldsymbol{p}_k^{j'} > 0 \\ \boldsymbol{c}_k^{i'} & \text{if } \boldsymbol{p}_k^{i'} > 0 \text{ and } \boldsymbol{p}_k^{j'} = 0 \\ \boldsymbol{c}_k^{j'} & \text{if } \boldsymbol{p}_k^{i'} = 0 \text{ and } \boldsymbol{p}_k^{j'} > 0 \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

where $\boldsymbol{p}_k^{i'}$ and $\boldsymbol{p}_k^{j'}$ represent the remapped part assignment values at point k for parts i and j respectively, and $\boldsymbol{c}_k^{i'}$ and $\boldsymbol{c}_k^{j'}$ are the corresponding remapped contact values. For cases where more than two human parts need to be merged into a single robot component, we repeat this pairwise merging process iteratively.

Following the predefined kinematic human-robot mapping, we apply the above merging operations to all required human part combinations to generate the robot-specific contact representations $[\boldsymbol{C}^r, \boldsymbol{P}^r]$. We then employ signed distance field-based contact optimization for human-robot contact alignment. The contact loss $\mathcal{L}_c$ is formulated as:

$$\mathcal{L}_c = \sum_{k=1}^{N} \boldsymbol{c}_k^r \sum_{b=1}^{B'} \boldsymbol{p}_{k,b}^r \cdot |\text{SDF}_b(\boldsymbol{o}_k)|, \tag{10}$$

where $\boldsymbol{c}_k^r$ is the contact value at object point $\boldsymbol{o}_k$, $\boldsymbol{p}_{k,b}^r$ indicates the assignment of point $\boldsymbol{o}_k$ to robot part b, and $\text{SDF}_b(\boldsymbol{o}_k)$ represents the signed distance from robotic hand part b to object point $\boldsymbol{o}_k$.

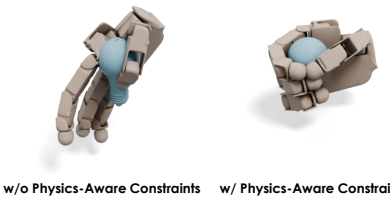

Fig. 4. An example of Allegro hand grasping a light bulb, demonstrating the robustness and stability provided by physics-aware constraints.

C. Physics-Aware Constraints

While human grasping primarily relies on natural kinematic principles, robot grasp synthesis requires additional stability guarantees. Therefore, beyond the contact loss $\mathcal{L}_c$ in human-robot contact alignment, we follow [21] and incorporate additional physics-aware constraints to enhance the robustness of synthesized grasps. We incorporate three key physical constraints: Surface Pulling Force (SPF) loss [13] that encourages proximity between robot parts and the object surface, External-penetration Repulsion Force (ERF) loss [3] that prevents hand-object collisions, and Self-penetration Repulsion Force (SRF) loss [13] that maintains realistic hand geometry by preventing finger self-intersections:

$$\mathcal{L}_{SPF} = \frac{\sum_{k\in\mathcal{S}} \sqrt{d_k^o}}{|\mathcal{S}| + \eta}, \tag{11}$$

TABLE I

COMPARISON OF DEXTEROUS GRASP DATASETS. OUR DATASET ACHIEVES THE LARGEST SCALE ON CROSS-EMBODIMENT DEXTEROUS HANDS TO THE BEST OF OUR KNOWLEDGE, WITH CONSIDERATIONS OF BOTH HUMAN-LIKE KINEMATICS AND KINEMATIC PHYSICAL AWARENESS.

Dataset	Hand Type	Object	Grasp	Human-Like Kinematics	Kinematic Physical Awareness
GRAB [15]	Human	51	1.64M	✓	✗
DexYCB [8]	Human	20	1K	✓	✗
DDG [18]	Single	565	6.9K	✗	✓
DexGraspNet [19]	Single	5.3K	1.32M	✗	✓
UniDexGrasp [13]	Single	5.5K	1.12M	✗	✓
DexGrasp Anything [21]	Single	15.6K	3.4M	✗	✓
RealDex [9]	Single	59K	52	✓	✗
GraspXL [20]	Anthropomorphic	**500K**	-	✗	✗
MultiDex [3]	**Cross-Embodiment**	58	436K	✗	✗
Multi-GraspLLM [23]	**Cross-Embodiment**	2.1K	140K	✗	✗
CEDex (Ours)	**Cross-Embodiment**	**500K**	**20M**	✓	✓

$$\mathcal{L}_{ERF} = \frac{1}{B'}\sum_{b=1}^{B'}\max_{k\in\mathcal{H}_b} |\min(0, \text{SDF}_{obj}(\boldsymbol{h}_k))|, \quad (12)$$

and

$$\mathcal{L}_{SRF} = \frac{1}{B'}\sum_{b=1}^{B'}\sum_{i,j\in\mathcal{H}_b, i\neq j} \max(0, d_{th} - ||\boldsymbol{h}_i - \boldsymbol{h}_j||), \quad (13)$$

where d_k^o represents the distance from robotic hand point $\boldsymbol{h}_k$ to the object surface, $\mathcal{S}$ is the set of hand points within threshold distance, $\mathcal{H}_b$ denotes the set of hand points belonging to robot part b, $\text{SDF}_{obj}(\boldsymbol{h}_k)$ represents the signed distance from the object surface to hand point $\boldsymbol{h}_k$, and d_{th} is the self-collision threshold distance. An example is showing in Fig. 4.

IV. CEDEX DATASET

Using the proposed CEDex method for synthesizing robust and diverse cross-embodiment dexterous grasps, we construct a large-scale cross-embodiment dexterous grasp dataset. For robotic hand selection, we choose four diverse robotic hands ranging from three to five fingers: Barrett Hand, Robotiq-3F, Allegro, and Shadow hand. We exclude two-finger grippers as we do not regard them as dexterous robotic hands, and two-finger grasps cannot achieve stable multi-directional force application and fail to pass rigorous multi-directional stability tests [3]. For object selection, we utilize 58 real-world objects from ContactDB [29] and YCB [30] datasets following [3], and 503,409 synthesized objects from the large-scale Objaverse [22] dataset following [20].

We generate our dataset using eight 32GB NVIDIA Tesla V100 GPUs. We first pretrain the human contact generation CVAE model using human grasp data from GRAB [15] and YCB Affordance [28] datasets on the 58 real-world objects. Then for each real-world object, we generate 64 human-like contact pairs consisting of contact maps and part maps. For each contact pair, we perform grasp optimization with 64 randomly sampled initial gripper wrist poses, which ensures grasp diversity even for each single human-like contact pairs, and select the top-16 grasps with the lowest optimization energy scores. This process yields 59,392 grasp candidates for each robotic hand, totaling 237,568 grasps, which is comparable to the scale of MultiDex [3]. To significantly expand the dataset scale, we process the 500K synthesized Objaverse objects with our CEDex. For each Objaverse object, we generate 4 human-like contact pairs, and for each contact pair, we perform grasp optimization with 16 randomly sampled initial gripper wrist poses, selecting the top-4 grasps with the lowest energy scores. This results in 8M grasp candidates for each robotic hand, totaling 32M grasps across all hands. We then apply rigorous filtering using Isaac Gym [17] based on comprehensive stability evaluation metrics detailed in Sec. V-A. After filtering, our final dataset contains 20M high-quality grasps. We provide both generated grasping poses and final grasping poses after Isaac Gym validation in our dataset.

As demonstrated in Tab. I, our CEDex dataset is the largest cross-embodiment dexterous grasp dataset to date in terms of both object diversity and grasp quantity. Crucially, we are the first dataset to simultaneously incorporate 1) human-like kinematics that leverages natural human grasping principles and contact patterns during the whole grasp synthesis process and 2) kinematic physical awareness that enforces physical constraints of hand kinematics, addressing limitations of existing datasets that typically focus on only one aspect.

V. EXPERIMENTS

A. Experimental Setup

We evaluate our CEDex framework on an 80GB NVIDIA A100 GPU. Following the evaluation protocol established in [3], we conduct experiments on a test set comprising 10 representative unseen daily objects from the ContactDB [29] and YCB [30] datasets. Our evaluation spans three diverse robotic hands with varying finger configurations: Barrett (3 fingers), Allegro (4 fingers), and Shadow hand (5 fingers). Note that we do not evaluate baseline methods on Robotiq-3F as they do not support this gripper, though we provide CEDex results on it in our analysis.

For implementation, we pretrain our CVAE model to generate human-like contact representations using training parameters from [1] on 58 real-world objects from the MultiDex dataset [3]. Subsequently, we perform grasp optimization over 200 iterations for final grasp synthesis.

We provide comprehensive comparisons against established cross-embodiment dexterous grasp synthesis baselines,

TABLE II

QUANTITATIVE RESULTS OF OUR CEDEX COMPARED WITH DIFFERENT CROSS-EMBODIMENT DEXTEROUS GRASP SYNTHESIS BASELINES ACROSS THREE ROBOTIC HANDS FROM THREE TO FIVE FINGERS: BARRETT, ALLEGRO, AND SHADOW HAND. WE EVALUATE SUCCESS RATE AND DIVERSITY.

Method	Success Rate (%) ↑				Diversity (rad.) ↑			
	Barrett	Allegro	ShadowHand	**Average**	Barrett	Allegro	ShadowHand	**Average**
DFC [7]	86.3	76.2	58.8	73.8	0.532	0.454	0.435	0.474
GenDexGrasp [3]	67.0	51.0	54.2	57.4	0.488	0.389	0.318	0.398
GeoMatch [5]	60.0	-	67.5	63.8	0.259	-	0.235	0.247
GeoMatch++ [6]	77.5	-	70.0	73.8	0.378	-	0.184	0.281
DRO-Grasp [4]	87.3	**92.3**	83.0	87.5	0.513	0.397	**0.441**	0.450
CEDex (Ours)	**93.1**	88.1	**85.0**	**88.7**	**0.624**	**0.473**	0.438	**0.512**

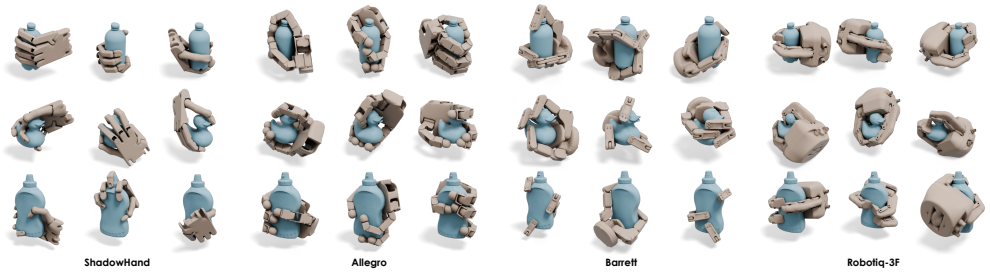

Fig. 5. Visualization of grasp results synthesized by our proposed CEDex, where our method generates robust and diverse grasps.

including optimization-based methods DFC [7] and GenDexGrasp [3], as well as learning-based approaches GeoMatch [6], GeoMatch++ [6], and DRO-Grasp [4]. To evaluate grasp synthesis performance, we employ success rate and diversity, defined as:

- **Success Rate**: We evaluate grasping success by applying external forces to the object and measuring its displacement. Using the Isaac Gym simulator [17], a simple grasp controller executes the predicted grasps [4]. Following the metric definition in [3], we sequentially apply forces along six orthogonal directions for 1 second each. A grasp is considered successful if the object's displacement remains below 2 cm once all forces are applied.
- **Diversity**: Grasp diversity is quantified by computing the standard deviation of joint configurations across all successful grasps, including the 6-DoF wrist pose and finger joint angles. Higher standard deviation indicates greater diversity in the generated grasp configurations.

B. Comparison with SoTAs

As shown in Tab. II, our quantitative results reveal distinct trade-offs between different approaches. Learning-based cross-embodiment methods (GeoMatch [5], GeoMatch++ [6], DRO-Grasp [4]) generally achieve higher success rates than optimization-based approaches (DFC [7], GenDexGrasp [3]), particularly on complex-structural robotic hands like Shadow hand. However, these data-driven methods suffer from limited diversity due to training data imbalances, where certain grasp orientations may be overrepresented in the dataset while others remain underexplored. This limitation is evident from diversity metrics, where optimization-based methods demonstrate superior diversity compared to learning-based counterparts.

In contrast, CEDex achieves the best performance across both success rate (88.7%) and diversity (0.512 rad.). This superior performance comes from two key advantages. First, CEDex improves grasp success rates with human-like kinematics. Unlike optimization-based methods that rely solely on physical constraints, our approach leverages kinematic human contact alignment by employing a learning-based CVAE model to generate kinematically grounded human-like contact representations with functional understanding that enables better performance on complex hand structures. Compared to optimization-based methods (DFC [7], GenDexGrasp [3]), our CEDex improves success rates by 26.2% to 30.8% on Shadow hand. In addition, CEDex improves grasp diversity through data-agnostic optimization. It avoids training on robotic grasp datasets and performs grasp optimization from spatially uniform initial poses around the object, which eliminates the aforementioned data imbalance problems. Compared to learning-based methods (GeoMatch [5], GeoMatch++ [6], DRO-Grasp [4]), our CEDex improves average diversity by 12.1% to 51.8%. Note that while Robotiq-3F results are not included in Tab. I due to the unavailability of baseline comparisons, CEDex achieves **91.9%** success rate and **0.401** diversity on it.

To evaluate the practical applicability of CEDex, we measure the computational time and GPU memory requirements for complete grasp synthesis, from 3D object and gripper input to final grasp pose generation. CEDex achieves remarkable efficiency with an average inference time of **7.8s** and **684 MiB** GPU memory per synthesis session, generating 64 grasps at **0.12s** per grasp. Compared to existing methods (batch size 64), CEDex demonstrates substantial speedup: DFC [7] requires 1800s per batch (230 times slower), GenDexGrasp [3] needs 19.7s per batch (2.5 times slower), and DRO-Grasp [4] takes 0.65s per single grasp with 4 GB GPU memory (batch size 1) (5.4 times slower per grasp), highlighting CEDex's superior computational efficiency for

TABLE III

QUANTITATIVE RESULTS OF DRO-GRASP [4] TRAINED ON OUR CEDEX DATASET COMPARED WITH ITS DATASET BASELINE ACROSS THREE ROBOTIC HANDS FROM THREE TO FIVE FINGERS: BARRETT, ALLEGRO, AND SHADOW HAND. WE EVALUATE SUCCESS RATE AND DIVERSITY.

Dataset	Success Rate (%) ↑				Diversity (rad.) ↑			
	Barrett	Allegro	ShadowHand	**Average**	Barrett	Allegro	ShadowHand	**Average**
MultiDex [3]	87.3	92.0	83.0	87.5	0.513	0.397	**0.441**	0.450
CEDex (Ours)	**91.2**	**95.4**	**86.4**	**91.0**	**0.524**	**0.419**	0.436	**0.460**

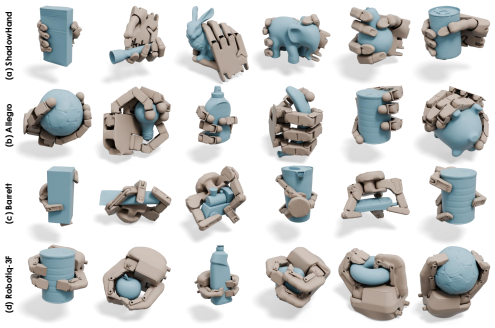

Fig. 6. Visualization of more grasp results synthesized by our CEDex.

practical robotic applications.

C. Grasp Synthesis Visualization

We provide qualitative results of our CEDex in Fig. 5, where our method generates robust and diverse grasps. It is worth mentioning that the multiple grasps shown for each object-hand pair in the figure are derived from a single generated human-like contact representation using different initial wrist poses. This illustrates how our method achieves grasp diversity by varying the wrist configuration while balancing human kinematic understanding and physical feasibility through physics-aware constraints. As a result, different initial wrist poses lead to diverse grasp outcomes, even from the same generated human-like contact representation. More results are available in Fig. 6, our demo, and our website.

D. Training Learning-based Networks on CEDex dataset

To validate the effectiveness of our proposed CEDex dataset, we train the learning-based cross-embodiment grasp synthesis network DRO-Grasp [4] using our CEDex dataset. For direct comparison of data quality, we evaluate on the ContactDB and YCB object sets, which have the same scale as the MultiDex dataset [3] originally used by DRO-Grasp. The results are reported in Tab. III, where DRO-Grasp trained on our CEDex dataset achieves superior metrics across all metrics except for diversity on Shadow hand when compared to the baseline trained on MultiDex. It shows 3.5% and 2.2% improvement on success rate and diversity, demonstrating the effectiveness and superiority of our constructed dataset for cross-embodiment grasp learning.

E. Ablation Study

Kinematic Human Contact Alignment. We evaluate our kinematic human contact alignment against the hand-agnostic contact loss from [3] as our baseline, where both approaches incorporate our physics-aware constraints to ensure fair comparison. Quantitative results are reported in Tab. IV, where our method with kinematic alignment achieves a success rate of 89.3% (a 61.1% improvement) and a diversity score of 0.484 (a 23.1% improvement). This substantial improvement stems from our kinematic alignment's ability to effectively adapt human-like contact representations to robot-specific morphologies, while the hand-agnostic baseline neglects critical embodiment differences that are essential for successful cross-embodiment transfer.

TABLE IV

ABLATION STUDY ON KINEMATIC HUMAN CONTACT ALIGNMENT ON FOUR ROBOTIC HANDS FROM THREE TO FIVE FINGERS: BARRETT, ROBOTIQ-3F, ALLEGRO, AND SHADOW HAND. WE REPORT AVERAGE SUCCESS RATE AND DIVERSITY.

Method	Success Rate (%) ↑	Diversity (rad.) ↑
w/o Alignment	27.7	0.372
w/ Alignment	**89.3**	**0.484**

TABLE V

ABLATION STUDY ON KINEMATIC PHYSICS-AWARE CONSTRAINTS ON FOUR ROBOTIC HANDS FROM THREE TO FIVE FINGERS: BARRETT, ROBOTIQ-3F, ALLEGRO, AND SHADOW HAND. WE REPORT AVERAGE SUCCESS RATE AND DIVERSITY.

SPF	ERF	SRF	Success Rate (%) ↑	Diversity (rad.) ↑
-	-	-	30.9	0.413
✓	-	-	86.7	0.405
✓	✓	-	87.2	0.412
✓	✓	✓	**89.3**	**0.484**

Physics-Aware Constraints. We systematically evaluate the contribution of our physics-aware constraints through incremental ablation studies. Starting from the proposed method without any constraints, we progressively add the three physics-aware constraints. As reported in Tab. V, each constraint contributes to the overall performance, especially the surface pulling loss (SPF) [13], which enhances the success rate from 30.9% to 86.7%. The results underscores the importance of physics-aware constraints in bridging the gap between kinematic alignment and physical realizability, ensuring that human-inspired grasps remain feasible when executed on diverse robotic embodiments.

F. Real-World Validation

We validate our CEDex in real world using a UR5-e robot equipped with a Leap Hand [31], as shown in Fig. 7, showcasing the effectiveness of our proposed CEDex methods in dexterous grasp

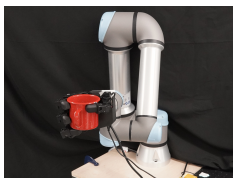

Fig. 7. Real-world setup.

synthesis. More results are in our demo and website.

VI. Conclusion

In this paper, we present CEDex, a novel cross-embodiment dexterous grasp synthesis method that bridges human grasping kinematics and robot kinematics through kinematic human contact alignment. By generating human-like contact representations and performing topological merging followed by SDF-based optimization, CEDex enables large-scale synthesis of physically feasible and kinematically plausible grasps across diverse robotic embodiments. We construct the largest cross-embodiment grasp dataset to date with 500K objects and 20M grasps across four gripper types, demonstrating superior performance over existing approaches. Extensive experiments validate the effectiveness of our approach and demonstrate that our method and dataset significantly advance the state-of-the-art in cross-embodiment dexterous grasping. For future work, we will focus on leveraging our large-scale dataset to develop advanced learning-based approaches, *e.g.*, diffusion models, to further enhance cross-embodiment dexterous grasping capabilities.

References

[1] S. Liu, Y. Zhou, J. Yang, S. Gupta, and S. Wang, "Contactgen: Generative contact modeling for grasp generation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 20 609–20 620.

[2] Q. She, S. Zhang, Y. Ye, R. Hu, and K. Xu, "Learning cross-hand policies of high-dof reaching and grasping," in *European Conference on Computer Vision*. Springer, 2024, pp. 269–285.

[3] P. Li, T. Liu, Y. Li, Y. Geng, Y. Zhu, Y. Yang, and S. Huang, "Gendexgrasp: Generalizable dexterous grasping," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 8068–8074.

[4] Z. Wei, Z. Xu, J. Guo, Y. Hou, C. Gao, Z. Cai, J. Luo, and L. Shao, "D (r, 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)*. IEEE, 2025.

[5] M. Attarian, M. A. Asif, J. Liu, R. Hari, A. Garg, I. Gilitschenski, and J. Tompson, "Geometry matching for multi-embodiment grasping," in *Conference on Robot Learning*. PMLR, 2023, pp. 1242–1256.

[6] Y. Wei, M. Attarian, and I. Gilitschenski, "Geomatch++: Morphology conditioned geometry matching for multi-embodiment grasping," in *CoRL Workshop on Learning Robot Fine and Dexterous Manipulation: Perception and Control*, 2024.

[7] T. Liu, Z. Liu, Z. Jiao, Y. Zhu, and S.-C. Zhu, "Synthesizing diverse and physically stable grasps with arbitrary hand structures using differentiable force closure estimator," *IEEE Robotics and Automation Letters*, vol. 7, no. 1, pp. 470–477, 2021.

[8] Y.-W. Chao, W. Yang, Y. Xiang, P. Molchanov, A. Handa, J. Tremblay, Y. S. Narang, K. Van Wyk, U. Iqbal, S. Birchfield *et al.*, "Dexycb: A benchmark for capturing hand grasping of objects," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 9044–9053.

[9] Y. Liu, Y. Yang, Y. Wang, X. Wu, J. Wang, Y. Yao, S. Schwertfeger, S. Yang, W. Wang, J. Yu *et al.*, "Realdex: towards human-like grasping for robotic dexterous hand," in *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, 2024, pp. 6859–6867.

[10] P. Grady, C. Tang, C. D. Twigg, M. Vo, S. Brahmbhatt, and C. C. Kemp, "Contactopt: Optimizing contact to improve grasps," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 1471–1481.

[11] H. Jiang, S. Liu, J. Wang, and X. Wang, "Hand-object contact consistency reasoning for human grasps generation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 11 107–11 116.

[12] L. Shao, F. Ferreira, M. Jorda, V. Nambiar, J. Luo, E. Solowjow, J. A. Ojea, O. Khatib, and J. Bohg, "Unigrasp: Learning a unified model to grasp with multifingered robotic hands," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2286–2293, 2020.

[13] Y. Xu, W. Wan, J. Zhang, H. Liu, Z. Shan, H. Shen, R. Wang, H. Geng, Y. Weng, J. 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*, 2023, pp. 4737–4746.

[14] W. Wan, H. Geng, Y. Liu, Z. Shan, Y. Yang, L. Yi, and H. 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*, 2023, pp. 3891–3902.

[15] O. Taheri, N. Ghorbani, M. J. Black, and D. Tzionas, "Grab: A dataset of whole-body human grasping of objects," in *European conference on computer vision*. Springer, 2020, pp. 581–600.

[16] A. T. Miller and P. K. Allen, "Graspit! a versatile simulator for robotic grasping," *IEEE Robotics & Automation Magazine*, vol. 11, no. 4, pp. 110–122, 2004.

[17] J. Liang, V. Makoviychuk, A. Handa, N. Chentanez, M. Macklin, and D. Fox, "Gpu-accelerated robotic simulation for distributed reinforcement learning," in *Conference on Robot Learning*. PMLR, 2018, pp. 270–282.

[18] M. Liu, Z. Pan, K. Xu, K. Ganguly, and D. Manocha, "Deep differentiable grasp planner for high-dof grippers," in *Robotics: Science and Systems*, 2020.

[19] R. Wang, J. Zhang, J. Chen, Y. Xu, P. Li, T. Liu, and H. Wang, "Dexgraspnet: A large-scale robotic dexterous grasp dataset for general objects based on simulation," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 11 359–11 366.

[20] H. Zhang, S. Christen, Z. Fan, O. Hilliges, and J. Song, "Graspxl: Generating grasping motions for diverse objects at scale," in *European Conference on Computer Vision*. Springer, 2024, pp. 386–403.

[21] Y. Zhong, Q. Jiang, J. Yu, and Y. Ma, "Dexgrasp anything: Towards universal robotic dexterous grasping with physics awareness," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 22 584–22 594.

[22] M. Deitke, D. Schwenk, J. Salvador, L. Weihs, O. Michel, E. VanderBilt, L. Schmidt, K. Ehsani, A. Kembhavi, and A. Farhadi, "Objaverse: A universe of annotated 3d objects," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 13 142–13 153.

[23] H. Li, W. Mao, W. Deng, C. Meng, H. Fan, T. Wang, Y. Osamu, P. Tan, H. Wang, and X. Deng, "Multi-grasplm: A multimodal llm for multi-hand semantic guided grasp generation," *arXiv preprint arXiv:2412.08468*, 2024.

[24] J. Romero, D. Tzionas, and M. J. Black, "Embodied hands: modeling and capturing hands and bodies together," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 6, pp. 1–17, 2017.

[25] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 652–660.

[26] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," *Advances in neural information processing systems*, vol. 30, 2017.

[27] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *Int. Conf. on Learning Representations*, 2013.

[28] E. Corona, A. Pumarola, G. Alenya, F. Moreno-Noguer, and G. Rogez, "Ganhand: Predicting human grasp affordances in multi-object scenes," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 5031–5041.

[29] S. Brahmbhatt, C. Ham, C. C. Kemp, and J. Hays, "Contactdb: Analyzing and predicting grasp contact via thermal imaging," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 8709–8719.

[30] B. Calli, A. Singh, A. Walsman, S. Srinivasa, P. Abbeel, and A. M. Dollar, "The ycb object and model set: Towards common benchmarks for manipulation research," in *2015 international conference on advanced robotics (ICAR)*. IEEE, 2015, pp. 510–517.

[31] K. Shaw, A. Agarwal, and D. Pathak, "Leap hand: Low-cost, efficient, and anthropomorphic hand for robot learning," *Robotics: Science and Systems*, 2023.