VRobotix: A Scalable and Cost-Effective Virtual-Reality-Based Robotic Manipulation Dataset Generation Framework

Xinmin Fang[1], Zheshuo Li[1], Lingfeng Tao[2], Zhengxiong Li[1]

***Abstract*— Large-scale, diverse datasets are essential for training robust learning-based robotic manipulation models; however, their acquisition typically requires controlled environments and specialized hardware in research laboratories. This paper presents VRobotix, a virtual reality (VR)-based framework that enables cost-effective and scalable robotic dataset generation through immersive human-in-the-loop control within a physics-accurate robot simulation. By leveraging off-the-shelf VR headsets (e.g., Oculus Quest 3), VRobotix eliminates the need for physical robots while supporting a URDF-compatible, physics-based simulator that accommodates adaptable robotic platforms and egocentric control interfaces, including handheld controllers and body posture tracking. Benefiting from the physics-based simulation, a unique contribution of VRobotix is the replay module, which can regenerate synchronized multi-modal dataset (kinematic states, RGB-D streams) with multiple dataset formats based on the replayable trajectory, supporting various robotic applications. Additionally, an imitation learning module is developed to train control policies using the data collected by VRobotix. Experiments on three initial tasks—pushing, grasping, and stacking—demonstrate a high data collection success rate, averaging 92.0%. Furthermore, policies trained on just 50 trials achieve a 100% task success rate. VRobotix reduces infrastructure costs while generating ROS-compatible datasets, democratizing scalable robotic data acquisition.**

I. Introduction

High-quality, large-scale datasets are increasingly crucial for robotic learning and manipulation. These datasets help mitigate the risk of overfitting to limited operational scenarios, thereby enhancing policy adaptability to dynamic environments and improving robustness in real-world deployments. Moreover, recent advancements in robotic manipulation have shown that larger and more diverse training datasets significantly improve policy generalization capabilities [1]. As a result, the availability of such datasets serves as a fundamental pillar for advancing data-driven robotic manipulation.

Traditional methods for acquiring large-scale robotic manipulation datasets rely on extensive infrastructure, including specialized robotic hardware, sensor arrays, and controlled laboratory environments designed to simulate real-world conditions [2]. However, these requirements introduce substantial financial and logistical costs. Additionally, the labor-intensive setup for task-specific well-controlled environments hinders scalability and reproducibility. These challenges highlight the urgent need for cost-effective, scalable alternatives to traditional data collection paradigms.

*This work is partially supported by US NSF Awards 2426269 and 2426470, as well as the 2022 Meta Research Award.

[1]X. Fang, Z. Li, and Z. Li are with the Department of Computer Science and Engineering, University of Colorado Denver, Denver, CO 80217 USA (e-mail: xinmin.fang@ucdenver.edu, zheshuo.li@ucdenver.edu, zhengxiong.li@ucdenver.edu)

[2]L. Tao are with the Oklahoma State University, 563 Engineering North, Stillwater, OK, 74078, USA (e-mail: lingfeng.tao@oksate.edu)

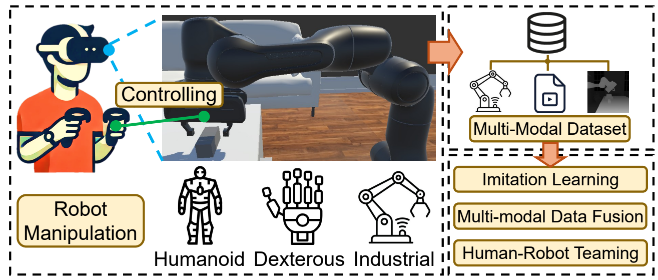

Fig. 1. VRobotix is a cost-effective and scalable robotic manipulation dataset generation framework via physics-accurate robotic simulation.

Therefore, we wonder about a question: *Can high-quality, large-scale robotic datasets be acquired without prohibitive costs?* To address the challenges of high cost and significant time investment, we require a cost-effective and efficient approach to robotic dataset collection. The proliferation of virtual reality (VR) technologies offers a compelling solution. The inherent capabilities of VR, such as unconstrained egocentric observation and multi-degree-of-freedom interaction, enable immersive real-time control of robotic twins in virtual environments. By decoupling data collection from physical hardware constraints, VR allows large-scale generation of diverse manipulation trajectories (e.g., grasping, pushing, and stacking) with minimal infrastructure.

In this paper, we propose VRobotix (Fig. 1), which leverages VR to enable cost-effective and scalable robotic manipulation dataset generation. Using intuitive VR interfaces, users control robots in simulated environments by mapping human motions (e.g., handheld controllers, body posture) to precise robot kinematics, eliminating costly hardware dependencies or complex physical environments. This strategy supports scalable data generation, which captures multi-modal dataset (trajectories, RGB-D streams, etc.) with multiple formats, significantly reducing operational costs and enhances the applicability of the dataset. By democratizing high-quality data collection through accessible VR hardware, VRobotix establishes a scalable, cost-efficient paradigm for training robust robotic policies.

Our work includes the following key contributions:

- We are one of the first to systematically exploit VR for generating large-scale, high-quality robotic manipulation datasets, eliminating physical hardware limitations while preserving task diversity and kinematic fidelity.

- We design a VR framework that can be used to generate large-scale, high-quality robotic manipulation datasets. It eliminates reliance on external hardware, integrates real-time physics simulation, and supports Unified Robot Description Format (URDF)-compliant robot models. This framework enables rapid, cost-effective data collection across heterogeneous robotic platforms and control modalities.
- VRobotix enables scalable data generation, which captures multi-modal data (motion trajectories and visual observations) with multiple formats. By replaying the replayable trajectories collected during user demonstrations, we can collect data from multiple sensor modalities (e.g., RGB-D cameras) to generate a comprehensive multi-modal dataset. This dataset, in turn, supports policy training using a variety of approaches.
- Data collection shows a high success rate (92.0%) and a short collection time (9.6s), demonstrate that VRobotix can efficiently conduct large-scale robotic data collection. Furthermore, the imitation learning results (100% success rate) confirm the datasets gathered through VRobotix are highly effective for policy training.
- Our experimental findings underscore that leveraging VR, complemented by well-designed UI, can achieve efficient large-scale robotic data collection. Furthermore, our results indicate that handheld controllers currently offer superior performance due to their higher state accuracy and reduced latency, underscoring their utility in robust data acquisition workflows.

II. Related Work

A. Data Collection of Robot Manipulation Dataset

Prior efforts in collecting robotic datasets mainly followed two paradigms: physical hardware deployments and synthetic simulation. (1) **Physical Hardware-Driven Approaches**: Large-scale real-world datasets like DROID [2] demonstrate the value of distributed human teleoperation across diverse tasks (350 hours, 76k trajectories). However, such efforts require costly infrastructure (50 operators, 12 months) and face scalability bottlenecks due to hardware maintenance and environmental constraints. Similarly, many platforms like RoboNet [1], BridgeData [3], and others [4], [5], [6], [7] suffer from inherent trade-offs between data diversity and hardware expenditure. (2) **Simulation Methods**: Photorealistic virtual environments (e.g., UnrealROX [8], RobotriX [9]) generate synthetic data but prioritize visual realism over robotic interaction fidelity. While RobotriX provides trajectories, its humanoid focus and scripted actions limit dexterous manipulation applicability. Recent VR systems (e.g., [10]) capture human poses but lack integration with robotic kinematics or physics-aware interaction modeling. Existing VR frameworks either neglect robotic embodiment (e.g., 3D-ARM-Gaze [10] for human motion) or oversimplify physical interactions (e.g., Garment [11], DemoGen [12]).

Crucially, none addresses the requirements of high-fidelity robot dynamics environment and cost-effective, scalable, human-in-the-loop data acquisition. Our work advances these directions by unifying high-fidelity physics simulation, precise human control schemes, and scalable data generation within a single VR system.

B. Robot Manipulation using VR

Robot manipulation using VR has advanced through immersive teleoperation and shared-control paradigms. Early work by Burdea et al. [13] identified VR's potential to address teleoperation challenges like delayed feedback. Subsequent systems, such as [14], enabled real-time 3D manipulation for industrial robots but lacked physics-based grounding. Recent efforts prioritize realism and usability: Sergiu et al. [15] developed a VR grasping system with adaptive hand-object interactions, while ROS Reality [16] demonstrated immersive teleoperation for physical robots, albeit with hardware dependencies.

Shared-control methods further refine efficiency. Shiyu et al. [17] combined VR with robot autonomy for manipulation tasks, yet required motion-capture devices. Similarly, [18] trained users in virtual environments but omitted multi-modal data capture. While these systems enhance interaction, most rely on external hardware, lack URDF support, or prioritize visual fidelity over physical accuracy. Our framework fills these gaps by enabling physics-based, device-agnostic manipulation with scalable data generation for large-scale dataset collection using diverse robots and control schemes.

III. System Design

VRobotix is a VR framework based on unity engine, which is compatible with various VR platforms. As shown in Fig. 2, the VRobotix comprises four core components: 1) VR Physics-based Virtual Robot Environment, 2) Controller Module, 3) Data Capture Module, and 4) Replay Module.

A. VR Physics-based Virtual Robot Environment

To ensure physically plausible interactions, we establish a high-fidelity simulation environment powered by *NVIDIA PhysX*, a widely adopted physics engine optimized for real-time rigid-body dynamics. To achieve a high-fidelity robotic simulation environment, we utilize PhysX to implement rigid-body simulation, incorporating friction, restitution, linear and angular damping, as well as articulation structures. The simulation runs at 100Hz with solver iteration counts set to 12 to minimize penetration and contact errors during complex manipulation tasks, which strikes a good balance between the computational resource constraints of mobile VR and the precision requirements. To support heterogeneous robotic platforms, VRobotix incorporates an URDF importer [19], enabling direct import of robot kinematic chains, inertial properties, and visual meshes. This standardized workflow guarantees that simulated robots exhibit motion fidelity congruent with their real-world counterparts.

B. Controller Module

Since the human body structure may not necessarily match that of the robot, and high-precision robot control is essential for executing complex tasks, it is vital to provide

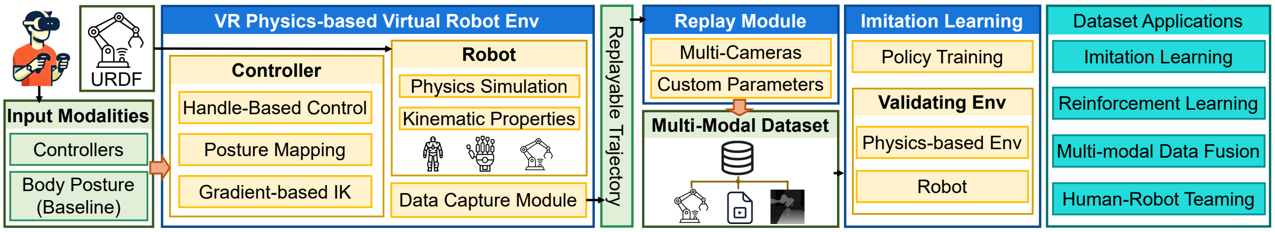

Fig. 2. VRobotix is a cost-effective and scalable robotic manipulation dataset generation framework via physics-accurate robotic simulation. Users intuitively control robots by multiple input methods, producing task-agnostic manipulation multi-modal datasets for many applications.

a user-friendly, efficient, and accurate control interface. In VRobotix, we design and implement a **Controller Module** that supports multiple user input modalities (e.g., handheld controllers, body posture) and flexibly maps them to various robot control methods. These control methods include: (1) **Handle-Based Control**: Thanks to the high positioning accuracy and low latency of handheld controllers, they are ideally suited for controlling robots in high-precision tasks. In this mode, buttons and controller states are mapped to the robot's joint movements, affording intuitive, gamepad-like control. (2) **Posture Mapping (Baseline)**: Leveraging VR's unique capability for full-body tracking, we develop a control method based on body posture. By mapping the user's physical movements to the robot's actions, this approach offers an exceptionally high degree of freedom. We leverage Meta's Movement SDK to capture the user's full-body motion. This eliminates the need for external trackers or specialized devices beyond a standard VR headset. (3) **Inverse Kinematics (IK) Control**: To simplify control and shorten the learning curve in VRobotix, we integrated a fast, gradient-based Inverse Kinematics (IK) algorithm from [20]. This algorithm's speed is essential for maintaining the high frame rates required by VR and our 100Hz physics simulation, different from other methods such as Fast IK [21], which cannot meet these temporal demands. The IK module is designed to be modular, allowing for easy integration of alternative solvers to handle more complex robots and constraints. Both controller and body posture can serve as target inputs to this IK module.

These three control methods can be freely combined to control different components of a single robot simultaneously. For instance, users might deploy IK to manage the robotic arm of a Franka Panda arm, use the controller to position the wrist, and trigger gripper actions via controller buttons. This hybrid strategy greatly enhances the flexibility and accuracy of robot manipulation in virtual space and ensures that VRobotix can generate heterogeneous, high-quality data for a wide array of robots and manipulation tasks. By accommodating these control modalities under one framework, VRobotix significantly extends the scope and precision of data collection, thereby establishing a robust foundation for large-scale robotic manipulation research.

C. Data Capture Module

To meet the frame rate and computational constraints on VR, we decouple data recording from data generation, ensuring that real-time performance remains unaffected by additional overhead, such as rendering and video encoding. Specifically, it records the full state of the scene at a frequency of 100 Hz, including all objects and robots. The logging result is a replayable trajectory that accurately captures users' interactions. The generated trajectory is subsequently provided to the Replay Module, which can reproduce the user's operations in detail, enabling thorough analysis and further data processing. This design allows VRobotix to reliably gather high-fidelity manipulation data while preserving an immersive and responsive virtual environment.

D. Replay Module

To make the datasets applicable to a wide range of robots and application scenarios, we designed the Replay Module. During this process, all physics simulations and interactions are disabled, as the objects' states are directly sourced from the captured trajectory. At each frame of the replay, the module sets object poses according to the recorded trajectory, updates virtual cameras, and produces the requested sensor data. This Module then generates the required multi-modal data based on the user's needs, such as RGB images, depth maps, or any other modality. It also supports multiple dataset formats (including LeRobot [22] and Open X-Embodiment [4]). The extensible design allows VRobotix to accommodate diverse research and application requirements.

E. Imitation Learning

To verify the quality of VRobotix's dataset, we design and implement an imitation learning framework to evaluate whether it can effectively train a policy to accurately complete specified tasks. We select behavior cloning [23], essentially a supervised learning approach that directly maps states to actions using expert trajectories. Compared to other imitation learning methods, such as inverse reinforcement learning (which requires additional online data), behavior cloning best reflects the quality of an offline dataset. Besides, selecting appropriate model inputs is crucial for ensuring robust policy performance. Therefore, we include timestamps, current task stages, object poses, end-effector poses, distances from the end-effector to the object, distances from the object to the goal, robot joint angles, and end-effector contact states. The timestamp of the input helps the model move forward rather than repeating actions. The model's output is the target joint angles of the robot. For example, there will be a 38-dimensional input and a 8-dimensional output for a Franka Panda robot equipped with a gripper manipulating a single object.

Different operators can exhibit distinct motion preferences, yielding highly divergent demonstrations. To capture this

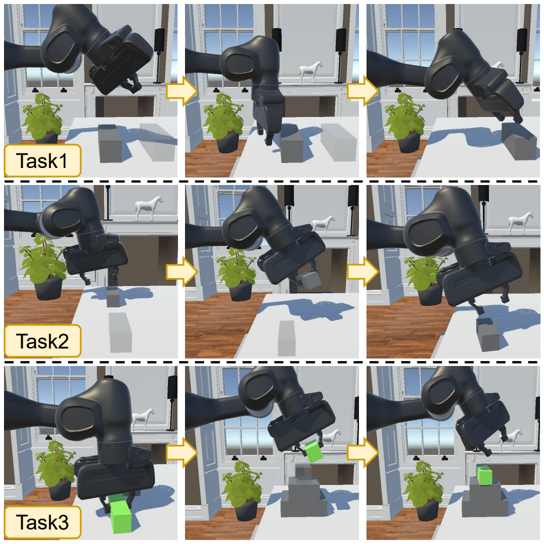

Fig. 3. Demonstrations of three robotic manipulation tasks: 1) Cube Pushing 2) Grasp and Place 3) Precision Stacking.

diversity without averaging them into a potentially infeasible policy, we adopt a Mixture of Experts (MoE) model that preserves behavioral variability and improves accuracy. The MoE model consists of a gating network, which comprises an LSTM layer followed by a dense layer and multiple expert networks, each incorporating an encoder-decoder framework built from two LSTM layers, followed by a dense layer. The final policy output is a weighted sum of each expert's output, as determined by the gating network.

Additionally, VRobotix provides a comprehensive model validation mechanism. By launching an external inference service and connecting to it, VRobotix transmits model inputs and receives outputs, executing the learned policy in the identical scene used for data collection. This architecture not only ensures compatibility with most deep neural networks, but also simplifies model development and testing.

IV. Experiments

A. Device

To assess the effectiveness of VRobotix in collecting high-quality robotic manipulation data, we utilize the Meta Oculus Quest 3, a state-of-the-art VR product equipped with onboard cameras for body motion tracking. This standalone headset has already sold more than one million units worldwide, underscoring its widespread adoption and user familiarity. Priced at approximately $499, the Quest 3 offers a cost-effective yet robust VR solution. This all-in-one nature significantly reduces the setup complexity, enabling an accessible and scalable robotic data collection approach.

B. Baseline

In our experimental evaluation, we adopt the body posture modality as the baseline, enabled by the official Meta Movement SDK. Using this input modality, we set the user's wrist as the IK target of the robotic arm and maps the user's finger to the end-effector.

C. Task Selection

As shown in Fig. 3, we evaluate our VRobotix framework using three representative manipulation tasks that vary in complexity and motor demands. The robot platform we utilize for illustration is the Franka Panda, which has nine DoFs. In the evaluation, we set the state of the controller as the IK target of the robotic arm, with button triggers enabling discrete end-effector actuation. These tasks are carefully designed to assess both the efficacy of our controlling method and the fidelity of our physics simulation. (1) **Task 1 Cube Pushing**: In this task, participants push a single cube from its initial position to a designated target zone. This task focuses on evaluating the basic translational control and interaction dynamics of the system. Success is measured by the cube's final position relative to the target. (2) **Task 2 Grasp and Place**: Participants grasp a cube and then transport and place it at a specified target location. This task assesses the system's capability to accurately map user gestures into precise robotic manipulation. (3) **Task 3 Precision Stacking**: The most complex of the three tasks, this scenario requires participants to pick up a cube and stack it on top of an existing tower constructed from five cubes. This task demands a higher level of dexterity and precision, as the participant must ensure that the cube is stably and accurately positioned atop the tower. For all tasks, both the initial and target positions of the manipulated objects are randomly varied within a radius of 3 centimeters.

D. Data Collection

For evaluating the performance of VRobotix in generating high-quality robotic manipulation data, we conduct a data collection study involving 10 participants following the Institutional Review Board (IRB) protocol. Half of the participants were familiar with VR operations, while the other half were not. Participants will familiarize themselves with the system before starting data collection. During the experiment, participants performed all three designated tasks using both the control methods. For each task, we collect five success trials, resulting in 50 trails for each task and each controlling method, ensuring a diverse set of demonstrations across multiple trials. All interaction data, including joint trajectories, object positions, and virtual sensor feedback, were captured by our integrated Data Capture Module. These data are used for the following imitation training.

E. Imitation Training

We evaluate the scalability and robustness of our approach by training imitation learning models using varying amounts of demonstration data. Specifically, subsets containing 30, 40, and 50 trails per task. The hidden dim size is 256, and the number of experts is five. The input and output sequence length is 30 and executes with a length of two per inference. To enhance generalization, Gaussian noise with a standard deviation of 0.005 is added to the input data. Upon completion of training, the trained policies are evaluated using VRobotix's integrated model validation functionality. Each policy is deployed in the same virtual environment used for data collection.

TABLE I

PERFORMANCE OF IMITATION LEARNING ACROSS TWO INPUT METHODS AND THREE TASKS MEASURES 100 TRAILS.

	Handheld Controllers									Body Posture (Baseline)								
	Task1			Task2			Task3			Task1			Task2			Task3		
Data Amount	30	40	50	30	40	50	30	40	50	30	40	50	30	40	50	30	40	50
Task Success Rate%	100	100	100	100	100	100	100	100	100	65.0	83.0	100	61.0	81.0	91.0	66.0	78.0	95.0
Grasp Success Rate%	N/A	N/A	N/A	100	100	100	100	100	100	N/A	N/A	N/A	89.0	100	96.0	96.0	100	98.0
Completion Time (s)	9.7	5.2	6.8	5.3	6.8	7.4	6.9	7.5	8.1	7.3	7.5	5.1	6.8	6.9	9.7	74.6	29.4	25.6
Cube Error (cm)	5.51	0.47	3.48	2.04	1.68	0.87	1.37	1.55	3.33	5.89	6.04	4.47	2.65	4.62	2.68	2.35	2.56	2.27
$RMSE_{cube\_a}$ (m/s^2)	0.20	0.36	0.34	1.31	1.34	1.05	0.38	0.56	0.34	0.36	0.68	0.53	0.93	1.12	1.48	1.34	1.00	0.92
$RMSE_{joint\_a}$ (m/s^2)	0.48	0.43	0.65	0.59	0.51	0.50	0.34	0.44	0.65	1.37	1.65	1.43	1.19	1.01	1.34	0.54	0.65	0.72

F. Metrics

To rigorously evaluate the performance of our system, we define the following metrics: (1) **Task Success Rate**: The percentage of trials in which the designated task is successfully completed. (2) **Grasp Success Rate**: For tasks involving object manipulation, this metric measures the percentage of successful grasps. (3) **Completion Time**: The duration required to collect one trail for each task. (4) **Cube Error**: It is defined as the Euclidean distance between the cube's final position and the designated target location at the end of the task. (5) $RMSE_{cube\_a}$: It quantifies the steadiness of the cube during the entire task execution by calculating its Root Mean Square (RMS) acceleration. Let a_i represent the instantaneous acceleration of the cube at time step i over N time steps. The $RMSE_{cube\_a}$ is then defined as: $RMSE_{cube\_a} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} a_i^2}$. A lower $RMSE_{cube\_a}$ implies the cube maintained a more stable position throughout the manipulation process. (6) $RMSE_{joint\_a}$: Similarly, $RMSE_{cube\_a}$ is assessed by computing the RMS acceleration of the robot's joints during task execution. Denote j_i as the instantaneous acceleration of a joint at time step i. The $RMSE_{cube\_a}$ is given by: $RMSE_{joint\_a} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} j_i^2}$.

V. RESULTS AND DISCUSSION

A. Data Collection Performance

Table II presents the comparative performance of handheld controllers and body posture input modalities in three manipulation tasks in VRobotix. The results indicate that handheld controllers consistently achieve higher task success rates across all tasks. As task complexity increases, the performance gap becomes more pronounced, with advantages of 4.6%, 14.5%, and 35.9% observed across the three tasks, respectively. This progression highlights fundamental challenges in body posture: despite employing IK to mitigate control complexity, the inherent morphological mismatch between the Franka Panda manipulator and human arm kinematics imposes significant cognitive load and adaptation requirements on users.

TABLE II

DATA COLLECTION PERFORMANCE OF TEN PARTICIPANTS ACROSS TWO INPUT MODALITIES AND THREE TASKS.

	Handheld Controllers			Body Posture		
	Task1	Task2	Task3	Task1	Task2	Task3
Task Success Rate%	94.5	90.1	91.3	90.3	78.7	67.2
Grasp Success Rate%	N/A	98.0	91.3	N/A	91.7	95.2
Completion Time (s)	9.1	10.1	9.6	11.6	22.6	20.4
Cube Error (cm)	4.03	1.33	1.51	3.99	2.80	2.15
$RMSE_{cube\_a}$ (m/s^2)	0.74	1.32	1.25	0.8	1.36	1.28
$RMSE_{joint\_a}$ (m/s^2)	1.03	0.85	0.77	1.25	1.15	0.99

The grasp success rate further validates this observation. For Tasks 2&3, the disparity between task and grasp success rates under controllers (7.9% and 0%, respectively) remains substantially lower than body posture (13% and 28%), indicating superior fine control precision. The completion time reinforces this advantage, with controllers taking only 9-10s per task, suggesting more intuitive spatial targeting and reduced adjustments. It also shows VRobotix's low cost compared to datasets based on physical hardware [1], [2].

The $RMSE_{cube_a}$ and $RMSE_{joint_a}$ of the controller-based policy are lower than the body posture-based policy; it might be due to the inherent limitations in body motion tracking systems (Meta Quest 3). Specifically, tracking latency and positional jitter introduce cumulative errors in manipulator control, requiring frequent compensatory adjustments. These findings collectively demonstrate that while body posture mapping offers naturalistic interaction paradigms, handheld controllers currently provide superior data collection fidelity, particularly for complex manipulation tasks.

B. Imitation Learning Performance

Table I quantifies the policy performance trained on datasets collected via VRobotix's two input methods. Notably, policies trained with handheld controller trails achieve 100% task success rates across all tasks and dataset sizes (30/40/50 trajectories), outperforming body posture-based policies by 21.0%, 28.8%, and 25.5% on Tasks 1, 2, and &3, respectively. These results demonstrate that VRobotix efficiently generates diverse robotic datasets while enabling the training of policies that successfully accomplish complex manipulation tasks with a high success rate.

The disparity in grasp success rates further validates this finding. While body posture-trained policies exhibit 0–12% lower grasp success rates compared to handheld policies, the margin narrows relative to task success rates. We attribute this to systemic failures during transportation and placement phases in body posture demonstrations, policies often fail to maintain stable grasps or achieve precise final positioning. Temporal metrics reinforce this analysis: body posture-trained policies require 35.7s longer on average for Task 3, with frequent pauses and corrective motions indicative of suboptimal trajectory learning. These performance limitations stem from inherent noise in body posture data collection. VR body motion tracking introduces kinematic jitter during arm motion capture, which propagates into two policy defects: 1) Body posture policies exhibit 113.4% higher $RMSE_{joint}$ than handheld controllers, reflecting unstable joint angle predictions caused by tracking inaccuracies. 2) The 74.8% increase in $RMSE_{cube}$ for body posture policies demonstrates compromised spatial precision, particularly detrimental in precision stacking (Task 3).

C. Discussion and Key Takeaways

Overall, our findings underscore the potential of VRobotix as a practical, cost-effective solution for large-scale robotic manipulation data collection by harnessing a standalone VR headset—eliminating the need for extensive external hardware. Our results highlight the relative advantages of handheld controllers, more precise positional tracking, reduced jitter, and lower latency. The resulting datasets enable effective imitation learning (100% success rate), illustrating that immersive, multi-degree-of-freedom interactions in VR can reliably capture the nuances of human expertise. In the future, we can leverage domain randomization and other techniques to achieve robust sim-to-real transfer, further bridging the gap between virtual demonstrations and real-world robotic performance.

VI. Conclusion

VRobotix introduces a VR framework for scalable robotic dataset collection using off-the-shelf hardware (Meta Quest 3). Our system enables cost-effective capture of high-precision human demonstrations. The framework supports URDF robots, multi-modal data capture, and diverse control schemes without external devices. Experiments show 92.0% task success rate while data collection and policies trained on 50 trails achieve 100% success rates across tasks with high accuracy and stability. VRobotix democratizes large-scale data collection, bridging the gap between accessible VR tools and robust robot learning.

References

[1] S. Dasari, F. Ebert, S. Tian, S. Nair, B. Bucher, K. Schmeckpeper, S. Singh, S. Levine, and C. Finn, "Robonet: Large-scale multi-robot learning," *arXiv preprint arXiv:1910.11215*, 2019.

[2] A. Khazatsky, K. Pertsch, S. Nair, A. Balakrishna, S. Dasari, S. Karamcheti, S. Nasiriany, M. K. Srirama, L. Y. Chen, K. Ellis, *et al.*, "Droid: A large-scale in-the-wild robot manipulation dataset," *arXiv preprint arXiv:2403.12945*, 2024.

[3] H. R. Walke, K. Black, T. Z. Zhao, Q. Vuong, C. Zheng, P. Hansen-Estruch, A. W. He, V. Myers, M. J. Kim, M. Du, *et al.*, "Bridgedata v2: A dataset for robot learning at scale," in *Conference on Robot Learning*. PMLR, 2023, pp. 1723–1736.

[4] A. O'Neill, A. Rehman, A. Maddukuri, A. Gupta, A. Padalkar, A. Lee, A. Pooley, A. Gupta, A. Mandlekar, A. Jain, *et al.*, "Open x-embodiment: Robotic learning datasets and rt-x models: Open x-embodiment collaboration 0," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 6892–6903.

[5] A. Mandlekar, Y. Zhu, A. Garg, J. Booher, M. Spero, A. Tung, J. Gao, J. Emmons, A. Gupta, E. Orbay, *et al.*, "Roboturk: A crowdsourcing platform for robotic skill learning through imitation," in *Conference on Robot Learning*. PMLR, 2018, pp. 879–893.

[6] S. Liu, L. Wu, B. Li, H. Tan, H. Chen, Z. Wang, K. Xu, H. Su, and J. Zhu, "Rdt-1b: a diffusion foundation model for bimanual manipulation," *arXiv preprint arXiv:2410.07864*, 2024.

[7] N. M. M. Shafiullah, A. Rai, H. Etukuru, Y. Liu, I. Misra, S. Chintala, and L. Pinto, "On bringing robots home," *arXiv preprint arXiv:2311.16098*, 2023.

[8] P. Martinez-Gonzalez, S. Oprea, A. Garcia-Garcia, A. Jover-Alvarez, S. Orts-Escolano, and J. Garcia-Rodriguez, "Unrealrox: an extremely photorealistic virtual reality environment for robotics simulations and synthetic data generation," *Virtual Reality*, vol. 24, pp. 271–288, 2020.

[9] A. Garcia-Garcia, P. Martinez-Gonzalez, S. Oprea, J. A. Castro-Vargas, S. Orts-Escolano, J. Garcia-Rodriguez, and A. Jover-Alvarez, "The robotrix: An extremely photorealistic and very-large-scale indoor dataset of sequences with robot trajectories and interactions," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 6790–6797.

[10] B. Lento, E. Segas, V. Leconte, E. Doat, F. Danion, R. Péteri, J. Benois-Pineau, and A. de Rugy, "3d-arm-gaze: a public dataset of 3d arm reaching movements with gaze information in virtual reality," *Scientific Data*, vol. 11, no. 1, p. 951, 2024.

[11] A. Boix-Granell, S. Foix, and C. Torras, "Garment manipulation dataset for robot learning by demonstration through a virtual reality framework," in *Artificial Intelligence Research and Development*. IOS Press, 2022, pp. 199–208.

[12] Z. Xue, S. Deng, Z. Chen, Y. Wang, Z. Yuan, and H. Xu, "Demogen: Synthetic demonstration generation for data-efficient visuomotor policy learning," *arXiv preprint arXiv:2502.16932*, 2025.

[13] G. C. Burdea, "Invited review: the synergy between virtual reality and robotics," *IEEE Transactions on Robotics and Automation*, vol. 15, no. 3, pp. 400–410, 1999.

[14] Y. Su, Y. Xu, S. Cheng, C. Ko, and K.-Y. Young, "Development of an effective 3d vr-based manipulation system for industrial robot manipulators," in *2019 12th Asian Control Conference (ASCC)*. IEEE, 2019, pp. 1–6.

[15] S. Oprea, P. Martinez-Gonzalez, A. Garcia-Garcia, J. A. Castro-Vargas, S. Orts-Escolano, and J. Garcia-Rodriguez, "A visually realistic grasping system for object manipulation and interaction in virtual reality environments," *Computers & Graphics*, vol. 83, pp. 77–86, 2019.

[16] D. Whitney, E. Rosen, E. Phillips, G. Konidaris, and S. Tellex, "Comparing robot grasping teleoperation across desktop and virtual reality with ros reality," in *Robotics Research: The 18th International Symposium ISRR*. Springer, 2019, pp. 335–350.

[17] S. Xu, S. Moore, and A. Cosgun, "Shared-control robotic manipulation in virtual reality," in *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*. IEEE, 2022, pp. 1–6.

[18] B. R. Galarza, P. Ayala, S. Manzano, and M. V. Garcia, "Virtual reality teleoperation system for mobile robot manipulation," *Robotics*, vol. 12, no. 6, p. 163, 2023.

[19] U. Technologies, "Urdf importer for unity," https://github.com/Unity-Technologies/URDF-Importer, 2024, accessed: 2025-07-04.

[20] A. Zucconi, "Inverse kinematics for robotic arms," Aug 2024. [Online]. Available: https://www.alanzucconi.com/2017/04/10/robotic-arms/

[21] J. Tenhumberg, A. Mielke, and B. Bäuml, "Efficient learning of fast inverse kinematics with collision avoidance," in *2023 IEEE-RAS 22nd International Conference on Humanoid Robots (Humanoids)*. IEEE, 2023, pp. 1–8.

[22] R. Cadene, S. Alibert, A. Soare, Q. Gallouedec, A. Zouitine, and T. Wolf, "Lerobot: State-of-the-art machine learning for real-world robotics in pytorch," https://github.com/huggingface/lerobot, 2024.

[23] C. Sammut and G. I. Webb, *Encyclopedia of machine learning*. Springer Science & Business Media, 2011.