IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

UniFucGrasp: Human-Hand-Inspired Unified Functional Grasp Annotation Strategy and Dataset for Diverse Dexterous Hands

Haoran Lin[1,2,*], Wenrui Chen[1,2,†], Xianchi Chen[1,*], Fan Yang[1,2], Qiang Diao[1], Wenxin Xie[3], Sijie Wu[3], Kailun Yang[1,2], Maojun Li[3], and Yaonan Wang[1,2]

***Abstract*—Dexterous grasp datasets are vital for embodied intelligence, but mostly emphasize grasp stability, ignoring functional grasps needed for tasks like opening bottle caps or holding cup handles. Most rely on bulky, costly, and hard-to-control high-DOF Shadow Hands. Inspired by the human hand's underactuated mechanism, we establish UniFucGrasp, a universal functional grasp annotation strategy and dataset for multiple dexterous hand types. Based on biomimicry, it maps natural human motions to diverse hand structures and uses geometry-based force closure to ensure functional, stable, human-like grasps. This method supports low-cost, efficient collection of diverse, high-quality functional grasps. Finally, we establish the first multi-hand functional grasp dataset and provide a synthesis model to validate its effectiveness. Experiments on the UFG dataset, IsaacSim, and complex robotic tasks show that our method improves functional manipulation accuracy and grasp stability, demonstrates improved adaptability across multiple robotic hands, helping to alleviate annotation cost and generalization challenges in dexterous grasping. The project page is at https://haochen611.github.io/UFG.**

I. Introduction

Functional dexterous grasping has attracted increasing attention due to its critical role in enabling robots to perform complex tasks such as tool use and human-like daily activities [1], [2], [3], [4]. Unlike conventional stable grasps, functional grasps require not only secure holding but also task-specific coordination between the hand and the object [1]. For example, a hammer is typically grasped by the handle during use, but may be held by the head when being handed to another person. These nuanced differences highlight the need for fine-grained, semantically meaningful hand-object pose alignment.

Despite their importance, the development of functional dexterous grasping has long been hindered by the lack of large-scale annotated datasets. This is mainly due to the high Degrees of Freedom (DoF) in dexterous hands, which makes the annotation process extremely costly and complex. Early studies [5], [6], [7], [8] have focused primarily on grasp stability, overlooking the role of task-specific semantic alignment in manipulation.

In the vision community, a common practice is to use the MANO hand model [10], [11], [12], [13] for synthesizing grasp motions. However, due to the absence of physical embodiment, the MANO model must be post-processed to map its output to real robotic hands, limiting its applicability in embodied scenarios.

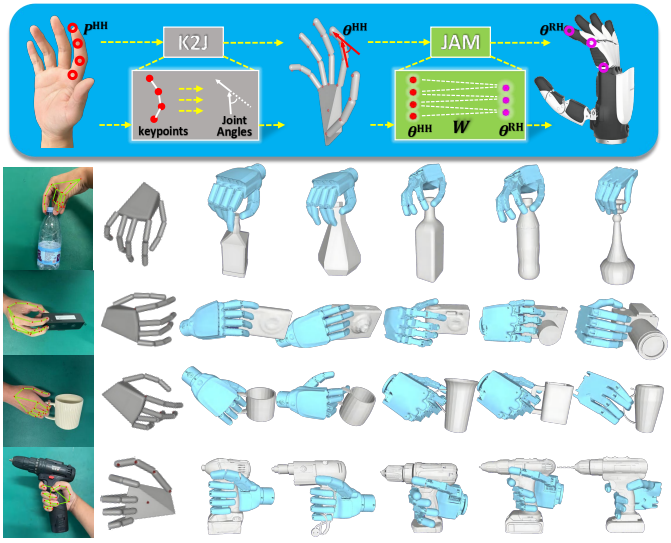

Fig. 1: Mapping natural human motions to an anthropomorphic hand model and unified control across representative robotic hands(ShadowHand, InspireHand), and HnuHand [9]. Visualization of functional grasps from UniFucGrasp dataset, including interactions with *bottle*, *camera*, *mug*, and *drill*.

Zhu *et al.* [2] were among the first to propose a functional grasp dataset for dexterous hands. Their method used binary encodings to annotate contact relationships between object surfaces and finger joints, but suffered from low efficiency and limited data scale. Later, Yang *et al.* [14] introduced a triplet-based semantic graph linking functional fingers to grasp gestures, enabling human-like behavior synthesis. However, their approaches were based on symbolic knowledge encoding and lacked real pose supervision.

Recently, DexVLG [15] and DexFuncGrasp [16] proposed large-scale functional grasp datasets (DexGraspNet 3.0 and DFG), providing valuable resources for training vision-language-action systems. Nonetheless, both datasets only support annotation for ShadowHand, a fully-actuated and high-cost robotic hand. This restricts their accessibility and hinders generalization to real-world applications due to the expensive hardware requirements.

This leads to a key question: Can we design a cost-efficient and generalizable annotation method that enables functional dexterous grasping across various hand types, facilitating broader adoption?

In fact, the design of robotic hands is often inspired by human motion coordination principles. Fully-actuated hands,

*Equal contribution.
†Corresponding author.
[1]The authors are with the School of Artificial Intelligence and Robotics, Hunan University, China.
[2]The authors are also with the National Engineering Research Center of Robot Visual Perception and Control Technology, Hunan University, China.
[3]The authors are with the College of Mechanical and Vehicle Engineering, Hunan University, China.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

such as ShadowHand, replicate independent joint control, while underactuated hands like InspireHand leverage mechanical linkages to simplify control. Inspired by this, we propose a novel human-to-robot grasp mapping framework that reformulates human motion transfer as a sparse matrix optimization problem. This unified formulation serves as a bridge between human demonstration and diverse dexterous hand architectures (both fully- and under-actuated), facilitating efficient and versatile functional grasp annotation.

Specifically, we propose a general mapping function that uses the human hand posture as an intermediary to bridge the structural differences between the human hand and various heterogeneous robotic hands. This function explicitly establishes the correspondence of Degrees of Freedom (DoFs) between the human hand and robotic hands through an adjustable mapping matrix W. Based on this mapping function, by incorporating the degrees of freedom of the target robotic hand d_{RH}, the degrees of freedom of the human hand d_{HH}, and their coupling relationships, the weight parameters in the mapping matrix W linking the joint angles of different heterogeneous robotic hands can be adjusted, while synchronously tuning the coupling matrix J, directly generating control commands for the corresponding robotic hand. Leveraging this design, we achieve structural decoupling in terms of DoF and actuation, and unified mapping modeling across diverse hands, eliminating the dependency on specific mechanical architectures, thereby facilitating efficient and precise adaptation of multiple heterogeneous robotic hands through the human hand as an intermediary.

Building on a general and efficient human-to-robot pose mapping method and using MuJoCo [17], we constructed and released UniFucGrasp (see Fig. 1)—a large-scale functional grasp dataset with over $100K$ high-quality annotations across $1,108$ objects from 21 daily-use categories. Supporting both fully-actuated and under-actuated dexterous hands, including ShadowHand, InspireHand, and HnuHand [9], the dataset enables stable grasp transfer on selected representative robotic hands and generalization to unseen instances within some categories, supporting improved cross-platform consistency. By employing a unified and novel pose mapping strategy, UniFucGrasp accurately replicates human hand motions on diverse robotic hands, providing stable, consistent functional grasp representations to support task-driven dexterous manipulation research.

In addition, we present a functional gesture generation model conditioned on point clouds for multiple dexterous hands. The backbone CVAE [18] learns shared grasp latent features, which a classification head maps to each hand's DOF space, facilitating a generalizable grasping strategy. Experiments in both IsaacSim [19] and real-world scenarios demonstrate significant improvements in functional manipulation accuracy and grasp stability, as well as efficient generalization across different robotic hands on identical tools and tasks.

Our main contributions are summarized as follows:

- We propose an annotation strategy and adopt a general, efficient human-to-robot pose mapping method that, using sparse matrix optimization and force-closure analysis, enables stable and reliable functional grasp transfer across diverse dexterous hands, effectively bridging structural and actuation differences.
- We construct the large-scale **UniFucGrasp** dataset, containing 1108 objects from 21 categories and over $100K$ functional grasp pose annotations, supporting dexterous hands with diverse structures and actuation types, including both fully- and under-actuated designs.
- We propose a functional gesture generation model conditioned on hand–object point clouds. Leveraging human prior annotations and joint training across multiple dexterous hands, it achieves unified functional grasping generation with improved precision, stability, and generalization, verified in both simulation and real-world experiments.

II. Related Work

A. Dexterous Robot Grasp Datasets

Existing dexterous hand grasping datasets [5] primarily focus on grasp stability, typically by directly sampling contact points on the object surface and evaluating grasp robustness using the GraspIt platform [20]. One approach [21] is to use the force closure criterion as the optimization objective to improve the stability and quality of the grasp. Another approach [7] focuses on sampling better initial grasp poses to further optimize the final target hand configurations. Although these methods have improved grasp performance to some extent, their reliance on simple grasping strategies and datasets still limits their ability to scale toward functional manipulation tasks. Zhu *et al.* [2] proposed a functional grasping dataset for dexterous hands. They used manually annotated binary codes to represent the contact relationships between the hand and the object, but this method is inefficient and lacks pose data from real robotic platforms. Although recent research [16] introduced a method for collecting functional grasping data from human hand motions in real time, it relies on deep learning models specifically trained for the shadow hand [22], resulting in significant hand-type dependency and limited generalization performance to other types of robotic hands. In addition, the lack of systematic evaluation of grasp stability leads to unreliable generated gestures, making it difficult to effectively support complex grasping and manipulation tasks across dexterous robotic hands.

Unlike existing datasets shown in Table I, this work establishes a unified functional grasping dataset for diverse dexterous hands, integrating grasp stability and functionality, with grasp data collected via human hand mapping, to alleviate the limitations of existing datasets in generalization capability and real-pose acquisition, thereby advancing complex dexterous manipulation tasks.

B. Human-to-Robotic Hand Motion Mapping

One of the key challenges in constructing dexterous hand grasp datasets is efficiently mapping natural human hand motions to various robotic hands. Existing methods mainly include joint mapping for power grasps [23], [24], [25], fingertip mapping for precision grasps [26], [27], and pose mapping for conveying functional intent [28]. Additionally, dimensionality reduction strategies based on postural synergies

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

TABLE I: Comparison of dexterous grasp datasets (F: Fully-actuated, U: Under-actuated).

Dataset	Robot Hand Type (F/U)	Grasp Method	Observations	Sim./Real	Grasps	Obj. (Cat.)	Collection Method	Data Generalization Across Diverse Hands
HO3D [10]	MANO (F)	Stable	RGBD	Real	77K	10	Estimation	✗
DexYCB [11]	MANO (F)	Stable	RGBD	Real	582K	20	Manual Annotation	✗
DexGraspNet [7]	ShadowHand (F)	Stable	-	Sim	1.32M	5355 (133)	Optimization	✓
AffordPose [12]	MANO (F)	Functional	-	Sim	26k	641 (13)	Optimization	✗
OakInk [13]	MANO (F)	Functional	RGBD	Sim	1K	100 (12)	Optimization	✗
Toward human-like grasp [2]	ShadowHand (F)	Functional	Semantic Knowledge	Real	-	129 (18)	Manual Annotation	✗
F2F [14]	InspireHand (U)	Functional	Semantic Knowledge	Real	14	127 (18)	Manual Annotation	✗
DexFuncGrasp [16]	ShadowHand (F)	Functional	RGBD	Real-Sim	14K	559 (12)	Optimization	✗
UniFucGrasp (Ours)	**ShadowHand (F), InspireHand (U), HnuHand (U)**	**Stable, Functional**	RGBD	Real-Sim	**100K**	**1108 (21)**	**Human Hand Mapping**	✓

have been applied to plan and control fully actuated robotic hands [29], [30], [31]. However, these methods are typically designed for a single type of robotic hand and overly rely on a single mapping paradigm, which limits their ability to scale in terms of generalizability and functional effectiveness. To address this, we propose a unified mapping strategy that models skeletal keypoints, joint angles, and actuation values as a sparse matrix, preserving natural human motion patterns while aiming for compatibility with various dexterous hands and alleviating dependence on large-scale training data. Based on this, we constructed the Unified Functional Grasp (UFG) dataset, providing annotated functional grasp poses and object category labels for multiple robotic hands.

III. Method

In this work, the goal is to generate reliable, functional grasp poses for diverse robotic hands via Human-Hand mapping. Given the object mesh, robot hand URDF, and 21 3D keypoints from an RGB-D camera, we build a unified action mapping represented by pose (rotation R, translation T) and joint angles J. We aim to provide an efficient pipeline to ensure reliable grasp generation and strong generalization for dexterous hands. An overview of our method is shown in Fig. 2.

Method Overview. First, we extract human hand skeletal keypoints and build a kinematic model, designing a keypoint-to-joint-angle conversion network (K2J) (see Sec. III-A) to faithfully replicate human hand motions. The K2J module uses MediaPipe [32] for keypoint detection, transforms them to the camera coordinate system, and maps them onto a biomimetic model closely matching human hand proportions, accurately reconstructing hand kinematics. Next, we represent the mapping from the human hand to various robotic hands (JAM) (see Sec. III-B) as a sparse matrix optimization, capturing anthropomorphic grasp poses. Using this mapping, we derive robotic joint angles and convert them from joint space to actuation space via a standardized method (see Sec. III-C), enabling stable control of the simulated hand and producing high-quality functional grasps. Finally, based on this annotation method, we build a multi-hand functional grasp dataset and validate its effectiveness within the functional gesture generation framework (see Sec. III-D).

A. Human-Hand Kinematic Modeling

To enable functional grasping, robots must understand the human-hand structure and motion. We decompose this into: (1) anthropomorphic hand model alignment, abstracting the hand into a biomimetic model to determine joint relative positions; (2) joint angle estimation, constructing a kinematic model for high-fidelity motion reconstruction.

Using MediaPipe [32], we detect 21 2D hand keypoints $\{\mathbf{k}_i\}_{i=1}^{21}$ from RGB images, then project them into 3D camera-frame points $\{\mathbf{p}_i\}_{i=1}^{21}$ with depth maps and camera intrinsics $\mathbf{K}_{\text{int}}$. Keypoints are registered onto the biomimetic hand model, replacing the wrist keypoint with the palm center for stability and accurate palm normal vector estimation. As shown in Fig. 2, the palm normal vector $\mathbf{n}_{\text{PALM}}$ is computed by the cross product of vectors formed by the ring finger base, index finger base, and palm center, using the index finger as an example:

$$\mathbf{n}_{\text{Palm}} = (\mathbf{p}_{\text{Ring}} - \mathbf{p}_{\text{Palm}}) \times (\mathbf{p}_{\text{Index}} - \mathbf{p}_{\text{Palm}}). \tag{1}$$

Next, for the i-th finger, we define the vector from joint n to joint $n+1$ as the n-th joint vector of the finger, denoted as $\mathbf{q}_{i,n}^{\text{HH}}$. This MCP joint vector is projected onto the palm plane defined by the palm normal vector $\mathbf{n}_{\text{Palm}}$, and the resulting projection vector is denoted as $\mathbf{n}_{i,1}^{\text{HH}}$. The abduction-adduction angle of the i-th finger is the angle between the palm-to-MCP vector $\mathbf{q}_{i,0}^{\text{HH}}$ and The projection vector $\mathbf{n}_{i,1}^{\text{HH}}$, given by:

$$\theta_{i,\text{abd}}^{\text{HH}} = \arccos\left(\frac{\mathbf{q}_{i,0}^{\text{HH}} \cdot \mathbf{n}_{i,1}^{\text{HH}}}{\|\mathbf{q}_{i,0}^{\text{HH}}\| \cdot \|\mathbf{n}_{i,1}^{\text{HH}}\|}\right). \tag{2}$$

The flexion-extension angle $\theta_{i,\text{flex}}^{\text{HH}}$ is defined as the angle between the current joint vector and the reverse extension of the adjacent joint vector:

$$\theta_{i,\text{flex}}^{\text{HH}} = \arccos\left(\frac{\mathbf{q}_{i,n}^{\text{HH}} \cdot \mathbf{q}_{i,(n+1)}^{\text{HH}}}{\|\mathbf{q}_{i,n}^{\text{HH}}\| \cdot \|\mathbf{q}_{i,(n+1)}^{\text{HH}}\|}\right). \tag{3}$$

The overall joint angle prediction error is calculated by:

$$E = \sqrt{\frac{1}{N}\sum_{i=0}^{f}\sum_{n=0}^{d_i}\left(\hat{\theta}_{i,n}^{\text{HH}} - \theta_{i,n}^{\text{HH}}\right)^2}. \tag{4}$$

where N denotes the total number of data samples, f is the total number of fingers, and d_i represents the number of joints in the i-th finger. Here, $\hat{\theta}_{i,n}^{\text{HH}}$ and $\theta_{i,n}^{\text{HH}}$ denote the predicted and ground-truth joint angles of the n-th joint in the i-th finger of the Human-Hand, respectively.

B. Human-Hand Mapping Representation

Given the human hand grasping pose represented by the joint angles $\theta \in \mathbb{R}^{20\times 1}$, our objective is to construct a general mapping function that enables accurate replication of the Human-Hand posture on the robotic platform. Specifically,

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

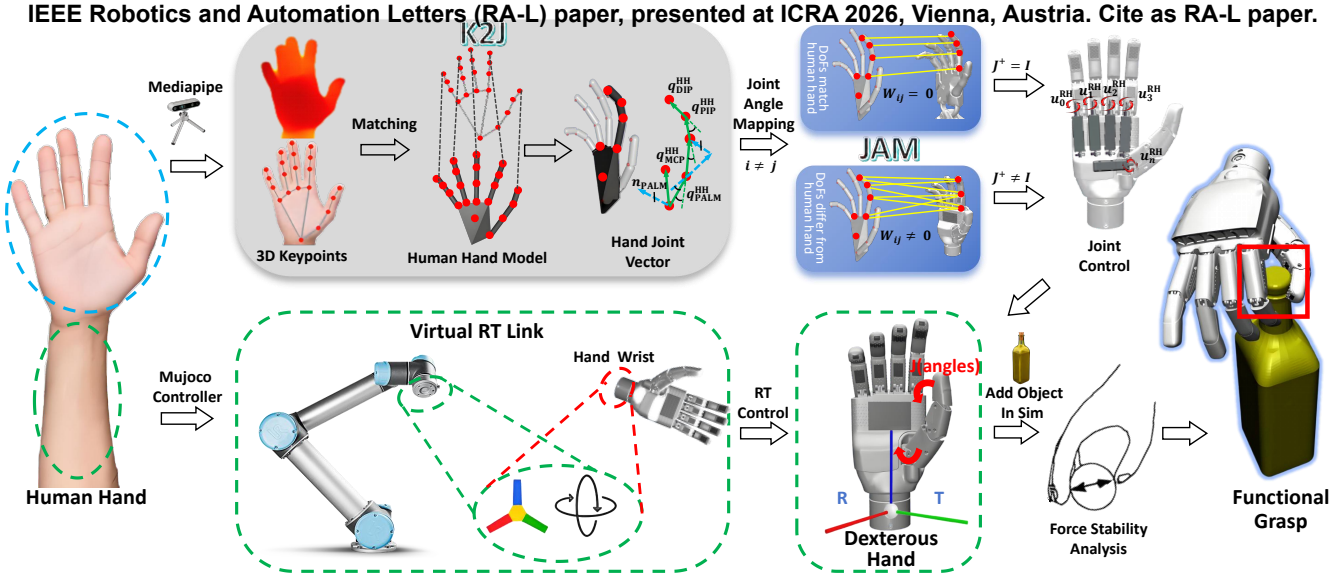

Fig. 2: Illustration of our annotation strategy for transferring human motions to robotic hands, with the left showing capture, the top depicting gesture mapping, the bottom showing pose collection, and the right displaying the resulting functional grasps.

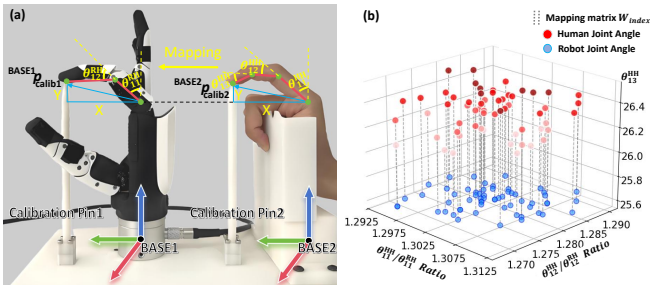

Fig. 3: Flowchart of data acquisition for mapping joint angles from the human index finger to the robotic index finger.

the n-th joint angle of the i-th finger in the Human-Hand is denoted as $\theta_{i,n}^{\mathrm{HH}}$, and the corresponding joint angle in the robotic hand is denoted as $\theta_{i,n}^{\mathrm{RH}}$. The mapping relationship is formulated as:

$$\theta_{i,n}^{\mathrm{RH}} = W\theta_{i,n}^{\mathrm{HH}} + \epsilon, \tag{5}$$

where W is the mapping matrix and ϵ is the error term capturing possible deviations. The dimension of the mapping matrix W depends on the relationship between the degrees of freedom of the robotic hand d_{RH} and the Human-Hand d_{HH}, specifically expressed as:

$$\dim(W) = \begin{cases} \mathbb{R}^{d_{\mathrm{RH}} \times d_{\mathrm{HH}}} = \mathbb{R}^{d_{\mathrm{HH}} \times d_{\mathrm{HH}}}, & \text{if } d_{\mathrm{RH}} = d_{\mathrm{HH}}, \\ \mathbb{R}^{d_{\mathrm{RH}} \times d_{\mathrm{HH}}}, & \text{if } d_{\mathrm{RH}} \neq d_{\mathrm{HH}}. \end{cases} \tag{6}$$

When $d_{\mathrm{RH}} = d_{\mathrm{HH}}$, the mapping matrix W is square, enabling one-to-one joint correspondence. When $d_{\mathrm{RH}} \neq d_{\mathrm{HH}}$, W becomes non-square, performing compression or expansion to adapt the Human-Hand joint space to the robotic hand.

For the ShadowHand ($d_{\mathrm{RH}} = d_{\mathrm{HH}}$), W is a diagonal matrix that scales joint angles based on size differences. For the InspireHand ($d_{\mathrm{RH}} = 12$, $d_{\mathrm{HH}} = 20$), Human-Hand joints $\theta^{\mathrm{HH}} \in \mathbb{R}^{20\times 1}$ are mapped to robotic hand joints $\theta^{\mathrm{RH}} \in \mathbb{R}^{12\times 1}$ via $W \in \mathbb{R}^{12\times 20}$.

$$\theta^{\mathrm{HH}} \in \mathbb{R}^{20\times 1} \xrightarrow{W} \theta^{\mathrm{RH}} \in \mathbb{R}^{12\times 1}. \tag{7}$$

The mapping matrix W can be decomposed into submatrices corresponding to each finger:

$$W = \begin{bmatrix} W_{\mathrm{Thumb}} & W_{\mathrm{Index}} & W_{\mathrm{Middle}} & W_{\mathrm{Ring}} & W_{\mathrm{Little}} \end{bmatrix}^T, \tag{8}$$

where each submatrix W_{Thumb}, W_{Index}, W_{Middle}, W_{Ring}, and W_{Little} maps the joint angles of a specific finger from the Human-Hand to the corresponding finger on the robotic hand. Taking the index finger as an example, since the InspireHand lacks the abduction degree of freedom and has one fewer flexion/extension DoF compared to the Human-Hand, it has 2 DoFs instead of 4. Thus, the two joint angles of the InspireHand index finger, $\theta_{11}^{\mathrm{RH}}$ and $\theta_{12}^{\mathrm{RH}}$, are linearly mapped from the three joint angles of the Human-Hand index finger, $\theta_{11}^{\mathrm{HH}}$, $\theta_{12}^{\mathrm{HH}}$, and $\theta_{13}^{\mathrm{HH}}$, through the index finger mapping submatrix W_{Index}, as follows:

$$\begin{bmatrix} \theta_{11}^{\mathrm{RH}} & \theta_{12}^{\mathrm{RH}} \end{bmatrix}^{\mathrm{T}} = W_{\mathrm{Index}} \begin{bmatrix} \theta_{11}^{\mathrm{HH}} & \theta_{12}^{\mathrm{HH}} & \theta_{13}^{\mathrm{HH}} \end{bmatrix}^{\mathrm{T}}, \tag{9}$$

and the mapping submatrix $W_{\mathrm{Index}} \in \mathbb{R}^{2\times 3}$ is defined and optimized as:

$$W_{\mathrm{Index}} = \begin{bmatrix} \alpha & \beta & \gamma \\ \delta & \varepsilon & \zeta \end{bmatrix}, \quad \min_{W_{\mathrm{Index}}} \sum_{i=1}^{N} \left\| W_{\mathrm{Index}} \boldsymbol{\theta}_i^{HH} - \boldsymbol{\theta}_i^{RH} \right\|_2^2. \tag{10}$$

The parameters $\alpha, \beta, \gamma, \delta, \varepsilon, \zeta$ are obtained through a fingertip alignment-based mapping optimization method. By synchronously collecting motion data of the human and robotic hands under consistent fingertip contact and pressing postures(Fig. 3(a)), a linear mapping W_{Index} is established between their joint angle spaces(Fig. 3(b)). Finally, the parameters are optimized via least squares, constrained by fingertip position alignment. The mapping must satisfy both joint feasibility and fingertip position consistency:

$$\theta_{i,\min}^{RH} \leq \theta_i^{RH} \leq \theta_{i,\max}^{RH}, \tag{11}$$

$$f^{HH}(\theta_i^{HH}) = f^{RH}(\theta_i^{RH}) = {}^{\mathrm{BASE}}p_{\mathrm{calib}}. \tag{12}$$

where $f^{HH}(\cdot)$ and $f^{RH}(\cdot)$ are the forward kinematics func-

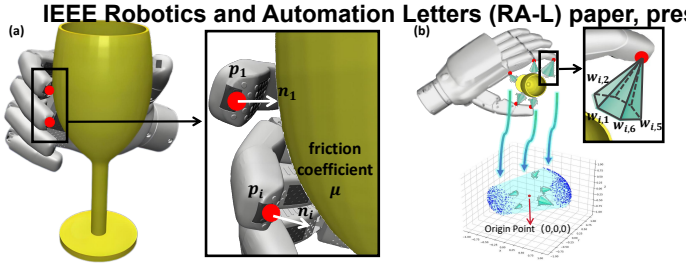

Fig. 4: For each contact i $(\mathbf{n}_i, \mathbf{p}_i)$, we show its six friction cone directions $\mathbf{w}_{i,j}$ and the reduced wrench spaces (blue regions).

tions of the human and robotic hands, respectively, mapping joint angles to fingertip positions; ${}^{\text{BASE}}p_{\text{calib}}$ represents the fingertip position expressed in the base coordinate frame(Fig. 3(a)).

C. Functional Dexterous Hand Control via RTJ Mapping

Six virtual links connect the dexterous hand base to the world frame, enabling explicit rotation and translation control. The simulation provides real-time hand pose feedback (quaternions and position) relative to the object. Accurate control requires mapping joint space to actuator space, accounting for motor inputs and constraints. This mapping is direct for fully actuated hands and accounts for coupling in underactuated hands. To unify both, the joint-to-actuator mapping is defined as:

$$u_{i,n}^{\text{RH}} = J^{+}\theta_{i,n}^{\text{RH}}, \tag{13}$$

where J^{+} is the generalized inverse of the mapping matrix J, which is commonly computed as the Moore-Penrose [33] pseudoinverse:

$$J^{+} = (J^{\top}J)^{-1}J^{\top}. \tag{14}$$

For the underactuated InspireHand, the joint coupling matrix $J \in \mathbb{R}^{12\times 6}$ was obtained via manual measurements, and its generalized inverse J^{+} was computed to map joint space to actuator space.

After actively mapping and collecting gestures, grasp performance is post-processed and evaluated using geometry-based force-closure analysis [29], [34]. We record hand–object collision points via open interfaces [17] to define contacts, and approximate each friction cone with six rays to model contact forces. As shown in Fig. 4(a), from the grasp contact points, we discretize each friction cone to approximate feasible contact forces. For each contact point $\mathbf{p}_i \in \mathbb{R}^3$, with normal $\mathbf{n}_i \in \mathbb{R}^3$ and friction coefficient μ, the friction cone half-angle is computed as:

$$\theta = \arctan(\mu). \tag{15}$$

Given a vector $\mathbf{r}$ that is not parallel to the normal vector $\mathbf{n}_i$, we construct two unit vectors $\mathbf{t}_1$ and $\mathbf{t}_2$ orthogonal to $\mathbf{n}_i$ via the cross product:

$$\mathbf{t}_1 = \frac{\mathbf{n}_i \times \mathbf{r}}{\|\mathbf{n}_i \times \mathbf{r}\|}, \quad \mathbf{t}_2 = \mathbf{n}_i \times \mathbf{t}_1. \tag{16}$$

Based on these, the j-th approximate friction cone direction is generated as:

$$\mathbf{w}_{i,j} = \cos(\theta)\mathbf{n}_i + \sin(\theta)\big(\cos(\phi_j)\mathbf{t}_1 + \sin(\phi_j)\mathbf{t}_2\big), \tag{17}$$

where $\phi_j = \frac{2\pi(j-1)}{6}$ for $j = 1, 2, \ldots, 6$.

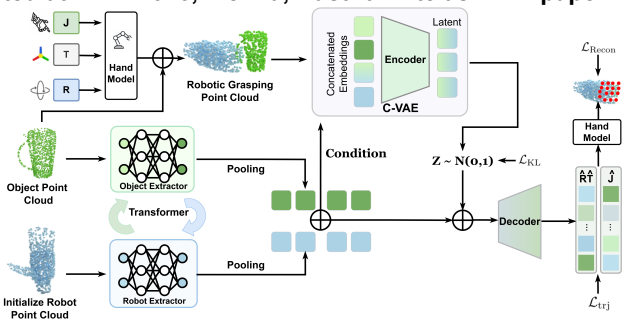

Fig. 5: Functional gesture generation for diverse robotic hands, with point cloud feature extraction, CVAE [18] latent encoding and sampling, and MLP decoder for gesture parameters.

For n contact points, each with 6 directions, the wrench and the grasp matrix $\mathbf{G}$ are computed as follows, where $i = 1, \ldots, n$ and $j = 1, \ldots, 6$:

$$\mathbf{wrench}_{i,j} = \begin{bmatrix} \mathbf{w}_{i,j} \\ \mathbf{p}_i \times \mathbf{w}_{i,j} \end{bmatrix} \in \mathbb{R}^6, \tag{18}$$

$$\mathbf{G} = [\mathbf{wrench}_{i,j}] \in \mathbb{R}^{6\times 6n}. \tag{19}$$

where i and j iterate over contact points and directions, respectively. Then, as shown in Fig. 4(b), the force-closure condition is checked by determining whether the origin lies inside the convex hull of the wrench vectors.

D. Functional Grasp Generation

Functional Grasp Synthesis Model: Both stable and functional grasping fundamentally depend on accurately predicting the dexterous hand's pose (R, T) and joint configurations (J). To assess the effectiveness of our dataset for functional grasping, we develop a task-driven, lightweight deep neural network. As shown in Fig. 5, the designed network takes hand and object point clouds with input dimensions of 2500×3 and 2000×3, which are separately processed by Robot Extractor and Object Extractor modules based on DGCNN [35]. It effectively extracts local geometric features from sparse 3D data by leveraging spatial relationships among neighboring points, providing stable and spatially-aware feature encodings for the subsequent CVAE [18] to generate diverse and plausible hand configurations.

In the designed functional grasp generation network, we adopt a lightweight Transformer-based architecture following DCP [36] for cross-object embedding and cross-modal alignment. Fused features from a 4-head encoder-decoder with 128 feedforward hidden dimensions are fed into the CVAE [18] encoder. A latent vector z is sampled and concatenated with max-pooled hand and object features to form a 260-dimensional joint representation. This vector is fed into the grasp generation network to predict hand rotation (r), translation (t), and joint angles (j).

Loss Functions: We adopt a compact loss to evaluate functional grasping, where $\alpha_1, \alpha_2, \alpha_3$ weight the KL divergence, trajectory, and reconstruction terms, respectively:

$$\mathcal{L} = \alpha_1\mathcal{L}_{\text{KL}} + \alpha_2\mathcal{L}_{\text{trj}} + \alpha_3\mathcal{L}_{\text{Recon}}. \tag{20}$$

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

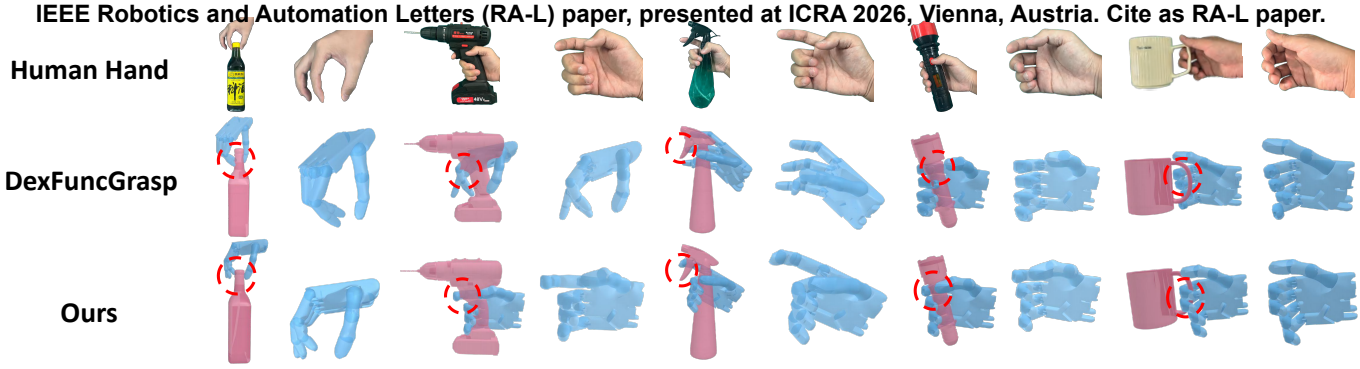

Fig. 6: Qualitative comparison with the state-of-the-art method DFG [16] on functional grasping shows that our method better captures human hand poses and more consistently aligns with functional regions across diverse manipulation tools.

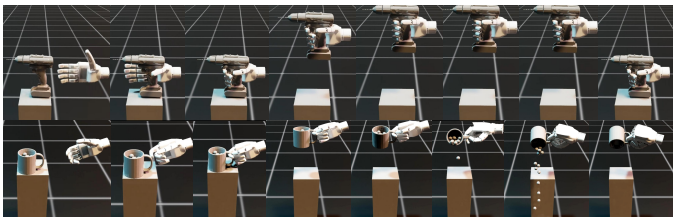

Fig. 7: Simulation results of the method for functional tasks: *Press Button and Pour Water*.

The KL divergence encourages the latent distribution $Q(z \mid o, g)$ to align with a standard Gaussian prior $\mathcal{N}(0, I)$, facilitating structured and continuous latent space learning:

$$\mathcal{L}_{\mathrm{KL}} = -\mathrm{KL}\left(Q(z \mid o, g) \parallel \mathcal{N}(0, I)\right). \tag{21}$$

We first define the L1 loss on grasp parameters: hand rotation ($\hat{r}$), translation ($\hat{t}$), and joint angles ($\hat{j}$):

$$\mathcal{L}_{\mathrm{trj}} = \lambda_1 \|\hat{r} - r\|_1 + \lambda_2 \|\hat{t} - t\|_1 + \lambda_3 \|\hat{j} - j\|_1. \tag{22}$$

$$\mathcal{L}_{\mathrm{Recon}} = \sum_{i=0}^{index} \left| p_i^{pred} - p_i^{gt} \right|. \tag{23}$$

In addition, we design a reconstruction loss L_{Recon} in Fig. 5, which supervises training by comparing keypoints of the predicted grasp gestures—generated from $\hat{r}, \hat{t}, \hat{j}$ via the hand model—with the corresponding ground-truth gestures. We adopt L1 loss for robustness and stable gradients, and apply it on gesture keypoints to enhance prediction accuracy and model stability.

IV. Experiments

A. Experiment Setup

Annotation and Dataset: We collected 60 sets of human joint angle data from six volunteers using the K2J module and refined the measurements with a bone joint goniometer, constraining joint angle errors according to Eq. (4) to improve accuracy. The JAM module was then used to establish fingertip mapping between human and robotic hands. Mapping parameters W were estimated via least-squares optimization to minimize joint angle differences (Eq. (10)), with constraints on joint range (Eq. (11)) and the absolute coordinates of the fingertips and hand base ${}^{\mathrm{BASE}}p_{\mathrm{calib}}$ (Eq. (12)). Taking the InspireHand as an example, since its thumb shares the same degrees of freedom as the human thumb, a direct mapping was used. For the other fingers, due to differences in degrees

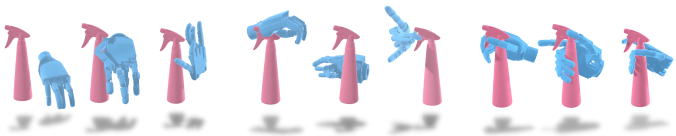

Fig. 8: Qualitative ablation results for functional grasping of *spraybottle* under three different hand-conditioned inputs.

of freedom, we employed a calibration experimental setup for fingertip mapping, using the functional index finger as a representative for calibration (Fig 3), to achieve a more accurate dimensionality-reducing mapping. The mapping matrix W_{Index} include α=0.3530, β=0.4310, γ=0.2827, δ=0.2584, ε=0.4130, and ζ=−0.0018. To directly translate the mapping results into dexterous hand control and enable efficient annotation of functional grasp postures, we measured and modeled the joint coupling of the InspireHand, constructed the Jacobian matrix $J \in \mathbb{R}^{12\times6}$ mapping joint to actuator space (Eq. (13)), and computed its pseudoinverse J^{+} to map joint-level commands to actuator-level signals. Based on this annotation method, we constructed the UFG dataset in MuJoCo [17] by controlling three dexterous hands via tracked natural hand motions in Fig. 1. The dataset covers 21 categories with $1,108$ object instances, each having over 70 validated functional grasp demonstrations, totaling more than $100K$ annotations. Grasp stability was ensured via force feedback and collision detection, with each contact point defined as the collision between the object and hand, its 3D position and surface normal recorded, and the friction cone discretized into six rays with a coefficient of friction μ=0.5 [37], [38], typical of common materials [39]. The dataset was split into training and testing sets at an 8.5:1.5 ratio per category, with the test set consisting entirely of unseen objects that are structurally distinct from those in the training set. Our method enhances the DFG dataset [16], which helps capture realistic human motion priors and common functional grasp patterns.

Implementation Details: The model employs DGCNN [35] for feature extraction and is trained within a conditional variational autoencoder CVAE [18] using the Adam optimizer, with a learning rate of 0.00001 over 15 epochs. Experiments are conducted on two NVIDIA RTX 3090 GPUs. To quantitatively evaluate the performance of our dataset and the functional grasp synthesis model, we apply Kullback-Leibler divergence to regularize the latent space and promote structured grasp representations, alongside an L1-based reconstruction loss to

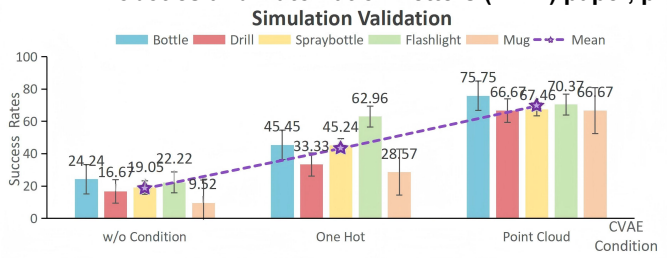

Fig. 9: Quantitative results of three hand-conditioned input settings under the same random seed (41, 42, 43).

TABLE II: Quantitative analysis of simulation test results across different random seeds for each object category.

Category	SR (DFG [16])	SR (Seed 41 / 42 / 43)	Mean (Ours) ± Std	Train / Test
Bottle	0.6862	0.7272 / 0.8181 / 0.7272	**0.7575 ± 0.0428**	54 / 11
Drill	0.5500	0.6250 / 0.7500 / 0.6250	**0.6667 ± 0.0589**	48 / 8
Spraybottle	0.5807	0.6428 / 0.6667 / 0.7143	**0.6746 ± 0.0364**	85 / 14
Flashlight	0.9103	0.7777 / 0.6667 / 0.6667	0.7037 ± 0.0523	44 / 9
Mug	0.5462	0.5714 / 0.7143 / 0.7143	**0.6667 ± 0.0674**	38 / 7
Total	0.6552	0.6732 / 0.7207 / 0.6934	**0.6955 ± 0.0219**	269 / 48

measure the accuracy of predicted hand rotation, translation, and joint angles, effectively reflecting the validity and stability of generated functional gestures.

B. Comparison of Grasping Performance

Quantitative Results: To verify the effect of hand point cloud conditioning in the CVAE framework, we performed an ablation study across different categories with all other components fixed (see Fig. 9). Results comparing no condition, One-hot identity encoding, and point cloud conditioning show clear gains in success rates: 19.05%, 44.21%, and 69.55%, respectively—with point cloud significantly improving performance by 57.34%, indicating that the point cloud offers richer geometric and spatial cues for generating stable, functional, and unified human-like dexterous grasps. The model's predicted grasp poses and joint angles are visualized and executed in IsaacSim [19] simulation, where both hand and object are treated as rigid bodies. A grasp is deemed successful if the object remains held after lifting the hand by $10cm$, and a manipulation is considered successful if the task can be completed under predefined disturbances. All other hyperparameters were kept unchanged, and seeds 41, 42, and 43 were used for testing. Table II shows our method achieves an average success rate of 69.55%, about 6.15% higher than DFG [16]. Although the grasp success rates of the mug and drill are relatively low across all methods due to handle shapes and narrow gaps, our method achieves higher success rates on these objects (66.67%), surpassing 55% (drill) and 54.62% (mug) in DFG [16]. while also generating gestures that precisely align the index finger with the drill button or execute the mug pouring action, ensuring successful functional manipulations. This improvement stems from modeling based on human hand priors, bringing the unified dexterous hand closer to human grasping patterns and aligning gestures with functional intentions, further validating the method's effectiveness and stability in functional grasping tasks.

Qualitative Analysis: Based on the quantitative analysis, we visualized the test gestures under three settings: without condition, One-hot identity encoding, and hand point cloud encoding. Fig. 8 shows that hand point cloud conditioning captures common geometric features across different hands, enabling unified human-like functional gestures. As shown in Fig. 6, compared to natural human hand motions and DFG [16], our method localizes gestures more accurately at functional contact regions while maintaining stable finger envelopment, and executes tasks such as pressing the drill button (6/20) or pouring (9/20) with higher precision and reliability in Fig. 7 and Fig. 10.

Fig. 10: Real-world functional manipulation results in multiple robotics hands. DFG [16] results are shown in brackets.

In contrast, DFG [16] depends purely on network prediction without structural or motion priors, resulting in less reliable finger coordination. Our method achieves more stable and human-consistent dexterous performance.

C. Real-World Experiments

In real-world experiments, we adopted a cost-effective setup combining a UR5 robotic arm with the InspireHand and HnuHand [9] for functional grasping validation. As illustrated in Fig. 11, the platform consists of two dexterous hands, a UR5 arm, a RealSense camera, a calibration board, Aruco codes, a 3D scanner, and a control computer. We first scanned several target objects (*e.g.*, *bottle*, *drill*, *spraybottle*, *flashlight*, and *mug*) using a FreeScan X3 scanner for modeling and post-processing. After calibrating the intrinsics and extrinsics of the RealSense camera, object poses were estimated using FoundationPose [40]. Uniform point cloud sampling and registration were performed on object surfaces. The processed point clouds were input to the functional grasp model, generating end-effector poses R, T and gesture parameters Q. Thanks to the two robotics hands' coupling, only relevant active joints are controlled, with success criteria consistent with simulation. After the grasps, minor predefined joint and rotation changes were applied to verify gesture stability and their effectiveness for subsequent functional tasks, such as pressing a button or pouring water. As shown in Table III, Our method outperformed the latest DFG [16] across five unseen object categories, improving the success rate by 11.54% on InspireHand and 23.08% on HnuHand [9]. Moreover, as shown in Fig. 10, our method, evaluated on different dexterous hands after successful grasps, outperformed DFG [16] by over 75% on key functional tasks (*e.g.*, *press button* and *pour water*), demonstrating more precise control of functional regions via hand motion priors. Despite the complexity of dexterous hands limiting overall success, our results show that reliable, human-like motions can improve functional grasping generalization.

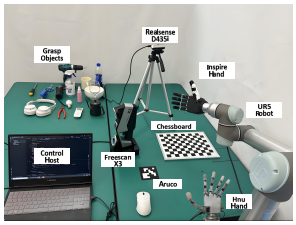

Fig. 11: Real-world experiment setting.

TABLE III: Real-world functional grasping and manipulation results on different robotic hands.

Method	Bottle	Spraybottle	Flashlight	Drill	Mug	Press Button	Pour Water	Total
DFG [16]	10/10	4/10	8/10	0/10	4/10	0/10	4/10	30/70
Ours (Inspire)	8/10	5/10	6/10	5/10	5/10	4/10	4/10	37/70
Ours (Hnu)	7/10	6/10	8/10	5/10	6/10	2/10	5/10	39/70

V. Conclusion

This work presents an efficient human hand mapping strategy, modeling hand motions as a sparse matrix to enable human-centered, real-time functional gesture transfer from robots to human hand. Combined with geometric force closure analysis, it effectively evaluates grasp stability. Based on this, we built a large-scale functional grasp dataset. Experiments show our strategy and dataset accurately capture grasp quality, supporting diverse and stable grasps that outperform existing methods. Real-world tests indicate that, while reliable functional grasps can be achieved on unseen instance, hand differences pose challenges for broader generalization. Future work will integrate physics simulation and multimodal sensing to further improve gesture precision, multi-hand generalization, and grasp performance.

References

[1] S. Brahmbhatt, A. Handa, J. Hays, and D. Fox, "ContactGrasp: Functional multi-finger grasp synthesis from contact," in *Proc. IROS*, 2019, pp. 2386–2393.

[2] T. Zhu, R. Wu, X. Lin, and Y. Sun, "Toward human-like grasp: Dexterous grasping via semantic representation of object-hand," in *Proc. ICCV*, 2021, pp. 15 721–15 731.

[3] Y. Zhang *et al.*, "FunctionalGrasp: Learning functional grasp for robots via semantic hand-object representation," *IEEE Robotics and Automation Letters*, vol. 8, no. 5, pp. 3094–3101, 2023.

[4] Y. Liu *et al.*, "RealDex: Towards human-like grasping for robotic dexterous hand," in *Proc. IJCAI*, 2024, pp. 6859–6867.

[5] M. Liu, Z. Pan, K. Xu, K. Ganguly, and D. Manocha, "Generating grasp poses for a high-DOF gripper using neural networks," in *Proc. IROS*, 2019, pp. 1518–1525.

[6] W. Wei *et al.*, "DVGG: Deep variational grasp generation for dextrous manipulation," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1659–1666, 2022.

[7] R. Wang *et al.*, "DexGraspNet: A large-scale robotic dexterous grasp dataset for general objects based on simulation," in *Proc. ICRA*, 2023, pp. 11 359–11 366.

[8] J. Ye *et al.*, "Dex1B: Learning with 1B demonstrations for dexterous manipulation," in *Proc. RSS*, 2025.

[9] Z. Zhou, W. Chen, Z. Hu, Q. Diao, Q. Gao, and Y. Wang, "Design of an adaptive modular anthropomorphic dexterous hand for human-like manipulation," *arXiv preprint arXiv:2511.22100*, 2025.

[10] S. Hampali, M. Rad, M. Oberweger, and V. Lepetit, "HOnnotate: A method for 3D annotation of hand and object poses," in *Proc. CVPR*, 2020, pp. 3193–3203.

[11] Y. Chao *et al.*, "DexYCB: A benchmark for capturing hand grasping of objects," in *Proc. CVPR*, 2021, pp. 9044–9053.

[12] J. Jian, X. Liu, M. Li, R. Hu, and J. Liu, "AffordPose: A large-scale dataset of hand-object interactions with affordance-driven hand pose," in *Proc. ICCV*, 2023, pp. 14 667–14 678.

[13] L. Yang *et al.*, "OakInk: A large-scale knowledge repository for understanding hand-object interaction," in *Proc. CVPR*, 2022, pp. 20 921–20 930.

[14] F. Yang *et al.*, "Task-oriented tool manipulation with robotic dexterous hands: A knowledge graph approach from fingers to functionality," *IEEE Transactions on Cybernetics*, vol. 55, no. 1, pp. 395–408, 2025.

[15] J. He *et al.*, "DexVLG: Dexterous vision-language-grasp model at scale," *arXiv preprint arXiv:2507.02747*, 2025.

[16] J. Hang *et al.*, "DexFuncGrasp: A robotic dexterous functional grasp dataset constructed from a cost-effective real-simulation annotation system," in *Proc. AAAI*, vol. 38, no. 9, 2024, pp. 10 306–10 313.

[17] E. Todorov, T. Erez, and Y. Tassa, "MuJoCo: A physics engine for model-based control," in *Proc. IROS*, 2012, pp. 5026–5033.

[18] K. Sohn, H. Lee, and X. Yan, "Learning structured output representation using deep conditional generative models," in *Proc. NeurIPS*, vol. 28, 2015, pp. 3483–3491.

[19] F. F. Monteiro, S. Silva, and P. N. Lima, "Simulating real robots in virtual environments using NVIDIA's Isaac SDK," in *Proc. SVR*, 2019, pp. 248–251.

[20] A. T. Miller and P. K. Allen, "GraspIt!: A versatile simulator for grasp analysis," in *Proc. IMECE*, vol. 26652, 2000, pp. 1251–1258.

[21] 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, 2022.

[22] S. Li *et al.*, "Vision-based teleoperation of shadow dexterous hand using end-to-end deep neural network," in *Proc. ICRA*, 2019, pp. 416–422.

[23] F. Kobayashi *et al.*, "Multiple joints reference for robot finger control in robot hand teleoperation," in *Proc. SII*, 2012, pp. 577–582.

[24] H. Liu *et al.*, "High-fidelity grasping in virtual reality using a glove-based system," in *Proc. ICRA*, 2019, pp. 5180–5186.

[25] M. V. Liarokapis, P. K. Artemiadis, and K. J. Kyriakopoulos, "Telemanipulation with the DLR/HIT II robot hand using a dataglove and a low cost force feedback device," in *Proc. MED*, 2013, pp. 431–436.

[26] R. N. Rohling, J. M. Hollerbach, and S. C. Jacobsen, "Optimized fingertip mapping: A general algorithm for robotic hand teleoperation," *Presence: Teleoperators & Virtual Environments*, vol. 2, no. 3, pp. 203–220, 1993.

[27] L. Cui, U. Cupcic, and J. S. Dai, "An optimization approach to teleoperation of the thumb of a humanoid robot hand: Kinematic mapping and calibration," *Journal of Mechanical Design*, vol. 136, no. 9, p. 091005, 2014.

[28] C. Meeker, T. Rasmussen, and M. Ciocarlie, "Intuitive hand teleoperation by novice operators using a continuous teleoperation subspace," in *Proc. ICRA*, 2018, pp. 5821–5827.

[29] M. Ciocarlie, C. Goldfeder, and P. Allen, "Dimensionality reduction for hand-independent dexterous robotic grasping," in *Proc. IROS*, 2007, pp. 3270–3275.

[30] F. Ficuciello, G. Palli, C. Melchiorri, and B. Siciliano, "Planning and control during reach to grasp using the three predominant UB hand IV postural synergies," in *Proc. ICRA*, 2012, pp. 2255–2260.

[31] G. Palli *et al.*, "The DEXMART hand: Mechatronic design and experimental evaluation of synergy-based control for human-like grasping," *The International Journal of Robotics Research*, vol. 33, no. 5, pp. 799–824, 2014.

[32] C. Lugaresi *et al.*, "MediaPipe: A framework for building perception pipelines," *arXiv preprint arXiv:1906.08172*, 2019.

[33] J. C. A. Barata and M. S. Hussein, "The Moore–Penrose pseudoinverse: A tutorial review of the theory," *Brazilian Journal of Physics*, vol. 42, pp. 146–165, 2012.

[34] M. T. Ciocarlie and P. K. Allen, "Hand posture subspaces for dexterous robotic grasping," *The International Journal of Robotics Research*, vol. 28, no. 7, pp. 851–867, 2009.

[35] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph CNN for learning on point clouds," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 5, pp. 1–12, 2019.

[36] Y. Wang and J. Solomon, "Deep closest point: Learning representations for point cloud registration," in *Proc. ICCV*, 2019, pp. 3522–3531.

[37] J. M. Inouye, J. J. Kutch, and F. J. Valero-Cuevas, "A novel synthesis of computational approaches enables optimization of grasp quality of tendon-driven hands," *IEEE Transactions on Robotics*, vol. 28, no. 4, pp. 958–966, 2012.

[38] K. Yao and A. Billard, "Exploiting kinematic redundancy for robotic grasping of multiple objects," *IEEE Transactions on Robotics*, vol. 39, no. 3, pp. 1982–2002, 2023.

[39] D. R. Lide, *CRC handbook of chemistry and physics: a ready-reference book of chemical and physical data*. CRC press, 1995.

[40] B. Wen, W. Yang, J. Kautz, and S. Birchfield, "FoundationPose: Unified 6D pose estimation and tracking of novel objects," in *Proc. CVPR*, 2024, pp. 17 868–17 879.