MagiClaw: A Dual-Use, Vision-Based Soft Gripper for Bridging the Human Demonstration to Robotic Deployment Gap

Tianyu Wu, Xudong Han, Haoran Sun, Zishang Zhang, Bangchao Huang
Design + Learning Research Group
Southern University of Science and Technology

Chaoyang Song
asRobotics
songcy@ieee.org

Fang Wan*
SUSTech
wanfang@ieee.org

Abstract:
The transfer of manipulation skills from human demonstration to robotic execution is often hindered by a "domain gap" in sensing and morphology. This paper introduces MagiClaw, a versatile two-finger end-effector designed to bridge this gap. MagiClaw functions interchangeably as both a handheld tool for intuitive data collection and a robotic end-effector for policy deployment, ensuring hardware consistency and reliability. Each finger incorporates a Soft Polyhedral Network (SPN) with an embedded camera, enabling vision-based estimation of 6-DoF forces and contact deformation. This proprioceptive data is fused with exteroceptive environmental sensing from an integrated iPhone, which provides 6D pose, RGB video, and LiDAR-based depth maps. Through a custom iOS application, MagiClaw streams synchronized, multi-modal data for real-time teleoperation, offline policy learning, and immersive control via mixed-reality interfaces. We demonstrate how this unified system architecture lowers the barrier to collecting high-fidelity, contact-rich datasets and accelerates the development of generalizable manipulation policies. Please refer to the iOS app at https://apps.apple.com/cn/app/magiclaw/id6661033548 for further details.

Keywords: Robot Learning from Demonstration, Vision-based Deformable Perception, Soft Robotics, Teleoperation

1 Introduction

The success of modern robot learning paradigms, from Learning from Demonstration (LfD) [1, 2] to offline reinforcement learning, is fundamentally dependent on the quality and richness of the underlying data [3]. For contact-rich manipulation tasks, robust policies require more than just kinematic trajectories; they demand a holistic understanding of interaction forces, tactile feedback, and environmental context [4, 5]. Consider a human deftly handling a delicate object: the action is a symphony of precise motion, modulated forces, and continuous tactile adjustments [6]. Replicating such skills requires capturing this multi-modal information stream in its entirety.

However, existing data collection methodologies present significant challenges. First, they often rely on a patchwork of disparate, expensive sensors—such as external motion capture systems, wrist-mounted force/torque sensors, and complex tactile skins [7, 8]—resulting in cumbersome and costly setups. This high barrier to entry limits the scale and diversity of data collection efforts [9]. Second, and more critically, a persistent **domain gap** exists between the human demonstrator and the robotic

9th Conference on Robot Learning (CoRL 2025), Seoul, Korea.

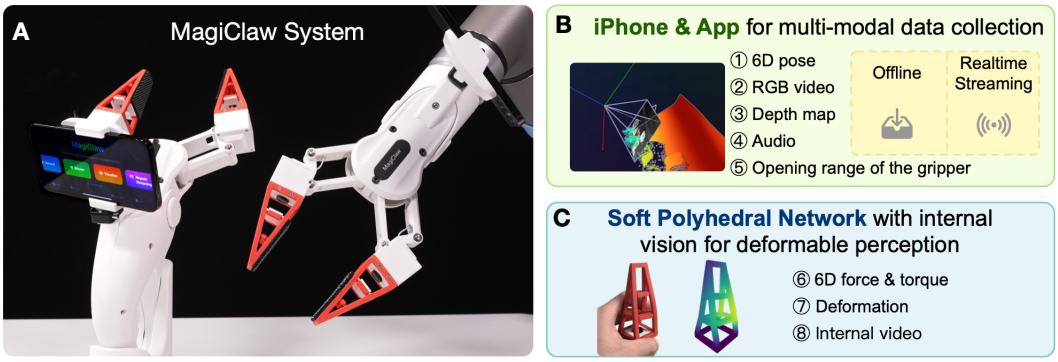

Figure 1: **Overview of the MagiClaw System.** (A) MagiClaw's dual-purpose design: a hand-held device for intuitive demonstration, data collection, and an identical end-effector mounted on a robotic arm for policy deployment. (B) An integrated iPhone provides exteroceptive sensing (6D pose, RGB, Depth) and a user interface for data management. (C) The Soft Polyhedral Network (SPN) fingertip contains an internal camera for vision-based proprioceptive sensing of 6D forces, torque, and local deformation.

learner [10]. Data is often collected using one set of hardware (e.g., an instrumented glove) and deployed on a robot with entirely different sensor suites and end-effector morphology. This mismatch necessitates complex domain adaptation techniques and is a primary reason why policies trained on demonstration data often fail to generalize to physical hardware [11].

To address these challenges, we present **MagiClaw**, a unified hardware platform designed to seamlessly bridge the gap from human demonstration to robotic deployment. MagiClaw is a dual-purpose, two-fingered gripper that merges three key innovations:

1. **Unified Hardware Form Factor:** The exact same MagiClaw device can be used as a hand-held tool for human demonstration or mounted on a robot arm for autonomous execution. This hardware consistency minimizes the sensor and morphological domain gap, facilitating direct policy transfer.
2. **Vision-Based Proprioceptive Fingertips:** Each finger integrates a *Soft Polyhedral Network* (SPN) [12] with an embedded miniature camera. This novel design enables *visuo-tactile perception*, inferring 6-DoF forces, torque, and high-resolution contact deformation from the distortion of the internal lattice structure, thereby obviating the need for costly external force sensors.
3. **Integrated Multi-Modal Exteroception:** An attached iPhone leverages its powerful sensor suite (LiDAR, RGB cameras, IMU) and ARKit framework [13] to provide synchronized, rich environmental context, including gripper pose, depth maps, and high-resolution video.

Our primary contribution is an integrated system that fundamentally streamlines the collection of holistic, contact-centric data for robot learning. By fusing proprioceptive force/tactile data from the fingertips with exteroceptive visual and spatial data from a commodity smartphone, MagiClaw offers a low-cost, powerful, and user-friendly solution for both teleoperation and autonomous policy development. We posit that by democratizing access to such high-fidelity, multi-modal data, MagiClaw can serve as a catalyst for developing more robust and generalizable manipulation skills, advancing the pursuit of universal action embodiment in robotics.

2 The MagiClaw Gripper System

MagiClaw is designed to bridge the gap between human demonstration and autonomous robotic manipulation through a *unified*, *dual-purpose* gripper system. As shown in Fig. 1, it can function

either as a **hand-held tool** for collecting multi-modal data in human-guided demonstrations or as a **robotic end-effector** mounted onto an industrial or collaborative robot arm. By sharing identical hardware and sensor layouts in these two modes, MagiClaw minimizes sensor disparities that often hinder the transfer of learned skills from humans to robots.

2.1 Engineering Highlights

Mechanical Design as a Dual-Purpose EOAgenT Fig. 1A shows the overall design, which is inspired by widely adopted industrial solutions such as OnRobot's RG2 gripper. However, we completely redesigned the entire mechanical system for use in robot learning scenarios, ensuring it is suitable for dual-purpose usage by both human operators and robotic arms.

The ***base design*** features a parallel four-bar gripper mechanism, actuated by a back-driable motor, with a detachable iPhone mount, and two omni-adaptive fingertips with an in-finger miniature camera capable of Vision-based Deformable Perception.

Two variations are currently available, including a *Hand-Held Mode* for data collection and an *End-Effector mode* for robotic manipulation, formulating an End-of-Arm-Agent (EOAgenT) system.

- **Hand-Held Mode**
  - **Ergonomic Handle and Trigger**: When used in *hand-held* mode (Fig. 1A), a molded handle accommodates the user's grip, and a trigger mechanism directly manipulates the finger openings. This setup enables an operator to perform everyday tasks, such as lifting, placing, sliding, or rotating objects, just as they would with a normal tool. Meanwhile, MagiClaw's onboard sensors continuously log force, pose, and environmental context without impeding the user's natural motions.
  - **Live Data Capture for Demonstrations**: Because the same system can later be mounted on a robot, the hand-held demonstration data (trajectories, forces, tactile events) directly translate into robotic replay or training sets for *imitation learning*. Operators can also leverage real-time visual or force feedback to refine their demonstrations in real-time. This approach significantly *lowers the entry barrier* to collecting rich multi-modal data, even outside specialized lab environments.
- **End-Effector Mode**
  - **Mounting and Interface**: In robotic deployments (Fig. 1B), the handle and trigger assembly can be detached, and the same mechanical finger unit is secured onto a standard robotic flange (e.g., an ISO 9409-1 mount). The iPhone remains attached to the gripper, maintaining the same sensing configurations. A single cable or wireless link connects the microcontroller to the robot's main controller, issuing commands that open/close or apply force.
  - **Closing the Demonstration-to-Deployment Loop**: This *dual-use design* is key to minimizing discrepancies between human-collected data and final robotic execution. By ensuring sensor placement, geometry, and compliance remain the same, MagiClaw helps learned policies replicate the human-demonstrated strategies more accurately. Tasks initially *taught* to the robot in hand-held mode—like gently grasping a delicate object or manipulating flexible packaging—can be re-executed on the robot with high fidelity since the gripper's mechanical and sensing characteristics are unchanged.

The entire design is 3D printable, offering a low-cost solution that can also be fabricated using metallic parts for enhanced reliability. We have open-sourced this design for the research community's use, with an iOS app available for free download at `https://apps.apple.com/cn/app/magiclaw/id6661033548`, along with accompanying documentation.

Parallel Four-Bar as the Driving Mechanism Although there is no universally "better" choice of design for the drive mechanism, the parallel four-bar design offers several key advantages that may be suitable for this dual-purpose application:

- **Field Use at a Low Cost**: Since its mechanism is driven, no rails are needed, which is great for small robots or end-effectors with tight mass budgets. The use of rolling joints means that it's less sensitive to dust, chips, or sprays than sliding guides, which is also 3D-printing-friendly.
- **A Big Stroke in a Thin Package**: The mechanism itself deploys and folds during usage, making it a more compact solution for field use while being capable of dealing with large-width objects during manipulation, covering most object sizes in daily life or even industrial scenarios.
- **Tunable Force Curve & Compliance**: The jaw speed and force changes with the angular input, meaning that the motion is an arc (the fingers' orientation stays parallel but their paths aren't perfectly straight), which provides a high mechanical advantage near full closure for strong holding with a small motor.

Backdrivable Actuation for Tunable Interaction Each finger is driven by a small motor with an encoder for accurate position feedback. The gearing ratio is tuned to deliver sufficient gripping force for everyday objects while preserving *back-drivability*, allowing the system to sense external contact forces and accommodate unmodeled variations. This design choice is especially helpful when switching between human-held demonstrations (where the user demands a responsive device) and robotic operation (where compliance prevents accidental damage to objects or the environment).

Adaptive Fingertips with Omni-directional Perception At the distal end of each four-bar linkage lies a *Soft Polyhedral Network* (SPN) [12], seen in Fig. 1C. These fingertips have a flexible lattice pattern (e.g., a TPU-based 3D-printed mesh) that can *conform in omni-directions* around diverse object profiles. Unlike silicone or purely elastic membranes, this lattice architecture provides both structural integrity and localized deformation nodes, thereby enhancing grip stability on irregular or deformable items.

Inside each fingertip, we embed a miniature camera (e.g., a wide-angle micro camera) that observes the lattice from within. As external forces act on the SPN, the lattice geometry distorts. By tracking the shifting pattern of these lattice elements in real-time, a lightweight neural network infers *6D force/torque* at the fingertip. Compared to conventional force sensors:

- **Low Cost**: The hardware cost is dominated by commodity micro cameras rather than specialized force-torque transducers.
- **High Spatial Resolution**: Deformation is recorded across the entire fingertip, reflecting *where* and *how* contact forces are applied.
- **Minimal Additional Bulk**: The sensing mechanism is entirely contained within the existing soft structure, maintaining a low overall profile.

Smartphone Sensing for Robotics A hallmark of MagiClaw is the integration of a consumer smartphone, specifically an iPhone Pro (although other smartphone brands or series may offer similar functionalities, depending on their hardware capabilities), into the gripper assembly (Figs. 1A&B). Currently, this smartphone offers the following capabilities. By leveraging off-the-shelf smartphone hardware, MagiClaw benefits from on-device processing capabilities, integrated communication (Wi-Fi, Bluetooth, cellular), and a user-friendly interface for real-time monitoring.

- **LiDAR Depth Sensing**: High-frame-rate 3D mapping of the environment, enabling real-time object detection, scene reconstruction, or augmented-reality overlays.
- **RGB Video**: High-resolution images or videos from the rear camera for visual context, teleoperation views, or training data.
- **Gripper Pose Tracking**: Using Apple's ARKit framework and provided APIs to track the gripper's orientation and movement relative to the global frame [13].
- **Audio Capture**: Potentially useful for event detection (e.g., collision sounds or object rattles).

Motor Drivers and Microcontroller Beyond the smartphone, a microcontroller (such as the Raspberry Pi 5) handles low-level tasks. This architecture isolates time-critical control from higher-level processes on the smartphone, maintaining robust performance despite the smartphone's variable compute load. It 1) receives setpoints (e.g., desired grip width) and commands the servo or DC motor drivers accordingly, 2) reads encoder values to relay finger positions and detect contact or stalling conditions, 3) measures interaction forces and transmits them to the handheld gripper, enabling haptic feedback for the user, and 4) coordinates timestamping of finger data with the smartphone's sensor streams.

System Communication Architecture The communication architecture of the MagiClaw system is illustrated in Fig. 2. An iPhone, mounted on the hand-held or motorized gripper, connects to a Wi-Fi router over a wireless network. Both the handheld and motorized grippers contain a motor, each wired to a Raspberry Pi. These two Raspberry Pi boards communicate with their respective motors via the CAN protocol.

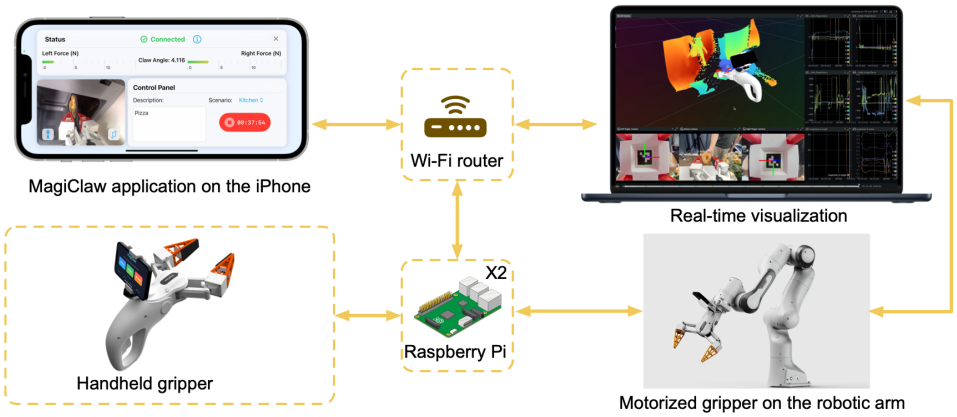

Figure 2: **Data visualization interface (via Rerun) and communication topology.** Multi-modal data streams (pose, depth, fingertip images, etc.) are synchronized across devices for real-time or offline analysis.

Each gripper also integrates two SPNs, each equipped with an internal camera. The cameras connect to the Raspberry Pi using Wi-Fi hotspots that they create themselves. The Raspberry Pis, in turn, communicate wirelessly with the central Wi-Fi router.

A computer running Rerun for real-time data visualization is also connected to the same router. It receives data streams from all devices in the local network. The MagiClaw app, running on the iPhone, serves a dual purpose. First, it displays 6D force/torque data from the SPN. Second, it broadcasts its own 6D pose, RGB images, and depth images, which are accessible to all devices within the local network. Additionally, the app offers direct control of the gripper motors, allowing for start and stop commands.

3 Experimental Validation: The Imitation Game

We validate the MagiClaw system through a series of use cases framed as *The Imitation Game*—a spectrum of tasks that demonstrate the system's capacity to capture holistic human actions and transfer them to a robot. These experiments serve to confirm the utility of our unified hardware and multi-modal sensing approach.

3.1 High-Fidelity Teleoperation and Immersive Demonstration

A primary use case for MagiClaw is real-time teleoperation, where a human operator's actions are mirrored by a robot-mounted gripper (Fig. 3). The operator uses a hand-held MagiClaw, and its

state (6D pose and grip width) is streamed to the robot. This setup validates several key system capabilities:

- **Intuitive Control and Data Capture:** The direct physical interface allows for natural and dexterous manipulation. Simultaneously, the system logs a complete, synchronized dataset of the operator's actions and the resulting environmental interactions.
- **Closed-Loop Force Feedback:** The vision-based force estimation from the robot's SPN fingertips is streamed back to the operator's hand-held device, providing haptic feedback. This allows the operator to "feel" the interaction forces, enabling delicate tasks that would be impossible with visual feedback alone.

To further enhance the operator's situational awareness, we integrate this system with an **Apple Vision Pro** mixed-reality headset (Fig. 3A). The headset provides a first-person view from the robot's perspective, overlaying real-time sensor data (e.g., force vectors, depth maps) to enhance situational awareness. This immersive interface significantly reduces the cognitive load on the operator, enabling the demonstration of highly precise and complex maneuvers. This capability validates our claim of creating a user-friendly and powerful interface for demonstration.

Figure 3: **MagiClaw in Action.** (A) An operator performs teleoperation with an immersive, first-person view provided by an Apple Vision Pro headset, which overlays real-time sensor data. (B) The system's versatility is demonstrated across various manipulation tasks, both in the real world and in simulation, showcasing its adaptability.

3.2 Learning from Multi-Modal Replays

The rich datasets collected during teleoperation or offline hand-held demonstrations (Fig. 4) form the foundation for policy learning. This workflow validates the core hypothesis that unified hardware reduces the domain gap. Please refer to the Supplementary Video for further demonstration.

- **Direct Policy Transfer:** Because the demonstration and deployment hardware are identical, simple behavioral cloning policies can be trained on the collected data and directly deployed on the robot with minimal performance degradation from sensor or kinematic mismatch.
- **Seeding for Advanced Learning Algorithms:** The multi-modal data is ideally suited for training more sophisticated models. For example, synchronized force and visual data can be utilized in offline reinforcement learning to learn reward functions that encourage gentle contact, or to train predictive models that anticipate contact events based on visual input.

The ability to replay demonstrations on the physical robot allows for iterative debugging and policy refinement. Discrepancies between the original demonstration and the robotic execution can be logged and used to further improve the learned model.

3.3 Validation in Contact-Rich Scenarios

We further validate MagiClaw's utility in advanced tasks where force and tactile feedback are critical. Please refer to the Supplementary Video for further demonstration.

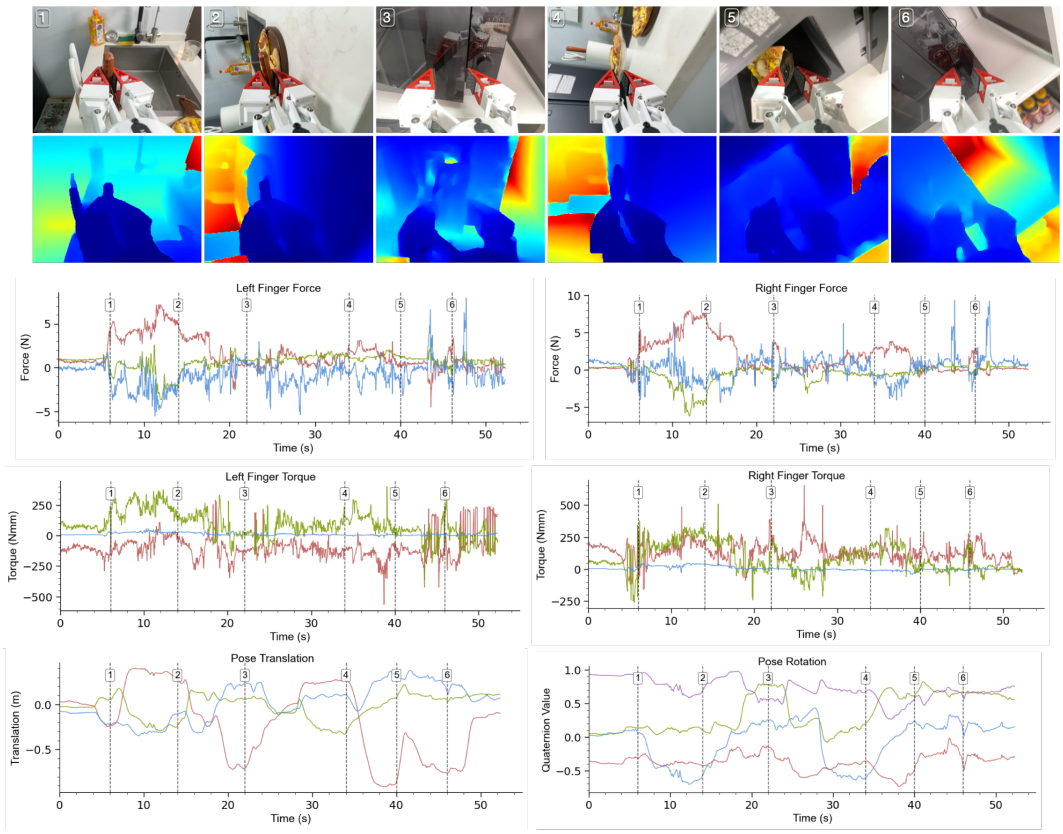

Figure 4: **Collected Multi-Modal Data.** Task of transferring a slice of pizza to a plate and heating it in the microwave, showing RGB and depth images alongside temporal variations of left and right fingertip 6D force/torque and MagiClaw gripper's 6D pose.

These scenarios confirm that the high-fidelity contact data captured by MagiClaw enables the learning of skills that are beyond the reach of systems relying solely on kinematic and visual information.

4 Conclusion

This paper introduced *MagiClaw*, a multi-modal gripper system designed to accelerate research in robot learning by directly addressing the demonstration-to-deployment gap. Its novel dual-purpose design, which unifies the hardware for data collection and policy execution, fundamentally minimizes domain shift. By integrating vision-based proprioceptive force sensing in soft fingertips with comprehensive exteroceptive sensing capabilities from a commodity smartphone, MagiClaw provides a low-cost yet powerful turnkey solution for generating rich, contact-centric datasets.

We have demonstrated how this integrated system enables high-fidelity teleoperation with immersive feedback, streamlines data collection for policy learning, and proves effective in challenging, contact-rich tasks. By simplifying and democratizing access to holistic action data, we believe MagiClaw will catalyze progress in data-driven robotics, paving the way for more dexterous, adaptable, and human-like manipulation.

Limitations and Future Work. The current system depends on low-latency wireless communication, which may bottleneck in congested networks. Vision-based force estimation is cost-effective but requires fingertip-specific calibration and training, suggesting room for streamlining. Reflective surfaces challenge LiDAR depth accuracy. As the iPhone is not a hard real-time system, iOS scheduling and thermal throttling limit the safety-critical control, making the integration of off-the-shelf cooling solutions a preferable option.

Our future work will focus on improving the system's robustness and expanding its capabilities. We plan to explore onboard policy learning directly on the integrated smartphone, investigate more sample-efficient calibration methods for the SPN fingertips, and develop a library of pre-trained models for common manipulation tasks. Crucially, we intend to open-source the hardware designs and core software modules to foster collaboration and empower the wider robotics research community to build upon our work.

Acknowledgments

This work was supported by Shenzhen AncoraSpring Robotics Technology Co., Ltd. and the National Natural Science Foundation of China (62206119 and 62473189).

References

[1] B. D. Argall, S. Chernova, M. Veloso, and B. Browning. A Survey of Robot Learning from Demonstration. *Robotics and Autonomous Systems*, 57(5):469–483, 2009.

[2] A. Billard, S. Calinon, R. Dillmann, and S. Schaal. *Robot Programming by Demonstration*, pages 1371–1394. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[3] A. Barekatain, H. Habibi, and H. Voos. A Practical Roadmap to Learning from Demonstration for Robotic Manipulators in Manufacturing. *Robotics*, 13(7), 2024.

[4] P. Jin, B. Huang, W. W. Lee, T. Li, and W. Yang. Visual-Force-Tactile Fusion for Gentle Intricate Insertion Tasks. *IEEE Robotics and Automation Letters*, 9(5):4830–4837, 2024.

[5] W. Liu, J. Wang, Y. Wang, W. Wang, and C. Lu. ForceMimic: Force-Centric Imitation Learning with Force-Motion Capture System for Contact-Rich Manipulation. *arXiv preprint arXiv:2410.07554*, 2024.

[6] A. Billard and D. Kragic. Trends and Challenges in Robot Manipulation. *Science*, 364(6446): eaat8414, 2019.

[7] M. H. Lee. Tactile Sensing: New Directions, New Challenges. *The International Journal of Robotics Research*, 19(7):636–643, 2000.

[8] M. Meribout, N. A. Takele, O. Derege, N. Rifiki, M. El Khalil, V. Tiwari, and J. Zhong. Tactile Sensors: A Review. *Measurement*, 238:115332, 2024.

[9] S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz, and D. Quillen. Learning Hand-Eye Coordination for Robotic Grasping with Deep Learning and Large-Scale Data Collection. *The International Journal of Robotics Research*, 37(4-5):421–436, 2018.

[10] H. Ravichandar, A. S. Polydoros, S. Chernova, and A. Billard. Recent Advances in Robot Learning from Demonstration. *Annual Review of Control, Robotics, and Autonomous Systems*, 3(1):297–330, 2020.

[11] M. Zare, P. M. Kebria, A. Khosravi, and S. Nahavandi. A Survey of Imitation Learning: Algorithms, Recent Developments, and Challenges. *IEEE Transactions on Cybernetics*, 54 (12):7173–7186, 2024.

[12] X. Liu, X. Han, W. Hong, F. Wan, and C. Song. Proprioceptive Learning with Soft Polyhedral Networks. *The International Journal of Robotics Research*, 43(12):1916–1935, 2024.

[13] A. Inc. Arkit — apple developer documentation, 2025. URL `https://developer.apple.com/documentation/arkit/`.