AnyDexRT: Calibration-Free Dexterous Hand Retargeting with Few-Shot Human Guidance

Chenxi Wang*,1, Ying Feng*,2,3, Hongjie Fang2,†, Shangning Xia1,
Lixin Yang2, Chuan Wen2, Cewu Lu1,2,3,†
1Noematrix 2Shanghai Jiao Tong University 3Shanghai Innovation Institute
*Equal Contribution †Corresponding Authors

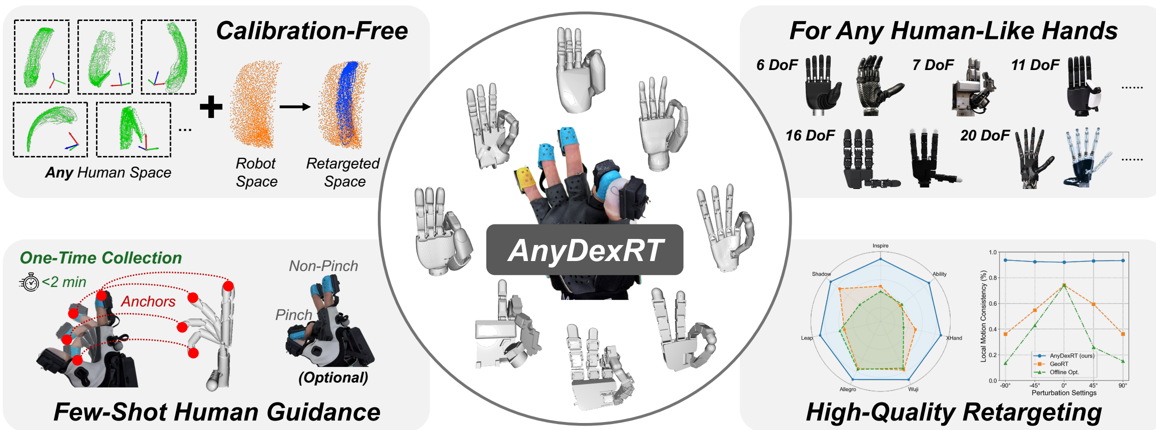

Fig. 1: AnyDexRT System. AnyDexRT is a calibration-free dexterous hand retargeting system. With few-shot human guidance, our system achieves high-quality retargeting across diverse human-like hands.

***Abstract*—Teleoperation is a key interface for controlling dexterous robotic hands and collecting demonstrations for imitation learning. Its effectiveness largely depends on kinematic retargeting, which maps operator hand motions to feasible and intuitive robot hand motions. Existing methods often require hand-crafted objectives, precise calibration, or global shape matching between human and robot hand spaces, making them sensitive to hand-specific tuning and less reliable across different dexterous hands. We propose AnyDexRT, a calibration-free retargeting method for intuitive dexterous teleoperation across human-like dexterous hands. AnyDexRT combines self-supervised fingertip correspondence learning with few-shot human guidance to anchor the mapping in task-relevant regions, and further refines pinch-related poses using a contact classifier. Experiments on diverse dexterous hands and real-world teleoperation tasks show that AnyDexRT improves retargeting quality, reduces manual tuning, and provides more intuitive and efficient control than prior retargeting methods. Project website: https://chenxi-wang.github.io/projects/anydexrt.**

I. Introduction

Dexterous manipulation is a key capability for general-purpose robots, enabling rich and adaptive physical interactions with objects and environments [2, 14, 30]. Compared with grippers, dexterous hands provide greater kinematic expressiveness and contact versatility, supporting diverse grasping [13, 53], in-hand object reorientation [5, 6], and contact-rich behaviors [22, 44] in unstructured environments. However, the same expressiveness that makes dexterous hands powerful also makes them difficult to control. Their high-dimensional action spaces, joint couplings, and contact dynamics make it challenging to manually design effective hand motions. Teleoperation provides a natural interface for accessing human dexterity, where an operator controls a robotic hand through their own hand motions while observing the robot's response in real time [33, 43, 55]. In this process, retargeting translates operator hand kinematics into feasible and intuitive robot hand motions. The resulting teleoperated interactions can further serve as high-quality demonstrations for imitation learning [21, 48, 57], making retargeting important for both real-time control and data-driven robot learning.

As this interface, retargeting should not be reduced to direct pose matching between human and robot hands. Because the operator and robot hands may differ in scale, motion range, joint coupling, and feasible configurations [29, 51], a natural operator motion may not directly correspond to a natural robot motion. A practical retargeting algorithm should therefore satisfy three requirements:

- **(R1) Intuitiveness.** It should preserve the operator's motion intent and produce feasible, predictable robot motions that are easy to control.
- **(R2) Calibration Efficiency.** It should reduce dependence on precise calibration or manually tuned hyperparameters such as scale factors, offsets, and task weights.
- **(R3) Generality.** The same operator hand space should control different human-like dexterous hands without hand-specific redesign or handcrafted tuning.

Existing methods still fall short of these requirements. Traditional retargeting methods [7, 18, 19, 32, 33, 54] often rely on inverse kinematics with hand-crafted task vectors or keypoints, requiring careful calibration of coordinate frames, vector origins, scale factors, and objective weights. Neural

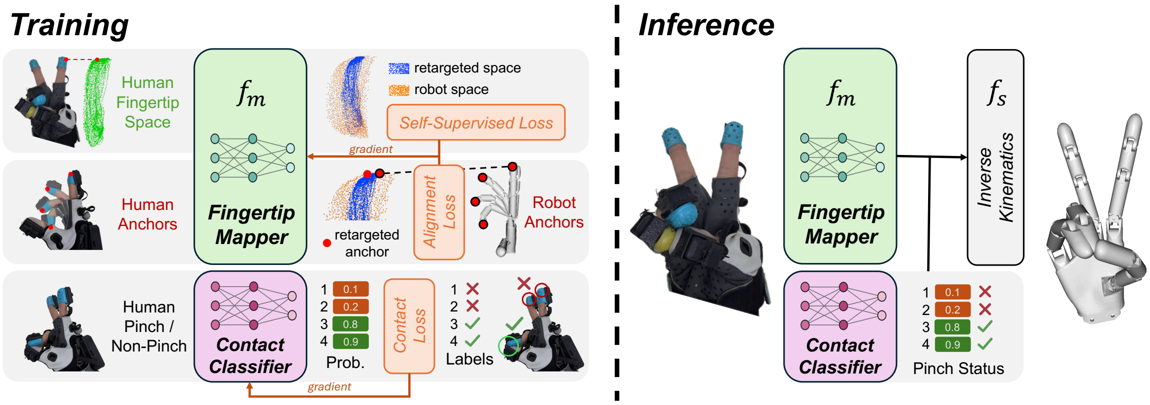

Fig. 3: Overview of AnyDexRT. *(Left)* During training, AnyDexRT learns a fingertip mapper from collected human fingertip samples with few paired human-robot anchors as guidance, and trains a contact classifier using collected pinch/non-pinch poses. *(Right)* During deployment, the fingertip mapper produces retargeted targets, which are refined by the contact classifier and converted into robot joint commands.

variants [20, 23, 43] reduce runtime cost by imitating such retargeting pipelines, but remain tied to calibrated targets and hand-specific assumptions. GeoRT [55] instead learns correspondence by globally aligning the reachable fingertip spaces of human and robot hands. However, global shape matching can be problematic when dexterous hands contain redundant feasible regions with no natural counterpart in the operator hand space, potentially distorting task-relevant mappings and yielding geometrically plausible but less intuitive teleoperation.

To this end, we propose AnyDexRT, a calibration-free retargeting method that enables intuitive dexterous teleoperation across different human-like dexterous hands with few-shot manual tuning. AnyDexRT learns fingertip-level correspondence via self-supervised shape matching and uses few-shot human guidance to anchor the mapping in task-relevant regions, reducing ambiguity caused by redundant robot-hand motion spaces. It further refines pinch-related poses with a contact classifier to improve the performance of pinch motions. These designs reduce hand-specific calibration and tuning while maintaining an intuitive operator-to-robot mapping across different hands. Experiments across multiple dexterous hands and diverse real-world teleoperation tasks demonstrate that AnyDexRT improves retargeting quality, reduces manual tuning effort, and provides more intuitive teleoperation than prior methods.

II. AnyDexRT

In this section, we present AnyDexRT, a calibration-efficient pipeline for mapping operator fingertip motions to robot hand commands. After formulating the retargeting problem and assumptions (§II-A), we introduce how AnyDexRT learns fingertip correspondence through self-supervised shape matching (§II-B), anchors task-relevant regions with few-shot human guidance (§II-C), and refines pinch-related poses using a contact classifier (§II-D). The overview of the system is shown in Fig. 3.

A. Preliminary

For a hand with F fingertips and D joints, let $\mathbf{C}^H, \mathbf{C}^R \subset \mathbb{R}^{F\times 3}$ be the set of 3D human and robot fingertip positions respectively, and $\mathbf{J}^R \subset \mathbb{R}^D$ be the set of the robot hand joints. We aim to find a mapping $f: \mathbf{C}^H \to \mathbf{J}^R$ from human fingertips to robot hand joints. The mapping f can be modeled as a composite of two functions, *i.e.*, $f = f_s \circ f_m$, where $f_m: \mathbf{C}^H \to \mathbf{C}^R$ maps human fingertip positions to robot fingertip positions, and $f_s: \mathbf{C}^R \to \mathbf{J}^R$ solves robot hand joints using robot fingertip positions.

To simplify the formulation while preserving generality, we make the following assumptions:

- **(A1) The robotic hand is structurally similar to the human hand.** For human-like dexterous hands, stable finger coupling and low-dimensional postural synergies [11, 34, 39] allow fingertip positions and reference joint angles to sufficiently constrain inverse kinematics (IK), so f_s can be approximated as a one-to-one mapping.
- **(A2) The human fingertip motion space can be covered by the robotic fingertip space after a suitable geometric transformation.** This coverage is only required from human to robot, ensuring that natural human manipulation motions can be retained, while redundant robot-only regions need not be covered by the human hand.

In this work, we mainly focus on the implementation of the fingertip mapper f_m. The function f_s can be implemented through inverse kinematics, nearest neighbor search, or neural networks.

B. Fingertip Mapping by Self-Supervised Shape Correspondence

Fingertip trajectories reside on a manifold, which allows hand retargeting to be interpreted as shape correspondence between human and robotic fingertips. This correspondence can be learned in a self-supervised manner using several criteria.

*1) **Partial Correspondence**:* The Chamfer loss is commonly used for shape correspondence [4, 12], and has been adopted in prior retargeting work [55]. Its bidirectional formulation encourages full coverage between the mapped human fingertip space $f_m^i(\mathbf{C}^{H,i})$ and the robot fingertip space $\mathbf{C}^{R,i}$ for the i-th finger. However, this assumption is not always suitable for dexterous hand retargeting. According to **(A2)**, not every

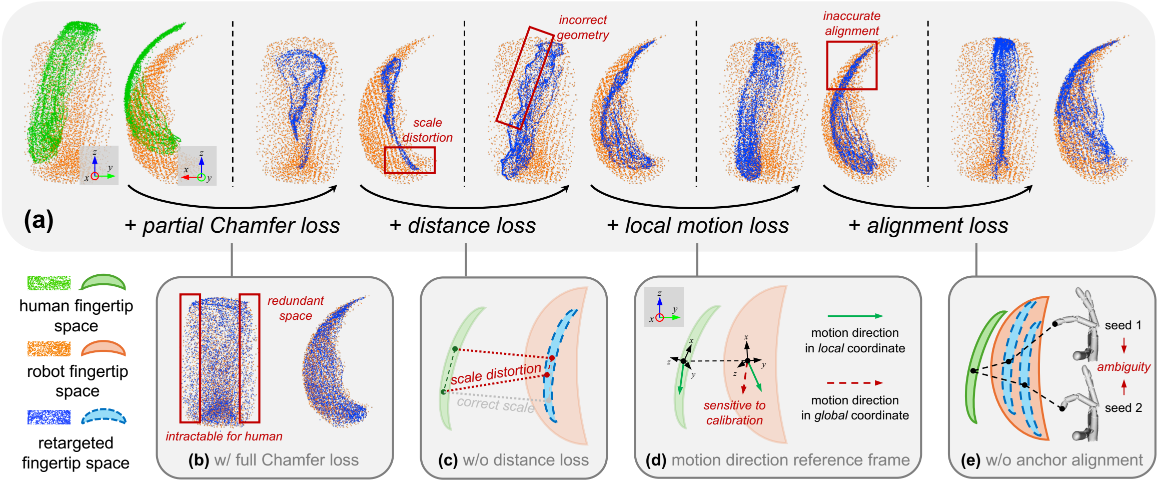

Fig. 4: Retargeting Objectives. (a) Combining all objectives produces a geometrically consistent and intuitive mapping from the human fingertip space to the robot fingertip space. **(b)** Full Chamfer loss can force unnatural coverage of redundant robot fingertip regions. **(c)** Distance loss preserves the geometric structure of the retargeted space and reduces mapping distortion. **(d)** Local motion preservation encourages consistent motion directions between human and robot fingertips, and is less sensitive to calibration compared to global motion preservation. **(e)** Few-shot anchor alignment resolves mapping ambiguity and stabilizes retargeting.

feasible robotic fingertip position has a natural counterpart in the human fingertip space. For highly dexterous hands with redundant feasible regions, forcing $f_m^i(\mathbf{C}^{H,i})$ to cover the entire $\mathbf{C}^{R,i}$ may distort the learned mapping and produce an unnatural spatial distribution, as illustrated in Fig. 4(b). We therefore employ a partial Chamfer loss [50]:

$$\mathscr{L}_{\text{P-Chamfer}}(\mathbf{C}^{H,i},\mathbf{C}^{R,i}) = \frac{1}{|\mathbf{C}^{H,i}|}\sum_{j=1}^{|\mathbf{C}^{H,i}|}\min_k \left\| f_m^i(x_j^{H,i}) - x_k^{R,i} \right\|, \tag{1}$$

where $\|\cdot\|$ stands for L2 normalization. This asymmetric objective maps the human fingertip space into the feasible robot fingertip space without requiring the robot space to be fully covered, thereby preserving the structure of natural operator motions while reducing distortions caused by redundant robot-hand regions.

2) ***Distance Preservation:*** Although the partial Chamfer loss enforces a one-way mapping, it does not adequately preserve the original geometric distribution in $\mathbf{C}^H$. Fig. 4(c) illustrates an example of the geometric distortion. To mitigate this effect, we introduce a pairwise distance preservation objective. For any $x_{j_1}^{H,i}, x_{j_2}^{H,i} \in \mathbf{C}^{H,i}$, we have

$$\mathscr{L}_{\text{dist}}(\mathbf{C}^{H,i}) = \frac{1}{|\mathbf{C}^{H,i}|(|\mathbf{C}^{H,i}|-1)} \sum_{j_1 \neq j_2} \left(\left\| f_m^i(x_{j_1}^{H,i}) - f_m^i(x_{j_2}^{H,i}) \right\| - \left\| x_{j_1}^{H,i} - x_{j_2}^{H,i} \right\| \right)^2. \tag{2}$$

This objective regularizes the distribution of $f_m(\mathbf{C}^H)$ and alleviates geometric distortion.

3) ***Local Motion Preservation:*** Partial correspondence aligns fingertip spaces, but does not guarantee that local motion directions are preserved. GeoRT [55] imposes directional consistency in the global hand-base frame, implicitly assuming well-calibrated fingertip measurements. In practice, glove-sensor calibration errors may introduce translational or rotational discrepancies between measured and actual fingertip positions, causing globally measured directions to be misaligned. We therefore enforce motion preservation in local coordinate frames, where the relative movement directions of the two spaces remain more consistent, as illustrated in Fig. 4(d). Let $\mathbf{T}(x)$ denote the local coordinate frame at position x. For a small perturbation Δx around $x_j^{H,i}$, the induced displacement in the robot space is $\Delta f_m^i(x_j^{H,i}) = f_m^i(x_j^{H,i}+\Delta x) - f_m^i(x_j^{H,i})$, and the local motion direction is $\mathbf{D}(x) = \mathbf{T}^{-1}(x)\frac{\Delta x}{\|\Delta x\|}$. We define the local motion loss as

$$\mathscr{L}_{\text{motion}}(\mathbf{C}^{H,i}) = -\frac{1}{|\mathbf{C}^{H,i}|}\sum_{j=1}^{|\mathbf{C}^{H,i}|} \left\langle \mathbf{D}\left(x_j^{H,i}\right), \mathbf{D}\left(f_m^i(x_j^{H,i})\right) \right\rangle, \tag{3}$$

where $\langle\cdot,\cdot\rangle$ denotes the vector inner product. Since $f_m^i(x_j^{H,i})$ does not provide a local rotation, we assign it the rotation of its nearest neighbor in $\mathbf{C}^{R,i}$. Together with $\mathscr{L}_{\text{P-Chamfer}}$ and $\mathscr{L}_{\text{dist}}$, this local motion loss encourages retargeting behavior that is less sensitive to calibration errors and more consistent with operator intent.

C. Geometric Alignment with Few-Shot Human Guidance

Self-supervised learning allows the model to learn human-robot shape correspondence from unpaired fingertip samples, but the solution is not unique. As illustrated in Fig. 4(e), the same human fingertip distribution $\mathbf{C}^H$ may be mapped to multiple plausible regions in the robot space $\mathbf{C}^R$, leading to unstable training and mappings that are sensitive to random sampling or initialization. Such ambiguity can also introduce translation or scale offsets in fingertip positions, resulting in unnatural joint configurations and less reliable teleoperation.

To anchor the correspondence without requiring large-scale paired data, we introduce few-shot human guidance to the system. Operators imitate a small set of reference gestures shown in Fig. 3, from which we collect paired human-robot fingertip anchors; the collection process is described in Appendix. Given these anchors, we define the alignment loss

Retargeting Method	Inspire Hand [38]		Ability Hand [31]		XHand [36]		Wuji Hand [45]	
	GMC	LMC	GMC	LMC	GMC	LMC	GMC	LMC
Optimization [33]*	53.9	43.2	47.7	39.5	46.9	33.8	80.2	74.2
GeoRT [55]	$71.9_{\pm 3.5}$	$50.6_{\pm 3.6}$	$65.0_{\pm 4.7}$	$36.2_{\pm 11.2}$	$75.4_{\pm 4.1}$	$51.3_{\pm 4.2}$	$87.3_{\pm 2.6}$	$77.0_{\pm 4.6}$
AnyDexRT *(ours)*	$\mathbf{80.7}_{\pm 0.2}$	$\mathbf{89.5}_{\pm 0.2}$	$\mathbf{84.1}_{\pm 0.2}$	$\mathbf{88.5}_{\pm 0.1}$	$\mathbf{76.0}_{\pm 0.3}$	$\mathbf{88.5}_{\pm 0.2}$	$\mathbf{93.1}_{\pm 0.2}$	$\mathbf{92.3}_{\pm 0.2}$
Retargeting Method	**Allegro Hand** [1]		**Leap Hand** [40]		**Shadow Hand** [35]		**Average** (7 Hands)	
	GMC	LMC	GMC	LMC	GMC	LMC	GMC	LMC
Optimization [33]*	**89.5**	76.1	70.8	60.0	45.0	38.5	62.0	52.2
GeoRT [55]	$85.5_{\pm 1.9}$	$74.8_{\pm 1.6}$	$\mathbf{73.4}_{\pm 3.2}$	$53.2_{\pm 8.8}$	$\mathbf{89.8}_{\pm 1.6}$	$75.2_{\pm 4.1}$	78.3	59.8
AnyDexRT *(ours)*	$87.4_{\pm 0.2}$	$\mathbf{92.1}_{\pm 0.1}$	$54.5_{\pm 0.4}$	$\mathbf{89.0}_{\pm 0.2}$	$83.6_{\pm 0.1}$	$\mathbf{91.4}_{\pm 0.3}$	**79.9**	**90.2**

TABLE I: Simulation Results ($\times 10^{-2}$) on Retargeting Quality. AnyDexRT achieves better global and local motion consistency across diverse dexterous hands. *Deterministic methods do not have standard deviations.

as

$$\mathscr{L}_{\text{align}}(\mathbf{C}^{H,i}, \mathbf{C}^{R,i}) = \frac{1}{M} \sum_{j=1}^{M} \left\| f_m^i(\bar{x}_j^{H,i}) - \bar{x}_j^{R,i} \right\|, \tag{4}$$

where $\bar{x}_j^{H,i}$ and $\bar{x}_j^{R,i}$ denote the j-th paired anchor for the i-th finger, and M is the number of anchors. This sparse supervision resolves mapping ambiguity while preserving the scalability of self-supervised correspondence learning. It also enables personalized retargeting, as operators can specify preferred scales and reference poses with only a few intuitive gestures.

D. Pinch Pose Refinement using Contact Classifier

Grasping tiny objects requires higher retargeting accuracy, as even small positional errors can lead to task failure. This is especially critical for pinch motions, where precise fingertip contact is required. As shown in Fig. 5, due to sensor limitations and potential electromagnetic interference, the measured hand pose may still deviate significantly from the actual hand gesture, even when the operator makes clear fingertip contact. As a result, the measured pose fails to capture the geometric characteristics of a pinch.

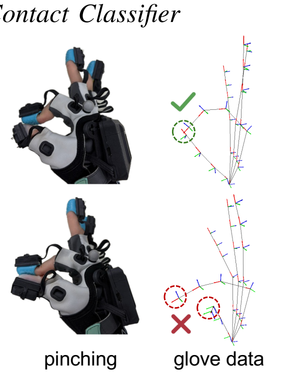

Fig. 5: Sensor Failure during Pinches. Measured data fail to capture hand pinches.

To address this issue, we no longer rely solely on fingertip position mapping to infer fine contact patterns. Since finger contact signals are easy to collect and annotate, and can directly reflect key manipulation intentions such as pinching, we instead train a classifier f_c to recognize such contact patterns. Let 0 be the thumb index and $i > 0$ be the index of other fingers, we have

$$\mathscr{L}_{\text{contact}}(\mathbf{C}^{H,i}) = \frac{1}{N} \sum_{j=1}^{N} \text{BCE}\left(y_j^i, f_c^i(x_j^{H,0}, x_j^{H,i})\right), \tag{5}$$

where N denotes the number of training samples, $\text{BCE}(\cdot,\cdot)$ denotes the binary cross entropy loss, and $y_j^i \in \{0, 1\}$ denotes the label of the j-th sample.

The contact classifier provides a more reliable retargeting signal for grasping tiny objects. During inference, when a human fingertip contact is detected, we search the neighborhood of the corresponding mapped robotic position for a robotic pinch pose, ensuring stable grasping.

III. Experiments

In this section, we assess AnyDexRT through simulation experiments and a real-world teleoperation evaluation. The evaluation aims to verify whether AnyDexRT satisfies the three requirements introduced in §I, including **(R1)** intuitiveness, **(R2)** calibration efficiency, and **(R3)** generality, and to compare its retargeting quality, efficiency, and operator-perceived intuitiveness with prior methods.

We conduct simulated experiments on seven human-like dexterous hands listed in Tab. I, and evaluate the retargeting performance in four real-world teleoperation tasks (Fig. 6) in Tab. II. Please refer to the Appendix for evaluation metrics, analyses, and more experiments.

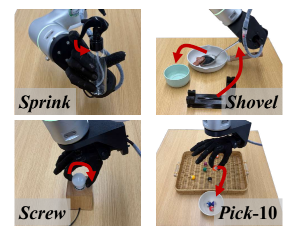

Fig. 6: Real-World Evaluation Tasks.

Retargeting Method	Time per Episode (s) ↓				Pinch Success Rate ↑
	Sprink	*Screw*	*Shovel*	*Pick-10*	
Optimization [33]	29.0	25.3	36.4	150.4	39.6%
GeoRT [55]	32.1	22.8	38.5	220.4	29.2%
AnyDexRT *(ours)*	**10.6**	**17.0**	**28.0**	**105.8**	**62.0**%

TABLE II: Real-World Teleoperation Performance. AnyDexRT improves task efficiency and pinch success rates over prior retargeting methods, demonstrating its effectiveness in real-world teleoperation and its applicability to collecting dexterous manipulation data.

IV. Conclusion

We present AnyDexRT, a calibration-free retargeting method for intuitive teleoperation across different human-like dexterous hands. AnyDexRT learns fingertip mapping through self-supervised shape matching, while using few-shot human guidance to anchor task-relevant regions and reduce ambiguity caused by redundant robot fingertip spaces. It further refines pinch-related poses with a contact classifier, improving the responsiveness and controllability of contact-critical motions. Experiments across multiple dexterous hands demonstrate that AnyDexRT achieves strong retargeting quality with lower tuning effort and better stability than representative baselines. Real-world teleoperation evaluations further show that AnyDexRT enables more efficient task completion and more reliable pinch control, supporting its applicability to dexterous teleoperation and data collection.

Acknowledgement

We would like to thank Hao-Shu Fang from the Massachusetts Institute of Technology for the insightful discussions. We are also grateful to Shirun Tang, Junchao Zhang, Zelin Ye from Noematrix, and Zihao He from Shanghai Jiao Tong University for their support during the real-world teleoperation evaluation, and to Shixuan Huang from Noematrix for his contributions to the mechanical design of the connector between the arm and the dexterous hand.

References

[1] Allegro. *Allegro Hand V4*. May 2026. URL: https://www.allegrohand.com/sub/product/p.php?idx=1.

[2] Shan An et al. "Dexterous Manipulation Through Imitation Learning: A Survey". In: *IEEE Transactions on Automation Science and Engineering* 23 (2025), pp. 1760–1792.

[3] Dafni Antotsiou, Guillermo Garcia-Hernando, and Tae-Kyun Kim. "Task-Oriented Hand Motion Retargeting for Dexterous Manipulation Imitation". In: *European Conference on Computer Vision Workshops*. 2018.

[4] Harry G. Barrow et al. "Parametric Correspondence and Chamfer Matching: Two New Techniques for Image Matching". In: *International Joint Conference on Artificial Intelligence*. William Kaufmann, 1977, pp. 659–663.

[5] Tao Chen et al. "Visual Dexterity: In-Hand Reorientation of Novel and Complex Object Shapes". In: *Science Robotics* 8.84 (2023), eadc9244.

[6] Tao Chen et al. "Vegetable Peeling: A Case Study in Constrained Dexterous Manipulation". In: *IEEE International Conference on Robotics and Automation*. 2025, pp. 4542–4550.

[7] Xuxin Cheng et al. "Open-TeleVision: Teleoperation with Immersive Active Visual Feedback". In: *Conference on Robot Learning*. 2024.

[8] Cheng Chi et al. "Universal Manipulation Interface: In-The-Wild Robot Teaching Without In-The-Wild Robots". In: *Robotics: Science and Systems*. 2024.

[9] Eunsuk Chong, Lionel Zhang, and Veronica J Santos. "A Learning-Based Harmonic Mapping: Framework, Assessment, and Case Study of Human-to-Robot Hand Pose Mapping". In: *The International Journal of Robotics Research* 40.2-3 (2021), pp. 534–557.

[10] Runyu Ding et al. "Bunny-VisionPro: Real-Time Bimanual Dexterous Teleoperation for Imitation Learning". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2025, pp. 12248–12255.

[11] Roel Duits, Arjan Egges, and A. Frank van der Stappen. "A Closed-Form Solution for Human Finger Positioning". In: *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games, MIG 2015, Paris, France, November 16-18, 2015*. ACM, 2015, pp. 73–78.

[12] Haoqiang Fan, Hao Su, and Leonidas J. Guibas. "A Point Set Generation Network for 3D Object Reconstruction from a Single Image". In: *CVPR*. IEEE Computer Society, 2017, pp. 2463–2471.

[13] Hao-Shu Fang et al. "AnyDexGrasp: General Dexterous Grasping for Different Hands with Human-Level Learning Efficiency". In: *arXiv preprint arXiv:2502.16420* (2025).

[14] Hao-Shu Fang et al. "DEXOP: A Device for Robotic Transfer of Dexterous Human Manipulation". In: *arXiv preprint arXiv:2509.04441* (2025).

[15] Hongjie Fang et al. "AirExo: Low-Cost Exoskeletons for Learning Whole-Arm Manipulation in the Wild". In: *IEEE International Conference on Robotics and Automation*. 2024.

[16] Hongjie Fang et al. "AirExo-2: Scaling up Generalizable Robotic Imitation Learning with Low-Cost Exoskeletons". In: *Conference on Robot Learning*. 2025.

[17] Ying Feng et al. "Learning Dexterous Manipulation with Quantized Hand State". In: *IEEE International Conference on Robotics and Automation*. 2026.

[18] Ankur Handa et al. "DexPilot: Vision-Based Teleoperation of Dexterous Robotic Hand-Arm System". In: *IEEE International Conference on Robotics and Automation*. 2020.

[19] Guanqi He and Wentao Zhang. *WujiHand Retargeting*. 2026. URL: https://github.com/wuji-technology/wuji-retargeting.

[20] Liang Heng et al. "HumDex: Humanoid Dexterous Manipulation Made Easy". In: *arXiv preprint arXiv:2603.12260* (2026).

[21] Guangqi Jiang et al. "Cross-Hand Latent Representation for Vision-Language-Action Models". In: *arXiv preprint arXiv:2603.10158* (2026).

[22] Jinzhou Li et al. "Adaptive Visuo-Tactile Fusion with Predictive Force Attention for Dexterous Manipulation". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2025, pp. 3232–3239.

[23] Shuang Li et al. "Vision-Based Teleoperation of Shadow Dexterous Hand Using End-to-End Deep Neural Network". In: *IEEE International Conference on Robotics and Automation*. 2019, pp. 416–422.

[24] Toru Lin et al. "Learning Visuotactile Skills With Two Multifingered Hands". In: *IEEE International Conference on Robotics and Automation*. 2025, pp. 5637–5643.

[25] Yuhao Lin et al. "TypeTele: Releasing Dexterity in Teleoperation by Dexterous Manipulation Types". In: *Conference on Robot Learning*. 2025, pp. 4975–4993.

[26] Hangxin Liu et al. "A Glove-Based System for Studying Hand-Object Manipulation via Joint Pose and Force Sensing". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2017, pp. 6617–6624.

[27] Hangxin Liu et al. "High-Fidelity Grasping in Virtual Reality using a Glove-Based System". In: *IEEE International Conference on Robotics and Automation*. 2019, pp. 5180–5186.

[28] MANUS. *MANUS Quantum Metagloves*. May 2026. URL: https://www.manus-meta.com/products/quantum-metagloves.

[29] Roberto Meattini et al. "Human to Robot Hand Motion Mapping Methods: Review and Classification". In: *IEEE Transactions on Robotics* 39.2 (2022), pp. 842–861.

[30] Allison M. Okamura, Niels Smaby, and Mark R. Cutkosky. "An Overview of Dexterous Manipulation". In: *IEEE International Conference on Robotics and Automation*. 2000.

[31] Psyonic. *Ability Hand*. May 2026. URL: https://www.psyonic.io/ability-hand.

[32] Yuzhe Qin et al. "DexMV: Imitation Learning for Dexterous Manipulation from Human Videos". In: *European Conference on Computer Vision*. 2022, pp. 570–587.

[33] Yuzhe Qin et al. "AnyTeleop: A General Vision-Based Dexterous Robot Arm-Hand Teleoperation System". In: *Robotics: Science and Systems*. 2023.

[34] Hans Rijpkema and Michael Girard. "Computer animation of knowledge-based human grasping". In: *Proceedings of the 18th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1991, Providence, RI, USA, April 27-30, 1991*. ACM, 1991, pp. 339–348.

[35] Shadow Robot. *Shadow Dexterous Hand Series*. May 2026. URL: https://shadowrobot.com/dexterous-hand-series/.

[36] RobotEra. *RobotEra XHand1*. May 2026. URL: https://www.robotera.com/en/goods1/4.html#/product/XHAND.

[37] Flexiv Robotics. *Flexiv Rizon Arm*. May 2026. URL: https://www.flexiv.com/product/rizon.

[38] Inspire Robots. *Inspire Hand RH56BFX*. May 2026. URL: https://en.inspire-robots.com/product/rh56bfx.

[39] Marco Santello, Martha Flanders, and John F. Soechting. "Postural Hand Synergies for Tool Use". In: *Journal of Neuroscience* 18.23 (1998), pp. 10105–10115.
[40] Kenneth Shaw, Ananye Agarwal, and Deepak Pathak. "LEAP Hand: Low-Cost, Efficient, and Anthropomorphic Hand for Robot Learning". In: *Robotics: Science and Systems*. 2023.
[41] Kenneth Shaw et al. "Bimanual Dexterity for Complex Tasks". In: *Conference on Robot Learning*. 2024.
[42] Zilin Si et al. "Tilde: Teleoperation for Dexterous In-Hand Manipulation Learning with a DeltaHand". In: *Robotics: Science and Systems*. 2024.
[43] Aravind Sivakumar, Kenneth Shaw, and Deepak Pathak. "Robotic Telekinesis: Learning a Robotic Hand Imitator by Watching Humans on Youtube". In: *Robotics: Science and Systems*. 2022.
[44] HJ Terry Suh et al. "Dexterous Contact-Rich Manipulation via the Contact Trust Region". In: *The International Journal of Robotics Research* (2025), p. 02783649251398875.
[45] Wuji Technology. *Wuji Hand*. May 2026. URL: https://wuji.tech/en/hand.
[46] HTC Vive. *HTC Vive Tracker 3.0*. May 2026. URL: https://www.vive.com/us/accessory/tracker3/.
[47] Chen Wang et al. "DexCap: Scalable and Portable Mocap Data Collection System for Dexterous Manipulation". In: *Robotics: Science and Systems*. 2024.
[48] Zhenyu Wei, Yunchao Yao, and Mingyu Ding. "One Hand to Rule Them All: Canonical Representations for Unified Dexterous Manipulation". In: *Robotics: Science and Systems*. 2026.
[49] Ruoshi Wen et al. "Dexterous Teleoperation of 20-DoF ByteDexter Hand via Human Motion Retargeting". In: *arXiv preprint arXiv:2507.03227* (2025).
[50] Xin Wen et al. "Cycle4Completion: Unpaired Point Cloud Completion Using Cycle Transformation With Missing Region Coding". In: *CVPR*. Computer Vision Foundation / IEEE, 2021, pp. 13080–13089.
[51] Chendong Xin et al. "Analyzing Key Objectives in Human-to-Robot Retargeting for Dexterous Manipulation". In: *IEEE Robotics and Automation Practice* (2026).
[52] Mengda Xu et al. "DexUMI: Using Human Hand as the Universal Manipulation Interface for Dexterous Manipulation". In: *Conference on Robot Learning*. 2025.
[53] Yinzhen Xu et al. "UniDexGrasp: Universal Robotic Dexterous Grasping via Learning Diverse Proposal Generation and Goal-Conditioned Policy". In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023, pp. 4737–4746.
[54] Shiqi Yang et al. "ACE: A Cross-Platform and Visual-Exoskeletons System for Low-Cost Dexterous Teleoperation". In: *Conference on Robot Learning*. PMLR. 2024, pp. 4895–4911.
[55] Zhao-Heng Yin et al. "Geometric Retargeting: A Principled, Ultrafast Neural Hand Retargeting Algorithm". In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2025, pp. 17376–17382.
[56] Di Zhang et al. "KineDex: Learning Tactile-Informed Visuomotor Policies via Kinesthetic Teaching for Dexterous Manipulation". In: *Conference on Robot Learning*. 2025, pp. 4123–4138.
[57] Gu Zhang et al. "UniDex: A Robot Foundation Suite for Universal Dexterous Hand Control from Egocentric Human Videos". In: *arXiv preprint arXiv:2603.22264* (2026).
[58] Han Zhang et al. "DOGlove: Dexterous Manipulation with a Low-Cost Open-Source Haptic Force Feedback Glove". In: *Robotics: Science and Systems*. 2025.

Appendix

A. Video Demo

We provide a video demo for the four real-world teleoperation tasks and an overview of AnyDexRT. See "video.mp4" for details.

B. Implementation Details

*1) **Networks**:* Both the finger mapper f_m and the contact classifier f_c are implemented using MLPs. For a hand with F fingers, we employ F sub-nets for f_m and $(F-1)$ sub-nets for f_c. All sub-nets of f_m have the same size of $(3,128,128,3)$, whose input and output are human fingertip positions and retargeted fingertip positions, respectively. All sub-nets of f_c have the same size of $(6,128,128,1)$ followed by the Sigmoid function. The input of $f_c^i(i>0)$ is a fingertip position concatenation of the thumb and the i-th finger, and the output is a one-dimensional score distributed in $[0,1]$.

*2) **Data Preprocessing**:* For each human/robot finger, we centralize the fingertip position, compute the ranges along each axis, and select the axis with the maximum range. The selected range is then used to normalize the positions. The normalized positions are distributed in $[-1,1]$ without geometric distortion.

*3) **Training**:* The finger mapper f_m and the contact classifier f_c can be trained in parallel since they are independent from each other. Both networks are trained for 20 epochs with the learning rate of 0.0001, and the batch size of 2048. During the training of f_m, we randomly perturb each sample with a generated delta movement and compute $\mathscr{L}_{\text{motion}}$ using the jittored positions. Anchor data with the batch size of 32 is also fed into f_m for the computation of $\mathscr{L}_{\text{align}}$. The loss for training f_m is $\mathscr{L}_{\text{mapping}} = \mathscr{L}_{\text{P-Chamfer}} + \mathscr{L}_{\text{dist}} + \mathscr{L}_{\text{motion}} + \mathscr{L}_{\text{anchor}}$, and we do not specifically tune the weights of each loss term.

*4) **Inference**:* The retargeted fingertip positions predicted by f_m need to be transformed to joint configurations. We simply use nearest neighbor search (NNS) to get the corresponding joint configurations from the given fingertip positions and the last joint configurations. It can also be implemented using inverse kinematics or neural networks, but we find NNS is sufficient for our experiments. We use a threshold of 0.5 for the scores predicted by f_c. For the predicted contact pattern of two fingers, we simply search the nearest fingertip positions from pre-generated templates using the corresponding positions predicted by f_m.

C. Simulation Experiments

*1) **Setup**:* We conduct simulated experiments on seven human-like dexterous hands listed in Tab. I to evaluate **(R3)**. These hands span diverse kinematic structures and range from 6 to 20 DoFs, testing whether each method can adapt to different hand embodiments. To evaluate retargeting quality, we measure motion consistency [55], which reflects whether the robot hand responds consistently with the operator's motion intent **(R1)**. We report global motion consistency (GMC) and local motion consistency (LMC). GMC follows GeoRT [55] and compares displacement directions in a shared coordinate frame, assuming an ideal, well-calibrated setting. LMC compares directions in local frames, making it less calibration-dependent and more aligned with operator control. We omit whole-space coverage as a main metric, since covering redundant robot regions does not necessarily improve teleoperation intuitiveness. We compare AnyDexRT with a representative optimization-based retargeting method [33] (offline version) and the neural retargeting method GeoRT [55]. For methods involving stochastic training or sampling, we report results over 5 random seeds to evaluate their stability and sensitivity to initialization.

*2) **Retargeting Quality**:* Tab. I reports retargeting quality across seven dexterous hands. AnyDexRT achieves strong performance across different hand embodiments, improving the average local motion consistency from 59.8% to 90.2%. This shows that AnyDexRT better preserves the operator's motion intent, supporting **(R1)**. Although optimized for local consistency, AnyDexRT also maintains competitive global motion consistency under an ideal, well-calibrated shared frame. Moreover, AnyDexRT achieves consistently high scores with small standard deviations across hands, while GeoRT exhibits larger variations across hands and random seeds. These results show that AnyDexRT provides more stable and general retargeting, supporting **(R3)**.

*3) **Manual Tuning Effort and Retargeting Efficiency**:* We further compare manual tuning effort and the runtime efficiency of different retargeting methods in Tab. III. Similar to GeoRT, AnyDexRT requires only 3 hyperparameters and runs at about 300Hz, outperforming the optimization baseline in both tuning effort and speed. In practice, the default hyperparameters of AnyDexRT rarely need adjustment and can be directly reused across different hands. These results show that AnyDexRT reduces tuning effort while maintaining real-time retargeting efficiency, supporting **(R2)**.

*4) **Calibration Sensitivity**:* We use Wuji Hand to evaluate calibration sensitivity. To simulate frame misalignment between glove measurements and the robot hand, we rotate the input human data around the y-axis by $-90°$, $-45°$, $45°$, and $90°$. We then calculate LMC under each perturbation to assess whether the mapping preserves operator motion intent despite calibration errors. As shown in Fig. 7, the optimization-based baseline and GeoRT degrade substantially under these perturbations, indicating their dependence on accurate coordinate alignment. In contrast, AnyDexRT maintains stable retargeting quality across all rotations, demonstrating its calibration-free ability to adapt to misaligned hand coordinate systems and better satisfy **(R2)**.

*5) **Training Stability**:* Fig. 8 reports the training stability of AnyDexRT and GeoRT [55], which is computed by averaging the standard deviations of the retargeted positions over 5 random seeds. We can observe that AnyDexRT achieves orders-of-magnitude better stability than GeoRT under different training initializations across all hands. Such stable mapping ensures the reproducibility of the results and substantially reduces the difficulty of teleoperation.

Method	# HP	Speed (Hz)
Optimization [33]	≥ 10	93.4
GeoRT [55]	4	281.7
ours	**3**	**293.0**

TABLE III: Retargeting Comparisons. HP is hyperparameter.

Objective	LMC ($\times 10^{-2}$)
$\mathcal{L}_{\text{P-Chamfer}}$	4.1
+ $\mathcal{L}_{\text{dist}}$	83.8
+ $\mathcal{L}_{\text{motion}}$	89.1
+ $\mathcal{L}_{\text{align}}$	92.3

TABLE IV: Ablation Results.

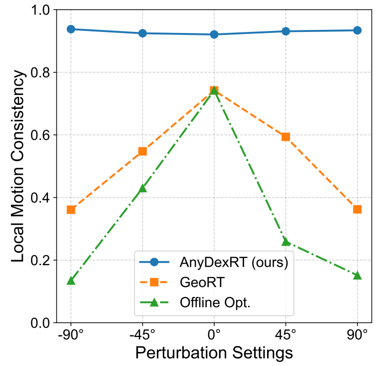

Fig. 7: Calibration Sensitivity. AnyDexRT maintains consistent retargeting quality, demonstrating its calibration-free capability.

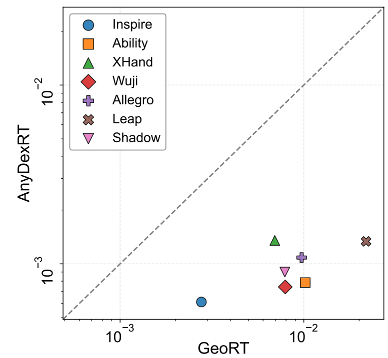

Fig. 8: Retargeting Standard Deviations. AnyDexRT shows strong mapping stability over GeoRT across different training initializations.

*6) **Ablation**:* We provide qualitative ablations in Fig. 4(a) and quantitative results on Wuji Hand in Tab. IV. $\mathcal{L}_{\text{dist}}$ and $\mathcal{L}_{\text{motion}}$ improve LMC by preserving the global geometric structure and local motion directions, respectively. Adding $\mathcal{L}_{\text{align}}$ further anchors the mapping with few-shot guidance, contributing to more natural and intuitive retargeting.

D. Real-World Teleoperation Performance

*1) **Setup**:* We use a Flexiv Rizon 4 arm [37] equipped with a Wuji Hand [45] as the robot platform. The operator wears a Manus glove [28] with an HTC Vive Tracker [46]; the glove provides fingertip poses for hand retargeting, and the tracker controls the arm. Each retargeting method converts the glove fingertip poses into real-time Wuji Hand joint commands. We evaluate four real-world tasks: spray-bottle triggering (***Sprink***), light-bulb screwing (***Screw***), steak shoveling (***Shovel***), and small-ball picking (***Pick*-10**). These tasks cover finger-specific actuation, grasping, tool use, and repetitive pinch manipulation. 8 operators with varying teleoperation experience participated in the evaluation. The task descriptions are as follows:

a) Spray-Bottle Triggering (Sprink): Grasp the spray-bottle with thumb, middle, ring, and pinky fingers. Trigger the bottle with the index finger, then release it. This task involves both power grasping and in-hand manipulation.

b) Light-Bulb Screwing (Screw): Use the thumb and index finger to screw in the light-bulb to turn it on. The progress is finished when the light is stable without flashing. This task is designed to evaluate the intuitiveness of the retargeting methods.

c) Steak Shoveling (Shovel): Grasp the spatula with the full hand, use it to shovel up the steak, pour it into the bowl, and then place the spatula back. The task difficulty mainly depends on the stability of the power grasp.

d) Small-Ball Picking (Pick-10): Use the thumb and index finger to grasp the small balls in the basket and place them into the bowl until all 10 balls are successfully transferred. This task is designed to evaluate the accuracy of the retargeted pinch poses. The pinch success rate is computed as $10/A$ where A denotes the number of attempts.

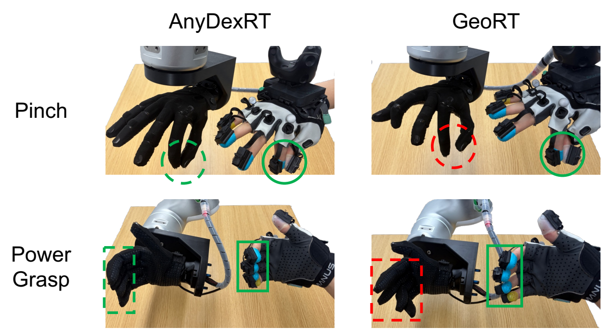

Fig. 9: Analysis of Real-World Teleoperation.

*2) **Intuitiveness and Efficiency**:* As shown in Tab. II, AnyDexRT achieves the shortest completion time on all tasks, suggesting more efficient and predictable teleoperation than the baselines. The improvements are especially clear in ***Sprink*** and ***Shovel***, which require accurate finger response and stable grasp adjustment. These results show that AnyDexRT improves real-world dexterous control quality during teleoperation, supporting **(R1)**.

*3) **Pinch Performance**:* We evaluate pinch control with the ***Pick*-10** task, which requires reliable thumb-finger contacts. AnyDexRT achieves the highest pinch success rate, significantly outperforming both the optimization method and GeoRT. This shows that contact-classifier refinement improves contact-critical pinch motions and supports higher-quality dexterous data collection.

*4) **Case Analysis**:* We compare the retargeted pinch poses and power grasps output by AnyDexRT and GeoRT in Fig. 9. The accuracy of pinch poses decides the operation speed in ***Pick-10***, and the stability of power grasps ensures the success of ***Sprink*** and ***Shovel***. GeoRT fails to predict accurate poses due to the widely distributed redundant fingertip space of the Wuji Hand. With human guidance, AnyDexRT successfully reduces this ambiguity and improves the intuitiveness of teleoperation.

Fig. 10: Qualitative Results of AnyDexRT.

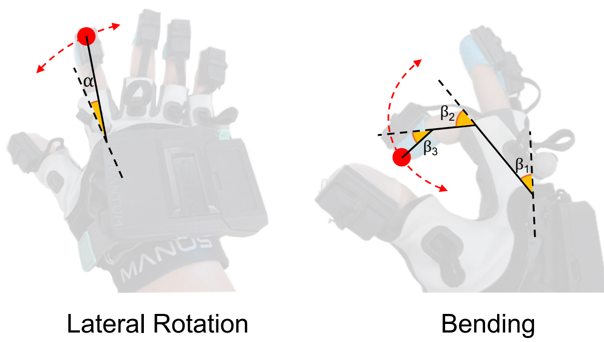

Fig. 11: Anchor Types and Collection.

E. Qualitative Results

Fig. 10 illustrates real-world qualitative results of AnyDexRT on Wuji Hand [45]. We can observe that AnyDexRT provides precise and intuitive hand retargeting on multiple types of gestures.

F. Anchor Data Collection

1) Anchor Type: The anchors can be defined according to the operators' habits. In Fig. 11, we provide the configuration in our experiments, which contains two types of paired anchors: **lateral rotation** and **bending**.

*a) **Lateral Rotation**:* Let $\beta_1 = \beta_2 = \beta_3 = 0$, we rotate the finger by changing α. This type of data can be easily collected by putting one's hand on the table.

*b) **Bending**:* Let $\alpha = 0$, the finger is rotated by changing β_1, β_2 and β_3. To simplify the collection process, we set $\beta_1 = \beta_2 = \lambda\beta_3$. According to [34], λ is usually a constant in bending, and we use $\lambda = 2$ in our experiments. Operators may use other constants ([34] uses $3/2$) based on the characteristics of their own hands. It's hard to precisely define and collect the bending anchor for the thumb, so we just sample several positions from the pre-generated bending trajectory of the robotic thumb finger and let operators imitate the finger pose. We find it is sufficient to conduct the task in our experiments.

2) Collection Process: We describe the collection process for anchor data as follows:

- **(P1) Human Anchor Collection.** We uniformly sample K_0 joint configurations according to the constraint of each anchor type, and ask the operator to wear the glove to collect the corresponding poses for each finger. For simplicity, one can also generate the finger poses in simulation and imitate them during data collection. We set $K_0 = 5$ for both anchor types. Specifically, for bending anchors, we increase β_1 from 0 to $\pi/2$ in increments of $\pi/8$.
- **(P2) Human Anchor Interpolation.** For two neighboring human anchors of the same finger, we generate a continuous trajectory via spatial interpolation and increase the number of anchors to K. We find that linear interpolation is sufficient for our experiments, although more complex interpolation methods can be used for higher precision. We set $K = 50$ for lateral rotation and $K = 100$ for bending.
- **(P3) Robot Anchor Generation.** The robot anchor data is generated in simulation by uniformly sampling K joint configurations and collecting the corresponding fingertip positions. This yields paired human-robot anchor data.

During the process, only **(P1)** needs human participation, while **(P2)** and **(P3)** are executed autonomously by computers. Note that human guidance is designed to make teleoperation more intuitive according to the operator's own habits. Therefore, the operator does not need to strictly align their finger with the defined anchor poses, and slight offsets are acceptable in practice.

G. Related Works

*1) **Kinematic Retargeting for Dexterous Hands**:* Kinematic retargeting converts human hand motion signals into robot joint commands for dexterous hands. Some systems use hand-specific gloves [26, 27] or exoskeletons [14, 52] whose kinematics are designed to match the target hand. While direct and responsive, such hardware is tied to a particular morphology and does not easily generalize to hands with different scales, proportions, or joint layouts. Most systems instead decouple sensing from retargeting: they estimate human hand poses from vision-based trackers [7, 10, 18, 32, 33, 43, 54], motion-capture gloves [17, 20, 41, 47, 49], or other wearable sensors [9, 58], and then pass fingertip poses to a retargeting module.

For the retargeting module, traditional optimization-based methods [3, 7, 10, 18, 19, 32, 33, 47, 49, 54, 58] manually define a set of task-vector pairs $\{(\mathbf{v}_i^H, \mathbf{v}_i^R)\}$ for the human and robotic hands, and solve for robot hand configurations by minimizing objectives such as $\sum_i \left\|\alpha_i \mathbf{v}_i^H - \mathbf{v}_i^R\right\|^2$, where $\{\alpha_i\}$ denotes the scaling factors for each finger. Several methods [9, 19] further introduce hand-specific objectives or constraints for particular robot hands, but such designs often lack generality. In practice, online optimization is commonly stabilized with joint limit, smoothness, or temporal regularization terms, while offline neural approaches [20, 23, 43] train networks to accelerate the retargeting process. However, when these networks are trained from calibrated optimization targets, they still inherit the assumptions and tuning requirements of the underlying retargeting formulation.

Recent work has explored learning human-robot correspondence with less manual correspondence design. GeoRT [55] reduces manual correspondence design by treating retargeting as space alignment and amortizing the mapping into a neural policy. However, its formulation introduces several practical limitations, including the use of a frozen neural forward-kinematics model, wrist-frame motion preservation that can be affected by fingertip misalignment, and bidirectional Chamfer matching over global reachable spaces. These factors can make the learned mapping sensitive to training samples, calibration errors, and redundant robot-hand motion regions.

*2) **Data Collection for Dexterous Manipulation**:* High-quality demonstrations are essential for learning dexterous manipulation skills, but remain costly to collect for complex hand behaviors. Prior work leverages specialized hardware [14, 42, 52], gloves [47], or in-the-wild human data [8, 15, 16], yet such data may exhibit visual gaps from robot executions and kinematic gaps across different dexterous hands. KineDex [56] collects data directly on dexterous hands through kinesthetic teaching, but human-guided collection can introduce occlusions and becomes difficult when the target hand differs substantially from the human hand in scale. Teleoperation therefore remains a primary paradigm for capturing expert behavior. Some methods simplify control with predefined motion patterns, grasp primitives, or low-dimensional action spaces [24, 25], trading expressiveness for lower operator burden. Recent systems instead use direct hand-motion interfaces and kinematic retargeting to capture richer contact-rich behaviors [7, 10, 18, 20, 32, 33, 43, 54, 58]. In these systems, retargeting directly affects demonstration quality: an unintuitive or poorly calibrated mapping can make teleoperation difficult and produce unnatural robot motions, limiting the usefulness of collected data for downstream imitation learning.