

UniBiDex: A Unified Teleoperation Framework for Robotic Bimanual Dexterous Manipulation

Zhongxuan Li*, Zeliang Guo*, Jun Hu, David Navarro-Alarcon, Jia Pan, Hongmin Wu[†] and Peng Zhou[†]

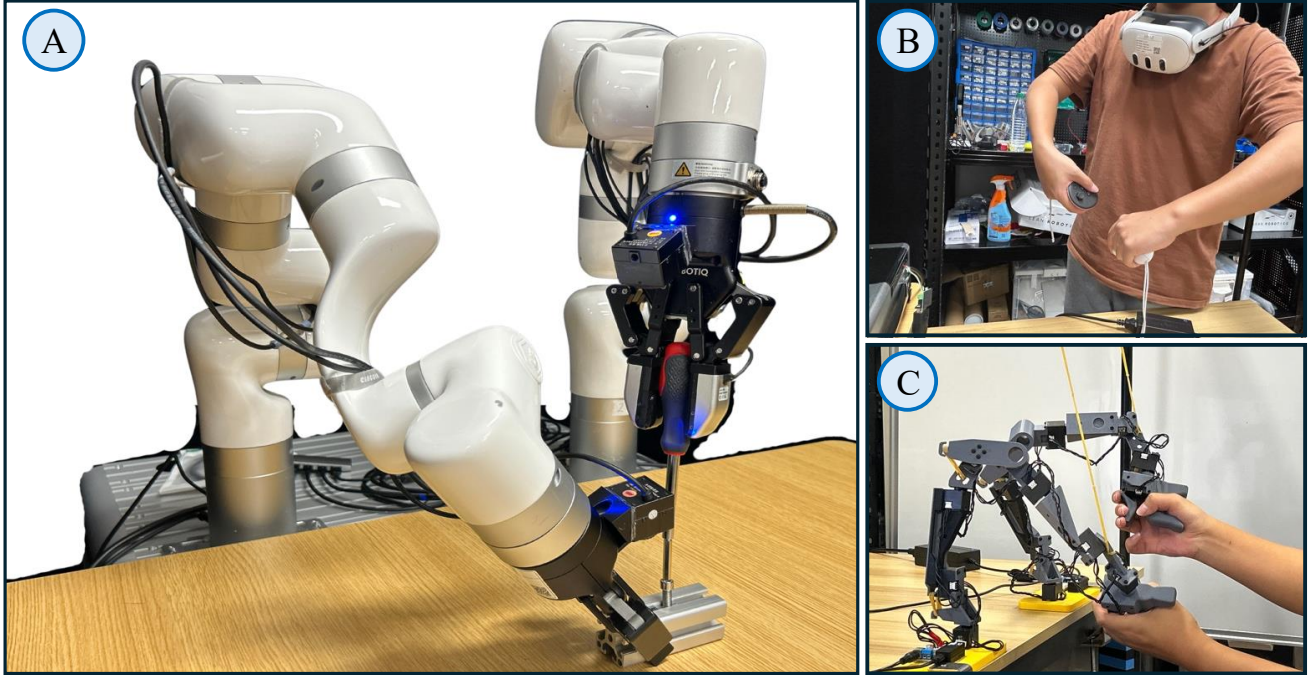


Fig. 1: We show the UniBiDex framework built for contact-rich bimanual teleoperation. The user can teleoperate the robot arms by controlling either the VR headset or leader arms. The system processes heterogeneous input devices with a unified kinematics module.

Abstract—We present UniBiDex, a unified teleoperation framework for robotic bimanual dexterous manipulation that supports both VR-based and leader-follower input modalities. UniBiDex enables real-time, contact-rich dual-arm teleoperation by integrating heterogeneous input devices into a shared control stack with consistent kinematic treatment and safety guarantees. The framework employs null-space control to optimize bimanual configurations, ensuring smooth, collision-free and singularity-aware motion across tasks. We validate UniBiDex on a long-horizon kitchen-tidying task involving five sequential manipulation subtasks, demonstrating higher task

success rates, smoother trajectories, and improved robustness compared to strong baselines. By releasing all hardware and software components as open-source, we aim to lower the barrier to collecting large-scale, high-quality human demonstration datasets and accelerate progress in robot learning.

I. INTRODUCTION

Robot learning methods have transformed robotics by replacing hand-crafted control policies with behaviors learned directly from data across perception, planning, and control [1], [2], [3]. Within this paradigm, imitation learning [4], [5], [6] enables robots to acquire complex manipulation skills from human demonstrations, and its effectiveness increases with richer datasets. Recent studies [7], [8], [9] show that expanding datasets with diverse object types [10], [11], force interactions, and bimanual coordination patterns leads to more accurate task execution and improved generalization [12]. Therefore, developing low-cost, data-efficient teleoperation systems is critical for collecting data to train more advanced imitation learning policies. However, existing teleoperation systems often fall short of capturing the nuance and

* denotes equal contribution and [†] denotes corresponding author.

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant No. 62403211, and in part by National Key Research and Development Program of China (2024YFB4709700), Guangdong Basic and Applied Basic Research Foundation (2025A1515011114, 2024A1515240007). *Corresponding author: Hongmin Wu, Peng Zhou.*

Z. Li and J. Pan are with the Department of Computer Science, The University of Hong Kong, Pokfulam, Hong Kong.

Z. Guo, J. Hu and P. Zhou are with the Dongguan Key Laboratory of Intelligent Equipment and Smart Industry, School of Advanced Engineering, The Great Bay University, Dongguan, China.

D. Navarro-Alarcon is with the Department of Mechanical Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong.

H. Wu is with the Guangdong Academy of Sciences, Guangzhou, China.

dexterity required for contact-rich bimanual manipulation. They frequently lack coordinated dual-arm control and robust safety mechanisms, limiting the precision, speed, and overall quality of the demonstration data. Additionally, existing teleoperation systems often lack force feedback, preventing users from perceiving subtle contact forces experienced by the follower arm. Finally, many teleoperation algorithms are tied to a single modality, which restricts their extensibility across different input devices.

To overcome these limitations, we introduce **UniBiDex**: a Unified Teleoperation Framework for Robotic Bimanual Dexterous Manipulation. Our framework enables precise, coordinated two-arm teleoperation with built-in safety guarantees, facilitating the scalable collection of high-fidelity human demonstrations in contact-rich tasks. The name “UniBiDex” reflects two core principles. First, the system offers universal device support: any input modality—such as VR controllers or leader arms—can be integrated into the same teleoperation framework, enabling user-friendly teleoperation and haptic feedback. Second, the system provides unified constraint handling: all safety and task-related constraints are enforced through a single bimanual control module, regardless of the input modality. This module leverages null-space control to exploit the redundancy of dual 7-DoF arms, enabling the computation of optimal, coordination-aware bimanual configurations. Our main contributions are summarized as follows:

- We present a practical implementation of UniBiDex, featuring modular software, dual-mode hardware integration (VR and leader–follower), and a unified kinematics module for real-time, contact-rich bimanual teleoperation.
- We conduct a comprehensive user study that demonstrates the system’s effectiveness in collecting complex manipulation data compared with other widely used, low-cost teleoperation systems reported in the literature.
- We release all hardware designs and software components as open-source resources to ensure reproducibility and enable the research community to build upon our work.

II. RELATED WORK

A. Teleoperation Devices

Beyond specialized options such as 3D mice or generic 6-DoF motion-tracking devices [13], the two most accessible teleoperation interfaces are VR headsets [14], [15], [16] and leader–follower arms [17], [18], [19], [20], [21]. VR systems are lightweight, portable, and provide an immersive first-person view with multi-DoF tracking—making them well-suited for complex, free-form tasks. However, discrepancies between human and robot kinematics can yield invalid inverse-kinematics solutions, resulting in unintended commands that exceed joint limits, reduce demonstration throughput, and increase failures near singularities or self-collisions. Leader–follower control employs secondary robot arms that are isomorphic to the target manipulator, enabling true one-to-one motion mapping and force feedback.

This matched morphology allows operators to intuitively perceive and respect the robot’s kinematic constraints, enabling precise, low-latency control. Nonetheless, these systems have notable drawbacks, including bulky hardware, restricted workspaces, and substantial integration effort, along with limited flexibility to adapt to new robots or tasks without redesign.

B. Kinematics Constraints in Bimanual Manipulation

Kinematic constraints present a significant challenge in bimanual teleoperation, and recent systems have incorporated solutions directly into control algorithms. Some approaches formulate inverse kinematics (IK) as constrained optimization problems that address singularity and collision avoidance. For example, [14] introduced a VR-based dual-arm system that optimizes singularity and self-collision penalties. [22] proposed a coordination heuristic that mitigates IK conflicts by assigning different roles to each arm. [23] presented a neural-based whole-body controller that integrates VR controller inputs to safely coordinate multiple limbs. However, key limitations remain. First, many systems are overly conservative, enforcing large safety margins that shrink the effective workspace and restrict dexterity. Second, constraint parameters are often robot- or task-specific, requiring manual tuning and limiting adaptability. Third, inter-arm coordination typically relies on simple heuristics rather than unified, real-time optimization, which can fail during tightly synchronized, contact-rich manipulation. Addressing these challenges is essential for enabling robust and adaptable bimanual teleoperation.

III. THE METHODOLOGY

A. Overall System Design

The proposed system supports two teleoperation input modalities: VR headsets and leader–follower arms. For the VR mode, we use the Meta Quest 3 [24], while the leader–follower mode is based on the hardware design of the GELLO system [19]. A unified dual-arm control layer minimizes collisions and inverse-kinematics failures in both modes, enabling precise and reliable bimanual manipulation.

Figure 2 illustrates the overall architecture of the proposed teleoperation framework, which consists of four decoupled modules: *input preprocessing*, *motion retargeting*, *bimanual motion control*, and *haptic feedback*. The *input preprocessing* module receives the target end-effector poses p_{in} and converts them into the robot base frame. If the user operates via a leader arm, it also receives joint angles $q_{0:k}$, where k is the number of joints. This module filters raw input to smooth spikes and suppress abnormal values, ensuring stable downstream control. The *motion retargeting* module translates input poses into joint configurations via inverse kinematics. In VR-based teleoperation, this step is necessary because only end-effector poses are available. For leader–follower arms, although joint angles are directly observed, inverse kinematics is still applied to identify and correct potentially invalid configurations (e.g., self-collisions or singularities). The *bimanual motion control* module computes the final joint

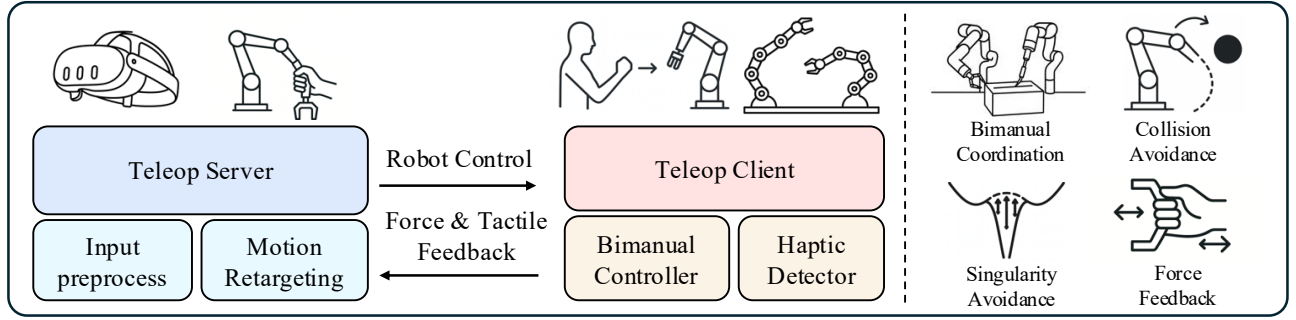


Fig. 2: Illustration of the proposed teleoperation framework. The system processes input commands from heterogeneous devices and formulates task-related and safety constraints for bimanual control as a single optimization problem.

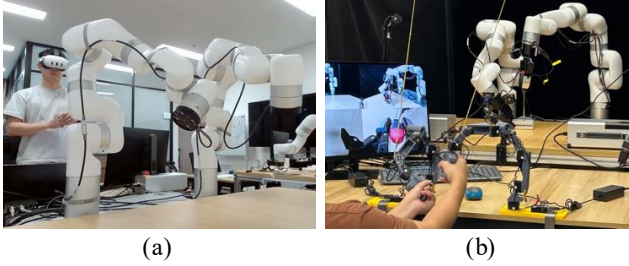


Fig. 3: (a) VR mode. (b) Leader-follower arm mode. UniBiDex supports both VR- and leader arm-based control with nearly identical performance.

commands by optimizing the configurations of both arms. We employ null-space control to exploit the redundancy of dual 7-DoF arms, enabling secondary objectives such as maintaining safe inter-arm distances and maximizing manipulability. Finally, the *haptic feedback* module estimates external contact forces using joint current measurements from the robot arms. After compensating for gravity effects, the processed signals are transmitted back to the teleoperation device, providing intuitive force feedback to the human operator.

B. Input Preprocessing and Motion Retargeting

To integrate heterogeneous input devices into a unified control framework, we define a *virtual base frame* at the user's initial input pose (either the VR-controller origin or the leader-arm base). All subsequent teleoperation commands are expressed relative to this frame and then retargeted to the robot, ensuring consistent behavior across devices and sessions.

1) *Notation:* For each side $i \in \{L, R\}$:

- ${}^V T_{C_i}(t) \in SE(3)$: current control pose in the virtual base frame.
- ${}^V T_{C_i}^0$: initial controller pose at $t = 0$.
- ${}^R T_{E_i}^{\text{des}}(t) \in SE(3)$: desired end-effector pose.
- ${}^R T_{E_i}^0$: robot end-effector pose at $t = 0$.

2) *Relative Motion Retargeting:* We first compute the user's relative controller motion,

$${}^V \Delta T_{C_i}(t) = ({}^V T_{C_i}^0)^{-1} {}^V T_{C_i}(t). \quad (1)$$

The desired robot end-effector poses are then calculated as

$${}^R T_{E_i}^{\text{des}}(t) = {}^V \Delta T_{C_i}(t) {}^R T_{E_i}^0 \quad (2)$$

C. Bimanual Teleoperation Coordination

Before initiating the inverse kinematics computation, we define a set of optimal bimanual configurations, denoted as $\mathcal{Q}_{\text{ref}} = \{(q_L^{(1)}, q_R^{(1)}), \dots, (q_L^{(M)}, q_R^{(M)})\}$. These reference poses can be obtained either through dragging the follower arms or by recording motions from the leader arms. As illustrated in Figure 4b, \mathcal{Q}_{ref} captures preferred dual-arm configurations for tasks such as inserting a seasoning box into a shelf. Notably, the configuration demonstrates superior manipulability in cluttered environments, offering greater reach and collision avoidance capability.

After obtaining the desired end-effector targets for each arm through relative motion retargeting, the system computes the joint commands by solving a safety-aware inverse kinematics (IK) optimization. For each arm i , at every control step we solve:

$$\Delta \mathbf{q}_i = \arg \min_{\Delta \mathbf{q}} \left(\underbrace{\|J_i \Delta \mathbf{q} - \mathbf{e}_i^{\text{cart}}\|^2}_{(1) \text{ Cartesian tracking}} + \underbrace{\omega_q \|\Delta \mathbf{q} - \Delta \mathbf{q}_{C_i}\|^2}_{(2) \text{ Joint matching}} + \underbrace{\mu^2 \|\Delta \mathbf{q}\|^2}_{(3) \text{ Damping}} \right) \quad (3)$$

where $\mathbf{e}_i^{\text{cart}} = \text{Log}({}^R T_{E_i}(t)^{-1} {}^R T_{E_i}^{\text{des}}(t))^\vee$ represents the small incremental end-effector motion at time t , and $\Delta \mathbf{q}_{C_i}$ encodes leader-follower joint increments (set to zero in VR mode).

The optimization is implemented in two stages. First, an unconstrained update is performed by solving the regularized least-squares problem in Eq. (3) without considering collisions, yielding a candidate increment $\Delta \mathbf{q}_i$ for each arm $i \in \{L, R\}$. To further improve bimanual consistency, a null-space control term is incorporated to attract each arm toward a set of predefined optimal bimanual poses \mathcal{Q}_{ref} .

At each control step, we select the closest full configuration pair $(q_L^*, q_R^*) \in \mathcal{Q}_{\text{ref}}$ to the current robot state (q_L, q_R) via:

$$(q_L^*, q_R^*) = \arg \min_{(q_L^{(k)}, q_R^{(k)}) \in \mathcal{Q}_{\text{ref}}} \|q_L^{(k)} - q_L\|^2 + \|q_R^{(k)} - q_R\|^2.$$

For each arm $i \in \{L, R\}$, we then compute the optimal null-space increment:

$$\Delta q_{i,\text{null}}^{\text{opt}} = \arg \min_{\Delta q_{i,\text{null}}} \|q_i + \Delta q_{i,\text{null}} - q_i^*\|$$

where the null-space increment is represented by

$$\Delta q_{i,\text{null}} = k_n (I - J_i^\dagger J_i) \Delta q_i^{\text{task}}$$

Here $J_i^\dagger = (J_i^\top J_i)^{-1} J_i^\top$ is the pseudoinverse of the arm Jacobian and $k_n > 0$ is a scalar gain. Finally, we augment the unconstrained solution:

$$\Delta q_i \leftarrow \Delta q_i + \Delta q_{i,\text{null}}^{\text{opt}}$$

By doing the above steps, each arm is subtly guided toward its corresponding reference pose, improving configuration consistency and reducing anomalous IK behavior without disrupting primary end-effector tracking. This unified procedure ensures accurate, real-time, and collision-aware bimanual teleoperation across all input modalities.

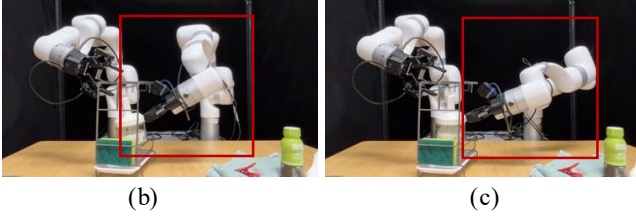


Fig. 4: Illustration of the optimal reference configuration. The right arm joint configuration in (c) provides greater degrees of freedom compared to the arm in (b), resulting in more compliant control in subsequent motions.

In practice, \mathcal{Q}_{ref} is obtained in a one-time calibration lasting 5–10 minutes per manipulation task. An expert operator uses the leader arm to teleoperate the robot and records 8–12 canonical bimanual configurations (e.g., reaching, approaching, handover) as joint-angle pairs on the followers. The resulting set is stored as a configuration file. In our experiments, we consistently used 10 reference poses without per-subtask tuning.

We did not enforce full-body collision detection across all links of the dual-arm robot, as doing so was found to hinder precise control in contact-rich manipulation tasks. Instead, we observed that null-space control naturally guided the system toward collision-free configurations by exploiting kinematic redundancy. To maintain safety without compromising dexterity, we implemented a lightweight watchdog thread that monitors abrupt joint movements.

Our framework ensures accurate end-effector tracking and compliant motion across all input modalities. After computing the optimal joint increments, a PD controller is

applied to generate joint velocity commands, which are then transmitted to the follower arm.

D. Haptic Feedback

In the teleoperation loop, both kinesthetic and vibrotactile feedback are generated for the operator using joint current measurements from the follower arm. At each control step, motor currents from all joints are recorded. These raw signals include contributions from gravity compensation as well as external interaction forces. By subtracting the precomputed gravity torques—derived from the robot’s mass and geometry—the system isolates the net interaction torque caused by contact or payload.

This torque is rendered directly on the leader arm, allowing the operator to intuitively perceive the forces experienced by the follower arm and providing kinesthetic feedback. In parallel, the magnitude of the interaction torque is mapped to a vibration signal on the VR controllers. This vibrotactile cue complements the kinesthetic channel by highlighting subtle contact events or dynamic interactions that might otherwise go unnoticed.

We note that while recent work such as [20] provides force feedback to the leader arm, it relies on high-end robotic platforms equipped with joint torque sensors (e.g., Franka [25]). In contrast, our approach infers external interaction forces directly from joint motor currents, enabling a more general and cost-effective solution applicable to low-cost robotic arms without built-in force sensing. This makes our system broadly accessible while still supporting rich haptic feedback.

Together, these two feedback channels create a rich, multi-modal haptic interface that enhances operator awareness and improves manipulation performance in contact-rich tasks.

IV. EXPERIMENTS

A. User Study Procedure

We recruited three male and one female participants, all with prior teleoperation experience. Each participant was asked to complete a single, long-horizon bimanual manipulation task under two input modalities: VR-based and leader–follower. The task simulates a household kitchen-tidying routine and involves sequential, contact-rich subtasks requiring dexterous coordination:

- 1) *Item Unpacking*: Pick up a group of assorted items (seasoning box, drink bottle, cleaning sponge, towel, and clamp) from a cloth bag and place them on the table.
- 2) *Shelf Organization*: Sort and place the items into their designated positions in a multi-level kitchen shelf.
- 3) *Towel Folding*: Fold the towel neatly along a marked midline using both grippers.
- 4) *Towel Placement*: Place the folded towel onto the rack section of the shelf.
- 5) *Clamp Attachment*: Attach the clamp onto the rack such that it secures the towel in place.

Before starting the task, participants received a five-minute training session per modality to familiarize themselves with



Fig. 5: Overview of the leader–follower arm teleoperation workflow for the household kitchen–tidying task, showing the five sequential subtasks: (1) item unpacking, (2) shelf organization, (3) towel folding, (4) towel placement, and (5) clamp attachment.

the control interface and virtual base frame calibration. Rest breaks were allowed between trials to prevent fatigue. All experiments were conducted on a dual-arm XArm-7 robot equipped with parallel-jaw grippers. In VR mode, participants used a Meta Quest 3 headset. We reimplemented the teleoperation logic of Bunny Vision Pro [14] and GELLO [19] on our hardware to enable a fair and meaningful comparison. For clarity, we define the baseline implementations as follows. *Naive VR*: a direct position-based IK solver mapping the VR controller pose to the robot, without null-space coupling or attraction toward Q_{ref} . *Naive LF*: a direct 1:1 joint mapping from leader to follower arms with light smoothing, again without null-space coordination or Q_{ref} . In contrast, *UniBiDex* employs a safety-aware IK solver with null-space attraction toward Q_{ref} , which regularizes the arms

into consistent, safe bimanual postures. System performance was evaluated in terms of success rate and task completion time among all successful trials.

B. Teleoperation Results

We collected 40 trials per modality (1 long-horizon task \times 4 users \times 2 modes \times 5 repetitions), for a total of 80 trials. Table I reports the mean and standard deviation of completion time and success rate for each subtask.

Overall, our unified framework delivered markedly more consistent performance across the entire manipulation sequence compared to both VR-only and leader–follower baselines. Specifically, baseline VR control suffered from erratic end-effector trajectories when recovering from large Cartesian errors, and occasionally stalled near kinematic singularities. In cluttered scenes—most notably during the towel

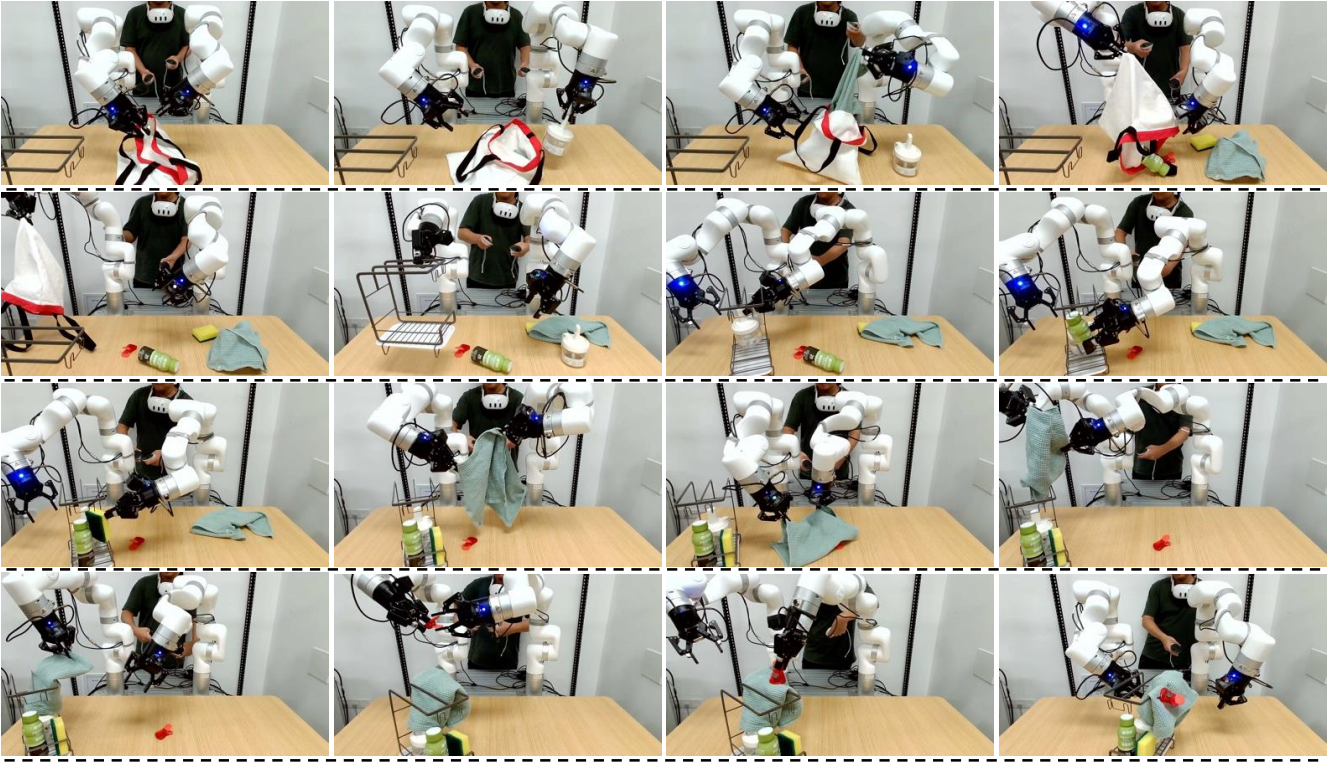


Fig. 6: Overview of the VR-based teleoperation workflow for the household kitchen-tidying task.

TABLE I: Completion time (s) and success rate across five subtasks and overall task performance

Method	Item Unpacking		Shelf Organization		Towel Folding		Towel Placement		Clamp Attachment		Overall Task	
	Time (s)	Succ	Time (s)	Succ	Time (s)	Succ	Time (s)	Succ	Time (s)	Succ	Time (s)	Succ
<i>VR-based Methods</i>												
<i>UniBiDex (VR)</i>	238 ± 12	31/40	198 ± 10	30/40	118 ± 6	32/40	80 ± 4	24/40	38 ± 4	24/40	672 ± 20	24/40
<i>Naive VR</i>	284 ± 16	28/40	236 ± 12	26/40	142 ± 10	25/40	96 ± 6	22/40	58 ± 4	22/40	816 ± 24	18/40
<i>Leader-Follower Methods</i>												
<i>UniBiDex (LF)</i>	113 ± 5	37/40	94 ± 4	36/40	56 ± 3	36/40	38 ± 2	30/40	18 ± 2	30/40	319 ± 8	30/40
<i>Naive LF</i>	117 ± 6	33/40	97 ± 5	32/40	57 ± 4	31/40	42 ± 3	25/40	22 ± 2	25/40	335 ± 9	24/40

placement and clamp attachment subtasks—these instabilities led to frequent trajectory overshoot, oscillations, and safety-margin violations. The leader–follower baseline, while more stable in free space, also suffered from reachability issues due to suboptimal joint configurations. These limitations contributed to frequent task failures under challenging spatial constraints.

In contrast, our method maintained smooth and reliable performance across all subtasks. The integration of null-space guided configuration optimization helped mitigate abrupt joint updates and enabled more stable arm behavior, particularly during fine manipulation stages. Task success rates also improved: overall task completion rose to 60% (24/40) in VR mode and 75% (30/40) in leader–follower mode, compared to 45% and 57% for the respective baselines.

To facilitate qualitative review, we have archived all experiment recordings in the supplementary materials, illustrating the smooth, coordinated behavior achieved by our framework

under both teleoperation modalities.

C. Failure Case Analysis and Discussion

We analyzed task failures across all five subtasks to identify the limitations of our system. Errors during *item unpacking* primarily occurred when objects within the cloth bag were occluded or improperly oriented, causing incorrect grasping. In the *shelf organization* subtask, our null-space-augmented inverse kinematics (IK) method reliably positioned items into their designated slots. The infrequent failures that did occur typically resulted from transient tracking drift or calibration inaccuracies. For *towel folding*, the main challenge emerged when lifting the folded towel off the table, as the layers tended to separate and slip, leading to partial unfolding. This instability directly affected the subsequent *towel placement* step, as misaligned or unfolded edges frequently missed the rack bar, necessitating additional time for regripping and realignment. The *clamp attachment* subtask, despite demanding high precision, demonstrated ro-

bust performance due to the teleoperation system’s accuracy, highlighting the effectiveness of our control approach.

Overall, these results underscore the importance of robust IK solutions and coordinated control frameworks for precision manipulation tasks, while also pointing toward the necessity for enhanced strategies in handling deformable objects. Future efforts will focus on incorporating tactile feedback and adaptive grasp techniques to preserve structural integrity during manipulation, thereby improving real-world task success rates.

V. CONCLUSIONS

In this paper, we presented UniBiDex, a unified teleoperation framework that supports both VR and leader–follower inputs through a shared kinematic and safety-aware control module. Our system enables precise, real-time bimanual manipulation across diverse devices and tasks. User studies show that UniBiDex outperforms strong baselines in task success, motion smoothness, and robustness, particularly in contact-rich scenarios. We hope UniBiDex will lower the barrier to collecting large-scale, high-quality demonstration datasets and thereby accelerate progress in robot learning.

REFERENCES

- [1] B. Zitkovich, T. Yu, S. Xu, P. Xu, T. Xiao, F. Xia, J. Wu, P. Wohlhart, S. Welker, A. Wahid, *et al.*, “Rt-2: Vision-language-action models transfer web knowledge to robotic control,” in *Conference on Robot Learning*, pp. 2165–2183, PMLR, 2023.
- [2] K. Black, N. Brown, D. Driess, and *et al.*, “ π_0 : A vision–language–action flow model for general robot control,” 2024. [arXiv:2410.24164](https://arxiv.org/abs/2410.24164) [cs.LG].
- [3] Z. Weng, P. Zhou, H. Yin, A. Kravberg, A. Varava, D. Navarro-Alarcon, and D. Kragic, “Interactive perception for deformable object manipulation,” *IEEE Robotics and Automation Letters*, 2024.
- [4] S. Schaal, “Is imitation learning the route to humanoid robots?,” *Trends in cognitive sciences*, vol. 3, no. 6, pp. 233–242, 1999.
- [5] 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 *2018 IEEE international conference on robotics and automation (ICRA)*, pp. 5628–5635, IEEE, 2018.
- [6] P. Zhou, J. Qi, A. Duan, S. Huo, Z. Wu, and D. Navarro-Alarcon, “Imitating tool-based garment folding from a single visual observation using hand-object graph dynamics,” *IEEE Transactions on Industrial Informatics*, vol. 20, no. 4, pp. 6245–6256, 2024.
- [7] A. Mandlekar, D. Xu, J. Wong, S. Nasiriany, C. Wang, R. Kulkarni, L. Fei-Fei, S. Savarese, Y. Zhu, and R. Martín-Martín, “What matters in learning from offline human demonstrations for robot manipulation,” *arXiv preprint arXiv:2108.03298*, 2021.
- [8] F. Lin, Y. Hu, P. Sheng, C. Wen, J. You, and Y. Gao, “Data scaling laws in imitation learning for robotic manipulation,” *arXiv preprint arXiv:2410.18647*, 2024.
- [9] C. Chi, Z. Xu, C. Pan, E. Cousineau, B. Burchfiel, S. Feng, R. Tedrake, and S. Song, “Universal manipulation interface: In-the-wild robot teaching without in-the-wild robots,” *arXiv preprint arXiv:2402.10329*, 2024.
- [10] J. Qi, L. Lu, L. Yang, Y. Ding, P. Zheng, D. Navarro-Alarcon, J. Pan, and P. Zhou, “A coarse-to-fine robotic fabric alignment system integrating visual servoing and admittance control,” *IEEE Transactions on Automation Science and Engineering*, 2025.
- [11] Z. Zhang, G. Chen, W. Chen, R. Jia, G. Chen, L. Zhang, J. Pan, and P. Zhou, “A joint learning of force feedback of robotic manipulation and textual cues for granular materials classification,” *IEEE Robotics and Automation Letters*, 2025.
- [12] P. Zhou, P. Zheng, J. Qi, C. Li, H.-Y. Lee, Y. Pan, C. Yang, D. Navarro-Alarcon, and J. Pan, “Bimanual deformable bag manipulation using a structure-of-interest based neural dynamics model,” *IEEE/ASME transactions on mechatronics*, 2024.
- [13] K. Darvish, L. Penco, J. Ramos, R. Cisneros, J. Pratt, E. Yoshida, S. Ivaldi, and D. Pucci, “Teleoperation of humanoid robots: A survey,” *IEEE Transactions on Robotics*, vol. 39, no. 3, pp. 1706–1727, 2023.
- [14] R. Ding, Y. Qin, J. Zhu, C. Jia, S. Yang, R. Yang, X. Qi, and X. Wang, “Bunny-visionpro: Real-time bimanual dexterous teleoperation for imitation learning,” 2024.
- [15] A. Iyer, Z. Peng, Y. Dai, I. Guzey, S. Haldar, S. Chintala, and L. Pinto, “Open teach: A versatile teleoperation system for robotic manipulation,” 2024.
- [16] C. Wang, H. Shi, W. Wang, R. Zhang, L. Fei-Fei, and C. K. Liu, “Dexcap: Scalable and portable mocap data collection system for dexterous manipulation,” *arXiv preprint arXiv:2403.07788*, 2024.
- [17] T. Z. Zhao, V. Kumar, S. Levine, and C. Finn, “Learning fine-grained bimanual manipulation with low-cost hardware,” *arXiv preprint arXiv:2304.13705*, 2023.
- [18] Z. Fu, T. Z. Zhao, and C. Finn, “Mobile aloha: Learning bimanual mobile manipulation with low-cost whole-body teleoperation,” *arXiv preprint arXiv:2401.02117*, 2024.
- [19] P. Wu, Y. Shentu, Z. Yi, X. Lin, and P. Abbeel, “Gello: A general, low-cost, and intuitive teleoperation framework for robot manipulators,” in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 12156–12163, IEEE, 2024.
- [20] J. J. Liu, Y. Li, K. Shaw, T. Tao, R. Salakhutdinov, and D. Pathak, “Factr: Force-attending curriculum training for contact-rich policy learning,” *arXiv preprint arXiv:2502.17432*, 2025.
- [21] P. Zhou, J. Zhu, S. Huo, and D. Navarro-Alarcon, “Lasesom: A latent and semantic representation framework for soft object manipulation,” *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 5381–5388, 2021.
- [22] J. Grannen, Y. Wu, B. Vu, and D. Sadigh, “Stabilize to act: Learning to coordinate for bimanual manipulation,” in *Conference on Robot Learning*, pp. 563–576, PMLR, 2023.
- [23] T. He, Z. Luo, X. He, W. Xiao, C. Zhang, W. Zhang, K. Kitani, C. Liu, and G. Shi, “Omnih2o: Universal and dexterous human-to-humanoid whole-body teleoperation and learning,” *arXiv preprint arXiv:2406.08858*, 2024.
- [24] Reality Labs (Meta), “Meta quest 3,” <https://www.meta.com/quest/quest-3/>, 2023. Standalone mixed reality headset; 2064×2208 px per eye, Snapdragon XR2 Gen 2, released October 10, 2023.
- [25] Franka Emika GmbH, “Franka Emika Panda Robot Arm,” <https://www.franka.de>, 2018. Accessed: 2025-07-29.