Any-ttach: Quick End-effector Swapping Enables Manipulation Dexterity with Simplicity

Author Names Omitted for Anonymous Review. Paper-ID [8]

***Abstract*— Robotic manipulation dexterity is often pursued by building increasingly complex high-DoF multifingered hands. While many robotic hands are designed to replicate the morphology of human hands, the functional role of human hands suggests a different perspective: much of their complexity may exist to enable tool use and tool making. This observation motivates *Any-ttach*: a tool-centric manipulation framework that leverages mechanical modularity and treats end-effector swapping as a primary mechanism for dexterity. *Any-ttach* combines three components: 1) a low-cost, automatic, and easily deployable swapping mechanism for a 1-DoF parallel gripper, 2) a handheld device for collecting human demonstrations, and 3) a task planning framework that composes learned, parameterized, and planned skills for flexible use of diverse tools. The system supports everyday tools, articulated tools such as scissors, and a low-cost anthropomorphic hand through the same end-effector interface. Our experiments show that *Any-ttach* improves tool-swapping reliability, enables more efficient demonstration collection, reduces tool-pose variability, and supports diverse tools and end-effector modules. In two long-horizon tasks, making a sandwich and preparing a cucumber, *Any-ttach* executes 6 tool use subskills and demonstrates that hierarchical tool-skill decomposition can improve complex task reliability. More details and videos are available at `https://any-ttach.github.io/`.**

I. Introduction

General-purpose dexterity remains a central challenge in robotic manipulation. Most existing approaches treat dexterity as an intrinsic property of the end-effector, seeking broader capability by increasing mechanical complexity through high degree of freedom hands, comprehensive contact surfaces, and dense sensing [1]–[4]. Although these systems demonstrate impressive performance, they tightly couple manipulation dexterity with hardware complexity and intricate contact-rich control. Even simple tasks require precise regulation of high dimensional contact dynamics, and the limitation becomes especially clear in long horizon tasks where robots must handle both low-level contact control and high-level skill sequencing.

This motivates an alternative view: manipulation capability need not come only from increasing end-effector complexity. The functional role of human hands offers an evolutionary perspective: much of the dexterity comes from enabling tool use and tool making [5], [6]. Tools expand manipulation capability by embedding task-specific contact geometry into simple physical interfaces, rather than requiring all dexterity to reside in the hand itself. This view aligns with extrinsic dexterity [7], where simple end-effectors achieve diverse behaviors by exploiting environmental constraints, object geometry, and contact interactions. Both perspectives point to a tool-centric route: robots can expand capability by choosing the right physical interface for each interaction and switching between tools and end-effectors as the task changes to achieve manipulation dexterity.

Fig. 1. **Tool-centric design achieves dexterity and end-effector swapping with simplicity.** Any-ttach enables robots to perform diverse manipulation skills by rapidly switching between interchangeable tool modules through a standardized mechanical interface. By externalizing task-specific contact geometry into tools, the system reformulates manipulation dexterity as tool selection and skill execution rather than complex end-effector control.

We propose ***Any-ttach***, a manipulation system that enables a simple 1-DoF parallel gripper to automatically switch among and securely operate everyday tools. Instead of relying on dexterous grasping or permanent attachment to use tools, *Any-ttach* provides efficient automatic end-effector swapping and reliable mechanical attachment, supporting stable tool use in contact-rich interactions. Through this design, we aim to explore an alternative path of manipulation dexterity with simplicity: dexterity can emerge from composing skills across end-effectors and exploiting contact interactions within each skill, allowing even a simple 1-DoF gripper to achieve rich manipulation capabilities. Our contributions are summarized as follows:

- **An automatic quick-swap mechanical interface** that enables a 1-DoF parallel gripper and handheld interface to directly attach and exchange diverse everyday tools, including 0-DoF tools, 1-DoF articulated tools, robot fingers, and a low-cost anthropomorphic hand.
- **A hierarchical framework** that integrates task planning, tool selection, skill composition, and skill execution. The framework supports heterogeneous skill implementations and flexible integration for different end-effectors.
- **Preliminary experiments** demonstrating reliable tool

swapping, compatibility with diverse tools and skill types, and long-horizon manipulation through composed tool-use behaviors.

II. Related Work

A. Dexterity in Robotic Manipulation

A dominant paradigm in manipulation dexterity pursues intrinsic dexterity, embedding capability directly in the end-effector morphology. High-DoF manipulators [4], [8] provide rich actuation but require coordinating many joints under complex contact dynamics, often relying on large-scale simulation and learning infrastructure [9], [10]. Prior work [8] demonstrated in-hand cube rotation via reinforcement learning, and subsequent work advances dexterous grasping [3], object-centric manipulation [2], and functional grasping policies [11]. In contrast, extrinsic dexterity [7] shows that simple grippers can achieve rich behaviors, with extensions through structured primitives [12], [13] and contact-mode control [14]. While intrinsic approaches embed capability within the hand itself, extrinsic approaches leverage structure available in the environment [15]. Our approach differs from both directions. Instead of increasing in-hand complexity or relying on environmental fixtures, we externalize task-specific contact geometry into interchangeable tool modules. This shifts part of the interaction structure from continuous control into simple geometric design, enabling dexterity through controlled reconfiguration of the end-effector rather than higher morphological complexity.

B. Tool Manipulation

Tool use extends manipulation capabilities beyond the native end-effector geometry [16]. Prior work has studied affordance learning [17], pose estimation [18], [19], contact-rich skill acquisition through imitation learning [20], dynamics learning [21], and compliant control [22], as well as tool selection and sequencing [23], [24]. Most systems grasp tools using a general-purpose gripper or anthropomorphic hand, introducing a gripper–tool interface in addition to the tool–object interaction. Disturbances at this interface can cause tool slip or rotation, reducing execution repeatability [25]–[27]. Existing solutions either employ high-DoF hands to regulate the grasp [11], [16], [17] or permanently mount task-specific end-effectors at the cost of flexibility. *Any-ttach* instead uses a mechanically constrained quick-swap interface, preserving tool-switching flexibility while reducing grasp-induced pose variability.

C. Industrial Tool Changers

Industrial tool changers enable quick swapping through various mechanisms such as pneumatically-actuated cam-and-ball locks [28], [29], bayonet locks driven by robot wrist motion [30], magnetic couplings [31], [32], 3D-printed kinematic couplings [33], but they typically require extra actuation, are designed for pre-programmed industrial workflows, only pair with specially designed tools, and are hard to be applied for everyday tools. In contrast, our approach provides a passive and automatic quick-swap interface that

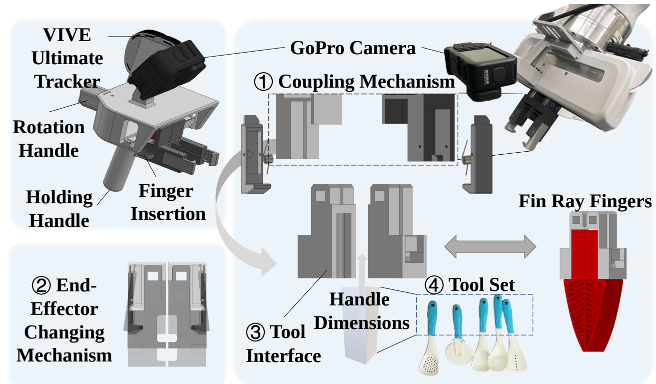

Fig. 2. **Hardware Design.** *Any-ttach* uses a shared mechanical interface to couple diverse tools and end-effector modules to both the robot arm and the handheld demonstration device. The system includes: (1) a mechanically constrained coupling mechanism for repeatable attachment, (2) an automatic end-effector changing mechanism for locking and release, (3) tool-side adapters that convert different handle dimensions into a standardized connector, and (4) an expandable tool set. This shared interface preserves consistent tool geometry across demonstration and deployment.

does not require extra actuation, is compatible with planning and learning-based manipulation frameworks, and can be easily mounted various everyday tools on robot fingers and allow 1-DoF actuation on the tools.

III. Method

Any-ttach integrates a standardized mechanical interface with hierarchical task execution. The hardware interface provides repeatable tool attachment, autonomous tool exchange, and consistent tool geometry between human demonstration and robot execution (Sec. III-A, Sec. III-B). At the software level, a task planner composes tool-conditioned skills, while runtime monitoring verifies tool attachment and task progress (Sec. III-C).

A. Unified Tool Interface and Quick-Swap Mechanism

Fig. 2 presents our hardware design of the system. The robot and handheld demonstration device share a standardized receptacle, while each tool module terminates in a mating connector. Passive guide rails and tapered constraints guide insertion toward a unique engagement pose, yielding a repeatable transform between the robot flange and tool frame.

New tools are incorporated through lightweight tool-specific adapters. We approximate a tool handle by the envelope $\mathbf{d}_h = [l_h, w_h, t_h]$ and subtract its tolerance-inflated mesh from a shared adapter template: $\mathscr{A}_h = \mathscr{A}_0 \setminus \mathscr{H}(\mathbf{d}_h + \varepsilon)$.

The resulting adapter conforms to the tool handle while preserving the standardized connector. This separates tool-specific geometry from the robot-side mechanism, allowing the tool library to expand without redesigning the coupling interface.

The same interface supports autonomous tool exchange without actuated docks. To attach a tool, the robot inserts its connector into the receptacle, where the alignment features guide it into a passively locked configuration. To detach it, the robot presses the release feature against the dock, rotating the latch and disengaging the lock before retreating. When no module is attached, the robot retains its native parallel-gripper configuration.

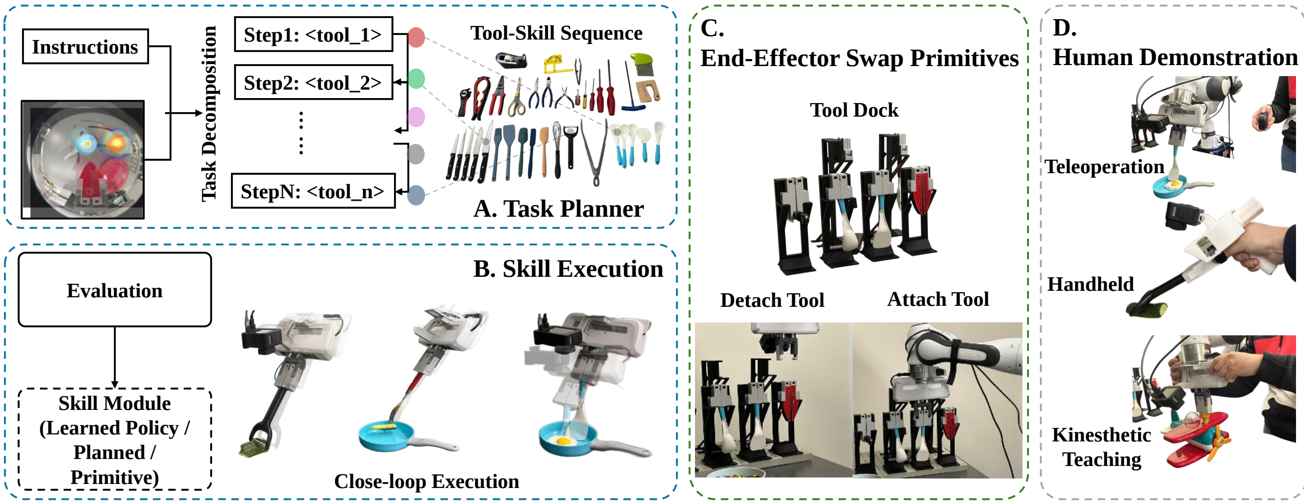

Fig. 3. **System pipeline of Any-ttach.** (A) **Task Planner:** a vision language model decomposes instruction into an ordered sequence of tool–skill pairs. (B) **Skill Execution:** learning-based policies execute each skill in closed loop using visual and proprioceptive observations. (C) **End-effector Swap Primitives:** the robot autonomously docks, attaches, and detaches tool modules through the standardized quick-swap interface. (D) **Human demonstration:** a handheld interface compatible with the robot coupling enables demonstration collection through handheld manipulation, kinesthetic teaching, or teleoperation. The unified tool interface ensures consistent tool geometry across demonstration and execution, enabling reliable tool use and long-horizon manipulation.

B. Handheld Demonstration Interface

The handheld device uses the same coupling geometry as the robot, ensuring that demonstrations and robot execution share an identical tool configuration. Unlike trigger-based or handle-pivoted devices such as UMI [34], our design places the coupling near the operator's fingers, allowing direct manipulation of the tool while a supporting handle provides wrist-level stability.

A VIVE tracker [35] records the coupled tool's 6-DoF pose in a calibrated robot reference frame. The shared mechanical interface preserves the tool frame across embodiments, reducing grasp-induced geometric variation and simplifying the transfer of demonstrations to robot skill learning.

C. Hierarchical Planning and Execution

As shown in Fig. 3, we equip the robot with a quick-swap interface $\mathscr{I}$, a tool library $\mathscr{Z} = \{z_1, \dots, z_M\}$, and a skill library $\Sigma = \{\sigma_1, \dots, \sigma_N\}$. Each tool provides a distinct contact geometry, and each skill defines a corresponding control primitive. Given a task instruction, the system generates and executes an ordered tool–skill sequence $\mathscr{P} = [(z^{(k)}, \sigma^{(k)})]_{k=1}^{K}$.

a) Task-level planning: A vision–language model (OpenAI GPT-5 [36]) receives the task instruction, current scene observation, and available tools, and decomposes the task into ordered steps specifying the subtask, tool, and skill. This decomposition enables long-horizon behavior through the composition of short-horizon, tool-conditioned skills.

b) Skill-level execution: The execution layer supports heterogeneous implementations, including learned policies and planning-based controllers. Learned skills use Diffusion Policy [37] in a receding-horizon manner, benefiting from the repeatable manipulator–tool transform provided by the mechanical interface. Skills defined by target poses or motion primitives use conventional motion planning, implemented with MoveIt [38]. This modular design combines contact-rich learned behaviors with geometric planning within a common execution framework.

c) Execution monitoring: The system verifies both the pre-condition and post-condition of each skill. Before execution, vision–language reasoning and tool segmentation confirm that the planned tool is attached. The tool name serves as a semantic prompt for SAM3 [39]; if the attachment is incorrect, the robot returns the current tool and acquires the planned one. After execution, a VLM evaluates the scene against the intended skill outcome. Failed skills are retried up to three times before the system reports failure.

IV. Experiments

Any-ttach system integrates three key design elements: 1) a kinematically constrained tool interface, 2) a shared human–robot demonstration setup, and 3) a hierarchical planning-execution-evaluation architecture. We therefore organize our experiments around three questions that evaluate how these elements affect tool swapping, demonstration collection, skill execution, and multi-step task robustness:

- **Q1: Swapping mechanism effectiveness.** How reliable and efficient is the proposed end-effector swapping mechanism compared to grasp-based tool changing?
- **Q2: Tool coverage and skill capability.** What range of tools and manipulation skills can the standardized interface support, and how reliably can these skills be executed?
- **Q3: Hierarchical Robustness.** Does the hierarchical tool–skill planning and evaluation architecture improve robustness in long-horizon multi-tool tasks?

A. Experimental Setup

System setup: We evaluate *Any-ttach* on a Franka Research 3 (FR3) 7-DoF robot arm and a tool dock for autonomous end-effector changing. The same coupling geometry is integrated into a handheld demonstration device used for human data collection.

Tool sets: We evaluate 15 tools across two tool sets in our experiments (Appendix B Fig. 4). (i) *Long-horizon task tool set:* six interchangeable modules used for autonomous multi-step tasks: gripper, spatula, spoon, peeler, knife, and fork. (ii) *Single-skill tool set:* additional tools used to evaluate tool coverage and single-skill capability, including daily tools (e.g., brush, pizza wheel, whisk, screwdriver) and unconventional end-effectors (e.g., Fin Ray gripper fingers, toy hand, and scissors). All can be coupled to both the robot end-effector and the handheld device through same interface.

Tasks: We evaluate *Any-ttach* through two representative full-system tasks (Appendix Fig. 5) and additional extended-tool demonstrations. The full-system tasks span different tool-use conditions. *Sandwich Assembly* represents object transfer and handling with both an active gripper and passive tools, requiring pick-and-place with the gripper, scooping with the spoon, and flipping with the spatula. *Cucumber Preparation* represents contact-rich object processing, requiring peeling with the peeler, cutting with the knife, and spearing with the fork. We further use the extended tool set to demonstrate broader tool coverage and single-skill capability.

Skill implementation: Each tool-use step is executed by a corresponding skill module with a defined input and output. For learning-based skills, the observations consist of the workspace RGB image, the 6-DoF tool-interface pose in the robot frame, and the binary gripper state; the output is a receding-horizon sequence of end-effector actions. The tool pose is computed from forward kinematics through the fixed transform defined by the coupling interface, giving each skill a consistent tool frame. For pose-driven skills, the input is a target end-effector pose or waypoint sequence, and the output is a collision-aware joint-space trajectory or scripted motion primitive. Tool coupling and decoupling follow predefined approach, docking, insertion/ release, and retreat waypoints relative to the tool dock during swapping.

B. Preliminary Experiment Results

We present preliminary results that evaluate the feasibility and performance of the proposed system. Additional experiment details are provided in the Appendix.

1) End-effector Swapping Performance: We evaluate the end-effector swapping subsystem from two perspectives: (i) autonomous swapping performance and (ii) effect on demonstration collection efficiency. Preliminary evaluation shows that *Any-ttach* improves autonomous tool-swapping success from 55.6% to 87.5% over grasp-based tool changing, while maintaining a comparable successful-trial swapping time (17.00 ± 3.61 s vs. 18.59 ± 2.94 s). For demonstration efficiency, human-assisted *Any-ttach* achieves 83.3% success in 11.09 ± 1.29 s, while direct human swapping achieves 100% success in 3.23 ± 0.27 s. These references suggest that the remaining autonomous overhead primarily arises from robot alignment and docking rather than the coupling interface itself. Detailed evaluation and failure analysis are provided in Appendix A.

2) Single Skill Capability: Compared with grasp-based tool holding, our preliminary experiments show that *Any-ttach* reduces attachment-pose variance by up to $5.1\times$ in position and $5.5\times$ in rotation. Using the same Diffusion Policy backbone with 30 demonstrations per skill, it improves the average success rate of sandwich-assembly skills from 44.4% to 71.1% over 15 evaluation trials and enables contact-rich peeling, cutting, and spearing skills that cannot be reliably demonstrated with the grasp-based baseline. Detailed protocols and every skill results are provided in Appendix B.

3) Hierarchical Robustness on Long-Horizon Tasks: On two long-horizon tasks, *Any-ttach* composes autonomous tool changes and tool-conditioned skills within a hierarchical execution pipeline. Because local errors can alter the scene state and propagate to subsequent steps, the system monitors each skill outcome and retries failed steps before proceeding. The results suggest that combining repeatable tool exchange with execution monitoring enables a simple gripper to perform and recover within multi-tool manipulation sequences. Detailed results are provided in Appendix C.

V. Conclusion and Future Work

We presented *Any-ttach*, a tool-centric manipulation system that treats quick end-effector swapping as a mechanism for composing diverse tool-use capabilities. Across tool swapping, demonstration collection, single-skill execution, and long-horizon tasks, our preliminary results show that mechanically constrained attachment reduces grasp-induced variability and enables more reliable tool-use behavior than grasp-based tool holding. More broadly, *Any-ttach* suggests that manipulation capability and dexterity does not need to be concentrated entirely in the robot hand. A shared interface can shift part of dexterity into interchangeable tools, allowing the robot to change not only what it grasps, but also the physical interface through which it acts on the world.

Future work will focus on improving the mechanical robustness of *Any-ttach*, automating its adaptation to new tools, and expanding its experimental validation. First, we will characterize the swapping mechanism under docking uncertainty and evaluate its load capacity, wear, and long-term durability. Second, although all adapters share a common template, each must currently be customized to the tool's attachment region. We plan to automate this process using SAM3 segmentation and 3D reconstruction within the adapter auto generation pipeline. Third, we will extend the quantitative evaluation from primarily 0-DoF tools to articulated 1-DoF tools. Finally, we will conduct further ablation studies to isolate the contributions of the planning, monitoring, and retry components.

In the long term, we hope *Any-ttach* can open a practical path toward dexterous manipulation systems, where shared end-effector interfaces and exchangeable sets can help robots move toward dexterity with simplicity: expanding what robots can do by adapting the tools they attach, exchange, and reuse as flexible as the policies that control them.

References

[1] H. Qi, A. Kumar, R. Calandra, Y. Ma, and J. Malik, "In-hand object rotation via rapid motor adaptation," in *Conference on Robot Learning*. PMLR, 2023, pp. 1722–1732.

[2] Y. Chen, C. Wang, Y. Yang, and C. K. Liu, "Object-centric dexterous manipulation from human motion data," *arXiv preprint arXiv:2411.04005*, 2024.

[3] R. Wang, J. Zhang, J. Chen, Y. Xu, P. Li, T. Liu, and H. Wang, "Dexgraspnet: A large-scale robotic dexterous grasp dataset for general objects based on simulation," *arXiv preprint arXiv:2210.02697*, 2022.

[4] K. Shaw, A. Agarwal, and D. Pathak, "Leap hand: Low-cost, efficient, and anthropomorphic hand for robot learning," *arXiv preprint arXiv:2309.06440*, 2023.

[5] M. W. Marzke, "Tool making, hand morphology and fossil hominins," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 368, no. 1630, p. 20120414, 2013.

[6] M. M. W, "Precision grips, hand morphology, and tools," *American Journal of Physical Anthropology*, vol. 102, no. 1, pp. 91–110, 1997.

[7] N. C. Dafle, A. Rodriguez, R. Paolini, B. Tang, S. S. Srinivasa, M. Erdmann, M. T. Mason, I. Lundberg, H. Staab, and T. Fuhlbrigge, "Extrinsic dexterity: In-hand manipulation with external forces," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 1578–1585.

[8] O. M. Andrychowicz, B. Baker, M. Chociej, R. Jozefowicz, B. McGrew, J. Pachocki, A. Petron, M. Plappert, G. Powell, A. Ray, *et al.*, "Learning dexterous in-hand manipulation," *The International Journal of Robotics Research*, vol. 39, no. 1, pp. 3–20, 2020.

[9] V. Makoviychuk, L. Wawrzyniak, Y. Guo, M. Lu, K. Storey, M. Macklin, D. Hoeller, N. Rudin, A. Allshire, A. Handa, *et al.*, "Isaac gym: High performance gpu-based physics simulation for robot learning," *arXiv preprint arXiv:2108.10470*, 2021.

[10] E. Todorov, T. Erez, and Y. Tassa, "Mujoco: A physics engine for model-based control," in *2012 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 2012, pp. 5026–5033.

[11] A. Agarwal, S. Uppal, K. Shaw, and D. Pathak, "Dexterous functional grasping," *arXiv preprint arXiv:2312.02975*, 2023.

[12] S.-M. Yang, M. Magnusson, J. A. Stork, and T. Stoyanov, "Learning extrinsic dexterity with parameterized manipulation primitives," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 5404–5410.

[13] N. Chavan-Dafle and A. Rodriguez, "Prehensile pushing: In-hand manipulation with push-primitives," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 6215–6222.

[14] M. Oller, D. Berenson, and N. Fazeli, "Tactile-driven non-prehensile object manipulation via extrinsic contact mode control." in *Robotics: Science and Systems*, 2024.

[15] W. Zhou and D. Held, "Learning to grasp the ungraspable with emergent extrinsic dexterity," 2022. [Online]. Available: https://arxiv.org/abs/2211.01500

[16] C. C. Kemp, A. Edsinger, and E. Torres-Jara, "Challenges for robot manipulation in human environments [grand challenges of robotics]," *IEEE Robotics & Automation Magazine*, vol. 14, no. 1, pp. 20–29, 2007.

[17] A. Z. Ren, B. Govil, T.-Y. Yang, K. R. Narasimhan, and A. Majumdar, "Leveraging language for accelerated learning of tool manipulation," in *Conference on Robot Learning*. PMLR, 2023, pp. 1531–1541.

[18] H.-S. Fang, M. Gou, C. Wang, and C. Lu, "Robust grasping across diverse sensor qualities: The graspnet-1billion dataset," *The International Journal of Robotics Research*, vol. 42, no. 12, pp. 1094–1103, 2023.

[19] H.-S. Fang, C. Wang, M. Gou, and C. Lu, "Graspnet-1billion: A large-scale benchmark for general object grasping," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 11 444–11 453.

[20] H. Chen, C. Zhu, Y. Li, and K. Driggs-Campbell, "Tool-as-interface: Learning robot policies from human tool usage through imitation learning," *arXiv e-prints*, pp. arXiv–2504, 2025.

[21] H. Shi, H. Xu, S. Clarke, Y. Li, and J. Wu, "Robocook: Long-horizon elasto-plastic object manipulation with diverse tools," *arXiv preprint arXiv:2306.14447*, 2023.

[22] A. Orsula, M. Geist, M. Olivares-Mendez, and C. Martinez, "Learning tool-aware adaptive compliant control for autonomous regolith excavation," *arXiv preprint arXiv:2509.05475*, 2025.

[23] Y. Dong, Y. Zhang, S. Calinon, and F. T. Pokorny, "Robustness-aware tool selection and manipulation planning with learned energy-informed guidance," *arXiv preprint arXiv:2506.03362*, 2025.

[24] M. Xu, P. Huang, W. Yu, S. Liu, X. Zhang, Y. Niu, T. Zhang, F. Xia, J. Tan, and D. Zhao, "Creative robot tool use with large language models," *arXiv preprint arXiv:2310.13065*, 2023.

[25] A. Bicchi, "Hands for dexterous manipulation and robust grasping: A difficult road toward simplicity," *IEEE Transactions on robotics and automation*, vol. 16, no. 6, pp. 652–662, 2000.

[26] M. Qin, J. Brawer, and B. Scassellati, "Robot tool use: A survey," *Frontiers in Robotics and AI*, vol. 9, p. 1009488, 2023.

[27] R. Holladay, T. Lozano-Pérez, and A. Rodriguez, "Force-and-motion constrained planning for tool use," in *2019 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2019, pp. 7409–7416.

[28] ATI Industrial Automation, "Robotic tool changers," https://www.ati-ia.com/products/toolchanger/robot_tool_changer.aspx.

[29] SCHUNK, "SWS quick change system," https://schunk.com/de/en/automation-technology/tool-changer/sws/c/PGR_1135.

[30] B. Dhakal, "Design of automatic tool changer for universal robots UR5," Master's thesis, Tampere University of Applied Sciences, 2019.

[31] D. Cheong, H. Park, and N. Kim, "Design and maneuver of a tool-changer using switchable magnet for a tool hung by a cable," in *IEEE International Conference on Automation Science and Engineering (CASE)*, 2024, pp. 1252–1257.

[32] H. Song, J. Hur, and S. Jeong, "Coaxial magnetic gear-based tool-changing system," *IEEE Access*, vol. 12, pp. 33 749–33 756, 2024.

[33] D. Mourtzis, J. Angelopoulos, M. Papadokostakis, and N. Panopoulos, "Design for 3D printing of a robotic arm tool changer under the framework of Industry 5.0," in *Procedia CIRP*, vol. 115, 2022, pp. 178–183.

[34] C. Chi, Z. Xu, C. Pan, *et al.*, "Universal manipulation interface: In-the-wild robot teaching without in-the-wild robot data," in *Robotics: Science and Systems (RSS)*, 2024.

[35] J. Kulozik and N. Jarrassé, "Evaluating the precision of the htc vive ultimate tracker with robotic and human movements under varied environmental conditions," *arXiv preprint arXiv:2409.01947*, 2024.

[36] A. Singh, A. Fry, A. Perelman, *et al.*, "Openai gpt-5 system card," 2025. [Online]. Available: https://arxiv.org/abs/2601.03267

[37] C. Chi, Z. Xu, S. Feng, E. Cousineau, Y. Du, B. Burchfiel, R. Tedrake, and S. Song, "Diffusion policy: Visuomotor policy learning via action diffusion," *The International Journal of Robotics Research*, vol. 44, no. 10-11, pp. 1684–1704, 2025.

[38] D. Coleman, I. A. Sucan, S. Chitta, and N. Correll, "Reducing the barrier to entry of complex robotic software: a MoveIt! case study," *Journal of Software Engineering for Robotics*, vol. 5, no. 1, pp. 3–16, 2014.

[39] N. Carion, L. Gustafson, Y.-T. Hu, S. Debnath, R. Hu, D. Suris, C. Ryali, K. V. Alwala, H. Khedr, A. Huang, *et al.*, "Sam 3: Segment anything with concepts," *arXiv preprint arXiv:2511.16719*, 2025.

APPENDIX

A. Evaluation on End-Effector Swapping Performance

We evaluate the end-effector swapping subsystem from two perspectives: (i) autonomous swapping performance and (ii) effect on demonstration collection efficiency.

Swapping effectiveness: *Any-ttach* improves swapping reliability compared with grasp-based tool changing. We evaluate swapping performance using swapping success rate and successful-trial swapping time. Swapping success measures whether the robot releases the current tool and attaches the target tool, while swapping time is computed only over successful trials. Compared with grasp-based tool changing, the autonomous *Any-ttach* interface increases swapping success from 55.6% to 87.5%, while maintaining a similar successful-trial swapping time (17.00 ± 3.61 s vs. 18.59 ± 2.94 s). Grasp-based tool changing must coordinate pose alignment, gripper open-close, and possible regrasping to reach a functional grasp, so failures often come from grasp misalignment or unstable adjustment. *Any-ttach* instead reduces tool changing to following a predefined docking trajectory into a mechanically constrained coupling, so failures mainly come from inaccurate arrival at the docking pose. Once coupled, the tool is mechanically locked in a fixed functional pose, while a grasped tool can still slip or rotate during later tool-use interactions. This reduces intermediate uncertainty and makes tool exchange more repeatable without increasing successful-trial swapping time.

We further compare against a human-assisted *Any-ttach* condition to separate the interface design from autonomous execution overhead. Human-assisted use achieves 83.3% success with a shorter swapping time of 11.09 ± 1.29 s, suggesting that the interface itself supports efficient tool exchange and that the remaining delay in autonomous swapping mainly comes from robot alignment and docking motions. Direct human hand swapping provides an upper-bound reference, with 100% success and 3.23 ± 0.27 s swapping time. Overall, these results show that *Any-ttach* improves swapping robustness over grasp-based tool changing while keeping successful swaps within a comparable time range.

Demonstration effectiveness: The standardized *Any-ttach* interface improves demonstration efficiency over gripper-based teleoperation, while direct handheld demonstration provides the fastest data collection mode overall. We evaluate demonstration efficiency using two metrics: average demonstration time per collected trial and usable demonstration rate, defined as the fraction of collected demonstrations that are successful and retained for policy training.

This suggests that the standardized interface makes robot-mediated demonstrations more repeatable by fixing the tool pose and reducing the grasp-dependent variability introduced by gripper-based tool acquisition. Direct handheld demonstrations further reduce the average collection time to 10.03 s and achieve a 100.00% usable rate, showing the benefit of removing robot teleoperation overhead while preserving the same tool coupling geometry used during deployment. The results indicate that the *Any-ttach* interface not only improves the efficiency of robot-mediated data collection relative to gripper-based teleoperation, but also narrows the gap between teleoperated and direct human demonstration, motivating its use as a practical demonstration modality for downstream policy learning.

B. Tool Coverage and Single-Skill Capability

Tool coverage: *Any-ttach* interface is designed to support interchangeable tools with diverse geometries and functional purposes. In our experiments, we attach multiple categories of tools to the standardized interface, including Fin Ray gripper fingers; kitchen utensils such as a spatula, spoon, fork, peeler, knife, brush, pizza wheel, and whisk; assembly tools such as a screwdriver; and unconventional end-effectors such as a low-cost anthropomorphic toy hand and scissors (Fig. 4). All of these tools can be coupled to both the handheld demonstration device and the robot end-effector using the same mechanical interface. This coverage shows that the interface is not tied to a single tool geometry or task family, but can support a broad range of manipulation functions through a unified coupling mechanism.

Single-skill capability: *Any-ttach* reduces attachment pose variance by up to $5.1\times$ in position and $5.5\times$ in rotation over a grasp-based baseline, where the same tools are held by the parallel gripper. This provides a more consistent tool frame for demonstration and deployment. To test whether this consistency improves execution, we train a policy for each skill with the same Diffusion Policy backbone and data collected under the two attachment strategies. Each skill is trained from 30 demonstrations and evaluated over 15 trials. We report *Deployment Success Rate*, defined as the fraction of trials in which a skill completes its goal. As shown in Table I, *Any-ttach* improves the average success rate of the sandwich assembly skills from 44.4% to 71.1%. It enables contact-rich cucumber preparation skills, peeling, cutting, and spearing, that cannot be reliably demonstrated and deployed with the grasp-based baseline.

Peeling and cutting require sustained tool–object contact. To approximate this behavior without explicit force control,

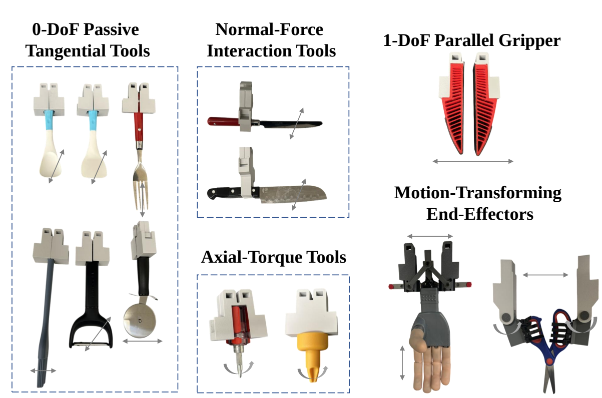

Fig. 4. **Tools covered.** The same coupling mechanism supports diverse tool categories, including passive kitchen tools, articulated tools, assembly tools, and unconventional end-effectors.

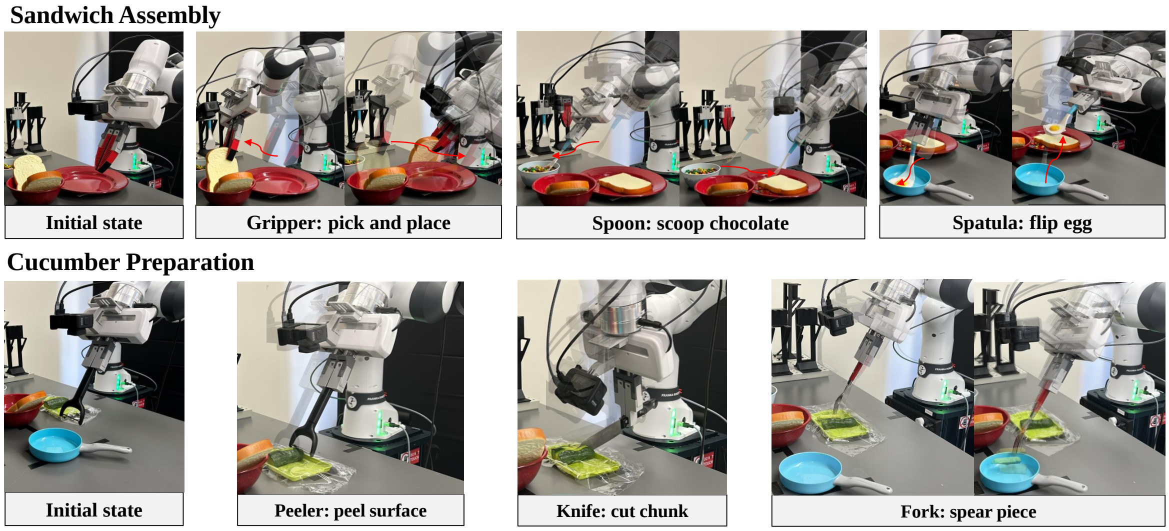

Fig. 5. We evaluate our system on two long-horizon tasks. **Sandwich Assembly:** the robot 1) picks and places bread, 2) scoops filling, 3) flips the fried egg onto the bread. **Cucumber Preparation:** the robot 1) uses the peeler to peel the cucumber, 2) cuts it in half with a knife, 3) uses a fork to spear the cucumber chunk into the pot.

TABLE I
SKILL SUCCESS RATE (15 TRIALS PER SKILL).
× INDICATES THE SKILL CANNOT BE EXECUTED WITH THE METHOD DUE TO UNACHIEVABLE DEMONSTRATIONS WITH THIS METHOD FOR THE SKILL, PREVENTING POLICY TRAINING AND DEPLOYMENT.

Skill	Gripper-tool	Any-ttach
Pick & Place	10/15	10/15
Flip	7/15	**12/15**
Scoop	6/15	**8/15**
Spread	12/15	13/15
Peel	×	**9/15**
Cut	×	**14/15**
Spear	×	**8/15**

we use a simple impedance-control heuristic implemented as a 5–7 mm positional offset along the tool normal. For these skills, success is defined as completing a continuous tool motion along the intended interaction direction, such as peeling from one end of the cucumber surface to the other or executing a full cutting stroke. Failures occur when the tool repeatedly slides on the surface without making progress.

Failure analysis: As shown in Fig. 6, contact-rich skills such as spatula flipping and fork spearing fail with grasp-based tool holding because interaction forces can cause tool slip or rotation within the gripper. By coupling the tool through a standardized interface, *Any-ttach* reduces grasp-dependent tool motion and improves repeatability under contact-rich interactions. For skills like brushing or scooping, the dominant forces are better supported by the gripper's closing direction and frictional contact, so the performance gap is smaller. This suggests that skills with off-axis tool loads are most revealing tests of tool-interface stability.

Single tool-use skill demonstrations: As a supplement, we test the same coupling interface on an extended tool set; videos of these demonstrations are available on the project website. With the same robot-side mechanism, *Any-ttach* can access single-skill behaviors such as brushing, rolling, whisking, screwing, alternative grasping, and articulated-tool cutting. Compared with grasp-based tool holding using a parallel gripper, these behaviors are difficult to realize reliably because the gripper must both secure the tool and maintain a functional tool pose under contact. Compared with complex anthropomorphic or multi-fingered hands, *Any-ttach* avoids placing all dexterity in the end-effector itself; instead, it exposes tool-specific capabilities through interchangeable modules with a simple robot-side coupling. These demonstrations serve as evidence of the interface's extensibility. The same design can be extended to additional tools through similar adapters while reusing the robot-side coupling and demonstration interface. This points to a practical route for expanding robot capability: instead of redesigning the end-effector for each new behavior, new capabilities can be introduced by adapting existing tools and end-effector modules that the robot can attach and use.

C. Hierarchical Robustness on Long-Horizon Tasks

We evaluate *Any-ttach* through two representative full-system tasks (Fig. 5), sandwich assembly and cucumber preparation. This experiment tests whether the system can complete a sequence of tool changes and tool-use skills in the planned order. Each full-task trial requires the robot to select the correct tool, attach it, execute the corresponding skill, monitor the outcome, and proceed to the next step. The results show that *Any-ttach* can complete multi-tool manipulation tasks by composing individual tool-use skills within the hierarchical execution pipeline. However, full-task success remains more challenging than single-skill execution because errors can accumulate across tool changes and skill boundaries. An incomplete scoop, inaccurate flip, failed peel, or unstable spear can alter the scene state for subsequent steps, making the next skill harder to execute. This highlights a key difficulty in long-horizon tool-use manipulation: the system must not only execute each skill, but also manage

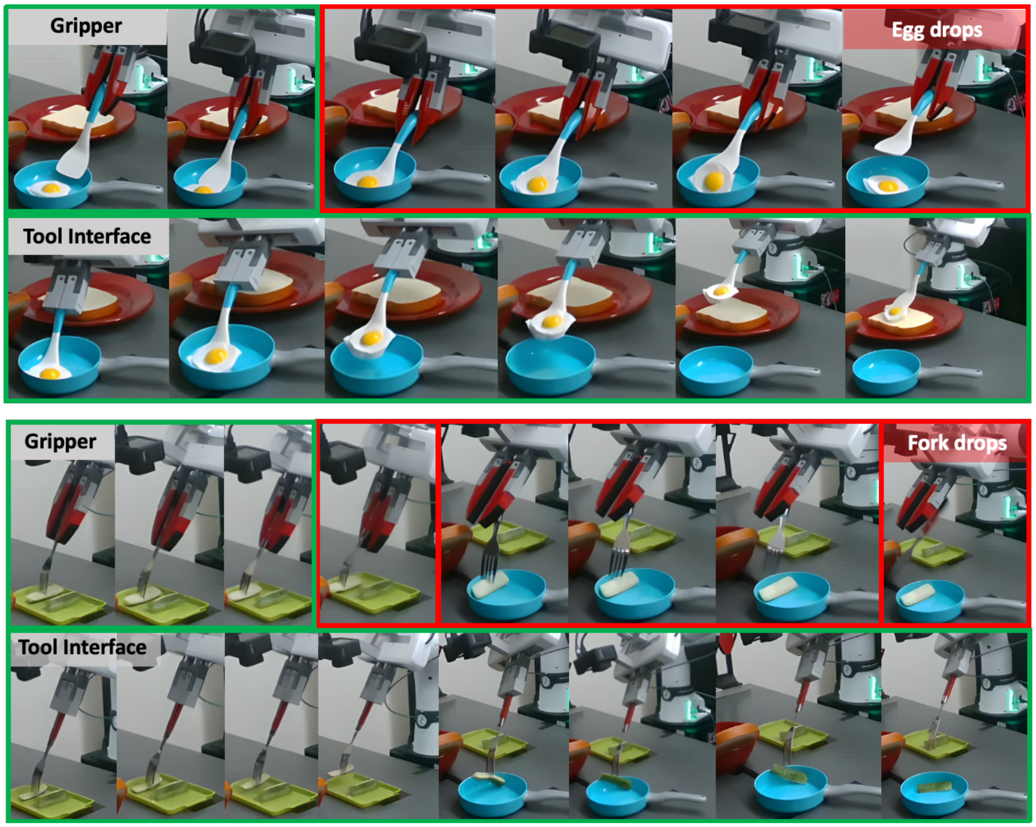

Fig. 6. **Gripper failure cases. Top:** during spatula flipping, contact forces induce tool rotation within the parallel-jaw gripper, causing the grasped tool pose to tilt and the egg to drop (red box). **Bottom:** during fork spearing, similar grasp-induced pose drift accumulates over the skill execution and leads to tool loss and task failure (red box). In contrast, our kinematically constrained tool interface maintains a fixed tool pose under the same interactions, enabling stable contact and successful completion (green boxes). These examples illustrate how grasp-based tool holding introduces a second unstable interface (manipulator–tool) whose errors accumulate under extrinsic, force-transmitting manipulation.

how intermediate errors affect downstream execution. Thus, simply composing individually trained skills is not sufficient for robust long-horizon behavior.

The execution monitoring module in *Any-ttach* mitigates this issue by checking the outcome of each skill before the system proceeds. When the intended outcome is not achieved, the system retries the same step instead of continuing with an incorrect scene state. This monitoring-and-retry mechanism improves task-level robustness by converting intermediate failures into recoverable events.

These results highlight that long-horizon tool-use manipulation is not only a problem of learning better individual skills. It also requires a reliable way to change the physical interface through which the robot acts on the world. In *Any-ttach*, repeatable tool attachment and autonomous end-effector switching allow a simple open-close actuation mechanism to access and compose diverse tool-use skills within a single task. Execution monitoring adds a layer of robustness that prevents some local failures from propagating into full task failures. This supports the central idea of *Any-ttach*: manipulation capability can be expanded not only through more complex end-effectors, but also through a shared interface that allows diverse tools and end-effectors to be attached, exchanged, and reused.