

Tactile-Based Robotic Manipulation: From Mechanics-Driven to Data-Driven Perspectives

Jingyi Xiang, Yixuan Wang, and Wenzhen Yuan

Siebel School of Computing and Data Science, University of Illinois Urbana-Champaign,
Urbana, Illinois, USA; email: jingyx4@illinois.edu, yixuan41@illinois.edu, yuanwz@illinois.edu

Annu. Rev. Control Robot. Auton. Syst. 2027.
10:3.1–3.27

The *Annual Review of Control, Robotics, and
Autonomous Systems* is online at
control.annualreviews.org

<https://doi.org/10.1146/annurev-control-032125-030330>

Copyright © 2027 by the author(s).
All rights reserved

Keywords

manipulation, tactile sensing, robot learning, visual–tactile learning

Abstract

Tactile sensing provides robots with direct access to physical interactions at the contact interface, making it a critical modality for robust and dexterous manipulation. Over the past several decades, tactile-based manipulation research has evolved from mechanics-driven approaches, which explicitly interpret tactile signals through contact mechanics and task-specific reasoning, to data-driven approaches that learn tactile representations and manipulation policies directly from data. This article reviews the development of tactile-based manipulation from these two complementary perspectives. We first introduce the fundamentals of tactile sensing and then review, from a mechanics-driven perspective, how those signals form manipulation-relevant state representations and subsequently influence the robot action. We then examine the recent advances in data-driven tactile manipulation, covering imitation learning, reinforcement learning, vision–language–action models, and predictive models. Finally, we discuss open challenges and future opportunities toward scalable, generalizable, and robust tactile-enabled robotic systems.

1. INTRODUCTION

Manipulation is the process by which a robot physically interacts with and changes the state of objects in its environment. Because these interactions occur through contact, many manipulation-relevant quantities—including contact state, force distribution, slip, and local object geometry—can only be observed at the contact interface. Tactile sensing provides direct measurements of these interactions and therefore plays a critical role in enabling robust and dexterous manipulation.

Research on robotic tactile sensing and manipulation has existed for decades. A central question throughout this history has been how tactile signals can be transformed into actionable information for robot control. Early research largely approached this problem from a mechanics-driven perspective, seeking explicit connections between tactile measurements, descriptions of physical interactions, and manipulation actions. For example, tactile feedback can be used to detect slip so as to secure a grasp, guide insertion and assembly based on force feedback, or measure local geometry for contact alignment. Although different approaches have been proposed—including analytical models, handcrafted feature extraction, and supervised learning techniques—the common principle is the use of explicit knowledge about contact mechanics and task structure to interpret tactile signals and guide robot behavior. In this review, we collectively refer to these approaches as mechanics-driven tactile-based manipulation.

In recent years, the rapid progress of data-driven robot learning has profoundly reshaped tactile-based manipulation research. In contrast to mechanics-driven approaches, these methods typically do not require explicit modeling of intermediate physical quantities. Instead, they learn task-relevant implicit representations and action policies directly from data, often through imitation learning (IL), reinforcement learning (RL), or large-scale multimodal pretraining. This trend emerged alongside the remarkable success of vision-based robot learning, where end-to-end learning frameworks have demonstrated impressive performance across a wide range of manipulation tasks. Consequently, a major research direction in tactile-based manipulation has become the integration of tactile observations into learning architectures originally developed for vision. This direction is particularly promising because vision and touch provide highly complementary information: Vision captures global scene context and long-range spatial relationships, while tactile sensing provides precise local measurements of contact and physical interaction. Recent advances in sensor development, scalable robot learning, and multimodal foundation models have further accelerated this trend, creating new opportunities for tactile-enabled manipulation while also introducing new challenges.

In this review, we examine tactile-based manipulation from both mechanics-driven and data-driven perspectives. We first discuss how tactile sensing captures the physics of contact and how tactile signals can be transformed into manipulation-relevant information through explicit reasoning. We then review recent advances in learning-based manipulation, including IL, RL, vision-language-action (VLA) models, and predictive models that leverage tactile observations as part of large-scale multimodal learning frameworks. By bringing together these two perspectives, we aim to provide a unified understanding of how tactile information contributes to robotic manipulation, identify the connections between classical and modern approaches, and highlight the key challenges and opportunities that will shape the future of tactile-enabled robotics. Owing to the breadth of the field, we emphasize representative works that illustrate major ideas and trends rather than providing an exhaustive catalog of existing research.

2. TACTILE SENSING FUNDAMENTALS

2.1. Tactile Sensors and Signals

The first step toward tactile-based manipulation is the tactile sensors. Tactile sensors measure touch, in the form of either (a) single-point force/torque (F/T) sensing or (b) distributed sensing that provides information across a fingertip or an area of skin. Commonly seen output signals from tactile sensors include the following:

- Normal force, which is the component of contact force perpendicular to the sensor surface. This is the most commonly measured tactile quantity.
- Shear force, which corresponds to tangential loads at the contact interface and plays a critical role in grasping and in-hand manipulation. The capability to measure shear force is much less common, but some modern tactile sensors, including magnetic (1, 2) and vision-based designs (3, 4), can estimate local shear distributions.
- Torque, which is available for F/T sensors. For distributed tactile sensors, torque can be inferred from the spatial distribution of forces.
- Contact geometry, which is typically achieved from the spatial distribution of the deformation or measured normal force across the sensing surface.
- Vibration in tactile signals, which contains important information about contact dynamics and is typically generated during sliding motions. In addition to extracting temporal features of tactile signals, some tactile systems explicitly incorporate accelerometers or vibration sensors to capture high-frequency signals.
- Auxiliary modalities, such as temperature and proximity. These are less commonly used in robotic manipulation and are therefore outside the scope of this review.

Over the past few decades, a large number of tactile sensors have been developed (5). They vary in characteristics such as sensitivity, spatial resolution, force range, bandwidth, robustness, and manufacturing complexity. One rough way to categorize the sensors is based on the number of their measurement units:

- F/T sensors have only one measuring unit but can typically measure six- or three-axis forces and torques. They are installed at the wrist or fingertip and can usually measure forces much more precisely than other types of sensors.
- Array-based tactile sensors typically contain 4–100 sensing units to measure normal force with a spatial resolution of several millimeters. A rough shape estimate is available, but not precise shapes or textures.
- Vision-based tactile sensors use an embedded camera to capture raw images, so the spatial resolution can be very high. The downside is their bulkiness. Some sensors in this category are designed to measure the surface geometry of every image pixel (4, 6), while others track the motion of the printed markers (3, 7), so their spatial resolution depends on the number of markers.

One major challenge in tactile manipulation is the strong dependency on sensor hardware. The quality and availability of tactile signals vary substantially across sensors, often requiring sensor-specific denoising and preprocessing methods, as well as training data when using machine learning methods. As a result, algorithms and learning models developed for one sensor frequently transfer poorly to another, and the performance of one manipulation system depends greatly on the choice of sensors (an issue that we discuss further in Section 5.1). Throughout this review, we therefore focus on reviewing manipulation methodologies based on how they are built on tactile signals rather than individual sensors.

To abstract out the implementation differences, it is useful to adopt a unified physical view of tactile sensing. Tactile sensors consist of two components: a deformable contact medium, typically made of compliant materials such as elastomers or structures filled with air or liquid, and a transducer that converts the deformation of the medium into measurable electrical signals. Transduction mechanisms may be resistive, capacitive, magnetic, piezoelectric, or optical. Based on this description, tactile sensing is fundamentally indirect: Rather than measuring contact itself, sensors measure contact-induced deformation so as to infer information about the interaction. Consequently, the differences among tactile sensors can be characterized by the design of the contact medium and the choice of transduction mechanism.

2.2. Tactile Simulation and Closing Sim-to-Real Gaps

The growing adoption of data-driven manipulation methods has created a strong demand for tactile sensor simulation. Modern learning-based approaches often require large amounts of training data, while collecting tactile data using physical robots is costly and difficult to scale. Simulation, therefore, provides an attractive alternative for generating large datasets and training manipulation models in a safe and highly parallelized manner. It also provides a good platform for quickly evaluating the design of robotic systems at low cost.

Accurately reproducing tactile sensor outputs while maintaining high simulation speed remains a long-standing challenge. While existing technologies cannot perfectly reproduce real sensor responses, improving simulation fidelity also comes at the cost of computational complexity. Since simulation efficiency is also critical for large-scale data generation, practical tactile simulators must balance accuracy and speed. Accordingly, current efforts in tactile simulation generally follow two complementary directions: developing more accurate and efficient sensor models and designing algorithms to mitigate the sim-to-real gap.

Unlike conventional robot simulation, tactile simulation must model both the contact mechanics between objects and the sensor and the sensor transduction process that converts deformation into measurable signals. Most robotic simulators employ simplified contact models that cannot fully capture the fine-grained contact phenomena required for tactile sensing, while sensor noise, calibration errors, and manufacturing variations further complicate accurate simulation. Existing tactile simulators roughly range from physics-based methods (8–11) to data-driven (12–14) and hybrid approaches (15, 16), each offering different trade-offs among fidelity, computational cost, and adaptability to individual sensors. In some applications, particularly those that rely primarily on geometric information, the sim-to-real gap can be sufficiently small to permit zero-shot transfer or transfer with only minor calibration (17, 18). However, tasks involving force estimation, frictional interactions, slip detection, or dynamic contact typically require substantially more accurate modeling.

To mitigate the sim-to-real gap, one common strategy is to reduce the dimensionality of the tactile representation. Rather than relying on high-resolution raw signals, many manipulation models take the input of lower-dimensional contact features such as binary contact states, contact locations, estimated contact geometry, or other task-specific abstractions (19–23). These representations are often easier to simulate consistently and therefore transfer more reliably to real systems. However, this simplification may discard fine-grained tactile information that could otherwise improve manipulation performance.

When higher-fidelity tactile representations are required, researchers have explored techniques to explicitly bridge the domain gap. One approach is domain randomization, where sensor parameters, contact properties, and measurement noise are randomized during training to improve robustness to simulation errors (24–26). Another approach is domain adaptation, which employs transfer learning techniques or generative models to align simulated and real

tactile observations (27–31). Despite these advances, reliable sim-to-real transfer remains an open challenge for tactile-based manipulation.

3. TACTILE-BASED MANIPULATION: A MECHANICS-DRIVEN PERSPECTIVE

A classical paradigm in tactile-based manipulation is to explicitly model the role that tactile sensing plays in the manipulation process. Under this view, tactile information is incorporated through a pipeline of perception $\rightarrow$ action. The perception module transforms raw tactile measurements into interpretable descriptions of the contact state that reveal task-relevant information. As a result, it defines what aspects of the physical interaction become observable to the robot and can therefore be exploited for manipulation. The action module then utilizes these perceptual estimates to guide planning, decision-making, and control.

A defining characteristic of this perspective is the use of explicit and human-defined representations that connect tactile measurements to robot actions. Although researchers have employed analytical models, handcrafted features, and supervised learning techniques, they share a common philosophy: Tactile signals are first interpreted through explicit reasoning based on human prior knowledge about contact and interaction before being used for manipulation.

3.1. Perception

A perception module takes the raw tactile signals (see Section 2.1) as input and outputs a perceptual descriptor that is sensor-agnostic and tied to some particular physical meaning. This general perceptual descriptor, however, can reveal rich information about the state of the manipulation process or the object's properties when interpreted in context. Depending on the sensor choices and system design preferences, the perception module can be a straight-in, straight-out function; an analytical modeling process; or a machine learning model. We note that tactile perception more broadly encompasses the general understanding of objects and environments. In this review, however, we restrict our focus to perceptual quantities that are directly relevant to manipulation. We introduce the commonly used perceptual descriptors below.

3.1.1. Contact force/torque. Force and torque measurements are among the most fundamental outputs of tactile and haptic sensors (Section 2.1). Although they are low-dimensional, they often provide rich information about the state of physical interaction. Force and torque values may be obtained directly from F/T sensors or estimated from raw tactile observations. Importantly, these measurements can characterize both the direct contact between a robot and an object and indirect interactions occurring elsewhere in the manipulation chain. As a result, force and torque measurements have been widely used to infer task-relevant information and guide robot actions. Examples include the following:

- Maintaining proper contact and grasp stability: Force feedback can be used to regulate contact force, preventing object slip while avoiding excessive force (32). In some tasks, controlling force also enables sliding motions along an object's surface (33).
- Manipulating articulated objects: Articulated objects, such as doors, drawers, or mechanisms with constrained motion, typically move most easily along specific directions. By monitoring interaction forces, robots can identify directions associated with minimal resistance and follow the underlying kinematic constraints of the object (34).
- Indicating that a task stage limit has been reached: When a robot reaches the end of a manipulation limit—say, after fully opening or closing a door, or after reaching the bottom of a hole during peg-in-hole insertion—a large impedance force can be detected (35).

- **Inferring extrinsic contact states:** Force measurements can reveal contact events between the objects in the grip and the environment. This can be used to infer task-specific state information, such as whether a cable has been successfully routed around a pillar (36). A classic example is peg-in-hole insertion, where interaction forces generated by contact between the peg and the hole can indicate alignment errors (37). Note that in some cases, extrinsic contact states are more precisely estimated from higher-dimensional tactile signals (discussed further in Section 3.1.6).
- **Performing tool-mediated manipulation:** When a robot indirectly manipulates objects through a tool, force and torque measurements provide important clues about the manipulation state (38) and even the state of the objects being in contact (39).
- **Handing over objects:** During human–robot handover, characteristic changes in a measured force often indicate that the receiving person has established a stable grasp. Robots can use these cues to determine when it is safe to release the object (40).

Notably, in many cases, the precision of the force and torque measurements is not essential for perception purposes. For example, for peg-in-hole insertion, even a very imprecise measurement of the direction of the contact force and torque can successfully guide the robot to adjust the motion to the correct directions (41).

3.1.2. Shape in contact. Contact geometry is among the basic outputs of many tactile sensors. While a large body of geometry-based tactile perception research focuses on object or material recognition, the role of shape measurement in manipulation is often about verifying whether the correct object or object part has been grasped (42, 43) or estimating the contact location on the objects when the objects’ shapes are known (17, 44). But more commonly, tactile shape measurements are indirectly used for in-hand localization (details in Section 3.1.3) and to estimate the pose of manipulated objects (details in Section 3.1.4).

3.1.3. Contact location. Based on tactile geometry measurements, a robot can estimate where it is being touched—either the location of contact on its body or the position of a grasped object within the gripper. For humanoid robots equipped with whole-body tactile skins, contact localization, combined with force measurements, can help detect unexpected interactions and support impedance control (45). For robotic grippers with fingertip sensors, contact localization can reveal undesirable grasp configurations, such as contacts near the finger edge, allowing the robot to adjust the grasp to enlarge the safety margin (33).

3.1.4. Object pose. By matching measured contact geometry with a known object shape, a robot can estimate the relative pose between the object and the gripper. Combined with robot kinematics, this estimate can be further transformed to the object’s pose in the world frame. Tactile pose estimation may target either the global pose of the object (44) or the pose of local geometric features, such as edges and corners (46). Accurate pose estimation enables manipulation tasks such as the following:

- **Precise manipulation:** Improved pose estimates support accurate object placement (47, 48), and the tactile-based pose estimates can be used as a major reference for repositioning objects in a hand (49–51).
- **Edge following:** The local orientation of edge features provides important guidance for robots when they need to move along an edge (52) or along deformable objects such as cables or cloth (33, 36, 42).

3.1.5. Slip. Slip refers to relative motion between an object and the robot at the contact interface. During grasping, both translational and rotational slip are major causes of grasp failure. To

prevent failure before it occurs, robots often seek to detect incipient slip, a transitional state in which partial detachment occurs at the contact surface before global slip develops. Tactile sensing is particularly well-suited for detecting slip and incipient slip.

From a mechanics perspective, slip can be predicted by comparing normal and tangential forces according to Coulomb's friction law (53). In practice, however, precise force measurements and friction coefficients are often unavailable. Instead, researchers typically infer slip from vibration signals (54–56), object micromotions (57), or local deformation and stretching of the contact surface (58–61). In addition, learning-based methods have been developed to directly predict slip or incipient slip from tactile observations (62–65).

In most manipulation tasks, slip serves as a warning signal of grasp instability, prompting the robots to prevent it (see Section 3.2.2). However, some tasks intentionally exploit controlled slip, such as sliding along an object's surface to achieve stable contact transitions during manipulation (33, 49, 51).

3.1.6. Extrinsic contact state. Beyond perceiving contacts directly occurring at the sensor surface, tactile sensing can be used to infer interactions between a grasped object and the external environment. This capability enables robots to reason about contacts that occur away from the sensing location, which is particularly useful in tool use and indirect manipulation tasks. We discussed simple examples based on force and torque measurements in Section 3.1.1; more generally, robots may seek to estimate the location, force, or mode of external contacts from tactile observations. Mechanics-driven approaches typically combine tactile measurements with estimates of object pose, geometry, robot kinematics, and contact mechanics to infer the external contact state (66–68). While effective, these methods often rely on accurate object models and precise physical reasoning, which can limit their scalability to diverse objects and tasks. To reduce this dependence on explicit modeling, researchers have also explored learning-based approaches that directly infer external contact states from tactile observations (69).

3.1.7. Object properties. Tactile sensing can be used to infer object properties that are important for manipulation. Such information can help robots select appropriate manipulation strategies and tune low-level controllers to precisely manipulate objects with unknown or challenging properties (discussed further in Section 3.2.3).

Although tactile object perception is itself a broad research topic, several properties are particularly relevant to manipulation, including center of mass (70–72), hardness (73, 74), surface roughness, friction, and material-specific properties such as liquid viscosity (75, 76). In some cases, these properties are not explicitly estimated but are instead implicitly encoded in learned manipulation policies (77, 78). Estimating such properties often requires interactive perception, where the robot actively probes an object and interprets the resulting tactile responses. Consequently, perception and action are often tightly coupled, with exploratory motions serving as an integral part of the perception process.

3.2. Action

The action module takes in the perceptual descriptor from the perception module and outputs actions executable by a robot. Depending on the task, the resulting actions may range from short-horizon control commands to longer-horizon motion sequences. While the underlying control and decision-making frameworks are often similar to those used in conventional robotic manipulation, tactile feedback provides additional information about the ongoing physical interaction and enables robots to react to contact events in real time. In this subsection, we review the most common strategies for incorporating tactile perception into robot actions.

3.2.1. Tactile-driven closed-loop control. For many contact-rich manipulation tasks, robots must react to tactile feedback in real time. A common strategy is to incorporate tactile perceptual descriptors into a closed-loop control framework, where the current contact state is represented by tactile descriptors and compared with a desired state. The resulting error is then used to generate corrective actions, enabling real-time regulation of contact interactions.

Li et al. (79) introduced a representative formulation of this idea. Using an array-based tactile sensor, the authors represented the contact state using contact location, object orientation, and normal force. Actions in the form of the robot's joint motion command were generated to reduce the error between the current and target tactile states, where inverse kinematics mapped desired tactile changes onto end-effector motions. This tactile servoing framework enabled the robot to maintain desired contact states and track object motion during interaction.

By appropriately defining the tactile state, reference state, and control algorithms, this paradigm can be applied to a wide range of manipulation tasks based on different choices of tactile descriptors, including the following:

- Obstacle avoidance (80) based on contact location (Section 3.1.3)
- Peg-in-hole insertion (37) based on contact force (Section 3.1.1)
- Edge and surface following, including cable manipulation (33, 36, 42, 52, 81), based on contact location (Section 3.1.3)
- Manipulation of articulated objects (34) based on contact force (Section 3.1.1)
- Object reorientation within the gripper (49, 51) based on object pose (Section 3.1.4)
- Object rotation with multi-degree-of-freedom (multi-DOF) contact control (82) based on contact force (Section 3.1.1) and contact location (Section 3.1.3)

The perception-to-action mapping can also be realized implicitly. For example, in RL, tactile perceptual signals may be used to define reward functions. The learned RL policy is then optimized to drive the robot motion toward desirable tactile states. For instance, Zheng et al. (83) trained an RL-based policy for robotic page flipping using rewards derived from the similarity between observed tactile signals and human expert demonstrations, eliminating the need for a hand-designed tactile controller.

3.2.2. State transition based on tactile perception. Unlike closed-loop tactile control, where actions are continuously adjusted according to tactile feedback, some manipulation tasks use tactile perception to estimate the current manipulation state and trigger transitions to appropriate actions. In this paradigm, tactile information serves as a decision variable that determines how the robot should proceed in the task.

Grasp stabilization and regrasping are common examples. While tactile sensing is widely used for detecting slip and grasp failure (Section 3.1.5), many systems employ the feedback to initiate corrective actions. The most common strategy is to increase the gripping force when slip is detected (32, 49, 57, 65, 84–86). More sophisticated approaches adjust the grasp configuration itself. For example, when rotational slip or excessive torque indicates an unstable grasp, the robot may reposition its contact location toward the object's center of mass (87). Learning-based approaches have also been developed to predict improved grasp configurations directly from tactile observations (88, 89).

A related application arises in long-horizon manipulation tasks that consist of multiple stages. In such settings, tactile perception can be used to detect the completion of a subtask or identify the current interaction state. The manipulation planner then transitions to the next stage, potentially invoking a different motion trajectory, controller, or manipulation strategy (35, 36).

3.2.3. Perception to update low-level control strategies. Many manipulation tasks involve objects with unknown or challenging physical properties, where the effectiveness of a controller depends strongly on properties such as mass, center of mass, friction, stiffness, or viscosity. In these cases, tactile perception can be used to estimate object properties and subsequently adapt controller parameters or control strategies. We discussed tactile-based property perception in Section 3.1.7, the results of which can be inserted directly into this framework. However, in some more challenging cases, the perceived information may be represented implicitly as a latent feature that captures the aspects of object dynamics most relevant to control. Although these representations may not have a direct physical interpretation, they often encode information similar to physical properties and serve the same purpose of informing controller design.

One example is the dynamic manipulation of objects with unknown weight distributions. In Reference 77, a robot first performs exploratory actions to collect tactile observations and infer a latent representation of a rod's physical properties. This representation is then used to select controller parameters for swinging the rod to a desired angle. Similar ideas have been applied to manipulating objects with challenging material properties. For example, Luo et al. (76) and Huang et al. (78) used tactile exploration to estimate properties of non-Newtonian fluids such as sauces and batter, allowing the robot to adjust control parameters for dispensing motion and achieve precise and consistent pouring behavior. Zhang et al. (90) inferred food material with properties regarding softness or skin mechanics from contact vibration and F/T feedback, then set the stroke-amplitude and downward-displacement parameters of a slicing primitive accordingly.

3.3. Discussion

In this section, we reviewed the mechanics-driven perspective of tactile-based manipulation, organized through a perception → action framework. The performance of those methods varies and depends heavily on the choice of sensor hardware, the application conditions, and the quality of the algorithm design. Their main strengths lie in their interpretability, reliability, and efficiency: Because the underlying contact representations and reasoning processes are explicitly modeled, it is often possible to understand when and why a method would succeed or fail. Many of those methods are based on analytical models or lightweight processing modules, so they can quickly process tactile data to meet the requirements of low-latency feedback control. Some methods exhibit strong generalization across objects and tasks when the underlying assumptions about contact interactions remain valid. However, their performance can degrade when contact interactions become highly complex or difficult to model, especially in real-world settings. These limitations have partially motivated the growing interest in data-driven methods.

The reviewed works also demonstrated fundamental properties of tactile sensing. First, tactile sensing provides local and precise measurements of contact, making it particularly effective for monitoring manipulation state and supporting high-frequency feedback control. Second, tactile perception is inherently active and interactive. Unlike passive sensing modalities such as vision, tactile information is acquired through physical interaction and often evolves as the robot acts on the environment. As a result, perception and action are frequently coupled, with exploratory motions serving not only manipulation objectives but also perception objectives.

Finally, effective manipulation often requires the integration of multiple sensing modalities. Vision and touch provide highly complementary information. In mechanics-driven methods, these modalities are typically integrated through explicitly designed representations and reasoning pipelines. As we discuss in the next section, recent data-driven approaches increasingly seek to learn such multimodal representations directly from data.

4. TACTILE-BASED MANIPULATION: A DATA-DRIVEN PERSPECTIVE

Data-driven methods learn policies—functions that map observations directly to actions—from data. This is in contrast to mechanics-driven approaches, where the perception-to-action process is explicitly modeled. Driven by advances in robot learning, computational resources, and large-scale data collection, this paradigm has achieved remarkable success in vision-based manipulation and has been increasingly adopted in tactile manipulation. Beyond applying existing learning frameworks to tactile inputs, recent research has focused on integrating vision and touch to improve performance in contact-rich tasks. In this section, we review major data-driven paradigms for tactile manipulation, including IL, RL, VLA systems, and predictive models, with a focus on how tactile information is utilized within these frameworks.

4.1. Tactile-Augmented Imitation Learning

IL, a supervised learning approach in which the policy learns to imitate the actions an expert demonstrator performed in similar observed situations, has emerged as a promising direction for learning visuomotor control. Popularized by methods like Diffusion Policy (91) that generate a sequence of future actions conditioned on current observations, IL has demonstrated remarkable results in a wide range of traditionally challenging tasks. These methods have so far been vision-dominant, and integrating tactile sensing into them has become an active research direction. With tactile-inclusive demonstration data collected with teleoperation or handheld devices, tactile-augmented IL has been shown to handle diverse contact-rich tasks, ranging from industrial assembly to everyday manipulation.

A central question in tactile-augmented IL is how tactile signals should be encoded and fused with existing modalities such as vision and proprioception. It has been widely reported across the literature that naive integration strategies—such as applying vision encoders and their pre-trained weights on tactile images, or simply treating tactile signals as additional observation inputs to the policy—often do not improve policy performance (92–98). In response, researchers have developed many novel integration strategies, which we organize and discuss below.

4.1.1. Novel tactile representations. A representation is a learned or designed encoding that maps raw, high-dimensional sensory inputs into a compact feature space that captures its task-relevant structure. Suitable representations can greatly improve the effectiveness of downstream policy learning (99). Developing novel tactile representations is especially active for high-dimensional tactile signals that are often treated as images, such as those from tactile arrays and vision-based tactile sensors. A few visual–tactile policy learning methods have shown that careful preprocessing of tactile signals, such as constructing visual–tactile point clouds (100, 101) and extracting dense force maps from GelSight Mini images (102), allows standard vision encoders to be applied effectively to tactile signals. A larger body of work has learned dedicated tactile encoder weights through a pretraining stage that precedes policy learning. Objectives used for pretraining include visual–tactile feature alignment (103, 104), future tactile signal prediction (105), masked autoencoding (95, 106, 107), and other self-supervised approaches (108–112). While low-dimensional F/T signals are often encoded with simpler neural networks without pretraining, applying standard vision encoders to 2D spectrogram images converted from F/T signals has proven to be more effective in some cases (113), and self-supervised pretraining yields better results for tactile skin F/T readings distributed across a dexterous hand (106).

4.1.2. Refining vision-generated trajectories with reactive tactile policies. Modern IL methods often predict and execute a sequence of future actions, often called an action chunk, in a single inference step (91, 114). This improves the temporal consistency of generated actions,

with the cost that the robot operates in an open-loop manner within each action chunk, unable to react to events that occur before the next prediction. This differs from how humans use touch for contact-rich manipulation, where fast tactile feedback drives rapid responses to contact events such as making and breaking contact, slip, and sudden changes in force (115). Inspired by human behaviors, a growing body of work has paired a vision-dominant policy that runs at the standard vision policy rate with a second, tactile-driven policy that runs at a higher frequency to make reactive adjustments between chunk predictions (93, 112, 116).

A structural limitation of this hierarchical design is that the fast tactile policy's high-frequency requirement constrains how complex its inputs and architecture can be. In practice, the fast policies operate with low-dimensional tactile signals such as F/T readings (93, 116) or tactile marker displacements (93, 112), rather than the high-resolution tactile imagery that many of the representation learning methods discussed in Section 4.1.1 were designed to exploit.

4.1.3. Adapting policy output for force-based control. Aside from methods that incorporate tactile sensing into the policy's input, another line of work extends the policy's output. While recent IL methods adopt position-based control and predict a six-DOF pose (91) or joint-angle targets (114), force-based control has long been the standard approach in classical robotics for handling contact-rich manipulation. By expanding the action space to include force targets alongside position targets, predicted actions can be executed through classical force-based controllers such as impedance and admittance control, allowing learned policies to inherit a mature body of control techniques rather than reinventing them (117). With an F/T sensor at the robot wrist, the policy can be trained to predict force-informed virtual position targets (113, 118–120) or to predict force targets directly alongside position targets (92, 116, 121). Similarly, when the gripper fingers are equipped with fingertip tactile sensors, the policy can be trained to predict target grasping force alongside target finger width to enable precise grasp control (102, 119).

Less settled is the trade-off between compliant behaviors and how accurately a robot can follow the given position targets. Some works have selected a fixed set of compliant controller gains suitable to solve all chosen tasks (116, 118), while others have found it necessary to have a different set of preset gains for each task (120, 121). It is also common to apply compliant control when the robot is in contact with the environment and high-accuracy position control otherwise, conditioned on a manually set force threshold to determine whether contact is happening (92, 97, 102, 113).

4.1.4. Additional approaches. Beyond the three patterns above, several other integration mechanisms have emerged. Drawing from the empirical observation that naive concatenation causes policies to ignore tactile feedback, one line of work manipulates the training process to force greater tactile reliance, randomly masking vision inputs throughout training (94) or blurring vision inputs at the early stages of training (96). Recognizing that tactile signals become more informative after contact happens, another line of work has found it beneficial to dynamically weight vision and tactile features based on the predicted likelihood of contact (97, 112). A related work has explored policy composition rather than feature fusion, dynamically combining the score functions of separate vision and tactile experts (98). A more recent direction uses tactile signals to steer a vision-only base policy at inference time (122).

4.1.5. Discussion. Tactile-augmented IL methods have brought performance improvements to vision-only policies along several dimensions: appropriate force regulation (92–94, 97, 113, 122), reliable in-hand state estimation (98, 100, 105, 122), robustness to visual occlusion (95, 98, 108), detection of contact events and transitions (95, 102, 122), and generalization to unseen objects (96, 104, 108, 119). However, these gains are uneven across sensor types and integration strategies,

motivating the discussion on tactile sensor heterogeneity in Section 5.1 and the open question of the functional roles of different tactile signal types in IL.

The available evidence suggests that while the gains in force regulation and contact event detection are largely sensor-agnostic (92–95, 97, 102, 113, 122), in-hand state estimation and visual occlusion robustness tend to benefit from the geometric details available from high-resolution tactile sensors (95, 98, 100, 105, 108, 122). However, it has also been reported that tactile resolution gains can saturate (100). These observations call for a more principled understanding of when richer tactile signals yield meaningful policy gains and when simpler sensing suffices.

4.2. Tactile-Augmented Reinforcement Learning

RL is a framework in which the robot learns a policy through trial-and-error interaction with its environment, optimizing its actions based on a reward signal that defines task success. For contact-rich manipulation, RL is well-suited to learning policies that react to high-frequency tactile signals and removes the need for hand-engineered contact strategies. The combination of tactile sensing and RL has a long history in the field (123, 124), with tactile signals most frequently entering the RL pipeline as part of the policy’s observations. This trial-and-error paradigm comes at a cost, however, as shaping an effective policy typically requires a large amount of interaction data. We discuss below how this challenge manifests for tactile-driven manipulation and how the field is addressing it, followed by a discussion of combining RL with IL—a growing direction in manipulation research—for tactile-driven manipulation.

4.2.1. Real-world reinforcement learning. Training RL policies directly in the real world allows robots to acquire manipulation skills from the distribution of interactions they will encounter at deployment. This is appealing for tactile-based manipulation since obtaining accurate contact and rich tactile signals in simulation can be costly. Real-world tactile RL has been demonstrated both on tasks that require reacting to subtle changes in contact state, such as peg-in-hole insertion (125, 126) and typing on a braille keyboard (127), and on tasks under challenging dynamic conditions, such as overhand throwing of deformable darts (128). Data efficiency is critical in this setting for several reasons: Running episodes on real hardware is slow, automatic resets between episodes require careful engineering (125, 128), and tactile sensors are fragile and degrade under repeated contact. Choosing compact and task-appropriate tactile representations, such as marker flows extracted from the raw images of vision-based tactile sensors (125, 127, 128) or pretrained multimodal representations that fuse vision and F/T signals (126), has proven to be an effective strategy. Appropriate RL design choices can also improve data efficiency for real-world tactile RL, including curriculum learning (125), stage-based dense rewards (126), and hindsight experience replay for sparse-reward settings (127). However, current real-world tactile RL is confined largely to small action spaces, such as reduced end-effector degrees of freedom, limiting the range of tasks for which the approach is currently viable.

4.2.2. Sim-to-real reinforcement learning. As simulation becomes faster and more physically accurate, a growing body of work is first training RL policies in simulation before deploying them on real hardware, bypassing the constraints of real-world RL. Simulation also provides access to privileged information difficult to obtain in the real world, such as ground truth object pose and dynamics, enabling data-efficient methods such as asymmetric actor-critic and teacher-student distillation.

One line of work that has benefited the most from sim-to-real RL is dexterous in-hand manipulation, which often takes the form of rotating or translating objects in the robot hand through finger movements. OpenAI’s pivotal work (129) demonstrated the viability of sim-to-real RL for in-hand object reorientation, with real-world state estimation provided by a multicamera setup.

Adding tactile sensing to the system helps reduce such reliance on cameras (50). Subsequent works have demonstrated blind in-hand object rotation and translation (20, 23, 130), simultaneous in-hand object rotation and 3D shape reconstruction (19), and rotation under arbitrary palm orientations (22). These methods often choose to bypass complicated sim-to-real procedures for raw tactile signals by extracting simplified tactile features, such as binary contact (20, 23, 130) and discretized contact location (19). Even at reduced fidelity, including touch consistently improves generalization to unseen objects (19, 20, 23).

In contrast, a growing body of recent work has simulated and leveraged richer and sometimes raw tactile sensor signals, including contact force (26), marker displacements (10, 25, 131), and raw tactile images (10, 27, 131). Using advanced simulation and sim-to-real techniques described in Section 2.2, these works have demonstrated successful zero-shot transfer of various types of insertion and nonprehensile tasks. So far, such results have been demonstrated primarily on parallel-jaw grippers and similar systems.

4.2.3. Combining reinforcement learning and imitation learning for tactile-driven manipulation. RL and IL have complementary strengths: RL can discover suitable strategies but suffers from sample inefficiency and costly real-world training, while IL leverages demonstrations but struggles with hard-to-demonstrate behaviors. Combining them has thus become an active research area in robotic manipulation (132).

Recent works have shown that a residual policy can be learned through real-world RL with dense rewards to provide small corrections to the base tactile-augmented IL policy, improving overall policy performance (133, 134). Leveraging recent advances in tactile simulation (10) and simulation-friendly sensors (135), VT-Refine (101) demonstrates that tactile-augmented diffusion policies trained on real-world data (100) can be fine-tuned directly in simulation using only a binary task-success reward. The fine-tuned policies transfer zero-shot to the real world and exhibit emergent tactile behaviors absent from the demonstrations, significantly outperforming the base policy on various bimanual assembly tasks. Another recent work has proposed an IL system augmented with RL-learned tactile primitives such as dexterous in-hand rotation and stable grasping (136). By delegating these complex low-level tactile skills to RL, the IL policy can focus on high-level planning, yielding significantly better results than end-to-end IL pipelines that attempt to learn such skills from scratch.

For tactile-driven manipulation, hybrid RL + IL approaches remain a much smaller body of work than stand-alone tactile RL or IL methods. This likely reflects that tactile RL + IL inherits many bottlenecks from tactile RL: Real-world RL can degrade tactile sensors quickly, and sim-to-real transfer remains a delicate and complicated procedure for many tactile sensor types.

4.2.4. Discussion. The three threads discussed above highlight one of the central challenges of tactile-augmented RL: the cost of data. Much of the progress in tactile-augmented RL can be viewed as attempts to improve the efficiency of acquiring, generating, and leveraging such data. Beyond these efforts, however, we believe another fundamental question remains largely unexplored: Choosing tactile signals and representations for RL involves fundamental task-dependent trade-offs among signal richness, data efficiency, and simulatability.

Several considerations push toward compact, low-dimensional tactile features: They improve data efficiency for RL training and reduce sim-to-real gaps. Although simulation for complex tactile signals is maturing rapidly, whether richer signals actually yield meaningful RL gains remains an open question. Dexterous in-hand manipulation works have demonstrated impressive results with sparse or binary tactile signals, and the evidence on whether additional signal richness improves performance is mixed (137). Some report that binary or low-dimensional contact signals are sufficient or even competitive with denser feedback (19, 20, 130), while others find that

dense-featured touch yields concrete gains for specific tasks, particularly when subtle contact information is needed to detect events such as incipient slip (22). Taken together, these discrepancies suggest that the benefit of richer tactile feedback is task-dependent rather than universal. The resulting trade-offs among signal richness, data efficiency, and simulatability motivate further research into mapping task demands to suitable tactile features.

4.3. Tactile-Augmented Vision–Language–Action Models

Sections 4.1 and 4.2 discussed learning paradigms that are typically trained from scratch on human demonstration or robot exploration data. Although these methods are applicable to a wide range of tasks, they require training a specific policy for each task. To reduce this burden, the broader robotic manipulation community has begun exploring a different paradigm: VLA models (138). Trained with supervised objectives similar to those in Section 4.1 but on larger-scale multitask demonstration data, VLAs are built by equipping a pretrained vision–language model (VLM), often called a backbone, to output robot actions from visual observations and natural language task specifications. This mapping is often handled by a dedicated action expert, which is a separate network that decodes the backbone’s representations into robot control commands. Inheriting the rich representations and commonsense knowledge from the backbone pretrained on internet-scale data, VLAs are designed with the ultimate goal of generalizing across different tasks, objects, embodiments, environments, and language instructions through a single set of policy weights, with recent systems making encouraging progress in this direction (139).

The improvements tactile sensing brings to vision-only IL policies (Section 4.1) motivate analogous efforts in the VLA setting. However, this paradigm shift sets up new challenges for tactile sensing integration. In a policy trained from scratch, the policy’s internal representations for both vision and touch are mostly shaped end-to-end by the task-specific datasets available. In a VLA, by contrast, the powerful backbone arrives with semantic priors learned from internet-scale image–text data with little notion of tactile or contact-level information. The architectural question is no longer how to fuse tactile information with vision in a jointly trained policy, but where and how to engineer the tactile modality into a representation space that was trained without it. In contrast to Section 4.1, where high-resolution tactile imagery paired with representation learning has proven highly effective and remains a dominant approach, many tactile-augmented VLA methods opt to work with low-dimensional signals—such as F/T and pressure-array readings—encoded through linear projections or shallow multilayer perceptrons. Below, we further discuss the two major strategies for integrating tactile sensing into VLA pipelines.

4.3.1. The tactile modality as a coequal input modality in the backbone. One line of work treats the tactile modality equally to the pretrained backbone’s native input modalities—vision and language—and adapts the backbone to accommodate them. Therefore, the technical question becomes how to reshape the backbone’s representation space without degrading its pretrained knowledge. One approach is adding a multimodal contrastive learning structure to the VLA training pipeline to maximize the mutual information shared between vision, language, and touch (140, 141). Another is to directly use noncausal attention masks in the VLM backbone, allowing tactile input to fuse with other modalities, either at all times (142) or only during contact (143). This method shapes a deeply integrated representation space. Tactile-VLA (142) observes that the pretrained VLM’s prior knowledge already includes a semantic understanding of physical interaction and that connecting it to tactile signals via multimodal demonstrations is sufficient to unlock this understanding for downstream robot control.

4.3.2. The tactile modality as add-on modules to the backbone. In contrast, other works have taken the perspective that the pretrained VLM backbone’s representation space is not only

valuable but also fragile under fine-tuning. Therefore, their design commitment is to separate the tactile stream from those of vision and language, preventing it from disrupting the backbone's pretrained representation space. Directly contrasting the multimodal fusion-friendly attention mask design mentioned in Section 4.3.1, FD-VLA (144) designs directional attention masks in the VLM to limit fusion between tactile and VLM-native modalities. TA-VLA (145) provides a systematic exploration of where and how to introduce torque signals into a VLA, empirically finding that integrating torque into the VLA's action expert rather than its VLM backbone yields the strongest results. Similar to the concept described in Section 4.1.2, T-Rex (146) and AT-VLA (147) achieve rapid closed-loop reactions by injecting high-frequency tactile streams into the base VLA's action expert through cascaded flow matching and contact-gated adaptive cross-attention. Another method of tactile integration uses mixture-of-experts (148) either to fuse tactile features with the output of the backbone in order to form the input to the action expert (149, 150) or to compute residual corrections for the VLA's raw action predictions (136). Building on the intuition that tactile perception contributes to manipulation on both the planning and control level, VLA-Touch (151) uses a tactile-language model (152) to help the VLM planner generate object property-informed instructions for a vision-only VLA and uses a tactile-based interpolant controller to refine the raw actions predicted by the VLA. The recent work FTP-1 (153) introduces a dedicated tactile expert parallel to the VLM and the action expert and pretrains it on approximately 3,000 hours of data across 21 types of heterogeneous tactile sensors, offering a reusable initialization for downstream tactile policy training.

4.3.3. Discussion. Tactile sensing has improved the performance of vision-only VLA policies across diverse contact-rich tasks similar to those in Section 4.1. Although the two integration strategies discussed above make opposite architectural commitments, they share the following broader pattern when viewed across all works: Existing tactile-augmented VLAs establish architectural feasibility, but substantial opportunities remain for realizing the VLA paradigm's core promise of generalization.

While a few works have shown that their tactile-augmented VLAs generalize across various novel objects (149) and language instructions (140, 142), most train one policy per task and evaluate it in the same narrow setting as the trained-from-scratch IL policies in Section 4.1, with weak emphasis on generalization. More systematic evaluations in genuinely multitask, language-conditioned, and novel-object settings would demonstrate the distinctive promises of the tactile-augmented VLA paradigm itself.

4.4. Tactile Predictive Models for Manipulation

A predictive model for manipulation takes the current state and a candidate action as input and outputs the predicted future state. Such models support a predict-then-control paradigm: Candidate actions are evaluated through their predicted outcomes, and the action most aligned with the task objective is selected. This idea underlies classical model predictive control (MPC), which optimizes a finite-horizon action sequence against a dynamics model (154, 155). World models, a prominent paradigm emerging more broadly in robot learning, build on the same predict-then-control idea while reworking how the predictive model and its representations are built. Specifically, we distinguish two cases: A dynamics model typically predicts transitions over a compact, task-specific state and is coupled to an external optimizer such as MPC, whereas a world model extends the prediction target to high-dimensional observations or latent encodings learned by pretrained encoders. In both cases, action-conditioned prediction is what makes the model useful for control. These predictive models matter most in contact-rich tasks where success depends on how contact evolves—slip, force change, making or breaking contact—rather than on what the

camera sees. Since that decisive future is local to the contact interface and often occluded, predicting it from touch is what informs control. Consistent with this section’s data-driven focus, the predictive models we cover are learned from interaction data rather than derived from analytical contact mechanics. We use tactile predictive models as an umbrella term for this learned lineage, treating recent tactile world models as a scaling of predict-then-control to tactile or visual–tactile latents rather than a new paradigm.

4.4.1. Predictive control with tactile dynamics models. Because contact is local and its mechanics are comparatively well-understood, touch was a natural target for forward modeling. Initial efforts on tactile prediction models were developed largely as task-specific predictive-control pipelines rather than as general-purpose world models. Tian et al. (156) trained an action-conditioned model that forecasts future GelSight tactile images of a small object under candidate finger movement. They then embedded this predictor into sampling-based MPC to servo objects such as a ball, an analog stick, and a die toward a goal position specified directly as a raw tactile image. Sutanto et al. (157) followed a similar servoing paradigm but represented tactile observations in a learned latent space rather than using raw tactile signals directly. LeTac-MPC (158) takes a different route by embedding a differentiable MPC layer over a learned tactile representation and training the controller end-to-end for reactive grasping. This enables stable grasping of objects with varying physical properties under external disturbance. These predictive models are trained with interaction data on the same embodiment, following either an automatic exploration pattern or a human-defined trajectory. Together, these works established the appeal that the rest of this subsection inherits: Contact is structured enough to model forward and informative enough for closed-loop correction.

4.4.2. Tactile world models for control. World models are now a leading approach in vision-based control, where learning to predict how observations evolve under action has driven strong results on robotic manipulation (159, 160). The visual–tactile world models below build on that success, extending visual world models to settings where contact carries the decisive signal. The works differ mainly in how the prediction is turned into an action: optimized over by an external planner, or consumed by a learned policy.

One line of work hands the forecast to an external action planner. RoboPack (161) learns a tactile-informed dynamics model that predicts future object states, using touch to infer latent physical properties that vision alone cannot reveal, and plans action with MPC. VT-WM (162) carries the same scheme into pretrained latents for both modalities, then plans by sampling action sequences and rolling them through the predicted latents. A second line uses forecasts more directly inside the policy. OmniVTA (112) forecasts future vision and touch and feeds the tactile forecast to a 60-Hz reflexive controller that corrects deviations between predicted and observed contact. By contrast, ViTacFormer (163) and DreamTacVLA (164) predict only the future tactile signal and condition action generation on that forecast. Recent works such as Dream-Tac (165) attempt to harness the power of large pretrained video prediction models by fine-tuning them to predict future tactile readings and robot actions alongside robot camera views.

So far, these methods work best on short-horizon, contact-reactive tasks where the decisive signal is immediately available, such as pushing (161, 162), wiping (112, 162), and stacking (162), or correcting an action as soon as observed contact departs from the prediction (112, 164). The longer-horizon prediction that originally motivated vision world models has not been demonstrated for touch, and whether it suits touch’s inherently local nature is itself unclear.

4.4.3. Tactile prediction as an auxiliary training objective. A different strategy uses tactile prediction only during training. An auxiliary loss that forecasts future tactile signals or contact is added to the policy-learning objective in the hope that the policy learns a more contact-aware

representation. At deployment, the policy acts on current observations alone. This method applies to policies that are trained in different ways, including IL (105, 166), RL (126, 137), and VLA (141, 145) policies, consistently reporting that the auxiliary tactile prediction objective improves policy performance on contact-rich tasks over a no-prediction baseline, at no added inference cost.

4.4.4. Discussion. While predict-then-control is a long-standing idea, its tactile and visual-tactile world-model instantiation is recent, and the optimal design choices are yet to be explored. Rather than offering firm conclusions, we close by raising two open issues that we view as crucial to advancing this direction.

First, a tactile world model should be more than a high-fidelity signal predictor. Many tactile predictive methods (112, 156, 163) apply pixel- or token-level similarity metrics between predicted and ground truth frames, which are inherited from the vision modality. But these methods might not be suitable to the tactile modality. For touch, subtle spikes in contact may barely affect the averaged error but are essential for task-relevant information. We therefore believe that developing a more suitable metric for tactile signals will significantly improve the effectiveness of using touch in contact-rich manipulation.

Second, it is unclear whether foundation-scale prediction can keep pace with contact dynamics. Robots need to quickly react to many tactile events, such as making and breaking contact, subtle changes in force, and incipient slip, whereas foundation-scale predictors (112, 162) run at much lower rates. It is essential for tactile world models to address this temporal mismatch between slow prediction and fast contact dynamics. We believe the solutions include developing either faster prediction models or mechanisms that explicitly separate slow predictive planning from fast reactive correction.

5. OPEN PROBLEMS AND FUTURE DIRECTIONS

The rapid development of tactile sensing, robot learning, and multimodal foundation models has created new opportunities for tactile-based manipulation. But the new tools and new possibilities also introduce new challenges that may play a critical role in determining the future of tactile-based manipulation.

5.1. Tactile Sensor Heterogeneity and Transferability

As discussed in Section 2.1, tactile sensors exhibit substantial variation in signal formats and output characteristics. This heterogeneity has become a major obstacle to scaling tactile datasets and sharing knowledge across the community. In contrast to vision, where large datasets can often be reused across platforms, tactile datasets and models are frequently tied to specific sensors, making it difficult to transfer learned representations and policies across hardware. A growing consensus is that addressing this challenge requires the development of sensor-agnostic tactile representations. One approach is learning shared latent representations directly from data collected from multiple vision-based tactile sensors (107, 110, 167, 168), with recent work extending this effort to additionally include array- and state-based tactile sensors (153). Another uses physically meaningful quantities, such as contact geometry (169) or contact force (170), as an intermediate representation shared across sensors. While promising progress has been made, developing representations that generalize across diverse tactile sensors and tasks remains an important open challenge.

5.2. Making Tactile Sensing Lightweight

Many manipulation tasks require robots to react rapidly to contact events, such as collisions, slip, or changes in contact conditions. As tactile sensors and algorithms become increasingly complex,

the latency introduced by communication and computation can become a bottleneck for real-time manipulation. We therefore believe that making tactile sensing more lightweight remains an important research direction. Possible approaches include (a) applying task-appropriate sensors rather than maximizing sensor signal richness, (b) developing low-dimensional tactile representations that support efficient perception and control (93), and (c) co-designing sensor hardware and algorithms, for example, by performing signal preprocessing and feature extraction directly on the sensor. Such approaches may improve the responsiveness and deployability of tactile-enabled robotic systems.

This also raises a broader question about sensor design: What properties should future tactile sensors prioritize? Despite the availability of rich tactile signals, many projects ended up using only low-dimensional tactile information. In current practice, we believe that sensor characteristics such as repeatability, durability, latency, and ease of integration are important for building an effective manipulation system. We therefore argue that, rather than pursuing increasingly complex designs and higher signal bandwidth alone, the community would benefit from robust, accessible, and reliable tactile sensors.

5.3. How Should We Evaluate Tactile-Driven Manipulation?

For tactile-driven manipulation, a fundamental question remains: What does it mean to achieve general-purpose manipulation? Most existing work evaluates success rates on narrowly scoped tasks that are difficult to compare across studies and may not reflect broader manipulation requirements. As a result, it remains challenging to assess how much progress individual methods contribute toward general tactile-enabled manipulation.

Recent work has begun to consider broader evaluation criteria, including training data efficiency (104, 137), task efficiency (22, 26), interaction force (26, 102), and generalization across objects (23, 96, 149). These metrics better reflect the practical requirements of real-world manipulation. However, the community has yet to converge on a common suite of representative evaluation tasks. Such benchmarks would not only facilitate more meaningful comparison between methods but also help align the trajectory of the field itself. Evaluation tasks shape the systems built around them; agreeing on representative tactile manipulation challenges would encourage researchers to develop methods that address shared and broadly meaningful goals rather than optimizing for isolated tasks. Recent benchmarking efforts in simulation-based tactile manipulation (137, 171, 172) represent encouraging progress, and extending these efforts across a wider range of sensors and tasks remains an important direction for future research.

5.4. Lessons from Mechanics-Driven to Data-Driven Methods

While recent progress in tactile-driven manipulation has been driven largely by data-driven approaches, the underlying physics of touch and contact-rich manipulation remain unchanged. Thus, insights accumulated from mechanics-driven methods continue to provide valuable structure for understanding contact interactions and designing effective manipulation systems. We believe that future advances will require not only scalable learning techniques but also a deeper understanding of the physical principles underlying touch. Such knowledge may further inform the design of more efficient tactile representations and learning architectures.

Some concepts established in mechanics-driven methods also remain underexplored in data-driven approaches. One example is active touch, where tactile perception is coupled with exploratory robot actions to acquire task-relevant information. As briefly discussed in Section 3.1.7, active touch has primarily been studied for perception (173). Estimates of object states and physical properties obtained by active exploration can also inform low-level

manipulation strategies (Section 3.2.3). Such approaches are particularly useful when interacting with objects with unknown or challenging physical properties. Existing data-driven manipulation methods often do not include components for actively gaining object properties or other task-relevant information through exploration, but this capability can improve manipulation policies' generalizability when handling new environments or novel objects. As the field moves toward manipulation in more diverse and complex physical environments, actively exploring the environment to gain task-relevant information may become increasingly important.

6. CONCLUSION

Tactile sensing provides robots with direct access to physical interactions at the contact interface, making it a unique and valuable modality for robotic manipulation. In this article, we have reviewed the development of tactile-based manipulation from both mechanics-driven and data-driven perspectives. Viewed through the history of the field, tactile manipulation has evolved from explicit reasoning about contact phenomena toward learning tactile representations and manipulation policies directly from data. Despite their methodological differences, both paradigms seek to answer the same fundamental question: How can robots use touch to better understand and interact with the physical world? We hope that this review provides a useful reference for researchers working toward that goal.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This material is based on work supported in part by the US National Science Foundation under award number 2442310. The authors thank Ruohan Zhang for insightful discussions and feedback on the manuscript.

LITERATURE CITED

1. Tomo TP, Regoli M, Schmitz A, Natale L, Kristanto H, et al. 2018. A new silicone structure for uSkin—a soft, distributed, digital 3-axis skin sensor and its integration on the humanoid robot iCub. *IEEE Robot. Autom. Lett.* 3(3):2584–91
2. Yan Y, Hu Z, Yang Z, Yuan W, Song C, et al. 2021. Soft magnetic skin for super-resolution tactile sensing with force self-decoupling. *Sci. Robot.* 6(51):eabc8801
3. Ward-Cherrier B, Pestell N, Cramphorn L, Winstone B, Giannaccini ME, et al. 2018. The TacTip family: soft optical tactile sensors with 3D-printed biomimetic morphologies. *Soft Robot.* 5(2):216–27
4. Yuan W, Dong S, Adelson EH. 2017. GelSight: high-resolution robot tactile sensors for estimating geometry and force. *Sensors* 17(12):2762
5. Dahiya RS, Metta G, Valle M, Sandini G. 2010. Tactile sensing—from humans to humanoids. *IEEE Trans. Robot.* 26(1):1–20
6. Lin C, Lin Z, Wang S, Xu H. 2023. DTact: a vision-based tactile sensor that measures high-resolution 3D geometry directly from darkness. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
7. Schneiter JL, Sheridan TB. 1984. An optical tactile sensor for manipulators. *Robot. Comput.-Integr. Manuf.* 1(1):65–71
8. Narang Y, Sundaralingam B, Macklin M, Mousavian A, Fox D. 2021. Sim-to-real for robotic tactile sensing via physics-based simulation and learned latent projections. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE

9. Agarwal A, Man T, Yuan W. 2021. Simulation of vision-based tactile sensors using physics based rendering. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
10. Akinola I, Xu J, Carius J, Fox D, Narang Y. 2025. TacSL: a library for visuotactile sensor simulation and learning. *IEEE Trans. Robot.* 41:2645–61
11. Dang A, Lee J, Mukadam M, Wu XA, Bucher B, et al. 2026. HydroShear: hydroelastic shear simulation for tactile sim-to-real reinforcement learning. In *Robotics: Science and Systems XXII*, ed. M Ciocarlie, L Carlone, J Strader. Robotics: Science and Systems Foundation
12. Ruppel P, Jonetzko Y, Görner M, Hendrich N, Zhang J. 2019. Simulation of the SynTouch BioTac sensor. In *Intelligent Autonomous Systems 15*, ed. M Strand, R Dillmann, E Menegatti, S Ghidoni. Springer
13. Sferrazza C, Bi T, D’Andrea R. 2020. Learning the sense of touch in simulation: a sim-to-real strategy for vision-based tactile sensing. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE
14. Ding Z, Lepora NF, Johns E. 2020. Sim-to-real transfer for optical tactile sensing. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
15. Pezzementi Z, Jantho E, Estrade L, Hager GD. 2010. Characterization and simulation of tactile sensors. In *2010 IEEE Haptics Symposium*. IEEE
16. Si Z, Yuan W. 2022. Taxim: an example-based simulation model for GelSight tactile sensors. *IEEE Robot. Autom. Lett.* 7(2):2361–68
17. Suresh S, Si Z, Mangelson JG, Yuan W, Kaess M. 2022. ShapeMap 3-D: efficient shape mapping through dense touch and vision. In *2022 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
18. Si Z, Zhu Z, Agarwal A, Anderson S, Yuan W. 2022. Grasp stability prediction with sim-to-real transfer from tactile sensing. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE
19. Qi H, Yi B, Suresh S, Lambeta M, Ma Y, et al. 2023. General in-hand object rotation with vision and touch. In *Proceedings of the 7th Conference on Robot Learning*, ed. J Tan, M Toussaint, K Darvish. Proceedings of Machine Learning Research 229. PMLR
20. Yin ZH, Huang B, Qin Y, Chen Q, Wang X. 2023. Rotating without seeing: towards in-hand dexterity through touch. In *Robotics: Science and Systems XIX*, ed. K Bekris, K Hauser, S Herbert, J Yu. Robotics: Science and Systems Foundation
21. Yuan Y, Che H, Qin Y, Huang B, Yin ZH, et al. 2024. Robot synesthesia: in-hand manipulation with visuotactile sensing. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
22. Yang M, Lu C, Church A, Lin Y, Ford CJ, et al. 2025. AnyRotate: gravity-invariant in-hand object rotation with sim-to-real touch. In *Proceedings of the 8th Conference on Robot Learning*, ed. P Agrawal, O Kroemer, W Burgard. Proceedings of Machine Learning Research 270. PMLR
23. Yin J, Qi H, Malik J, Pikul J, Yim M, et al. 2025. Learning in-hand translation using tactile skin with shear and normal force sensing. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
24. Gomes DF, Paoletti P, Luo S. 2021. Generation of GelSight tactile images for Sim2Real learning. *IEEE Robot. Autom. Lett.* 6(2):4177–84
25. Chen W, Xu J, Xiang F, Yuan X, Su H, et al. 2024. General-purpose Sim2Real protocol for learning contact-rich manipulation with marker-based visuotactile sensors. *IEEE Trans. Robot.* 40:1509–26
26. Noseworthy M, Tang B, Wen B, Handa A, Kessens C, et al. 2025. FORGE: force-guided exploration for robust contact-rich manipulation under uncertainty. *IEEE Robot. Autom. Lett.* 10(5):4436–43
27. Church A, Lloyd J, Hadsell R, Lepora NF. 2022. Tactile sim-to-real policy transfer via real-to-sim image translation. In *Proceedings of the 5th Conference on Robot Learning*, ed. A Faust, D Hsu, G Neumann. Proceedings of Machine Learning Research 164. PMLR
28. Wang S, Lambeta M, Chou PW, Calandra R. 2022. TACTO: a fast, flexible, and open-source simulator for high-resolution vision-based tactile sensors. *IEEE Robot. Autom. Lett.* 7(2):3930–37
29. Jianu T, Gomes DF, Luo S. 2022. Reducing tactile Sim2Real domain gaps via deep texture generation networks. In *2022 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
30. Lin Y, Lloyd J, Church A, Lepora NF. 2022. Tactile Gym 2.0: sim-to-real deep reinforcement learning for comparing low-cost high-resolution robot touch. *IEEE Robot. Autom. Lett.* 7(4):10754–61

31. Zhao Y, Jing X, Qian K, Gomes DF, Luo S. 2023. Skill generalization of tubular object manipulation with tactile sensing and Sim2Real learning. *Robot. Auton. Syst.* 160:104321
32. Romano JM, Hsiao K, Niemeyer G, Chitta S, Kuchenbecker KJ. 2011. Human-inspired robotic grasp control with tactile sensing. *IEEE Trans. Robot.* 27(6):1067–79
33. She Y, Wang S, Dong S, Sunil N, Rodriguez A, et al. 2021. Cable manipulation with a tactile-reactive gripper. *Int. J. Robot. Res.* 40(12–14):1385–401
34. Zhao Z, Li Y, Li W, Qi Z, Ruan L, et al. 2025. Tac-Man: tactile-informed prior-free manipulation of articulated objects. *IEEE Trans. Robot.* 41:538–57
35. Su Z, Kroemer O, Loeb GE, Sukhatme GS, Schaal S. 2016. Learning to switch between sensorimotor primitives using multimodal haptic signals. In *From Animals to Animats 14*, ed. E Tuci, A Giagkos, M Wilson, J Hallam. Springer
36. Wilson A, Jiang H, Lian W, Yuan W. 2023. Cable routing and assembly using tactile-driven motion primitives. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
37. Van Wyk K, Culleton M, Falco J, Kelly K. 2018. Comparative peg-in-hole testing of a force-based manipulation controlled robotic hand. *IEEE Trans. Robot.* 34(2):542–49
38. Karayiannidis Y, Smith C, Viña FE, Kragic D. 2014. Online contact point estimation for uncalibrated tool use. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
39. Yu KT, Bauza M, Fazeli N, Rodriguez A. 2016. More than a million ways to be pushed. A high-fidelity experimental dataset of planar pushing. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE
40. Eguíluz AG, Rañó I, Coleman SA, McGinnity TM. 2017. Reliable object handover through tactile force sensing and effort control in the Shadow Robot hand. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
41. Li X, Li R, Qiao H, Ma C, Li L. 2017. Human-inspired compliant strategy for peg-in-hole assembly using environmental constraint and coarse force information. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE
42. Sunil N, Wang S, She Y, Adelson E, Garcia AR. 2023. Visuotactile affordances for cloth manipulation with local control. In *Proceedings of the 6th Conference on Robot Learning*, ed. K Liu, D Kulic, J Ichnowski. Proceedings of Machine Learning Research 205. PMLR
43. Pai S, Chen T, Tippur M, Adelson E, Gupta A, et al. 2023. TactoFind: a tactile only system for object retrieval. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
44. Bauza M, Bronars A, Rodriguez A. 2023. Tac2Pose: tactile object pose estimation from the first touch. *Int. J. Robot. Res.* 42(13):1185–209
45. Dean-Leon E, Guadarrama-Olvera JR, Bergner F, Cheng G. 2019. Whole-body active compliance control for humanoid robots with robot skin. In *2019 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
46. Huang HJ, Kaess M, Yuan W. 2025. NormalFlow: fast, robust, and accurate contact-based object 6DoF pose tracking with vision-based tactile sensors. *IEEE Robot. Autom. Lett.* 10(1):452–59
47. Li R, Platt R, Yuan W, ten Pas A, Roscup N, et al. 2014. Localization and manipulation of small parts using GelSight tactile sensing. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE
48. Bauza M, Bronars A, Hou Y, Taylor I, Chavan-Dafle N, et al. 2024. SimPLE, a visuotactile method learned in simulation to precisely pick, localize, regrasp, and place objects. *Sci. Robot.* 9(91):eadi8808
49. Hogan FR, Ballester J, Dong S, Rodriguez A. 2020. Tactile dexterity: manipulation primitives with tactile feedback. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
50. Pitz J, Röstel L, Sievers L, Bäuml B. 2023. Dextrous tactile in-hand manipulation using a modular reinforcement learning architecture. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
51. Bronars A, Kim S, Patre P, Rodriguez A. 2024. TEXterity: Tactile Extrinsic deXterity. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
52. Lepora NF, Aquilina K, Crampthorn L. 2017. Exploratory tactile servoing with active touch. *IEEE Robot. Autom. Lett.* 2(2):1156–63

53. Melchiorri C. 2000. Slip detection and control using tactile and force sensors. *IEEE/ASME Trans. Mechatron.* 5(3):235–43
54. Howe R, Cutkosky M. 1989. Sensing skin acceleration for slip and texture perception. In *1989 IEEE International Conference on Robotics and Automation*. IEEE
55. Yamada Y, Cutkosky M. 1994. Tactile sensor with 3-axis force and vibration sensing functions and its application to detect rotational slip. In *Proceedings of the 1994 IEEE International Conference on Robotics and Automation*. IEEE
56. Su Z, Hausman K, Chebotar Y, Molchanov A, Loeb GE, et al. 2015. Force estimation and slip detection/classification for grip control using a biomimetic tactile sensor. In *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*. IEEE
57. Dong S, Yuan W, Adelson EH. 2017. Improved GelSight tactile sensor for measuring geometry and slip. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE
58. Ikeda A, Kurita Y, Ueda J, Matsumoto Y, Ogasawara T. 2004. Grip force control for an elastic finger using vision-based incipient slip feedback. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE
59. Yuan W, Li R, Srinivasan MA, Adelson EH. 2015. Measurement of shear and slip with a GelSight tactile sensor. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
60. Khamis H, Izquierdo Albero R, Salerno M, Shah Idil A, Loizou A, et al. 2018. PapillArray: an incipient slip sensor for dexterous robotic or prosthetic manipulation—design and prototype validation. *Sens. Actuators A* 270:195–204
61. Sui R, Zhang L, Li T, Jiang Y. 2021. Incipient slip detection method with vision-based tactile sensor based on distribution force and deformation. *IEEE Sens. J.* 21(22):25973–85
62. Reinecke J, Dietrich A, Schmidt F, Chalon M. 2014. Experimental comparison of slip detection strategies by tactile sensing with the BioTac[®] on the DLR hand arm system. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
63. Meier M, Patzelt F, Haschke R, Ritter HJ. 2016. Tactile convolutional networks for online slip and rotation detection. In *Artificial Neural Networks and Machine Learning—ICANN 2016*, ed. AE Villa, P Masulli, AJ Pons Rivero. Springer
64. James JW, Pestell N, Lepora NF. 2018. Slip detection with a biomimetic tactile sensor. *IEEE Robot. Autom. Lett.* 3(4):3340–46
65. Veiga F, Peters J, Hermans T. 2018. Grip stabilization of novel objects using slip prediction. *IEEE Trans. Haptics* 11(4):531–42
66. Hou Y, Jia Z, Mason M. 2020. Manipulation with shared grasping. In *Robotics: Science and Systems XVI*, ed. M Toussaint, A Bicchi, T Hermans. Robotics: Science and Systems Foundation
67. Kim S, Rodriguez A. 2022. Active extrinsic contact sensing: application to general peg-in-hole insertion. In *2022 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
68. Oller M, Berenson D, Fazeli N. 2024. Tactile-driven non-prehensile object manipulation via extrinsic contact mode control. In *Robotics: Science and Systems XX*, ed. D Kulic, G Venture, K Bekris, E Coronado. Robotics: Science and Systems Foundation
69. Molchanov A, Kroemer O, Su Z, Sukhatme GS. 2016. Contact localization on grasped objects using tactile sensing. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE
70. Atkeson CG, An CH, Hollerbach JM. 1985. Rigid body load identification for manipulators. In *1985 24th IEEE Conference on Decision and Control*. IEEE
71. Guo X, Huang HJ, Yuan W. 2023. Estimating properties of solid particles inside container using touch sensing. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE
72. Jin S, Mo Y, Yuan W. 2025. Learning to double guess: an active perception approach for estimating the center of mass of arbitrary objects. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
73. Su Z, Fishel JA, Yamamoto T, Loeb GE. 2012. Use of tactile feedback to control exploratory movements to characterize object compliance. *Front. Neurobot.* 6:7
74. Yuan W, Zhu C, Owens A, Srinivasan MA, Adelson EH. 2017. Shape-independent hardness estimation using deep learning and a GelSight tactile sensor. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE

75. Huang HJ, Guo X, Yuan W. 2022. Understanding dynamic tactile sensing for liquid property estimation. In *Robotics: Science and Systems XVIII*, ed. K Hauser, D Shell, S Huang. Robotics: Science and Systems Foundation
76. Luo X, Jin S, Huang HJ, Yuan W. 2024. An intelligent robotic system for perceptive pancake batter stirring and precise pouring. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE
77. Wang C, Wang S, Romero B, Veiga F, Adelson E. 2020. SwingBot: learning physical features from in-hand tactile exploration for dynamic swing-up manipulation. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE
78. Huang HJ, Xiang J, Yuan W. 2024. Kitchen Artist: precise control of liquid dispensing for gourmet plating. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
79. Li Q, Schürmann C, Haschke R, Ritter H. 2013. A control framework for tactile servoing. In *Robotics: Science and Systems IX*, ed. P Newman, D Fox, D Hsu. Robotics: Science and Systems Foundation
80. Jain A, Killpack MD, Edsinger A, Kemp CC. 2013. Reaching in clutter with whole-arm tactile sensing. *Int. J. Robot. Res.* 32(4):458–82
81. Lepora NF, Church A, de Kerckhove C, Hadsell R, Lloyd J. 2019. From pixels to percepts: highly robust edge perception and contour following using deep learning and an optical biomimetic tactile sensor. *IEEE Robot. Autom. Lett.* 4(2):2101–7
82. Li Q, Haschke R, Ritter H. 2015. A visuo-tactile control framework for manipulation and exploration of unknown objects. In *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*. IEEE
83. Zheng Y, Veiga FF, Peters J, Santos VJ. 2022. Autonomous learning of page flipping movements via tactile feedback. *IEEE Trans. Robot.* 38(5):2734–49
84. Ajoudani A, Hocaoglu E, Altobelli A, Rossi M, Battaglia E, et al. 2016. Reflex control of the Pisa/IIT SoftHand during object slippage. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
85. Veiga F, Edin B, Peters J. 2020. Grip stabilization through independent finger tactile feedback control. *Sensors* 20(6):1748
86. Zhang Y, Yuan W, Kan Z, Wang MY. 2020. Towards learning to detect and predict contact events on vision-based tactile sensors. In *Proceedings of the Conference on Robot Learning*, ed. LP Kaelbling, D Kragic, K Sugiura. Proceedings of Machine Learning Research 100. PMLR
87. Kolamuri R, Si Z, Zhang Y, Agarwal A, Yuan W. 2021. Improving grasp stability with rotation measurement from tactile sensing. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE
88. Calandra R, Owens A, Jayaraman D, Lin J, Yuan W, et al. 2018. More than a feeling: learning to grasp and regrasp using vision and touch. *IEEE Robot. Autom. Lett.* 3(4):3300–7
89. Hogan FR, Bauza M, Canal O, Donlon E, Rodriguez A. 2018. Tactile regrasp: grasp adjustments via simulated tactile transformations. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE
90. Zhang K, Sharma M, Veloso M, Kroemer O. 2019. Leveraging multimodal haptic sensory data for robust cutting. In *2019 IEEE-RAS 19th International Conference on Humanoid Robots (Humanoids)*. IEEE
91. Chi C, Xu Z, Feng S, Cousineau E, Du Y, et al. 2025. Diffusion Policy: visuomotor policy learning via action diffusion. *Int. J. Robot. Res.* 44(10–11):1684–704
92. Liu W, Wang J, Wang Y, Wang W, Lu C. 2025. ForceMimic: force-centric imitation learning with force-motion capture system for contact-rich manipulation. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
93. Xue H, Ren J, Chen W, Zhang G, Yuan F, et al. 2025. Reactive diffusion policy: slow-fast visual-tactile policy learning for contact-rich manipulation. In *Robotics: Science and Systems XXI*, ed. L Carlone, D Kulic, G Venture, J Strader. Robotics: Science and Systems Foundation
94. Romero B, Fang HS, Agrawal P, Adelson E. 2024. EyeSight Hand: design of a fully-actuated dexterous robot hand with integrated vision-based tactile sensors and compliant actuation. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE

95. Zhu X, Huang B, Li Y. 2025. Touch in the wild: learning fine-grained manipulation with a portable visuo-tactile gripper. In *Advances in Neural Information Processing Systems 38*, ed. D Belgrave, C Zhang, H Lin, R Pascanu, P Koniusz, et al. Curran Associates
96. Liu JJ, Li Y, Shaw K, Tao T, Salakhutdinov R, et al. 2025. FACTR: force-attending curriculum training for contact-rich policy learning. In *Robotics: Science and Systems XXI*, ed. L Carlone, D Kulic, G Venture, J Strader. Robotics: Science and Systems Foundation
97. He Z, Fang H, Chen J, Fang HS, Lu C. 2025. FoAR: force-aware reactive policy for contact-rich robotic manipulation. *IEEE Robot. Autom. Lett.* 10(6):5625–32
98. Chen H, Xu J, Chen H, Hong K, Huang B, et al. 2026. Multi-modal manipulation via multi-modal policy consensus. In *2026 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
99. Pari J, Shafullah N, Arunachalam S, Pinto L. 2022. The surprising effectiveness of representation learning for visual imitation. In *Robotics: Science and Systems XVIII*, ed. K Hauser, D Shell, S Huang. Robotics: Science and Systems Foundation
100. Huang B, Wang Y, Yang X, Luo Y, Li Y. 2025. 3D-ViTac: learning fine-grained manipulation with visuo-tactile sensing. In *Proceedings of the 8th Conference on Robot Learning*, ed. P Agrawal, O Kroemer, W Burgard. Proceedings of Machine Learning Research 270. PMLR
101. Huang B, Xu J, Akinola I, Yang W, Sundaralingam B, et al. 2025. VT-Refine: learning bimanual assembly with visuo-tactile feedback via simulation fine-tuning. In *Proceedings of the 9th Conference on Robot Learning*, ed. J Lim, S Song, HW Park. Proceedings of Machine Learning Research 305. PMLR
102. Helmut E, Funk N, Schneider T, de Farias C, Peters J. 2026. Tactile-conditioned diffusion policy for force-aware robotic manipulation. In *2026 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
103. George A, Gano S, Katragadda P, Farimani AB. 2025. VTaL pretraining: visuo-tactile pretraining for tactile and non-tactile manipulation policies. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
104. Wu L, Yu C, Ren J, Chen L, Jiang Y, et al. 2026. FreeTacMan: robot-free visuo-tactile data collection system for contact-rich manipulation. In *2026 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
105. Xu Y, Wei L, An P, Zhang Q, Li YL. 2025. exUMI: extensible robot teaching system with action-aware task-agnostic tactile representation. In *Proceedings of the 9th Conference on Robot Learning*, ed. J Lim, S Song, HW Park. Proceedings of Machine Learning Research 305. PMLR
106. Wu T, Li J, Zhang J, Wu M, Dong H. 2025. Canonical representation and force-based pretraining of 3D tactile for dexterous visuo-tactile policy learning. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
107. Zhao J, Ma Y, Wang L, Adelson E. 2025. Transferable tactile transformers for representation learning across diverse sensors and tasks. In *Proceedings of the 8th Conference on Robot Learning*, ed. P Agrawal, O Kroemer, W Burgard. Proceedings of Machine Learning Research 270. PMLR
108. Guzey I, Evans B, Chintala S, Pinto L. 2023. Dexterity from touch: self-supervised pre-training of tactile representations with robotic play. In *Proceedings of the 7th Conference on Robot Learning*, ed. J Tan, M Toussaint, K Darvish. Proceedings of Machine Learning Research 229. PMLR
109. Xu Z, Uppuluri R, Zhang X, Fitch C, Crandall PG, et al. 2025. UniT: data efficient tactile representation with generalization to unseen objects. *IEEE Robot. Autom. Lett.* 10(6):5481–88
110. Higuera C, Sharma A, Bodduluri CK, Fan T, Lancaster P, et al. 2025. Sparsh: self-supervised touch representations for vision-based tactile sensing. In *Proceedings of the 8th Conference on Robot Learning*, ed. P Agrawal, O Kroemer, W Burgard. Proceedings of Machine Learning Research 270. PMLR
111. Higuera C, Sharma A, Fan T, Bodduluri CK, Boots B, et al. 2025. Tactile beyond pixels: multisensory touch representations for robot manipulation. In *Proceedings of the 9th Conference on Robot Learning*, ed. J Lim, S Song, HW Park. Proceedings of Machine Learning Research 305. PMLR
112. Zheng Y, Gu S, Li W, Zheng Y, Zang Y, et al. 2026. OmniVTA: visuo-tactile world modeling for contact-rich robotic manipulation. Preprint, arXiv:2603.19201v3 [cs.RO]
113. Hou Y, Liu Z, Chi C, Cousineau E, Kuppuswamy N, et al. 2025. Adaptive compliance policy: learning approximate compliance for diffusion guided control. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE

114. Zhao TZ, Kumar V, Levine S, Finn C. 2023. Learning fine-grained bimanual manipulation with low-cost hardware. In *Robotics: Science and Systems XIX*, ed. K Bekris, K Hauser, S Herbert, J Yu. Robotics: Science and Systems Foundation
115. Johansson RS, Flanagan JR. 2009. Coding and use of tactile signals from the fingertips in object manipulation tasks. *Nat. Rev. Neurosci.* 10(5):345–59
116. Xu X, Hou Y, Liu Z, Song S. 2025. Compliant Residual DAGger: improving real-world contact-rich manipulation with human corrections. In *Advances in Neural Information Processing Systems 38*, ed. D Belgrave, C Zhang, H Lin, R Pascanu, P Koniusz, et al. Curran Associates
117. Villani L, De Schutter J. 2016. Force control. In *Springer Handbook of Robotics*, ed. B Siciliano, O Khatib. Springer
118. Chen C, Yu Z, Choi H, Cutkosky M, Bohg J. 2025. DexForce: extracting force-informed actions from kinesthetic demonstrations for dexterous manipulation. *IEEE Robot. Autom. Lett.* 10(6):6416–23
119. Choi H, Hou Y, Pan C, Hong S, Patel A, et al. 2026. In-the-wild compliant manipulation with UMI-FT. In *2026 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
120. Kamijo T, Beltran-Hernandez CC, Hamaya M. 2024. Learning variable compliance control from a few demonstrations for bimanual robot with haptic feedback teleoperation system. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE
121. Zhou B, Jiao R, Li Y, Yuan X, Fang F, et al. 2025. Admittance visuomotor policy learning for general-purpose contact-rich manipulations. *IEEE Trans. Ind. Electron.* 72(11):11520–29
122. Zhang Z, Ma J, Yang X, Wen X, Zhang Y, et al. 2026. TouchGuide: inference-time steering of visuomotor policies via touch guidance. In *Robotics: Science and Systems XXII*, ed. M Ciocarlie, L Carlone, J Strader. Robotics: Science and Systems Foundation
123. Chebotar Y, Hausman K, Su Z, Sukhatme GS, Schaal S. 2016. Self-supervised regrasping using spatio-temporal tactile features and reinforcement learning. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE
124. van Hoof H, Hermans T, Neumann G, Peters J. 2015. Learning robot in-hand manipulation with tactile features. In *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*. IEEE
125. Dong S, Jha DK, Romeres D, Kim S, Nikovski D, et al. 2021. Tactile-RL for insertion: generalization to objects of unknown geometry. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
126. Lee MA, Zhu Y, Zachares P, Tan M, Srinivasan K, et al. 2020. Making sense of vision and touch: learning multimodal representations for contact-rich tasks. *IEEE Trans. Robot.* 36(3):582–96
127. Church A, Lloyd J, Hadsell R, Lepora NF. 2020. Deep reinforcement learning for tactile robotics: learning to type on a braille keyboard. *IEEE Robot. Autom. Lett.* 5(4):6145–52
128. Aslam S, Kumar K, Zhou P, Yu H, Wang MY, et al. 2025. DartBot: overhand throwing of deformable objects with tactile sensing and reinforcement learning. *IEEE Trans. Autom. Sci. Eng.* 22:13644–61
129. Andrychowicz OM, Baker B, Chociej M, Józefowicz R, McGrew B, et al. 2020. Learning dexterous in-hand manipulation. *Int. J. Robot. Res.* 39(1):3–20
130. Khandate G, Shang S, Chang ET, Saidi TL, Adams J, et al. 2023. Sampling-based exploration for reinforcement learning of dexterous manipulation. In *Robotics: Science and Systems XIX*, ed. K Bekris, K Hauser, S Herbert, J Yu. Robotics: Science and Systems Foundation
131. Zhao Y, Qian K, Duan B, Luo S. 2024. FOTS: a fast optical tactile simulator for Sim2Real learning of tactile-motor robot manipulation skills. *IEEE Robot. Autom. Lett.* 9(6):5647–54
132. Lei K, Li H, Yu D, Wei Z, Guo L, et al. 2026. Performant robotic manipulation with real-world reinforcement learning. *Sci. Robot.* 11(116):eae6267
133. Yu K, Han Y, Wang Q, Saxena V, Xu D, et al. 2025. MimicTouch: leveraging multi-modal human tactile demonstrations for contact-rich manipulation. In *Proceedings of the 8th Conference on Robot Learning*, ed. P Agrawal, O Kroemer, W Burgard. Proceedings of Machine Learning Research 270. PMLR
134. Zhao Z, Haldar S, Cui J, Pinto L, Bhirangi R. 2026. Touch begins where vision ends: generalizable policies for contact-rich manipulation. *IEEE Robot. Autom. Lett.* <https://doi.org/10.1109/LRA.2026.3730288>
135. Huang B, Li Y. 2026. FlexiTac: a low-cost, open-source, scalable tactile sensing solution for robotic systems. Preprint, arXiv:2604.28156v1 [cs.RO]

136. Tang T, Ji X, Xing W, Hao C, Xu W, et al. 2026. Towards human-like manipulation through RL-augmented teleoperation and mixture-of-dexterous-experts VLA. Preprint, arXiv:2603.08122v1 [cs.RO]
137. Miller E, McInroe T, Abel D, Mac Aodha O, Vijayakumar S. 2025. Enhancing tactile-based reinforcement learning for robotic control. In *Advances in Neural Information Processing Systems 38*, ed. D Belgrave, C Zhang, H Lin, R Pascanu, P Koniusz, et al. Curran Associates
138. Zitkovich B, Yu T, Xu S, Xu P, Xiao T, et al. 2023. RT-2: Vision-language-action models transfer web knowledge to robotic control. In *Proceedings of the 7th Conference on Robot Learning*, ed. J Tan, M Toussaint, K Darvish. Proceedings of Machine Learning Research 229. PMLR
139. Black K, Brown N, Darpinian J, Dhabalia K, Driess D, et al. 2025. $\pi_{0.5}$: a vision-language-action model with open-world generalization. In *Proceedings of the 9th Conference on Robot Learning*, ed. J Lim, S Song, HW Park. Proceedings of Machine Learning Research 305. PMLR
140. Jones J, Mees O, Sferrazza C, Stachowicz K, Abbeel P, et al. 2025. Beyond sight: finetuning generalist robot policies with heterogeneous sensors via language grounding. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
141. Liu Z, Liu J, Xu J, Han N, Gu C, et al. 2026. MLA: a multisensory language-action model for multimodal understanding and forecasting in robotic manipulation. In *2026 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
142. Huang J, Wang S, Lin F, Hu Y, Wen C, et al. 2025. Tactile-VLA: unlocking vision-language-action model's physical knowledge for tactile generalization. Preprint, arXiv:2507.09160v1 [cs.RO]
143. Zhang K, Zhang H, Xu Z, Zhang Z, Prince MRI, et al. 2026. TacVLA: contact-aware tactile fusion for robust vision-language-action manipulation. In *2026 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. Forthcoming
144. Zhao R, Wang W, Ma Y, Li X, Tay FE, et al. 2026. FD-VLA: force-distilled vision-language-action model for contact-rich manipulation. In *2026 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
145. Zhang Z, Xu H, Yang Z, Yue C, Lin Z, et al. 2025. Elucidating the design space of torque-aware vision-language-action models. In *Proceedings of the 9th Conference on Robot Learning*, ed. J Lim, S Song, HW Park. Proceedings of Machine Learning Research 305. PMLR
146. Niu D, Liu Z, Wang Z, Shao B, Yin ZH, et al. 2026. T-Rex: tactile-reactive dexterous manipulation. In *Proceedings of the 10th Conference on Robot Learning*. Forthcoming
147. Li X, Cai M, Xu J, Zhu J, Fan H, et al. 2026. AT-VLA: adaptive tactile injection for enhanced feedback reaction in vision-language-action models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE
148. Shazeer N, Mirhoseini A, Maziarz K, Davis A, Le Q, et al. 2017. Outrageously large neural networks: the sparsely-gated mixture-of-experts layer. In *5th International Conference on Learning Representations*. OpenReview.net. <https://openreview.net/forum?id=B1ckMDqlg>
149. Yu J, Liu H, Yu Q, Ren J, Hao C, et al. 2025. ForceVLA: enhancing VLA models with a force-aware MoE for contact-rich manipulation. In *Advances in Neural Information Processing Systems 38*, ed. D Belgrave, C Zhang, H Lin, R Pascanu, P Koniusz, et al. Curran Associates
150. Li Y, Zhaxizhuoma Z, Jiang H, Xia J, Zhang H, et al. 2026. ForceVLA2: unleashing hybrid force-position control with force awareness for contact-rich manipulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE
151. Bi J, Ma KY, Hao C, Zheng MS, Soh H. 2026. VLA-Touch: enhancing vision-language-action model with dual-level tactile feedback. *IEEE Robot. Autom. Lett.* 11(7):8487–94
152. Yu S, Kelvin L, Xiao A, Duan J, Soh H. 2024. Octopi: object property reasoning with large tactile-language models. In *Robotics: Science and Systems XX*, ed. D Kulic, G Venture, K Bekris, E Coronado. Robotics: Science and Systems Foundation
153. Yuan C, Zhang Z, Zhou M, Chen W, Wang Y, et al. 2026. FTP-1: a generalist foundation tactile policy across tactile sensors for contact-rich manipulation. In *Proceedings of the 10th Conference on Robot Learning*. Forthcoming
154. Richalet J, Rault A, Testud J, Papon J. 1978. Model predictive heuristic control: applications to industrial processes. *Automatica* 14(5):413–28

155. Mayne D, Rawlings J, Rao C, Scokaert P. 2000. Constrained model predictive control: stability and optimality. *Automatica* 36(6):789–814
156. Tian S, Ebert F, Jayaraman D, Mudigonda M, Finn C, et al. 2019. Manipulation by feel: touch-based control with deep predictive models. In *2019 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
157. Sutanto G, Ratliff N, Sundaralingam B, Chebotar Y, Su Z, et al. 2019. Learning latent space dynamics for tactile servoing. In *2019 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE
158. Xu Z, She Y. 2024. LeTac-MPC: learning model predictive control for tactile-reactive grasping. *IEEE Trans. Robot.* 40:4376–95
159. Guo Y, Shi L, Chen J, Finn C. 2026. Ctrl-World: a controllable generative world model for robot manipulation. In *14th International Conference on Learning Representations*. OpenReview.net. <https://openreview.net/forum?id=748bHL2BAv>
160. Wu P, Escontrela A, Hafner D, Abbeel P, Goldberg K. 2023. DayDreamer: world models for physical robot learning. In *Proceedings of the 6th Conference on Robot Learning*, ed. K Liu, D Kulic, J Ichnowski. Proceedings of Machine Learning Research 205. PMLR
161. Ai B, Tian S, Shi H, Wang Y, Tan C, et al. 2024. RoboPack: learning tactile-informed dynamics models for dense packing. In *Robotics: Science and Systems XX*, ed. D Kulic, G Venture, K Bekris, E Coronado. Robotics: Science and Systems Foundation
162. Higuera C, Arnaud S, Boots B, Mukadam M, Hogan FR, et al. 2026. Visuo-tactile world models. Preprint, arXiv:2602.06001v1 [cs.RO]
163. Heng L, Geng H, Zhang K, Abbeel P, Malik J. 2026. ViTacFormer: learning cross-modal representation for visuo-tactile dexterous manipulation. In *Robotics: Science and Systems XXII*, ed. M Ciocarlie, L Carlone, J Strader. Robotics: Science and Systems Foundation
164. Ye G, Zhang Z, Zhao X, Wu S, Lu H, et al. 2026. Learning to feel the future: DreamTacVLA for contact-rich manipulation. Preprint, arXiv:2512.23864v4 [cs.RO]
165. Lou Y, Ye Y, Fu Y, Cen J, Chi X, et al. 2026. Dream-Tac: a unified tactile world action model for contact-rich robot manipulation. Preprint, arXiv:2606.08737v1 [cs.RO]
166. Niu Y, Fang Z, Chen B, Zhou S, Senthilkumaran RK, et al. 2026. Learning versatile humanoid manipulation with touch dreaming. In *2026 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. Forthcoming
167. Feng R, Hu J, Xia W, Gao T, Shen A, et al. 2025. AnyTouch: learning unified static-dynamic representation across multiple visuo-tactile sensors. In *13th International Conference on Learning Representations*. OpenReview.net. <https://openreview.net/forum?id=XToAemis1h>
168. Gupta H, Mo Y, Jin S, Yuan W. 2025. Sensor-invariant tactile representation. In *13th International Conference on Learning Representations*. OpenReview.net. <https://openreview.net/forum?id=RnJY9WcpA3>
169. Rodriguez S, Dou Y, Oller M, Owens A, Fazeli N. 2025. Cross-sensor touch generation. In *Proceedings of the 9th Conference on Robot Learning*, ed. J Lim, S Song, HW Park. Proceedings of Machine Learning Research 305. PMLR
170. Chen Z, Ou N, Zhang X, Wu Z, Zhao Y, et al. 2026. Training tactile sensors to learn force sensing from each other. *Nat. Commun.* 17:2101
171. Luu QK, Zhou P, Xu Z, Zhang Z, Qiu Q, et al. 2026. ManiFeel: benchmarking and understanding visuotactile manipulation policy learning. Preprint, arXiv:2505.18472v2 [cs.RO]
172. Chen B, Wan W, Chen T, Guo X, Xu C, et al. 2026. UniVTAC: a unified simulation platform for visuo-tactile manipulation data generation, learning, and benchmarking. Preprint, arXiv:2602.10093v1 [cs.RO]
173. Seminara L, Gastaldo P, Watt SJ, Valyear KF, Zuher F, et al. 2019. Active haptic perception in robots: a review. *Front. Neurobot.* 13:53