Towards Learning Whole Body Mobile Manipulation for Grasping Any Object Anywhere

Jason Liu Changwei Yao Andrew Wang

***Abstract*—Generalizable whole-body mobile manipulation remains challenging because robots must operate across diverse objects and environments while handling high-dimensional control, collision avoidance, and partial observability. In this work, we present a scalable, fully simulation-based framework for learning generalizable whole-body mobile manipulation without human demonstrations. Our method decomposes the task into free-space motion and contact-rich interaction. For free-space motion, we introduce a GPU-vectorized whole-body controller (WBC) that coordinates end-effector tracking, active vision, and collision avoidance, substantially simplifying policy learning by abstracting away core whole-body behaviors. On top of this controller, we train a suite of reinforcement learning policies that focus on object interaction local to the end effector, reducing the effective degrees of freedom, task complexity, and horizon of the learning problem. We then distill the WBC and these local policies into a single vision-based policy controlling the mobile base, arm, dexterous hand, and actuated neck for active perception. We evaluate the framework on language-conditioned mobile dexterous grasping and demonstrate robust performance across diverse objects and scenes, along with emergent collision-aware and active-vision behaviors. Extensive comparisons and ablations show that every component of the pipeline contributes meaningfully to final performance, validating the proposed design as a scalable recipe for learning generalizable whole-body mobile manipulation.**

***Index Terms*—Mobile Manipulation, Sim-to-Real**

I. Introduction

Mobile manipulation has long been viewed as a central goal of robotics: a robot that can navigate through unseen scenes, perceive its surroundings, and carry out diverse tasks would unlock a broad range of real-world applications. Yet, despite substantial progress in robot learning, building mobile manipulators that operate robustly outside of curated lab settings remains a major challenge. Such systems must integrate navigation, perception, whole-body control, and dexterous manipulation under partial observability, relying only on onboard sensors while adapting to cluttered, diverse, and unstructured environments.

A common approach to learning mobile manipulation skills is to imitate expert demonstrations. While imitation learning has enabled impressive progress, scaling this paradigm for general mobile manipulation is difficult for many reasons. A robot must generalize not only across objects, but also across the environments in which those objects appear. This combinatorial diversity makes demonstration collection extremely data hungry. The challenge is further exacerbated for platforms whose morphology differs substantially from that of humans. In our setting, the robot consists of a holonomic mobile base, a high-degree-of-freedom arm, a dexterous hand, and an actuated neck for active perception, totaling 32 degrees of freedom. Collecting large-scale, high-quality demonstrations for such a system is cumbersome, and teleoperation or human video data do not directly map cleanly onto the robot's embodiment.

Simulation-based reinforcement learning offers an appealing alternative because it can generate large amounts of robot data without the need for human demonstrations. However, applying reinforcement learning (RL) directly to whole-body mobile manipulation is itself extremely challenging. Exploration becomes exponentially more difficult as the dimensionality of the action space grows, and long-horizon tasks require the robot to discover both free-space motion and contact-rich interaction. Moreover, unconstrained whole-body policies can produce motions that are difficult to regularize and may not transfer reliably to a physical robot. These issues are especially pronounced in mobile manipulation, where the robot must simultaneously coordinate its mobile base, arm, dexterous hand, and sensors while avoiding both self-collisions and collisions with the surrounding scene.

In this work, we propose a scalable, fully simulation-based framework for learning whole-body mobile manipulation without human demonstrations. Our key idea is to decompose mobile manipulation into free-space motion and contact-rich interaction. For free-space motion, we introduce a GPU-vectorized whole-body controller based on geometric fabrics that coordinates end-effector reaching, active vision, and collision avoidance while respecting robot motion constraints. This controller abstracts away core whole-body behaviors and allows reinforcement learning to focus on the local, contact-rich phase of the task, where the dexterous hand interacts with the object and nearby environment. We train local state-based dexterous grasping policies on top of this controller and distill them into a single closed-loop vision-based student policy, yielding an end-to-end system that controls the full mobile manipulator from onboard point-cloud observations and proprioception.

We evaluate our framework on mobile dexterous grasping across diverse objects and scenes. The resulting policy exhibits robust whole-body behavior, including collision-aware motion around obstacles, active camera motion that keeps the target object visible, and dexterous grasping across diverse object geometries. As summarized in the real-world examples in Figure 3 and the simulation ablations in Figure 4, the full system outperforms policies that remove active vision or

directly learn joint-space control.

In summary, our main contributions are:

1) A framework for learning whole-body mobile dexterous manipulation policies without human demonstrations.
2) A scalable GPU-vectorized whole-body controller that coordinates mobile-base, arm, dexterous-hand, and active-camera motion for end-effector reaching, visibility maintenance, and collision avoidance.
3) A vision-based distillation pipeline that converts controller-guided local dexterous RL teachers into a single closed-loop whole-body policy for language-conditioned grasping across diverse objects, placements, and environments.

II. Related Works

A. Sim-to-real for Manipulation

Recent advances in high-throughput physics simulation and GPU-accelerated reinforcement learning have made simulation an increasingly effective method for learning manipulation policies.

Prior works have demonstrated that dexterous policies trained entirely in simulation can transfer to real hardware [1], [2]. More recent works have extended this paradigm to vision-based policies, where the policy operates purely on point clouds [3], depth [4], or RGB pixels [5].

Additionally, local manipulation policies trained in simulation have been shown to compose effectively with perception, planning, and language-conditioned task decomposition to solve long-horizon real-world manipulation tasks [6]. However, these methods primarily study fixed-base or tabletop manipulation, where the robot operates from a largely static viewpoint and does not need to jointly reason about mobility, onboard sensing, and whole-body collision avoidance. Our work extends simulation-based manipulation learning to the mobile setting, where grasping requires coordinated control of the base, arm, dexterous hand, and active camera across diverse scenes.

B. Robot Learning for Mobile Manipulation

Learning-based mobile manipulation has largely been driven by demonstration-based approaches, where policies are trained from teleoperated robot trajectories or human demonstrations. Mobile ALOHA demonstrates that whole-body teleoperation and imitation learning can enable mobile bimanual robots to perform long-horizon household tasks [7]. Works have also shown that mobile robots can operate articulated objects such as doors, drawers, and cabinets in open-world environments by combining behavior cloning with online adaptation [8]. HoMMI further improves the scalability of data collection by learning whole-body mobile manipulation from robot-free human demonstrations, using cross-embodiment policy designs to bridge the gap between human demonstrations and robot execution [9]. Other works study active perception for manipulation, showing that learned camera motion can improve bimanual manipulation by enabling task-relevant search, tracking, and focusing behaviors [10]. Beyond demonstration-based methods, recent visual sim-to-real approaches for humanoid loco-manipulation show that whole-body policies trained in simulation can transfer to real hardware for mobile interaction tasks [11].

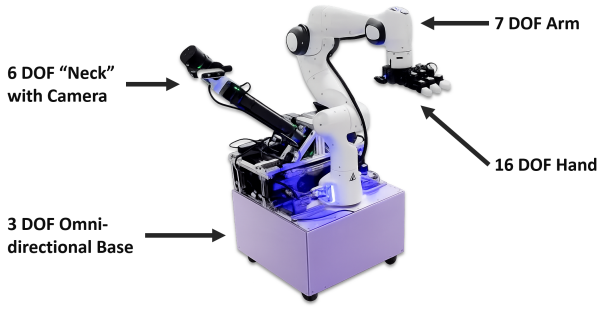


Fig. 1. System overview. The robot combines a holonomic mobile base, Franka Panda arm, LEAP dexterous hand, actuated active-vision arm, wrist-mounted RGB-D camera, and 3D LiDAR. Our framework decomposes mobile dexterous grasping into controller-guided free-space motion and local contact-rich interaction before distilling both stages into one onboard vision-based whole-body policy. Back to text.

While these works highlight the promise of learning-based mobile manipulation, they typically depend on human demonstrations and task-specific abstractions. In this work, we target a complementary setting: learning a whole-body mobile dexterous grasping policy entirely in simulation, without human demonstrations. This requires the policy to conduct closed-loop control from onboard partial observations, collision-aware motion through clutter, active camera control, and generalization across both objects and surrounding environments.

III. Method

Our goal is to learn a closed-loop whole-body dexterous mobile manipulation policy that can generalize across diverse objects and scenes. The robot must identify an object of interest, navigate through obstacles while avoiding collisions, keep the object within view using an actuated active vision camera, and then execute a dexterous grasp. Directly learning this behavior with RL in simulation over the full robot action space is difficult due to the high dimensionality of the system and long task horizon.

We therefore decompose the problem into two stages: free-space motion and contact-rich interaction. During free-space motion, a whole-body controller moves the robot end effector near the target object while coordinating the mobile base, arm, and active camera and enforcing collision avoidance. During contact-rich interaction, local RL policies control the hand and local end-effector motion to complete the grasp. Finally, we distill the controller and local policies into a single vision-based whole-body student policy that runs directly from onboard point-cloud observations and proprioception.

A. Hardware Platform

We create a robotic system consisting of a TidyBot 2 holonomic base [12], a 7-DoF Franka Panda arm [13], a 16-DoF LEAP hand [14], and a 6-DoF ARX PiPER active vision

arm [15]. The full system is shown in Figure 1. Attached to the end of the active vision arm is an Intel RealSense D435 [16], which provides RGB-D input. The robot is also equipped with an onboard Unitree L2 3D LiDAR [17]. The full system has 32 degrees of freedom, including the mobile base, arm, hand, and active vision arm.

B. Geometric Fabrics Controller

Whole-body mobile manipulation requires several coupled behaviors beyond the primary manipulation objective. The robot must move the end effector toward the target, keep the object visible, avoid collisions with the surrounding scene, avoid self-collisions, and maintain smooth motions that remain within joint and velocity limits. Learning all of these behaviors directly with sparse task reward is difficult because the policy must discover a multi-objective control strategy in a high-dimensional action space. We therefore introduce a non-learning-based whole-body controller that handles free-space behavior and exposes a simpler local interface to the learned policies.

Let $q \in \mathbb{R}^n$ denote the generalized whole-body configuration, including the planar base, arm joints, and active-vision joints. For each control objective i, we define a task map $x_i = \phi_i(q)$ with Jacobian $J_i(q) = \partial\phi_i/\partial q$. The controller specifies a desired task-space velocity v_i and positive semi-definite metric W_i for each objective. We solve a damped weighted least-squares problem at every control step:

$$\dot{q}^\star = \left(\sum_i J_i^\top W_i J_i + \lambda I\right)^{-1} \left(\sum_i J_i^\top W_i v_i\right), \qquad (1)$$

where λ is a damping term. This form allows objectives with different units and priorities to be combined through their metrics. The resulting whole-body velocity is integrated and tracked by low-level joint controllers.

The end-effector reaching objective tracks a desired pose $x_{ee}^\star$ supplied either by the free-space planner or by the local dexterous policy:

$$v_{ee} = K_p^{ee}(x_{ee}^\star - x_{ee}) - K_d^{ee}\dot{x}_{ee}. \qquad (2)$$

The active-vision objective aligns the camera optical axis with the target object centroid p_{obj} while maintaining a useful viewing distance. If $r_{cam}(q)$ is the camera optical axis and $\hat{r}_{obj}$ is the unit vector from the camera to the target, the gaze objective minimizes the angular error between these directions:

$$e_{gaze} = r_{cam}(q) \times \hat{r}_{obj}, \qquad v_{gaze} = -K_p^{gaze} e_{gaze}. \qquad (3)$$

This term encourages the robot to actively move the camera during approach rather than relying on a fixed viewpoint.

Collision avoidance is represented with repulsive geometric objectives. For each robot collision sphere or capsule and each obstacle primitive, we compute a signed distance $d_j(q)$. When the distance is below a safety margin d_{safe}, the controller adds a repulsive velocity along the distance gradient:

$$v_{col,j} = \begin{cases} \eta\left(\frac{1}{d_j} - \frac{1}{d_{safe}}\right)\frac{1}{d_j^2}\nabla_q d_j, & d_j < d_{safe}, \\ 0, & d_j \geq d_{safe}, \end{cases} \qquad (4)$$

where η controls repulsion strength. We use the same formulation for self-collision and environment collision, with separate safety margins and weights.

The controller also includes posture and smoothness objectives that keep the arm and neck away from joint limits and penalize abrupt base motion. All objectives are evaluated in parallel on the GPU across the simulated environments. Batched kinematics, distance queries, Jacobians, and linear solves make the controller compatible with large-scale reinforcement learning in Isaac Gym. As a result, the learned policies do not need to rediscover basic reaching, visibility, and collision-avoidance behavior; instead, they only need to provide local manipulation targets that the WBC can execute safely.

C. Learning Contact-Rich Dexterous Local Policies in Simulation

We set up the simulation scene with a random target object surrounded by obstacles such as shelves, bins, tables, and clutter. The robot is initialized in front of a table at a randomized pose relative to the object. The objective is to approach, grasp, and lift the target object while avoiding collisions with the environment and the robot itself.

Once the end effector is near the target object, the task becomes dominated by local contact-rich interaction between the dexterous hand, the object, and nearby scene geometry. We train a suite of local RL policies in Isaac Gym [18] using PPO [19].

Each local policy receives privileged state information during training, including object pose relative to the hand, robot proprioception, fingertip positions relative to the hand, and rough bounding-box information of the object. The policy outputs relative end-effector targets and dexterous hand commands:

$$a_t^{local} = \pi^{local}(s_t^{privileged}). \qquad (5)$$

The whole-body controller then tracks these end-effector targets while continuing to handle whole-body coordination, active vision, and collision avoidance. This design prevents the RL policy from needing to directly control the full robot joint space.

The reward encourages the hand to approach, make stable contact, lift the object, and maintain a secure grasp while discouraging collisions and unnecessary motion:

$$r_t = w_{reach} r_{reach} + w_{contact} r_{contact} + w_{grasp} r_{grasp} + w_{lift} r_{lift} + w_{stal} - w_{col} r_{col} - w_{act}\|a_t\|^2 - w_{smooth}\|a_t - a_{t-1}\|^2. \qquad (6)$$

The reaching term rewards reducing the distance between the hand and object. The contact term rewards fingertip-object contacts while penalizing contacts with non-target obstacles. The grasp term rewards opposing-finger contact patterns and hand configurations that enclose the object. The lift term rewards increasing the target object's height above the support surface. The stability term penalizes object slipping, excessive angular velocity, and loss of contact after lift-off. The collision and action terms regularize the policy for transfer.

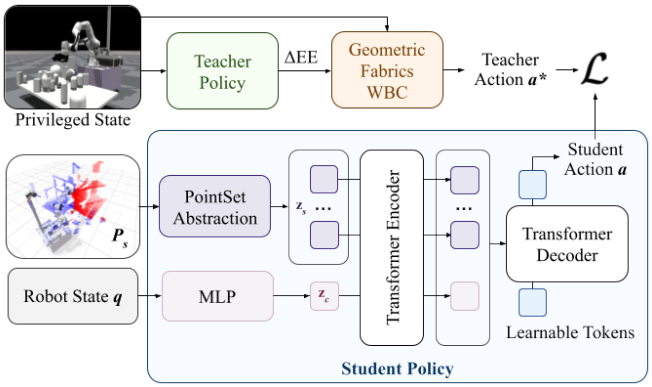


Fig. 2. Visual distillation pipeline. Teacher rollouts are generated by composing the GPU-vectorized whole-body controller with privileged-state local dexterous policies. The student policy receives only onboard observations: partial depth-camera point clouds, LiDAR point clouds, proprioception, and language-conditioned target information. Behavior cloning distills free-space reaching, active vision, collision avoidance, and contact-rich grasping into one whole-body policy. Back to text.

Episodes are randomized over object geometry, object pose, friction, mass, local obstacle layout, support-surface height, robot initialization, controller gains, and contact parameters. Different local policies specialize to different local geometries and grasp strategies. For example, a shelf-like scene may favor a side grasp, while a bin-like scene may favor a top-down grasp.

Because free-space motion is handled by the whole-body controller, these local policies only need to solve the final interaction phase. This reduces the effective action dimensionality, simplifies exploration, and shortens the training horizon. In practice, this enables efficient training of dexterous grasping behaviors despite the high-dimensional robot morphology.

D. Visual Distillation

We distill the controller and local policy teachers into a single closed-loop student policy, as shown in Figure 2. The student policy directly maps onboard point clouds, proprioception, and target conditioning to whole-body actions. It learns a unified behavior that approaches the object, keeps it visible, avoids obstacles, and executes the grasp to pick up the object.

1) Teacher Rollout Generation: We first generate teacher trajectories by composing the whole-body controller with the learned local policies. Each trajectory begins with the controller moving the end effector toward a pre-grasp pose near the target object. Once the robot reaches the local interaction region, one of the trained local policies takes over and produces hand commands and local end-effector targets. The controller remains active throughout the rollout to enforce whole-body coordination, active gaze, and collision avoidance.

This produces a dataset of successful whole-body trajectories:

$$\mathcal{D} = \{(o_t, a_t, p_t^{obj})\}_{t=1}^{T}, \tag{7}$$

where o_t contains the onboard observations available to the student policy and a_t is the corresponding teacher action. Importantly, the student does not receive privileged state information such as ground-truth object pose or complete scene geometry. Instead, it must act from the same partial onboard observations available at deployment time.

2) GPU Vectorized Sensor Simulation: Since mobile manipulation must be performed using onboard sensing, accurate modeling of partial observability is critical. To address this, we implement vectorized simulated renderers for both the depth camera and the LiDAR. The camera renderer produces partial point clouds from the articulated neck viewpoint, while the LiDAR renderer produces sparse geometric observations around the mobile base. Both renderers account for occlusion, field of view, sensor pose, and environment geometry. The resulting observations more closely match the partial and viewpoint-dependent sensing available on the physical robot.

The point clouds are generated for all parallel simulation environments on the GPU. This allows us to randomize thousands of distinct object and obstacle configurations while still producing realistic onboard observations at training time.

3) Policy Learning: Given teacher action a_t^* and student prediction $\pi_\theta(o_t)$, the primary distillation objective is

$$\mathcal{L}_{BC} = \mathbb{E}_{(o_t, a_t^*)\sim\mathcal{D}} \left[\|\pi_\theta(o_t) - a_t^*\|^2 \right]. \tag{8}$$

The action loss is applied to the mobile base, arm, hand, and active-vision joints. We additionally weight the hand and end-effector targets more heavily near contact, where small action errors can determine grasp success.

We apply domain randomization during distillation over object shape, scale, mass, friction, object pose, table and shelf geometry, obstacle placements, lighting-independent depth noise, point-cloud dropout, camera extrinsics, LiDAR sparsity, robot initial configuration, joint latency, controller gains, and observation delay. This randomization exposes the student to the sensing and dynamics variations expected at deployment and encourages the policy to rely on robust geometric cues rather than memorized simulator artifacts.

The student uses a point-cloud transformer architecture. Separate PointNet-style encoders first process the depth camera point cloud and the LiDAR point cloud into compact point tokens. These visual tokens are concatenated with proprioceptive tokens, previous-action tokens, and target-conditioning tokens. A transformer fusion module then attends over the visual, proprioceptive, and target tokens. The decoder outputs actions for the holonomic base, Franka arm, LEAP hand, and active-vision arm. The final policy runs closed-loop and controls all robot degrees of freedom directly.

IV. Results

A. Real World Grasping

We evaluate the learned whole-body policy on real-world mobile dexterous grasping trials with diverse objects and placements. The examples in Figure 3 show continuous grasping behavior in which the robot approaches the target, maintains visibility using the active camera, avoids nearby obstacles, closes the dexterous hand around the object, and lifts it from the support surface.

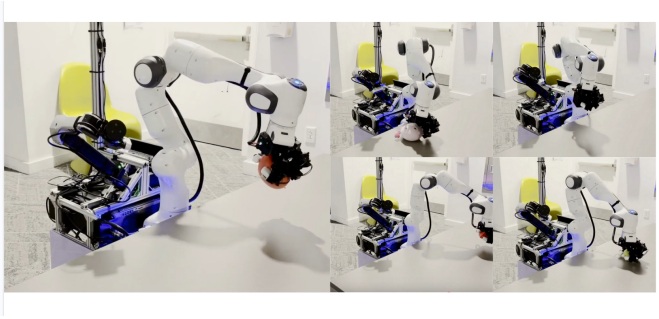
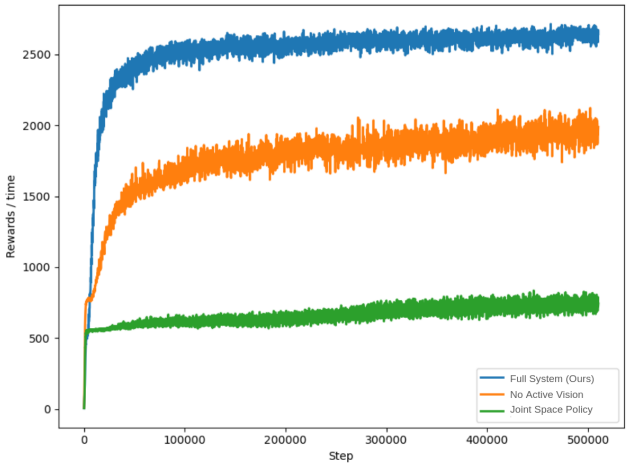


Fig. 3. Real-world grasping examples. The learned policy coordinates the mobile base, arm, dexterous hand, and active-vision arm to grasp diverse objects from cluttered tabletop scenes. The sequence illustrates whole-body approach, active target tracking, hand preshaping, object contact, and successful lift. Back to text.

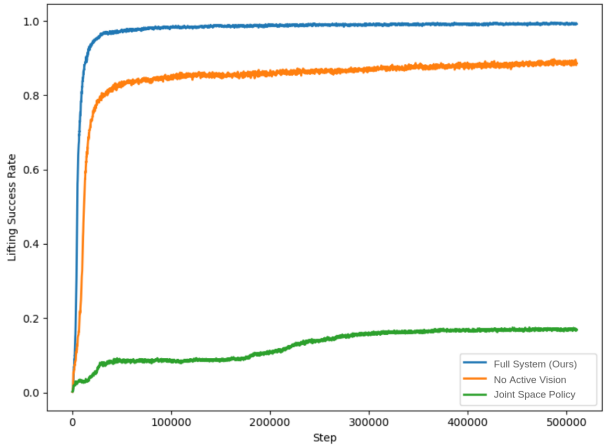
B. Simulation Ablations

We compare the full system against two ablations: a policy without active vision and a direct joint-space policy. The reward curves in Figure 4(a) show that the full system achieves the highest final reward. The no-active-vision ablation improves initially but saturates at a lower reward, indicating that controllable viewpoint selection is important for closed-loop grasping under partial observability. The joint-space policy plateaus much earlier, suggesting that directly learning high-dimensional whole-body control makes exploration and optimization substantially harder.

Figure 4(b) compares lifting success rate across methods. The full system reaches nearly 100% lifting success, while the no-active-vision ablation reaches roughly 80%. The joint-space policy reaches only about 20%, reflecting the difficulty of simultaneously learning base motion, arm reaching, dexterous grasping, active sensing, and collision avoidance without structured control.



(a) Training reward.



(b) Lifting success rate.

Fig. 4. Simulation ablations. The full controller-guided active-vision system achieves the highest reward and lifting success. Removing active vision reduces final performance because the policy loses the ability to maintain target visibility during approach and grasping. Direct joint-space learning performs worst, demonstrating the importance of decomposing free-space whole-body behavior from local contact-rich dexterous interaction. Back to text.

V. Conclusion

We presented a scalable simulation-based framework for learning whole-body mobile dexterous grasping without human demonstrations. By decomposing the task into controller-guided free-space motion and local contact-rich interaction, the method reduces the complexity of reinforcement learning while preserving full-body coordination. A GPU-vectorized geometric-fabrics controller provides reaching, active vision, and collision avoidance, while local RL policies learn dexterous contact behavior near the object. Visual distillation then combines these components into one closed-loop policy that acts from onboard point-cloud observations and proprioception. Real-world examples and simulation ablations indicate that active vision, structured whole-body control, and local dexterous learning are all important for robust mobile manipulation.

References

[1] ↩ OpenAI et al., "Solving Rubik's Cube with a Robot Hand," *arXiv preprint arXiv:1910.07113*, 2019.

[2] ↩ A. Handa, A. Allshire, V. Makoviychuk, A. Petrenko, R. Singh, J. Liu, D. Makoviichuk, K. Van Wyk, A. Zhurkevich, B. Sundaralingam, Y. Narang, J.-F. Lafleche, D. Fox, and G. State, "DeXtreme: Transfer of Agile In-Hand Manipulation from Simulation to Reality," *arXiv preprint arXiv:2210.13702*, 2022.

[3] ↩ Y. Qin, B. Huang, Z.-H. Yin, H. Su, and X. Wang, "DexPoint: Generalizable Point Cloud Reinforcement Learning for Sim-to-Real Dexterous Manipulation," in *Proceedings of the 6th Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, vol. 205, pp. 594–605, 2023.

[4] ↩ T. G. W. Lum, M. Matak, V. Makoviychuk, A. Handa, A. Allshire, T. Hermans, N. D. Ratliff, and K. Van Wyk, "DextrAH-G: Pixels-to-Action Dexterous Arm-Hand Grasping with Geometric Fabrics," *arXiv preprint arXiv:2407.02274*, 2024.

[5] ↩ R. Singh, A. Allshire, A. Handa, N. Ratliff, and K. Van Wyk, "DextrAH-RGB: Visuomotor Policies to Grasp Anything with Dexterous Hands," *arXiv preprint arXiv:2412.01791*, 2024.

[6] ↩ M. Dalal, M. Liu, W. Talbott, C. Chen, D. Pathak, J. Zhang, and R. Salakhutdinov, "Local Policies Enable Zero-Shot Long-Horizon Manipulation," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2025.

[7] ↩ Z. Fu, T. Z. Zhao, and C. Finn, "Mobile ALOHA: Learning Bimanual Mobile Manipulation with Low-Cost Whole-Body Teleoperation," *arXiv preprint arXiv:2401.02117*, 2024.

[8] ↩ H. Xiong, R. Mendonca, K. Shaw, and D. Pathak, "Adaptive Mobile Manipulation for Articulated Objects in the Open World," *arXiv preprint arXiv:2401.14403*, 2024.

[9] ↩ X. Xu, J. Park, H. Zhang, E. Cousineau, A. Bhat, J. Barreiros, D. Wang, and S. Song, "HoMMI: Learning Whole-Body Mobile Manipulation from Human Demonstrations," *arXiv preprint arXiv:2603.03243*, 2026.

[10] ↩ H. Xiong, X. Xu, J. Wu, Y. Hou, J. Bohg, and S. Song, "Vision in Action: Learning Active Perception from Human Demonstrations," in *Proceedings of the 9th Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, vol. 305, pp. 5450–5463, 2025.

[11] ↩ T. He, Z. Wang, H. Xue, Q. Ben, Z. Luo, W. Xiao, Y. Yuan, X. Da, F. Castañeda, S. Sastry, C. Liu, G. Shi, L. Fan, and Y. Zhu, "VIRAL: Visual Sim-to-Real at Scale for Humanoid Loco-Manipulation," *arXiv preprint arXiv:2511.15200*, 2025.

[12] ↩ J. Wu, W. Chong, R. Holmberg, A. Prasad, Y. Gao, O. Khatib, S. Song, S. Rusinkiewicz, and J. Bohg, "TidyBot++: An Open-Source Holonomic Mobile Manipulator for Robot Learning," in *Conference on Robot Learning*, 2024.

[13] ↩ Franka Robotics, "Franka Control Interface Documentation (Franka Research 3 and Panda)." https://frankarobotics.github.io/docs/. Accessed: 2026-04-20.

[14] ↩ K. Shaw, A. Agarwal, and D. Pathak, "LEAP Hand: Low-Cost, Efficient, and Anthropomorphic Hand for Robot Learning," *Robotics: Science and Systems*, 2023.

[15] ↩ AgileX Robotics, "PiPER 6-DOF Robotic Arm." https://global.agilex.ai/products/piper. Accessed: 2026-04-20.

[16] ↩ Intel RealSense, "Intel RealSense Depth Camera D435." https://www.intelrealsense.com/depth-camera-d435/. Accessed: 2026-04-20.

[17] ↩ Unitree Robotics, "Unitree 4D LiDAR L2." https://www.unitree.com/mobile/L2. Accessed: 2026-04-20.

[18] ↩ V. Makoviychuk, L. Wawrzyniak, Y. Guo, M. Lu, K. Storey, M. Macklin, D. Hoeller, N. Rudin, A. Allshire, A. Handa, and G. State, "Isaac Gym: High Performance GPU-Based Physics Simulation for Robot Learning," in *Advances in Neural Information Processing Systems Datasets and Benchmarks Track*, 2021.

[19] ↩ J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[20] ↩ N. Carion, L. Gustafson, Y.-T. Hu, S. Debnath, R. Hu, D. Suris, C. Ryali, K. V. Alwala, H. Khedr, A. Huang, J. Lei, T. Ma, B. Guo, A. Kalla, M. Marks, J. Greer, M. Wang, P. Sun, R. Rädle, T. Afouras, E. Mavroudi, K. Xu, T.-H. Wu, Y. Zhou, L. Momeni, R. Hazra, S. Ding, S. Vaze, F. Porcher, F. Li, S. Li, A. Kamath, H. K. Cheng, P. Dollár, N. Ravi, K. Saenko, P. Zhang, and C. Feichtenhofer, "SAM 3: Segment Anything with Concepts," *arXiv preprint arXiv:2511.16719*, 2025.