HERMES: Human-to-Robot Embodied Learning from Multi-Source Motion Data for Mobile Dexterous Manipulation

Leveraging human motion data to impart robots with versatile manipulation skills has emerged as a promising paradigm in robotic manipulation. Nevertheless, translating multi-source human hand motions into feasible robot behaviors remains challenging, particularly for robots equipped with multi-fingered dexterous hands characterized by complex, high-dimensional action spaces. In this paper, we introduce HERMES, a human-to-robot learning framework for mobile bimanual dexterous manipulation. First, HERMES formulates a unified reinforcement learning approach capable of seamlessly transforming heterogeneous human hand motions from multiple sources into physically plausible robotic behaviors. Subsequently, to mitigate the sim2real gap, we devise an end-to-end, depth image-based sim2real transfer method for improved generalization to real-world scenarios. Furthermore, to enable autonomous operation in varied and unstructured environments, we augment the navigation foundation model with a closed-loop Perspective-n-Point (PnP) localization mechanism, ensuring precise alignment of visual goals and effectively bridging autonomous navigation and dexterous manipulation. Extensive experimental results demonstrate that HERMES consistently exhibits generalizable behaviors across diverse, in-the-wild scenarios, successfully performing numerous complex mobile bimanual dexterous manipulation tasks. Project Page https://hermes-manipulation.github.io/

ACM Reference Format:
. 2018. HERMES: Human-to-Robot Embodied Learning from Multi-Source Motion Data for Mobile Dexterous Manipulation. *ACM Trans. Graph.* 37, 4, Article 111 (August 2018), 5 pages. XXXXXXX.XXXXXXX

1 Introduction

Humans continuously generate diverse bimanual manipulation data, inherently serving as natural guidance for robots to emulate human-like behaviors. Several previous studies [Dan et al. 2025; Kim et al. 2025; Lum et al. 2025; Wang et al. 2023; Zhou et al. 2025] have attempted to extract trajectories of human hands and manipulated objects from video data, subsequently applying them to robotic manipulation tasks. Nevertheless, these methods have predominantly targeted robots equipped with simple gripper-based end effectors, failing to generalize effectively to dexterous hands due to the vastly greater complexity of action space. Despite recent advances that utilize kinematic retargeting approaches to produce human-like robotic motions [Qin et al. 2023; Qiu et al. 2025; Shaw et al. 2023, 2024; Yang et al. 2025], these approaches still fall short in achieving physically-aware pose retargeting and bridging the embodiment gap to derive feasible robot actions capable of successfully accomplishing the intended tasks. ***A critical limitation lies in the omission of explicit modeling of interactions between robotic hands and manipulated objects, a fundamental component of manipulation tasks.*** Consequently, neglecting these interactions undermines the robot's ability to fully understand and adapt to the dynamics of manipulation scenarios.

Motivated by these challenges, we propose HERMES, a versatile human-to-robot embodied learning framework tailored for mobile bimanual dexterous hand manipulation. HERMES offers the following three advantages: **1. Diverse sources of human motion**: Our framework supports several human motion sources, including teleoperated simulation data, motion capture (mocap) data, and raw human videos. We also provide corresponding approaches for data acquisition, enabling HERMES to efficiently transform varied human motion data into robot-feasible behaviors through RL. Furthermore, these tasks share a uniform set of reward terms, obviating the necessity of designing intricate and task-specific reward functions. In contrast to the methods that depend on collecting a large amount of demonstrations, we can achieve generalizable policy by editing a single reference human motion trajectory coupling with RL training. **2. End-to-end vision-based sim2real transfer**: HERMES facilitates robust vision-based sim2real transfer by employing DAgger distillation, which converts state-based expert policies into vision-based student policies. Moreover, we introduce a generalized, object-centric depth image augmentation and hybrid control approach, effectively bridging the perception and dynamic sim2real gap. **3. Mobile manipulation capability**: Our method endows robots with mobile manipulation skills. Building upon ViNT [Shah et al. 2023], we develop a RGB-D based module for precise localization wherein the task is modeled as a Perspective-n-Point (PnP) problem and addressed through an iterative process. This ensures seamless integration with subsequent manipulation tasks and unlock the policy's capacity to operate autonomously across a broad spectrum of real-world environments.

2 Method

2.1 Collect One-shot Human Motion

To validate the effectiveness and robustness of HERMES, we employ three distinct sources of human motion: teleoperation in simulation, motion capture data obtained from public datasets, and hand-object poses extracted from raw videos. Moreover, by leveraging merely a single human reference trajectory in conjunction with RL training, we are able to derive the generalizable robot policy without the need for collecting extensive demonstrations.

Synthesize multiple trajectories: To obtain a more generalizable policy, we perform the trajectory editing for the one-shot human motion reference by randomizing the object's position and orientation in a predefined range. The hand and object poses across the augmented trajectories are transformed as follows:

$$\hat{\mathbf{A}}^{\text{pose}}\left[\tau_k\right] = \mathbf{T}^{\text{trans}} \cdot \mathbf{A}^{\text{pose}}\left[\tau_k\right]. \tag{1}$$

For any given frame k in the trajectory τ, we apply a transformation matrix $\mathbf{T}^{\text{trans}}$ to alter its pose, where $\mathbf{A}^{\text{pose}}$ may represent either the

Author's Contact Information:

ACM 1557-7368/2018/8-ART111
XXXXXXX.XXXXXXX

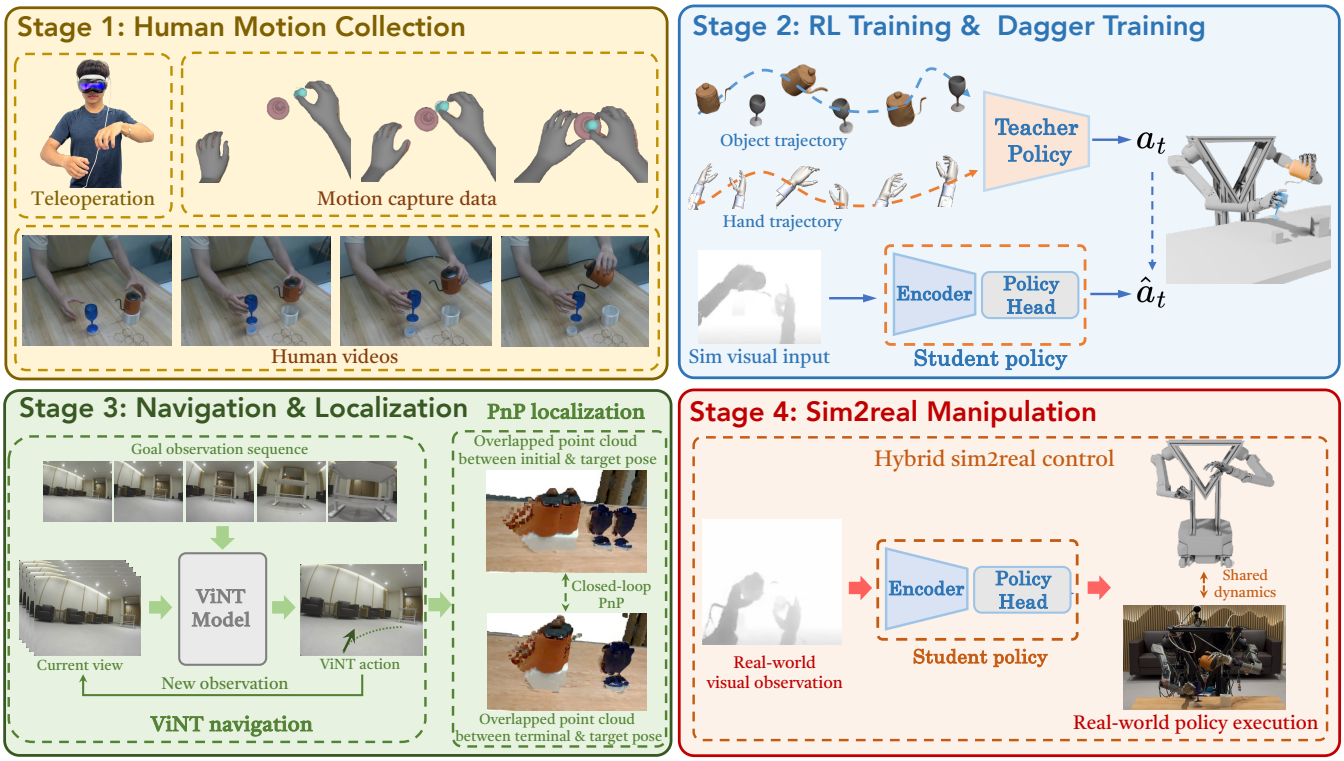

Fig. 1. **The main pipeline of HERMES.** HERMES comprises a four-stage pipeline for achieving mobile bimanual dexterous manipulation through sim2real transfer. First, we acquire a one-shot human demonstration drawn from diverse sources. Then, in stage 2, we train a state-based RL teacher policy, then apply DAgger to distill into a vision-based student policy. Following this, HERMES execute long-horizon navigation using ViNT, followed by real-time PnP to finely adjust the robot's pose and achieve precise alignment in stage 3. Once localization is achieved, the student policy is deployed in a zero-shot fashion directly in the real world.

object pose or the hand pose. By editing the reference trajectory, we enable spatial generalization from a single human motion demonstration, obviating the need to manually collect large numbers of teleoped demonstrations.

Upon obtaining synthesized object and hand trajectories from various data sources, we initially employ the DexPilot retargeting method [Handa et al. 2020] to map the captured human hand poses onto corresponding robot hand configurations. Subsequently, reinforcement learning is leveraged to refine and adapt the initialized robot behaviors.

2.2 Generalizable Reward Design for Manipulation

Standard reinforcement learning typically relies on hand-crafted reward functions tailored to each specific task. However, designing such complicated reward structures often impedes scalability and usability, particularly for the dexterous hand. To alleviate this issue, we leverage one-shot human demonstration combined with a generalizable reward formulation, enabling the reuse of a unified reward function across tasks and facilitating the straightforward construction of challenging, long-horizon manipulation tasks.

Object-centric Distance chain: Capturing the dynamic spatial relationships between the human hands and the object stands as a pivotal factor in enabling the policy to acquire fine-grained hand-object interaction skills. We designate the coordinates of the fingertips and palm of the hand, along with the center of the object's collision mesh, as keypoints. By modeling the temporal evolution of vectors between these keypoints, we formulate the following reward function:

$$r_{\text{chain}} = \begin{cases} \exp\left\{-\frac{1}{n}\sum_{i=1}^{n}\left\|\vec{r}_{\text{ref}}^{(i)} - \vec{r}^{(i)}\right\|\right\}, & \text{if } N_{\text{contact}} \geq N_{\text{num}} \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where $\vec{r}^{(i)}$ is the vector from object center to the fingertip or palm. Furthermore, we incorporate contact information into this reward term. Specifically, during the computation of the distance chain, we also evaluate the number of contact points between the fingertips and palms of both hand mesh $\mathbf{C}_{\text{hand}}$ and the object's collision mesh $\mathbf{C}_{\text{obj}}$. This reward component is activated only when the number of contact points N_{contact} exceeds a predefined threshold N_{num}, ensuring that the policy attends to physically meaningful hand-object interactions.

We also incorporate an object trajectory tracking and a power-penalty term to align the policy's behavior with the target object's trajectory and enhance the smoothness of policy execution and to alleviate the jittering actions. We adopt DrM [Xu et al. 2024], an off-policy method, leverages a dormant ratio mechanism [Sokar et al. 2023] to enhance exploration capabilities and demonstrates high sample efficiency.

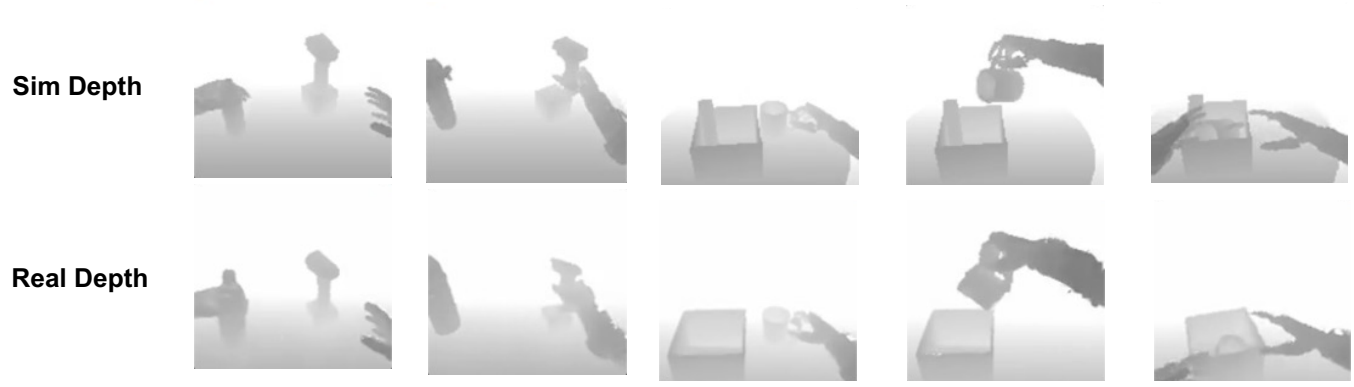

Fig. 2. **Depth image visualization.** After applying our preprocessing pipeline, the depth representations of the hand and object exhibit a strong semantic correspondence, highlighting the efficacy of HERMES in bridging the sim2real gap.

3 Sim-to-real Transfer

The training of state-based RL policies typically relies on privileged information which is not accessible in real-world deployment scenarios. Consequently, it is imperative to distill the state-based policy into a visual policy for achieving sim2real transfer. We leverage the depth image as visual input.

DAgger Distillation Training: In DAgger training, the state-based expert policy acts as the teacher to guide the learning of a visual student policy. In contrast to prior approaches that distill to object masks or segmented images, HERMES directly distills the state into raw visual observations of entire visual scenarios. This design obviates the need for explicit camera calibration and facilitates the acquisition of the robot's in-the-wild generalization capabilities. Furthermore, we introduce a series of auxiliary design choices aimed at enhancing both the asymptotic performance of DAgger training.

Hybrid sim2real control: To mitigate the gap between simulation and real-world dynamics as well as proprioceptive information, we adopt a hybrid control strategy: real-world visual observations are used to infer the actual action, which is then applied to the simulation environment to perform a forward step. The updated joint positions of the simulated robot are subsequently transferred to the real robot for execution. By sharing the same Inverse Kinematics (IK) method and dynamic parameters across simulation and the real world, this approach not only enables the policy to adapt its behavior based on real-world environmental variations but also effectively narrows the sim2real discrepancy.

4 Navigation Methodology

We choose ViNT [Shah et al. 2023] for achieving image-goal robotic navigation. ViNT not only enables long-range, in-the-wild navigation but also demonstrates effective zero-shot generalization capability without necessitating model fine-tuning. For our mobile manipulation tasks, moderate discrepancies between the robot's final pose and the target pose can lead to the manipulation policy failing to finish the task. However, ViNT does not guarantee termination within a sufficiently tight error bound. To address this, we introduce a local refinement step after ViNT completes navigation: a closed-loop Perspective-n-Point (PnP) localization algorithm is employed to adjust the robot's pose, ensuring closer alignment with the goal image pose.

5 Sample Efficiency of HERMES

We evaluate the training sample efficiency of HERMES across seven tasks. For each task, the source of the one-shot human motion demonstration is indicated in the title of each sub-figure in Figure 3. The vertical axis in the figure represents the proportion of the trajectory length successfully executed by the current policy relative to the total length of the trajectory. As demonstrated in Figure 3, regardless of the origin of the human motion data, HERMES reliably succeeds in converting human hand and arm actions into generalizable robot-executable behaviors.

Additionally, we compare training performances with ObjDex [Chen et al. 2024]. ObjDex defines its reward based on the tracking of the object's joint movement, translations, and orientations. We re-implement this reward formulation within our own algorithmic framework. Figure 3 indicates that HERMES exhibits superior performance relative to ObjDex across all tasks. In tasks such as *Bottle Handover*, *Flower Vase*, and *Putoff Burner*, where interactions involve only a single object, ObjDex is able to complete the tasks; however, HERMES can achieve higher sample efficiency during training. Furthermore, in more intricate tasks involving multi-object interactions, ObjDex consistently fails, irrespective of the type of human motion data provided. Contributed to our object-centric distance chain, HERMES is capable of robustly acquiring diverse manipulation skills even in long-horizon, multi-object environments. Moreover, HERMES demonstrates high sample efficiency and successfully learns policies in 3M training steps.

6 Mobile Manipulation Experiments

To evaluate the mobile manipulation ability of HERMES, we integrate the entire pipeline across all tasks. Each trained policy is tested over 10 runs. As illustrated in Figure 4, HERMES demonstrates strong real-world navigation, precise localization, and dexterous manipulation capabilities. We also apply the identical manipulation policy equipped with ViNT as a baseline. Figure 4 reveals that,

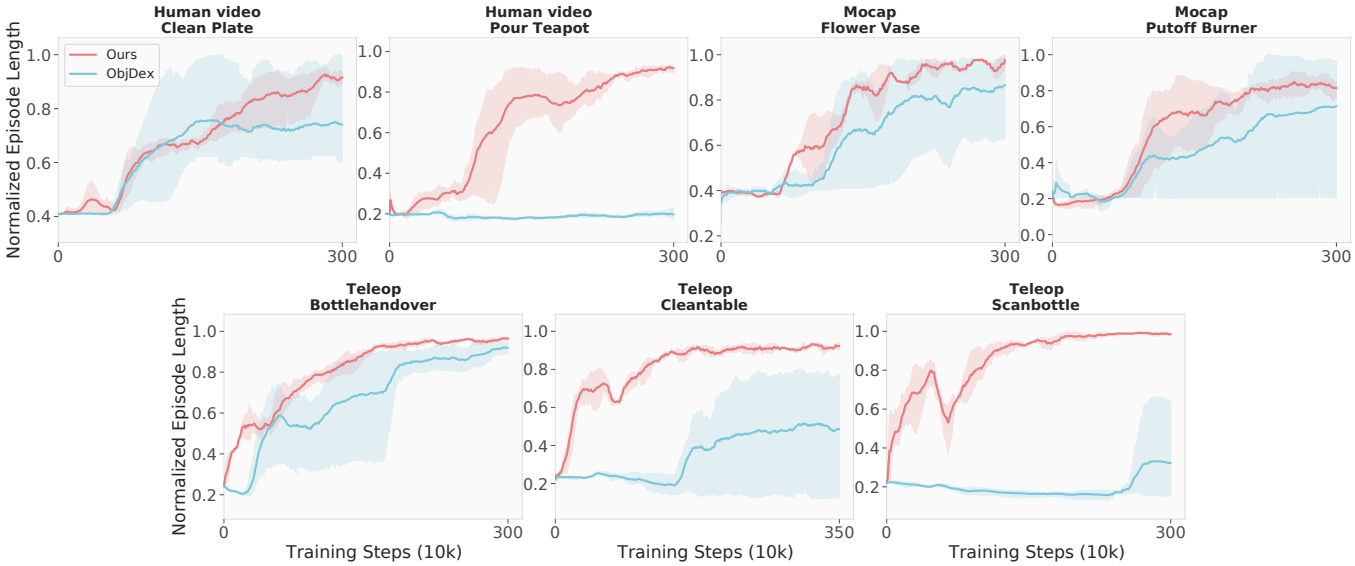

Fig. 3. **The training curve of HERMES.** The horizontal axis denotes the training steps, while the vertical axis represents the normalized task length successfully accomplished by the policy. *Teleop* refers to one-shot human motion teleoperation in simulation, *Human video* denotes trajectories extracted from video data, and *Mocap* corresponds to motion derived from mocap datasets. All results are evaluated across 3 seeds.

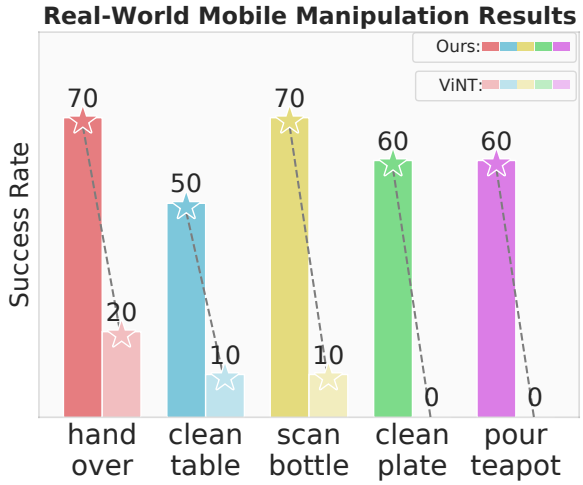

Fig. 4. **Real-world mobile manipulation results.**

without closed-loop PnP localization, the policy cannot generalize or successfully complete tasks when faced with significant positional and rotational shifts. Conversely, HERMES achieves a notable **+54.0%** improvement in manipulation success rate compared to pure ViNT. These findings underscore that closed-loop PnP localization is the essential bridge linking navigation and manipulation, enabling both modules to synergize for enhanced performance.

References

Yuanpei Chen, Chen Wang, Yaodong Yang, and Karen Liu. 2024. Object-Centric Dexterous Manipulation from Human Motion Data. In *8th Annual Conference on Robot Learning*. PMLR.

Prithwish Dan, Kushal Kedia, Angela Chao, Edward Weiyi Duan, Maximus Adrian Pace, Wei-Chiu Ma, and Sanjiban Choudhury. 2025. X-Sim: Cross-Embodiment Learning via Real-to-Sim-to-Real. *arXiv preprint arXiv:2505.07096* (2025).

Ankur Handa, Karl Van Wyk, Wei Yang, Jacky Liang, Yu-Wei Chao, Qian Wan, Stan Birchfield, Nathan Ratliff, and Dieter Fox. 2020. Dexpilot: Vision-based teleoperation of dexterous robotic hand-arm system. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 9164–9170.

Hanjung Kim, Jaehyun Kang, Hyolim Kang, Meedeum Cho, Seon Joo Kim, and Youngwoon Lee. 2025. UniSkill: Imitating Human Videos via Cross-Embodiment Skill Representations. *arXiv preprint arXiv:2505.08787* (2025).

Tyler Ga Wei Lum, Olivia Y Lee, C Karen Liu, and Jeannette Bohg. 2025. Crossing the Human-Robot Embodiment Gap with Sim-to-Real RL using One Human Demonstration. *arXiv preprint arXiv:2504.12609* (2025).

Yuzhe Qin, Wei Yang, Binghao Huang, Karl Van Wyk, Hao Su, Xiaolong Wang, Yu-Wei Chao, and Dieter Fox. 2023. Anyteleop: A general vision-based dexterous robot arm-hand teleoperation system. *arXiv preprint arXiv:2307.04577* (2023).

Ri-Zhao Qiu, Shiqi Yang, Xuxin Cheng, Chaitanya Chawla, Jialong Li, Tairan He, Ge Yan, David J Yoon, Ryan Hoque, Lars Paulsen, et al. 2025. Humanoid Policy˜ Human Policy. *arXiv preprint arXiv:2503.13441* (2025).

Dhruv Shah, Ajay Sridhar, Nitish Dashora, Kyle Stachowicz, Kevin Black, Noriaki Hirose, and Sergey Levine. 2023. ViNT: A Foundation Model for Visual Navigation. In *Conference on Robot Learning*. PMLR, 711–733.

Kenneth Shaw, Shikhar Bahl, and Deepak Pathak. 2023. Videodex: Learning dexterity from internet videos. In *Conference on Robot Learning*. PMLR, 654–665.

Kenneth Shaw, Shikhar Bahl, Aravind Sivakumar, Aditya Kannan, and Deepak Pathak. 2024. Learning dexterity from human hand motion in internet videos. *The International Journal of Robotics Research* 43, 4 (2024), 513–532.

Ghada Sokar, Rishabh Agarwal, Pablo Samuel Castro, and Utku Evci. 2023. The dormant neuron phenomenon in deep reinforcement learning. In *International Conference on Machine Learning*. PMLR, 32145–32168.

Chen Wang, Linxi Fan, Jiankai Sun, Ruohan Zhang, Li Fei-Fei, Danfei Xu, Yuke Zhu, and Anima Anandkumar. 2023. MimicPlay: Long-Horizon Imitation Learning by Watching Human Play. In *7th Annual Conference on Robot Learning*.

Guowei Xu, Ruijie Zheng, Yongyuan Liang, Xiyao Wang, Zhecheng Yuan, Tianying Ji, Yu Luo, Xiaoyu Liu, Jiaxin Yuan, Pu Hua, et al. 2024. DrM: Mastering Visual Reinforcement Learning through Dormant Ratio Minimization. In *The Twelfth International Conference on Learning Representations*.

Ruihan Yang, Qinxi Yu, Yecheng Wu, Rui Yan, Borui Li, An-Chieh Cheng, Xueyan Zou, Yunhao Fang, Hongxu Yin, Sifei Liu, et al. 2025. EgoVLA: Learning Vision-Language-Action Models from Egocentric Human Videos. *arXiv preprint arXiv:2507.12440* (2025).

Huayi Zhou, Ruixiang Wang, Yunxin Tai, Yueci Deng, Guiliang Liu, and Kui Jia. 2025. You Only Teach Once: Learn One-Shot Bimanual Robotic Manipulation from Video Demonstrations. *arXiv preprint arXiv:2501.14208* (2025).