Exo-ViHa: A Cross-Platform Exoskeleton System with Visual and Haptic Feedback for Efficient Dexterous Skill Learning

Xintao Chao[*1], Shilong Mu[*†1], Yushan Liu[1], Shoujie Li[1], Chuqiao Lyu[†1],
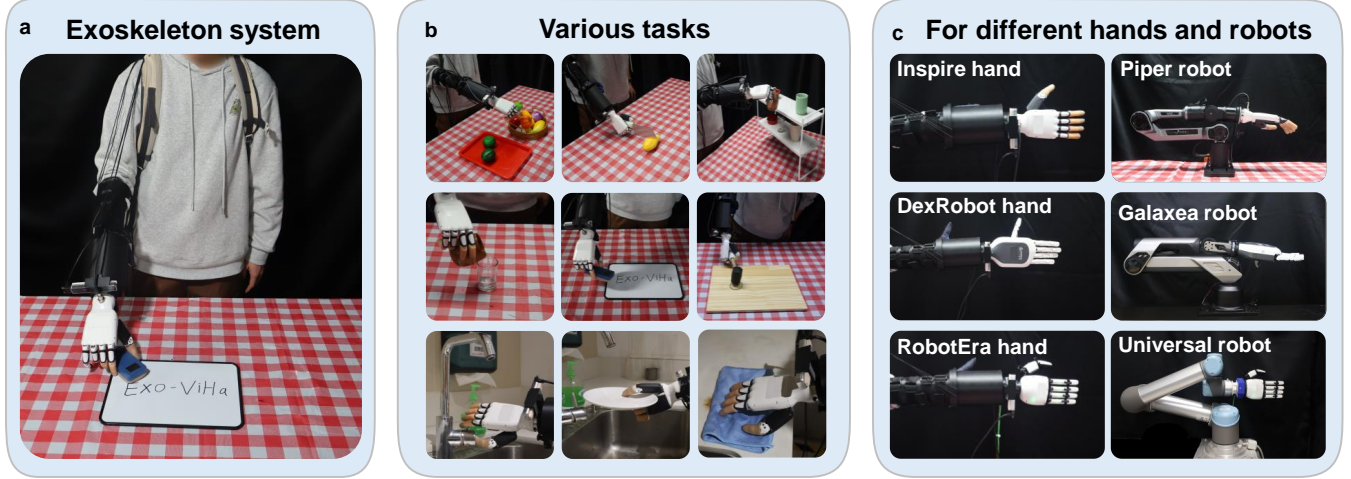Wenbo Ding[†1], Xiao-Ping Zhang[1], *Fellow, IEEE*

Fig. 1: **Overview of the Exo-ViHa system.** (a) Demonstration of the exoskeleton-based data collection system. (b) Data collection of diverse tasks in various scenarios. (c) Compatibility of the Exo-ViHa system with different dexterous robotic hands and arms.

Abstract—**Imitation learning has emerged as a powerful paradigm for robot skills learning. However, traditional data collection systems for dexterous manipulation face challenges, including a lack of balance between acquisition efficiency, consistency, and accuracy. To address these issues, we introduce Exo-ViHa, an innovative 3D-printed exoskeleton system that enables users to collect data from a first-person perspective while providing real-time haptic feedback. This system combines a 3D-printed modular structure with a slam camera, a motion capture glove, and a wrist-mounted camera. Various dexterous hands can be installed at the end, enabling it to simultaneously collect the posture of the end effector, hand movements, and visual data. By leveraging the first-person perspective and direct interaction, the exoskeleton enhances the task realism and haptic feedback, improving the consistency between demonstrations and actual robot deployments. In addition, it has cross-platform compatibility with various robotic arms and dexterous hands. Experiments show that the system can significantly improve the success rate and efficiency of data collection for dexterous manipulation tasks. Webpage: https://exo-viha2025.github.io/.**

I. INTRODUCTION

Dexterous manipulation remains a key challenge in the field of robotics, particularly for tasks requiring fine control and high degree-of-freedom coordination, such as pen spinning [1], ball throwing [2], and folding clothes [2]. To enable robots to learn these complex tasks, imitation learning provides an effective solution by allowing robots to learn various skills from human demonstration [3], [4]. However, the effectiveness of imitation learning largely depends on high-quality, large-scale demonstration data, which imposes higher requirements on data collection systems.

Traditional data collection methods often rely on teleoperation systems, such as virtual reality (VR) devices [5], [6], [7], handheld controllers [8], [9], or vision-based teleoperation [10], [11], [12]. These approaches improve the flexibility of teleoperate to some extent, allowing humans to intuitively control robots and collect corresponding manipulation data. However, they typically suffer from limitations such as the lack of low-latency first-person visual feedback, the absence of haptic feedback, and inconsistencies between the data collection and deployment phases. These limitations may lead to inefficient data collection or degraded model performance during real-world deployment, thus reducing the generalizability of imitation learning.

To address these constraints, we introduce a novel exoskeleton device designed to efficiently collect data for dexterous hand imitation learning tasks. The proposed exoskeleton is a 3D-printed component designed to accommodate various dexterous hands. During data collection, the user first mounts a Intel RealSense tracking camera T265 and a wrist-mounted camera onto the exoskeleton's camera base. Then, the user wears a motion capture glove and attaches the exoskeleton to the forearm to initiate data collection. The

---

TABLE I: Comparison of Data Collection Systems for Dexterous Manipulation

System	End-Effector	Visual	Haptic	VC	BC	Data Type	Cross-Platform	Cost
DexCap [14]	Dexterous Hand	FP	★★★★★	★★	★★	Off-line	✗	Low
UMI [15]	Gripper	FP	★★★	★★★★★	★★★★	Off-line	✗	Low
ForceMimic [16]	Gripper	FP	★★★	★★★★★	★★★★	Off-line	✗	Medium
Fast-UMI [17]	Gripper	FP	★★★	★★★★★	★★★★	Off-line	✗	Low
Mobile ALOHA [18]	Gripper	TP	★★	★★★★★	★★★★★	On-line	✗	High
Feelit [19]	Gripper	TP	★★★★	★★★★★	★★★★★	On-line	✗	High
AnyTeleop [11]	Dexterous Hand	VR	★	★★★★★	★★★★★	On-line	✓	High
Open-TeleVision [7]	Dexterous Hand	VR	★	★★★★★	★★★★★	On-line	✓	High
Bunny-VisionPro [20]	Dexterous Hand	VR	★★★	★★★★★	★★★★★	On-line	✓	High
ACE [21]	Hand/Gripper	TP	★	★★★★★	★★★★★	On-line	✓	High
HIRO [22]	Dexterous Hand	FP	★★★★	★★	★★	On-line	✗	Low
Our System	**Dexterous Hand**	**FP**	★★★	★★★★★	★★★★	**Off-line**	✓	**Low**

FP: First Person, VR: Virtual Reality, TP: Third Person, VC: Visual Consistency (rated ★ to ★★★★★),
BC: Body Consistency (rated ★ to ★★★). Higher star count indicates better performance.

system collects multiple types of data, including end-effector pose data from the T265 camera, motion data from the dexterous hand, and frame data from the cameras. We employ the Action Chunking Transformer (ACT) [13] to train the imitation learning algorithm. The output end-effector pose and dexterous hand motion data from the model are then used to control the robotic arm and dexterous hand for movement and manipulation.

To properly evaluate our contributions, we evaluate our system against existing data collection solutions, considering several critical dimensions as illustrated in Table I.

The advantages of our system are highlighted through comparisons with existing data collection solutions:

- **First-Person visual feedback**: Our exoskeleton offers true first-person visual feedback. This direct engagement allows users to operate in their natural field of view, enhancing task realism and operational accuracy.
- **Haptic feedback**: Since the exoskeleton makes direct contact with the user's arm, users can distinctly perceive tactile sensations and force feedback while operating the objects, which is crucial for tasks that require force adjustment during the interaction with multiple objects.
- **Visual consistency**: By placing identical camera bases on both the exoskeleton and the robotic arm, cameras can collect sufficient visual information while ensuring a consistent visual perspective during both data collection and deployment phases.
- **Body consistency**: Our system directly collects data from the dexterous hand when performing tasks. This contrasts with other platforms where data might be indirectly mapped or transformed, ensuring that our deployment results are more reliable and closely aligned with the demonstration conditions.

II. RELATED WORK

A. Imitation Learning

Imitation Learning (IL) has gained significant traction in recent years, primarily due to its ability to mimic expert behavior in complex environments, providing a pathway to efficient learning without exhaustive trial-and-error exploration. This paradigm demonstrates versatility across various domains, including robotics, autonomous driving, and game playing, showcasing its ability to generalize expert knowledge effectively [23], [24]. The IL method can have input from different sources of demonstrations, such as human demonstrations [25], [26], real-Robot expert data [13], [18], [27], and simulation date [28]–[30], thus enhancing its practicality in real production. Several algorithms have been developed to improve the performance of IL systems. Among them, the Differential Privacy (DP) [3] algorithm and the ACT algorithm stand out, which has proven effective in tasks that require high temporal precision and accurate control.

B. Data Collection Systems

In recent years, several data collection systems have been developed for robotic applications. One common method is teleoperation [3], [11], [31], where robots are controlled manually from a distance. Another approach involves the use of haptic devices, which provide tactile feedback, allowing the operator to feel the robot's interactions with the environment. Virtual reality systems also offer an immersive experience for data collection [5], [6], [7], [20], enabling users to control robots in a simulated environment.

For imitation learning, it is crucial to ensure that input-output data remain consistent to achieve effective learning. High-quality data collection should prioritize precision, which often requires tactile feedback and first-person perspectives to accurately capture the robot's interactions. However, many existing systems lack these features, which limits the effectiveness of the collected data for training sophisticated models.

III. EXO-VIHA SYSTEM ARCHITECTURE

A. Hardware Design

The design of the exoskeleton system is illustrated in Fig. 2. The entire exoskeleton is constructed using 3D printing technology with Polylactic Acid (PLA) material, ensuring a balance of durability and lightweight structure.
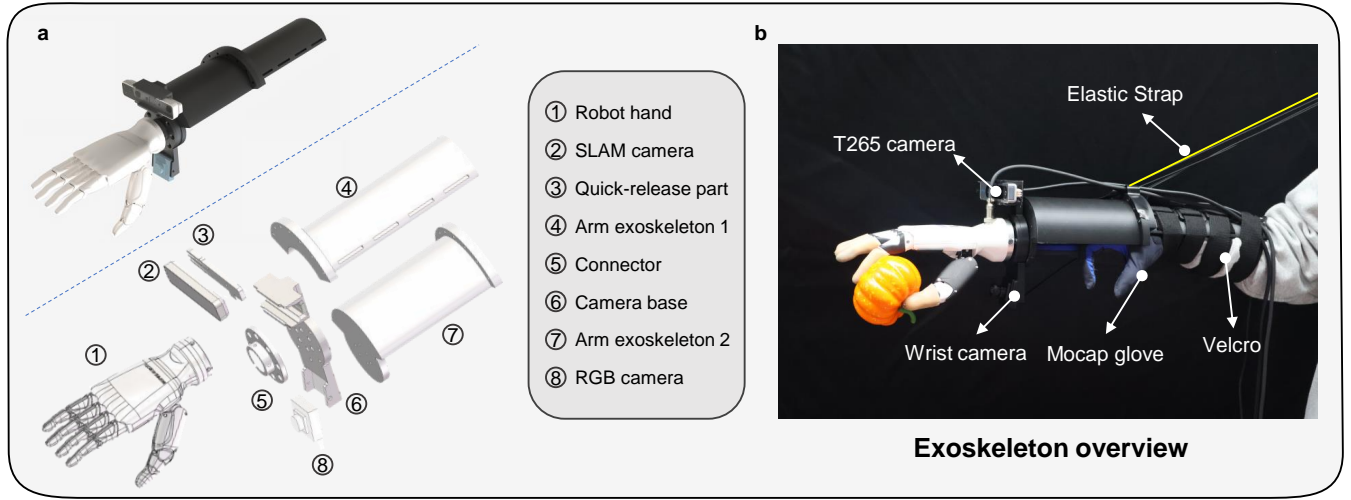
Fig. 2: **Details of the exoskeleton system.** (a) The components of the exoskeleton system. (b) Exoskeleton wearing demonstration.

The overall weight of the system is about 425 grams. Upon first use, users can assemble all components within a few minutes. The precise 3D printing process ensures that all components fit seamlessly and securely, providing both functionality and comfort for the user.

The forearm exoskeleton 1 is secured to the user's hand using velcro straps, while forearm exoskeleton 2 provides cover and finger movement space for the user's hand during data collection, ensuring that the external cameras can only view the dexterous hand and not the user's hand. Additionally, the camera base component includes slots for the T265 camera and the wrist camera below. To facilitate quick installation and removal of the cameras, the slots feature a rapid insertion clip design, allowing the T265 camera to be easily attached and detached using a compatible quick-connect adapter.

Moreover, the system is designed to accommodate a wide range of dexterous hands by 3D printing specific connectors tailored for each particular hand or two-finger gripper. The entire exoskeleton system weighs approximately 425 grams; however, when a dexterous hand is attached, the total weight increases to around 1100 grams. To address this, elastic bands are utilized to provide a degree of gravitational compensation, thereby reducing the burden on the user's arm.

B. System Design

The operational demonstration of the entire Exo-ViHa system is illustrated in Fig. 3. During the data collection phase, in addition to the SLAM camera and wrist camera mounted on the exoskeleton, we also employed two external cameras positioned at different angles to comprehensively capture environmental information. As shown in Fig. 3 (a), during data collection, experimenters can access first-person visual information, allowing them to observe the current workspace through their own eyes and control the position and posture of the dexterous hand, facilitating a more direct acquisition of spatial information.

In addition, when performing tasks with the exoskeleton and dexterous hand, we experience a noticeable resistance through the dexterous hand to our forearm when it makes contact with objects, even if there are no haptic sensors to capture specific contact pressure. This force feedback enables us to perceive the interactions and friction between objects effectively during multi-contact tasks, allowing for better control of forearm positioning and the use of the dexterous hand, thereby enhancing the quality of the data demonstration.

While wearing the entire exoskeleton, our arms maintain a significant range of motion despite the presence of elastic bands used for gravitational compensation. As shown in Fig. 3 (b), wearers can easily rotate their arms 180° from side to side. The vertical motion is limited to approximately 120° primarily due to the length of the connecting cables. However, this range of motion is effectively unrestricted for performing basic tasks.

Another advantage of the system lies in the consistency of perspective. During data collection, the arm wearing the exoskeleton is largely obscured, making it difficult to view it directly. Consequently, external cameras can capture information about the positioning and tasks of the dexterous hand more effectively. Additionally, since our camera base is mounted on both the exoskeleton and the robotic arm, as illustrated in Fig. 3 (c), the information obtained by the wrist camera is nearly identical as long as the end-effector's pose remains consistent during both collection and deployment. This alignment allows the model's outputs to closely fit the collected data, thereby reducing visual discrepancies during actual deployment and improving the success rate of task execution.

Moreover, due to the lightweight and agile design of the exoskeleton, we can conveniently place the power source and a compact computer in a backpack. As illustrated in Fig. 3 (d), the exoskeleton system is worn for data collection in various scenarios. Considering the necessity for a third-
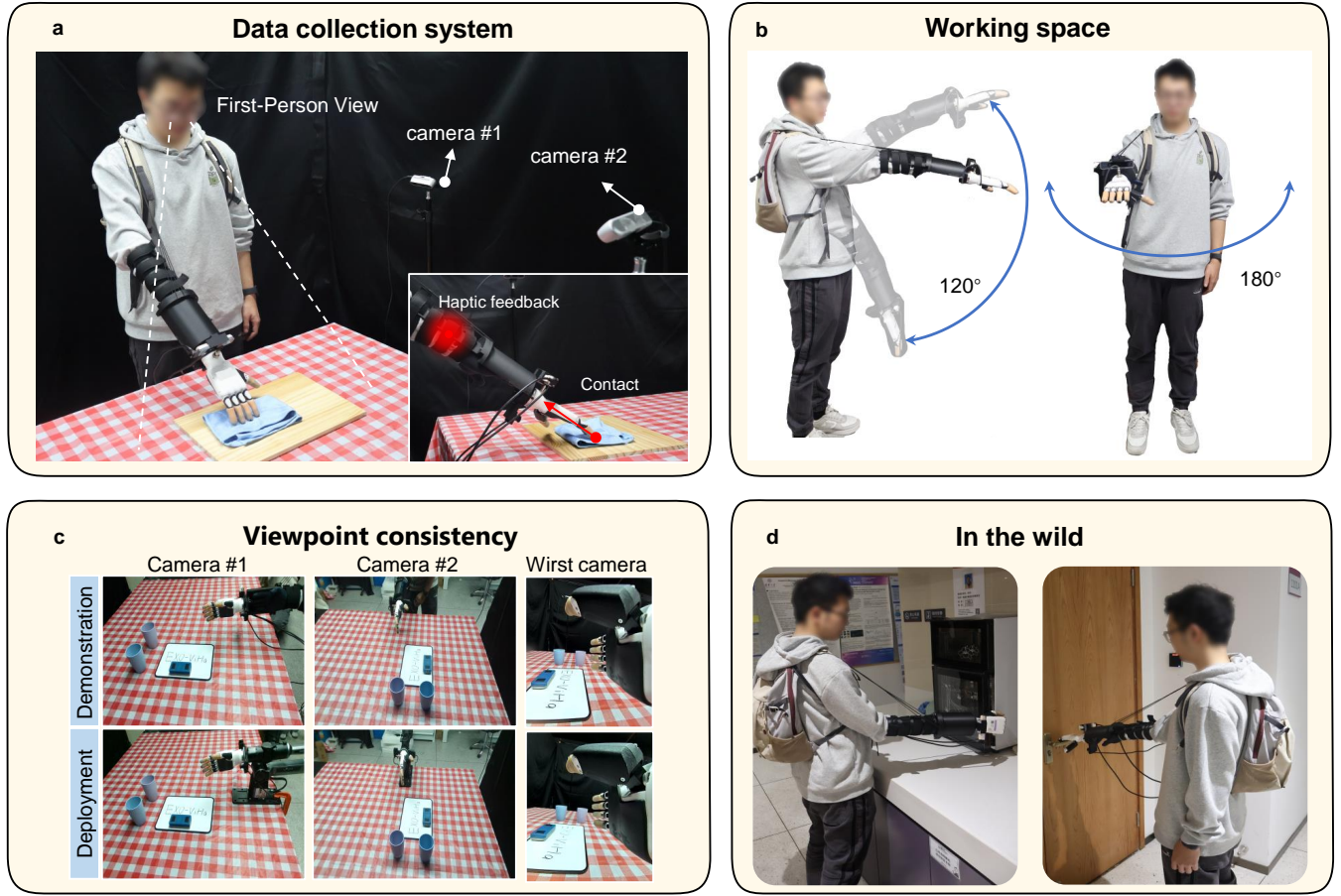
Fig. 3: **Details of the Exo-ViHa system.** (a) Demonstration of the exoskeleton used for data collection and schematic diagram of the first-person perspective and haptic feedback. (b) Illustration of the working space range. (c) The visual information captured by the external cameras and the wrist camera is basically consistent during both the data collection and deployment phases. (d) The power source and microcomputer are placed into a backpack, allowing the system to be used for data collection in outdoor environments.
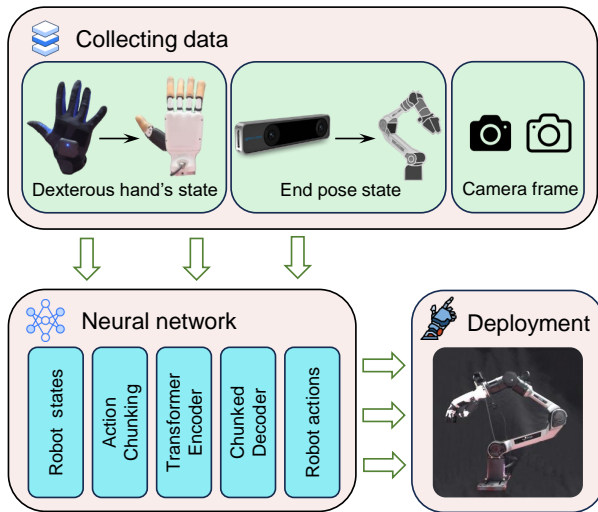

Fig. 4: **Workflow of data collection and deployment of Exo-ViHa.** The collected data include end-effector states, 6D end-effector pose and camera frames.
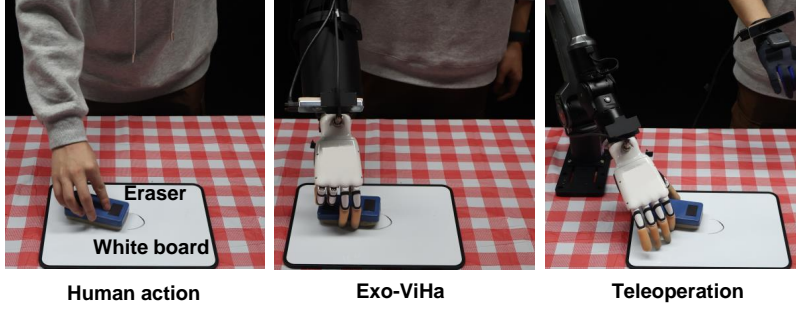
person perspective camera, we can either place the camera at a fixed position in the external environment or mount the external camera on the chest to record the pose during data collection(similar to Dexcap [14]).
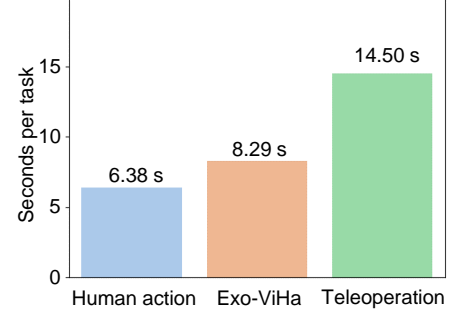
C. Learning Framework

Fig. 4 illustrates the overall learning framework of the proposed system.

During data collection using the exoskeleton, the user wears a motion capture glove to obtain real-time posture information of the hand beneath the arm exoskeleton 2. The data from the motion capture glove is then mapped onto the current dexterous hand, enabling it to replicate the user's hand posture. Simultaneously, real-time data from the dexterous hand is recorded to serve as input for the neural network model. After training, the model's output directly controls the dexterous hand, ensuring alignment between the input and output data formats. Although there may be some discrepancies in posture fitting between the motion capture glove and the dexterous hand, as long as the experimenter can control the dexterous hand to perform the corresponding
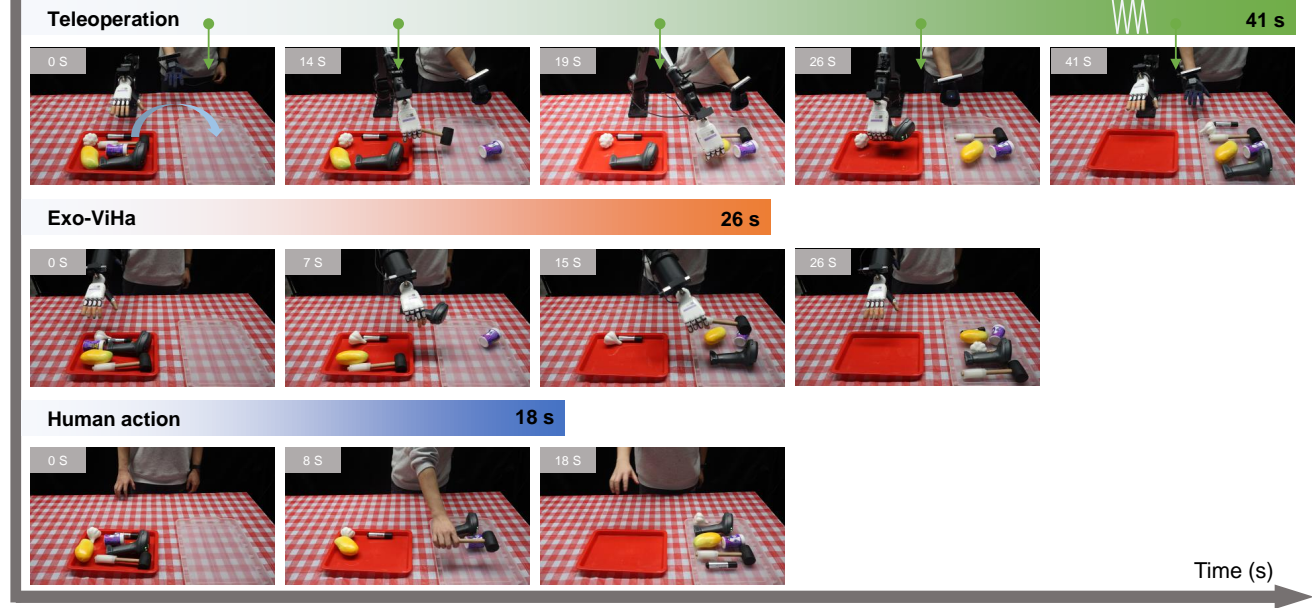
Fig. 5: **Data collection efficiency comparision across different scenarios.** (a-b) Comparative analysis of whiteboard erasing tasks: (a) Multi-contact interaction scenarios demonstrating human action, Exo-ViHa and teleoperation; (b) Quantitative time efficiency comparison across methods. (c) Pick and place scenario demonstrating the execution process for teleoperation, Exo-ViHa, and human action.

tasks during data collection, the model's output will be sufficient to direct the dexterous hand in completing those tasks.

Meanwhile, we employ an Intel T265 SLAM camera to record the end-effector pose data of the exoskeleton, while calibration and measurement procedures are utilized to fit this data to the actual end-effector pose of the robotic arm:

$$\mathbf{T}\text{robot} = \mathbf{T}\text{calib} \cdot \mathbf{T}_{\text{SLAM}} \quad (1)$$

where Tcalib represents the calibration matrix aligning the SLAM coordinate system with the target robotic arm's workspace.

Visual inputs comprise RGB streams from three perspectives: two third-view cameras and a wrist camera.

Finally, we chose to use the ACT model for training within the imitation learning framework. By utilizing the ACT model, we can ensure that the dexterous hand exe-

cutes predetermined tasks efficiently and smoothly, thereby enhancing the task completion rate.

IV. EXPERIMENT

In this section, to evaluate the advantages of the proposed system, we conduct experiments on two key aspects of imitation learning: the efficiency of the data collection phase and the performance during real-world robot operation. The efficiency comparison of data collection using this system is shown in Fig.5 and Table II, while the task examples of real robotic arm deployment after training with imitation learning are presented in Fig.6.

A. Collection Efficiency

Currently, the primary form of data acquisition for imitative learning in dexterous manipulation tasks is represented by teleoperation. In this section, we compare the
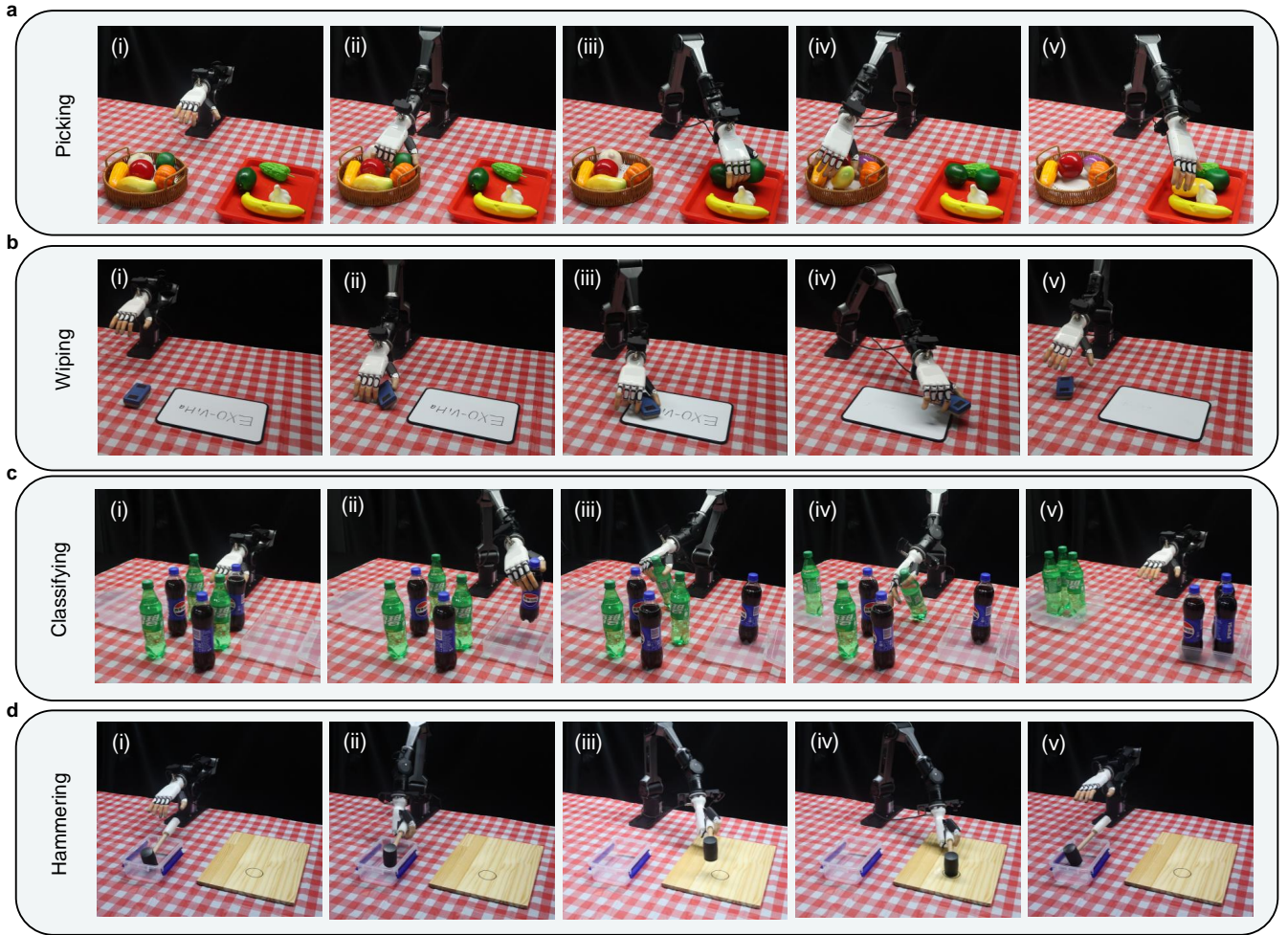
Fig. 6: **Deployment of various robotic tasks.** (a) Picking: pick and place multiple objects of different shapes. (b) Wiping: hold the whiteboard wipe to wipe the whiteboard writing. (c) Classifying: sort the bottles based on their types. (d) Hammering: hold the hammer and hit the designated position.

time required for different tasks using human hands versus traditional teleoperation with the timing of our system.

Fig. 5 illustrates the time taken by the three different data acquisition methods to complete the task of wiping a whiteboard with an eraser. The results indicate that, for this multi-contact manipulation task, the completion time of our system is only a few seconds slower than that of human hands. In contrast, teleoperation, due to the lack of first-person perspective for observing object contact, often led to issues where the dexterous hand exerted excessive force, resulting in hard contact with the surface or gripping the eraser without making actual contact with the whiteboard. These factors contributed to an increased task completion time, as detailed in Fig. 5 (b).

Furthermore, when performing simpler tasks such as picking up and placing objects, our system demonstrates high efficiency. As shown in Fig. 5 (c), this task involved placing items of various sizes and shapes. It is evident that the time taken to complete the task using our system is only 8 seconds longer than that of autonomous human movement, while the

use of teleoperation nearly doubled the time required.

TABLE II: Performance metrics (average time and success rate) for robotic manipulation tasks during data collection and robot deployment using the Exo-ViHa system.

Task	Data Collection		Imitation
	Avg. time (s)	Succ./Trials	Succ.Rate
Pick-place objects	4.8 ± 0.9	29/30	86.6%
Sort six bottles	41.8 ± 7.8	27/30	76.6%
Hammer manipulation	12.4 ± 3.3	28/30	83.3%
Wipe whiteboard	12.9 ± 2.1	27/30	80.0%

Avg. time represent mean $\pm$ standard deviation (n=30 trials)

Finally, we conducted tests on various task types for 30 iterations, as summarized in Table II. The system demonstrates reduced execution times and higher task completion rates across nearly all tasks

B. Manipulation Performance

For imitation learning, we conducted training on the LeRobot [32] platform using an NVIDIA RTX 4090 GPU. The training process consisted of a total of 160,000 offline

steps. For each action, we conducted 30 episodes, with each episode completing the corresponding task. The batch size was set to 16, and the learning rate was configured to 2e-5.

In the Fig. 6,We deployed the Piper robotic arm along with the Inspire dexterous hand to accomplish the tasks depicted. In Fig. 6 (a), The dexterous hand was used to grasp multiple fruit models and transfer them to the adjacent red plate. In Fig. 6 (b), Starting from the initial position, the dexterous hand picked up an eraser from the right side of the whiteboard and subsequently wiped the markings on the whiteboard from right to left. In Fig. 6 (c), On the table, there were six bottles of different colors; the dexterous hand sorted the bottles by color, placing each type into its corresponding plastic container. In Fig. 6 (d), The dexterous hand first grasped a hammer from the box, then lifted the hammer and moved it to a designated location, striking down into the target circle.

During the deployment phase, we conducted 30 trials for each task. As shown in Table II, the success rate for each task was generally around 80%. The success rate was highest for the relatively simple grasping tasks, while the success rate for the classifying task was lower. This was due to the large size of the bottle, combined with the use of a full-grip method during data collection, which required precise positioning of the dexterous hand. A slight positional errors during deployment could cause the bottle to be knocked over. Additionally, in the Hammering and Wiping tasks, most failures occurred during the initial step of grasping the hammer and whiteboard eraser. In most cases, once the hammer and eraser were successfully grasped, the robotic arm was able to complete the subsequent tasks.

V. DISCUSSION AND FUTURE WORK

In this study, we introduced the Exo-ViHa system, a novel 3D printed exoskeleton designed to achieve a balance between data collection efficiency, consistency, and accuracy in dexterous tasks. The system enables users to have first-person visual feedback and direct tactile sensations during data collection. Additionally, the exoskeleton design ensures that the visual information captured by the cameras during both data collection and deployment phases is highly consistent, while the dexterous hand data is directly derived from the hand itself, significantly reducing discrepancies in visual and body alignment. Extensive experiments have demonstrated that the data collection efficiency achieved with this system is nearly identical to that of human actions, significantly surpassing the efficiency of traditional teleoperation methods. Furthermore, the success rate of automatic task deployment in multi-contact tasks after training reaches about 80%, demonstrating the effectiveness and efficiency of the system.

Future work will focus on further structural improvements to the system: the forearm exoskeleton will be optimized to provide additional working space and comfort for the user's hand. Additionally, as data collection is a relatively lengthy process, improving the gravity compensation mechanism will be a key area of focus. Meanwhile, we will focus on evaluating the Exo-ViHa system in real-world settings, assessing

its performance in collaborative tasks. Through extensive field studies, we aim to ensure the system's practicality and effectiveness across diverse applications.

REFERENCES

[1] S. Nakatani and Y. Yamakawa, "Dynamic manipulation like normal-type pen spinning by a high-speed robot hand and a high-speed vision system," in *2023 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, pp. 636–642, IEEE, 2023.

[2] H. Munn, B. Tidd, D. Howard, and M. Gallagher, "Whole-body dynamic throwing with legged manipulators," *arXiv preprint arXiv:2410.05681*, 2024.

[3] C. Chi, Z. Xu, S. Feng, E. Cousineau, Y. Du, B. Burchfiel, R. Tedrake, and S. Song, "Diffusion policy: Visuomotor policy learning via action diffusion," *The International Journal of Robotics Research*, p. 02783649241273668, 2023.

[4] Y. Ze, G. Zhang, K. Zhang, C. Hu, M. Wang, and H. Xu, "3d diffusion policy: Generalizable visuomotor policy learning via simple 3d representations," *arXiv preprint arXiv:2403.03954*, 2024.

[5] S. P. Arunachalam, S. Silwal, B. Evans, and L. Pinto, "Dexterous imitation made easy: A learning-based framework for efficient dexterous manipulation," in *2023 ieee international conference on robotics and automation (icra)*, pp. 5954–5961, IEEE, 2023.

[6] S. P. Arunachalam, I. Güzey, S. Chintala, and L. Pinto, "Holo-dex: Teaching dexterity with immersive mixed reality," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5962–5969, IEEE, 2023.

[7] X. Cheng, J. Li, S. Yang, G. Yang, and X. Wang, "Open-television: Teleoperation with immersive active visual feedback," *arXiv preprint arXiv:2407.01512*, 2024.

[8] D. Rakita, B. Mutlu, and M. Gleicher, "Remote telemanipulation with adapting viewpoints in visually complex environments," *Robotics: Science and Systems XV*, 2019.

[9] B. E. Khadir, J. Varley, and V. Sindhwani, "Teleoperator imitation with continuous-time safety," *ArXiv*, vol. abs/1905.09499, 2019.

[10] S. Li, N. Hendrich, H. Liang, P. Ruppel, C. Zhang, and J. Zhang, "A dexterous hand-arm teleoperation system based on hand pose estimation and active vision," *IEEE Transactions on Cybernetics*, vol. 54, no. 3, pp. 1417–1428, 2022.

[11] 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," *arXiv preprint arXiv:2307.04577*, 2023.

[12] A. Handa, K. V. Wyk, W. Yang, J. Liang, Y.-W. Chao, Q. Wan, S. Birchfield, N. D. Ratliff, and D. Fox, "Dexpilot: Vision-based teleoperation of dexterous robotic hand-arm system," *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 9164–9170, 2019.

[13] J. Lee, O. Sushkov, T. Pevny, D. Morrison, J. Leitner, and P. Corke, "Learning fine-grained bimanual manipulation with low-cost hardware," *arXiv preprint arXiv:2304.13705*, 2023.

[14] S. Huang, H. Wang, Y. Wu, J. Li, H. Liu, and H. Su, "Dexcap: Scalable and portable mocap data collection system for dexterous manipulation," 2024.

[15] S. Bahl, R. Mendonca, L. Chen, U. Jain, and D. Pathak, "Universal manipulation interface: In-the-wild robot teaching without in-the-wild robots," 2024.

[16] J. Lee, O. Sushkov, T. Pevny, D. Morrison, J. Leitner, and P. Corke, "Forcemimic: Force-centric imitation learning with force-motion capture system for contact-rich manipulation," 2024.

[17] Z. Wu, T. Wang, C. Guan, Z. Jia, S. Liang, H. Song, D. Qu, D. Wang, Z. Wang, N. Cao, *et al.*, "Fast-umi: A scalable and hardware-independent universal manipulation interface," *arXiv preprint arXiv:2409.19499*, 2024.

[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] O. Yu and Y. She, "Feelit: Combining compliant shape displays with vision-based tactile sensors for real-time teletaction," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 13853–13860, IEEE, 2024.

[20] 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," *arXiv preprint arXiv:2407.03162*, 2024.

[21] H. Wang, S. Huang, J. Li, Y. Wu, H. Liu, and H. Su, "Ace: A cross-platform visual-exoskeletons system for low-cost dexterous teleoperation," 2024.

[22] D. Wei and H. Xu, "A wearable robotic hand for hand-over-hand imitation learning," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 18113–18119, IEEE, 2024.

[23] J. Cheng, Y. Chen, and Q. Chen, "Pluto: Pushing the limit of imitation learning-based planning for autonomous driving," *arXiv preprint arXiv:2404.14327*, 2024.

[24] Y.-F. Zhang, F.-M. Luo, and Y. Yu, "Improve generated adversarial imitation learning with reward variance regularization," *Machine Learning*, pp. 1–19, 2022.

[25] H. Xiong, R. Mendonca, K. Shaw, and D. Pathak, "Adaptive mobile manipulation for articulated objects in the open world," *arXiv preprint arXiv:2401.14403*, 2024.

[26] R. Mendonca, S. Bahl, and D. Pathak, "Structured world models from human videos," *arXiv preprint arXiv:2308.10901*, 2023.

[27] A. Brohan, N. Brown, J. Carbajal, Y. Chebotar, J. Dabis, C. Finn, K. Gopalakrishnan, K. Hausman, A. Herzog, J. Hsu, *et al.*, "Rt-1: Robotics transformer for real-world control at scale," *arXiv preprint arXiv:2212.06817*, 2022.

[28] J. Hathaway, R. Stolkin, and A. Rastegarpanah, "Imitation learning for sim-to-real transfer of robotic cutting policies based on residual gaussian process disturbance force model," *arXiv preprint arXiv:2311.04096*, 2023.

[29] X. Liu, "Interactive imitation learning in robotics based on simulations," *arXiv preprint arXiv:2209.03900*, 2022.

[30] K. Guo, Z. Miao, W. Jing, W. Liu, W. Li, D. Hao, and J. Pan, "Lasil: learner-aware supervised imitation learning for long-term microscopic traffic simulation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15386–15395, 2024.

[31] M. Caeiro-Rodríguez, I. Otero-González, F. A. Mikic-Fonte, and M. Llamas-Nistal, "A systematic review of commercial smart gloves: Current status and applications," *Sensors*, vol. 21, no. 8, p. 2667, 2021.

[32] H. Face, "lerobot." `https://github.com/huggingface/lerobot`, 2023. Accessed: March 1, 2025.