Learning Various Assistive Skills for Manipulation

Ruihai Wu

Peking University

Abstract. This research focuses on equipping robots with a diverse set of manipulation skills crucial for a wide range of assistive applications. We begin by exploring interactive perception for articulated object manipulation, enabling robots to handle objects with complex joints, such as doors and drawers. Building on this, we address deformable object and garment manipulation, where robots learn to manage more complex items like shirts. Next, we investigate collaborative bimanual manipulation, allowing robots to further perform coordinated tasks with both arms. Finally, we consider the complexity of real-world environments, empowering robots to operate effectively in diverse and dynamic scenes Together, these advancements facilitate the deployment of assistive robots in various real-world scenarios.

Interactive Perception for Articulated Object Manipulation

To help humans perform everyday tasks, future assistant robots need to gain the capabilities of perceiving and manipulating diverse objects in human environments. Articulated objects that contain functionally important and semantically interesting articulated parts (e.g., cabinets with drawers and doors) require significant attention, as they are more often interacted with by humans. Having much higher degree-of-freedom (DoF) state spaces, articulated objects are, however, generally more difficult to understand and subsequently to interact with

To manipulate an articulated object, traditional methods usually design heuristic policies, lacking generalization to diverse scenarios. For instance, to open a door, they detect the handle, grasp the handle and execute a calculated trajectory. However, the diverse handle geometries make it difficult to compute accurate grasp poses for different handles, let alone doors without handles.

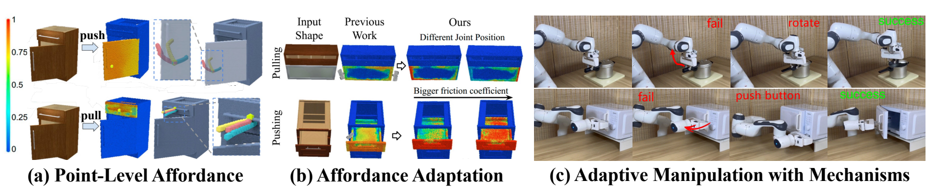


Fig. 1: Point-level affordance (a), affordance adaptation via active interactions tackling kinematics ambiguities (*e.g.*, on which side to open the door without a handle; the effect of surface friction in pulling the drawer) (b), and adaptive manipulation on objects with different mechanisms (*e.g.*, rotate or push button to unlock before manipulation) (c).

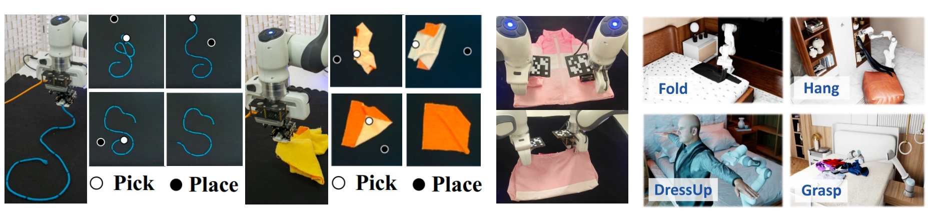
The success of Computer Vision reflects, learning based methods have promising performance in finding out essential features for a task using large-scale data with promising generalization capability. For many manipulation tasks, the most essential property of objects is whether the geometry around the target manipulation point (*e.g.*, a point on the handle), can afford the action to fulfill the task. Such property is in align with *Affordance* proposed by Gibson [3] in the per-point level. Demonstrated in Figure 1 (a), my research [13] proposes learning to **perceive** such geometry-aware point-level affordance using large-scale trial-and-error data generated by large-scale **interactions**, with actionable points colored in orange and red, and trajectory proposals for downstream tasks.

Given pure passive visual observations for perception and manipulation, there exist uncertainties that require robots to actively explore target objects and then adapt policies. Such uncertainties mainly come from two aspects, kinematics and geometries. Kinematic ambiguities include joint limits (*e.g.*, push inward or pull outward for a door) and axis (*e.g.*, where to pull, shown in Figure 1, b, upper), and physics parameters like friction and mass (Figure 1, b, lower). Geometry ambiguities refer to unseen local geometries. My studies [8,14] respectively integrate passive visual inputs with active actions by exploring points that mostly reflect kinematic and geometric uncertainties, using interaction results to adapt affordance and policies. Further, for objects with complex manipulation mechanisms, we set up environments with diverse mechanisms [6] and propose adaptive policy to tackle difficult mechanisms via interaction trials (Figure 1, c).

Deformable Object and Garment Manipulation

Compared with rigid and articulated objects, deformable objects pose much more challenges for manipulation, due to the exceptionally large and even infinite state and action spaces and highly complex dynamics. Besides, unlike tasks for rigid (like grasping) or articulated objects (like pushing a door) that require one or only a few steps, deformable object manipulation (like unfolding) usually requires many steps to accomplish, making it difficult for models with limited capacities to learn complex states in different steps.

My study [11] leverages point-level affordance to represent deformable objects, as per-point representation is a natural match for deformable objects with



(a) Long-Horizon Rope and Fabric Manipulation (b) Garment Manipulation (c) Garment Benchmark

Fig. 2: Diverse deformable object manipulation, from relatively simple ropes and fabrics (a), to more complex garments (b), and the corresponding unified benchmark (c).

complex states. We further empower learned affordance with the awareness of future steps to support multi-step long-horizon manipulation (Figure 2, a).

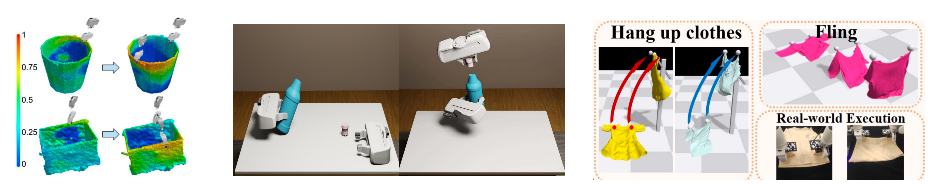
Among different kinds of deformable objects, garments, such as shirts, dresses and trousers, are highly essential for the various real-world applications, while typically challenging for their diverse shapes with large deformations. For garments with different deformations, to have an unified understanding of garments and thus facilitate a long range of downstream tasks, my study, UniGarmentManip [10], leverages the topological and structural similarity of garments in the category level, and thus can manipulate unseen garments only one- or few-shot demonstrations on a demonstration garment.

Furthermore, while most studies only focus on a certain garment type or a certain manipulation task, we propose a unified simulation environment and benchmark [7] supporting manipulating more than 10 categories of garments with different materials (such as shirts, hats, ties and socks) simulated by different methods, with diverse manipulators (such as mobile robots, dexterous hands and dual-arm robots), boosting the broad and thorough study, as well as future large-scale applications of garment manipulation.

Collaborative Bimanual Manipulation

Many tasks, such as steadily picking up a heavy basket and assembling broken parts into a whole shape, cannot be fulfilled using a single arm. However, it is much more challenging to conduct such bimanual manipulation, as the degree-of-freedom in action space is doubled. Besides, the proposed pairs of actions must be aware of collaborations.

My study, DualAfford [15], disentangles affordance and policy learning of two arms into two separate yet highly coupled subtasks, reducing the complexity from a intrinsically quadratic problem into a conditional problem by sequentially and conditionally predicting the two arms' affordance (Figure 3, a). Further studies tackle the more challenging geometric assembly task [1, 12] (Figure 3, b) by taking both the geometry of breaking parts and the collaboration of two arms into consideration. Additionally, for garments with both large state and bimanual action space, my work [10] leverages point-level correspondence as the guidance for few-shot bimanual manipulation (Figure 3, c).



(a) Bimanual Affordance (b) Geometric Assembly (c) Bimanual Garment Manipulation

Fig. 3: Bimanual affordance for picking up a bucket and a basket (a), bimanual collaboration for geometric assembly (b).

Manipulation in Complex Environments

Most existing works on manipulation primarily focus on single-object scenarios with homogeneous agents such as flying grippers. Consequently, these approaches tend to develop object-centric representations and policies, neglecting the realistic constraints imposed from or benefits brought by complex environments, which are commonplace in real-world scenarios. For example, successfully opening a cabinet door that is obstructed by occluders not only depends on the properties of the target door but also heavily relies on the robot's position and the way it interacts (e.g., colliding or bypassing) with the occluders.

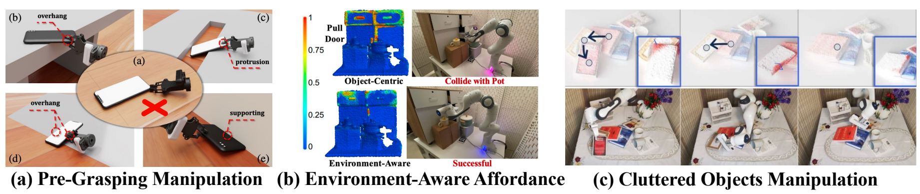


Fig. 4: Diverse environments for manipulation.

My study [9] formulates the task of environment-aware affordance learning to explore manipulation within environment constraints such as occlusions and robot configurations (Figure 4, b). The following work [5] leverages the environment-aware affordance for a mobile robot to walk around the occlusions and then safely accomplish the task. For cluttered scenes which also require safe manipulation, as disturbing and even breaking other objects will easily happen when retrieving a specific target object, we leverage support relations between nearby objects as the constraint to guarantee safety [4] (Figure 4, c). On the contrary, when the environment brings benefits for manipulation, my study, PreAfford [2], offers a robust solution (affordance learning for pre-grasping manipulation) for grasping hard-to-grasp objects with the support of the environment (Figure 4, a). For example, while a phone is difficult to grasp, pushing the phone until part of it is over the edge of the table allows us to robustly grasp that part.

References

1. Co-firstauthor: Ongoing project. 3
2. Ding, K., Chen, B., **Ruihai Wu**, Li, Y., Zhang, Z., Gao, H.a., Li, S., Zhu, Y., Zhou, G., Dong, H., et al.: Preafford: Universal affordance-based pre-grasping for diverse objects and environments. IROS (2024) 4
3. Gibson, J.J.: The theory of affordances. Hilldale, USA **1**(2), 67–82 (1977) 2
4. Li*, Y., **Ruihai Wu***, Lu, H., Ning, C., Shen, Y., Zhan, G., Dong, H.: Broadcasting support relations recursively from local dynamics for object retrieval in clutters. In: Robotics: Science and Systems (2024) 4

5. Li*, Y., Cheng*, K., **Ruihai Wu***, Shen, Y., Zhou, K., Dong, H.: Mobileafford: Mobile robotic manipulation through differentiable affordance learning. In: 2nd Workshop on Mobile Manipulation and Embodied Intelligence at ICRA 2024 (2024), https://openreview.net/forum?id=f6sc69sfrL 4
6. Li*, Y., Zhang*, X., **Ruihai Wu***, Zhang, Z., Geng, Y., Dong, H., He, Z.: Unidoormanip: Learning universal door manipulation policy over large-scale and diverse door manipulation environments. arXiv preprint arXiv:2403.02604 (2024) 2
7. Lu*, H., Li*, Y., **Ruihai Wu***, Ning, C., Shen, Y., Dong, H.: Unigarment: A unified simulation and benchmark for garment manipulation. In: ICRA Workshop on Deformable Object Manipulation (2024) 3
8. Ning, C., **Ruihai Wu**, Lu, H., Mo, K., Dong, H.: Where2explore: Few-shot affordance learning for unseen novel categories of articulated objects. NeurIPS (2023) 2
9. **Ruihai Wu***, Cheng*, K., Shen, Y., Ning, C., Zhan, G., Dong, H.: Learning environment-aware affordance for 3d articulated object manipulation under occlusions. NeurIPS (2023) 4
10. **Ruihai Wu***, Lu*, H., Wang, Y., Wang, Y., Hao, D.: Unigarmentmanip: A unified framework for category-level garment manipulation via dense visual correspondence. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2024) 3
11. **Ruihai Wu***, Ning*, C., Dong, H.: Learning foresightful dense visual affordance for deformable object manipulation. In: ICCV (2023) 2
12. **Ruihai Wu***, Tie*, C., Du, Y., Shen, Y., Dong, H.: Leveraging se(3) equivariance for learning 3d geometric shape assembly. In: ICCV (2023) 3
13. **Ruihai Wu***, Zhao*, Y., Mo*, K., Guo, Z., Wang, Y., Wu, T., Fan, Q., Chen, X., Guibas, L., Dong, H.: VAT-mart: Learning visual action trajectory proposals for manipulating 3d ARTiculated objects. In: ICLR (2022) 2
14. Wang*, Y., **Ruihai Wu***, Mo*, K., Ke, J., Fan, Q., Guibas, L., Dong, H.: Adaafford: Learning to adapt manipulation affordance for 3d articulated objects via few-shot interactions. ECCV (2022) 2
15. Zhao*, Y., **Ruihai Wu***, Chen, Z., Zhang, Y., Fan, Q., Mo, K., Dong, H.: Dualafford: Learning collaborative visual affordance for dual-gripper object manipulation. ICLR (2023) 3