

Tac-Man: Tactile-Informed Prior-Free Manipulation of Articulated Objects

Zihang Zhao[✉], Yuyang Li[✉], Wanlin Li[✉], Zhenghao Qi[✉], Lecheng Ruan[✉],
Yixin Zhu[✉], and Kaspar Althoefer[✉], *Senior Member, IEEE*

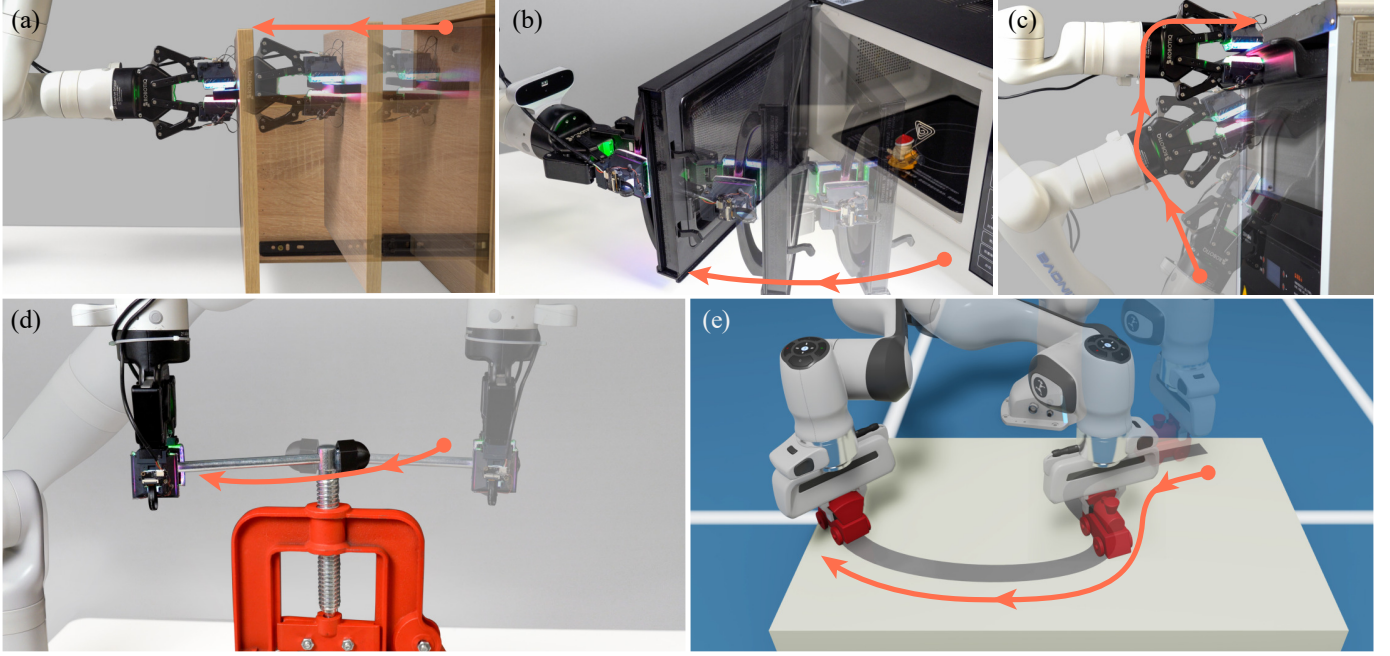


Fig. 1: **Tactile-informed prior-free manipulation of articulated objects.** Timely compensation for contact deviations allows robots to precisely manipulate a variety of articulated objects, eliminating the need for prior kinematic model knowledge. The capabilities are showcased by (a) manipulating objects with prismatic joints, such as drawers; (b) handling objects with revolute joints, such as microwave ovens; and managing complex mechanisms that involve simultaneous translational and rotational movements, as exemplified by (c) an ice machine and (d) a vise. (e) Additionally, the robot proficiently performs manipulation following arbitrary trajectories in simulation; see [Supp Video S0](#).

Abstract—Integrating robots into human-centric environments such as homes, necessitates advanced manipulation skills as

robotic devices will need to engage with articulated objects like doors and drawers. Key challenges in robotic manipulation of articulated objects are the unpredictability and diversity of these objects’ internal structures, which render models based on object kinematics priors, both explicit and implicit, inadequate. Their reliability is significantly diminished by pre-interaction ambiguities, imperfect structural parameters, encounters with unknown objects, and unforeseen disturbances. Here, we present a prior-free strategy, Tac-Man, focusing on maintaining stable robot-object contact during manipulation. Without relying on object priors, Tac-Man leverages tactile feedback to enable robots to proficiently handle a variety of articulated objects, including those with complex joints, even when influenced by unexpected disturbances. Demonstrated in both real-world experiments and extensive simulations, it consistently achieves near-perfect success in dynamic and varied settings, outperforming existing methods. Our results indicate that tactile sensing alone suffices for managing diverse articulated objects, offering greater robustness and generalization than prior-based approaches. This underscores the importance of detailed contact modeling in complex manipulation tasks, especially with articulated objects. Advancements in tactile-informed approaches significantly expand the scope of robotic applications in human-centric environments, particularly where accurate models are difficult to obtain. See additional material at <https://tacman-aom.github.io>.

This work is supported in part by the National Science and Technology Major Project (2022ZD0114900), the National Natural Science Foundation of China (62376009), the Beijing Nova Program, the State Key Lab of General AI at Peking University, the PKU-BingJi Joint Laboratory for Artificial Intelligence, the National Comprehensive Experimental Base for Governance of Intelligent Society, Wuhan East Lake High-Tech Development Zone, and the Horizon Europe Framework through the PALPABLE project (101092518). (Zihang Zhao and Yuyang Li contributed equally to this work. Corresponding authors: Lecheng Ruan and Yixin Zhu.)

Zihang Zhao and Yixin Zhu are with the Institute for Artificial Intelligence, Peking University, Beijing 100871, China (email: zhaozihang@stu.pku.edu.cn; yixin.zhu@pku.edu.cn).

Yuyang Li and Zhenghao Qi are with the Institute for Artificial Intelligence, Peking University, Beijing 100871, China, and the Beijing Institute for General Artificial Intelligence, Beijing 100080, China, and also with the Department of Automation, Tsinghua University, Beijing 100084, China (emails: {liyuyang20, qi-zh21}@emails.tsinghua.edu.cn).

Wanlin Li is with the Beijing Institute for General Artificial Intelligence, Beijing 100080, China (emails: liwanlin@bigai.ai).

Lecheng Ruan is with the College of Engineering, Peking University, Beijing 100871, China, and also with the PKU-Wuhan Institute for Artificial Intelligence, Wuhan 430075, China (emails: ruanlecheng@ucla.edu).

Zihang Zhao and Lecheng Ruan were also with the Beijing Institute for General Artificial Intelligence partially during this work.

Kaspar Althoefer is with the Centre for Advanced Robotics @ Queen Mary, within the School of Engineering & Materials Science, Queen Mary University of London, London E1 4NS, UK (email: k.althoefer@qmul.ac.uk).

Digital Object Identifier (DOI): 10.1109/TRO.2024.3508134.

Index Terms—Articulated object manipulation, prior-free, tactile sensing, contact regulation.

I. INTRODUCTION

ARTICULATED objects, defined by interconnected parts that move in conjunction with each other, include a vast array of everyday items, such as doors, drawers, and appliances [1], as illustrated in Fig. 1. A robot's capability to manipulate articulated objects is essential for effective integration into human-centric environments, assisting humans in diverse tasks while representing a significant challenge within robotics [2, 3].

To interact efficiently with these objects, robots traditionally rely on prior knowledge of their kinematics. This knowledge is often embedded either explicitly, as in kinematic models used for designing manipulation strategies via control or planning techniques [4–21], or implicitly, within policy models developed through machine learning techniques [22–38]. However, acquiring precise prior knowledge about such objects presents several spatial and temporal challenges:

- **Ambiguity:** Objects with similar external appearances may have vastly different internal structures, complicating the unique determination of their kinematics through perception. This issue is extensively explored by Zhu *et al.* [39].
- **Imperfectness:** Accurate acquisition of an object's structural parameters is challenging due to perception noise or deviation from standard models, such as the off-axis doors.
- **Unknown:** Certain objects feature unique and sophisticated kinematics, making them difficult to model and generalize, as exemplified by the objects in Fig. 1(c) and (d).
- **Obsolescence:** Kinematic models can become obsolescent due to active or passive changes over time, leading to a gap in knowledge when executing manipulation tasks.

These challenges prompt us to question the hitherto assumed indispensability of prior knowledge in the manipulation of articulated objects and explore alternative approaches.

Drawing inspiration from human tactile interaction, we propose a novel tactile-informed prior-free manipulation approach, Tac-Man. Tac-Man is centered around stable robot-object contact during manipulation. By accurately sensing contact geometry information, Tac-Man leverages deviations from the initial state to inform the robot of both the direction and magnitude of adjustments needed to effectively manipulate the articulated object. Without any prior-based modeling, Tac-Man addresses the challenges posed by ambiguity, imperfectness, unknowns, and obsolescence in object kinematics, which are prevalent in dynamic human-centric environments.

Our comprehensive validation includes a series of real-world experiments as well as extensive simulation studies. The experiments demonstrate Tac-Man's proficiency in handling objects with different joint types, from basic prismatic and revolute joints to complex mechanisms with simultaneous translational and rotational movements. We conducted further simulations to scale up these real-world experiments and probe Tac-Man's adaptability and robustness. Collectively, these studies confirm the effectiveness of Tac-Man, especially in scenarios where traditional reliance on priors is unreliable or inadequate.

The present paper makes three major contributions:

- We introduce Tac-Man, a novel tactile-informed prior-free manipulation framework for articulated objects. Unlike

traditional methods that depend on prior knowledge of object kinematics, Tac-Man leverages tactile feedback to dynamically maintain stable contact during the manipulation process. This strategy ensures robust interaction with objects by adapting to real-time tactile data, offering a novel perspective on robotic manipulation.

- Through comprehensive real-world experiments, we validate the superior performance of Tac-Man over conventional methods in scenarios characterized by ambiguous, imperfect, unknown, and obsolescent priors. These experiments not only underscore Tac-Man's adaptability and precision in handling complex manipulations but also demonstrate its potential in enhancing robotic autonomy in unpredictable environments.
- We conduct an extensive simulation study to evaluate the scalability and generalization capabilities of Tac-Man across a diverse range of articulated objects and trajectories, including those arbitrarily generated. This study, one of the largest of its kind, confirms Tac-Man's effectiveness in a broad spectrum of settings, illustrating its versatility and the feasibility of its application in real-world scenarios beyond the confines of controlled experiments.

The paper is organized to provide a thorough exploration of Tac-Man and its implications. Sec. II reviews relevant literature, positioning our work within the broader context of robotic manipulation of articulated objects. Sec. III details Tac-Man, our novel tactile-informed, prior-free methodology. Sec. IV describes the setup and results of our real-world experiments, while Sec. V presents our simulation studies, showcasing the scalability and generalizability of Tac-Man. In Sec. VI, we present a detailed discussion on the strengths, limitations, and potential applications of Tac-Man, as well as considerations for future research. Finally, Sec. VII concludes the paper, summarizing our key findings and contributions to this field.

II. RELATED WORK

A. Manipulation with Explicit Priors

Traditional robotic manipulation strategies have predominantly utilized manually predefined kinematic models, supplied externally, for interaction with articulated objects [4–6]. While this approach has been effective in certain controlled scenarios, it inherently limits the autonomy of the robotic system due to its reliance on pre-established, accurate kinematic information.

Seeking to enhance system autonomy, recent developments in robotic manipulation have shifted focus toward autonomously deriving object models via robot perception. This trend is exemplified by strategies employing visual inputs to predict kinematic models [10–12, 14–16, 40]. Nonetheless, this reliance on visual perception introduces ambiguity, as objects with similar appearances can harbor different internal articulations, challenging the accuracy of model predictions [14, 39]. To address these issues, several studies have integrated multi-frame analysis and wrist-mounted force/torque sensors for more accurate kinematic model estimation [7–9, 13, 17–20]. Despite these improvements, the dependency on being able

to identify specific joint types, such as revolute or prismatic, limits the efficacy of such methods.

Our work introduces a novel prior-free approach to address these limitations inherent in manipulation methods with explicit priors. Tac-Man, validated by both real-world experiments and simulations, demonstrates superior adaptability and efficacy, especially in scenarios where traditional methods, reliant on explicit priors, are found to be unreliable or insufficient.

B. Manipulation with Implicit Priors

Robotic manipulation utilizing implicit priors typically relies on extensive data to infer appropriate actions from the observed state of an object. There are two prominent approaches: imitation learning and reinforcement learning.

Imitation learning enables robots to acquire implicit priors through the observation of human demonstrations [22–30], sometimes facilitated by tele-operational systems [41]. One major challenge in this approach is the collection of diverse and rich data from humans [42]. Although transfer methods [43–45] have improved the applicability to similar objects, extending these priors to a wider range of objects remains a substantial hurdle.

In contrast, reinforcement learning allows robots to develop implicit priors through iterative interactions in simulated settings [31–34]. The advent of extensive datasets of articulated objects [35–38] has significantly propelled this approach forward. However, the variance in these datasets often falls short in terms of replicating the complexity encountered in real-world situations, leading to incomplete prior acquisition, in turn constraining the robots’ manipulation capabilities.

While both approaches show promise in articulated object manipulation, they grapple with the issue of incomplete priors due to limited data coverage. Tac-Man, sidestepping the need for laborious data collection and extensive training, addresses these challenges and has demonstrated its efficacy in manipulating a diverse array of articulated objects.

C. Robot-Object Contact Modeling

Current research tends to oversimplify and overlook the contact between robots and objects. The assumption of a fixed contact point is prevalent in numerous models [4, 5, 8, 13, 15, 17–20], yet this is often not feasible in practical applications. [6] have introduced the concept of a virtual joint [46] to represent this contact, but in actual implementation, they still rely on the assumption of a fixed joint. Works by [7, 9] have relaxed this constraint to some extent, permitting only specific types of movement between the robot and the object. [47] estimates contact using tactile sensors. However, it is limited to a specific type of handling based on rational motion and overlooks the potential for slippage.

Our research, however, uncovers that the data obtained from the contact site itself is adequate to support the task of articulated object manipulation, enabling a prior-free approach. By utilizing advanced tactile sensors to regulate contact, Tac-Man successfully demonstrates the ability to manipulate articulated objects under a diverse array of scenarios. This prior-free

approach significantly enhances robots’ manipulation skills in comparison to existing methodologies.

D. Contact Sensing

Sensing of robot-object contacts is crucial for contact regulation [48–53]. Decades of development have yielded various sensing technologies regarding this issue, broadly categorized as intrinsic or extrinsic styles. Intrinsic tactile sensors, such as 6D force/torque sensors [54], are integrated within the robot’s structure, providing indirect contact measurements through internal force sensing. Extrinsic tactile sensors, attached to the robot’s interacting surface, often measure contact information in a more direct manner, including conductive-fluid-based (*e.g.*, BioTac [55]), resistor-based [56, 57], piezoelectricity-based [58, 59], capacitance-based [60, 61], and vision-based [62–65] sensors, each with unique principles for measuring robot-object contacts.

The effectiveness of Tac-Man relies on timely compensation for contact deviations, which are modeled geometrically as the changes in the contact point positions, effectively informing both the direction and the magnitude of the necessary adjustments. Under this formulation, sensors with a primary focus on sensing contact geometry, such as GelSight [62], TacTip [63], and others, can be directly adopted within the Tac-Man framework. Additionally, sensors primarily designed to measure contact forces can also be utilized within our Tac-Man framework, provided that they are properly calibrated according to Hooke’s law [66] to obtain geometric information. A detailed discussion on sensor selection is available in Sec. VI.

In this research, we adopt the GelSight-type sensor for the pipeline illustration and experiment verification of the Tac-Man among the set of sensors with geometric measurements as one economical option. However, it should be noted that there are no methodological preferences among these sensors with geometric measurement ability, and the formulations should be very similar to each other, with only slight differences in the pre-processing procedure where contact deviations are extracted from the raw sensor data.

E. Tactile Servoing

Tactile servoing, which involves adjusting robot poses in response to tactile signals, has been a significant area of research since its inception by Weiss *et al.* [67] and first implementation by Berger and Khosla [68]. It has evolved into two main approaches: model-based and learning-based. Model-based tactile servoing uses explicit models such as the inverse tactile model [69], the tactile Jacobian model [70, 71], the projector matrix [72, 73], or other tactile-based controllers [74, 75] to map tactile feedback onto robot actions. Learning-based tactile servoing leverages modern machine learning techniques, including imitation learning [76] and deep learning [53, 77, 78], to enable implicit mapping for complex robot manipulations. Tac-Man builds upon the fundamental principles of model-based tactile servoing, demonstrating its efficacy in the intricate manipulation of articulated objects.

III. THE TAC-MAN METHOD

This section outlines our problem formulation for prior-free manipulation of articulated objects. At the core of Tac-Man is the ability to regulate stable contact throughout the manipulation process. This involves guiding the object from its initial to its final state using just tactile signals. We begin by introducing essential notations and preliminaries in Sec. III-A, as these provide the foundational basis for our problem formulation and the subsequent derivations detailed in Sec. III-B. We then describe the contact representation in Sec. III-C. This representation is crucial in determining whether a contact is stable and this is discussed in Sec. III-D. Finally, Sec. III-E presents a computational method for calculating the robot's pose adjustments. These adjustments are essential for maintaining stable contact, thereby enabling the successful manipulation of articulated objects.

A. Notation and Preliminaries

Before delving into the detailed formulation of Tac-Man, we start by defining the following notations:

- Bold lowercase letters represent vectors (e.g., $\mathbf{p}$), with the subscript i denoting the i -th entry of the vector (e.g., p_i).
- Bold uppercase letters represent matrices (e.g., $\mathbf{M}$). The identity matrix of dimension n is denoted by $\mathbf{I}_{n \times n}$. The superscript $\cdot^T$ indicates the transpose.
- The symbol $\hat{\cdot}$ indicates the axis of a frame. The generalized pose of a frame $\{i\}$ relative to another frame $\{j\}$ is described by a position vector $\mathbf{p}_i^j \in \mathbb{R}^3$ and a rotation matrix $\mathbf{R}_i^j \in SO(3)$. For simplicity, when frame $\{j\}$ is equivalent to the world frame $\{w\}$, the superscript is omitted. The homogeneous transformation matrix $\mathbf{T}_i^j \in SE(3)$, integrating the position vector and rotation matrix, is given by:

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

- Calligraphic font letters, such as $\mathcal{C}$, are used to denote sets. The notation $|\cdot|$ represents the cardinality, or number of elements, in a set. Specially, $\mathcal{K}_{ij}$ refers to a set of corresponding element pairs from sets $\mathcal{C}_i$ and $\mathcal{C}_j$, formulated as:

$$\mathcal{K}_{ij} = \{(\mathbf{u}, \mathbf{v}) \mid \mathbf{u} \in \mathcal{C}_i, \mathbf{v} \in \mathcal{C}_j\}.$$

B. Problem Formulation

In this subsection, we formulate the process of tactile-informed prior-free manipulation of articulated objects. Consider an articulated object with a movable part, which is assumed to be graspable through a rigidly attached “handle.” As illustrated in Fig. 2, We postulate that the object's articulation restricts the handle to a 1-Degree-of-freedom (DoF) trajectory. Let us denote the handle's trajectory as J , beginning at point J_s and concluding at J_e . Initially, in the absence of priors, only J_s is known. A robotic gripper, positioned in a global pose $\mathbf{T}_1$ is grasping the handle. The objective is to navigate the gripper through a series of poses, $\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3, \dots$, to maneuver the handle towards J_e . At any given pose $\mathbf{T}_i$, the contact state

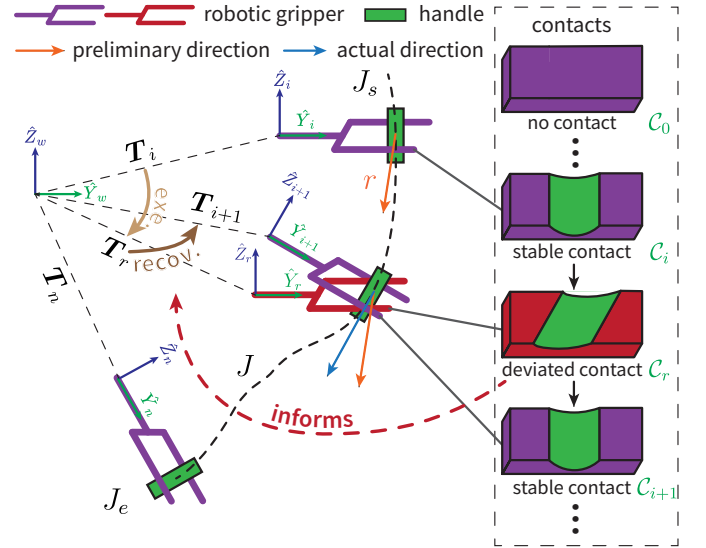


Fig. 2: **Schematic of Tac-Man.** Tac-Man starts from an initial pose and an estimated preliminary direction. Thereafter, the pose is dynamically adjusted so as to maintain stable contact throughout the manipulation process. This continuous adjustment is crucial to ensure the correct interaction direction is also maintained. This culminates in the successful manipulation of the articulated object.

$\mathcal{C}_i$ between the object handle and the robotic gripper can be ascertained via a mapping function f_c :

$$\mathcal{C}_i = f_c(J, \mathbf{T}_i). \quad (1)$$

In practical terms, $\mathcal{C}_i$ is directly measurable through tactile sensors. Importantly, the state prior to establishing any contact is denoted as $\mathcal{C}_0$.

Stable contact: To ensure successful manipulation, it is crucial that the contacts $\mathcal{C}_{1:n}$ meet certain constraints, guaranteeing that the gripper can consistently grasp the handle throughout the process. In this paper, we characterize a state that meets these constraints as maintaining *stable* contact.

Stable contact is characterized by two essential features. First, the gripper has to maintain a secure grip on the handle without causing damage to itself, while also preventing slippage. This ensures that any relative pose changes between the grip and handle are accurately mirrored in the contact changes. This concept is quantified by referencing state prior to establishing any contact $\mathcal{C}_0$ and evaluating subsequent contact states $\mathcal{C}_i$. We use the function $f_d(\mathcal{C}_i, \mathcal{C}_0)$ to calculate the deformation of $\mathcal{C}_i$; this deformation must remain within the elastic limits e of the gripper material. Moreover, the shear deformation, as calculated by the function $f_s(\mathcal{C}_i, \mathcal{C}_0)$, should not exceed s , a threshold determined by the friction ratio between the handle and gripper material. Formally, this can be expressed as:

$$\begin{bmatrix} f_e(\mathcal{C}_i, \mathcal{C}_0) \\ f_s(\mathcal{C}_i, \mathcal{C}_0) \end{bmatrix} \leq \begin{bmatrix} e \\ s \end{bmatrix}. \quad (2)$$

Second, a *stable* contact should guarantee that the movement of the gripper is synchronized with the handle's motion. To establish a benchmark for *stable* contact $\mathcal{C}_1$, an initial gripper pose $\mathbf{T}_1$ is manually configured. This methodology

is commonly employed in the manipulation of articulated objects [9]. A contact $\mathcal{C}_i$ is considered stable if its deviation from $\mathcal{C}_1$ stays within a specified threshold d :

$$f_d(\mathcal{C}_i, \mathcal{C}_1) \leq d, \quad (3)$$

where $f_d(\mathcal{C}_i, \mathcal{C}_1)$ computes the difference between $\mathcal{C}_i$ and $\mathcal{C}_1$.

Robot manipulation: Upon establishing stable contact, the manipulation process commences with a preliminary direction relative to frame $\{i\}$ for operating the articulated object. This direction, represented as r , is encapsulated by a rotation matrix $\mathbf{R}^i \in SO(3)$. A detailed discussion on obtaining it and its impact on the efficiency of the method is provided in Section VI-B. Then, the homogeneous transformation matrix $\mathbf{T}_r^i$, resulting from moving along r by a distance t , is defined as:

$$\mathbf{T}_r^i = \begin{bmatrix} \mathbf{I}_{3 \times 3} & \mathbf{R}^i \mathbf{t} \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (4)$$

where $\mathbf{t}$ denotes the translation vector whose Euclidean norm is equal to t :

$$\|\mathbf{t}\|_2 = t. \quad (5)$$

The robot follows the preliminary direction, executing a series of gripper frame poses $\mathbf{T}_{1:n}$ to guide the handle's transition from J_s to J_e while ensuring continuous stable contact. For each incremental step $i = 1, \dots, n-1$, the next pose $\mathbf{T}_{i+1}$ is computed as follows:

$$\begin{aligned} \mathbf{T}_{i+1} &= \mathbf{T}_i \mathbf{T}_{i+1}^i \\ &= \mathbf{T}_i \underbrace{\mathbf{T}_r^i}_{\text{execution}} \underbrace{\arg \min_{\mathbf{T}_{i+1}^r \in SE(3)} \|f_c(\mathbf{T}_i \mathbf{T}_r^i \mathbf{T}_{i+1}^r, J) - \mathcal{C}_1\|_2}_{\text{recovery}}. \end{aligned} \quad (6)$$

Execution: The execution term in Eq. (6) involves proceeding to move along the preliminary direction as described in Eq. (4). This navigation progresses under the assumption of no prior knowledge about the object handle's trajectory, posing a risk of exceeding the tolerance limits necessary for maintaining stable contact. To balance this risk and maintain efficiency, it is crucial to carefully select the magnitude of t in Eq. (4). The objective is to maximize t while ensuring that the constraints for stable contact are met. This optimization can be mathematically represented as:

$$\begin{aligned} &\underset{t}{\text{maximize}} \quad t \\ &\text{subject to} \quad f_e(f_c(\mathbf{T}_i \mathbf{T}_r^i, J), \mathcal{C}_0) \leq e, \\ &\quad f_s(f_c(\mathbf{T}_i \mathbf{T}_r^i, J), \mathcal{C}_0) \leq s, \\ &\quad f_d(f_c(\mathbf{T}_i \mathbf{T}_r^i, J), \mathcal{C}_1) \leq d. \end{aligned} \quad (7)$$

Recovery: The recovery term in Eq. (6) implies that at each step, the robot adjusts its pose to minimize any deviation from the stable contact state $\mathcal{C}_1$, which may arise due to the execution step.

C. Contact Representation

For the effective computation of Eq. (6) and optimization of Eq. (7), a suitable representation of $\mathcal{C}_i$ is crucial. To ensure compatibility with the homogeneous transformation matrix and to match the capabilities of current tactile sensors, we

represent the contact $\mathcal{C}_i$ as a set of positions of all discretized points at the contact site, relative to the reference frame $\{i\}$. The position of each contact point within $\mathcal{C}_i$ is denoted by $\mathbf{p}^i$. To facilitate alignment with the homogeneous transformation matrix, the column vector representation of $\mathbf{p}^i$ is defined in an augmented format as follows:

$$\mathbf{p}^i = [p_x^i \ p_y^i \ p_z^i \ 1]^\top, \quad (8)$$

where p_x^i , p_y^i , and p_z^i represent the coordinates relative to the gripper frame $\{i\}$.

Significantly, $\mathcal{C}_0$ includes all points on the gripper's outer surface. Assuming the total number of points is m , $\mathcal{C}_0$ is represented as:

$$\mathcal{C}_0 = \{\mathbf{p}_j^0 \mid j = 1, \dots, m\}. \quad (9)$$

For simplicity, we assume a flat gripper surface, a common assumption in current robotic grippers and easily achievable with tactile sensor fabrication. Consequently, p_x^0 is consistent across all points and set to 0, attainable through coordinate transformation. The position of a contact point at $\mathcal{C}_i$ is then characterized by the condition:

$$|p_1^i| \geq \epsilon \quad (10)$$

where ϵ is a small positive number, indicating that deformation in the $\hat{x}$ direction must exceed ϵ for it to be considered contact.

Therefore, $\mathcal{C}_i$ for $i = 1 \dots n$ can be defined as:

$$\mathcal{C}_i = \{\mathbf{p}^i \mid |p_1^i| \geq \epsilon\}. \quad (11)$$

D. Stable Contact Formulation

With the established representation of contact, we now turn to the computational process for determining stable contact.

The computation of stable contact involves addressing the gripper's maximum elastic deformation, encompassing both normal and shear components. These are captured in the expressions for $f_e(\mathcal{C}_i, \mathcal{C}_0)$ and e within the context of stable contact. Ensuring that the contact deformation remains within the material's elastic limits is crucial for maintaining a stable grip on the object:

$$f_e(\mathcal{C}_i, \mathcal{C}_0) = \left[\max \left\{ \left\| \begin{bmatrix} u_2 - v_2 \\ u_3 - v_3 \end{bmatrix} \right\|_2 \mid \begin{matrix} \max\{|u_1| \mid \mathbf{u} \in \mathcal{C}_i\} \\ (\mathbf{u}, \mathbf{v}) \in \mathcal{K}_{0i} \end{matrix} \right\} \right] \quad (12)$$

and

$$\mathbf{e} = [e_n \ e_s]^\top, \quad (13)$$

where d_n and d_s denote the maximum normal and shear elastic deformation, respectively. For notation simplicity, we denote $|u_1|$ as ΔN and $\left\| \begin{bmatrix} u_2 - v_2 \\ u_3 - v_3 \end{bmatrix} \right\|_2$ as ΔS henceforth.

Similarly, we formulate $f_s(\mathcal{C}_i, \mathcal{C}_0)$ and s to address concerns related to slipping in Eq. (2). Practically, to prevent sliding, the friction f between the object and the gripper should not surpass the maximum static friction f_M . In our model, the friction at $\mathcal{C}_i$ is positively correlated with the shear deformation observed from $\mathcal{C}_0$:

$$f \sim \Delta S. \quad (14)$$

Thus, the shear deformation must adhere to the following constraints:

$$\Delta S \leq \delta^M(f_N), \forall (\mathbf{u}, \mathbf{v}) \in \mathcal{K}_{0i}, \quad (15)$$

where δ^M represents the deformation associated with the maximum static friction f_M , proportional to the normal force f_N . Assuming linear elasticity in the normal direction, δ^M can be modeled as:

$$\delta^M(\Delta N) = \delta_0 \Delta N, (\mathbf{u}, \mathbf{v}) \in \mathcal{K}_{0i}, \quad (16)$$

with δ_0 being a constant determined experimentally:

$$\delta_0 = \frac{\delta^M(\Delta N_0)}{\Delta N_0}, (\mathbf{u}, \mathbf{v}) \in \mathcal{K}_{0i}. \quad (17)$$

The no-sliding constraint is thus defined as:

$$\frac{\Delta S}{\Delta N} \leq \delta_0, \forall (\mathbf{u}, \mathbf{v}) \in \mathcal{K}_{0i}, \quad (18)$$

leading to the expression for $f_s(C_i, C_0)$:

$$f_s(C_i, C_0) = \max \left\{ \frac{\Delta S}{\Delta N} \mid (\mathbf{u}, \mathbf{v}) \in \mathcal{K}_{0i} \right\}, \quad (19)$$

and setting

$$s = \delta_0. \quad (20)$$

Additionally, in the context of Eq. (3), we intuitively compute it as the difference between the corresponding points of two contacts, considering our established contact representation. To quantify this difference, we apply the loss function outlined in Besl *et al.* [79]. However, to ensure fairness in cases where a contact has a larger number of contact points, we normalize this loss by the number of corresponding pairs in $\mathcal{K}_{1i}$:

$$f_d(C_i, C_1) = \frac{\sum_{(\mathbf{u}, \mathbf{v}) \in \mathcal{K}_{1i}} \|\mathbf{u} - \mathbf{v}\|_2}{|\mathcal{K}_{1i}|}. \quad (21)$$

E. Robot Pose Update

Having established the contact representation and computation methods for stable contact constraints, we should address the optimization in Eq. (7) to compute the value of t for the determination of $\mathbf{T}_r^i$. However, in the absence of knowledge about the handle's trajectory J , determining t explicitly remains challenging. Therefore, we adopt an exploration mechanism by continuously executing with small increments in t until any term in the constraints of optimization Eq. (7) reaches its upper bound, scaled by a safety margin α ($0 < \alpha < 1$). This process is concluded as

$$\mathbf{T}_r = \mathbf{T}_i \mathbf{T}_r^i, \quad (22)$$

which is referred to as the *execution stage*.

To recover the contact to within the specified bounds after the execution stage, we employ the Kabsch algorithm [80] to compute the optimal transformation $\mathbf{T}_{i+1}^r$ for the recovery term in Eq. (6):

$$\mathbf{T}_{i+1}^r = \arg \min_{\mathbf{T}_{i+1}^r \in SE(3)} \sum_{(\mathbf{u}, \mathbf{v}) \in \mathcal{K}_{1r}} \|\mathbf{T}_{i+1}^r \mathbf{u} - \mathbf{v}\|_2. \quad (23)$$

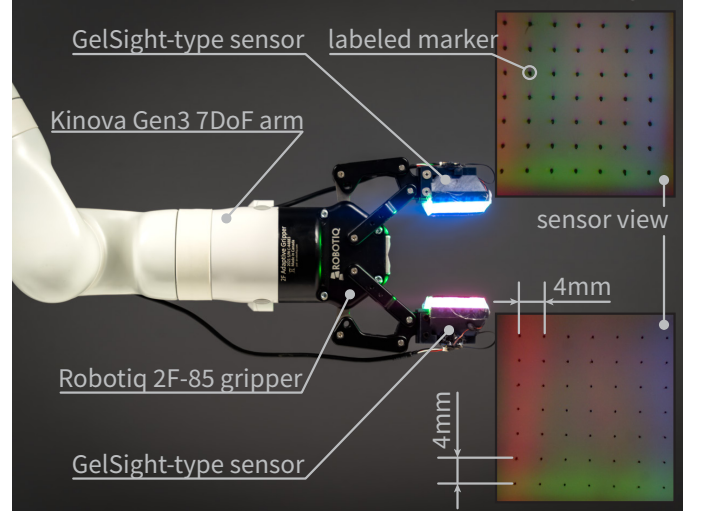


Fig. 3: **Real-world experiment setup.** The experiments are conducted in a real-world environment using a Gen3 7DoF robotic arm paired with a Robotiq 2F-85 gripper. To equip it with the ability to sense contact geometry, we adopt two GelSight-type tactile sensors, each featuring a 7×7 grid of labeled markers for point correspondence between frames.

The updated robot pose is then given by:

$$\mathbf{T}_{i+1} = \mathbf{T}_r \mathbf{T}_{i+1}^r. \quad (24)$$

This phase is referred to as the *recovery stage*. The update of the robot pose follows an iteration between the two stages.

IV. REAL-WORLD EXPERIMENTS

To assess the effectiveness of Tac-Man, our proposed approach to prior-free manipulation of articulated objects, we conducted a series of real-world experiments. These were designed with meticulous attention to detail, as outlined in Sec. IV-A. They aim to thoroughly evaluate key constraints and determine essential hyper-parameters for the task, as discussed in Sec. III. Our experiments showcase the strengths of Tac-Man, particularly its robustness and adaptability under various conditions. These include scenarios where priors for object articulation are ambiguous (Sec. IV-B), imperfect (Sec. IV-C), unknown (Sec. IV-D), or obsolescent (Sec. IV-E), as previously discussed in Sec. I.

A. Experimental Setup

Robot system setup: Tac-Man is designed for broad compatibility with robotic systems capable of facilitating 6D movement in the gripper. In experiments, we utilized a Kinova Gen3 7DoF robotic arm equipped with a Robotiq 2F-85 gripper (Fig. 3). To assess the contact constraints as outlined in Sec. III, we integrated two GelSight-type tactile sensors into the gripper.

Tactile sensor: To further enhance our ability to validate these constraints, we adopt GelSight-type tactile sensors, which effectively provide geometry information at the contact surface. The back of the deformable pad is rigidly attached to a rigid acrylic sheet. Each sensor was specially fabricated with a unique 7×7 grid of labeled markers (Fig. 4) on the

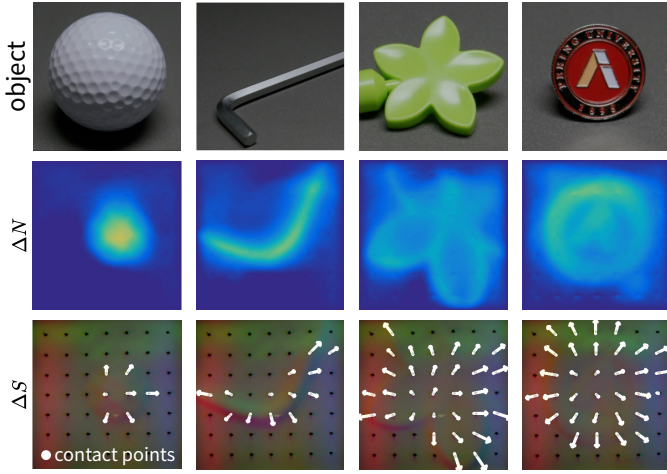


Fig. 4: **Visualization of the contact geometry detected by the sensors adopted.** The GelSight-type sensors are adept at providing contact geometry information ΔN in the normal direction during object interactions. Labeled markers are used for accurate point correspondence to further enhance their effectiveness. Utilizing these features, ΔS is calculated and visualized using white arrows in the illustration. For better visibility, the magnitude of these arrows is exaggerated by a factor of 5, ensuring a clearer visualization of the contact dynamics.

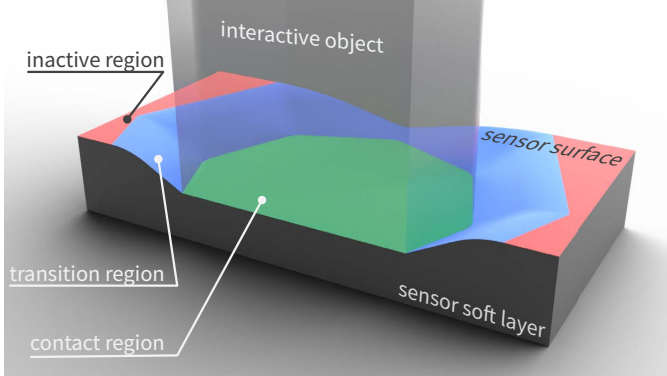


Fig. 5: **Selection concerns for the ϵ value.** The compliance inherent in the tactile sensor material results in the activation of both the direct contact region and the adjacent transitional areas when contacting the handle. This characteristic underlines the need for a carefully adaptable ϵ value. Such adaptability ensures the precise identification of true contact points, thereby reducing the likelihood of detecting unreliable points.

front of the pad. Without any surrounding structural geometry, these markers share the same mechanical characteristics, allowing for calibration using consistent parameters. This grid aids in accurately identifying point correspondences between two contacts. Although this modification slightly reduces the resolution and depth reconstruction accuracy as compared to conventional GelSight sensors, its effectiveness in practical scenarios and in validating the contact constraints detailed in Sec. III is notable.

System hyper-parameters: The determination of values for the defined parameters is crucial. The following outlines the chosen values, with a summary provided in Tab. I:

- The threshold for identifying contact points, ϵ in Eq. (10), is set to be adaptable. This is due to the material’s flexibility,

as shown in Fig. 5, which results in deformation not only in the contact region but also in the surrounding area upon contact with the object. This deformation relates to ΔN and the handle’s shape, among other factors. For adaptability, ϵ is defined as the minimal ΔN that ensures at least 8 contact points, exceeding the number of markers in a line (7 in our sensors), thus ensuring these points are non-colinear.

- The normal elastic deformation limit, e_n in Eq. (13), is set at 2 mm. This value is chosen based on observations that within this range, the system can revert to its initial state after releasing contact. It is important to note that this value represents a conservatively reduced normal elastic deformation for enforcing stricter constraints.
- The parameter δ_0 in Eq. (17) is linked to handle material properties. To set a lower bound, an experiment with a slippery Lego piece, depicted in Fig. 6(a), identifies the δ^M values at different ΔN . This is done by tracking the peak shear displacement for each marker, as shown in Fig. 6(b). The results from multiple trials, illustrated in Fig. 6(c), lead to setting δ_0 at 1.5 for safety.
- The shear elastic deformation limit, e_s in Eq. (13), is calculated as $e_n \delta_0$. Though smaller than the actual one, this value is deemed acceptable for stricter constraint enforcement.
- The upper bound for another stable contact constraint, d in Eq. (3), is experimentally determined to be 0.4 mm, ensuring it does not exceed e_n , e_s , and δ_0 during manipulation.
- The movement magnitude in the *execution stage*, the small increment in t established in Eq. (7), is set as $\epsilon \delta_0$ to preserve stable contact during execution.

TABLE I: **Values of system hyper-parameters**

hyper-parameter	ϵ	e_n	e_s	δ_0	d	α
value	adaptive	2 mm	3 mm	1.5	0.4 mm	0.6

The experiments described below are conducted in accordance with the setup and hyper-parameters detailed above. This ensures consistency and relevance in evaluating the efficacy of Tac-Man in real-world scenarios.

B. Manipulation under Ambiguous Priors

Ambiguous visual priors pose a significant challenge in robotic manipulation, especially when objects with visually indistinguishable appearances harbor varying articulation mechanisms. This dilemma is discussed in Sec. II-A and exemplified by Zhu *et al.* [39]. To illustrate, Fig. 7(a)–(d) features four objects that, despite their identical appearances, are equipped with different articulations: a cabinet with a revolute joint on the left (a), another with the joint on the right (b), a third with a joint at the bottom (c), and a drawer with a prismatic joint (d). This diversity underscores the complexity of inferring object kinematics solely from visual information.

Tac-Man effectively demonstrates proficient manipulation of distinct objects in all four cases, as shown in Fig. 7(e)–(h), each corresponding to the configurations in Fig. 7(a)–(d). Starting from the same initial state with stable contact $\mathcal{C}_1$, Tac-Man employs an iterative two-stage cycle, beginning with an identical preliminary direction $\mathbf{R}^1$. During the *execution*

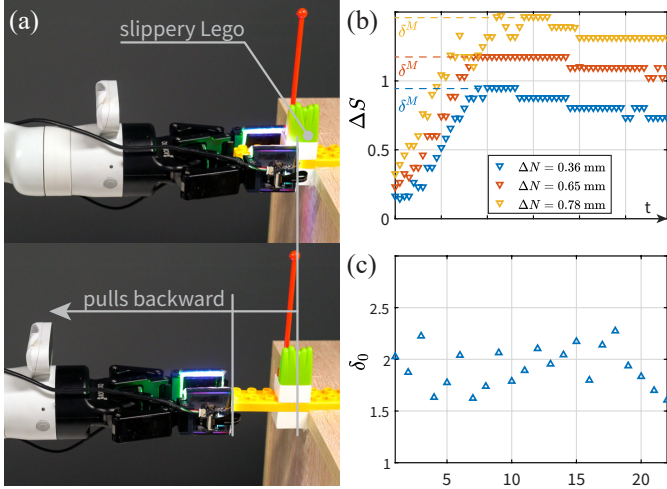


Fig. 6: δ_0 value acquisition. The experiment to obtain hyperparameter δ_0 , as defined in Eq. (17): (a) The experiment begins with the arm pulling backward after the gripper has firmly grasped a slippery Lego piece. (b) Multiple trials are conducted at different values of ΔN . Three representative trials are visualized to illustrate the process. The results follow the friction model well and δ^M can be easily identified. The results align well with the friction model, allowing for the straightforward identification of δ^M . (c) The δ_0 obtained from these multiple trials is visualized, and a value of 1.5 is selected for safety considerations.

stage, tactile feedback is utilized to discern deviations from the correct interaction direction by observing changes in the contact points. For example, the contact in (e) exhibits a shear counterclockwise deviation, indicating a need for an additional clockwise rotation in the robot's movement, and conversely for the shear clockwise deviation observed in (f). In the scenario depicted in (g), a deviation in the normal direction ΔN signals the requirement for an added rotation for recovery. Conversely, in (h), the contact remains stable as the arm's movement is congruent with the correct direction. When a predetermined deviation threshold is reached, the system enters the *recovery stage*, employing the optimization outlined in Eq. (23) to compute the optimal arm pose that compensates for the observed deviation, thereby realigning the gripper to the correct interaction direction. Following this adjustment, the process resumes along the preliminary direction $\mathbf{R}^i$. Through continual refinement to preserve stable contact, the arm successfully manipulates the objects into their intended states, as depicted in the last images in Fig. 7.

C. Manipulation under Imperfect Priors

Imperfect priors present a significant challenge in robotic manipulation, even when the articulation structure is known. Errors in model parameters can make these priors unreliable, as demonstrated by the example of an off-axis door, depicted in Fig. 8(a). Unlike conventional doors with hinges located along one edge, the off-axis door features a hinge positioned away from the edge, complicating the acquisition of accurate priors.

The impact of imperfect priors is illustrated through a challenging manipulation task, in which the robot is programmed to follow an arc trajectory to open a microwave

oven with a revolute joint, as shown in Fig. 8(b). However, a 10 % error in the trajectory's radius introduces a small but critical imperfection in the priors. This error, without any real-time corrective adjustments, accumulates a loss that eventually leads to the gripper losing contact with the handle, resulting in task failure.

In contrast, Fig. 8(c) demonstrates Tac-Man's approach, employing adaptive adjustments during the manipulation to effectively counteract the inaccuracies in the priors. By dynamically compensating for the unforeseen deviation, Tac-Man ensures alignment with the correct interaction direction. This adaptability allows Tac-Man to successfully complete the task, fully opening the microwave oven door, showcasing its efficacy in overcoming the limitations imposed by imperfect priors.

D. Manipulation under Unknown Priors

Encountering objects whose articulation mechanisms defy accurate perception introduces a scenario where priors on kinematics remain unknown. Such conditions, characterized by complex movements that cannot be succinctly represented by conventional kinematic models like prismatic or revolute joints, present a significant challenge. Existing research has scarcely addressed the manipulation of objects requiring combined translational and rotational motions due to the difficulty in modeling these actions explicitly. For example, Fig. 9(a) illustrates an object designed compactly, necessitating that its handle executes both linear and rotational movement simultaneously. Another example is shown in Fig. 9(b), where a vise handle follows a spiral path. The ability to accurately identify and model such complex motion patterns remains beyond the reach of current perception methods.

Tac-Man circumvents this challenge, as demonstrated with two examples of objects whose handles exhibit complex movements combining translational and rotational motions. Through iterative processing, the Tac-Man adjusts the robot to the correct interaction direction and successfully manipulates the object in both examples.

E. Manipulation under Obsolescent Priors

Dynamic environments often introduce unpredictable perturbations, such as human interventions or environmental turbulence, rendering initially accurate priors for object kinematics obsolescent. To evaluate Tac-Man's performance in this context, we conducted experiments depicted in Fig. 10 in which the position of the articulated object was deliberately altered during manipulation. This was achieved by applying force to one side of the object, leading to a shift in either the rotation axis of a cabinet (Fig. 10(a)) or the opening direction of a drawer (Fig. 10(b)), as indicated by the hand icons.

Despite encountering such unpredictable disturbances, Tac-Man demonstrates remarkable adaptability. It swiftly compensates for the changes by correctly adjusting the robot's interaction strategy, thus enabling successful manipulation. This resilience underscores Tac-Man's ability to effectively handle real-time perturbations, ensuring consistent manipulation performance even when initial kinematic priors are no longer valid due to unforeseen environmental interactions.

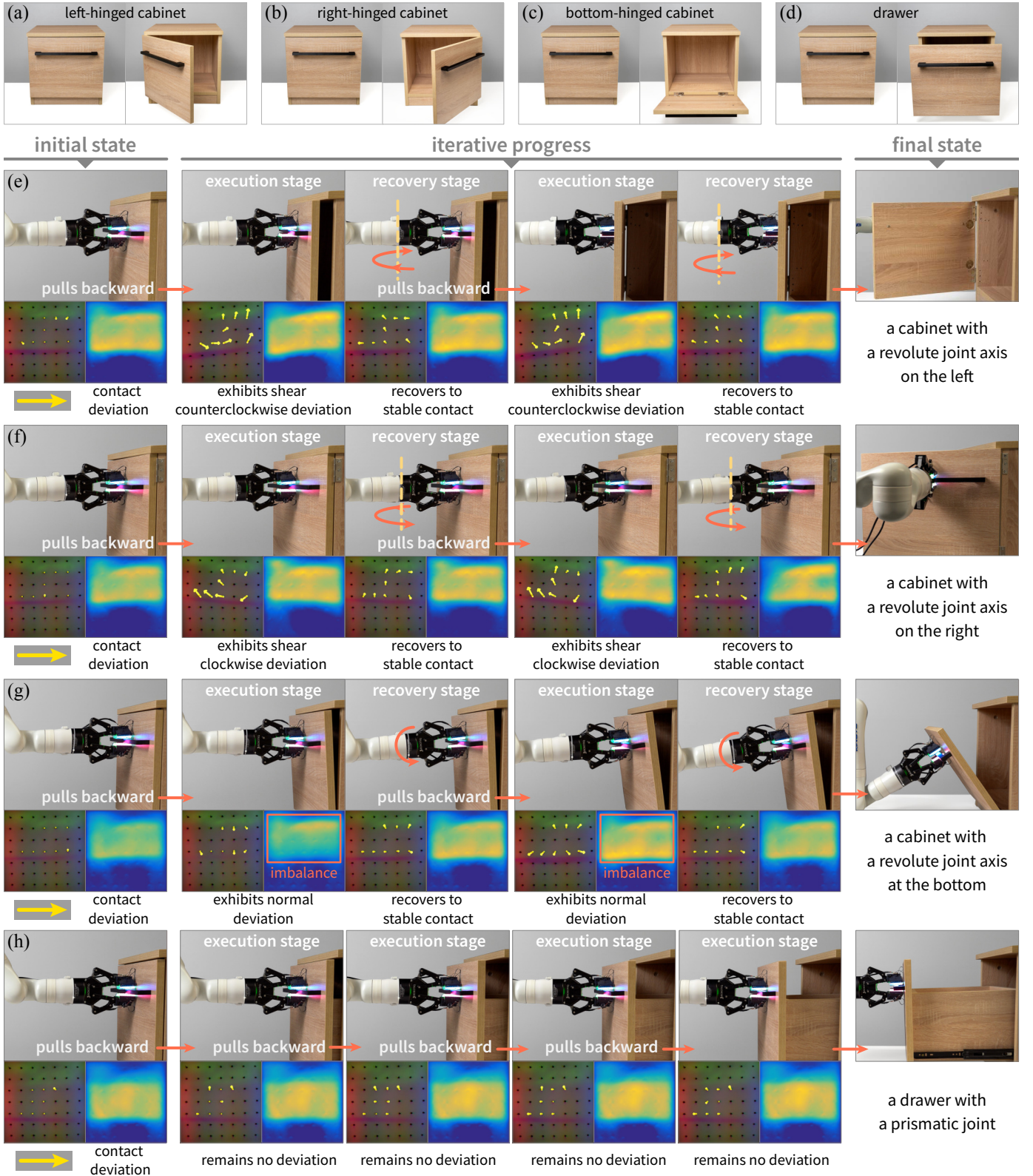


Fig. 7: Manipulation under ambiguous priors. Tac-Man showcases its proficiency in addressing the challenge caused by ambiguous priors, characterized by objects that, while visually indistinguishable, feature distinct articulation mechanisms. Our evaluation involves four articulated objects that look identical but differ in articulation: (a) a cabinet with a revolute joint on the left; (b) another cabinet with a revolute joint on the right; (c) a cabinet with a revolute joint at the bottom; and (d) a drawer with a prismatic joint. In scenarios (e)–(h), corresponding to objects (a)–(d), Tac-Man starts from the same position and in the same direction. The initial backward pull by the arm reveals variations in contact loss due to the differing articulations. Upon reaching a certain deviation threshold, the system triggers an adjustment in the arm’s pose. This adjustment, optimized by Eq. (23), aims to recover stable contact and re-align to the correct interaction direction. The arm then continues its backward pull. Through this iterative process of adjustment and execution, each object, starting from the same initial state, is maneuvered to a distinct final state reflective of its unique articulation. For video demonstrations, refer to [Supp Video S1](#).

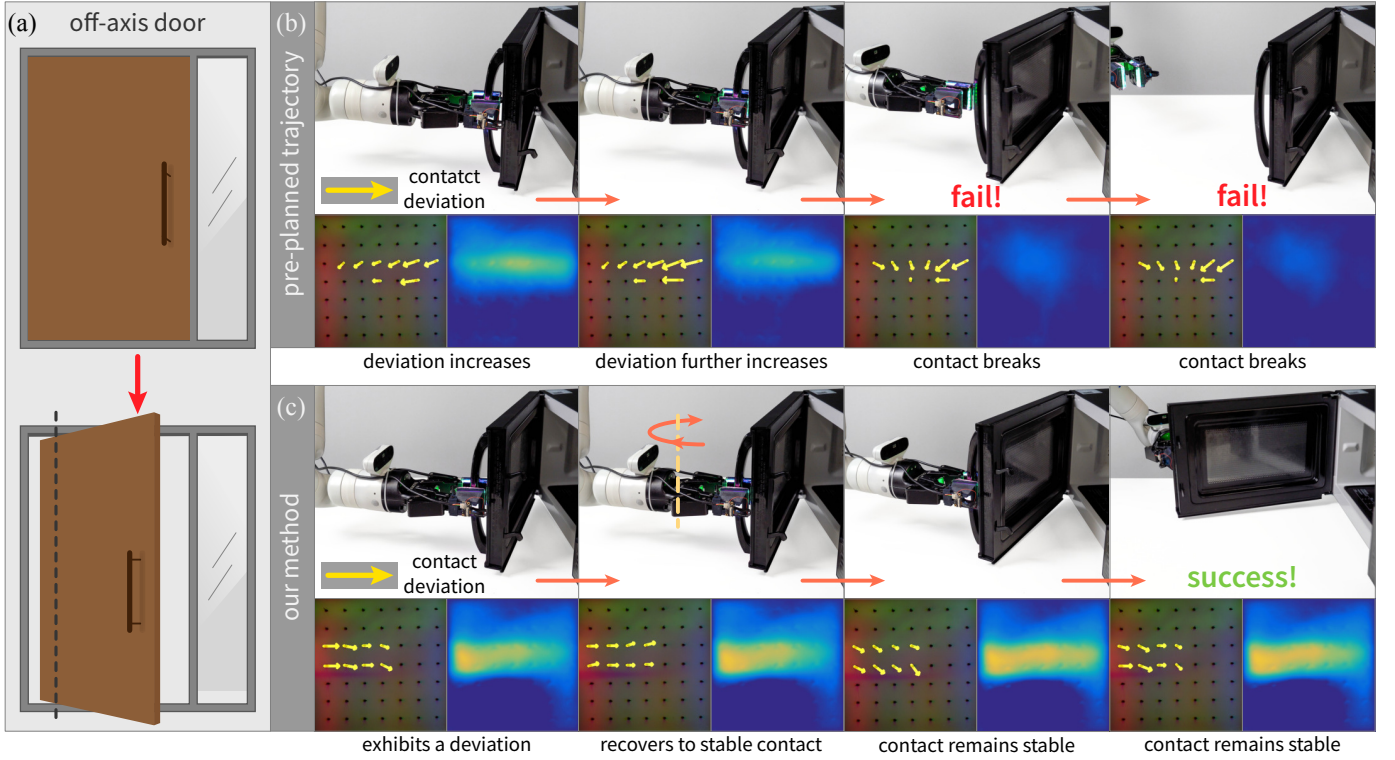


Fig. 8: **Manipulation under imperfect priors.** Tac-Man adeptly addresses challenges posed by imperfect priors, which accurately identify the model type but contain erroneous parameters. (a) An example is an off-axis door, designed for space and energy efficiency, diverging from traditional models by not placing the hinge directly on one side. This variation can mislead prior assumptions regarding the arc trajectory’s radius. (b) We mirror this scenario by assigning the robot a pre-programmed arc trajectory to open a microwave oven with a revolute joint, incorporating a 10 % radius error. This slight misalignment between the priors and actual parameters leads to a gradual loss, eventually causing the gripper to lose contact with the handle, resulting in a failed manipulation attempt. (c) In contrast, Tac-Man dynamically recalibrates the arm’s pose to correct any unexpected errors, ensuring alignment with the correct interaction direction, going on to successfully open the microwave oven. This demonstration of adaptability underscores Tac-Man’s ability to complete tasks even with imperfect priors. For video demonstrations, refer to [Supp Video S2](#).

V. LARGE-SCALE VERIFICATION IN SIMULATION

Given the logistical challenges of physically experimenting with a comprehensive variety of objects, we look to simulation studies to further validate the generalization capability of Tac-Man. Utilizing NVIDIA Isaac Sim, our simulations cover a wide range of objects, mirroring the setup of our real-world experiments as outlined in [Sec. V-A](#). For a thorough assessment, we incorporate objects with prismatic and revolute joints from the PartNet-Mobility dataset [36], reflecting the common articulations found in existing datasets.

Acknowledging the limitation of current datasets of objects with simpler joint types, we further extend our verification to include articulated objects designed with complex manipulation trajectories. These objects are uniquely characterized by trajectories that follow randomly sampled curves, introducing additional challenges to the manipulation task.

To showcase the effectiveness of Tac-Man, we select several representative prior-based and prior-free methods as baselines, as detailed in [Sec. V-B](#). In [Sec. V-C](#), we present both qualitative and quantitative results from our simulations. The findings compellingly demonstrate Tac-Man’s proficiency in manipulating objects with a wide range of articulations, highlighting its adaptability and superior generalization across diverse scenarios compared to the baseline methods.

A. Environmental Setup

Our simulation environment, depicted in [Fig. 11](#), features a robot arm equipped with a gripper poised to engage an object’s handle. Initially, the robot is programmed to follow a preliminary direction R^1 , applying Tac-Man to articulate the object. A critical aspect of Tac-Man is its reliance on the compliance of silica gel material to accurately track contact between the gripper and the handle. To mimic this in a rigid body simulation environment, we employ specialized designs for contact measurement, detailed in [Appx. A](#).

The simulation divides test objects into two categories: standard objects with prismatic or revolute joints featuring various handle designs from existing datasets, and randomly created custom objects with complex trajectories. From the PartNet-Mobility dataset [36], we select objects with detailed articulation annotations provided by GPartNet [38], as outlined in [Appx. B](#). For the evaluation, 25 objects with prismatic joints and another 25 with revolute joints were pre-processed. A manipulation is deemed successful if a revolute joint rotates beyond 60.0° or a prismatic joint extends over 250.0 mm. [Fig. 11\(a\)](#) showcases a subset of these objects and their initial grasps.

To assess Tac-Man’s adaptability beyond basic joint types, we also evaluate it on articulated objects designed with

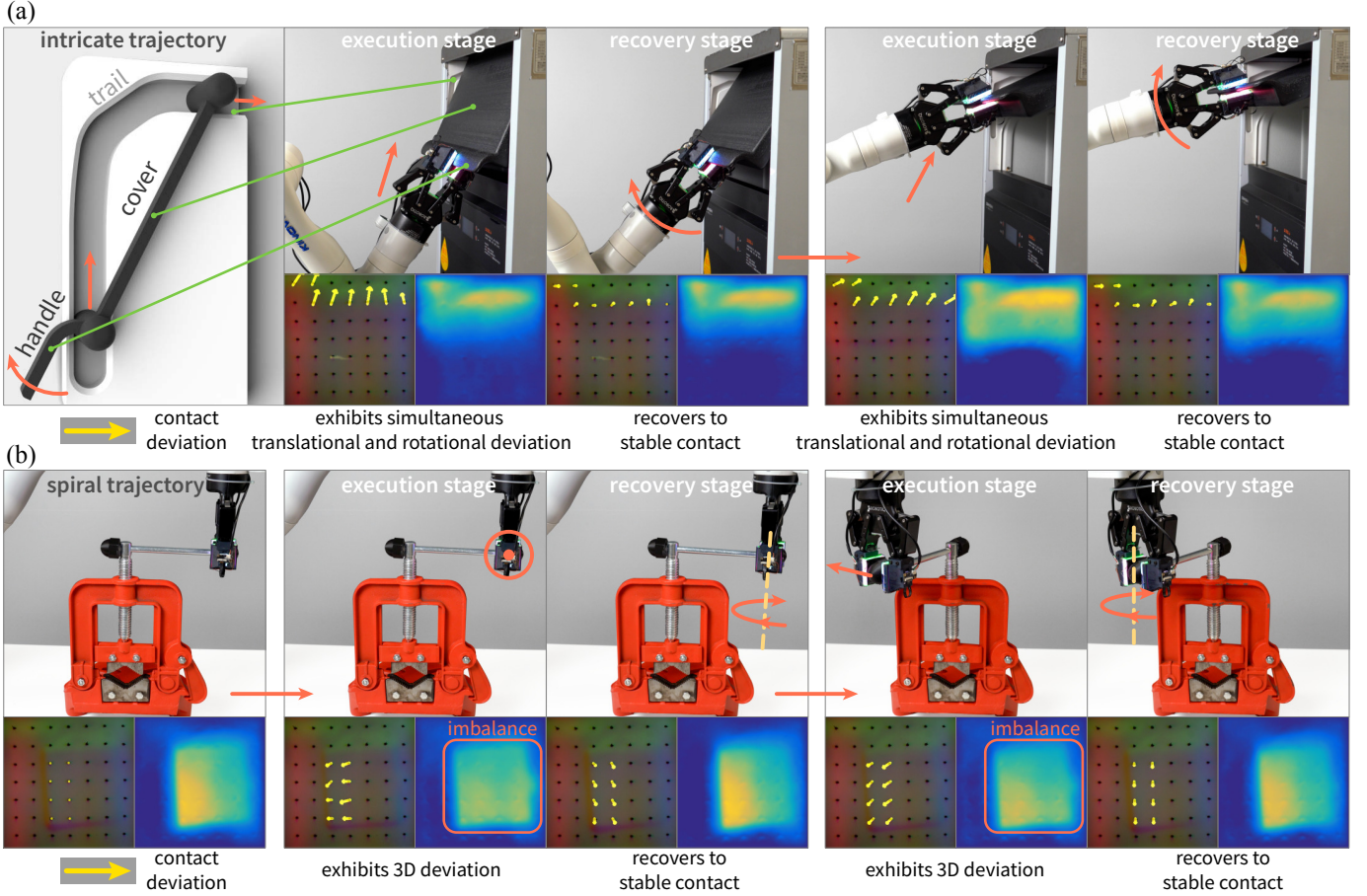


Fig. 9: Manipulation under unknown priors. Tac-Man showcases exceptional proficiency in managing scenarios characterized by unknown priors, particularly when object kinematics defy prior-based prediction or exceed the limits of current robotic perception. We illustrate this capability with two distinct examples. In scenario (a), the object’s articulation mechanism, which is concealed beneath a cover, involves motions that combine translation and rotation—a complexity not easily captured by standard prismatic or revolute joint models. Scenario (b) presents a handle executing a helical trajectory, with the precise determination of its pitch posing a substantial challenge. Through these examples, Tac-Man proves its adeptness at navigating the intricacies of complex kinematic patterns, underscoring its effectiveness in scenarios where traditional perception models may falter. For video demonstrations, refer to [Supp Video S3](#).

randomly generated trajectories. These are constructed by sampling Bézier curve segments of various orders (2, 3, 4, and 5) on a 2D plane, ensuring no self-intersections. Each curve inspires a 3D playboard that defines a manipulation path, with a toy train at the start point acting as the handle. Physical collisions are configured to restrict the handle’s movement along the trajectory, with the construction process detailed in [Appx. C. Fig. 11\(b\)](#) illustrates an example of a 3-order Bézier curve, its corresponding play-board, and the simulation’s initial state. Additional examples of generated play-boards are presented in [Fig. 11\(c\)](#), highlighting the increased complexity and manipulation challenge with higher-order curves. For the evaluation, 25 curves of orders 2 to 5 are sampled, each tested once. Success is achieved when the gripper moves the toy train from one end of the curve to the other.

B. Selected Baselines

To demonstrate the effectiveness of Tac-Man, we also select several representative prior-based or prior-free methods designed for the manipulation of articulated objects to conduct comparisons with Tac-Man. Specifically, we evaluate

the performance of three distinct techniques: the prior-based pre-planned trajectory generation method, impedance control in a prior-free setting, and a vision-based learning approach, FlowBot3D [19]. Below, we detail the setup for each of these methods:

Prior-based pre-planned trajectory: This method is highly effective when the kinematic priors of articulated objects are precisely known. However, meeting this requirement proves challenging in the real world due to pervasive uncertainties and unexpected perturbations [39]. To account for this in our simulations, we introduce noise ξ into the ideal kinematic model of the test objects. This noise ξ is modeled as a normal distribution with a mean of zero and a variance of 0.01. We explore the impact of this noise on the pre-planned trajectory generation method under the same initial conditions as Tac-Man for different joint types. For revolute joints, the relationship between the actual handle trajectory radius r_{gt} and the manipulated trajectory radius r_m (both in meters) is:

$$r_m = r_{gt} + \xi. \quad (25)$$

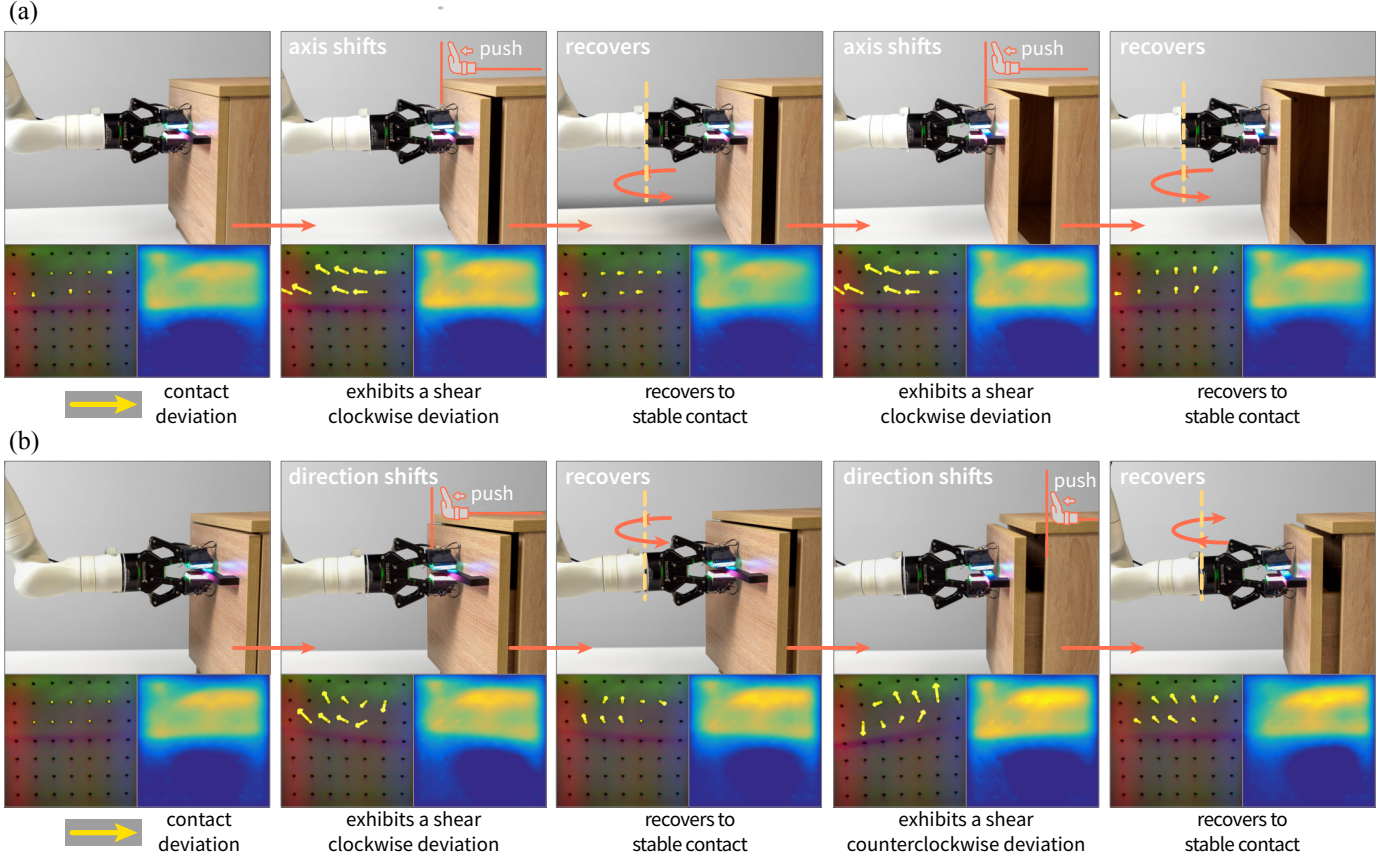


Fig. 10: **Manipulation under obsolescent priors.** Tac-Man demonstrates exceptional performance in scenarios characterized by obsolescent priors, often resulting from unpredictable perturbations that modify object kinematics. Such perturbations are common in dynamic, human-centric environments, frequently emerging from unanticipated human interactions. To assess Tac-Man’s resilience against these perturbations, we introduced random forces to one side of the object in scenarios (a) and (b), effectively altering the cabinet’s rotation axis in (a) and the drawer’s opening direction in (b). Despite the introduction of errors through these human-induced disturbances, Tac-Man adeptly adjusts, preserving the correct interaction direction and ensuring successful manipulation outcomes. For video demonstrations, refer to [Supp Video S4](#).

For prismatic joints, the relationship between the actual handle trajectory angle θ_{gt} and the manipulated trajectory angle θ_m (both in radians) is:

$$\theta_m = \theta_{gt} + \xi. \quad (26)$$

For Bézier curves, the relationship between the actual control point position $\mathbf{p}_{gt}$ and the manipulated trajectory control point position $\mathbf{p}_m$ (both in meters), excluding the starting point, is:

$$\mathbf{p}_m = \mathbf{p}_{gt} + [\xi \quad \xi]^T. \quad (27)$$

Impedance control: Impedance control, introduced by Hogan [81], is a robust technique widely used in robotic manipulation tasks. We compare impedance control with Tac-Man, following the same preliminary direction. In this context, impedance control serves as a prior-free approach for manipulating articulated objects and serves as a baseline in an ablation study to assess the essential role of tactile sensing in the Tac-Man.

FlowBot3D: FlowBot3D, introduced by Eisner *et al.* [19], is a recent advancement using modern learning techniques for generalized manipulation based on visual feedback. It employs a flow prediction model to predict part mobility

and compute actions accordingly. It should be noticed that, to align with our simulation settings, we apply only the flow prediction model and action computation strategy. We assess its performance using the same tasks described in [Sec. V-A](#) as those used for Tac-Man, without additional fine-tuning.

C. Results

Our simulation results substantiate the efficacy of the setup in evaluating Tac-Man. In [Fig. 12](#), we highlight five representative examples: (a–c) illustrate the manipulation of objects with standard prismatic and revolute joints, while (d–e) detail scenarios involving intricate trajectory manipulations. The direction of pull and subsequent adjustments are indicated by the orange arrows. The simulated tactile feedback is visualized in the grey area beneath each figure, in which inactive, reference and current contact points are depicted as grey dots, blue circles, and red dots, respectively. Orange arrows, enlarged five-fold for clarity, depict the displacement from reference to current markers, with their size increasing in proportion to the error magnitude.

Consistent with our real-world experiments, deviations in the manipulation direction lead to deviation ([Fig. 12\(a-c\)](#)) in

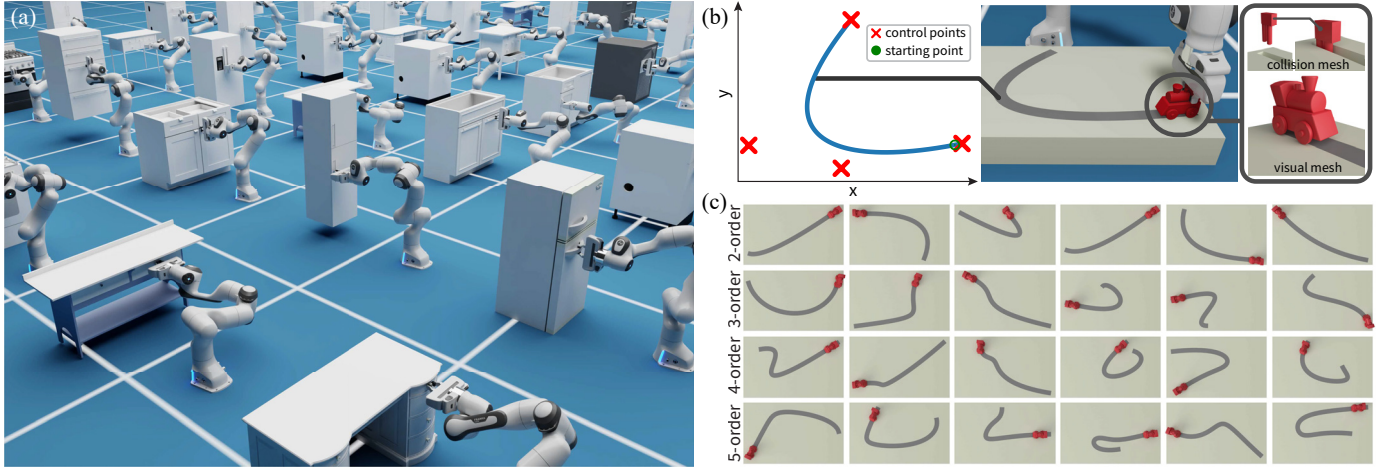


Fig. 11: **Simulation setup.** (a) The setup includes objects with prismatic and revolute joints from the PartNet-Mobility dataset [35], shown in their initial simulation states. (b) We further introduce a playbook featuring an intricate manipulation trajectory based on a randomly sampled Bézier curve; control points of the curve are marked with red “x”s. At the starting point of this curve, a toy train is placed to act as the manipulation handle (right), with a detailed collision setup (below). (c) Further examples of generated playboards, where rows 1-4 correspond to playboards with 2nd, 3rd, 4th, and 5th-order Bézier curves as the trajectories, showcasing the range of complexity in manipulation paths.

the tactile pattern, signaling the need for corrective movements beyond the preliminary direction. Once the contact deviation surpasses $\alpha\epsilon$ (Fig. 12(b–e)), Tac-Man adeptly adjusts to re-establish stable contact.

The quantitative results for the three baseline methods and Tac-Man are summarized in Fig. 13. To provide a more informed measure of performance beyond binary success or failure, we introduce a distance-weighted success rate SR_W , defined as:

$$SR_W = \frac{\min(d_A, d_E)}{d_E} \times 100\%, \quad (28)$$

where d_E represents the expected contact point moving distance required for success, as specified in Sec. V-A, and d_A is the actual handle moving distance calculated from the final handle position in each trial.

Fig. 13 presents both mean and standard deviation of the weighted success rate across objects under different categories. Specifically, columns 1-2 display weighted success rates for manipulating objects with prismatic and revolute joints, while columns 3-6 show performance on intricate manipulation trajectories, with PB- n indicating performance on Bézier trajectories of order n . Notably, Tac-Man achieves a 100% weighted success rate across all objects with various handle designs, markedly outperforming the three baseline methods and verifying its effectiveness. Key observations for the baseline methods include:

- **Pre-planned trajectory methods:** In scenarios lacking tactile feedback, these methods are vulnerable to noise, affecting their performance.
- **Impedance control:** Following the same preliminary direction as Tac-Man, yet without object kinematics priors and tactile feedback, this method proves effective for objects with prismatic joints. However, objects requiring movements beyond a straight trajectory may quickly surpass the compliance limits, resulting in failures.

- **FlowBot3D:** This method demonstrates effectiveness with familiar objects featuring prismatic and revolute joints. However, it underperforms on unfamiliar, intricate trajectories.

VI. DISCUSSION

A. Advancing Robotic Manipulation with Tac-Man

Our study has validated the efficacy of Tac-Man, a tactile-informed, prior-free approach for manipulating articulated objects. The core of Tac-Man is leveraging tactile feedback to discern discrepancies between the preliminary direction and actual interaction directions, informing dynamic adjustments to ensure precise manipulations. A standout feature of Tac-Man is its operational independence from pre-existing knowledge about object kinematics, presenting a considerable advantage over traditional methods reliant on priors. This capability is particularly beneficial in environments where priors may be ambiguous (Sec. IV-B), imperfect (Sec. IV-C), unknown (Sec. IV-D), or obsolescent (Sec. IV-E)—conditions prevalent in dynamic, human-centric settings. While classical control methods like impedance control can operate without object priors, they may fail with complex trajectories, as shown in Fig. 13.

Tac-Man enhances robotic autonomy by obviating the need for manually inputting accurate object kinematic models, a common requirement in earlier strategies [4–6]. It sidesteps the limitations of methods that depend on visual perception for determining object kinematics [10–12, 14–16, 21], which are vulnerable to the ambiguities and inaccuracies intrinsic to visual data [39], as demonstrated by the reduced effectiveness of pre-planned methods under noise in Fig. 13. Unlike approaches that utilize multi-frame observations to clarify ambiguities [7–9, 13, 17–20], Tac-Man does not presuppose the existence of specific articulation types, thus demonstrating its versatility in managing novel articulation mechanisms as evidenced in Sec. IV-D and Fig. 13.

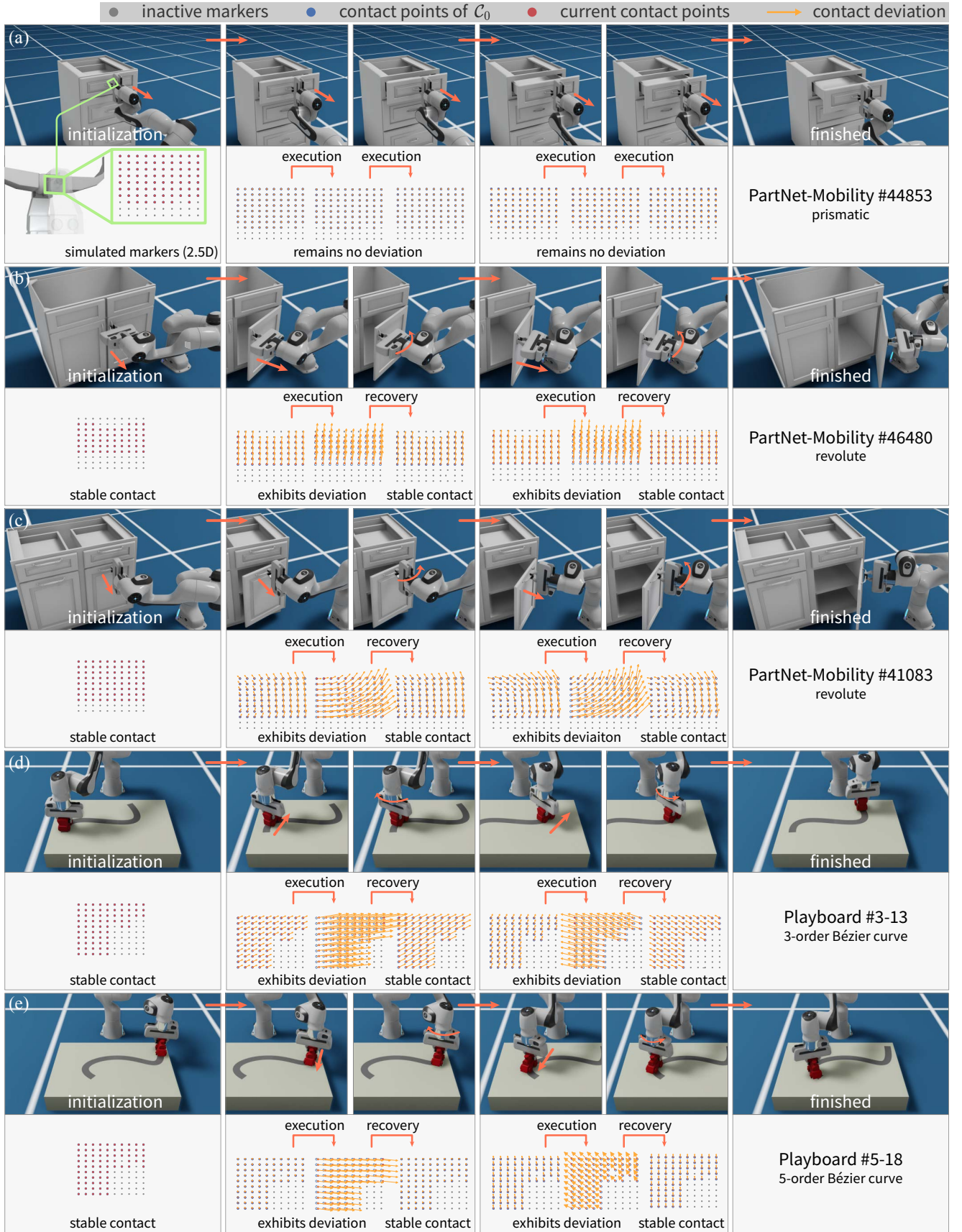


Fig. 12: **Examples of simulated manipulation.** Each row captures a sequence of manipulating an articulated object, showcasing scenarios with basic joint types (a-c) and intricate trajectories (d-e). For each scenario, the depiction includes the initial and final states, the *execution-recovery* iterations, and the tactile patterns observed. This underscores the effectiveness of our simulation setup in demonstrating Tac-Man’s capability to adapt and succeed across varying articulation challenges. For video demonstrations, refer to [Supp Video S5](#).

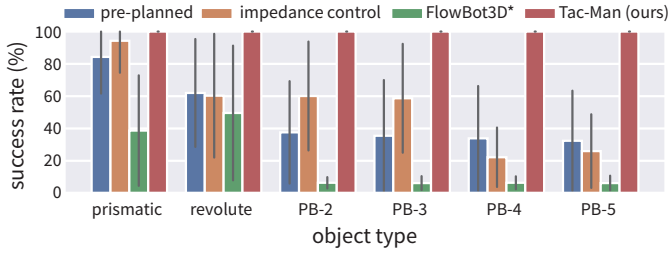


Fig. 13: **Barplot of the distance-weighted success rates of different methods across six articulated object types.** Tac-Man achieves 100% success rate, outperforming the other three baseline methods across all the articulated object types (* Tested on Isaac Sim using our settings without fine-tuning on PB- n).

Moreover, Tac-Man circumvents the exhaustive data collection process typically required by learning-based methods to tackle articulated object manipulation challenges [22–38]. The diversity and complexity of articulated objects, ranging from basic prismatic and revolute joints to more intricate mechanisms featuring simultaneous translation and rotation, pose a significant challenge for data-driven approaches, as shown by FlowBot3D’s lower success rates on unseen, intricate trajectories. Yet, Tac-Man adeptly addresses these challenges without relying on extensive datasets, thereby facilitating autonomous exploration and data collection for previously unseen objects. This autonomous data-gathering capability enhances the data collection process’s versatility and substantially reduces human labor.

Additionally, Tac-Man showcases remarkable adaptability to dynamic changes within environments, as explored in Sec. IV-E. Its ability to adjust to unforeseen human interventions underscores a vital strength for robots operating in human-centric environments, where static assumptions about object kinematics can lead to manipulation failures. This adaptability is essential for ensuring seamless human-robot co-existence, enabling robots to effectively navigate the inherent unpredictability of real-world settings.

B. Limitations and Future Work

Compatibility across different sensors: Tac-Man’s effectiveness primarily stems from its ability to compensate for contact deviation, which is modeled in a geometric perspective so that both direction and magnitude of deviation can be obtained, as detailed in Sec. III-E. This paper demonstrates the implementation of Tac-Man with GelSight-based sensor, as one economic representative for those sensors with the ability to measure contact geometry. We would like to clarify that the high-resolution property of GelSight is not a major consideration in this work, as indicated in the setup for experiments. Actually, all these sensors can be effectively adopted within the Tac-Man framework and should have a very similar formulation, with only minor variation in the pre-processing procedure where contact deviation geometry is extracted from raw sensor data. For example, the TacTip family [63] uses an array of pins to detect surface deformations, which can be mapped to contact points in a manner similar to our current implementation. Therefore, one important future work will

be to investigate whether the extraction of contact deviation geometry with different sensors can be generalized to some degree to improve the succinctness of Tac-Man formulation.

Force/torque sensors, while primarily measuring forces and torques, can also be adapted for use within the Tac-Man framework, with an extra step of calculating geometric contact deviation, and thus the magnitude for compensation, from the force/torque measurements by Hooke’s Law [66]:

$$d\mathbf{f} = K d\mathbf{c} \quad (29)$$

where $\mathbf{c} \in \mathbb{R}^6$ represents the geometric contact deviation, and $\mathbf{f} \in \mathbb{R}^6$ the force/torque measurement. K refers to the contact stiffness matrix, which can be constant or varying depending on the interaction surface, and often needs proper calibration before use. Therefore, another interesting future direction will be the investigation of contact stiffness calibration, which may further motivate the possibility of exploring force-based formulation for Tac-Man, where obtaining the magnitude of contact deviation from force becomes a major point.

Therefore, Tac-Man in principle has the ability to accommodate various types of sensors. Conducting comparative studies to evaluate the impact of different sensor types on Tac-Man’s performance in different scenarios remains an important area for future work. These investigations could assist in optimizing sensor selection under specific tasks and potentially lead to a more generalized and succinct formulation of Tac-Man that can seamlessly incorporate data from diverse sensor types.

Initial grasp generation and preliminary direction acquisition: The effectiveness of Tac-Man relies on obtaining a stable initial contact $\mathcal{C}_1$ and having a preliminary (rough) direction to manipulate an articulated object. In this study, we manually establish the initial (rough) grasp and preliminary direction for all experiments. Crucially, Tac-Man’s success stems from its ability to adjust control input on the fly based on tactile feedback; hence, this initial grasp need not be precise and can sometimes even be intentionally wrong. To demonstrate Tac-Man’s resilience, we conducted experiments on a drawer (see Fig. 14), providing the robot arm with intentionally incorrect grasps and directions. Despite these inaccuracies, Tac-Man adapts and successfully completes the task of opening the drawer.

Although Tac-Man does not require prior knowledge, such as object kinematics and learning policy, which was once deemed essential in the manipulation of articulated objects, its dependence on initial grasp and preliminary direction highlights a limitation but also suggests a potential area for future research. Improving the precision of learning-based initial grasps and directions [45, 82, 83], or developing capabilities within Tac-Man to autonomously refine them, could enhance its effectiveness and broaden its application across various scenarios involving articulated objects. These improvements could potentially lead to more adaptable and autonomous robotic manipulation systems.

Grasping force: The initial grasping force critically influences Tac-Man’s efficiency by directly impacting the stable contact state $\mathcal{C}_1$ and associated tactile feedback. While it minimally affects the computation of the optimal transformation matrix in Eq. (23) (which relies on relative changes

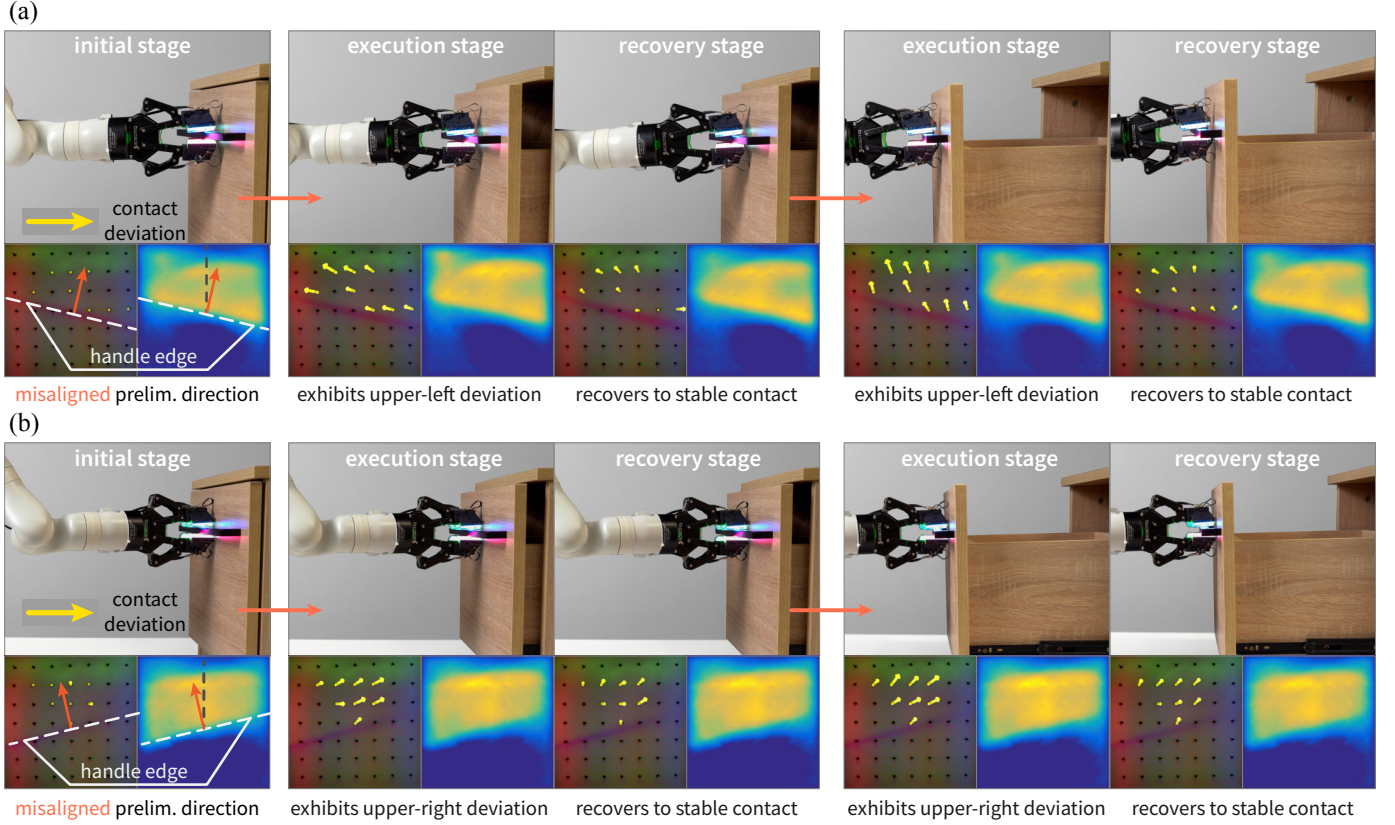


Fig. 14: **Manipulation under imperfect initial grasp and misaligned preliminary direction.** Even given an imperfect initial grasp and a misaligned preliminary direction, Tac-Man is still able to use it to successfully open the drawer.

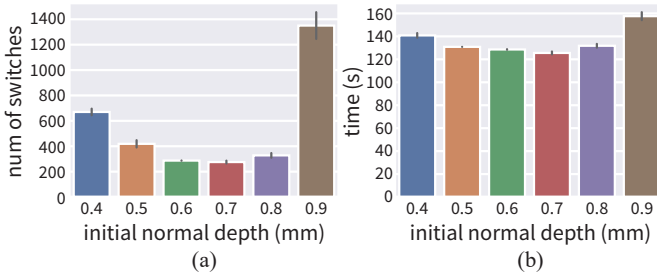


Fig. 15: **Impact of grasping force.** Extreme forces significantly reduce the efficiency of Tac-Man by affecting the computation of y , the maximum distance during the execution stage, in the optimization program (7). This impact results in an increase in both the number of (a) execution-recovery cycles and (b) execution times. Crucially, within the elastic deformation region, Tac-Man yields successful manipulations of articulated objects despite varied execution time, demonstrating Tac-Man’s robustness.

to C_1 for contact stabilization), it significantly influences the calculation of t , the maximum distance during the execution stage. This t value, derived from the optimization program (7), depends on absolute changes. Simulation results in Fig. 15 demonstrate that both excessively large and small grasping forces prolong manipulation times. This aligns with intuition: larger forces may exceed the gripper’s maximum elastic deformation capacity, while smaller forces increase slippage risk. Both scenarios are more likely to violate the stable contact constraint (2), reducing the maximum t value from program

(7) and necessitating additional iterations between execution and recovery stages. Crucially, within the elastic deformation region, Tac-Man yields successful manipulations of articulated objects despite varied execution time, demonstrating its robustness. Based on these findings, we practically select a grasping force that yields a deformation of approximately half the maximum normal elastic deformation e_n . While this approach is effective, future work should explore methods for adaptively selecting the optimal grasping force to further enhance performance across various manipulation scenarios.

Time efficiency: While the Tac-Man demonstrates notable capabilities in the prior-free manipulation of articulated objects, an area for potential enhancement is its time efficiency. As indicated in Tab. II and Fig. 16, the completion time for manipulation tasks using Tac-Man tends to exceed that of human performance. This disparity is largely attributed to the current implementation of Tac-Man, which operates without the benefits of closed-loop control.

TABLE II: **Time in second for completing tasks in the real world.**

Task	Fig. 7(e)	Fig. 7(f)	Fig. 7(g)	Fig. 7(h)	Fig. 8(b)	Fig. 9(a)	Fig. 9(b)
Time (s)	≈ 480	≈ 600	≈ 480	≈ 180	≈ 750	≈ 280	≈ 900

In its present form, Tac-Man employs a cautious, step-by-step exploratory approach. This method, devoid of pre-established object priors or learning, requires careful evaluation of, and response to, tactile feedback. The task duration figures reported in Tab. II and Fig. 16 primarily reflect this

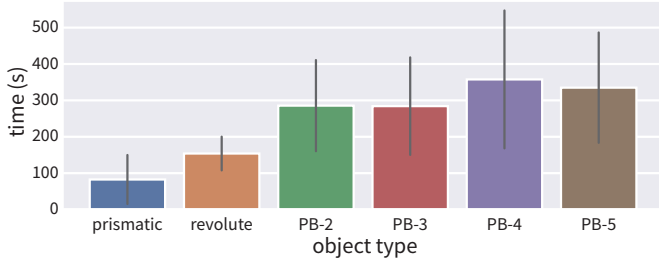


Fig. 16: Time in second for completing tasks in simulation studies.

initial phase of exploration, critical for the robot to accurately interpret and interact with the object.

Following this initial phase, the robot accumulates insights into the object’s kinematics, aiding in the development of effective priors for future manipulations. These priors facilitate quicker subsequent manipulations, yet the absence of closed-loop control could lead to inefficiencies, particularly when initial conditions vary between attempts.

Looking ahead, enhancing Tac-Man to include velocity-based control elements could significantly improve its time efficiency. Such a development would enable quicker and more responsive adaptations during initial exploration, potentially reducing the overall time required for manipulations. It would lead to a more balanced approach, achieving manipulation robustness and efficiency on a par with human performance, especially in dynamic environments as discussed in Sec. IV-E.

Integrating Tac-Man into learnable framework: Tac-Man, as discussed in Sec. II-E, is a model-based tactile servoing method. This approach offers benefits such as reduced data collection requirements and enhanced explainability [84]. However, integrating Tac-Man into a learnable framework represents a promising avenue for future research. This integration aims to generalize the strategy of utilizing tactile sensing in articulated object manipulation to a broader range of challenges, including dexterous manipulation [3], cable following [74], surface tracking [53], and others.

VII. CONCLUSION

This paper introduces a novel prior-free approach to the manipulation of articulated objects, characterized by its reliance on tactile feedback rather than predefined object models. Central to this approach is the regulation of stable contact throughout the manipulation process, enabling dynamic adjustments of the robot’s grip to ensure successful interaction with various objects.

The efficacy of Tac-Man is demonstrated through a series of experiments and simulations, highlighting its robustness in diverse scenarios. This includes its ability to handle ambiguous perceptions, adapt to imperfect parameter estimates, and respond effectively to unforeseen perturbations. The versatility of Tac-Man is further evidenced by its ability to manage objects with intricate trajectories in simulation.

In summary, this research makes a timely contribution to the field of robotic manipulation. It introduces a flexible and intuitive solution for interactions with articulated objects in environments marked by complexity and uncertainty. The

TABLE A1: Hyper-parameters defined in Appxs. A and B.

Para.	n_{res}	${}^{\text{sim}}\epsilon$	ζ_n	n_{ctrl}	H	W	w	r_{pin}	η
Value	10	0.25 mm	2 mm	3-6	40 cm	60 cm	4 cm	1.5 cm	0.02

adoption of a prior-free methodology marks a step towards robotic systems that are more adaptable and capable of performing sophisticated manipulation tasks, paving the way for enhanced automation and efficiency in various applications.

ACKNOWLEDGMENT

We thank Hongjie Li (PKU), Yida Niu (PKU), and Zimo He (PKU) for setting up the experiments; Prof. Hao Dong (PKU), Zeyi Li (PKU), and Ruihai Wu (PKU) for setting up the Franka Robot; Dr. Chi Zhang (BIGAI) and Dr. Muzhi Han (UCLA) for insightful discussions; and Mr. Mish Toszeghi (QMUL) for meticulously proofreading our manuscript.

APPENDIX

A. Details in Simulation Setup

As detailed in Sec. III-B, Tac-Man exploits the compliance of the elastic deformable layer on the gripper to guide manipulation. This deformation provides critical tactile feedback by informing the gripper-object contact, which Tac-Man utilizes to adjust the gripper to maintain stable contact. Following previous studies [85], we employ rigid-body simulations to efficiently and rigorously evaluate our method on a wide range of objects. Despite their limitation in handling deformable materials, we do not use soft-body materials due to much lower speed and higher computational demands. To closely replicate the tactile feedback mechanism under these constraints in the rigid-body simulation, we devised specific adaptations to estimate the behavior of a deformable layer, as detailed below. ${}^{\text{sim}}(\cdot)$ denotes simulated variables.

Simulating sensor markers: We simulate the positions of the markers of the tactile sensors with point cloud grids. Specifically, we affix a point cloud grid, ${}^{\text{sim}}\mathcal{C}_0 = \{\mathbf{p}_i^0 | i = 1, \dots, m\}$, on the grasping area of each gripper fingertip. We also add Gaussian noise $\gamma \sim \mathcal{N}(0, \zeta^2)$ to the point cloud to imitate real-world sensor noise.

Simulating contact deviation: Initially, the gripper stably grasps the handle, and the initial contact set ${}^{\text{sim}}\mathcal{C}_1$ is determined as the markers close enough to the handle:

$${}^{\text{sim}}\mathcal{C}_1 = \{\mathbf{p}^1 | d(\mathbf{p}^1) < {}^{\text{sim}}\epsilon\}, \quad (\text{A1})$$

where $d(\mathbf{p})$ denotes the distance from point $\mathbf{p}$ to the handle. This contact set is then transformed to the handle’s frame $\{h\}$ and affixed to it, denoted as ${}^{\text{sim}}\mathcal{C}_{h1}$. At each time step i during manipulation, the contact set ${}^{\text{sim}}\mathcal{C}_i$ comprises the coordinated of the points affixed to the handle relative to the current gripper frame:

$${}^{\text{sim}}\mathcal{C}_i = \left\{ \mathbf{T}_{h_i}^i \mathbf{p}^{h_i} + [\gamma \ \gamma \ \gamma \ 0]^T | \mathbf{p}^{h_i} \in {}^{\text{sim}}\mathcal{C}_{h_i} \right\}. \quad (\text{A2})$$

The discrepancy between ${}^{\text{sim}}\mathcal{C}_1$ and ${}^{\text{sim}}\mathcal{C}_i$ in the gripper’s frame quantifies the contact deviation under the non-sliding stable contact conditions (Eq. (2)). Based on this information,

the iterative process between the *execution* and *recovery* stages in the simulation adheres to the methodology detailed in [Sec. III](#) and utilizes the same hyper-parameters as in the real-world experiments ([Tab. I](#)).

Emulating the hypothetical deformable layer: In rigid body simulations, we cannot simulate the elastic deformable layer, which causes misaligned dynamics compared to real-world scenarios. The fingers' movement directly applies frictional forces to the object handle to move it, rather than through the contact between the layer and the handle. This can lead to unrealistic movements of the object part during the recovery stage, where [Eq. \(6\)](#) does not guarantee a successful recovery. To address this discrepancy, we have empirically increased the damping within the joint settings during this stage to emulate the effect of the deformable layer. This modification suppresses the unintended object movement, aligning the simulated interactions more closely with those observed in real-world conditions. Although these adaptations do not perfectly mirror real-world trajectories, they provide a reasonable framework to assess how our algorithm tracks and maintains contact stability, thereby validating our approach.

Determining hyper-parameters: Key hyper-parameters play a significant role in our simulation studies, and their values are determined based on empirical observations and testing, which we report in [Tab. A1](#) to ensure transparency and reproducibility of the results.

B. Preprocess PartNet-Mobility

The PartNet-Mobility dataset [\[36\]](#) comprises 1045 articulated objects, and GPartNet [\[38\]](#) enriches it with detailed part-level annotations. In this dataset, each object includes a base link and one or more moving links, connected via prismatic or revolute joints. Objects with multiple joints often necessitate sequential manipulations to fully access all their links (*e.g.* opening an oven door before pulling out the grill inside). For simplicity, we select one accessible handle on each of them for manipulation.

C. Generate Random 1-DoF Playboards

Our simulation setup incorporates playboards consisting of a baseboard (white) and a toy train (red), and the train is restricted to a predetermined trajectory (grey). To achieve this, we constructed a groove on a cuboid board and set up a distance joint between the board and the train, limiting its movement to within the plane of the board. Two cylindrical locating pins beneath the train's collision model are inserted into the groove to guide the train along the trajectory. The trajectories are sampled as Bézier curves, $B(s)$, $s \in [0, 1]$, with n_{ctrl} control points uniformly sampled and scaled to fit an $H_{\text{cm}} \times W_{\text{cm}}$ area, resulting in $(n_{\text{ctrl}} - 1)$ -order curves. Any self-intersecting curves, taking into account a padding of w , are discarded. The groove is carved into a $(H + 2w)$ cm $\times$ $(W + 2w)$ cm board following the curve. The initial position of the train is set at $B(\eta)$ with a small positive η to prevent initial pin-border collisions. The hyper-parameters for this setup are listed in [Tab. A1](#).

REFERENCES

- [1] 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.
- [2] E. Krotkov, D. Hackett, L. Jackel, M. Perschbacher, J. Pippine, J. Strauss, G. Pratt, and C. Orlowski, "The darpa robotics challenge finals: Results and perspectives," *Journal of Field Robotics*, vol. 34, no. 2, pp. 229–240, 2017.
- [3] A. Billard and D. Kragic, "Trends and challenges in robot manipulation," *Science*, vol. 364, no. 6446, p. eaat8414, 2019.
- [4] 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.
- [5] 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.
- [6] Z. Jiao, Z. Zhang, X. Jiang, D. Han, S.-C. Zhu, Y. Zhu, and H. Liu, "Consolidating kinematic models to promote coordinated mobile manipulations," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] B. Abbatematteo, S. Tellex, and G. Konidaris, "Learning to generalize kinematic models to novel objects," in *Conference on Robot Learning (CoRL)*, 2019.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] M. Han, Z. Zhang, Z. Jiao, X. Xie, Y. Zhu, S.-C. Zhu, and H. Liu, "Scene reconstruction with functional objects for robot autonomy," *International Journal of Computer Vision (IJCV)*, vol. 130, no. 12, pp. 2940–2961, 2022.
- [17] 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.
- [18] Z. Jiang, C.-C. Hsu, and Y. Zhu, "Ditto: Building digital twins of articulated objects from interaction," in *Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR)*, 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] 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.
- [21] Z. Zhang, L. Zhang, Z. Wang, Z. Jiao, M. Han, Y. Zhu, S.-C. Zhu, and H. Liu, "Part-level scene reconstruction affords robot interaction," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023.
- [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. Welschehold, 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] 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.
- [25] C. Lynch, M. Khansari, T. Xiao, V. Kumar, J. Tompson, S. Levine, and P. Sermanet, "Learning latent plans from play," in *Conference on Robot Learning (CoRL)*, 2020.
- [26] 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.
- [27] Y. Qin, Y.-H. Wu, S. Liu, H. Jiang, R. Yang, Y. Fu, and X. Wang, "Dexmv: Imitation learning for dexterous manipulation from human videos," in *Proceedings of European Conference on Computer Vision (ECCV)*, 2022.
- [28] J. Wong, A. Tung, A. Kurenkov, A. Mandlekar, 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.
- [29] 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.
- [30] J. Ye, J. Wang, B. Huang, Y. Qin, and X. Wang, "Learning continuous grasping function with a dexterous hand from human demonstrations," *IEEE Robotics and Automation Letters (RA-L)*, vol. 8, no. 5, pp. 2882–2889, 2023.
- [31] 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.
- [32] 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.
- [33] 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.
- [34] 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.
- [35] 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.
- [36] 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.
- [37] 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.
- [38] 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.
- [39] 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.
- [40] L. Ma, J. Meng, S. Liu, W. Chen, J. Xu, and R. Chen, "Sim2Real²: Actively building explicit physics model for precise articulated object manipulation," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 11698–11704, IEEE, 2023.
- [41] 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.
- [42] 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*, no. 99, pp. 1–16, 2022.
- [43] 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.
- [44] 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.
- [45] P. Li, T. Liu, Y. Li, Y. Geng, Y. Zhu, Y. Yang, and S. Huang, "Gendexgrasp: Generalizable dexterous grasping," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2023.
- [46] N. Likar, B. Nemec, and L. Žlajpah, "Virtual mechanism approach for dual-arm manipulation," *Robotica*, vol. 32, no. 6, p. E3, 2014.
- [47] A. J. Schmid, N. Gorges, D. Goger, and H. Worn, "Opening a door with a humanoid robot using multi-sensory tactile feedback," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2008.
- [48] R. S. Dahiya, G. Metta, M. Valle, and G. Sandini, "Tactile sensing—from humans to humanoids," *IEEE Transactions on Robotics (T-RO)*, vol. 26, no. 1, pp. 1–20, 2009.
- [49] R. S. Dahiya and M. Valle, *Robotic tactile sensing: technologies and system*, vol. 1. Springer, 2013.
- [50] R. S. Dahiya, P. Mittendorf, M. Valle, G. Cheng, and V. J. Lumelsky, "Directions toward effective utilization of tactile skin: A review," *IEEE Sensors Journal*, vol. 13, no. 11, pp. 4121–4138, 2013.
- [51] 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.
- [52] M. Kaboli and G. Cheng, "Robust tactile descriptors for discriminating objects from textural properties via artificial robotic skin," *IEEE Transactions on Robotics (T-RO)*, vol. 34, no. 4, pp. 985–1003, 2018.
- [53] J. Lloyd and N. F. Lepora, "Pose-and-shear-based tactile servoing," *International Journal of Robotics Research (IJRR)*, vol. 43, no. 7, p. 1024–1055, 2023.
- [54] ATI Industrial Automation, "ATI multi-axis force/torque sensor system." Available online: <https://www.ati-ia.com/products/ft/sensors.aspx>, 2024. Accessed: 2024.
- [55] 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.
- [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] G. Tian, W. Deng, Y. Gao, D. Xiong, C. Yan, X. He, T. Yang, L. Jin, X. Chu, H. Zhang, *et al.*, "Rich lamellar crystal baklava-structured pzt/pvdf piezoelectric sensor toward individual table tennis training," *Nano Energy*, vol. 59, pp. 574–581, 2019.
- [59] P. Lv, J. Qian, C. Yang, T. Liu, Y. Wang, D. Wang, S. Huang, X. Cheng, and Z. Cheng, "Flexible all-inorganic sm-doped pmn-pt film with ultrahigh piezoelectric coefficient for mechanical energy harvesting, motion sensing, and human-machine interaction," *Nano Energy*, vol. 97, p. 107182, 2022.
- [60] 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.
- [61] 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)*, pp. 614–619, IEEE, 2023.
- [62] 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.
- [63] 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.
- [64] W. Li, A. Alomainy, I. Vitanov, Y. Noh, P. Qi, and K. Althoefer, "F-touch sensor: Concurrent geometry perception and multi-axis force measurement," *IEEE Sensors Journal*, vol. 21, no. 4, pp. 4300–4309, 2020.
- [65] W. Li, M. Wang, J. Li, Y. Su, D. K. Jha, X. Qian, K. Althoefer, and H. Liu, "L³ f-touch: A wireless gelsight with decoupled tactile and

three-axis force sensing,” *IEEE Robotics and Automation Letters (RA-L)*, 2023.

- [66] P. L. Gould and Y. Feng, *Introduction to linear elasticity*, vol. 2. Springer, 1994.
- [67] 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.
- [68] 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.
- [69] N. Chen, H. Zhang, and R. Rink, “Edge tracking using tactile servo,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 84–89, IEEE, 1995.
- [70] H. Zhang and N. N. Chen, “Control of contact via tactile sensing,” *IEEE Transactions on Robotics and Automation*, vol. 16, no. 5, pp. 482–495, 2000.
- [71] 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.
- [72] Q. Li, C. Schürmann, R. Haschke, and H. J. Ritter, “A control framework for tactile servoing,” in *Robotics: Science and Systems (RSS)*, 2013.
- [73] 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.
- [74] 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.
- [75] 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)*, pp. 10408–10414, IEEE, 2023.
- [76] G. Sutanto, N. Ratliff, B. Sundaralingam, Y. Chebotar, Z. Su, A. Handa, and D. Fox, “Learning latent space dynamics for tactile servoing,” in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3622–3628, IEEE, 2019.
- [77] N. F. Lepora, A. Church, C. De Kerckhove, R. Hadsell, and J. Lloyd, “From pixels to percepts: Highly robust edge perception and contour following using deep learning and an optical biomimetic tactile sensor,” *IEEE Robotics and Automation Letters (RA-L)*, vol. 4, no. 2, pp. 2101–2107, 2019.
- [78] 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.
- [79] P. J. Besl and N. D. McKay, “Method for registration of 3-d shapes,” in *Sensor fusion IV: control paradigms and data structures*, vol. 1611, pp. 586–606, Spie, 1992.
- [80] 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.
- [81] N. Hogan, “Impedance control-an approach to manipulation. i-theory. ii-implementation. iii-applications,” *ASME Journal of Dynamic Systems and Measurement Control B*, vol. 107, pp. 1–24, 1985.
- [82] K. Mo, L. J. Guibas, M. Mukadam, A. Gupta, and S. Tulsiani, “Where2act: From pixels to actions for articulated 3D objects,” in *Proceedings of International Conference on Computer Vision (ICCV)*, 2021.
- [83] 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.
- [84] M. Edmonds, F. Gao, H. Liu, X. Xie, S. Qi, B. Rothrock, Y. Zhu, Y. N. Wu, H. Lu, and S.-C. Zhu, “A tale of two explanations: Enhancing human trust by explaining robot behavior,” *Science Robotics*, vol. 4, no. 37, p. eaay4663, 2019.
- [85] S. Wang, M. Lambeta, P.-W. Chou, and R. Calandra, “Tacto: A fast, flexible, and open-source simulator for high-resolution vision-based tactile sensors,” *IEEE Robotics and Automation Letters (RA-L)*, vol. 7, no. 2, pp. 3930–3937, 2022.



Zihang Zhao received his B.Eng. degree from the School of Mechatronic Engineering at Harbin Institute of Technology in 2020, followed by an M.S. degree from the Department of Mechanical and Aerospace Engineering at the University of California, Los Angeles (UCLA) in 2021. He is currently pursuing his Ph.D. degree, advised by Prof. Yixin Zhu at the Institute for Artificial Intelligence, Peking University. His current research primarily focuses on interactive intelligence.



Yuyang Li is a first-year PhD student at the Institute for AI, Peking University. He is also a research intern at the Beijing Institute for General AI. He received his bachelor's degree in engineering from the Department of Automation at Tsinghua University in 2024. He aspires to advance the development of embodied intelligence, enabling machines to interpret multimodal sensory input, and interact with the environment with human-level adeptness.



Wanlin Li received the B.Eng. and the Ph.D. degree from the School of Electronic Engineering and Computer Science at Queen Mary University of London (QMUL) in 2016 and 2021, respectively. He is currently a research scientist at the Beijing Institute for General Artificial Intelligence (BIGAI). His current research interests include force and tactile sensing.



Zhenghao Qi is a sophomore student from the Department of Automation at Tsinghua University. He is also a student researcher in the Institute for AI at Peking University. His research interests include dexterous manipulation and robot motion control.



Lecheng Ruan received the B.E. honor degree from the School of Mechatronic Engineering, Harbin Institute of Technology in 2015, and the Ph.D. degree from the Department of Mechanical and Aerospace Engineering, University of California, Los Angeles in 2020. He is now directing the Center of Intelligence in the Laboratory of Human-Robot Systems, Peking University. His research interests include perception, control and optimization of robotic systems.



Yixin Zhu is an Assistant Professor and the Assistant Dean at the Institute for Artificial Intelligence, Peking University. He received a Ph.D. degree from UCLA in 2018, advised by Prof. Song-Chun Zhu. His research builds interactive AI by integrating high-level common sense (functionality, affordance, physics, causality, intent) with raw sensory inputs (pixels and haptic signals) to enable richer representation and cognitive reasoning on objects, scenes, shapes, numbers, and agents.



Kaspar Althoefer (M'01-SM'19) is an expert roboticist conducting competitively funded research on soft robotics, intelligent tactile-sensing systems, and interaction dynamics modeling at Queen Mary University of London, with a focus on applications in minimally invasive surgery, assistive technologies, and manufacturing. Currently, he is a Professor of Robotics Engineering and Director of ARQ (Advanced Robotics @ Queen Mary) at Queen Mary University of London, U.K. He co-authored more than 500 refereed research papers in mechatronics

and robotics.