



I Know Kung Fu: Synthetic Dexterous Hand Demonstration Collection via VR Teleoperation

Kara Lu, Yanzi He, Cohen Lu, Peihao Li

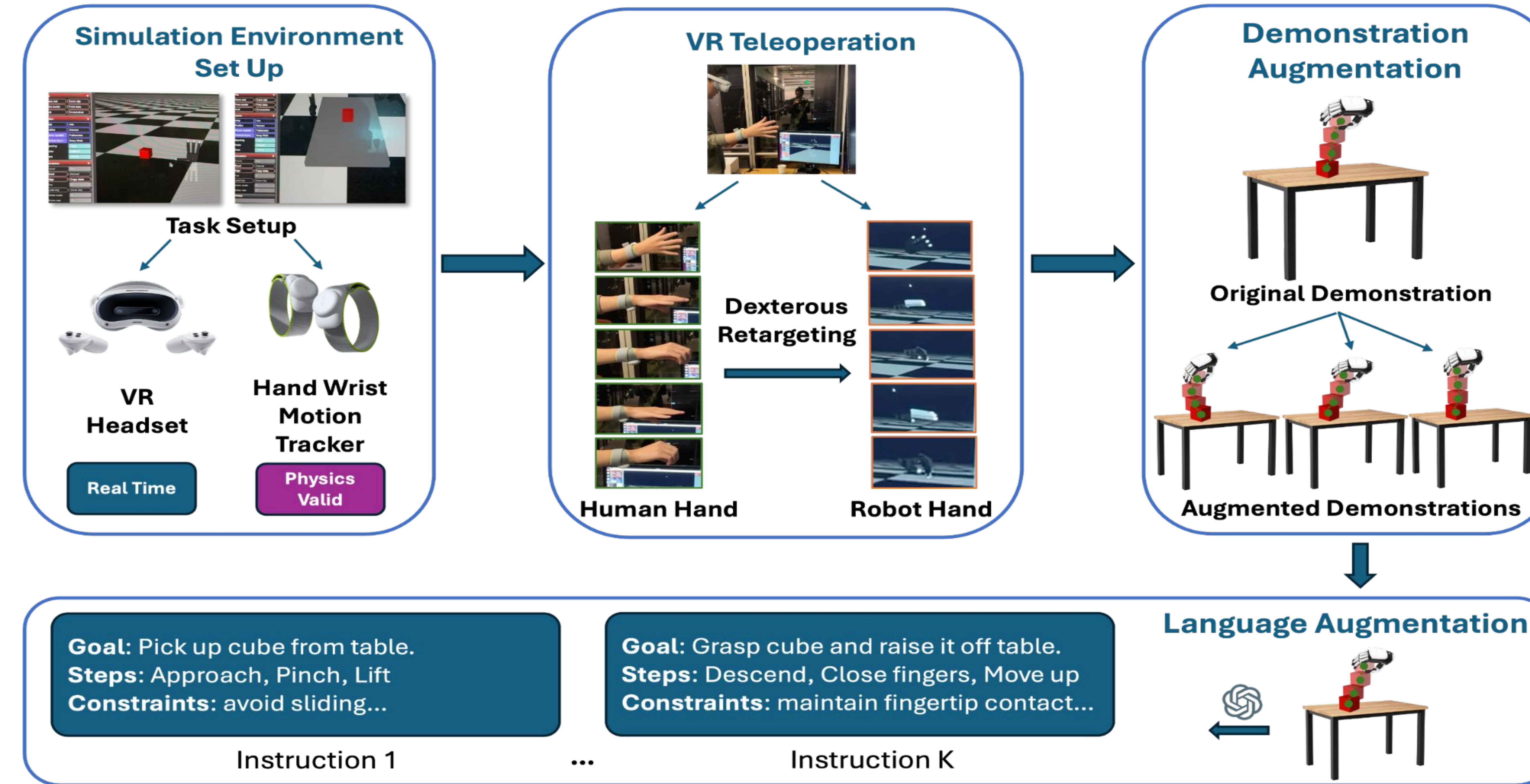


Abstract

In robotic dexterous manipulation learning, limitations in the acquisition of large and scalable amounts of high-quality demonstration data have been a **critical bottleneck**. Although recent works show significant advances in increasing efficiency through the use of VR technology, human video, or synthetic demonstration generation techniques, these methods have **limitations in accuracy and effectiveness**. In this project, we propose a data collection pipeline that uses a **VR-teleoperation system** to track human hand motions in order to collect dexterous hand demonstrations in simulation. Our method utilizes a **wrist motion tracker** for hand orientation tracking, followed by a **dexterous retargeting module** to sync human movement with robot movement in real time. We will also implement **demonstration augmentation**, ultimately yielding multiple distinct, successful trajectories across different scenarios from a single human demonstration.

Advantages

Our method offers several key advantages over traditional teleoperation and LLM-based planning approaches. By **using VR simulation**, it enables **scalable and low-cost data collection** without requiring physical robot hardware, while still ensuring demonstrations remain physically plausible through simulated physics and realistic hand-object interactions. The system naturally captures universal human hand motions, allowing the demonstrations to **generalize across different robotic hand embodiments** and avoiding the **need for custom teleoperation rigs**. Compared to conventional tools like gloves or gamepads, VR provides far more **intuitive control** for multi-fingered dexterous hands, making it possible to collect high-quality trajectories even for complex manipulation tasks. Unlike LLM-only planning methods, which often hallucinate actions and lack concrete low-level control signals, **our approach produces grounded, consistent, and physically validated motion data** thereby improving the reliability and real-world transferability of the collected demonstrations.



Methodology

Our system has three modules: (i) **hand-wrist pose tracking**, (ii) **dexterous retargeting from finger motion**, and (iii) **physics-valid demonstration augmentation**. Together they enable real-time VR teleoperation that produces **contact-aware trajectories** and a **scalable data engine** for imitation learning.

A 6-DoF tracker provides $(p_{\text{pico}}, q_{\text{pico}})$ in frame F_{pico} . A calibration rotation $R_{\text{pico_to_sim}}$ (with quaternion q_R) maps these into simulation:

$$p_{\text{sim}} = R_{\text{pico_to_sim}} * p_{\text{pico}}$$

$$q_{\text{sim}} = q_R * q_{\text{pico}} * \text{inverse}(q_R)$$
 We apply $T_{\text{hand}} = (p_{\text{sim}}, q_{\text{sim}})$ each step, with clamped motion for stability. Tracker/simulator rates are aligned by interpolation, with optional smoothing to reduce jitter. The HMD head pose drives an egocentric camera. We log both the applied pose and synchronized tracker data.

XR-Robotics SDK landmarks are normalized to a MediaPipe-style skeleton and mapped to joint targets q_{finger} . A linear model $q_{\text{finger}} = \text{Clamp}(W * \phi(\text{landmarks}) + b, q_{\text{min}}, q_{\text{max}})$ is fitted from short calibration motions. Joint synergies handle coupled fingers, and low-pass filtering plus velocity limits stabilize contact. Each step we send $(T_{\text{hand}}, q_{\text{finger}})$ to the simulator and log all signals, including object pose and camera parameters, for downstream learning.

Discussion & Future Works

Motivation and Design Rationale

We pursue scalable dexterous manipulation data through VR teleoperation, capturing fine-grained hand motion and mapping it in real time to simulator trajectories. To ensure **cross-embodiment portability**, we maintain a **canonical hand parameterization** with learned per-robot retargeting. By decoupling task logic from physics and I/O, RoboVerse enables **seamless cross-VR and cross-simulator reuse**, yielding contact-rich seeds ready for augmentation and policy learning.

Toward Scalable Augmentation

A DemoGen-style module expands seeds without extra teleoperation by decomposing trajectories into contact and free-space segments; contact portions are transformed to preserve **interaction frames**, while **free-space motions** are replanned to maintain feasible kinematics. Visual observations are synthesized through point-cloud editing to align actions with novel object configurations, retaining only variants that satisfy task predicates.

Experimental Plan

We evaluate across three axes: (i) **Cross-VR reproducibility**, measuring fingertip error, contact sequence alignment, and demo throughput across headsets/trackers; (ii) **Cross-simulator portability**, comparing terminal object error, contact-event alignment, and success rates when replaying seeds across engines; (iii) **Augmentation scaling**, analyzing success versus dataset size and reset coverage while ablating pose, instance, viewpoint, and disturbance sampling to compare policies trained on raw versus augmented seeds.

Limitations and Open Problems

VR noise can introduce **micro-oscillations** despite smoothing, and retargeting may distort **contact geometry** when kinematics differ substantially. Cross-simulator replay can also shift **contact timing**, while overly strict augmentation risks biasing toward easy configurations. We address these through **coverage heatmaps** and **tuned sampling**. Despite these caveats, VR-in-sim seeds with DemoGen-style expansion provide a principled path toward scalable dexterous datasets.