HOVER: Generalized Retargeting for Dexterous Manipulation

Jing (Daisy) Dai[1,2,*], Qian Shu (Helen) Wang[1,3,*], Shurui Zhang[1,3,*], Bin Zhao[1,4], Jiahong Zhang[1], Jianbo Yuan[1,2], Yiwen Lu[1,†]

[1]DexRobot Co. Ltd. [2]Shanghai Jiao Tong University [3] University of Cambridge [4]East China University of Science and Technology

*Equal contribution, sorted by alphabetical order †Corresponding author: lyw@dex-robot.com

Problem Statement

Retargeting—mapping human hand trajectories to robot joint commands—is fundamental to dexterous manipulation. It bridges the **embodiment gap** for both offline learning from demonstrations and online teleoperation.

Recent work leaves gaps. **Kinematics-based** methods [1, 2] achieve good pose matching across diverse hands but prioritize mimicry over task completion. **Object-centric** methods [3, 4] ensure task success but train separate models per demonstration, limiting generalization.

Method Type	Anthropomorphism	Task Fidelity	Real Deploy
Kinematics-based	High	Low	Limited
Object-centric	Low	High	Sim Only
Our Method	**High**	**High**	**Yes**

We build on object-centric learning but introduce a **virtual operator** that simulates human teleoperation adaptively. This enables **goal-directed** behaviors that **generalize across tasks** while preserving human-like motion, with direct hardware deployment capability.

The HOVER Framework

HOVER (Human-Operator Virtual-Enhanced Retargeting) bridges offline human demonstrations and online robotic execution through a key insight: human operators naturally close the control loop during teleoperation, but offline demonstrations lack this feedback. We introduce a **virtual operator** policy that simulates this closed-loop control during training.

The virtual operator and retargeting policy are jointly trained using PPO. The retargeting policy sees **only joint angles**—no object or task information—forcing generalizable joint-to-joint mappings. The virtual operator sees rich observations but is penalized for large corrections, absorbing dataset-specific artifacts while preserving generalization.

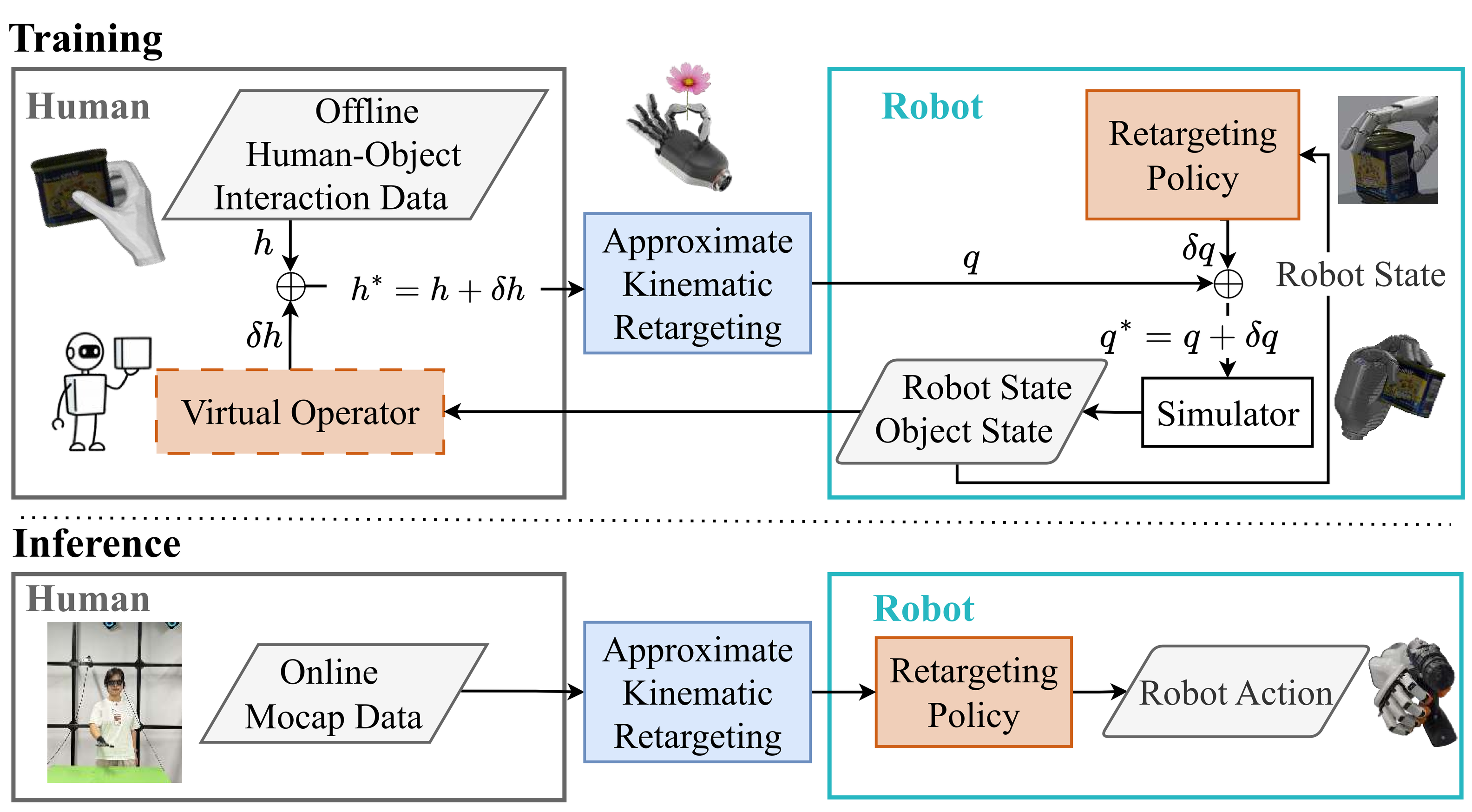

Figure 1. The HOVER framework. Virtual operator adjusts human joints h by δh. Approximate kinematic retargeting maps $h^* = h + \delta h$ to robot targets q. Retargeting policy refines to final commands $q^* = q + \delta q$.

Insights: (i) **Closing the loop**: The virtual operator provides essential closed-loop control for offline data. (ii) **Generalization through minimal interface**: Pure joint-to-joint mapping across diverse datasets ensures robust task generalization.

Contributions

1. **Generalizable Retargeting Policy:** Single policy across diverse tasks achieving both high task fidelity and anthropomorphism—metrics that existing methods trade off.

2. **Virtual Operator Methodology:** Novel training paradigm simulating closed-loop human control for offline demonstrations, maintaining generalization through minimal interface design.

3. **Comprehensive Evaluation:** Improved success rates across diverse manipulation tasks in simulation and successful real hardware deployment with DexHand021.

Watch Demo Videos
Scan for materials

Built by
DexRobot
Leading Innovation in Dexterous Manipulation
www.dex-robot.com

Experimental Setup

Training: We train a unified policy across diverse manipulation datasets: HO-3D, ARCTIC, DexYCB, and DexCanvas (our self-collected human manipulation dataset, to be released).

Offline Evaluation: Quantitative comparison with DexMachina [4] on task completion rates across standard benchmarks.

Online Deployment:

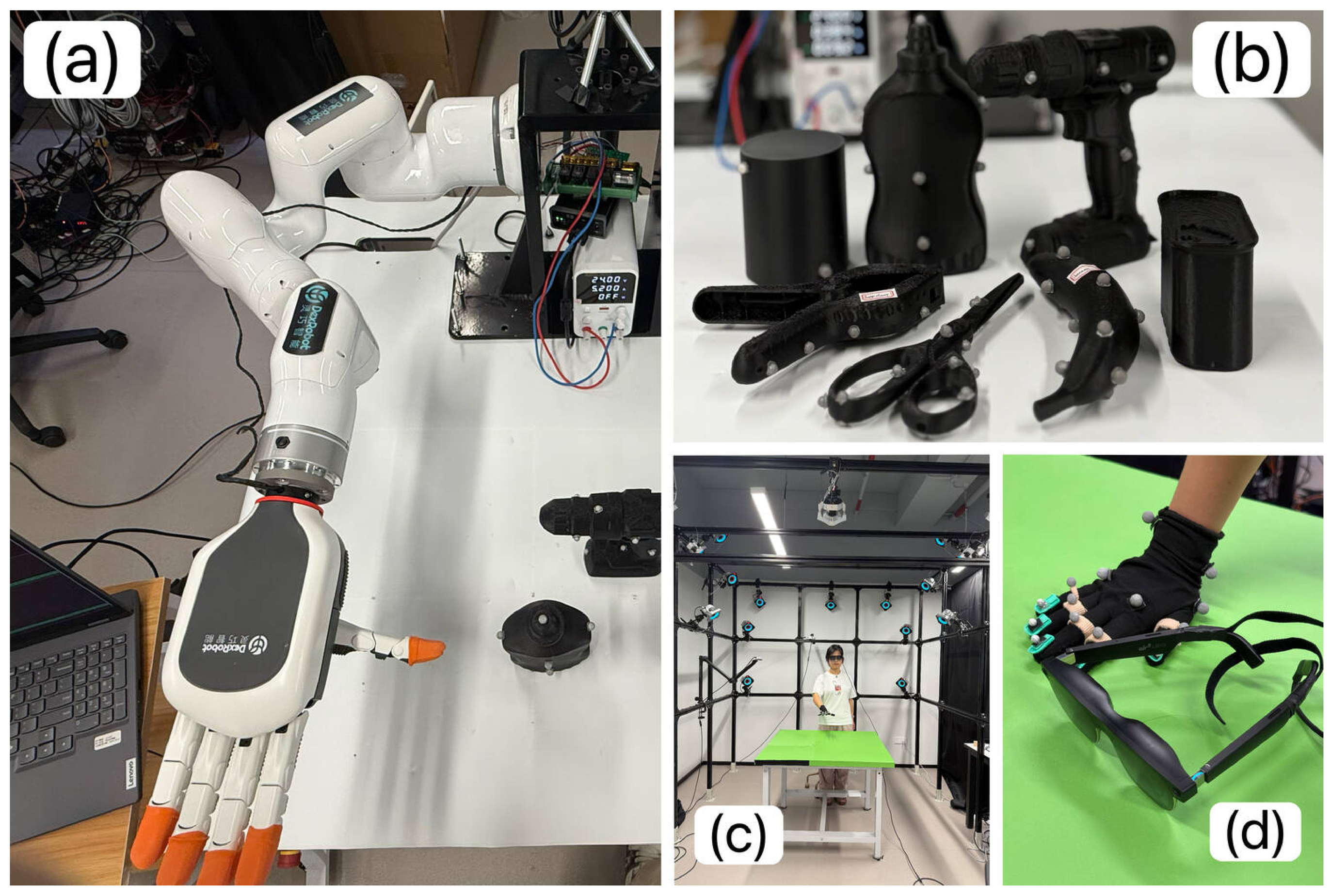

Figure 2: Hardware setup for real-world deployment: (a) DexHand021 end-effector with JAKA Mini arm, (b) 3D-printed YCB objects, (c) ChingMu motion capture system, (d) Xreal Air 2 glasses streaming third-person view from ZED Mini stereo camera.

Results

Offline Performance:

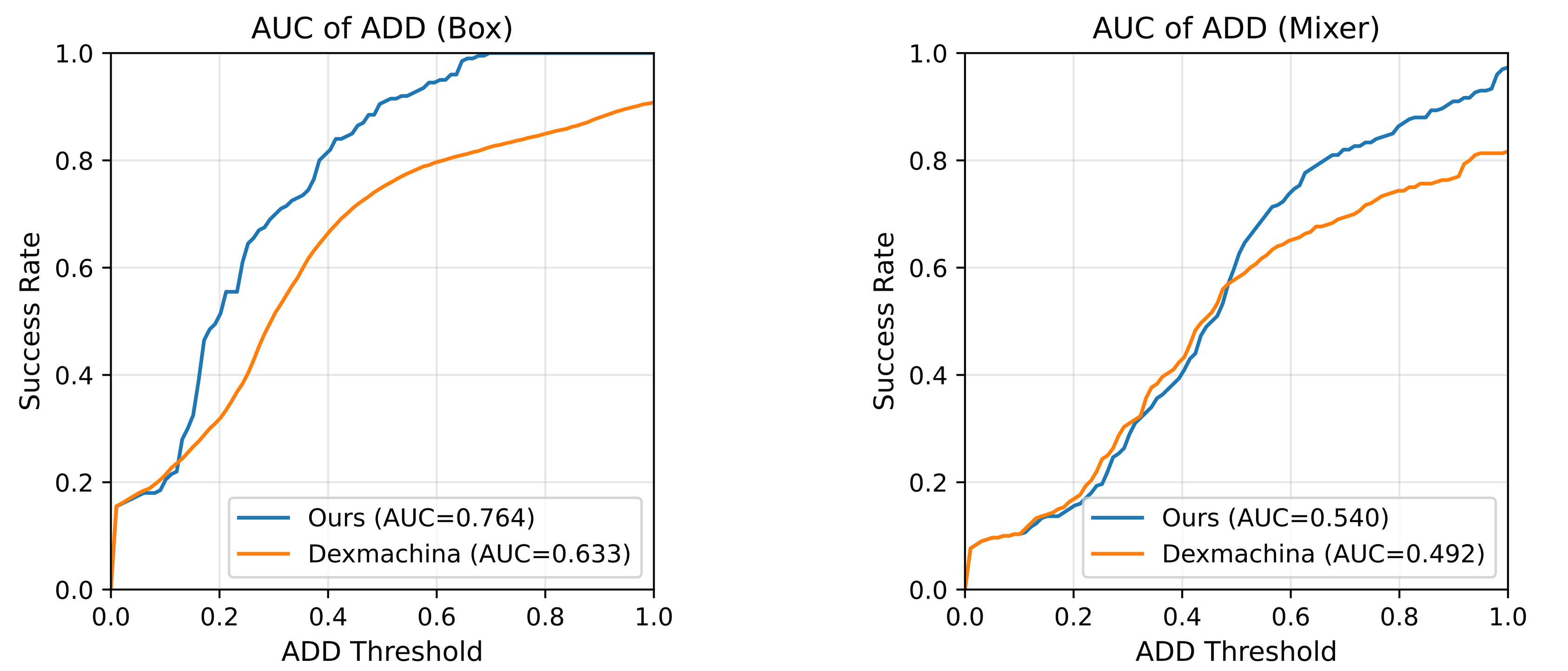

Figure 3. Quantitative evaluation using ADD-AUC metric [4], which measures area under the accuracy-threshold curve. ADD (Average Distance) computes mean 3D distance between predicted and ground-truth object points; lower thresholds (in cm) require more precise manipulation. HOVER consistently outperforms baselines across diverse tasks.

Online Deployment:

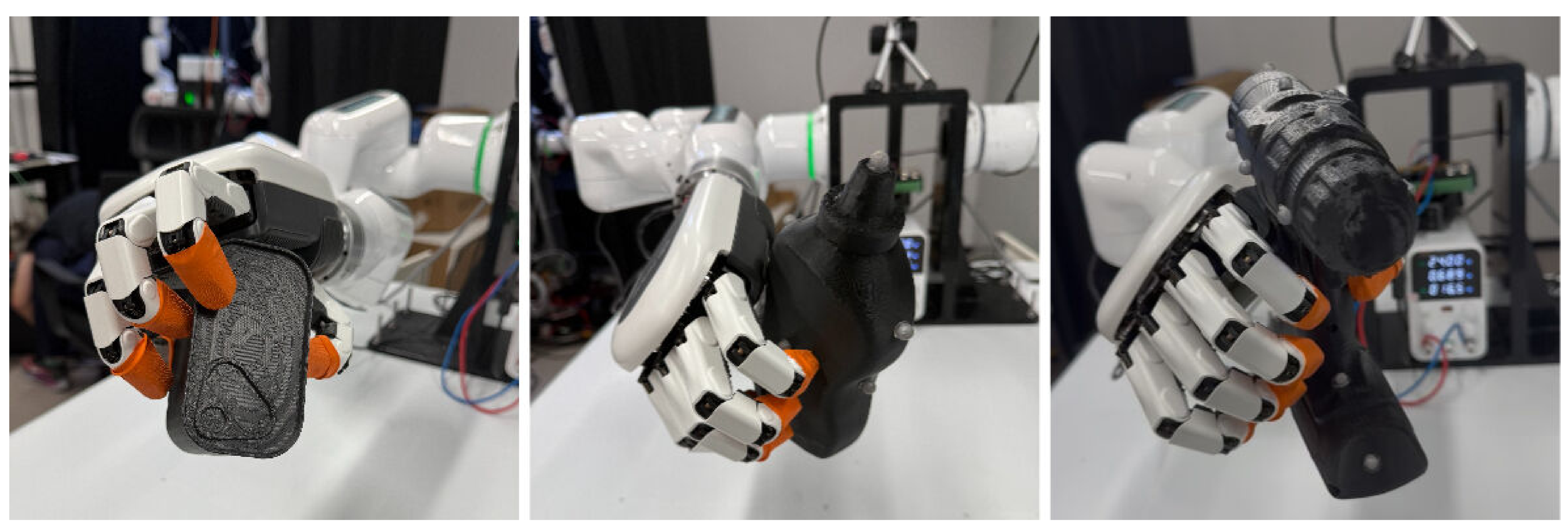

Figure 4: Successful real-world deployment demonstrating three grasp types: (i) power grasp for robust object control, (ii) enveloping grasp for stabilization, (iii) tool-handle grasp for functional manipulation.

Conclusion

HOVER enables unified retargeting across offline learning and online teleoperation through virtual operator training, achieving both task success and anthropomorphic motion.

Future Work: Bimanual manipulation; challenging tasks with intense in-hand reorientations; generating large-scale synthetic robotic datasets from human demonstrations; human studies on teleoperation ease and naturalness.

[1] Yuanpei Chen et al. "Object-Centric Dexterous Manipulation from Human Motion Data". In: *8th Annual Conference on Robot Learning*. 2024

[2] Mandi Zhao et al. "DexMachina: Functional Retargeting for Bimanual Dexterous Manipulation". In: *arXiv preprint arXiv:2505.24853* (2025)

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

[4] Chendong Xin et al. "Analyzing Key Objectives in Human-to-Robot Retargeting for Dexterous Manipulation". In: *arXiv preprint arXiv:2506.09384* (2025)