DexEXO: A Wearability-First Dexterous Exoskeleton for Operator-Agnostic Demonstration and Learning

Alvin Zhu[1,2*], Mingzhang Zhu[3*], Beom Jun Kim[3], Jose Victor S. H. Ramos[3], Yike Shi[1,2], Yufeng Wu[3], Raayan Dhar[1], Fuyi Yang[1], Ruochen Hou[3], Hanzhang Fang[1,2], Quanyou Wang[3], Yuchen Cui[1†], and Dennis W. Hong[3†]
dexexo-research.github.io

***Abstract*— Scaling dexterous robot learning is constrained by the difficulty of collecting high-quality demonstrations across diverse operators. Existing wearable interfaces often trade comfort and cross-user adaptability for kinematic fidelity, while embodiment mismatch between demonstration and deployment requires visual post-processing before policy training. We present DexEXO, a wearability-first hand exoskeleton that aligns visual appearance, contact geometry, and kinematics of an existing robotic hand at the hardware level. DexEXO features a pose-tolerant thumb mechanism and a slider-based finger interface analytically modeled to support hand lengths from 140 mm to 217 mm, reducing operator-specific fitting and enabling scalable cross-operator data collection. A passive hand visually matches the deployed robot, allowing direct policy training from raw wrist-mounted RGB observations. User studies demonstrate improved comfort and usability compared to prior wearable systems. Using visually aligned observations alone, we train diffusion policies that achieve competitive performance while substantially simplifying the end-to-end pipeline. These results show that prioritizing wearability and hardware-level embodiment alignment reduces both human and algorithmic bottlenecks without sacrificing task performance.**

I. Introduction

Scaling dexterous robot learning remains fundamentally limited by the difficulty of collecting high-quality demonstrations for multi-finger hands [1–3]. Unlike parallel-jaw grippers, dexterous manipulation requires capturing high-dimensional finger motion, fine hand-object contact, and closed-loop correction during task execution. Existing data collection interfaces often trade off wearability, cross-user adaptability, and motion fidelity, making large-scale demonstration collection difficult [4–6].

Wearable exoskeleton interfaces offer a promising solution for dexterous data collection by mechanically coupling human motion to the robot and reducing retargeting ambiguity [7–9]. Recent systems such as DexUMI enable scalable in-the-wild data collection through lightweight wearables and vision-based reconstruction, but still require segmentation and inpainting to address visual embodiment mismatch between demonstrations and deployment [7]. Other mechanically coupled systems such as DexOP improve hardware-level embodiment alignment through robot-specific co-design, but their tightly constrained linkage structures can limit ergonomic adaptability across users and robot morphologies [8]. More broadly, prior wearable interfaces often sacrifice comfort, cross-user fit, or natural thumb motion due to rigid alignment requirements [10, 11].

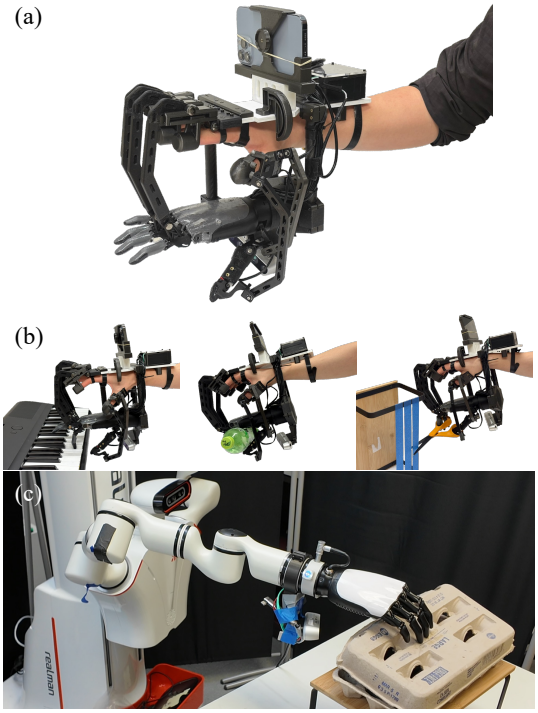

Fig. 1: **System overview of DexEXO.** (a) Full device worn on a user's hand. (b) Demonstrations of piano playing, full-hand grasping, and scissors cutting. (c) Policy deployment on the robot.

To address these limitations, we present DexEXO, a wearability-first hand exoskeleton for scalable dexterous demonstration and learning. DexEXO is designed around two principles: (1) cross-user wearability for sustained and operator-agnostic data collection, and (2) hardware-level embodiment alignment for direct end-to-end policy learning. Our system combines a slider-based finger interface for anthropometric adaptability, a pose-tolerant thumb mechanism that preserves natural thumb motion, and a passive hand that visually matches the deployed robot. This alignment allows policies to be trained directly from raw wrist-mounted RGB observations without visual post-processing [8, 9, 12].

* denotes equal contribution. † denotes equal advising.
[1]Department of Computer Science, [2]Department of Electrical and Computer Engineering, [3]Department of Mechanical and Aerospace Engineering, UCLA, Los Angeles, CA, USA.

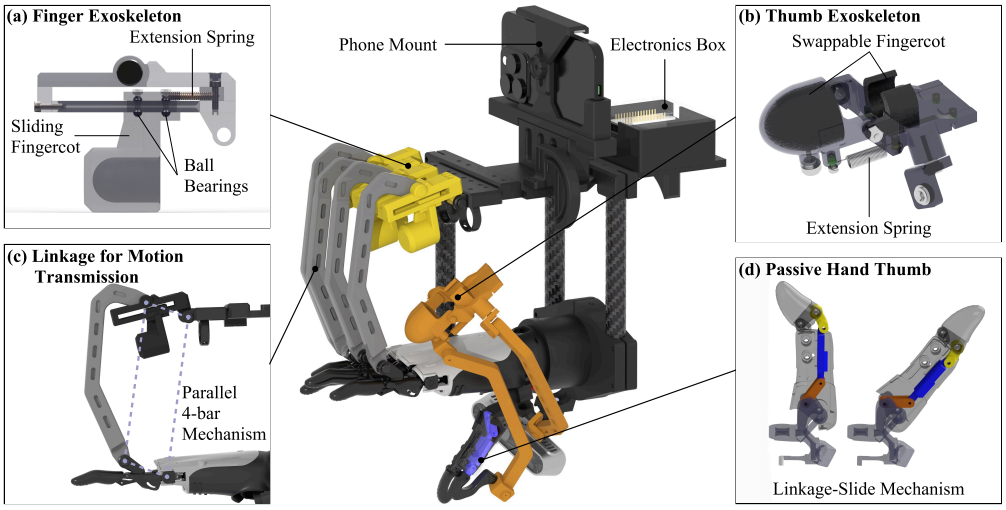

Fig. 2: Mechanical overview of DexEXO. DexEXO integrates a linkage-driven wearable exoskeleton, a passive data-capture hand, and an onboard sensing/power module. Insets highlight key subsystems: (a) passive finger slider for cross-user fit, (b) pose-tolerant thumb coupling interface, (c) parallel four-bar finger linkage for motion transmission, and (d) passive hand thumb that reproduces the intended thumb DOF

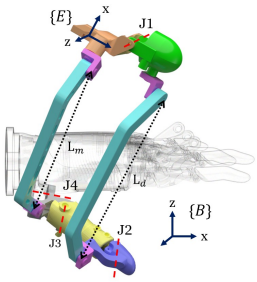

Fig. 3: Kinematic schematic of the exoskeleton thumb. Swivel joints (purple) enable self-alignment between frames E and B while preserving fixed linkage lengths L_d and L_m.

We validate DexEXO through user studies and policy learning experiments, showing improved comfort and usability over prior wearable systems while achieving competitive dexterous manipulation performance using only visually aligned RGB observations. In summary, our contributions are:

- A wearability-first dexterous exoskeleton with analytically validated cross-user compatibility.
- A pose-tolerant thumb mechanism that preserves natural thumb motion while maintaining structured robot correspondence.
- An embodiment-aligned data collection and policy learning pipeline that enables direct training from raw wrist-mounted RGB observations.

II. Hardware Design

A. Hardware Overview

DexEXO consists of (i) a linkage-driven wearable exoskeleton, (ii) a passive demonstration hand, and (iii) an onboard sensing and power module for untethered use. The passive hand follows the 6-DoF ROH-AP001 geometry [13], with a 2-DoF thumb and single-DoF finger flexion. As shown in Fig. 2, motion is transmitted via two coupling strategies: parallel linkages for the four fingers, enabling consistent flexion mapping across users, and a multi-DoF thumb coupling that allows relative translation and rotation between the exoskeleton and palm while preserving key thumb motions. This design improves comfort and adaptability across hand sizes. A dorsal-mounted electronics module enables onboard power and logging, while a wrist-mounted iPhone provides pose capture for in-the-wild data collection.

B. Passive Hand

To ensure high-fidelity proprioception, DexEXO integrates six joint encoders within a rigid structure that prevents sensor drift during dynamic manipulation. In parallel, we performed kinematic identification using a URDF to design a custom linkage-slide mechanism that matches the actual hand kinematics, ensuring consistent and physically accurate trajectory mapping.

C. Slider-based Finger Interface

DexEXO enables cross-operator use without rigid joint alignment or per-user calibration by incorporating passive tolerance into the transmission design. A passive spring-loaded linear slider accommodates variations in finger length. This decouples finger length from joint alignment, improving fit robustness while preserving effective motion transmission.

D. Pose-tolerant Thumb Mechanism

The thumb poses challenges for wearable interfaces due to anatomical variability, where rigid alignment can cause discomfort and restrict motion. DexEXO addresses this with a pose-tolerant coupling that preserves wearability while supporting IP flexion/extension and TM ab/ad.

1) Mechanism overview: As shown in Fig. 3, the exoskeleton thumb includes an instrumented IP joint J_1 (θ_1), while the passive thumb has IP (J_2, θ_2) and TM ab/ad (J_4, θ_4) joints, with J_3 coupled to J_2. Two rigid linkages connect the exoskeleton to the passive thumb. Instead of enforcing

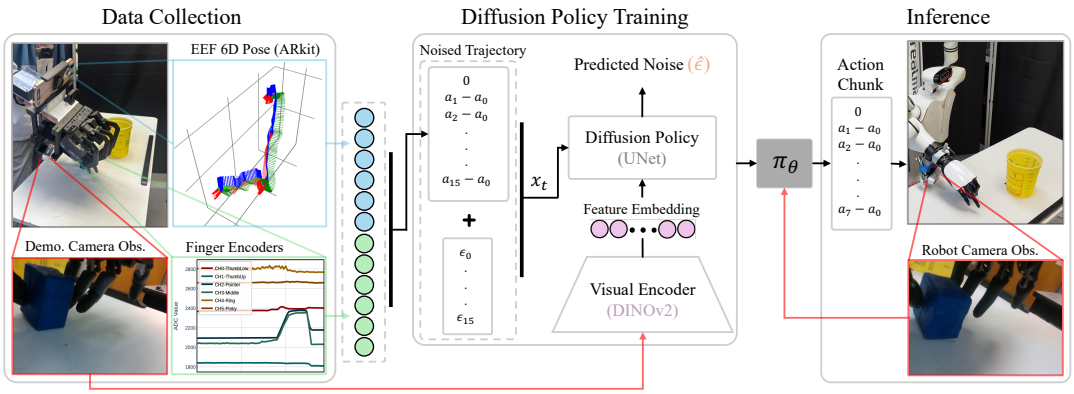

Fig. 4: An overview of the full demonstration data modalities, policy training, and inference with visual-aligned observations.

alignment, the design imposes distance constraints, allowing relative translation and rotation.

2) Simplified kinematic model: Let $\{B\}$ and $\{E\}$ denote the palm and exoskeleton frames with relative pose ${}^{B}T_{E} = ({}^{B}R_{E}, {}^{B}p_{E})$. The passive thumb configuration is $q_p = [\theta_2 \;\; \theta_4]^\top$, with $\theta_3 = f(\theta_2)$. Let ${}^{B}r_i(q_p)$ $(i \in \{d, m\})$ be the passive attachment points and ${}^{E}\bar{r}_i$ their fixed exoskeleton counterparts. Their positions satisfy

$$ {}^{B}r_i^E = {}^{B}R_E \, {}^{E}\bar{r}_i + {}^{B}p_E, \quad \|{}^{B}r_i^E - {}^{B}r_i(q_p)\| = L_i. \quad (1) $$

With a 6-DoF exoskeleton pose and two scalar constraints, the system admits a 4-DoF self-motion manifold for fixed q_p, corresponding to the observed "wiggle space," where the exoskeleton can move relative to the palm without altering thumb posture.

III. Data Collection and Policy Training

1) Data Collection: DexEXO records synchronized finger joint positions, end-effector pose, and wrist-mounted RGB observations during demonstrations. Six onboard analog encoders measure finger motion at 1 kHz and are mapped to robot joint commands through calibrated piecewise linear interpolation. End-effector pose is captured at 60 Hz using the iPhone-based TeleDex system [14]. A wrist-mounted Intel RealSense camera records RGB images at 640×480 resolution and 30 Hz. All modalities are temporally aligned using image timestamps as the reference.

A. Policy Architecture and Training Setup

Our aligned embodiment enables a direct pipeline from demonstration to policy training. The passive hand matches the geometry of the deployed robot, eliminating the visual embodiment gap and the need for segmentation, masking, or inpainting [7]. Policies are trained directly from raw wrist RGB with synchronized end-effector and finger signals.

a) Observations: Each training sample includes an RGB frame from the wrist-mounted camera and (optionally) a low-dimensional hand state. The RGB image is resized to 240×240, randomly cropped to 224×224, and augmented with color jitter during training. Visual features are extracted using a DINOv2 ViT-S/14 encoder [15], and the resulting embedding is used as the primary conditioning signal for the policy. When used, the hand state is the 6D absolute finger pose.

b) Actions: The policy outputs a 12D action consisting of a 6-DoF end-effector command and 6 finger commands. We train a diffusion policy [16] to predict an action horizon of 16 steps and execute the first 8 actions in a receding-horizon manner at inference time. Actions are expressed relative to the initial state of the horizon: the k-th predicted action corresponds to $T_k - T_0$. This representation supports reactive closed-loop control while retaining multi-step prediction capability.

All policies in this work use the same diffusion policy backbone and vision encoder; differences between action parameterizations and conditioning signals are evaluated in Sec. IV.

IV. Results

A. Demonstration User Studies

A user study was conducted to compare experience and performance across different demonstration methods. We recruited 14 university students (7 female, 7 male; aged 18–27) with hand sizes ranging from 165 mm to 195 mm. Participants engaged with 3 demonstration devices: DexEXO, DexUMI, and vision-based teleoperation [14]. Teleoperation served as a baseline, as it is the most commonly used approach in prior work. For each device, participants were asked to perform the following tasks:

Scissors cutting: Pick up scissors and cut a strip of tape.
Page flipping: Use the fingertip to flip a notebook page.
Cup stacking: Stack 3 cups facing up.
Piano playing: Play 16 notes on a piano using 4 fingers.

For each task, we recorded success rate and completion time as quantitative metrics. All tasks were performed under a 120-second time limit, and completion time was capped at 120 seconds if unfinished.

The quantitative results from the user study are shown in Table I. DexEXO was the only device capable of performing the scissors cutting task. DexUMI failed as its added exoskeleton geometry, absent in the original robot hand, prevented the fingers from fitting within the handles. Teleoperation failed at the same task due to a lack of precision,

TABLE I: Summary of quantitative results in user study (mean ± SEM). Bold indicates best performance per metric.

	Scissors Cutting		Page Flipping		Cup Stacking		Piano Playing	
Method	Success Rate	Time (s)†	Success Rate	Time (s)†	Success Rate	Time (s)†	Success Rate	Time (s)†
DexEXO	$\mathbf{0.79 \pm 0.10}$	$\mathbf{11.7 \pm 1.4}$	$\mathbf{0.88 \pm 0.03}$	5.4 ± 0.6	$\mathbf{0.82 \pm 0.07}$	12.0 ± 1.1	$\mathbf{0.96 \pm 0.02}$	$\mathbf{21.6 \pm 1.8}$
DexUMI	0.00 ± 0.00	—	0.86 ± 0.04	$\mathbf{4.7 \pm 0.7}$	0.80 ± 0.07	$\mathbf{8.9 \pm 1.0}$	0.62 ± 0.13	25.9 ± 2.5
Teleoperation	0.00 ± 0.00	—	0.51 ± 0.06	18.0 ± 2.1	0.33 ± 0.09	68.6 ± 13.1	0.60 ± 0.09	97.4 ± 7.8

† Completion time was defined as the average time from picking up scissors to finishing the cut, time for 5 page flips, average time for a successful 3-cup stack, and time to play 16 piano notes, respectively.

TABLE II: Policy success rates with and without finger-state conditioning.

Method	**Tasks**		
Finger Condition	Block	Carton	Bottle
No	**0.90**	0.90	**0.85**
Yes	0.85	**0.95**	0.80

responsiveness, and force feedback. While DexUMI outperforms DexEXO in page flipping and cup stacking in terms of completion time (13.0% and 25.8% faster, respectively), DexEXO achieves higher success rate in both tasks. DexEXO outperforms DexUMI significantly in the piano task, with 54.5% higher success and 16.6% faster completion time. It is also worth noting that teleoperation performs worse overall compared to both exoskeleton methods.

B. Policy Evaluation

We evaluate whether aligned visual and contact geometry embodiment enables effective end-to-end policy learning without visual post-processing, and whether explicit hand-state conditioning remains necessary under this setting.

a) Experimental Setup: Policies are trained on demonstrations collected using DexEXO as described in Sec. III. We evaluate three representative manipulation tasks:

Block: Grasp a block and place it into a cup, testing precision and fingertip alignment.

Carton: Open an egg carton lid using coordinated multi-finger interaction and distributed contact.

Bottle: Grasp a bottle and lift it above 50 mm, highlighting whole-hand grasping with a palm-assisted enclosure.

For each trained policy, we conduct 20 evaluation trials with randomized object initial poses. Success is defined as complete placement into the cup (Block), opening the lid beyond 30° (Carton), and lifting the bottle by at least 50 mm and holding it stably for 2 s (Bottle).

b) Ablation Study: We ablate the use of explicit hand-state conditioning by comparing policies trained with and without absolute finger pose inputs. All policies share the same diffusion architecture, visual encoder, and training configuration; only the observation inputs differ.

c) Quantitative Results: Policy success rates are reported in Table II. For the **Block** task, success primarily depends on precise fingertip alignment and stable grasp closure during placement. Under wrist-aligned embodiment, finger configuration remains visually observable from RGB input, allowing the policy to infer grasp posture directly from image features.

For the **Carton** task, which requires coordinated multi-finger interaction and distributed contact, visual cues such as lid deformation and relative hand pose provide sufficient information for closed-loop adjustment, resulting in similar performance with and without explicit finger-state inputs.

For the **Bottle** task, which emphasizes whole-hand grasping and stable lifting, performance remains similar with and without explicit finger-state conditioning. Because the task primarily relies on gross hand-object alignment and whole-hand grasping, the policy can recover sufficient hand configuration even under major occlusion.

These results suggest that when hardware handles geometric and visual alignment, raw RGB observations provide sufficient state information, making explicit finger-state conditioning redundant.

d) Comparison to Prior Wearable-Based Pipelines: We evaluate block placement and carton opening to compare with DexUMI, which reports success rates of 1.00 (Cube) and 0.85 (Carton) using relative actions, image and tactile conditioning, and segmentation with inpainting to address visual embodiment mismatch.

With our aligned pseudo-hand embodiment, we achieve 0.90 success on both tasks without segmentation, masking, inpainting, or tactile conditioning. DexUMI's raw image baseline without these components performs substantially worse, highlighting their reliance on visual post-processing. In contrast, our hardware-level alignment enables strong performance directly from raw RGB observations with a simpler end-to-end pipeline.

V. Conclusion

We presented **DexEXO**, a wearability-first dexterous exoskeleton for scalable, cross-operator demonstration collection with structured kinematic correspondence to a target robotic hand. Through analytically modeled finger interfaces and a pose-tolerant thumb mechanism, DexEXO accommodates anthropometric variation without rigid alignment or per-user calibration. Experiments demonstrate consistent transmission and self-alignment, while user studies show improved usability over prior systems. Policy results demonstrate that hardware-level embodiment alignment enables effective end-to-end learning from raw wrist-mounted RGB. Overall, prioritizing wearability and geometric alignment reduces both human and algorithmic bottlenecks in dexterous learning without sacrificing performance.

REFERENCES

[1] A. Rajeswaran et al., "Learning complex dexterous manipulation with deep reinforcement learning and demonstrations," in *Proceedings of Robotics: Science and Systems (RSS)*, 2018.

[2] M. Andrychowicz et al., "Learning dexterous in-hand manipulation," *The International Journal of Robotics Research*, vol. 39, no. 1, pp. 3–20, 2020.

[3] Y. Jiang, A. Stone, Z. Tan, et al., "Vima: General robot manipulation with multimodal prompts," in *International Conference on Machine Learning (ICML)*, 2023.

[4] C. Chi et al., "Universal manipulation interface: In-the-wild robot teaching without in-the-wild robots," in *Robotics: Science and Systems (RSS)*, 2024.

[5] Y. Liu, Y. Yang, Y. Wang, et al., "Realdex: Towards human-like grasping for robotic dexterous hand," in *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI)*, 2024.

[6] E. Welte et al., "Interactive imitation learning for dexterous robotic manipulation: Challenges and perspectives," *Frontiers in Robotics and AI*, 2025.

[7] M. Xu et al., *Dexumi: Using human hand as the universal manipulation interface for dexterous manipulation*, 2025. arXiv: 2505.21864 [cs.RO].

[8] H.-S. Fang et al., *Dexop: A device for robotic transfer of dexterous human manipulation*, 2025. arXiv: 2509.04441 [cs.RO].

[9] J. Du et al., *Mile: A mechanically isomorphic exoskeleton data collection system with fingertip visuotactile sensing for dexterous manipulation*, 2025. arXiv: 2512.00324 [cs.RO].

[10] R. Ge, Y. Liu, Z. Yan, Q. Cheng, S. Qiu, and D. Ming, "Design of a self-aligning four-finger exoskeleton for finger abduction/adduction and flexion/extension motion," in *2023 International Conference on Rehabilitation Robotics (ICORR)*, 2023.

[11] C. Brogi et al., "An original hybrid-architecture finger mechanism for wearable hand exoskeletons," *Mechatronics*, vol. 98, p. 103 117, 2024.

[12] Z. Si, K. L. Zhang, Z. Temel, and O. Kroemer, *Tilde: Teleoperation for dexterous in-hand manipulation learning with a deltahand*, 2024. arXiv: 2405.18804 [cs.RO].

[13] OYMotion Technologies Co., Ltd., *Roh-ap001 dexterous robotic hand*, https://www.oymotion.com/en/product62, Five-finger robotic hand with human-like proportions and independent finger motion, 2024.

[14] O. Rayyan, M. Gilles, and Y. Cui, *Teledex: Accessible dexterous teleoperation*, GitHub repository, 2026.

[15] M. Oquab et al., "Dinov2: Learning robust visual features without supervision," *arXiv preprint arXiv:2304.07193*, 2023.

[16] C. Chi et al., "Diffusion policy: Visuomotor policy learning via action diffusion," in *Proceedings of Robotics: Science and Systems (RSS)*, 2023.