DEXOP: Hardware for Collecting Contact-Rich and Dexterous Robotic Manipulation Data In-The-Wild

Hao-Shu Fang, Arthur Hu*, Branden Romero*, Edward Adelson and Pulkit Agrawal
Massachusetts Institute of Technology
contact: fhs@mit.edu

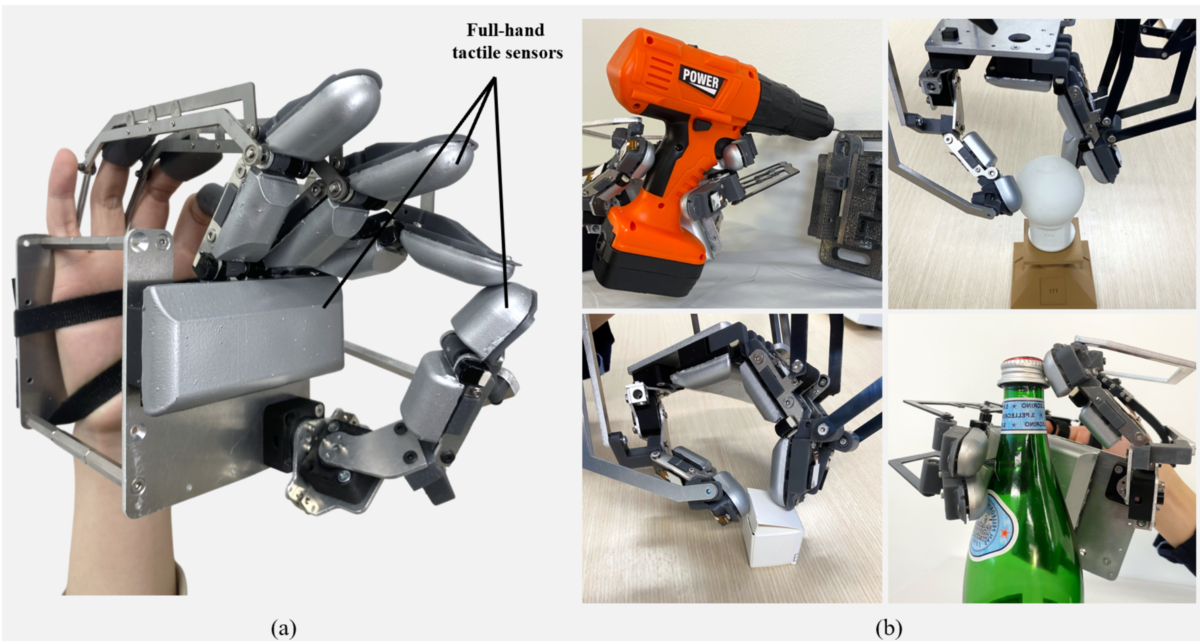

Fig. 1: (a) Picture of DEXOP. (b) DEXOP is used to collect demonstrations in-the-wild on various dexterous tasks: drilling, lamp installation, box packaging, and bottle opening.

***Abstract*— We present DEXOP, a novel passive hand exoskeleton system designed to enable scalable and intuitive teaching of dexterous robotic manipulation in real-world environments. Unlike traditional teleoperation methods that lack force feedback and are expensive to scale, DEXOP offers kinematic mirroring and force transparency, allowing human users to control robot hands naturally. By transmitting tactile and joint data during task execution, DEXOP supports efficient collection of high-quality demonstrations. We evaluate DEXOP across four dexterous tasks and show it significantly improves task throughput and success rate compared to standard teleoperation, highlighting its potential as a powerful tool for advancing robot learning.**

I. INTRODUCTION

Dexterous manipulation is fundamental to advancing robot autonomy, yet remains difficult due to challenges in control, sensing, and generalization. Existing approaches such as reinforcement learning, teleoperation, and video-based imitation struggle with issues like reward shaping, human-robot embodiment mismatch, or lack of tactile feedback. Moreover, collecting large-scale high-quality data for dexterous tasks is still an unsolved bottleneck.

To tackle this, we propose DEXOP: a passive hand exoskeleton system that allows humans to directly drive a passive robotic hand with natural motion and haptic feedback, *without needing a real robot*. The system achieves scalability using a low-cost, portable setup and enables high-fidelity in-the-wild demonstrations with tactile feedback.

DEXOP enables intuitive, in-the-wild demonstration collection with high-fidelity contact information. Our experiments demonstrate DEXOP's ability to facilitate complex real-world tasks while outperforming existing teleoperation baselines. This system provides a path toward scalable data-driven dexterous manipulation.

II. SYSTEM OVERVIEW

A. Design Principles

DEXOP prioritizes intuitive human-robot interaction through:

- **Kinematic mirroring:** Robot hand mimics human finger postures.
- **Force transparency:** Forces from robot-environment contact are relayed to the user.
- **Passive actuation:** The human operator drives the system without active motors.

B. Hardware Architecture

The system comprises a wearable exoskeleton and a passive robotic hand linked through a 4-bar linkage mechanism. Each joint includes angular encoders for joint state tracking. The mechanical design ensures low inertia and high fidelity

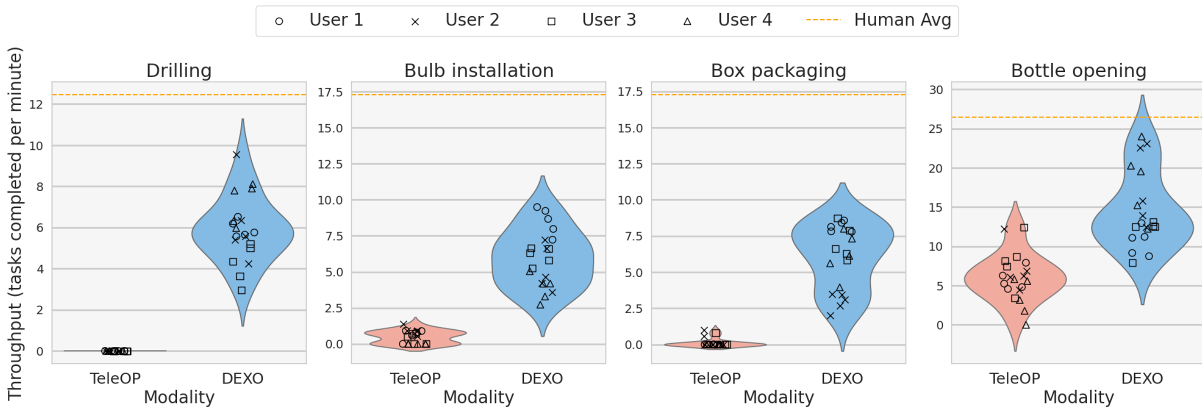

Fig. 2: Comparison of task throughput across different data collection modalities.

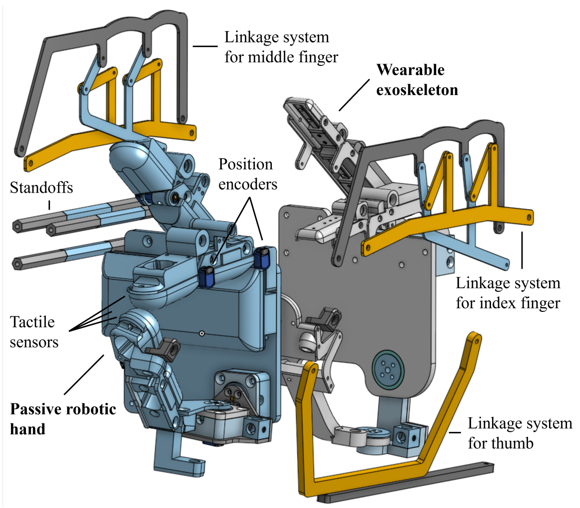

Fig. 3: Exploded view of the DEXOP system.

motion transfer. Fig. 3 gives an overview of our hardware design.

C. Sensor Integration

A full-hand vision-based tactile sensor (GelSim(ple)) [3] is mounted on the robot fingers, enabling capture of rich contact data. All joint and tactile signals are streamed to a host computer via custom electronics with RS-485 communication and Arducam multi-camera boards.

D. EyeSight Hand Compatibility

DEXOP shares the same mechanical design and sensors as the EyeSight Hand [3]—including joint kinematics, full-hand GelSim(ple) tactile sensors, and a wrist camera—enabling seamless transfer of collected data and policies to the real robot without domain adaptation.

E. Ease of Use

DEXOP supports untethered in-the-wild data collection, optionally using arm-based exoskeleton AirExo [2] or SLAM-based global tracking [1] to coordinate hand motion with an external arm.

III. Experiments

We evaluate DEXOP on four representative manipulation tasks: drilling, bulb installation, box packaging, and bottle opening. In a user study with four participants, we compare three modalities:

1) **DEXOP system** with haptic feedback.
2) **Visual teleoperation** using trakSTAR hand tracking.
3) **Direct human performance** as upper-bound.

Each task was repeated five times per user and modality. Task throughput (completions/minute) was measured during a 3-minute capped trial per task. A trial was considered a failure if the participant was unable to complete the task in time. Task throughput (completions/minute) was measured. Results show DEXOP achieves (see Fig. 2):

- **Only DEXOP** completed drilling within time cap.
- **6x faster** bulb installation than teleop.
- **7x better** performance in box packaging.
- **2.4x speedup** in bottle opening.

These results indicate DEXOP enables more natural and effective control for fine manipulation. Participants also provided qualitative feedback, highlighting the DEXOP system's ease of use and the benefits of haptic feedback. Users noted that force transparency allowed them to better sense contact conditions and adjust grip strategies in real time—an advantage that was lacking in visual-only teleoperation systems.

IV. Conclusion

DEXOP introduces a scalable, intuitive platform for teaching dexterous manipulation. By combining passive exoskeleton design, tactile sensing, and kinematic mirroring, it bridges key limitations in teleoperation and data collection. Our user study validates its effectiveness across diverse real-world tasks. Future work includes scaling data collection and integrating DEXOP into robot learning pipelines for foundation model training.

References

[1] Cheng Chi et al. "Universal manipulation interface: In-the-wild robot teaching without in-the-wild robots". In: *arXiv preprint arXiv:2402.10329* (2024).

[2] Hongjie Fang et al. “AirExo: Low-Cost Exoskeletons for Learning Whole-Arm Manipulation in the Wild”. In: *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2024, pp. 15031–15038.

[3] Branden Romero et al. “EyeSight Hand: Design of a Fully-Actuated Dexterous Robot Hand with Integrated Vision-Based Tactile Sensors and Compliant Actuation”. In: *arXiv preprint arXiv:2408.06265* (2024).

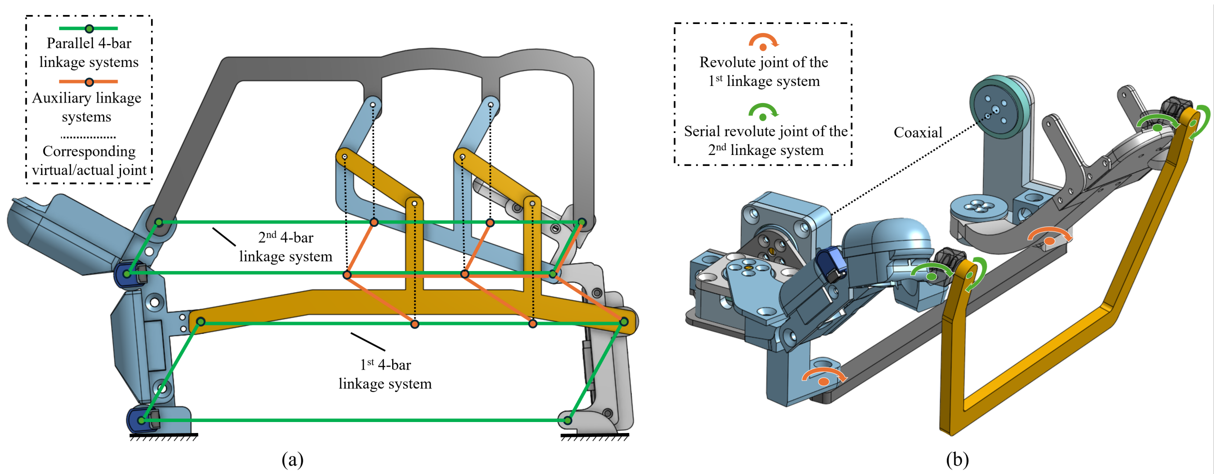

Fig. S1: (a) Annotated view of the 4-bar linkages coupling the index fingers of the robot and exoskeleton hands together (b) Annotated view of the rotary linkage system coupling the thumbs of the robot and exoskeleton hands together

APPENDIX

A. Mechanical Linkage Design

The exoskeleton-to-robot mapping is enabled by a series of parallel 4-bar linkages. For each finger, the proximal and middle phalanges are mirrored using two stacked 4-bar systems. These are constrained to prevent contra-parallelogram configurations that could cause control instability. To extend the workspace beyond 180 degrees, auxiliary coupler linkages maintain mechanical consistency and avoid collisions between joints. The thumb is actuated via a more complex spatial linkage that includes two perpendicular axes for abduction and flexion, enabling three degrees of freedom using layered linkages. Fig. S1 gives an overview of our linkage system design.

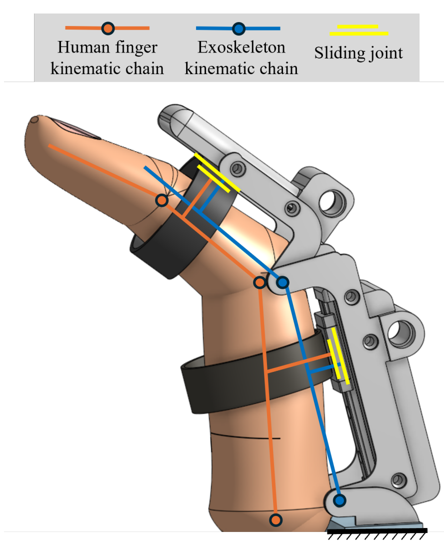

Fig. S2: The kinematics of the wearable exoskeleton match the kinematics of the robotic finger. The sliding joints serve as compensatory mechanisms, ensuring that despite the differences in size, shape, and range of motion between the human hand and the robotic finger, the exoskeleton can still perform synchronized and natural movements.

B. Joint Alignment and Human Fit

To align the exoskeleton's joints with human anatomy, the joints are mounted beside the user's fingers and connected using 1mm spring steel. A 3D-printed finger backing wraps around the dorsal side of the finger, allowing firm yet comfortable coupling. Linear sliders mounted on the backing provide compensation for finger length variation and skin deformation during flexion. This design ensures joint alignment without restricting natural motion or causing discomfort. An illustration is given in Figure S2.

C. Electronics and Sensing Pipeline

Each revolute joint on the passive robotic hand is equipped with an iC-MH16 12-bit absolute angular encoder, offering angular resolution of 0.0015 radians. These encoders communicate via RS-485 using a custom PCB that aggregates signals from all fingers and transmits them to a host computer over USB. The tactile sensors use IMX219 fisheye camera modules embedded inside the fingertips, illuminated with diffused LEDs. Tactile data is captured at 30 FPS using Arducam quad-sync kits and fused with joint data for real-time processing and storage.

D. Tactile Sensor and Vision System

The GelSim(ple) sensor design used in DEXOP is mechanically and optically identical to that of the EyeSight Hand. Each sensor consists of a soft elastomer with embedded reflective markers, a fisheye camera, and uniform internal lighting. The contact deformation is imaged and processed to infer contact geometry and shear force. Additionally, the wrist of the hand mounts a wide-angle camera, enabling egocentric video collection during manipulation tasks.

E. User Study Protocol and Task Details

We conducted a within-subject user study involving four participants. Each participant performed four dexterous tasks (drilling, bulb installation, box packaging, bottle opening) illustrated in Figure S3 using three control modalities:

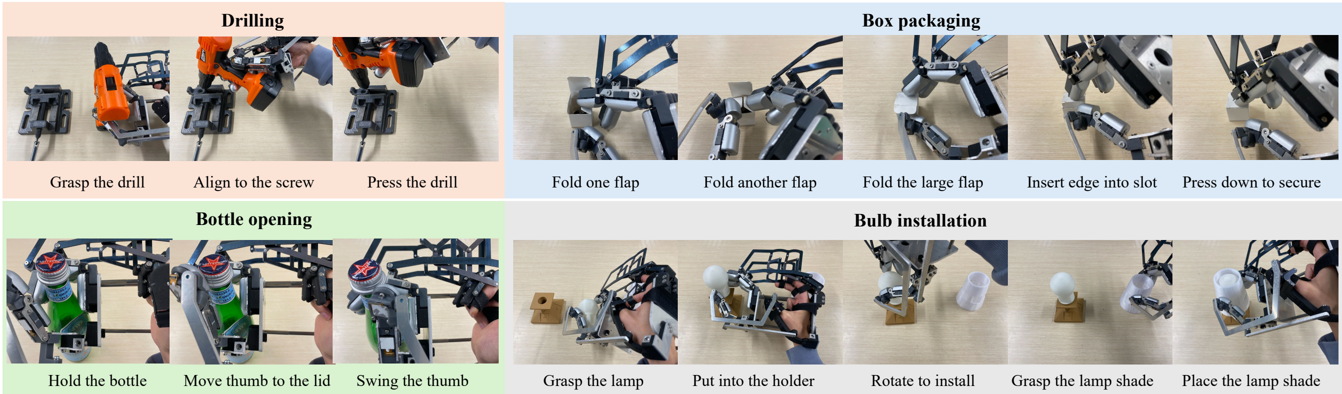

Fig. S3: Illustration of evaluation tasks. **Drilling**: the user must pick up a drill standing upright on a table, the user then inserts the drill bit into an M2 screw head and tightens it by actuating the drill. **Bottle opening**: with the bottle placed within the workspace of the hand, the user grasps the bottle and then uses the thumb to unscrew the cap. **Box Packaging:** the user approaches the an open box, and folds the side flaps before closing the top flap by folding the the securing flap into the box. **Bulb installation**: the task is composed of three parts, a lamp base, a light bulb, and a light shade. The user picks and screws the light bulb into the lamp base before placing the light shade over the entire assembly.

DEXOP, visual teleoperation (trakSTAR + EyeSight Hand), and direct hand use. Each task-modality pair was attempted five times, with a cap of 3 minutes per trial. Success was defined as completing the task within this cap. Throughput was computed as the number of successful completions per minute.

a) Drilling Task: Participants picked up a handheld drill, aligned it with a vertical screw, and actuated it. Teleoperation consistently failed due to misalignment and poor feedback. DEXOP enabled six successful completions per minute on average.

b) Bulb Installation Task: Participants screwed a bulb into a base and placed a lampshade. Precision and fine rotation were critical. DEXOP reduced completion time by a factor of 8 compared to teleop.

c) Box Packaging Task: Participants folded three flaps and inserted a securing tab into the slot. This required coordinated multi-finger motion. DEXOP enabled an average throughput of 5 completions per minute.

d) Bottle Opening Task: Participants grasped a bottle with two fingers and unscrewed the cap using the thumb. DEXOP allowed for consistent execution with 2.4x the speed of teleoperation.