Unity Factory Simulation with VR Teleoperation for Bimanual Manipulation and Data Collection

Jing Dai
School of Future Technology
Shanghai University
Shanghai, China
daijing18808224723@shu.edu.cn

Zeyan Huang
School of Future Technology
Shanghai University
Shanghai, China
ximogu0305@shu.edu.cn

Yaoguang Lei
School of Future Technology
Shanghai University
Shanghai, China
loupi233@shu.edu.cn

Linqi Ye
School of Future Technology
Shanghai University
Shanghai, China
yelinqi@shu.edu.cn

***Abstract*—Aiming at the prominent bottlenecks in industrial bimanual robot research, including scarce real-world manipulation data, high physical collection costs, hazardous on-site operation environments, and monotonous demonstration samples, this paper constructs a factory production line simulation platform based on Unity. The platform integrates conveyor-based logistics, autonomous box transportation of mobile bimanual robots, and dual-mode immersive VR human–robot teleoperation. Gamified task design is adopted to lower the operation threshold for repeated data collection, while an auditable recording pipeline automatically captures high-precision full-dimensional data including robot trajectories, joint poses, visual observations and sequential action signals. We validate the platform on a conveyor box-handling task with structured practice and progressive challenge modes. Experimental results show that the system can efficiently generate diverse embodied manipulation datasets. This work provides a low-cost, traceable and versatile data acquisition solution to support the research and deployment of industrial embodied intelligence.**

Keywords—virtual reality, robot teleoperation, bimanual manipulation, Unity factory, embodied intelligence

I. INTRODUCTION

Data serves as the core foundation of embodied intelligence. Large public datasets including DROID [1], Open X-Embodiment [2], and BridgeData V2 [3] have proven that sufficient data volume and environmental diversity greatly improve the practicability of robotic manipulation data. Although bimanual mobile robots have extensive application value in factory logistics and material handling, real-world on-site data collection faces multiple critical limitations: physical factory debugging brings heavy labor and equipment expenses, robotic operation scenarios contain potential safety hazards for staff, raw sensor motion data needs time-consuming manual annotation, and professional manual teaching significantly raises the acquisition cost of high-quality manipulation demonstrations.

Existing teleoperation systems reduce the threshold of demonstration collection via low-cost hardware, motion retargeting and immersive visual feedback [4]–[11]. Physics-based simulation environments further simplify data acquisition and support long-sequence mobile manipulation, object-oriented task setup and VR human demonstration recording [12]–[15]. Gamification can provide standardized task organization mechanisms for repeated data collection [17], though relevant studies indicate score and level settings may alter operators' task behaviors without enhancing intrinsic motivation [18], [19]. Nevertheless, current robotic simulation platforms still present prominent limitations: most rely on oversimplified isolated workstations lacking complete factory production line logic, few integrate gamified structured workflows to standardize repetitive collection, and they are not equipped with systematic, traceable pipelines for sorting and storing demonstration data.

To address the above-mentioned bottlenecks of real-world data collection and the deficiencies of existing simulation platforms, this paper constructs a Unity-based factory simulation platform integrated with immersive VR teleoperation for bimanual conveyor box transportation tasks. The platform reconstructs complete end-to-end assembly line logistics with automatic conveyor feeding, and develops two gamified task modes, namely low-pressure Practice Mode and progressive Challenge Mode, to standardize repeated human demonstration acquisition. Furthermore, a fault-aware four-layer session–trial–event–frame data provenance pipeline is embedded into the system, which automatically synchronously records high-precision multi-dimensional information covering robot base movement, dual-arm TCP (Tool Center Point) trajectories, joint angles, grasp states and frame-level motion signals, and supports full backward auditing from trial statistical results to original frame samples.

The main contributions of this work are summarized into three aspects:

(1) A complete industrial factory simulation environment built on Unity, which realizes cyclic conveyor material supply and mobile bimanual box handling. Combined with dual VR teleoperation gamified modes, it standardizes long-term repeated demonstration collection without relying on physical factory equipment.

(2) A hierarchical auditable data storage architecture with four linked layers of session, trial, event and frame, which distinguishes valid successful trials, timeout failures and incomplete interrupted records, and enables full traceability and anomaly diagnosis of all collected demonstration data.

(3) A set of quantitative evaluation metrics for bimanual manipulation demonstration quality, including task completion time, dual-TCP midpoint path length, grip-gated joint smoothness and active-action ratio. Single-operator repeated trials verify that the platform can stably generate diversified, verifiable embodied datasets to provide low-cost and safe data support for industrial robot learning.

II. RELATED WORK

A. Immersive and Whole-Body Teleoperation

Teleoperation systems balance hardware cost, operator immersion and portability. RoboTurk adopts consumer-grade devices to realize large-scale remote demonstration collection [4], while ALOHA leverages low-cost dual manipulators to record high-precision bimanual operation data [5]. Mobile ALOHA and TeleMoMa further expand research to full-body

We provide a video demo for this paper: https://linqi-ye.github.io/video/UnityFactory.mp4 coordinated manipulation, which requires simultaneous control of mobile bases and robotic arms without imposing excessive operational burden on users [6], [7]. Inspired by these studies, our platform supports synchronous control of base movement and two end-effectors, avoiding split independent control modes for box transportation tasks. Unlike existing research that mainly optimizes human-robot motion mapping, this work centers on building a complete task lifecycle to convert raw human operation into traceable, auditable trial records.

Another research direction focuses on matching human motion, robot morphology and visual feedback. AnyTeleop relies on vision-based motion retargeting to adapt to multi-arm and multi-scene requirements, covering both physical robots and virtual simulation environments [8]. Holo-Dex and Open-TeleVision adopt immersive interactive schemes to realize high-precision pose input paired with first-person stereoscopic visual feedback [9], [10]. UMI breaks the limitation of fixed workstations for data collection while reserving all motion information required for subsequent robot training [11]. This work provides complementary results with simplified controller mappings, prioritizing the synchronization of controllers, task state machines, visual feedback and auditable data logging for industrial simulation tasks.

B. Interactive Simulation and VR Demonstrations

Contemporary embodied simulation engines integrate high-fidelity rendering, physical engine calculation, multi-joint rigid bodies and standardized task logic. Habitat 2.0 realizes physically accurate mobile robot manipulation in interactive 3D virtual scenes [13]; iGibson 2.0 introduces object state management and a built-in VR module for collecting human operation demonstrations [12]; BEHAVIOR puts forward cross-platform universal task definitions and quantifies task efficiency based on 500 groups of VR operation samples [14]. Unity is also widely recognized as a flexible development tool for intelligent interactive agents [15]. All the above platforms prove that virtual simulation can effectively support human-in-the-loop robot manipulation experiments. Different from general universal simulators, this paper targets conveyor-driven factory box handling scenarios and takes auditable data collection as the core design goal; the expansion of multi-scene environments and cross-task benchmark testing are not within the scope of this research.

C. Traceable Data and Gamified Task Design

Simply expanding dataset volume cannot guarantee the interpretability of each single operation episode. Large-scale datasets including DROID, Open X-Embodiment and BridgeData V2 are committed to enriching manipulation scene diversity [1]–[3], and RLDS provides unified storage specifications for sequential episode-step data [16]. On this basis, we design a layered data structure that independently stores session configuration parameters, trial execution results, discrete task events and per-frame motion states. This hierarchical storage mechanism retains failed and interrupted trials completely, facilitating subsequent classification analysis of valid successes, task failures and invalid incomplete records.

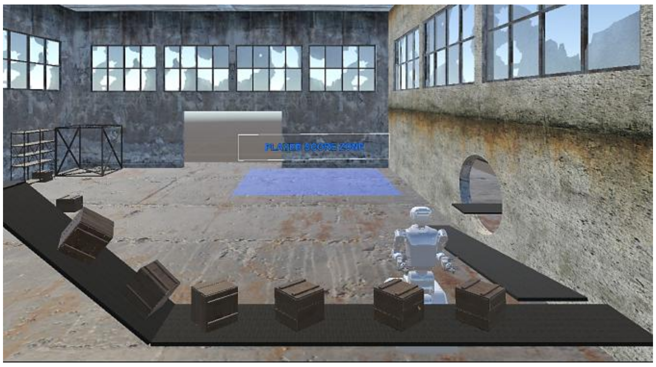
Game-based task design brings potential confounding factors for experimental analysis. Gamification refers to introducing game elements into non-game scenarios [17], yet various game mechanisms exert inconsistent psychological influences on operators. Relevant research confirms that scoring and level settings will change users' operation behaviors without improving their intrinsic task motivation [18], and different game elements satisfy distinct psychological demands of operators [19]. For this reason, we introduce countdown timers, scoring rules and level progression to construct two task execution modes. All experimental analysis in this paper focuses on objective task results and data traceability rather than subjective psychological motivation of operators. Compared with existing large-dataset research and gamification theory studies, our core innovation lies in data traceability: all task setup information is stored as contextual metadata, and every valid trial result is permanently associated with its full lifecycle event logs and original frame motion data.

III. SYSTEM DESIGN

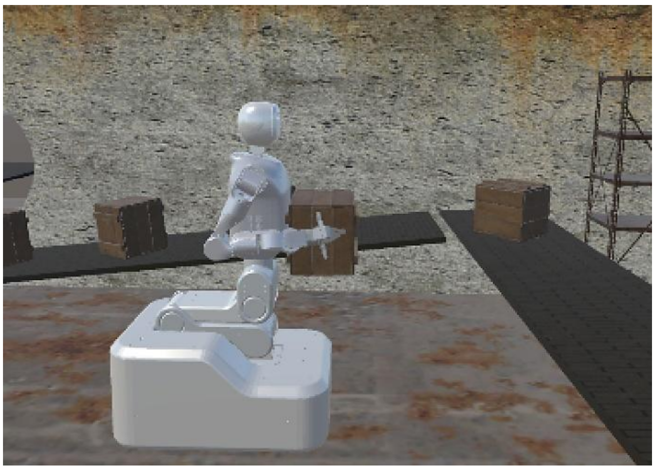
A. Design Objectives and System Architecture

The proposed framework is developed based on three core design requirements. First, the unified interaction interface supports all degrees of freedom for mobile bimanual box transfer, eliminating frequent control mode switching during operation. Second, the platform enables flexible configuration of task difficulty while maintaining consistent control logic and data recording specifications. Third, all trial-level task results support full backward traceability to the corresponding task events and frame-level motion sequences.

The entire simulation task environment is constructed in Unity, as visualized in Fig. 1. This customized factory workcell provides standardized physical environments and task constraints for subsequent human-in-the-loop teleoperation and embodied data collection.



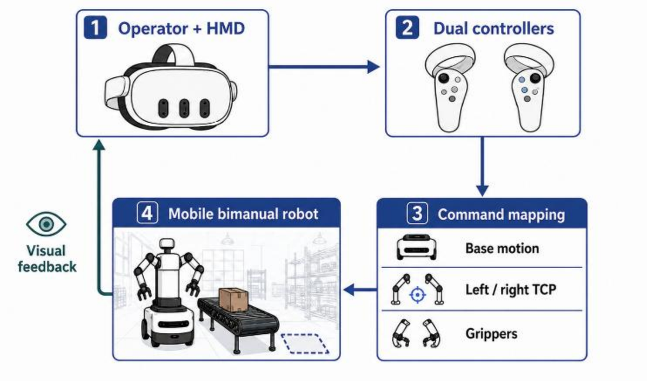
(a) Unity factory scene



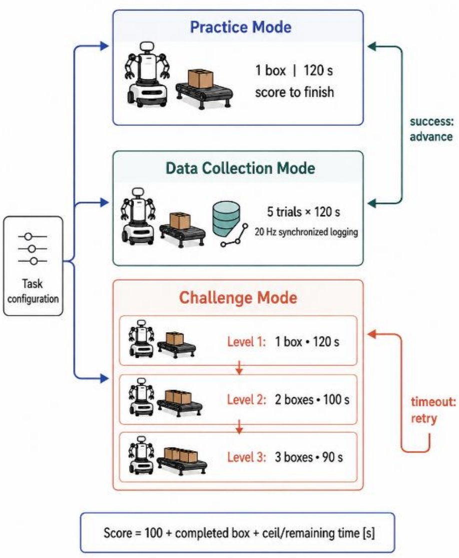
(b) Bimanual box-transfer interaction

Fig. 1. Unity task environment used for VR teleoperation.

As shown in Fig. 2, the system forms a complete closed-loop workflow covering interaction, simulation, feedback and recording. VR headset and dual-controller inputs are mapped to the mobile base motion, dual-arm TCP poses and gripper switching states, and the Unity simulation environment returns real-time visual rendering feedback to operators. Both the Practice Mode and progressive Challenge Mode are deployed to adjust task difficulty, while sharing a unified session–trial–event–frame hierarchical data pipeline. The task state machine continuously outputs real-time task feedback and termination signals, and the built-in shared recorder synchronously archives all motion and visual data streams to guarantee systematic and traceable demonstration collection.



(a) VR Teleoperation Closed Loop



(b) Progressive Task Design

Fig. 2. System overview of the VR teleoperation and task design

B. Factory Task and VR Control Loop

The virtual workcell consists of a conveyor feeding zone, dynamically spawned boxes, a mobile bimanual robot, and a delivery target. Conveyor-based box transfer is chosen for its similarity to real intralogistics workflows and well-defined success criteria. Compared with simple tabletop pick-and-place, this long-horizon task requires operators to wait for cargo, adjust the mobile base, perform bimanual grasping, transport goods and finish delivery. Moreover, it avoids assembly-related tolerance requirements that would interfere with pipeline evaluation, and the conveyor guarantees standardized task start timing.

Operators observe the Unity environment via a head-mounted display. Two tracked VR controllers define end-effector targets, and inverse kinematics computes joint motions for the simulated arms. Thumbsticks control planar base movement and yaw; grip buttons activate arm following; triggers actuate grippers. These parallel mappings support simultaneous base adjustment, bimanual grasping and transportation without mode switching.

The control loop is bounded by real-time task feedback instead of open-ended sandbox operation. Once a box arrives in the target area, the task manager updates completion status and scores. Successful scoring terminates Practice Mode, while timeout events are logged as valid failed trials. Manual program interruption is recorded independently, as it yields incomplete trajectories. This categorization supports downstream data filtering: timeouts constitute usable failure samples, whereas interrupted episodes are treated as incomplete records.

C. Progressive Task State Machine

Before entering the VR scene, the operator configures the task through the setup interface shown in Fig. 3. The interface provides Practice, Data Collection, and Gamified Transport options, where Gamified Transport corresponds to the progressive Challenge Mode described below. It also supports the configuration of recording and control parameters, including the sampling rate, grip threshold, base-motion thresholds, joint-rate threshold, campaign count, box score, required box numbers, and level time limits. The selected option and parameter values are stored as session-level metadata for subsequent trial annotation.

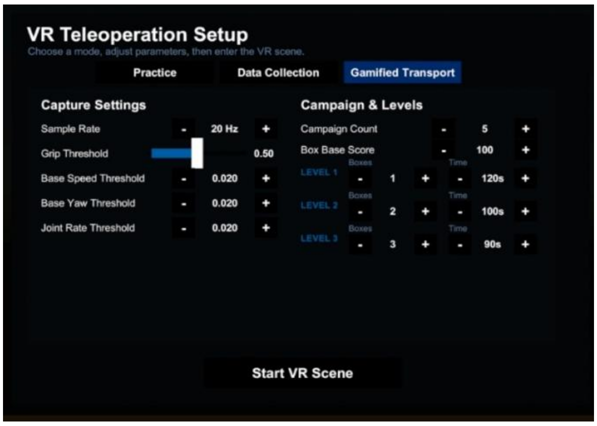


Fig. 3. VR teleoperation setup interface for task-mode selection and parameter configuration.

The platform implements two task modes with unified state-transition logic. Practice Mode consists of a single-box transportation task with a 120 s time limit and terminates immediately after successful delivery. It serves as a standardized tutorial and provides stable baseline conditions for demonstration collection. By contrast, Challenge Mode adopts a five-round campaign with progressive difficulty. Each round contains three levels requiring the transportation of one, two, and three boxes within 120 s, 100 s, and 90 s, respectively. Successful completion triggers level progression, whereas a timeout retains the current level for another attempt. The task score is calculated as 100 multiplied by the number of completed boxes plus the ceiling of the remaining time.

The progressive box counts and decreasing time limits are empirically selected to impose incremental task pressure rather than being analytically optimized. This monotonic increase in workload encourages repeated operation and increases motion-planning complexity without excessively extending task duration. These values therefore serve as configurable defaults and can be recalibrated for different tasks or operator groups.

Although task difficulty varies across modes, the underlying measurement and recording specifications remain unchanged. Practice and Challenge Modes share the same robot parameters, controller mappings, physical properties, task-event definitions, and data-storage schema. Mode labels and corresponding parameters are recorded as contextual metadata, ensuring unambiguous trial annotation while preserving a unified data structure across all collected demonstrations.

D. *Fault-Aware Logging and Data Hierarchy*

The system adopts a four-tier linked logging hierarchy for structured data recording. The session layer stores core configurations including data schema version, sampling rate, action thresholds, grip criteria and task parameters. The trial layer records task outcomes, elapsed time, sample quantity and derived evaluation metrics. The event layer captures key task transitions such as task start, scoring, timeout and manual interruption. The frame layer archives high-frequency raw data, including dual-controller states, mobile base poses and velocities, dual-TCP poses, 17-dimensional joint angles, box status, grasp signals and operational diagnostic information. Unique cross-layer identifiers enable full backward verification, allowing all statistical results to be reconstructed and validated from original frame data and event sequences.

This hierarchical structure avoids the loss of diagnostic information caused by simple outcome-based data filtering. Successfully scored trials can be excluded if their frame trajectories contain abnormal oscillations that degrade demonstration quality, while timeout-induced failed trials can be retained as valid negative samples despite irregular motions. Accordingly, the system decouples dataset inclusion status, task outcome type and data quality flags as independent attributes, achieving fine-grained and credible demonstration data screening.

IV. EXPERIMENTAL SETUP

A. *Data Collection Protocol*

Data collection was conducted under Practice Mode with a single operator performing repeated single-box transportation trials. Each trial was limited to 120 s. All trials retain their original session ordering to preserve traces of practice effects, fatigue and device variations, rather than aggregating records indiscriminately.

The operator is the system developer with prior familiarity with the task and interface. Accordingly, the measured success rate reflects performance of an experienced user under the given configuration; it cannot be generalized to novice operators or broader populations.

B. *Trial Filtering & Outcome Rules*

As shown in Fig. 4, the raw dataset contains 57 trials. Three trials terminated via ApplicationQuit (manual shutdown of the simulation during an ongoing trial) are discarded due to incomplete task outcomes. One nominally successful trial is excluded, as sustained arm oscillation renders the trajectory unsuitable for robot learning. The filtered dataset finally includes 53 valid trials: 37 successes and 16 timeouts.

Timeout trials with abnormal motions are retained as valid failure samples and marked in the audit log. This scheme strictly distinguishes incomplete recordings, low-quality successful demonstrations, and authentic task failures.

C. *Evaluation Metrics*

Instead of constructing a unified quality score, we adopt complementary metrics for data auditing: task success rate and completion time reflect collection efficiency; bimanual TCP midpoint path length quantifies motion redundancy; grip-gated joint second differences characterize local motion smoothness; the active-action ratio measures continuous command occupancy over trajectories. We emphasize that the active-action ratio does not represent the operator's psychological engagement.

For each frame t, let $\boldsymbol{p}_{L,t}$ and $\boldsymbol{p}_{R,t}$ denote left and right TCP positions. The bimanual midpoint is $\boldsymbol{m}_t = (\boldsymbol{p}_{L,t} + \boldsymbol{p}_{R,t})/2$, and path length accumulates $\|\boldsymbol{m}_t - \boldsymbol{m}_{t-1}\|_2$ across frames, covering both base and arm movements. Joint smoothness is defined using the mean absolute second-order difference, evaluated only within grip-active intervals.

Additional logging metrics monitor sampling frequency, maximum inter-frame gaps and numerical validity to verify recording robustness. All trial-level statistics can be recalculated from raw frame records for cross-checking.

D. *Analysis Boundary and Reproducibility*

The analysis relies on descriptive statistics and Wilson 95% binomial confidence intervals. Repeated trials are treated as multiple attempts from one operator, not independent human samples. Every metric is recomputed from frame-level CSV files and validated against trial summaries. The exported dataset documents inclusion criteria, exclusion reasons, quality flags and session identifiers to support full external auditing.

All reproducibility artifacts are packaged, including the fixed Unity 2022.3.62f3c1 environment, data schema, frame and event logs, summary files, analysis scripts and figure-generation code. Configuration parameters such as thresholds, joint ordering and metric definitions are frozen in the manifest. Researchers can reproduce data filtering, metric computation and visualization without running the Unity simulator.

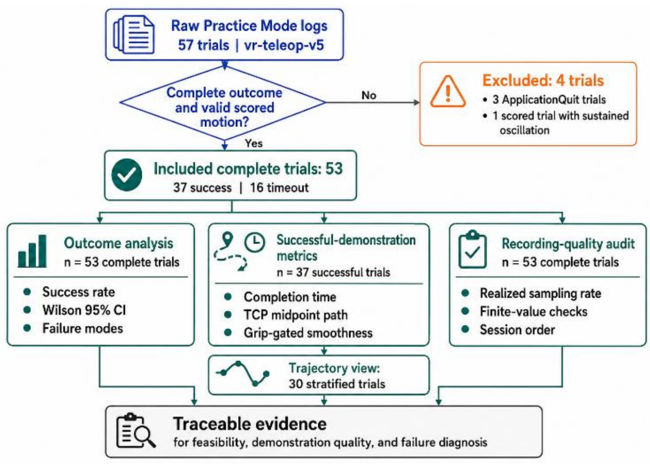


Fig. 4. Trial inclusion and analysis routing.

V. RESULTS

A. Trial Disposition and Task Outcome

Among the 53 valid trials, 37 achieved successful delivery and 16 ended in timeout, yielding a success rate of 69.81% (Wilson 95% CI: 56.46–80.48%). The relatively wide confidence interval indicates statistical uncertainty stemming from the limited number of repeated trials; this performance metric is not generalized to broader user populations. Instead, the outcome verifies that the integrated control logic, task state machine, termination rules and recording pipeline reliably generate both valid successful and failed demonstration samples under the experimental configuration.

B. Characteristics of Successful Demonstrations

Quantitative statistics for the 37 successful trials are summarized in Table I and Fig. 5: mean completion time of 76.06 ± 16.44 s, TCP-midpoint path length of 25.54 ± 6.98 m, grip-gated joint smoothness of 0.003445 ± 0.001041 rad/sample^2, and an active-action ratio of 57.65 ± 5.65%. The non-trivial standard deviations reflect substantial variation in operator motion strategies across repeated trials within the identical task setup.

Trial-level data reveal considerable variability even among successful attempts. These complementary metrics serve as diagnostic indicators to identify slow, overly circuitous, jerky, or frequently interrupted trajectories, rather than forming a unified composite score for demonstration quality.

TABLE I. SUCCESSFUL-TRIAL METRICS. VALUES ARE TRIAL-LEVEL DESCRIPTIVE STATISTICS; SD DENOTES SAMPLE STANDARD DEVIATION.

Measure	n	Mean ± SD	Median [IQR]
Completion time (s)	37	76.06 ± 16.44	69.09 [63.87, 88.73]
TCP path (m)	37	25.54 ± 6.98	23.95 [22.45, 25.99]
Smoothness (rad/sample^2)	37	0.003445 ± 0.001041	0.003211 [0.002640, 0.004034]
Active-action (%)	37	57.65 ± 5.65	58.29 [54.30, 60.70]

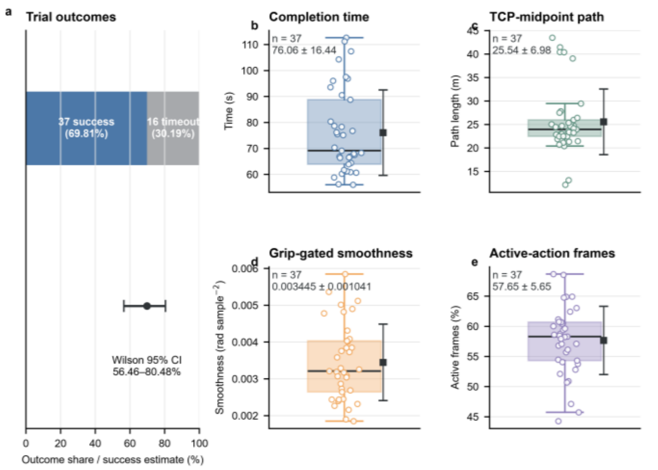


Fig. 5. Practice Mode experimental results. (A) Outcome count and success-rate Wilson 95% CI for all 53 complete trials. (B–E) Trial-level distributions for the 37 successful trials: completion time, dual-TCP midpoint path length, Grip-gated joint smoothness, and active-action ratio. Boxes show median and interquartile range; whiskers use the conventional 1.5-IQR rule; points are individual trials.

C. Trajectory Morphology

Figure 6 aligns successful TCP-midpoint trajectories by removing each trial's initial positional offset. These temporally ordered trials share a general transport direction but exhibit varying lateral and longitudinal paths. Endpoint spread and path-length data confirm the operator did not replicate an identical trajectory across successful runs. For this single-operator experiment, such dispersion arises from intra-operator behavioral variability, including alternative base placement and in-transit corrective motions. Trajectories are only spatially translated, without rotation or time normalization, preserving their original X–Z planar shape. Importantly, this variability differs from task coverage: true task coverage requires diverse objects, initial states and target layouts, rather than repeated attempts on an identical task configuration.

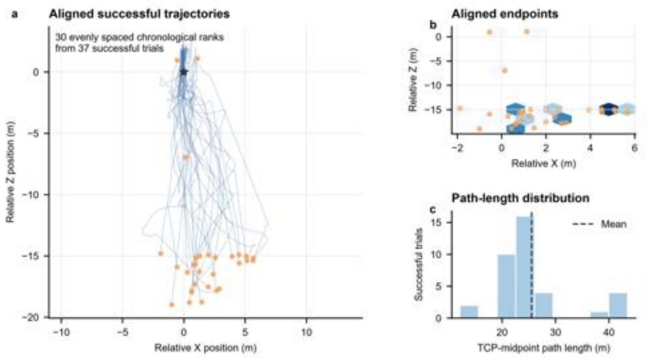


Fig. 6. Start-aligned X–Z projections for 30 temporally stratified successful trials (left) and their path-length distribution (right), characterizing within-condition execution variation.

D. Sampling Continuity and Anomaly Auditing

Over the 53 valid trials, the empirical sampling frequency reaches 19.316 ± 0.257 Hz, approaching the 20 Hz target. The median maximum inter-frame gap per trial is 0.146 s, with a mean value of 0.182 ± 0.155 s and the largest gap of 0.985 s. Although the average sampling rate closely matches the target, sporadic long sampling intervals exist. Average frequency alone cannot expose such transient discontinuities, which justifies reporting both realized sampling frequency and maximum frame gap jointly.

All frame datasets adopted for analysis conform to the predefined schema, contain monotonically increasing timestamps and valid finite values. Recalculated TCP path lengths, grip-gated smoothness and active-action indicators are consistent with trial-level aggregated results. The auditing tool further annotates the rejected oscillatory successful trial and preserved oscillatory timeout trials within their original session sequences. While these validations cannot guarantee the rationality of every robot pose, they render data inclusion rules and local recording anomalies fully reproducible.

VI. DISCUSSION

A. Outcomes of the Experimental Validation

This work demonstrates system-level engineering feasibility instead of comparative advantages against alternative methods. The unified simulation platform sustains immersive bimanual teleoperation, standardized task lifecycles and traceable data logging across repeated sessions, yielding 37 valid successful trials out of 53 complete attempts. The 16 timeout trials are retained as meaningful failure cases, while interrupted runs are discarded due to indeterminate task outcomes. This three-class categorization delivers richer information than simply discarding all unsuccessful attempts.

Decoupling task configurations, lifecycle events, frame trajectories and derived metrics establishes a reusable data collection framework. It facilitates iterative dataset building, benchmark development, comparative evaluation of teleoperation schemes, operator training, and quality-conscious imitation learning. Session manifests standardize experimental conditions, event logs pinpoint failure phases, and continuous metrics enable screening of inefficient, jerky or discontinuous motion traces. Although validated on the box-transfer scenario, this layered architecture can be adapted to other manipulation tasks equipped with defined triggers, event signals, termination logic and frame-level state recording.

B. Progressive Task Structure and Gamified Elements

Progressive task design functions as an interchangeable scheduling layer for repeated human data collection and is independent of the core logging pipeline. Practice Mode offers straightforward objectives and ample time to help operators adapt to the interface. Challenge Mode introduces rising task demands, time pressure, scoring and campaign progression to sustain continuous task engagement. Since both modes share identical control rules and data formats, task difficulty can be adjusted without breaking data consistency.

These gamified components exert conflicting influences. Stricter time limits may boost concentration yet trigger hasty adjustments, fatigue or frustration. Consistent with existing literature, behavioral performance and intrinsic motivation do not always correlate [18], and different gamification mechanisms lead to distinct user experiences [19]. For this reason, scores and task results are treated as objective experimental variables, not indirect measurements of subjective engagement.

C. Scope and Limitations

Experiments rely on a single experienced operator, multiple recording batches, in-session environment adjustments and Unity collision handling. Accordingly, results confirm reproducibility under the tested simulation setup, and cannot be generalized to broad user populations, prove superiority of the proposed control scheme, or guarantee performance on physical robots. Within this boundary, the data provenance mechanism and metric recalculation pipeline are fully validated by the collected trials.

VII. CONCLUSION

This work presents a VR teleoperation simulation platform equipped with a progressive state machine and fault-aware four-tiered logging hierarchy for collecting traceable bimanual mobile robot manipulation demonstrations. Experimental results verify that the system reliably generates both successful and timeout failure trials under standardized conditions, with fully auditable trajectories and consistent sampling performance. The layered data architecture separates task configuration, event records and frame-level motion data, supporting reusable dataset construction and trajectory quality inspection. Progressive Practice and Challenge Modes enable adjustable task difficulty without disrupting data schema consistency. The validation confirms the engineering feasibility of the pipeline, though tests are limited to one experienced operator. Future work will extend data collection to multiple participants and transfer the learned motion policies toward physical robotic platforms.

REFERENCES

[1] A. Khazatsky et al., "DROID: A Large-Scale In-the-Wild Robot Manipulation Dataset," in Proc. Robotics: Science and Systems, Delft, The Netherlands, 2024.

[2] Open X-Embodiment Collaboration et al., "Open X-Embodiment: Robotic Learning Datasets and RT-X Models," in Proc. IEEE Int. Conf. Robotics and Automation, 2024, pp. 6892–6903, doi: 10.1109/ICRA57147.2024.10611477.

[3] H. R. Walke et al., "BridgeData V2: A Dataset for Robot Learning at Scale," in Proc. 7th Conf. Robot Learning, PMLR, vol. 229, 2023, pp. 1723–1736.

[4] A. Mandlekar et al., "RoboTurk: A Crowdsourcing Platform for Robotic Skill Learning through Imitation," in Proc. Conf. Robot Learning, PMLR, vol. 87, 2018, pp. 879–893.

[5] T. Z. Zhao, V. Kumar, S. Levine, and C. Finn, "Learning Fine-Grained Bimanual Manipulation with Low-Cost Hardware," in Proc. Robotics: Science and Systems, 2023, doi: 10.15607/RSS.2023.XIX.016.

[6] Z. Fu, T. Z. Zhao, and C. Finn, "Mobile ALOHA: Learning Bimanual Mobile Manipulation with Low-Cost Whole-Body Teleoperation," in Proc. 8th Conf. Robot Learning, PMLR, vol. 270, 2024, pp. 4066–4083.

[7] S. Dass et al., "TeleMoMa: A Modular and Versatile Teleoperation System for Mobile Manipulation," arXiv:2403.07869, 2024.

[8] Y. Qin et al., "AnyTeleop: A General Vision-Based Dexterous Robot Arm-Hand Teleoperation System," in Proc. Robotics: Science and Systems, 2023.

[9] A. Arunachalam et al., "Holo-Dex: Teaching Dexterity with Immersive Mixed Reality," in Proc. IEEE Int. Conf. Robotics and Automation, 2023, pp. 5962–5969, doi: 10.1109/ICRA48891.2023.10160547.

[10] X. Cheng, J. Li, S. Yang, G. Yang, and X. Wang, "Open-TeleVision: Teleoperation with Immersive Active Visual Feedback," in Proc. 8th Conf. Robot Learning, 2024.

[11] C. Chi et al., "Universal Manipulation Interface: In-the-Wild Robot Teaching Without In-the-Wild Robots," in Proc. Robotics: Science and Systems, 2024, doi: 10.15607/RSS.2024.XX.045.

[12] C. Li et al., "iGibson 2.0: Object-Centric Simulation for Robot Learning of Everyday Household Tasks," in Proc. 5th Conf. Robot Learning, PMLR, vol. 164, 2022, pp. 455–465.

[13] A. Szot et al., "Habitat 2.0: Training Home Assistants to Rearrange their Habitat," in Advances in Neural Information Processing Systems 34, 2021.

[14] S. Srivastava et al., "BEHAVIOR: Benchmark for Everyday Household Activities in Virtual, Interactive, and Ecological Environments," in Proc. 5th Conf. Robot Learning, PMLR, vol. 164, 2022, pp. 477–490.

[15] A. Juliani et al., "Unity: A General Platform for Intelligent Agents," arXiv:1809.02627, 2018.

[16] S. Ramos et al., "RLDS: An Ecosystem to Generate, Share and Use Datasets in Reinforcement Learning," arXiv:2111.02767, 2021.

[17] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From Game Design Elements to Gamefulness: Defining Gamification," in Proc. 15th Int. Academic MindTrek Conf., 2011, pp. 9–15, doi: 10.1145/2181037.2181040.

[18] E. D. Mekler, F. Brühlmann, A. N. Tuch, and K. Opwis, "Towards Understanding the Effects of Individual Gamification Elements on Intrinsic Motivation and Performance," Comput. Human Behav., vol. 71, pp. 525–534, 2017, doi: 10.1016/j.chb.2015.08.048.

[19] M. Sailer, J. U. Hense, S. K. Mayr, and H. Mandl, "How Gamification Motivates: An Experimental Study of the Effects of Specific Game Design Elements on Psychological Need Satisfaction," Comput. Human Behav., vol. 69, pp. 371–380, 2017, doi: 10.1016/j.chb.2016.12.033.