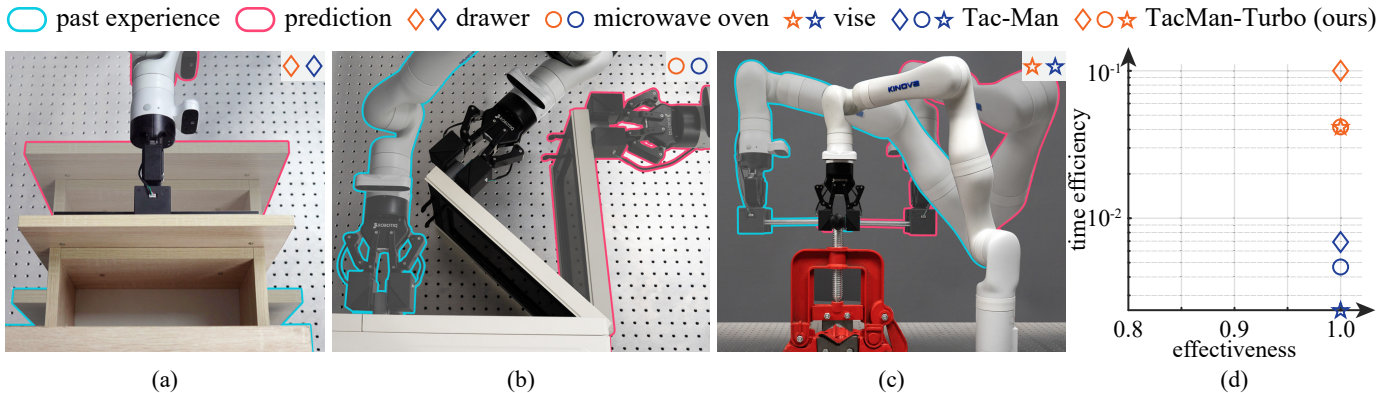


# TacMan-Turbo: Proactive Tactile Control for Robust and Efficient Articulated Object Manipulation

Zihang Zhao<sup>✉</sup>, Zhenghao Qi<sup>✉</sup>, Yuyang Li<sup>✉</sup>, Leiya Cui<sup>✉</sup>, Zhi Han<sup>✉</sup>, Lecheng Ruan<sup>✉</sup>, and Yixin Zhu<sup>✉</sup>



**Fig. 1: Robust and efficient tactile-informed manipulation of articulated objects.** TacMan-Turbo enables robots to achieve both effective and efficient manipulation of articulated objects through a novel tactile-based control framework. Unlike previous approaches, TacMan-Turbo leverages past tactile experience to extract local kinematic information, allowing the controller to predict optimal interaction positions and make proactive adjustments—all without requiring pre-provided object kinematics. This capability significantly enhances manipulation efficiency while maintaining robustness across diverse scenarios. Panels (a-c) demonstrate TacMan-Turbo’s versatility across various joint mechanisms: (a) manipulating a prismatic-joint drawer in just 10 seconds, (b) operating a revolute-joint microwave oven in 24 seconds, and (c) controlling a complex vise with a helical handle path in 24 seconds. Panel (d) quantifies these improvements, showing that while maintaining 100% effectiveness, our method achieves approximately 15×, 11×, and 18× higher time efficiency than the recent tactile-informed method, Tac-Man [1], as measured by average manipulation speeds. These order-of-magnitude improvements in efficiency without compromising reliability mitigate the longstanding trade-off between effectiveness and efficiency in articulated object manipulation.

**Abstract**—Adept manipulation of articulated objects is essential for robots to operate successfully in human environments. Such manipulation requires both effectiveness—reliable operation despite uncertain object structures—and efficiency—swift execution with minimal redundant steps and smooth trajectories.

This work is supported in part by the Brain Science and Brain-like Intelligence Technology–National Science and Technology Major Project (2025ZD0219400), the National Natural Science Foundation of China (62376009), the State Key Lab of General AI at Peking University, the PKU-Bingji Joint Laboratory for Artificial Intelligence, the Wuhan Major Scientific and Technological Special Program (2025060902020304), the Hubei Embodied Intelligence Foundation Model Research and Development Program, and the National Comprehensive Experimental Base for Governance of Intelligent Society, Wuhan East Lake High-Tech Development Zone. (Zihang Zhao and Zhenghao Qi contributed equally to this work. Corresponding authors: Zhi Han, Lecheng Ruan, and Yixin Zhu.)

Zihang Zhao, Zhenghao Qi, Yuyang Li, Lecheng Ruan, and Yixin Zhu are with the Peking University, Beijing 100871, China

Zihang Zhao, Zhenghao Qi, Yuyang Li, and Yixin Zhu are also with the State Key Lab of General AI, Peking University, and the Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing 100871, China.

Zihang Zhao is also with the LeapZenith AI Research, Shanghai 201707, China (email: zhaozihang@stu.pku.edu.cn).

Zhenghao Qi is also with the Tsinghua University, Beijing 100084, China (emails: qi-zh21@mails.tsinghua.edu.cn).

Yuyang Li and Yixin Zhu are also with the Embodied Intelligence Lab, PKU-Wuhan Institute for Artificial Intelligence (email: y.li@stu.pku.edu.cn; yixin.zhu@pku.edu.cn)

Leiya Cui is with the University of Chinese Academy of Sciences, Beijing 100049, China, and interned at the Institute for Artificial Intelligence, Peking University, Beijing 100871, China (email: cuileiyao24@mails.ucas.ac.cn).

Zhi Han is with the University of Chinese Academy of Sciences, Beijing 100049, China (email: hanzhi@sia.cn).

Data is available online at <https://tacman-aom.github.io>.

Existing approaches struggle to achieve both objectives simultaneously: methods relying on predefined kinematic models lack robustness when encountering structural variations, while the tactile-informed approach achieves robust manipulation but sacrifices efficiency through reactive, step-by-step execute-and-recover cycles. To address this challenge, this paper introduces TacMan-Turbo, a proactive tactile control framework that unifies the cycles into a continuous control loop. Our key insight is to interpret tactile signals temporally in addition to spatially: instead of treating contact deviations merely as instantaneous error signals requiring immediate compensation, we analyze sequential tactile observations to reveal local kinematic information. This temporal perspective enables our controller to predict future object states and proactively modulate manipulation velocities, eliminating the previous execute-and-recover cycle. In evaluations across 200 diverse simulated articulated objects and real-world experiments, our approach maintains a near-perfect performance while significantly enhancing time efficiency, action efficiency, and trajectory smoothness (all p-values < 0.0001). These results demonstrate that tactile feedback serves a dual purpose: providing not only spatial error signals for reactive control but also critical temporal structural information that enables proactive prediction and control. By enabling robots to manipulate articulated objects both reliably and efficiently, this work advances robot capabilities for seamless operation in dynamic, human-centric environments.

**Note to Practitioners**—Robots operating in human-centric environments need to adeptly manipulate articulated objects (e.g., drawers and doors), despite uncertain or inaccurate kinematic models, while finishing tasks quickly and smoothly without causing damage. Traditional planners that assume known kinematics are brittle in real-world settings, and recent tactile-reactive controllers, though robust, are slow due to repeated

execute-and-recover cycles. TacMan-Turbo introduces a proactive tactile control framework built on a temporal interpretation of tactile signals: by extracting local kinematic information from sequential tactile observations, the controller predicts short-horizon handle motion and proactively adjusts end-effector velocity to track these predictions, eliminating dedicated recovery phases. In evaluations across 200 simulated articulated objects and real-world experiments, the method achieved near-perfect performance while reducing task completion time by nearly an order of magnitude compared to a strong tactile-based baseline, delivering higher action efficiency and smoother joint trajectories. Overall, TacMan-Turbo provides a training-free, drop-in upgrade for tactile manipulation of articulated objects, while preserving robustness and delivering faster, smoother, and more efficient operation without prior object kinematic models.

*Index Terms*—Proactive control, temporary interpretation of tactile signals, manipulation of articulated objects, and efficient manipulation.

## I. INTRODUCTION

ADAPT manipulation of articulated objects, from doors and drawers to various household items, constitutes one of the basic capabilities that robots must master to function effectively in human-centric environments [2, 3]. Success in this domain hinges on two critical objectives: *effectiveness*—the reliable manipulation of objects despite uncertain internal structures, and *efficiency*—the execution of tasks with minimal redundant movements and smooth trajectories [4, 5].

Current approaches struggle to achieve both objectives simultaneously. Methods that rely on pre-provided kinematic information demonstrate efficiency in controlled scenarios but often fail when encountering discrepancies between assumed and actual object kinematics. Conversely, while the recent tactile-informed approach by Zhao *et al.* [1] successfully achieves robust manipulation despite kinematic uncertainty, it sacrifices efficiency—requiring several minutes to manipulate a single object due to its cautious, step-by-step exploration-compensation mechanism. This fundamental trade-off between effectiveness and efficiency remains unresolved in articulated object manipulation, as further detailed in Sec. II-A.

To address this challenge, we propose TacMan-Turbo, a novel proactive control framework that fundamentally reimagines how tactile feedback can guide manipulation. While leveraging Zhao *et al.*'s insight in using robot-object contact deviations as state representations [1], TacMan-Turbo introduces a temporal perspective: instead of treating contact deviations merely as spatial errors requiring immediate compensation, we analyze sequential tactile observations as time-series data that reveal local kinematic information. This temporal interpretation enables TacMan-Turbo to predict future object states and proactively modulate manipulation velocities to minimize predicted future deviations, streamlining the previous execute-and-recover cycle into a continuous control loop. As illustrated in Fig. 1, this novel approach preserves the effectiveness of contact-based manipulation while enabling substantially more efficient execution.

We validate TacMan-Turbo in both simulation and real-world settings. Our test scenarios systematically cover articulated objects with diverse joint types—from basic prismatic

and revolute joints to complex mechanisms involving concurrent translational and rotational movements. Across 200 simulated articulated objects and multiple real-world experiments, TacMan-Turbo achieves the same perfect manipulation success rate as the previous tactile-informed approach while significantly improving all efficiency metrics: time efficiency, action efficiency, and trajectory smoothness (all comparisons yielding p-values < 0.0001). These results confirm that TacMan-Turbo successfully mitigates the long-standing effectiveness-efficiency trade-off in articulated object manipulation.

Our main contributions are summarized below:

- **A temporal interpretation of tactile signals for local kinematic inference:** Rather than treating contact deviations merely as spatial errors requiring immediate compensation, we introduce a temporal perspective that analyzes sequential tactile observations as time-series data. This approach extracts local kinematic information—revealing object motion patterns from contact history—and enables prediction of future object states without requiring pre-provided kinematic models.
- **A proactive tactile control framework that unifies manipulation into a continuous process:** Building on our predictive capability, we reformulate articulated object manipulation from a reactive pose-correction problem (minimizing current contact deviations) to a proactive velocity-modulation problem (minimizing predicted future deviations). This reformulation streamlines the execute-and-recover cycle into a unified process, significantly enhancing the manipulation efficiency without compromising the effectiveness.

The remainder of this paper is organized as follows: Sec. II positions our work within the broader context of robotic manipulation research. Sec. III presents the theoretical framework and implementation details of TacMan-Turbo. Secs. IV and V evaluate TacMan-Turbo through simulation studies and real-world experiments, respectively. Sec. VI analyzes TacMan-Turbo's performance characteristics and discusses future research directions. Finally, Sec. VII summarizes our findings and contributions.

## II. RELATED WORK

This section contextualizes TacMan-Turbo's contributions within the broader landscape of robotic research. We begin by assessing current approaches to articulated object manipulation, establishing the fundamental trade-off between effectiveness and efficiency that motivates our work (Sec. II-A). In Sec. II-B, we explore contact modeling techniques, highlighting how the proposed method builds upon and advances previous work to address the effectiveness-efficiency trade-off. Next, Sec. II-C examines contact sensing technologies, categorizing them into force-based and geometry-based approaches while explaining their compatibility with our method. Finally, Sec. II-D reviews contact control strategies, tracing their evolution from early reactive feedback systems to our proposed proactive tactile control frameworks.

### A. Articulated Object Manipulation

Approaches to articulated object manipulation can be categorized based on their strategies for addressing the fundamental challenges of effectiveness and efficiency. Traditional methods rely on predefined kinematic models provided manually [6, 7]. While these approaches enable efficient manipulation through precise motion planning in controlled environments, they significantly restrict system autonomy by requiring accurate prior knowledge of object kinematics—limiting their effectiveness in real-world scenarios where such information is often unavailable or inaccurate.

To overcome this dependency on manually provided information, subsequent research has focused on the autonomous acquisition of object kinematics. A prominent line of work leverages visual perception to infer kinematic models through geometric reasoning and motion analysis [8–14]. However, visual information alone often proves insufficient, as visually similar objects may possess fundamentally different internal mechanisms [1, 10, 15]. This inherent limitation has led researchers to integrate additional sensing modalities to resolve ambiguity. By combining multi-frame analysis with wrist-mounted force/torque sensors, more accurate parameter estimation of kinematic models becomes possible through active interaction [4, 5, 16–21]. Nevertheless, these methods remain constrained by their underlying assumption that objects conform to predefined joint types (*e.g.*, revolute, prismatic), limiting their effectiveness when encountering complex or unconventional mechanisms.

The persistent challenges in model-based approaches have driven the development of learning-based techniques for articulated object manipulation. Imitation learning enables robots to acquire implicit manipulation strategies from human demonstrations, typically facilitated through teleoperation systems [22–29]. This approach naturally captures human manipulation expertise but faces significant challenges in collecting sufficiently diverse and high-quality demonstration datasets [30]. While transfer learning techniques have improved generalization to similar objects [31, 32], extending these methods to handle arbitrary articulated objects remains challenging due to the vast diversity of possible kinematic structures.

Reinforcement learning offers an alternative paradigm that attempts to overcome the data collection bottleneck by developing manipulation strategies through repeated interactions in simulated environments [33–38]. This approach has been significantly advanced by the creation of large-scale datasets of articulated objects [39–43], enabling training across diverse object categories. However, the inevitable sim-to-real gap often leads to degraded performance when these methods are deployed in real-world settings, particularly when encountering previously unseen object variations.

The limitations in both model-based and learning-based approaches recently led to the development of tactile-informed manipulation. Zhao *et al.* [1] introduced an approach that achieves reliable manipulation of articulated objects by compensating for robot-object contact deviations without requiring kinematic priors. While this method demonstrates robust

effectiveness with unknown objects, its reactive execute-and-recover cycle results in significant time overhead. TacMan-Turbo advances this foundation through a proactive control framework that unifies the execute-and-recover cycle into a continuous control loop, significantly enhancing manipulation efficiency.

### B. Contact Modeling

Contact modeling has emerged as a fundamental approach in robotics research, providing essential information about robot-environment interactions [44–48]. The effectiveness of contact-based strategies has been particularly demonstrated in robotic grasping, where accurate contact models enable reliable manipulation across diverse object geometries [49, 50].

Despite its proven success in grasping, contact modeling remains insufficiently explored for articulated object manipulation—a domain with inherently more complex interaction dynamics. Most existing approaches employ simplified contact assumptions, typically modeling fixed contact points between the robot and the object [6, 7, 11, 16–18, 51]. While this simplification enables efficient computation, it fundamentally fails to capture the dynamic nature of real-world interactions where contact points frequently shift during manipulation. Some progress has been made by relaxing these constraints to permit specific forms of relative motion between the robot and object [4, 5], enabling more effective control command generation. However, these approaches still rely on predefined kinematic models, limiting their effectiveness across diverse object types.

Zhao *et al.* [1] demonstrated that modeling contact as geometric deviations enables robust articulated object manipulation without requiring prior kinematic models. TacMan-Turbo advances this foundation by reinterpreting how these contact deviations are utilized: rather than treating them merely as spatial error signals requiring immediate compensation, we analyze sequential observations temporally to extract local kinematic information, enabling efficient proactive control.

### C. Contact Sensing

The successful implementation of TacMan-Turbo relies on accurate sensing of contacts between robots and objects, leveraging decades of advances in tactile sensing technology [52, 53]. Based on their primary output format, tactile sensors can be categorized into two main types: force-based and geometry-based sensors, each offering distinct advantages for manipulation tasks.

Force-based sensors output contact forces by converting mechanical interactions into electrical signals through various transduction mechanisms, including conductive fluid [54], resistance [55–58], capacitance [59–61], and others [62, 63]. Geometry-based sensors, conversely, directly capture the geometric characteristics of contact interfaces, with vision-based tactile sensors being prominent examples [64–71].

Since TacMan-Turbo’s core functionality centers on contact geometry modeling for local kinematic information extraction, it is naturally compatible with geometry-based sensors that

directly provide the required information. Although force-based sensors can also be effectively utilized, they require an additional calibration step using Hooke’s law [72] to convert force measurements into geometric information [1]. To demonstrate TacMan-Turbo’s capabilities, in this paper, we implement our system using the GelSight sensor [64], a well-established geometry-based tactile sensing technology that provides sufficient geometric information about contact surfaces.

#### D. Contact Control

Contact control, also known as tactile control, has evolved into a powerful strategy for robotic manipulation by modulating robot motion based on real-time tactile feedback. This approach was pioneered by Weiss *et al.* [73] and subsequently advanced by Berger and Khosla [74], who provided the first practical implementation demonstrating its potential for precise interaction tasks.

The fundamental challenge in tactile control lies in designing algorithms capable of effectively transitioning from current tactile observations to desired reference states. To address this challenge, researchers have developed two primary approaches: feedback control and direct mapping. Feedback control employs various controllers (*e.g.*, PID controllers) to systematically minimize the difference between observed and desired tactile features, requiring careful selection of task-specific control parameters [75–79]. Direct mapping, on the contrary, transforms tactile feature differences directly into robot motion commands without explicit parameter tuning. This approach was first introduced by Chen *et al.* [80], who proposed the concept of an inverse tactile Jacobian for 2D edge tracking using manually designed features. This foundational framework was subsequently refined and extended by numerous researchers, including Zhang *et al.* [81] and Sikka *et al.* [82], who demonstrated its versatility in different manipulation scenarios. More recently, Wilson *et al.* [83] advanced the field by introducing motion primitives for cable rearrangement tasks, providing a structured approach to handle deformable objects. Zhao *et al.* [1] made a contribution to this paradigm through their implementation of contact point registration, demonstrating robust manipulation of articulated objects despite their kinematic complexity.

The latest evolution in tactile control has seen the emergence of predictive frameworks that anticipate future tactile states rather than reacting only to current observations. For example, Xu *et al.* [84] employed learning-based techniques to predict future tactile states for reactive grasping. However, such data-driven approaches often face challenges in generalizing across diverse and dynamic tasks due to their dependence on training data distribution. In contrast, TacMan-Turbo introduces a model-free, training-free predictive framework grounded in analytical interpretation of tactile sequences. Rather than learning black-box prediction models from data, we extract local kinematic information directly from sequential tactile observations through geometric analysis. This analytical approach enables TacMan-Turbo to predict future object states and make proactive adjustments without prior training.

### III. THE PROACTIVE TACTILE CONTROL

At its core, TacMan-Turbo is a novel proactive tactile control framework designed for both effective and efficient manipulation of articulated objects. Unlike **reactive** approaches that respond only to current sensory information, TacMan-Turbo extracts local object kinematic patterns from past contact data to predict optimal subsequent interaction positions. This predictive capability enables proactive adjustments that systematically mitigate contact deviations before they significantly impact manipulation performance.

The development of this framework unfolds through four interconnected components. First, **Sec. III-A** establishes the essential mathematical notation and preliminaries that provide the foundation for our approach. Building on this mathematical framework, **Sec. III-B** formally defines the articulated object manipulation problem, specifically addressing the dual optimization objectives of effectiveness and efficiency. Next, **Sec. III-C** introduces our novel prediction framework that analyzes sequential patterns in contact deviations to forecast articulated object states during manipulation. Finally, **Sec. III-D** details how these kinematic predictions are translated into optimal gripper velocity adjustments while adhering to the robot’s physical constraints and ensuring consistent contact maintenance.

#### A. Notation

To establish a rigorous foundation for our formulation of effective and efficient manipulation, we define the following mathematical notation:

- $\mathbb{R}$  represents the set of real numbers, with  $\mathbb{R}^+$  denoting nonnegative reals. Matrix and vector spaces are denoted as  $\mathbb{R}^{m \times n}$  and  $\mathbb{R}^n$  respectively. The special orthogonal group  $SO(n)$  consists of rotation matrices with determinant 1, while the special Euclidean group  $SE(n)$  represents rigid body transformations. Additional sets use calligraphic letters (*e.g.*,  $\mathcal{C}$ ), with  $|\cdot|$  denoting set cardinality.
- Vectors are represented by bold lowercase letters (*e.g.*,  $\mathbf{q} \in \mathbb{R}^n$ ), with subscript  $i$  indicating the  $i$ -th element (*e.g.*,  $q_i$ ). The velocity vector  $\mathbf{u} \in \mathbb{R}^6$  specifically combines linear velocities  $\mathbf{v} = [v_x, v_y, v_z]^T \in \mathbb{R}^3$  and angular velocities  $\boldsymbol{\omega} = [\omega_x, \omega_y, \omega_z]^T \in \mathbb{R}^3$ , where  $\cdot^T$  denotes transpose.
- Matrices use bold uppercase letters (*e.g.*,  $\mathbf{M} \in \mathbb{R}^{m \times n}$ ), with  $m_{ij}$  denoting the entry at row  $i$  and column  $j$ . The trace operator is written as  $\text{tr}(\mathbf{M})$ , and  $\mathbf{I}_{n \times n}$  represents the  $n$ -dimensional identity matrix.
- For relative pose representation between frames  $\{i\}$  and  $\{j\}$ , we use position vector  $\mathbf{p}_i^j \in \mathbb{R}^3$  and rotation matrix  $\mathbf{R}_i^j \in SO(3)$ . When frame  $\{j\}$  is the world frame  $\{w\}$ , we omit the superscript. These components combine into the homogeneous transformation matrix:

$$\mathbf{T}_i^j = \begin{bmatrix} \mathbf{R}_i^j & \mathbf{p}_i^j \\ \mathbf{0} & 1 \end{bmatrix}.$$

#### B. Problem Formulation

We now formulate the mathematical framework for effective and efficient tactile-informed manipulation of articulated

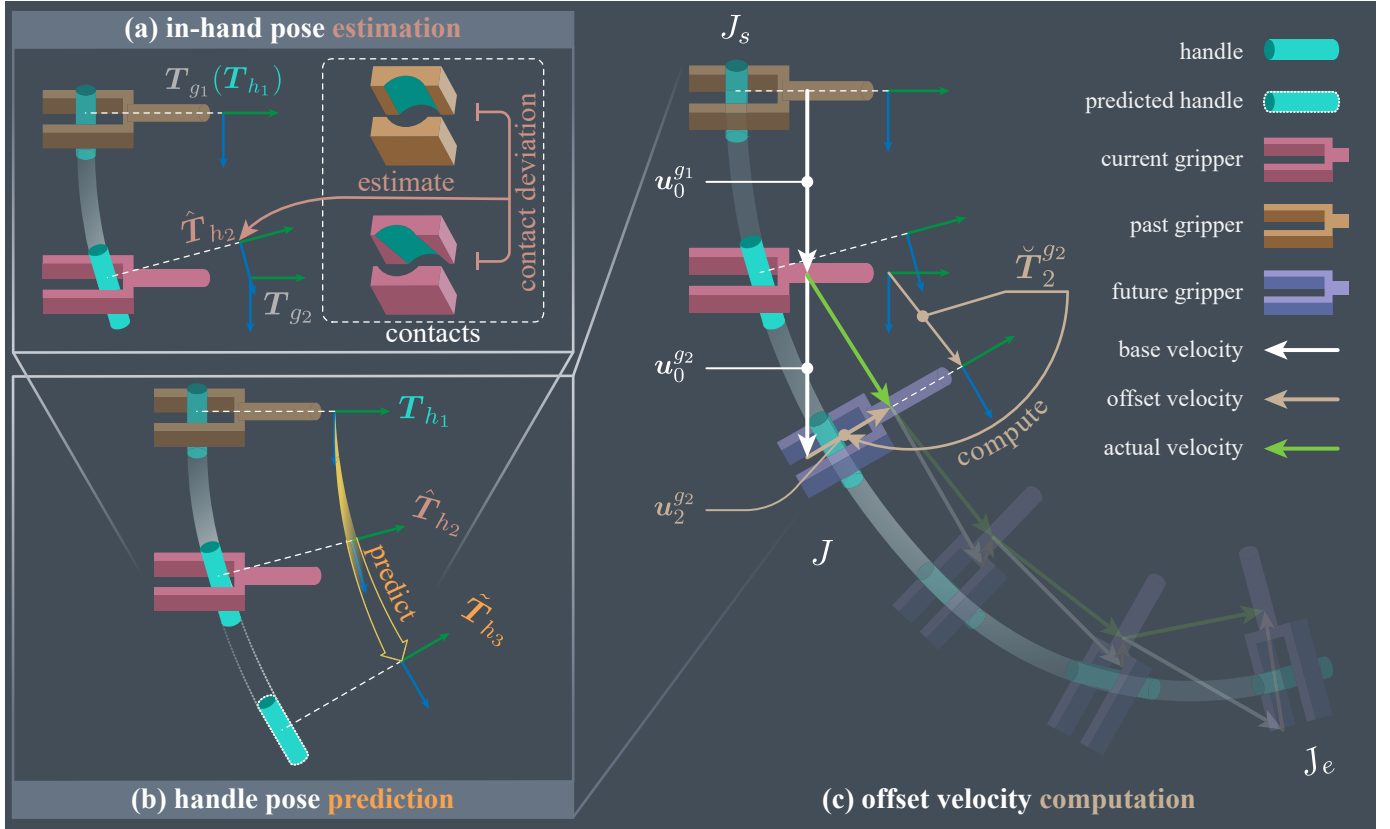


Fig. 2: **Schematic overview of proactive tactile proactive control framework, TacMan-Turbo.** Our framework integrates three sequential components that enable proactive manipulation: (a) **In-hand pose estimation** extracts tactile contact patterns from gripper sensors to determine the relative transformation between current gripper pose  $T_{g_i}$  and handle pose  $T_{h_i}$ . (b) **Handle pose prediction** analyzes sequential pose estimates ( $T_{h_1}, T_{h_2}$ ) to extract local kinematic patterns, enabling prediction of the future handle pose  $T_{h_3}$ . (c) **Offset velocity computation** generates the optimal control signal by calculating an offset velocity  $u_i^{g_i}$  that, when combined with base velocity  $u_0^{g_i}$ , creates a resultant motion  $u_0^{g_i} + u_i^{g_i}$  that aligns future gripper position with predicted handle motion along path  $J$ . By proactively adjusting for potential contact deviations, this approach ensures smooth, continuous manipulation from start point  $J_s$  to end point  $J_e$  without requiring separate correction phases.

objects. Consider an articulated object with a movable part and its rigidly attached “handle,” as illustrated in Fig. 2. The underlying articulation mechanism constrains the handle’s motion to a 1-degree of freedom (DoF) trajectory, represented as path  $J$  from initial point  $J_s$  to target point  $J_e$ . In our problem setting, only the initial position  $J_s$  is initially known, reflecting real-world scenarios where complete kinematic models are unavailable.

At the start of manipulation, a robot gripper at global pose  $T_{g_1}$  grasps the handle at pose  $T_{h_1}$ . For mathematical convenience and without loss of generality, we align the initial handle coordinate frame with that of the robot gripper:

$$T_{h_1} = T_{g_1}. \quad (1)$$

The physical interaction between the gripper and handle is characterized by their contact state, denoted by  $C_i$ . This state is represented geometrically as a set containing three-dimensional position vectors of activated contact points in the gripper’s coordinate frame, denoted as  $[p_x^{g_i}, p_y^{g_i}, p_z^{g_i}]^T$ . Contact points become activated when their normal deformation (illustrated along the  $x$ -axis in Fig. 2) exceeds the threshold

$$\epsilon: \quad C_i = \left\{ \left[ \begin{array}{c} p_x^{g_i} \\ p_y^{g_i} \\ p_z^{g_i} \end{array} \right] \mid |p_x^{g_i}| \geq \epsilon \right\}, \quad (2)$$

where the threshold  $\epsilon$  is adaptively selected in experiments, with implementation details provided in Sec. V-A.

The initial contact state  $C_1$  is assumed stable, providing a secure connection without material damage. Generally, any contact state  $C_i$  is defined through a mapping function  $f_c$  of handle and gripper poses, as the relative pose between these elements fully determines their interaction:

$$C_i = f_c(T_{h_i}, T_{g_i}). \quad (3)$$

In practice, this contact state is directly measured by tactile sensors mounted on the gripper.

Manipulation begins with a constant base velocity  $u_0^{g_i}$  in the gripper’s coordinate system, aligned with a known rough direction of the manipulation task. The acquisition of this base velocity and its impact on manipulation efficiency are examined in detail in Sec. VI.

Without prior knowledge of the articulation mechanism, the trajectories of the gripper and handle inevitably become misaligned during manipulation. This misalignment produces

contact deviations that may result in inefficient or even unstable interactions if not addressed. To mitigate this issue, an offset gripper velocity, denoted as  $\mathbf{u}_i^{g_i}$ , must be introduced to adapt the gripper's motion to the handle's constrained path.

The key challenge lies in determining this offset velocity. Efficiency considerations preclude simply halting motion to correct current contact deviations—instead, corrections must be planned proactively for subsequent steps, requiring anticipation of how the contact state will evolve. Therefore, the offset velocity cannot simply react to current deviations but must proactively minimize future contact deviations while maintaining the base velocity. Mathematically, this can be formulated as the following optimization problem:

$$\mathbf{u}_i^{g_i} = \arg \min_{\mathbf{u}_i^{g_i}} \sum_{(\mathbf{c}_1, \mathbf{c}_{i+1}) \in \mathcal{K}_{1,i+1}} \|\mathbf{c}_1 - \mathbf{c}_{i+1}\|_2, \quad (4)$$

where  $\mathcal{K}_{1,i+1}$  defines corresponding points between the reference (initial) and subsequent contact states:

$$\mathcal{K}_{1,i+1} = \{(\mathbf{c}_1, \mathbf{c}_{i+1}) \mid \mathbf{c}_1 \in \mathcal{C}_1, \mathbf{c}_{i+1} \in \mathcal{C}_{i+1}\}, \quad (5)$$

subject to the velocity limit constraint:

$$\mathbf{u}^m \leq \mathbf{u}_0^{g_i} + \mathbf{u}_i^{g_i} \leq \mathbf{u}^M, \quad (6)$$

where  $\leq$  denotes an element-wise "less than or equal to" relationship between vectors, while  $\mathbf{u}^m$  and  $\mathbf{u}^M$  represent the lower and upper bounds of the velocity that the robot can physically generate.

Given that the contact state depends solely on the poses of the gripper and handle (as established in Eq. (3)), maintaining stable contact requires aligning the gripper's motion with the handle's trajectory. Our solution addresses this fundamental challenge in two complementary steps: First, in Sec. III-C, we develop a method to predict the handle's next-step pose  $\mathbf{T}_{h_{i+1}}$  based on sequential contact patterns observed during interaction. Second, in Sec. III-D, we compute the control input  $\mathbf{u}_i^{g_i}$  that optimally aligns the gripper's motion with this predicted pose while ensuring compliance with the velocity limit constraints.

### C. Handle Pose Estimation and Prediction

To enable proactive control adjustments, our approach leverages past contact data to predict the optimal position for subsequent object interactions. This predictive capability builds upon two complementary components: first, estimating current handle poses from observed contact deviations and known gripper poses; second, leveraging the temporal sequence of these estimated poses to predict future handle positions. We begin by detailing our pose estimation approach.

At each time step, the handle pose  $\mathbf{T}_{h_i}$  relates to the gripper pose  $\mathbf{T}_{g_i}$  through their relative transformation:

$$\mathbf{T}_{h_i} = \mathbf{T}_{g_i} \mathbf{T}_{h_i}^{g_i}, \quad (7)$$

where  $\mathbf{T}_{h_i}^{g_i} \in SE(3)$  represents their relative homogeneous transformation. Initially at  $i = 1$ ,  $\mathbf{T}_{h_1}^{g_1} = \mathbf{I}_{4 \times 4}$  following our initial pose definition (Eq. (1)).

For subsequent time steps ( $i > 1$ ), assuming no slip between the sensor outer surface and the handle, any relative trans-

formation can be observed as deviations between the current contact state  $\mathcal{C}_i$  and the reference state  $\mathcal{C}_1$ . This relationship enables us to estimate  $\mathbf{T}_{h_i}^{g_i}$  through the following optimization:

$$\hat{\mathbf{T}}_{h_i}^{g_i} = \arg \min_{\hat{\mathbf{T}}_{h_i}^{g_i} \in SE(3)} \sum_{(\mathbf{c}_1, \mathbf{c}_i) \in \mathcal{K}_{1,i}} \left\| \hat{\mathbf{T}}_{h_i}^{g_i} \begin{bmatrix} \mathbf{c}_1 \\ 1 \end{bmatrix} - \begin{bmatrix} \mathbf{c}_i \\ 1 \end{bmatrix} \right\|_2, \quad (8)$$

where

$$\mathcal{K}_{1,i} = \{(\mathbf{c}_1, \mathbf{c}_i) \mid \mathbf{c}_1 \in \mathcal{C}_1, \mathbf{c}_i \in \mathcal{C}_i\} \quad (9)$$

defines the corresponding contact point pairs between the reference and current states.

This optimization problem (8) can be efficiently solved using the Kabsch algorithm [85], provided the contact state  $\mathcal{C}_i$  satisfies two key constraints:

First, it must contain at least three contact points:

$$|\mathcal{C}_i| \geq 3. \quad (10)$$

Second, these points must be non-collinear. Specifically, there must exist points  $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3 \in \mathcal{C}_i$  such that:

$$(\mathbf{p}_2 - \mathbf{p}_1) \times (\mathbf{p}_3 - \mathbf{p}_1) \neq \mathbf{0}. \quad (11)$$

Given these constraints, we can estimate the handle pose  $\hat{\mathbf{T}}_{h_i}$  at each time step  $i$  by combining the gripper pose with the estimated relative transformation:

$$\hat{\mathbf{T}}_{h_i} = \mathbf{T}_{g_i} \hat{\mathbf{T}}_{h_i}^{g_i}. \quad (12)$$

Building upon these estimated poses, we now focus on predicting future handle positions. Under the assumptions of locally smooth object trajectories and sufficiently high sampling frequencies, consecutive handle poses exhibit only slight variations, primarily driven by the object's instantaneous velocity. This local smoothness makes the constant velocity model a physically faithful approximation of the short-horizon dynamics [86–88]. Moreover, compared to higher-order alternatives (e.g., acceleration or jerk models), the constant-velocity formulation requires estimating fewer derivatives from finite-difference observations, which inherently reduces sensitivity to measurement noise. Based on this model, the next handle pose  $\tilde{\mathbf{T}}_{h_{i+1}}$  is predicted as:

$$\tilde{\mathbf{T}}_{h_{i+1}} = \hat{\mathbf{T}}_{h_i} \mathbf{T}_u, \quad (13)$$

where  $\mathbf{T}_u$  represents the velocity effects, which are assumed to remain constant over short time intervals. Given consistent sampling intervals,  $\mathbf{T}_u$  is computed from the two most recent pose estimates:

$$\mathbf{T}_u = (\hat{\mathbf{T}}_{h_{i-1}})^{-1} \hat{\mathbf{T}}_{h_i}. \quad (14)$$

While this constant velocity model effectively predicts handle poses, it may introduce minor residual errors relative to ground truth due to the omission of higher-order effects such as acceleration. However, these residual errors are detected by subsequent in-hand pose estimation and corrected during the next prediction step, preventing error accumulation. Moreover, for scenarios involving more dynamic motion or reduced sampling frequencies, this prediction framework can be extended to account for higher-order motion effects by incorporating additional past data—for example, transitioning to a constant

acceleration model by using one more observation to account for velocity changes.

As evident from Eq. (14), the proposed constant velocity prediction requires at least two previous observations ( $i \geq 2$ ). For the initial step ( $i = 1$ ), when past data is unavailable, the system relies solely on the base velocity  $\mathbf{u}_0^{g_1}$  defined in Sec. III-B. This integration of pose estimation and prediction provides the foundation for computing appropriate gripper velocities, which we address in the following section.

#### D. Offset Velocity Computation

Having predicted the next handle pose, we now develop a method for computing the offset gripper velocity that proactively compensates for contact deviations while maintaining operational efficiency. In the ideal scenario, the solution is straightforward: align the gripper's next pose precisely with the predicted handle pose to eliminate any potential contact deviation:

$$\mathbf{T}_{g_{i+1}} = \tilde{\mathbf{T}}_{h_{i+1}}. \quad (15)$$

This desired alignment requires the gripper to execute a specific transformation over a sampling interval  $\Delta t \in \mathbb{R}_{>0}$ . We compute this required transformation relative to the gripper's current coordinate frame:

$$\check{\mathbf{T}}_i^{g_i} = (\mathbf{T}_{g_i})^{-1} \mathbf{T}_{g_{i+1}}. \quad (16)$$

From this transformation, we derive the required velocity components:

$$\mathbf{u}_i^{g_i} = \frac{\log(\check{\mathbf{T}}_i^{g_i})^\vee}{\Delta t} - \mathbf{u}_0^{g_i}, \quad (17)$$

where  $\log(\cdot)$  denotes the logarithmic map and the operator  $\cdot^\vee$  transforms the skew-symmetric matrix into vector form.

While the above formulation provides the theoretically optimal velocity, practical implementation must account for the limitations imposed by robot hardware—specifically the maximum gripper velocities determined by the robot's joint motor limits, as defined in Eq. (6). The relationship between joint velocities  $\dot{\mathbf{q}}_i \in \mathbb{R}^n$  and the gripper velocity  $\mathbf{u}_i^{g_i}$  is given by:

$$\mathbf{J}(\mathbf{q}_i) \dot{\mathbf{q}}_i = \mathbf{u}_i^{g_i} + \mathbf{u}_0^{g_i}, \quad (18)$$

where  $\mathbf{J}(\mathbf{q}_i) \in \mathbb{R}^{6 \times n}$  is the robot Jacobian matrix. For analytical tractability, we assume  $n \geq 6$  and that  $\mathbf{J}(\mathbf{q}_i)$  maintains row full rank. Over small time intervals, treating  $\mathbf{q}_i$  as constant allows us to solve for joint velocities:

$$\dot{\mathbf{q}}_i = \mathbf{J}^\dagger(\mathbf{q}_i) (\mathbf{u}_i^{g_i} + \mathbf{u}_0^{g_i}), \quad (19)$$

using the pseudoinverse:

$$\mathbf{J}^\dagger(\mathbf{q}_i) = \mathbf{J}(\mathbf{q}_i)^\top (\mathbf{J}(\mathbf{q}_i) \mathbf{J}(\mathbf{q}_i)^\top)^{-1}. \quad (20)$$

To ensure joint velocities remain within their maximum limits  $\dot{\mathbf{q}}_M$ , we introduce a scaling factor  $\alpha$ :

$$\alpha = \min \left( \min \left( \left\{ \frac{\dot{q}_{M_j}}{\max(|\dot{q}_{i_j}|, \delta)} \beta \mid j = 1, \dots, n \right\}, 1 \right), 1 \right), \quad (21)$$

where  $\beta$  is a safety factor and  $\delta$  represents a small positive constant introduced to prevent division by zero. In this study,

we select  $\beta = 2$  and  $\delta = 1 \times 10^{-4}$  across all the experiments. This scaling yields the final executable offset velocity command:

$$\mathbf{u}_i^{g_i} = \alpha \mathbf{u}_i^{g_i} + (\alpha - 1) \mathbf{u}_0^{g_i} \quad (22)$$

with the base velocity also scaled by  $\alpha$ . Consequently, to achieve the desired handle position  $\tilde{\mathbf{T}}_{h_{i+1}}$  while respecting joint velocity limits, the execution duration must be adjusted from  $\Delta t$  to:

$$\Delta t' = \frac{\Delta t}{\alpha}. \quad (23)$$

While this scaling may marginally reduce manipulation efficiency, it represents a necessary adjustment to ensure the robot operates within its physical constraints, particularly when addressing large contact deviations that require substantial compensation.

## IV. SIMULATION STUDIES

To thoroughly evaluate TacMan-Turbo, we conducted extensive simulation studies focusing on both manipulation effectiveness and efficiency. This section presents our experimental approach and findings in four parts: we first detail our simulation setup (Sec. IV-A), establish comparative baselines (Sec. IV-B), define quantitative evaluation metrics (Sec. IV-C), and analyze the performance results (Sec. IV-D).

### A. Simulation Setup

Our evaluation framework integrates three core components designed to enable comprehensive assessment of manipulation capabilities under controlled yet realistic conditions.

**Simulation environment:** We implemented our experiments in NVIDIA Isaac Sim, operating at 60 Hz to ensure sufficient temporal resolution for capturing dynamic contact interactions. Critical to our study is the contact simulation method developed by Zhao *et al.* [1], which enables accurate modeling of tactile feedback by monitoring displacement patterns within designated contact regions. This approach captures the nuanced contact dynamics that emerge during the manipulation of articulated objects.

**Test articulated objects:** To systematically evaluate performance across varying kinematic structures, we assembled a three-category test suite with progressively increasing trajectory complexity:

- 50 objects with prismatic joints from PartNet-Mobility dataset [40] (Fig. 3(a)), featuring linear motion paths with varying physical characteristics.
- 50 objects with revolute joints also from PartNet-Mobility (Fig. 3(b)), representing rotational articulations that require continuous adjustment of manipulation direction.
- 100 custom-generated objects with Bézier curve trajectories designed to assess performance on complex, non-linear paths. This category includes 25 curves each of orders 2, 3, 4, and 5, sampled on a 2D plane and constrained to avoid self-intersections. Each curve defines a manipulation path on a 3D playboard, with a toy train positioned at the starting point serving as the handle (Fig. 3(c)).

Such a design from simple to complex kinematics provides systematic evaluation across a spectrum of trajectory

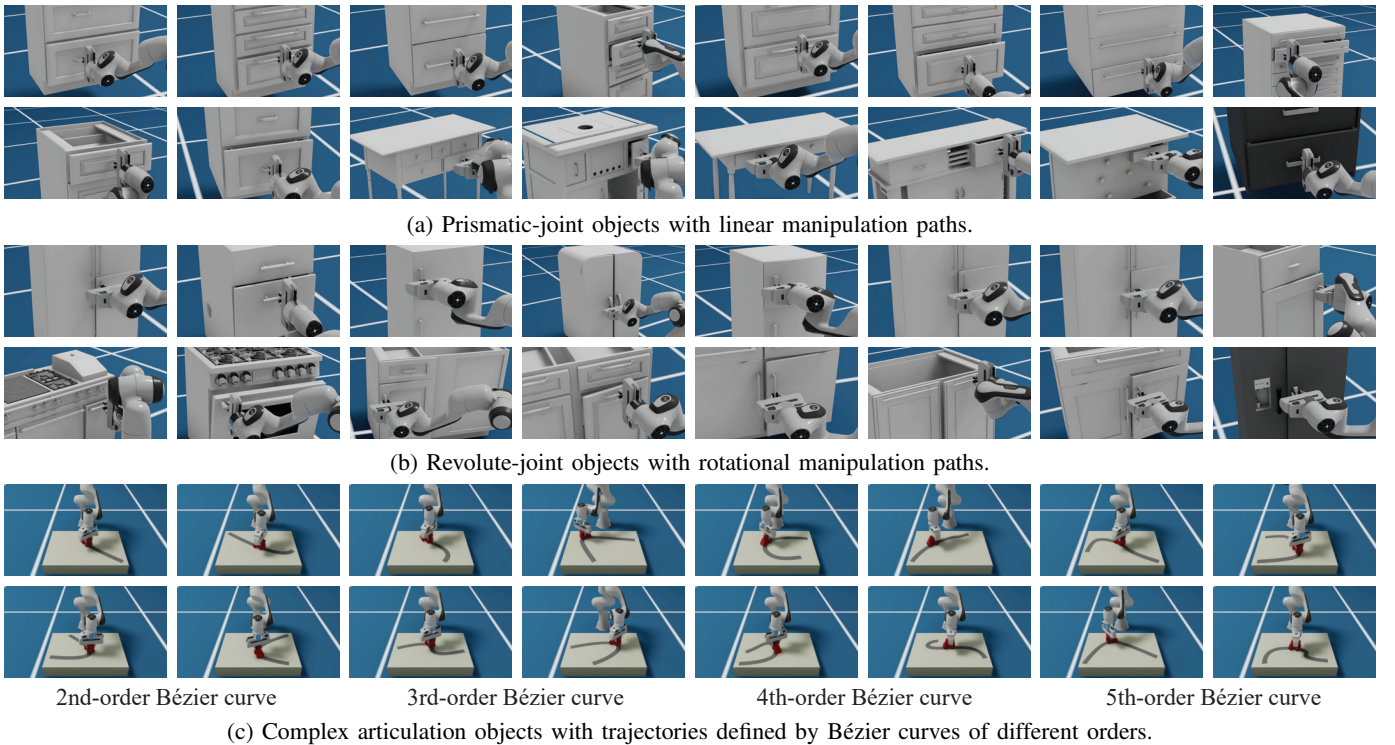


Fig. 3: **Simulation test environment and object categories.** Our evaluation encompasses three progressively complex articulated object categories: (a) 50 prismatic-joint objects featuring linear motion paths, representing the most kinematically straightforward mechanisms; (b) 50 revolute-joint objects with circular motion paths, introducing rotational complexity; and (c) 100 complex articulated objects with trajectories derived from Bézier curves of orders 2–5, simulating sophisticated real-world mechanisms with variable curvature. This systematically designed test suite enables rigorous performance assessment across a spectrum of kinematic complexity. Each case is standardized using initial robot configurations, ensuring fair comparison across manipulation strategies.

predictability and curvature characteristics. The generation process for the Bézier-based objects followed protocols in Zhao *et al.* [1], producing diverse trajectory complexities that challenge predictive capabilities in distinct ways.

**Robot:** Our experiments employ a 7-DoF Franka robotic arm, selected for its ability to execute full 6-DoF Cartesian motions—a capability essential for accommodating the diverse movement requirements of our test objects. We also integrate the B\* algorithm to place the robot, ensuring that hand trajectories during manipulation remain within the robot workspace [89]. We implement motion control using NVIDIA Isaac Sim’s built-in RMPFlow algorithm [90]. Each experimental trial begins with a standardized initialization: the robot establishes a secure grasp on the target handle, creating a well-defined reference contact state  $\mathcal{C}_1$  that serves as the foundation for subsequent manipulation and deviation measurements.

### B. Baseline Selection

For comparative analysis, we adopt Tac-Man [1] as our primary baseline, a selection driven by its demonstrated superiority in manipulation effectiveness. The key architectural distinction between Tac-Man and our proposed TacMan-Turbo framework lies in their fundamental approach to manipulation control. Tac-Man implements a two-stage **reactive** paradigm: an execution stage that performs the primary manipulation, followed by a separate recovery stage that corrects contact

deviations after they occur. In contrast, TacMan-Turbo integrates these functions into a unified **proactive** process that anticipates and preemptively addresses potential deviations, enabling more fluid and continuous manipulation.

Beyond the primary baseline, we evaluate TacMan-Turbo against three additional baselines spanning traditional pre-planned methods, impedance control strategies, and learning-based approaches to contextualize our contribution within the broader literature. The pre-planned method leverages known kinematics to generate trajectories. To model parameter estimation uncertainty in real-world deployment, we perturb the ground-truth kinematic parameters with Gaussian noise (variance equals 0.01), a level consistent with the accuracy of current vision-based estimation methods [20, 21]. The impedance control baseline operates under the same prior-free setting as Tac-Man but excludes tactile feedback, instead relying on impedance adaptation to accommodate deviations. Finally, the learning-based baseline employs FlowBot3D [19] in a zero-shot setting, which is trained to use visual feedback for manipulation guidance.

To ensure a fair and rigorous comparison, both methods operate under identical initialization conditions as specified in Sec. IV-A, including consistent robot configuration, initial grasp state, and interaction direction. This standardized experimental protocol effectively isolates the impact of their differing control strategies, ensuring that performance differences directly reflect the fundamental distinction.



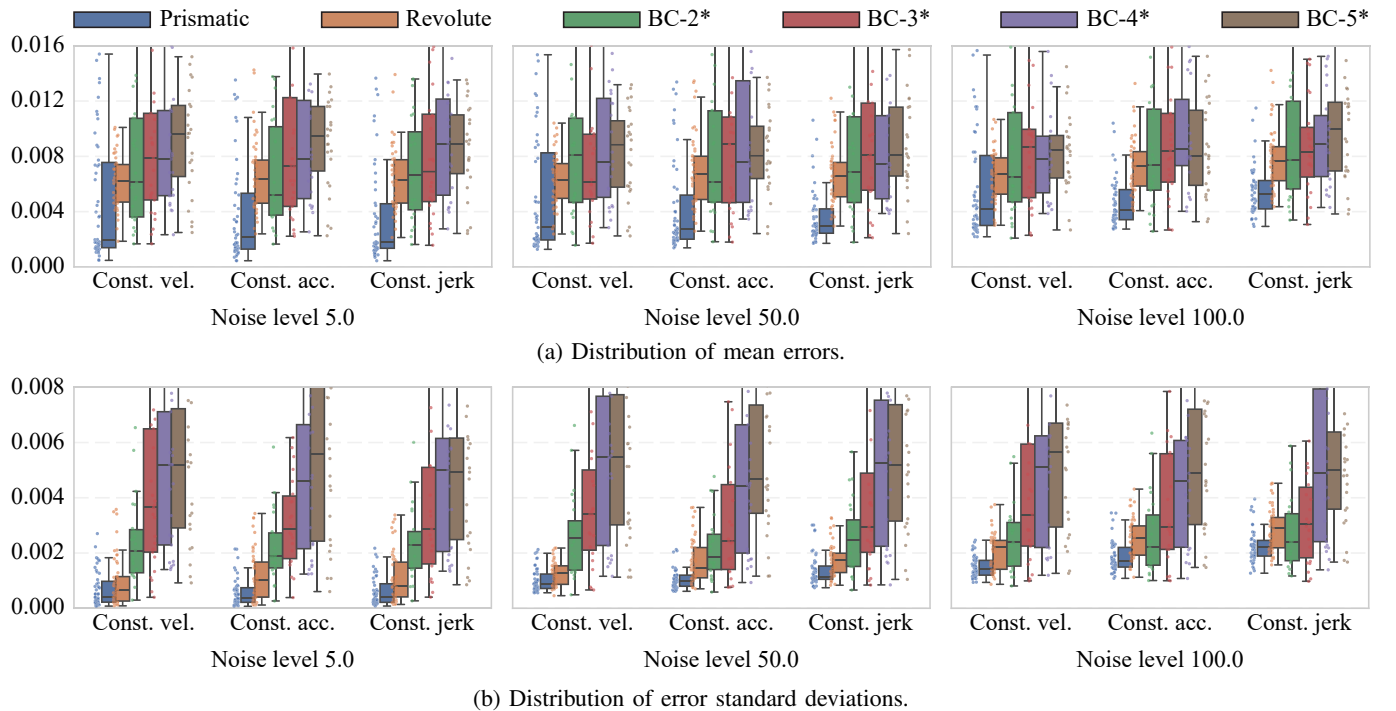


Fig. 5: **Prediction errors of various predictive models.** The prediction error distributions for three kinematic models—constant velocity, constant acceleration, and constant jerk—are presented and compared. All methods demonstrate comparable performance overall. However, higher-order models exhibit superior accuracy under low-noise conditions, as they better capture the underlying motion dynamics. Conversely, these same models show degraded performance in high-noise environments due to the cumulative effects of numerical differentiation, which amplifies measurement noise with each successive derivative computation. Box plots display time distributions with medians (center lines), interquartile ranges (boxes), and distribution extent (whiskers at  $1.5 \times \text{IQR}$ ). BC- $n$  represents Bézier curves of order  $n$ .

practicality and user experience.

**Action efficiency:** Action efficiency evaluates how directly each motion contributes to task progress, providing insights into the manipulation strategy’s precision and economy of movement. We implement joint-specific calculations: for revolute joints, we compute normalized angular displacement relative to a  $60^\circ$  threshold; for prismatic joints, we measure normalized linear displacement relative to a 250 mm threshold.

For the more complex Bézier curve trajectories, efficiency is determined by progression along the curve’s parameterization  $s \in [0, 1]$ , where the curve position is defined by:

$$B(s) = \sum_{i=0}^n b_i \binom{n}{i} s^i (1-s)^{n-i}, \quad (24)$$

where  $b_i$  represent the control points shaping the curve,  $\binom{n}{i}$  is the binomial coefficient, and  $n$  indicates the curve order. This unified parameterization enables consistent efficiency measurement across trajectories of varying complexity.

**Smoothness of robot joint trajectories:** Joint trajectory smoothness serves as both an efficiency indicator and a critical operational metric in robotic manipulation. Smoother trajectories offer multiple benefits: they reflect more efficient motion control, reduce mechanical stress and wear (extending robot lifespan), enhance control precision by limiting vibrations, and improve operational safety by minimizing abrupt movements—particularly important in human-robot interaction scenarios [91].

To quantify smoothness while avoiding the pitfall of arti-

cially smooth but excessively slow trajectories, we employ a time-weighted joint angle jerk. This metric measures the rate of change in joint angle acceleration scaled by task completion time, providing a balanced assessment that accounts for both smoothness and time efficiency.

Together, these four metrics provide a holistic performance profile, capturing the critical dimensions of manipulation capability from both effectiveness and efficiency perspectives.

#### D. Simulation Results

We begin by presenting qualitative results of TacMan-Turbo’s performance in manipulating articulated objects. Using representative examples from each category described in Sec. IV-A, Fig. 4 illustrates how the predicted handle poses effectively guide the gripper during manipulation tasks. Our observations reveal that when object motion exhibits smooth and consistent trends, the prediction model captures this behavior with high accuracy. However, as shown in the rightmost column of Fig. 4, abrupt changes in object motion can occasionally introduce prediction errors, stemming from the non-causal nature of our prediction model. Nevertheless, the system demonstrates robust performance through its iterative prediction process, which continually incorporates new observations to correct these errors. These results highlight TacMan-Turbo’s practical effectiveness in handling complex manipulation scenarios commonly encountered in real-world applications. Additional qualitative results are available in the [Supplementary Video S1](#).

To quantitatively assess the robustness and sensitivity of the predictive model, we evaluate the prediction residuals under varying noise levels and prediction models. While Sec. III establishes the constant velocity model as one approach, this section extends the analysis to include constant acceleration and constant jerk models.

Real tactile sensors, such as GelSight, exhibit inherent noise on the order of micrometers. To simulate realistic conditions while assessing the robustness under greater noise, we introduce Gaussian noise with standard deviations of 5.0  $\mu\text{m}$ , 50.0  $\mu\text{m}$ , and 100.0  $\mu\text{m}$ . The prediction residual is represented by the relative handle position with respect to the robot gripper after the robot executes the predicted action. To quantify this error, we map the transformation from the Lie group to its corresponding Lie algebra and compute the 2-norm of the resulting vector.

The results are shown in Fig. 5 with maximum prediction errors listed in Tab. I. All models achieve comparable performance overall. As the complexity of object kinematics increases, prediction errors increase correspondingly due to more varied motion patterns. Under low-noise conditions, higher-order models achieve better performance by capturing more complex underlying dynamics. However, they are more sensitive to noise due to the multiple differential computations involved. This tendency is clearly evident for basic prismatic and revolute joint types, but becomes less distinct in more complex kinematic scenarios. These results demonstrate that the constant-velocity model can be a robust choice, aligning with the intuition underlying previous works [86–88]. While all predictive models inevitably produce prediction residuals, TacMan-Turbo’s high-frequency operation enables rapid detection of deviations and adaptive adjustment. Rather than relying on perfect predictions, the system uses them as initial estimates within a closed-loop control framework that continuously corrects for modeling errors.

Our quantitative analysis evaluates TacMan-Turbo using the four metrics defined in Sec. IV-C.

**Success rate results:** As shown in Tab. II, the pre-planned method achieves considerable success rates across all tests with known kinematics. However, its performance degrades substantially under parameter estimation uncertainty, highlighting its sensitivity to modeling errors. Impedance control demonstrates effectiveness in handling moderate directional deviations without requiring known kinematics or tactile feedback. Nevertheless, its adaptive capacity becomes limited when large discrepancies exist between the planned and actual manipulation directions. The learning-based method performs well on objects with visual appearances similar to those encountered during training. However, its success rate drops to nearly zero when confronted with unfamiliar objects, revealing the generalization limitations inherent to current learning-based approaches.

Contrastingly, both tactile-informed methods, Tac-Man and TacMan-Turbo, achieve perfect 100% manipulation success rates across all 200 test objects spanning six categories. This consistent performance demonstrates the fundamental effectiveness of tactile-informed approaches for articulated object manipulation. Since successful task completion is a

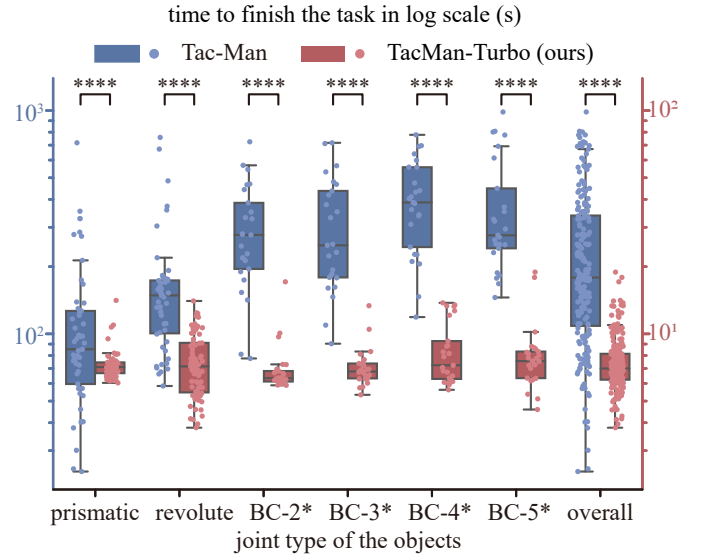


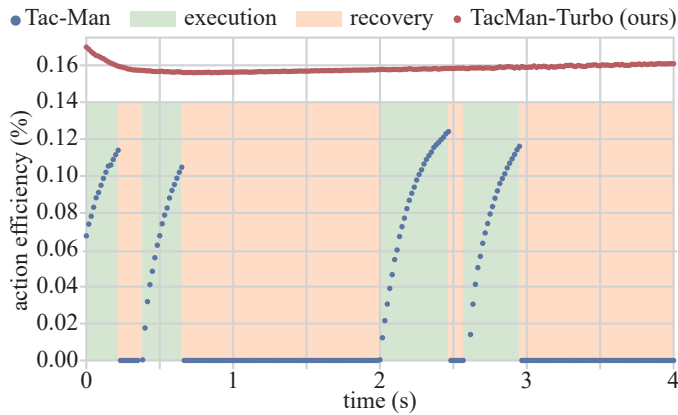
Fig. 6: **Time efficiency comparison (logarithmic scale, seconds).** Evaluation of task completion times reveals TacMan-Turbo’s superior efficiency compared to Tac-Man across all tested articulation types. Our method consistently completes manipulation tasks in significantly less time, with performance advantages particularly pronounced for complex articulations. Statistical analysis confirms these improvements are highly significant ( $****p < 0.0001$ ) throughout all joint categories. BC- $n$  denotes Bézier curves of order  $n$ ; box plots follow conventions established in Fig. 5.

prerequisite for meaningful efficiency evaluation, we focus our comparative analysis exclusively on methods that reliably satisfy this fundamental criterion.

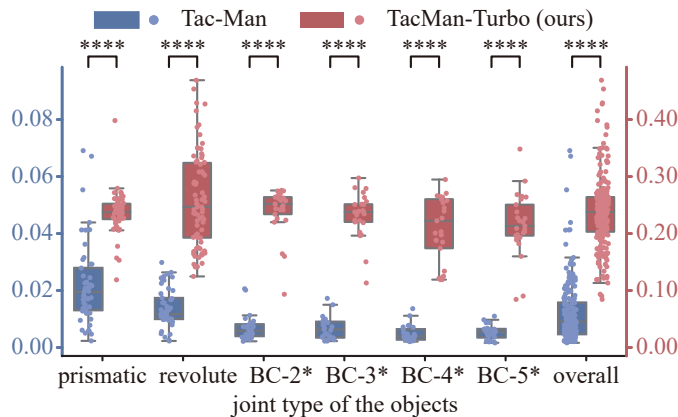
**Time efficiency results:** While maintaining identical success rates, TacMan-Turbo delivers dramatic improvements in execution speed, as detailed in Fig. 6. Average task completion times for TacMan-Turbo remain consistently around 10 s across all six object categories, compared to Tac-Man’s typical requirement of over 100 s. Using a one-sided paired t-test on the paired differences, performance was significantly faster for TacMan-Turbo ( $M = 7.6$  s,  $SD = 2.3$  s) than for Tac-Man ( $M = 250.7$  s,  $SD = 194.4$  s),  $t(199) = -17.7$ ,  $p < 0.0001$ . The one-sided 95% confidence interval for the mean paired difference was  $(-\infty, -220.4]$  s, and the effect size was Cohen’s  $d_z = -1.3$ .

**Action efficiency results:** The architectural differences between the two methods are most evident in their action efficiency profiles. As illustrated in Fig. 7(a), Tac-Man’s reactive approach requires dedicated recovery periods that, while necessary for correcting contact deviations, do not directly advance the manipulation task. In contrast, TacMan-Turbo’s proactive approach ensures that virtually all actions contribute directly to task progression, eliminating these non-productive recovery phases.

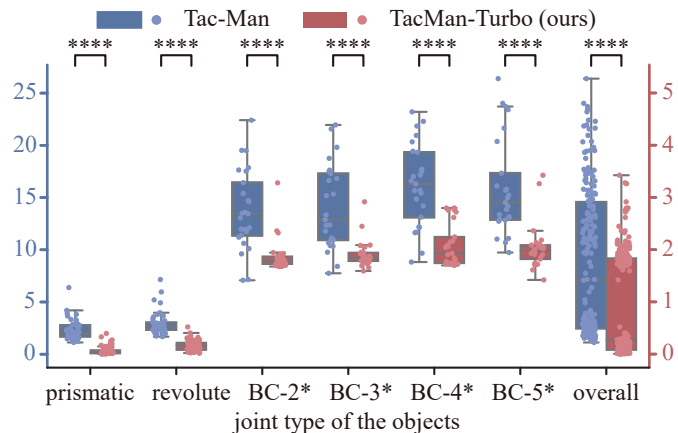
Our quantitative analysis examines both mean efficiency and the Relative Standard Deviation (RSD)—the ratio of standard deviation to mean—across all trials. Results presented in Figs. 7(b) and 7(c) demonstrate TacMan-Turbo’s consistently higher average efficiency with minimal variation, even when handling complex Bézier curve trajectories. Using a one-



(a) Temporal analysis of action efficiency during manipulation.



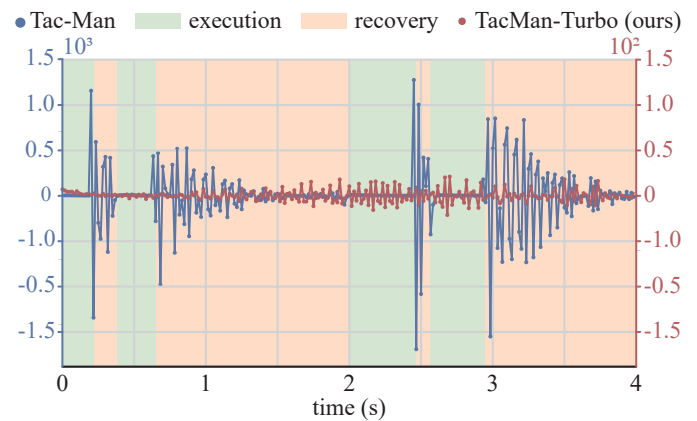
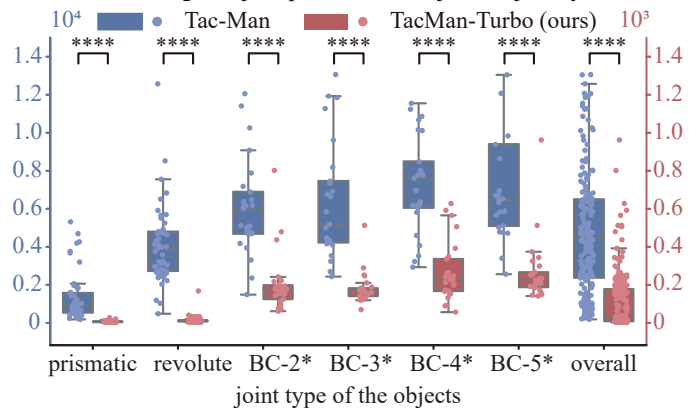
(b) Mean action efficiency (%) across object categories.



(c) Relative standard deviation of action efficiency.

**Fig. 7: Analysis of action efficiency across manipulation strategies.** The results reveal TacMan-Turbo’s superior motion efficiency through three complementary measures. Temporal profiles highlight TacMan-Turbo’s continuous productive motion vs. TacMan’s alternating execution-recovery cycles. Mean efficiency measurements confirm TacMan-Turbo’s consistently higher percentage of productive motion across all object categories. Relative standard deviation analysis demonstrates TacMan-Turbo’s more stable and predictable performance. Statistical tests across 200 objects spanning six joint categories show highly significant improvements (\*\*\*\*  $p < 0.0001$ ) in both efficiency magnitude and consistency. BC-n represents Bézier curves of order n; box plots follow conventions established in Fig. 5.

sided paired t-test on the paired differences, performance was significantly higher for TacMan-Turbo ( $M = 0.239\%$ ,  $SD =$

(a) Time-weighted jerk profile of base joint trajectory ( $s^{-3}$ ).(b) Mean time-weighted jerk magnitude across all categories ( $s^{-3}$ ).

**Fig. 8: Analysis of motion smoothness through jerk characteristics.** Motion smoothness evaluation reveals TacMan-Turbo maintains consistently lower jerk magnitudes by eliminating the disruptive transition phases inherent to TacMan’s approach. The temporal comparison shows TacMan-Turbo’s continuous motion profile contrasted with TacMan’s oscillatory pattern. Statistical analysis across 200 test objects confirms TacMan-Turbo achieves dramatic smoothness gains with standard joint types (prismatic and revolute), while improvements become more moderate with complex Bézier trajectories where challenging pose predictions occasionally necessitate corrective motions. Nevertheless, significant overall improvement (\*\*\*\*  $p < 0.0001$ ) in motion quality is observed across all categories. *Note:* BC-n denotes Bézier curves of order n; box plots follow conventions established in Fig. 5.

0.063%) than for TacMan ( $M = 0.012\%$ ,  $SD = 0.011\%$ ),  $t(199) = 52.6$ ,  $p < 0.0001$ . The one-sided 95% confidence interval for the mean paired difference was  $[0.2, \infty)\%$ , and the effect size was Cohen’s  $d_z = 3.7$ . Additionally, efficiency stability was significantly better for TacMan-Turbo ( $M = 1.0$ ,  $SD = 1.0$ ) than for TacMan ( $M = 8.8$ ,  $SD = 6.9$ ),  $t(199) = -15.9$ ,  $p < 0.0001$ . The one-sided 95% confidence interval for the mean paired difference was  $(-\infty, -7.0]$ , and the effect size was Cohen’s  $d_z = -1.1$ .

**Smoothness results:** Trajectory smoothness analysis reveals fundamental behavioral differences between the two approaches. As shown in Fig. 8(a), TacMan-Turbo demonstrates more consistent and significantly lower jerk variations compared to TacMan. This improvement stems directly from TacMan-Turbo’s proactive strategy, which eliminates the frequent transitions between execution and recovery stages

characteristic of Tac-Man. In the reactive approach, compensatory actions to address contact deviations typically require abrupt accelerations and decelerations, resulting in higher jerk measurements.

Comprehensive testing across all 200 objects, summarized in Fig. 8(b), reveals varying degrees of improvement across object categories. TacMan-Turbo shows the most dramatic smoothness gains with standard prismatic and revolute-joint objects, with more moderate improvements for Bézier-curve trajectories. This performance variation correlates with motion predictability—standard joint types follow consistent patterns that facilitate accurate prediction, while Bézier curves present more variable motion patterns that occasionally lead to mismatches between predicted and actual handle poses. Using a one-sided paired t-test on the paired differences, performance was significantly smoother for TacMan-Turbo ( $M = 109.8 \text{ s}^{-3}$ ,  $SD = 147.7 \text{ s}^{-3}$ ) than for Tac-Man ( $M = 4291.4 \text{ s}^{-3}$ ,  $SD = 3722.49 \text{ s}^{-3}$ ),  $t(199) = -18.3$ ,  $p < 0.0001$ . The one-sided 95% confidence interval for the mean paired difference was  $(-\infty, -3803.4] \text{ s}^{-3}$ , and the effect size was Cohen’s  $d_z = -1.3$ .

## V. REAL-WORLD EXPERIMENTS

While simulation studies establish TacMan-Turbo’s theoretical advantages, real-world deployment introduces critical challenges absent in simulation: sensor noise, execution uncertainty, and physical constraints. These factors can significantly impact control performance, necessitating comprehensive physical validation. This section presents our real-world experimental evaluation in four complementary parts: experimental setup (Sec. V-A), testing protocols (Sec. V-B), performance analysis (Sec. V-C), and validation on everyday objects (Sec. V-D)—including categories not represented in standard articulated object datasets.

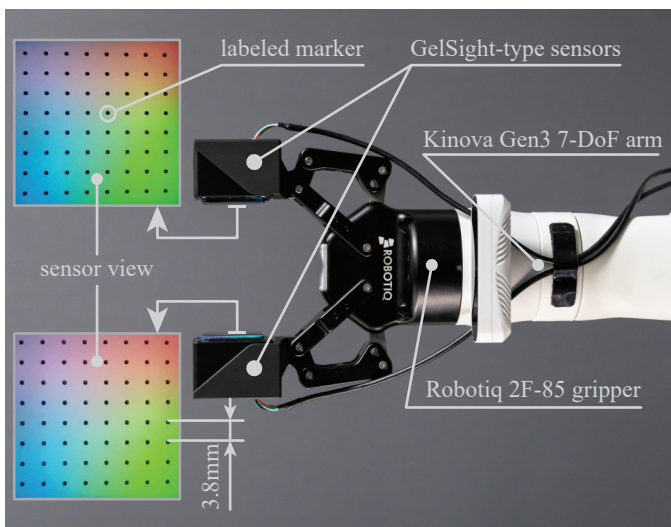


Fig. 9: **Physical experimental setup.** The experimental platform consists of a Kinova Gen3 7-DoF robotic arm equipped with a Robotiq 2F-85 parallel gripper. Each gripper finger incorporates a GelSight-type tactile sensor featuring an  $8 \times 8$  marker grid for precise contact force measurement and deformation tracking.

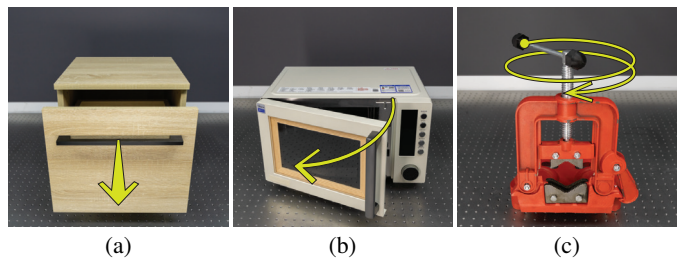


Fig. 10: **Test objects for real-world experiments.** The experimental validation employs three articulated objects with distinct kinematic characteristics: (a) a wooden drawer representing pure prismatic joint motion along a linear path, (b) a microwave oven featuring revolute joint dynamics through its door hinge mechanism, and (c) a bench vise with a hand crank generating complex helical motion. These objects collectively challenge the manipulation system across fundamental mechanical joint categories, from simple linear translation to complex compound movements requiring sophisticated trajectory prediction.

### A. Real-World Experimental Setup

Our physical validation framework integrates three essential components designed to evaluate TacMan-Turbo’s performance under realistic conditions: a robotic manipulation platform, high-fidelity tactile sensing, and diverse test objects with varied kinematic properties.

**Robot system setup:** Manipulating articulated objects with varied handle paths requires full 6-DoF movement capability in Cartesian space. Our platform employs a 7-DoF Kinova Gen3 robotic arm paired with a Robotiq 2F-85 gripper, providing the necessary workspace coverage and manipulation flexibility. To enable precise contact sensing—crucial for our tactile-based approach—we modified the gripper by replacing its standard pads with two GelSight-type tactile sensors. This integrated system, illustrated in Fig. 9, combines motion versatility with tactile feedback. The tactile sensor’s pose relative to the robot wrist is determined from the robot’s mechanical model. The internal camera of the GelSight-type sensor is then calibrated using OpenCV’s built-in routines: we estimate the camera intrinsics and lens distortion coefficients by imaging laser-cut black markers that serve as calibration targets.

Overall, the tactile sensor updates contact information at 120 Hz, while the offset velocity is computed at 50 Hz and transmitted to Kinova’s low-level controller, which operates at 1000 Hz. All computations were performed on Ubuntu 22.04 with an AMD Ryzen 9 5950X CPU and an NVIDIA GeForce RTX 3090 GPU. Over 10,000 test steps, the computation time from acquiring a new tactile image to producing the velocity command was  $3.70 \text{ ms} \pm 0.85 \text{ ms}$ . During these tests, the CPU load was  $13.36\% \pm 3.39\%$ , and the GPU load was  $56.75\% \pm 2.04\%$ . This low latency enables the controller to receive and react to new feedback in real time. Further discussion of the update frequency is provided in Sec. VI-B.

**Tactile sensors:** Accurate contact information between the object and robot is essential for TacMan-Turbo’s operation. Our implementation utilizes two GelSight-type tactile sensors [64], each featuring an  $8 \times 8$  grid of black markers (Fig. 9). These markers enable precise computation of the relative transformation matrix between gripper and handle through

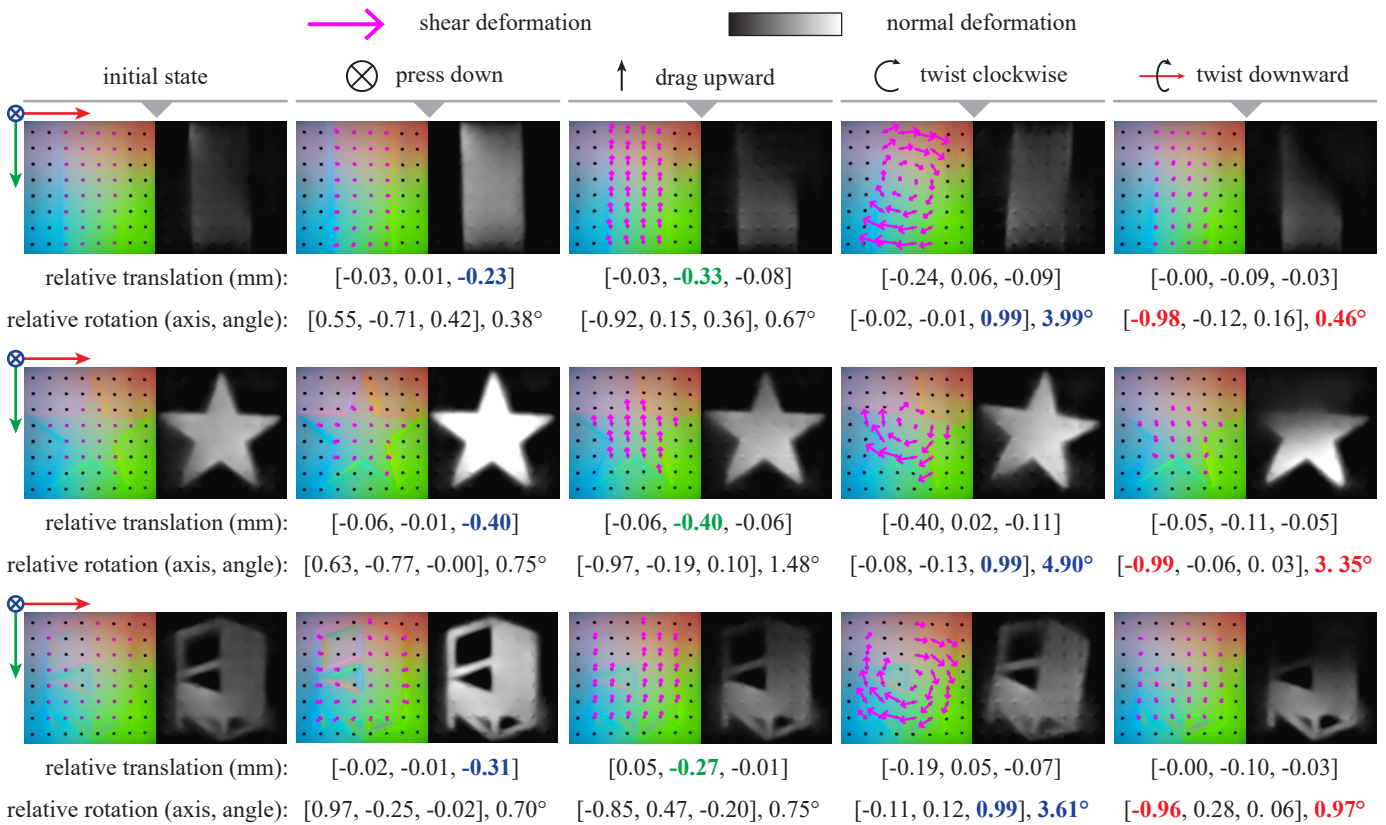


Fig. 11: **Tactile-informed pose estimation of in-hand objects.** Precise tracking of object movements during manipulation is demonstrated through tactile feedback across three different objects subjected to five distinct manipulations: initial state, pressing down, dragging upward, and two types of twisting motions. Contact points (marked with purple dots) are tracked to compute relative pose transformations using Eq. (8). The colored grids display marker displacement patterns while the grayscale images reveal normal force distribution. Quantitative relative translations (mm) and rotations (axis-angle representation) below each image pair confirm accurate capture of subtle pose changes. Visualization arrows (scaled 7 $\times$  for clarity) illustrate movement direction and magnitude, showing consistent correspondence between applied manipulations and detected transformations across varied object geometries.

point correspondence matching, as formalized in Eq. (8).

To validate the sensing system’s reliability, we conducted controlled perturbation tests with in-hand objects (Fig. 11). The computed transformations consistently aligned with applied perturbation directions, confirming the system’s ability to accurately estimate relative pose changes—a capability fundamental to the prediction framework described in Sec. III-C.

Selecting contact points requires careful parameter tuning to ensure accurate identification within the contact region; see Eq. (2). As noted by Zhao *et al.* [1], the transition zone between contact and non-contact regions can exhibit substantial normal deformation due to the soft sensor’s compliance, yet it does not reliably reflect changes in handle position. This makes a fixed threshold  $\epsilon$  ineffective.

To mitigate this, we adapt  $\epsilon$  in Eq. (2) during each trial, choosing the smallest value that yields exactly nine contact points in the initial contact set  $\mathcal{C}_i$ . Given that our setup places at most eight markers on any single line, this choice satisfies the constraints in Eqs. (10) and (11) by preventing collinearity, thereby enabling robust computation of contact deviations.

**Test articulated objects:** Following our simulation methodology, we evaluate TacMan-Turbo across three kinematically distinct categories: a drawer with prismatic joints (Fig. 10(a)), a microwave oven with revolute joints

(Fig. 10(b)), and a vise featuring a helical handle path (Fig. 10(c)). This selection enables systematic assessment across fundamental joint types while incorporating real-world mechanical complexities such as friction variations and mechanical play.

### B. Real-World Experiment Protocol

A fair comparison between manipulation approaches requires standardized testing conditions that isolate algorithmic differences from experimental variations. We implemented a rigorous protocol for evaluating TacMan-Turbo against TacMan across all real-world scenarios.

Each experimental trial began with identical initialization conditions: manually established stable grasps on object handles (illustrated in the left column of Fig. 12), ensuring consistent starting contact states. We maintained controlled experimental parameters by testing both methods on the same physical objects with identical initial interaction directions, eliminating potential confounding variables. For the grasping force, since a larger force ensures the non-slipping assumption, we used the lesser of two values: the gripper’s maximum force or the tactile sensor’s force tolerance. In the case of the selected Robotiq 2F-85 gripper, its maximum force is

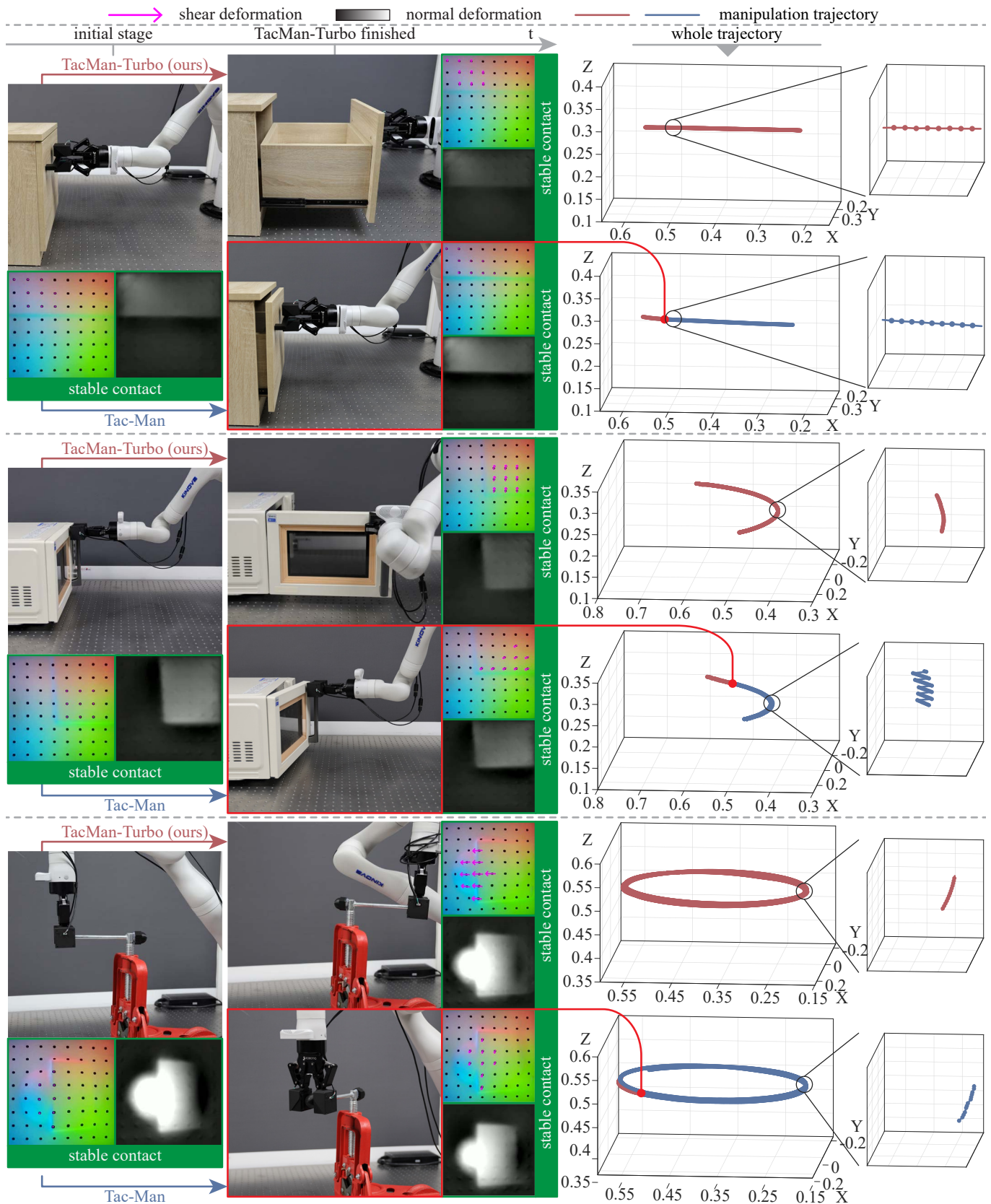


Fig. 12: **Real-world experiment results.** Comparative analysis across three articulated objects (drawer, microwave, vise) demonstrates successful manipulation by both methods with clear performance differences. Each row presents a complete manipulation sequence showing initial stable contact (left), manipulation stages (middle), and corresponding 3D trajectories (right). Tactile feedback visualizations display contact force distributions through marker displacement patterns (colored grids) and normal force intensity (grayscale images). While both approaches successfully complete all tasks, TacMan-Turbo consistently delivers more direct trajectories, resulting in reduced completion time and smoother motion profiles. Detailed comparisons are available in the [Supplementary Video S3](#).

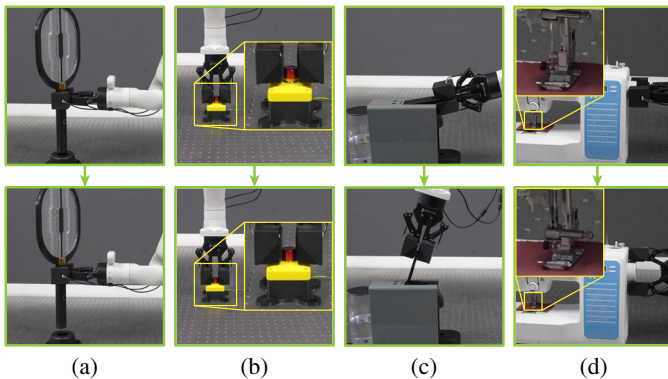


Fig. 13: **Tests on various common household objects.** Our TacMan-Turbo approach successfully generalizes to diverse everyday objects without requiring additional tuning. The system effectively manipulates: (a) a battery charging dock requiring precise device extraction, (b) a power switch with discrete on/off state transitions, (c) a coffee maker with a rotational lid mechanism, and (d) a sewing machine featuring complex mechanical linkages. Before (top) and after (bottom) images demonstrate successful manipulation for each object. These tests validate TacMan-Turbo’s ability to adapt to varied kinematic constraints and interaction requirements found in common household items, confirming its practical applicability beyond the primary test objects. Detailed manipulation processes are available in the [Supplementary Video S4](#).

significantly lower than the tactile sensor’s capacity. Therefore, we consistently applied the gripper’s maximum force.

Throughout each manipulation sequence, we recorded two complementary performance metrics: task completion time (measuring overall efficiency) and complete gripper trajectories (capturing motion quality). This dual measurement approach allows a comprehensive assessment of both the temporal efficiency and spatial characteristics of each manipulation strategy.

### C. Real-World Experiment Results

Our experimental results, presented in [Fig. 12](#), provide evidence of both methods’ manipulation capabilities while revealing significant performance distinctions between their approaches.

**Manipulation effectiveness:** Both TacMan-Turbo and Tac-Man successfully manipulated all test objects shown in [Fig. 10](#), confirming the fundamental effectiveness of tactile-based strategies for articulated object manipulation. Each method maintained stable contact throughout the manipulation process and completed all assigned tasks. The resulting motion trajectories (middle-right column, [Fig. 12](#)) closely followed expected kinematic patterns: linear motion for the prismatic-joint drawer, circular arcs for the revolute-joint microwave door, and extended helical paths for the vise handle. Minor deviations primarily stemmed from manufacturing tolerances inherent in real-world objects.

**Performance differences:** Despite similar success rates, time efficiency analysis revealed dramatic performance differences between the two approaches. As captured in the middle-left column of [Fig. 12](#), TacMan-Turbo completed manipulation tasks substantially faster, while Tac-Man had barely initiated object movement at the same point in time. This efficiency

gap stemmed from fundamental differences in operational strategies evident in detailed trajectory analysis (right column, [Fig. 12](#)).

The operational differences were most pronounced in objects requiring frequent directional adjustments. For the drawer’s linear motion, where the initial interaction direction naturally aligned with the object’s inherent trajectory, both methods demonstrated comparable path following. However, for revolute and helical motions requiring continuous directional adaptation, the methods exhibited markedly different behaviors:

Tac-Man employed a stepwise exploratory approach, advancing until a predefined threshold was reached. During this process, motion direction gradually drifted from the optimal trajectory, resulting in reduced action efficiency. While this drift was eventually corrected during recovery phases, the repeated cycles of exploration and correction produced non-smooth, zig-zag motion patterns with frequent velocity reversals.

In contrast, TacMan-Turbo leveraged predictive modeling to proactively adjust its motion direction. This continuous adaptation achieved consistently high action efficiency, smoother trajectories, and significantly faster task completion by eliminating the need for separate recovery phases.

These differences in execution time, action efficiency, and trajectory smoothness highlight TacMan-Turbo’s superior performance, successfully translating its theoretical advantages from simulation to real-world application. The results demonstrate TacMan-Turbo’s robustness in practical scenarios, effectively addressing challenges including sensor noise, execution uncertainty, and physical constraints. For detailed visualization of the manipulation process, including tactile sensor feedback throughout execution, we direct readers to the [Supplementary Video S2](#) and [Supplementary Video S3](#).

### D. Additional Tests on Everyday Objects

To assess TacMan-Turbo’s generalization capabilities beyond laboratory test objects, we extended our evaluation to include diverse household items encountered in daily environments. These objects, all amenable to secure manipulation using a parallel gripper, present varied kinematic structures and functional challenges not represented in standard datasets.

A battery charging dock ([Fig. 13\(a\)](#)) exemplifies objects with linear motion paths, where precise trajectory following is essential for secure device removal and to prevent potential damage. A power switch ([Fig. 13\(b\)](#)) represents safety-critical mechanisms with discrete state transitions, requiring reliable manipulation despite high resistance forces. A coffee maker ([Fig. 13\(c\)](#)) involves rotational motion, demanding adaptive handling to ensure smooth operation. Lastly, a sewing machine ([Fig. 13\(d\)](#)) introduces intricate motion patterns with varying resistance caused by its mechanical linkages and operational states.

These objects are not fully represented in standard articulated object datasets, highlighting TacMan-Turbo’s ability to generalize to novel, real-world mechanisms. Starting from stable initial grasps, TacMan-Turbo enabled the robot to successfully execute practical tasks: safely removing devices from

charging docks, reliably operating emergency power switches, preparing coffee makers for use, and manipulating sewing machines for basic textile operations.

The consistent success across this diverse set of previously unseen objects demonstrates TacMan-Turbo’s robustness and effectiveness in real-world applications. Its ability to handle varying physical properties, motion patterns, and functional requirements underscores its practical utility for everyday manipulation tasks. For detailed visualization of the manipulation process, including tactile sensor feedback throughout execution, we direct readers to the [Supplementary Video S4](#).

## VI. DISCUSSIONS

### A. Comparison with Tac-Man

Both TacMan-Turbo and Tac-Man leverage robot-object contact as the state representation. This approach ensures robustness without requiring pre-defined object kinematics for articulated object manipulation. The effectiveness of this strategy is comprehensively demonstrated through simulation results ([Tab. II](#)) and validated in real-world experiments ([Fig. 12](#)).

In the tactile control context, this modeling framework facilitates direct generation of robot Cartesian motions, eliminating the need for parameter tuning such as PID gain selection. Furthermore, unlike other approaches that also directly generate robot Cartesian motions, this framework avoids reliance on manually-defined tactile features. Instead, it leverages contact points for frame registration, making it inherently robust and agnostic to variations in handle shapes and structures.

Although both approaches share a similar state representation, they address the manipulation problem fundamentally differently. Tac-Man adopts a **reactive** strategy, prioritizing compensation for the current contact deviation—denoted as the deviation at step  $i$ . This approach temporarily halts object manipulation and focuses exclusively on deviation correction. Consequently, the process naturally divides into two distinct stages: execution and recovery, as evident in [Fig. 7\(a\)](#) and the zig-zag trajectories observed in [Fig. 12](#). While this formulation enhances robustness by ensuring timely compensation, it introduces inefficiencies, as the recovery stage does not directly advance the manipulation task. Additionally, transitions between execution and recovery stages produce non-smooth trajectories, as highlighted in [Fig. 8\(a\)](#) and [Fig. 12](#). Although Tac-Man incorporates multi-step exploration in real deployments to mitigate these limitations, the fundamental drawbacks of the reactive formulation remain.

In contrast, TacMan-Turbo’s core motivation explores whether robust manipulation can be achieved without interrupting the object interaction process. This question drives the development of TacMan-Turbo’s **proactive** formulation. Unlike reactive approaches, this formulation must account not only for the current  $i$ -step contact deviation but also anticipate the impact of the future  $i + 1$ -step contact deviation caused by continuous, non-halting manipulation, as described in [Eq. \(4\)](#).

This non-causal formulation requires estimating future object states—a challenge that reactive approaches cannot address, as they treat contact deviations solely as instantaneous

spatial error signals. TacMan-Turbo overcomes this through a temporal perspective: by analyzing sequential contact observations as time-series data, we extract object local kinematic information inherently embedded within tactile signals. This temporal interpretation enables TacMan-Turbo to predict future object states without requiring pre-specified kinematic models, allowing proactive velocity modulation that minimizes predicted future deviations.

This proactive formulation streamlines the previous execute-and-recover cycle into a continuous control loop. As a result, it delivers significant performance gains, including an order-of-magnitude improvement in time efficiency, consistently high action efficiency, and smooth trajectories free from acceleration-deceleration cycles—all while maintaining manipulation effectiveness. Furthermore, this framework maintains robustness against unexpected disturbances that frequently occur in human-centric environments while continuing the manipulation process without stopping for reactive compensation, as shown in [Fig. 14](#) and [Supplementary Video S5](#).

### B. Base Velocity Effects and Acquisition

The performance of TacMan-Turbo depends critically on the initial base velocity  $\mathbf{u}_0^{g_i}$  introduced in [Sec. III-B](#). While our primary experiments utilized manually assigned values for this parameter, understanding both its effects on system performance and methods for its automated acquisition has significant implications for practical deployment. We analyze the magnitude and direction components separately to provide a more structured understanding of their respective influences.

**Magnitude effect and acquisition:** Velocity magnitude directly impacts manipulation efficiency, with higher values generally enabling faster task completion. However, two fundamental constraints limit the applicable magnitude: the manipulator’s physical capabilities and the tactile sensors’ characteristics.

The manipulator constraint is straightforward— $\mathbf{u}_0^{g_i}$  cannot exceed the maximum permissible end-effector velocity. The tactile sensor constraint presents more nuanced considerations. For accurate contact deviation detection, differences in consecutive sensor readings (induced by  $\mathbf{u}_0^{g_i}$ ) must remain below the threshold that would trigger relative slippage between sensor and object. The most limiting scenario occurs when  $\mathbf{u}_0^{g_i}$  is perpendicular to the optimal interaction direction, as this configuration contributes maximally to contact deviation. Under these conditions, the constraint is formalized as:

$$\mathbf{u}_0^{g_i} \leq \gamma f, \quad (25)$$

where  $\gamma$  represents the critical slippage threshold and  $f$  denotes sampling frequency. This formulation reveals two pathways to enabling higher base velocities and thus improving manipulation efficiency: increasing surface roughness (which raises  $\gamma$ ) or elevating the sampling frequency.

Our implementation employed vision-based tactile sensors limited by their image sensors to approximately 120 Hz sampling frequency. Alternative sensing technologies, such as capacitive or 6D force/torque sensors, can operate at frequencies exceeding 1000 Hz, suggesting promising opportunities for

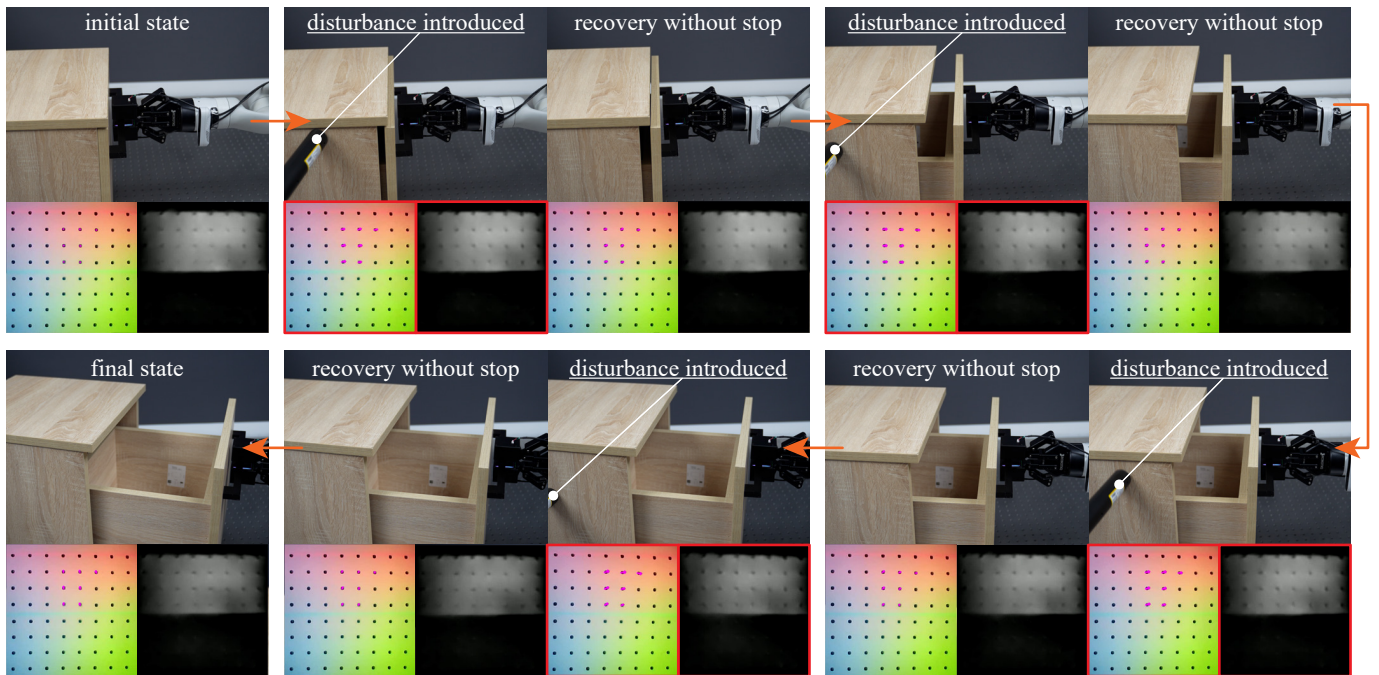


Fig. 14: **Manipulation under disturbance.** The efficiency gains enabled by TacMan-Turbo do not compromise its robustness against unexpected disturbances that frequently occur in human-centric environments. Unlike previous reactive paradigms [1] that require stopping for reactive adjustments, TacMan-Turbo can maintain manipulation progress continuously. Detailed manipulation processes are available in the [Supplementary Video S5](#).

substantial efficiency gains through TacMan-Turbo adaptation to these advanced modalities.

**Direction effect and acquisition:** The direction of base velocity profoundly influences manipulation efficiency. Greater alignment between  $u_0^{g_i}$  and the object’s optimal interaction direction yields higher first-step action efficiency (as defined in [Sec. IV-A](#)). This initial efficiency advantage propagates through subsequent steps, potentially maintaining elevated performance throughout the entire manipulation process.

To quantify this relationship, we conducted systematic simulations across 50 prismatic-joint objects ([Fig. 15\(a\)](#)), varying the angle  $\theta$  between  $u_0^{g_i}$  and the optimal direction. The results confirm our theoretical expectations: both time efficiency ([Fig. 15\(b\)](#)) and action efficiency ([Fig. 15\(c\)](#)) reach maximum values at  $\theta = 0^\circ$  and deteriorate as  $\theta$  increases. The time efficiency degradation closely follows  $1/\cos(\theta)$  trend, consistent with the scaling of the correct velocity component’s magnitude. This implies that as  $\theta$  approaches  $90^\circ$ , the time cost grows unboundedly—the manipulator maintains contact but ceases to produce meaningful object displacement. Furthermore, [Fig. 15\(d\)](#) reveals that per-step action efficiency maintains remarkable consistency throughout the manipulation process, closely tracking the first-step efficiency and achieving its peak at perfect alignment ( $\theta = 0$ ).

These findings not only validate our theoretical framework but also highlight the critical importance of appropriate base velocity direction selection. They further suggest promising research directions for automating preliminary directional information acquisition. Recent advances in vision-based direction estimation [[19–21](#)] offer particularly promising pathways.

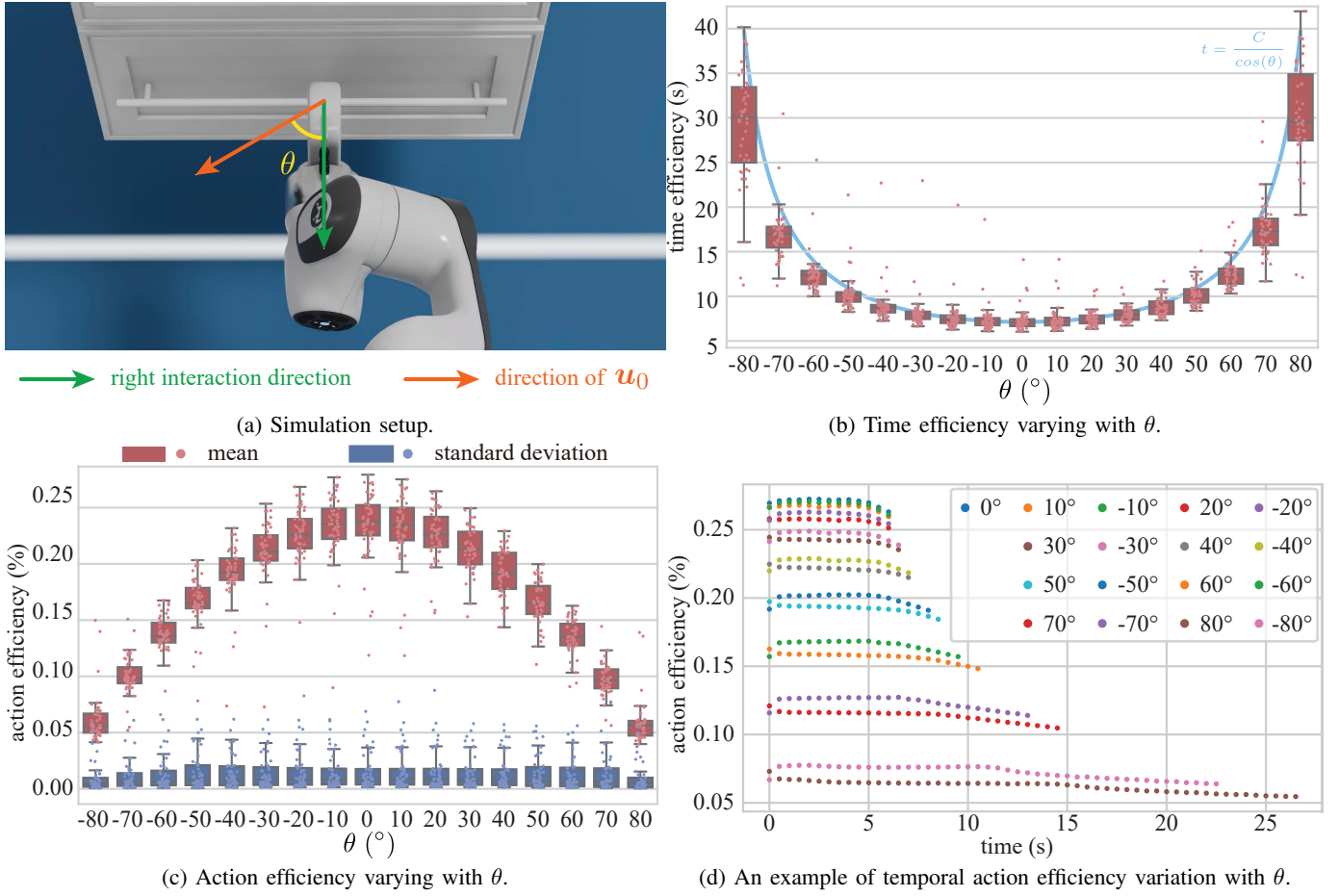
While such methods inevitably introduce some estimation noise, they enhance TacMan-Turbo’s operational autonomy. Moreover, our framework’s inherent robustness could effectively compensate for moderate directional inaccuracies, as demonstrated by the gradual performance degradation shown in [Fig. 15\(b\)](#).

## VII. CONCLUSION AND FUTURE WORK

This paper introduces TacMan-Turbo, a novel proactive control framework that successfully mitigates the persistent trade-off between efficiency and effectiveness in articulated object manipulation. By fundamentally reframing contact deviations as valuable kinematic information rather than mere error signals, our approach enables proactive control adjustments that significantly enhance manipulation efficiency while maintaining robust performance across diverse articulated mechanisms.

Our evaluation, including varied kinematic structures and multiple real-world experiments, demonstrates that TacMan-Turbo achieves perfect manipulation success rates (100%) while delivering substantial improvements in time efficiency, action efficiency, and trajectory smoothness compared to previous approaches. The statistical significance of these performance gains (p-values  $< 0.0001$ ) confirms that TacMan-Turbo effectively mitigates the effectiveness-efficiency trade-off that has challenged existing manipulation methods.

This work represents an important advancement toward practical robot deployment in human-centric environments, where robots must manipulate articulated objects reliably and efficiently without precise pre-provided kinematic models. The ability to extract and leverage kinematic insights directly from tactile feedback enables robots to adapt to uncertain object



**Fig. 15: Effects of base velocity direction.** We investigate how initial directional alignment impacts manipulation performance. (a) The simulation setup measures the angle  $\theta$  between the correct interaction direction (green arrow) and the initial motion direction  $\mathbf{u}_0^{gi}$  (orange arrow). (b) Time efficiency results reveal significant performance degradation as  $\theta$  increases in magnitude, with optimal performance near  $\theta = 0^\circ$ . The degradation closely follows a  $1/\cos(\theta)$  trend. The constant  $C$  in the figure denotes the median time efficiency at  $\theta = 0^\circ$ . (c) Action efficiency peaks when the initial direction closely aligns with the correct interaction direction and gradually decreases with increasing angular deviation. (d) The temporal progression further illustrates how different initial angles affect performance throughout the manipulation process, with properly aligned initial directions maintaining consistently higher efficiency. These findings highlight the critical importance of directional alignment in manipulation tasks. Box plots follow conventions established in Fig. 5.

structures while executing smooth, efficient movements—a capability essential for robots operating effectively in dynamic, everyday settings alongside humans.

Several promising directions for future research emerge from this work. First, developing adaptive methods for base velocity optimization could further enhance efficiency by dynamically adjusting control parameters in response to object characteristics. Second, extending the framework to support diverse tactile sensing modalities would broaden its applicability across robot platforms with different sensing capabilities. Finally, the current research primarily focuses on rigid, single-joint articulated objects. Third, while our current focus is on rigid, single-joint articulated objects, effectively handling soft, multi-part articulated systems in cluttered environments remains an important and promising challenge. These enhancements will not only further boost the framework’s efficiency but also facilitate its application in increasingly complex, unstructured real-world environments.

#### ACKNOWLEDGMENT

Our sincere thanks go to Yuanhong Zeng (UCLA) and Qian Long (UCLA) for their suggestion in driving Kinova, Lei Yan (LeapZenith AI Research) for mechanical design support, Yida Niu (PKU) for demo shooting, Saiyao Zhang (UCAS) for simulation support, and Ms. Hailu Yang (PKU) for her assistance in procuring every piece of raw material necessary for this research.

#### REFERENCES

- [1] Z. Zhao, Y. Li, W. Li, Z. Qi, L. Ruan, Y. Zhu, and K. Althoefer, “Tac-Man: Tactile-informed prior-free manipulation of articulated objects,” *IEEE Transactions on Robotics (T-RO)*, vol. 41, pp. 538–557, 2025.
- [2] W. Wang, Z. Zhao, Z. Jiao, Y. Zhu, S.-C. Zhu, and H. Liu, “Rearrange indoor scenes for human-robot co-activity,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2023.
- [3] J. Jeon, H.-r. Jung, N. Pico, T. Luong, and H. Moon, “Task-motion planning system for socially viable service robots based on object manipulation,” *Biomimetics*, vol. 9, no. 7, p. 436, 2024.
- [4] Y. Karayiannidis, C. Smith, F. E. Vina, P. Ogren, and D. Kragic, ““open sesame!” adaptive force/velocity control for opening unknown doors,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2012.

- [5] Y. Karayiannidis, C. Smith, F. E. V. Barrientos, P. Ögren, and D. Kragic, "An adaptive control approach for opening doors and drawers under uncertainties," *IEEE Transactions on Robotics (T-RO)*, vol. 32, no. 1, pp. 161–175, 2016.
- [6] S. Chitta, B. Cohen, and M. Likhachev, "Planning for autonomous door opening with a mobile manipulator," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2010.
- [7] F. Burget, A. Hornung, and M. Bennewitz, "Whole-body motion planning for manipulation of articulated objects," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2013.
- [8] R. Hu, W. Li, O. Van Kaick, A. Shamir, H. Zhang, and H. Huang, "Learning to predict part mobility from a single static snapshot," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 6, pp. 1–13, 2017.
- [9] X. Li, H. Wang, L. Yi, L. J. Guibas, A. L. Abbott, and S. Song, "Category-level articulated object pose estimation," in *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
- [10] V. Zeng, T. E. Lee, J. Liang, and O. Kroemer, "Visual identification of articulated object parts," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021.
- [11] M. Mittal, D. Hoeller, F. Farshidian, M. Hutter, and A. Garg, "Articulated object interaction in unknown scenes with whole-body mobile manipulation," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022.
- [12] T. Dai, J. Wong, Y. Jiang, C. Wang, C. Gokmen, R. Zhang, J. Wu, and L. Fei-Fei, "Acad: Automated creation of digital cousins for robust policy learning," in *Conference on Robot Learning (CoRL)*, 2024.
- [13] Y. Liu, B. Jia, R. Lu, J. Ni, S.-C. Zhu, and S. Huang, "Building interactable replicas of complex articulated objects via gaussian splatting," in *Proceedings of International Conference on Learning Representations (ICLR)*, 2025.
- [14] J. Ni, Y. Liu, R. Lu, Z. Zhou, S.-C. Zhu, Y. Chen, and S. Huang, "Compositional neural scene reconstruction with generative diffusion prior," in *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.
- [15] Y. Zhu, T. Gao, L. Fan, S. Huang, M. Edmonds, H. Liu, F. Gao, C. Zhang, S. Qi, Y. N. Wu, *et al.*, "Dark, beyond deep: A paradigm shift to cognitive ai with humanlike common sense," *Engineering*, vol. 6, no. 3, pp. 310–345, 2020.
- [16] K. Hausman, S. Niekum, S. Osentoski, and G. S. Sukhatme, "Active articulation model estimation through interactive perception," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2015.
- [17] R. Martín-Martín and O. Brock, "Coupled recursive estimation for online interactive perception of articulated objects," *International Journal of Robotics Research (IJRR)*, vol. 41, no. 8, pp. 741–777, 2022.
- [18] J. Lv, Q. Yu, L. Shao, W. Liu, W. Xu, and C. Lu, "Sageci-system: Towards sample-efficient, generalizable, compositional, and incremental robot learning," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2022.
- [19] B. Eisner, H. Zhang, and D. Held, "FlowBot3D: Learning 3D articulation flow to manipulate articulated objects," in *Robotics: Science and Systems (RSS)*, 2022.
- [20] Q. Yu, J. Wang, W. Liu, C. Hao, L. Liu, L. Shao, W. Wang, and C. Lu, "Gamma: Generalizable articulation modeling and manipulation for articulated objects," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2024.
- [21] J. Wang, W. Liu, Q. Yu, Y. You, L. Liu, W. Wang, and C. Lu, "RPMart: Towards robust perception and manipulation for articulated objects," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2024.
- [22] P. Pastor, H. Hoffmann, T. Asfour, and S. Schaal, "Learning and generalization of motor skills by learning from demonstration," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2009.
- [23] T. Welschhold, C. Dornhege, and W. Burgard, "Learning mobile manipulation actions from human demonstrations," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017.
- [24] H. Xiong, Q. Li, Y.-C. Chen, H. Bharadhwaj, S. Sinha, and A. Garg, "Learning by watching: Physical imitation of manipulation skills from human videos," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021.
- [25] R. Gong, J. Huang, Y. Zhao, H. Geng, X. Gao, Q. Wu, W. Ai, Z. Zhou, D. Terzopoulos, and S.-C. Zhu, "Arnold: A benchmark for language-grounded task learning with continuous states in realistic 3d scenes," in *Proceedings of International Conference on Computer Vision (ICCV)*, 2023.
- [26] T. Zhang, Z. McCarthy, O. Jow, D. Lee, X. Chen, K. Goldberg, and P. Abbeel, "Deep imitation learning for complex manipulation tasks from virtual reality teleoperation," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2018.
- [27] J. Wong, A. Tung, A. Kurenkov, A. Mandlikar, L. Fei-Fei, S. Savarese, and R. Martín-Martín, "Error-aware imitation learning from teleoperation data for mobile manipulation," in *Conference on Robot Learning (CoRL)*, 2022.
- [28] Y. Qin, W. Yang, B. Huang, K. Van Wyk, H. Su, X. Wang, Y.-W. Chao, and D. Fox, "Anyteleop: A general vision-based dexterous robot arm-hand teleoperation system," in *Robotics: Science and Systems (RSS)*, 2023.
- [29] Z. Fu, T. Z. Zhao, and C. Finn, "Mobile aloha: Learning bimanual mobile manipulation with low-cost whole-body teleoperation," in *Conference on Robot Learning (CoRL)*, 2024.
- [30] B. Zheng, S. Verma, J. Zhou, I. W. Tsang, and F. Chen, "Imitation learning: Progress, taxonomies and challenges," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 99, pp. 1–16, 2022.
- [31] Y. Qin, H. Su, and X. Wang, "From one hand to multiple hands: Imitation learning for dexterous manipulation from single-camera teleoperation," *IEEE Robotics and Automation Letters (RA-L)*, vol. 7, no. 4, pp. 10873–10881, 2022.
- [32] A. Tekden, M. P. Deisenroth, and Y. Bekiroglu, "Grasp transfer based on self-aligning implicit representations of local surfaces," *IEEE Robotics and Automation Letters (RA-L)*, 2023.
- [33] Y. Urakami, A. Hodgkinson, C. Carlin, R. Leu, L. Rigazio, and P. Abbeel, "Doorgym: A scalable door opening environment and baseline agent," in *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2019.
- [34] Z. Xu, Z. He, and S. Song, "Universal manipulation policy network for articulated objects," *IEEE Robotics and Automation Letters (RA-L)*, vol. 7, no. 2, pp. 2447–2454, 2022.
- [35] Y. Chen, T. Wu, S. Wang, X. Feng, J. Jiang, Z. Lu, S. McAleer, H. Dong, S.-C. Zhu, and Y. Yang, "Towards human-level bimanual dexterous manipulation with reinforcement learning," in *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2022.
- [36] H. Geng, Z. Li, Y. Geng, J. Chen, H. Dong, and H. Wang, "Partmanip: Learning cross-category generalizable part manipulation policy from point cloud observations," in *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.
- [37] K. Fan, Z. Chen, Q. Liu, G. Ferrigno, and E. De Momi, "A reinforcement learning approach for real-time articulated surgical instrument 3d pose reconstruction," *IEEE Transactions on Medical Robotics and Bionics*, 2024.
- [38] D. Honerkamp, H. Maheshka, J. O. von Hartz, T. Welschhold, and A. Valada, "Whole-body teleoperation for mobile manipulation at zero added cost," *IEEE Robotics and Automation Letters (RA-L)*, 2025.
- [39] K. Mo, S. Zhu, A. X. Chang, L. Yi, S. Tripathi, L. J. Guibas, and H. Su, "Partnet: A large-scale benchmark for fine-grained and hierarchical part-level 3D object understanding," in *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [40] F. Xiang, Y. Qin, K. Mo, Y. Xia, H. Zhu, F. Liu, M. Liu, H. Jiang, Y. Yuan, H. Wang, *et al.*, "Sapien: A simulated part-based interactive environment," in *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
- [41] L. Liu, W. Xu, H. Fu, S. Qian, Q. Yu, Y. Han, and C. Lu, "AKB-48: a real-world articulated object knowledge base," in *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.
- [42] H. Geng, H. Xu, C. Zhao, C. Xu, L. Yi, S. Huang, and H. Wang, "Gapartnet: Cross-category domain-generalizable object perception and manipulation via generalizable and actionable parts," in *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.
- [43] W. Cui, C. Zhao, S. Wei, J. Zhang, H. Geng, Y. Chen, H. Li, and H. Wang, "Gapartmanip: A large-scale part-centric dataset for material-agnostic articulated object manipulation," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 14791–14798, IEEE, 2025.
- [44] B. Siciliano, O. Khatib, and T. Kröger, *Springer handbook of robotics*, vol. 200. Springer, 2008.
- [45] Y. Li, W. Du, C. Yu, P. Li, Z. Zhao, T. Liu, C. Jiang, Y. Zhu, and S. Huang, "Tacccl: Scaling up vision-based tactile robotics via high-performance gpu simulation," in *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2025.
- [46] S. H. Lee, J. M. Baek, H. Moon, H. R. Choi, and J. C. Koo, "Enhancing free-space transparency with discrete energy-based compensation in physical human–robot interaction," *Robotics and Autonomous Systems*, vol. 187, p. 104940, 2025.

- [47] S. H. Park, M. Doh, C. Park, A. T. Luong, H. R. Choi, J. C. Koo, H. Rodrigue, and H. Moon, "Modeling and reinforcement learning-based control of simultaneous positive and negative pressure generation in pneumatic systems," *IEEE Robotics and Automation Letters (RA-L)*, 2025.
- [48] S. H. Lee, D. J. Oh, H. Moon, H. R. Choi, and J. C. Koo, "Improving transparency in physical human-robot interaction via unknown dynamics compensation," *Robotics and Computer-Integrated Manufacturing*, vol. 98, p. 103153, 2026.
- [49] S. Brahmabhatt, C. Ham, C. C. Kemp, and J. Hays, "Contactdb: Analyzing and predicting grasp contact via thermal imaging," in *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [50] Y. Li, B. Liu, Y. Geng, P. Li, Y. Yang, Y. Zhu, T. Liu, and S. Huang, "Grasp multiple objects with one hand," *IEEE Robotics and Automation Letters (RA-L)*, 2024.
- [51] C. Moses, M. Noseworthy, L. P. Kaelbling, T. Lozano-Pérez, and N. Roy, "Visual prediction of priors for articulated object interaction," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2020.
- [52] H. Yousef, M. Boukallel, and K. Althoefer, "Tactile sensing for dexterous in-hand manipulation in robotics—a review," *Sensors and Actuators A: physical*, vol. 167, no. 2, pp. 171–187, 2011.
- [53] S. Luo, N. F. Lepora, W. Yuan, K. Althoefer, G. Cheng, and R. Dahiya, "Tactile robotics: An outlook," *IEEE Transactions on Robotics (T-RO)*, 2025.
- [54] J. A. Fishel and G. E. Loeb, "Sensing tactile microvibrations with the biotac—comparison with human sensitivity," in *International Conference on Biomedical Robotics and Biomechanics (BioRob)*, 2012.
- [55] R. Ma, J. Lee, D. Choi, H. Moon, and S. Baik, "Knitted fabrics made from highly conductive stretchable fibers," *Nano letters*, vol. 14, no. 4, pp. 1944–1951, 2014.
- [56] Y. Yang, X. Wei, N. Zhang, J. Zheng, X. Chen, Q. Wen, X. Luo, C.-Y. Lee, X. Liu, X. Zhang, *et al.*, "A non-printed integrated-circuit textile for wireless theranostics," *Nature Communications*, vol. 12, no. 1, p. 4876, 2021.
- [57] Y. Yu, J. Li, S. A. Solomon, J. Min, J. Tu, W. Guo, C. Xu, Y. Song, and W. Gao, "All-printed soft human-machine interface for robotic physicochemical sensing," *Science Robotics*, vol. 7, no. 67, p. eabn0495, 2022.
- [58] H. Lee, H. Sun, H. Park, G. Serhat, B. Javot, G. Martius, and K. J. Kuchenbecker, "Predicting the force map of an ert-based tactile sensor using simulation and deep networks," *IEEE Transactions on Automation Science and Engineering (T-ASE)*, vol. 20, no. 1, pp. 425–439, 2022.
- [59] C. M. Boutry, M. Negre, M. Jorda, O. Vardoulis, A. Chortos, O. Khatib, and Z. Bao, "A hierarchically patterned, bioinspired e-skin able to detect the direction of applied pressure for robotics," *Science Robotics*, vol. 3, no. 24, p. eaau6914, 2018.
- [60] A. B. Dawood, C. Coppola, and K. Althoefer, "Learning decoupled multi-touch force estimation, localization and stretch for soft capacitive e-skin," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2023.
- [61] S. Zhao, Z. Li, H. Xia, and R. Cui, "Skin-inspired triple tactile sensors integrated on robotic fingers for bimanual manipulation in human-cyber-physical systems," *IEEE Transactions on Automation Science and Engineering (T-ASE)*, 2023.
- [62] H. Liu, Y. Yu, F. Sun, and J. Gu, "Visual-tactile fusion for object recognition," *IEEE Transactions on Automation Science and Engineering (T-ASE)*, vol. 14, no. 2, pp. 996–1008, 2016.
- [63] D. Guo, H. Liu, B. Fang, F. Sun, and W. Yang, "Visual affordance guided tactile material recognition for waste recycling," *IEEE Transactions on Automation Science and Engineering (T-ASE)*, vol. 19, no. 4, pp. 2656–2664, 2021.
- [64] W. Yuan, S. Dong, and E. H. Adelson, "Gelsight: High-resolution robot tactile sensors for estimating geometry and force," *Sensors*, vol. 17, no. 12, p. 2762, 2017.
- [65] B. Ward-Cherrier, N. Pestell, L. Cramphorn, B. Winstone, M. E. Giannaccini, J. Rossiter, and N. F. Lepora, "The tactip family: Soft optical tactile sensors with 3D-printed biomimetic morphologies," *Soft Robotics*, vol. 5, no. 2, pp. 216–227, 2018.
- [66] W. Li, M. Wang, J. Li, Y. Su, D. K. Jha, X. Qian, K. Althoefer, and H. Liu, "L<sup>3</sup> f-touch: A wireless gelsight with decoupled tactile and three-axis force sensing," *IEEE Robotics and Automation Letters (RA-L)*, 2023.
- [67] W. Li, Z. Zhao, L. Cui, W. Zhang, H. Liu, L.-A. Li, and Y. Zhu, "Minitac: An ultra-compact 8 mm vision-based tactile sensor for enhanced palpation in robot-assisted minimally invasive surgery," *IEEE Robotics and Automation Letters (RA-L)*, vol. 9, no. 12, pp. 11170–11177, 2024.
- [68] Z. Zhao, W. Li, Y. Li, T. Liu, B. Li, M. Wang, K. Du, H. Liu, Y. Zhu, Q. Wang, *et al.*, "Embedding high-resolution touch across robotic hands enables adaptive human-like grasping," *Nature Machine Intelligence*, vol. 7, no. 6, pp. 889–900, 2025.
- [69] B. Zhang, S. Cui, C. Zhang, J. Hu, R. Wang, and S. Wang, "Gelstereo tip: A spherical fingertip visuotactile sensor for multi-finger screwing manipulation," *IEEE Transactions on Automation Science and Engineering (T-ASE)*, 2025.
- [70] S. Athar, X. Zhang, J. Ueda, Y. Zhao, and Y. She, "Vibtac: A high-resolution high-bandwidth tactile sensing finger for multi-modal perception in robotic manipulation," *IEEE Transactions on Haptics*, 2025.
- [71] M. R. I. Prince, S. Athar, P. Zhou, and Y. She, "Tacscope: A miniaturized vision-based tactile sensor for surgical applications," *Advanced Robotics Research*, p. e202500117, 2025.
- [72] P. L. Gould and Y. Feng, *Introduction to linear elasticity*, vol. 2. Springer, 1994.
- [73] L. Weiss, A. Sanderson, and C. Neuman, "Dynamic sensor-based control of robots with visual feedback," *IEEE Journal on Robotics and Automation*, vol. 3, no. 5, pp. 404–417, 1987.
- [74] A. D. Berger and P. K. Khosla, "Using tactile data for real-time feedback," *International Journal of Robotics Research (IJRR)*, vol. 10, no. 2, pp. 88–102, 1991.
- [75] Q. Li, C. Schürmann, R. Haschke, and H. J. Ritter, "A control framework for tactile servoing," in *Robotics: Science and Systems (RSS)*, 2013.
- [76] Z. Kappassov, J.-A. Corrales, and V. Perdureau, "Touch driven controller and tactile features for physical interactions," *Robotics and Autonomous Systems*, vol. 123, p. 103332, 2020.
- [77] N. F. Lepora and J. Lloyd, "Pose-based tactile servoing: Controlled soft touch using deep learning," *IEEE Robotics and Automation Magazine (RA-M)*, vol. 28, no. 4, pp. 43–55, 2021.
- [78] Y. She, S. Wang, S. Dong, N. Sunil, A. Rodriguez, and E. Adelson, "Cable manipulation with a tactile-reactive gripper," *International Journal of Robotics Research (IJRR)*, vol. 40, no. 12-14, pp. 1385–1401, 2021.
- [79] J. Lloyd and N. F. Lepora, "Pose-and-shear-based tactile servoing," *International Journal of Robotics Research (IJRR)*, vol. 43, no. 7, pp. 1024–1055, 2024.
- [80] N. Chen, H. Zhang, and R. Rink, "Edge tracking using tactile servo," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 1995.
- [81] H. Zhang and N. N. Chen, "Control of contact via tactile sensing," *IEEE Transactions on Robotics and Automation (T-RA)*, vol. 16, no. 5, pp. 482–495, 2000.
- [82] P. Sikka, H. Zhang, and S. Sutphen, "Tactile servo: Control of touch-driven robot motion," in *Experimental Robotics III: The 3rd International Symposium, Kyoto, Japan, October 28–30, 1993*, pp. 219–233, Springer, 2005.
- [83] A. Wilson, H. Jiang, W. Lian, and W. Yuan, "Cable routing and assembly using tactile-driven motion primitives," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2023.
- [84] Z. Xu and Y. She, "Letac-mpc: Learning model predictive control for tactile-reactive grasping," *IEEE Transactions on Robotics (T-RO)*, vol. 40, pp. 4376–4395, 2024.
- [85] W. Kabsch, "A solution for the best rotation to relate two sets of vectors," *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, vol. 32, no. 5, pp. 922–923, 1976.
- [86] A. J. Davison, I. D. Reid, N. D. Molton, and O. Stasse, "Monoslam: Real-time single camera slam," *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 29, no. 6, pp. 1052–1067, 2007.
- [87] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "Orb-slam: A versatile and accurate monocular slam system," *IEEE Transactions on Robotics (T-RO)*, vol. 31, no. 5, pp. 1147–1163, 2015.
- [88] C. Schöller, V. Aravantinos, F. Lay, and A. Knoll, "What the constant velocity model can teach us about pedestrian motion prediction," *IEEE Robotics and Automation Letters (RA-L)*, vol. 5, no. 2, pp. 1696–1703, 2020.
- [89] Z. Zhao, L. Cui, S. Xie, S. Zhang, Z. Han, L. Ruan, and Y. Zhu, "B\*: Efficient and optimal base placement for fixed-base manipulators," *IEEE Robotics and Automation Letters (RA-L)*, vol. 10, no. 10, pp. 10634–10641, 2025.
- [90] C.-A. Cheng, M. Mukadam, J. Issac, S. Birchfield, D. Fox, B. Boots, and N. Ratliff, "Rmpflow: A geometric framework for generation of multitask motion policies," *IEEE Transactions on Automation Science and Engineering (T-ASE)*, vol. 18, no. 3, pp. 968–987, 2021.
- [91] S. Macfarlane and E. A. Croft, "Jerk-bounded manipulator trajectory planning: design for real-time applications," *IEEE Transactions on Robotics and Automation (T-RA)*, vol. 19, no. 1, pp. 42–52, 2003.