DexHiL: A Human-in-the-Loop Framework for Vision-Language-Action Model Post-Training in Dexterous Manipulation

Yifan Han*,[1] Zhongxi Chen*,[2] Yuxuan Zhao[2] Congsheng Xu[2]
Yanming Shao[3] Yichuan Peng[2] Yao Mu[2,†] Wenzhao Lian[2,†]

[1]CASIA [2]SJTU [3]Shanghai AI Laboratory
*Equal contribution [†]Corresponding author

***Abstract*— While Vision-Language-Action (VLA) model has demonstrated promising generalization capabilities in robotic manipulation, deploying them on specific and complex downstream tasks still demands effective post-training. In parallel, Human-in-the-Loop (HiL) learning has proven to be a powerful mechanism for refining robot policies. However, extending this paradigm to dexterous manipulation remains challenging: multi-finger control is high-dimensional, contact-intensive, and exhibits execution distributions that differ markedly from arm motion, leaving existing dexterous VLA systems limited in reliability and adaptability. We present DexHiL, the first integrated arm–hand human-in-the-loop framework for dexterous manipulation VLA , enabling coordinated interventions over the arm and the dexterous hand within a single system. DexHiL introduces an intervention-aware data sampling strategy that prioritizes corrective segments for post-training, together with a lightweight teleoperation interface that supports instant human corrections during execution. Real-robot experiments demonstrate that DexHiL serves as an effective post-training framework, yielding a substantial performance leap that outperforming standard offline-only finetuning baselines by an average of 25% in success rates across distinct tasks. Project page: https://chenzhongxi-sjtu.github.io/dexhil/**

I. Introduction

Vision-Language-Action (VLA) models represent a paradigm shift in robot learning, enabling cross-scenario semantic understanding and spatial perception by directly mapping multimodal inputs to control outputs [1], [2], [3], [4], [5]. While VLA models have demonstrated significant potential in general manipulation, their application to high-degree-of-freedom (DOF) dexterous hands remains challenging, particularly in the post-training and downstream adaptation phases. Existing VLA post-training strategies, which typically rely on Supervised Fine-Tuning (SFT) over offline datasets [3], [6], [7], struggle to bridge the gap between high-dimensional end-effector control and the intricate, contact-rich requirements of multi-fingered manipulation.

The primary bottleneck in adapting VLA models to dexterous tasks partially stems from hardware-level kinematic misalignment. Traditional teleoperation interfaces, such as exoskeletons and leader-follower arms [8], [9], [10], often fail to precisely map human hand motions to the complex joint configurations of robotic hands. This misalignment, coupled with the sharp discontinuities inherent in dexterous contact states, precludes the collection of high-quality, fine-grained demonstration data. Consequently, existing VLA paradigms lack the high-fidelity guidance necessary to optimize the nuanced motion manifold required for robust dexterous performance. Beyond hardware limitations, the algorithmic post-training process for dexterous VLAs faces three core systemic challenges: (i) Convergence difficulties in high-dimensional action spaces: The expansive action manifold of dexterous hands, compounded by complex contact dynamics, makes stable policy convergence exceptionally difficult; (ii) Sample efficiency bottlenecks: Offline datasets are frequently dominated by repetitive success data, forcing

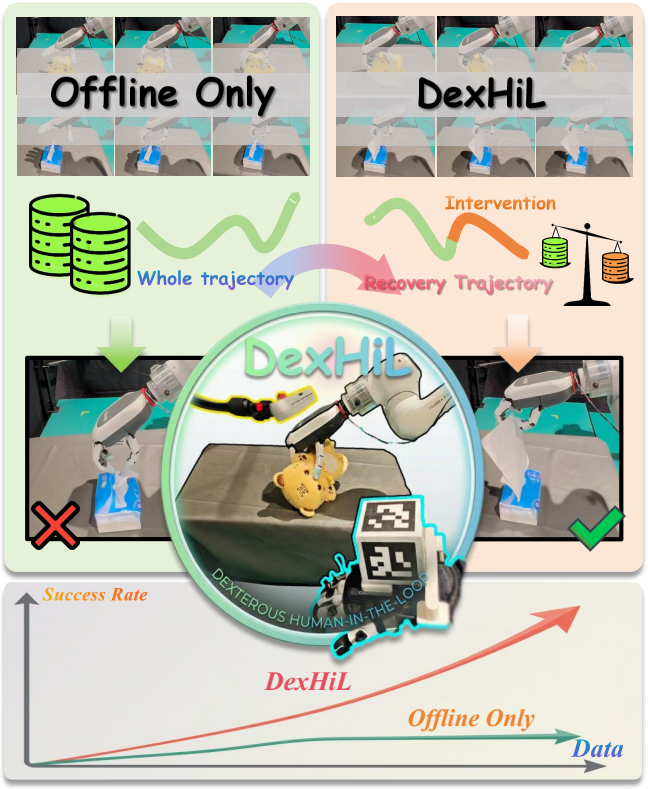

Fig. 1: While scaling offline data for VLA models yields slow accuracy gains and performance plateaus, DexHiL integrates offline training with online Human-in-the-Loop interventions. By strategically reweighting offline and corrective online data, our approach achieves high data efficiency and rapid accuracy growth.

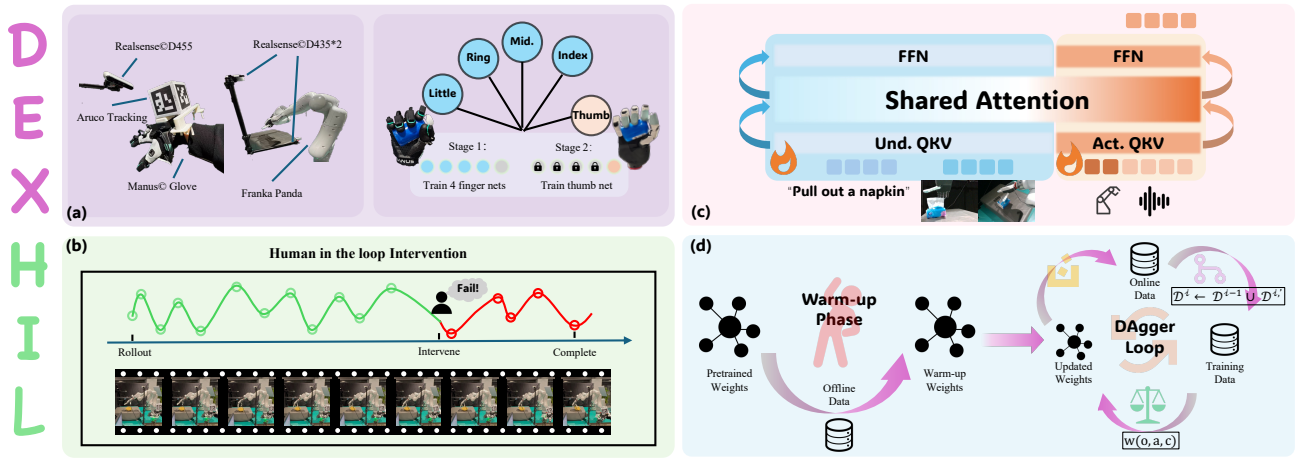

Fig. 2: **The DexHiL Framework.** Below we will introduce our overall framework of DexHiL from data acquisition system, human-in-the-loop intervention paradigm, dexterous manipulation VLA model structure to offline-to-online training process. (a) We propose an arm-hand data collection system both supporting teleoperation offline data collection and online human-in-the-loop policy intervention data collection. We also propose a two-stage training method for precise hand joint retargeting. (b) We propose an asynchronous human-in-the-loop policy intervention mechanism for online data collection, here we examplify the "Plush Toy Grasping" case. (c) Our dexterous manipulation VLA policy follows Being-H0.5 [11] structure, which utilizing MoT (Mixture of Transformer) to relate understanding model with action expert for multi-modal reasoning and action generation and inherit the open-source pretrained weights of it. (d) We propose a two-phase training framework that ultilize both offline dataset and online dataset. In the first warm-up phase, we finetuned the pretrained weights into warm-up model. In the DAgger loop, we utilize the system above to acquire online dataset and use reweighting training to update the policy which will be applied in the next DAgger loop.

the model to expend gradients on consolidating learned behaviors rather than exploring the critical transitions necessary for complex tasks; (iii) Covariate shift and error accumulation: During real-robot execution, open-loop VLA policies suffer from trajectory drift. Without an effective recovery mechanism, minor errors quickly push the system into Out-of-Distribution (OOD) states, leading to failure [12].

These challenges collectively suggest that relying solely on conventional offline post-training is insufficient for robust dexterous manipulation. To bridge this gap, we propose DexHiL, an interactive online training framework that integrates dexterous robotic hands with Human-in-the-Loop (HiL) intervention [13], [14]. DexHiL leverages a human-robot collaboration mechanism to facilitate precise manipulation and rapid error recovery in complex, contact-rich tasks. By incorporating real-time expert interventions and an intervention-aware weighting mechanism, DexHiL significantly enhances sample efficiency and policy convergence, ensuring robust performance in high-DOF real-world applications.

Our experimental results demonstrate that for a VLA model pre-trained on large-scale human videos with a cold-start action head, jointly fine-tuning on multiple downstream tasks using the DexHiL framework yields significant performance gains. Through three iterative stages of online refinement, DexHiL achieves 20% and 30% improvements in success rates across two distinct tasks, respectively, compared to a baseline model trained offline with an equivalent volume of data. Ablation studies further validate the necessity of our core components, demonstrating that the intervention-aware weighting mechanism is the primary driver for overcoming sample efficiency bottlenecks in high-DOF manipulation.

Our contributions are summarized as follows:

- **Human-to-Robot Hand Movement Retargeting**: We introduce a novel learning-based approach for precise retargeting of human hand movements to dexterous robotic hands. This method effectively maps human hand gestures to the robotic manipulator with high fidelity, ensuring accurate control over complex dexterous hand movements. Unlike traditional optimization methods, our approach provides a more adaptive, real-time mapping, improving the performance of dexterous manipulation tasks.
- **Integrated and HiL-Enabled Teleoperation System**: We propose a seamless human-in-the-loop Teleoperation System that addresses the challenge of intervention discontinuity in high-DOF dexterous hand manipulation. By enabling smooth real-time interventions during post-training, our framework ensures effective and high-quality error correction.
- **Iterative Human-in-the-Loop Post-Training for VLA**: We propose DexHiL, a post-training pipeline for dexterous VLA models that introduces a novel intervention-aware data sampling strategy. This strategy prioritizes corrective segments from expert interventions, ensuring that the most valuable data for error recovery and task refinement is efficiently used.

By dynamically re-weighting these corrective samples, DexHiL accelerates convergence, improves sample efficiency, and enhances model performance, particularly in high-DOF, contact-intensive tasks.

II. Related work

A. Dexterous Manipulation Data Collection System

Dexterous manipulation data collection systems typically employ two approaches. The first method directly collects human hand data and generates dexterous hand data in a simulation environment using optimization methods [15] or reinforcement learning [16], [17], [18]. The other relies on teleoperation with retargeting algorithms, typically utilizing exoskeletons or master-slave arms [19], [20]. However, both approaches face limitations when applied to Human-in-the-Loop systems. Exoskeletons or master-slave arms do not align precisely with the current dexterous hand's robotic arm, preventing the intervention process from being made in an incremental form. Additionally, teleoperation methods often rely on optimization [21], [22], [23] or network fitting [24], [25], [21], with the former struggling to control finger movement precisely, especially in grasping tasks, and the latter typically focusing on the thumb, leading to poor correspondence with the remaining fingers. Our framework addresses these issues by providing more accurate data collection and dexterous manipulation solutions.

B. Vision Language Action model for dexterous manipulation

Recent advancements in dexterous manipulation, exemplified by various frameworks, highlight the transformative potential of VLA models [11], [26] and Vision-Language Grasp(VLG) models [27], [28]. However, these approaches typically require the collection of offline real-world datasets, followed by fine-tuning to adapt to downstream robots and tasks. In the high-dimensional action spaces inherent to dexterous hands, such offline post-training strategies often encounter significant distribution shifts and suffer from low sample efficiency [29]. These challenges make achieving convergence, particularly for contact-intensive tasks, especially difficult. To overcome these limitations, we propose an online post-training framework that mitigates these issues.

C. Human-in-the-Loop Corrections for Robot Learning

Interactive correction learning aims to rectify robotic behaviors through real-time human intervention [12], [29], [30]. Interactive correction learning refines robot behaviors through real-time human intervention [12], [13], [31], [14]. HG-DAgger [13] and its subsequent work, Sirius [14], incorporate human corrections during critical moments to enhance policy efficiency [32], [33]. While these human-in-the-loop data aggregation paradigms have proven effective, they are predominantly confined to parallel gripper-arm setups and have not been successfully extended to dexterous manipulation [33], [32], [34]. A recent work closely related to ours is DexGrasp-VLA [34], which applies human-in-the-loop correction only to the robotic arm. However, it does not integrate HIi for the hand, leaving hand movements controlled by a separate grasping network. This lack of coordination between the arm and hand limits the effectiveness of the data, especially for contact-rich tasks.

III. Method

We present DexHiL, a framework designed to tackle the complexities of high-dimensional control and data inefficiency in dexterous VLA. The methodology comprises two synergistic components: (1) an Interactive Human-in-the-Loop Teleoperation System for Dexterous Manipulation, which provides a unified interface for seamless arm-hand coordination and real-time intervention; and (2) a Human-in-the-Loop Post-training Pipeline, which leverages an intervention-aware weighting mechanism to accelerate policy convergence and mitigate covariate shift during real-world execution. Together, these components enable the model to efficiently learn robust error-recovery behaviors in contact-rich scenarios.

A. Interactive Human-in-the-Loop Teleoperation System for Dexterous Manipulation

To support high-fidelity offline data collection and intuitive human-in-the-loop operation, we implement a lightweight arm–hand teleoperation interface. A handheld ArUco marker cube is tracked by a monocular camera to recover its 6D pose in real time, providing a low-overhead and fast-to-deploy solution that remains robust across setups. To account for the kinematic mismatch between the human operator and the robot, we use a **dual-path mapping**: one path aligns end-effector pose and global arm motion, while the other retargets finger articulations to the dexterous hand, jointly maintaining trajectory consistency and gesture-level controllability.

1) Hand Joint Retargeting: We propose a unified hand-joint retargeting pipeline for dexterous hands with coupled joints. Human keypoints $\mathbf{X}_{\text{keypoints}}$ are captured by a motion-capture glove, and the retargeting network f_θ maps them to the robot's actuated joint angles $\mathbf{X}_{\text{act}} \in \mathbb{R}^m$. The full joint configuration $\mathbf{q}$ is obtained by applying the hand's predefined coupling constraints to $\mathbf{X}_{\text{act}}$, and the robot fingertip positions are computed by forward kinematics as $\mathbf{P}_{\text{tip}} = \text{FK}(\mathbf{q})$.

To avoid the degenerate behavior observed in single-network five-finger retargeting, where the learned grasps collapse to pinch-like postures, we adopt a two-stage scheme.Empirically, optimizing all five fingers in a single network degrades the solution space of the four non-thumb fingers: the learned behaviors tend to concentrate on fingertip-to-thumb contacts, yielding stable opposition while the resulting "grasps" become pinch-like and fail to preserve enveloping grasp capability. In the first stage, we optimize only the index, middle, ring, and little fingers to obtain a stable and complete four-finger motion manifold.Supervision is imposed through intra-finger geometric constraints, which capture both articulation direction and extension length without relying on global five-finger couplings. Specifically, for

each $i \in \mathcal{I}$, let $\mathbf{r}_i^H, \mathbf{r}_i^R(\mathbf{q}) \in \mathbb{R}^3$ denote the human/robot root-to-tip vectors, with $d_i = \|\mathbf{r}_i^H\|_2$ and $\hat{\mathbf{r}}_i^H = \mathbf{r}_i^H/(d_i + \epsilon)$. We train with

$$\mathcal{L}_{\text{vec}} = \frac{1}{2} \sum_{i \in \mathcal{I}} s(d_i) \left\| \mathbf{r}_i^R(\mathbf{q}) - f(d_i)\, \hat{\mathbf{r}}_i^H \right\|_2^2 + \gamma \|\mathbf{q}\|_2^2, \quad (1)$$

where $s(\cdot)$ and $f(\cdot)$ are distance-dependent weight and scale functions.

After training the non-thumb fingers, we freeze their retargeting parameters and optimize only a thumb residual mapping. This stage aims to preserve the mapping from the human fingertip C-space to the robot fingertip C-space. Following the objective design of GeoRT [24], we optimize the thumb residual using a concise set of geometric regularizers, including the motion-preservation loss $\mathcal{L}_{\text{dir}}$, workspace-coverage loss $\mathcal{L}_{\text{cover}}$, flatness loss $\mathcal{L}_{\text{flat}}$, and pinch-preservation loss $\mathcal{L}_{\text{pinch}}$.We further introduce an inter-fingertip kinematic term $\mathcal{L}_{\text{kin}}$, which applies Eq. (1) to a predefined set of five-finger fingertip-pair vectors (thumb–finger and finger–finger) to match human and robot inter-finger geometry.We train the model with the following combined objective:

$$\begin{aligned} \mathcal{L} = \lambda_{\text{dir}}\mathcal{L}_{\text{dir}} + \lambda_{\text{cover}}\mathcal{L}_{\text{cover}} + \lambda_{\text{flat}}\mathcal{L}_{\text{flat}} \\ + \lambda_{\text{pinch}}\mathcal{L}_{\text{pinch}} + \lambda_{\text{kin}}\mathcal{L}_{\text{kin}}. \end{aligned} \quad (2)$$

2) Arm Pose Mapping: Let $\mathbf{T}_M \in SE(3)$ denote the 6-DoF pose of the ArUco cube expressed in the camera frame. Assuming the constant transformation matrix from the cube frame to the robot's end effector (EE) frame is known and indicated as $\mathbf{T}_{\text{robot}}^{\text{cube}}$. During policy rollout, the human operator requires a key press on the keyboard to trigger the mechanism of **human intervention**. Upon activation, we will record the current robot EE pose $\mathbf{T}_{EE_0}$ and ArUco cube pose $\mathbf{T}_{M_0}$ as **anchor poses**. To enable a smooth and intuitive takeover, we map the cube's subsequent motion to the robot by applying the cube's relative pose (w.r.t. the anchor) to the EE anchor pose. Hence, the target pose of the robot's EE in the base frame, $\mathbf{T}_{EE}$, is computed as:

$$\mathbf{T}_{EE} = \mathbf{T}_{EE_0} (\mathbf{T}_{\text{robot}}^{\text{cube}})^{-1} \mathbf{T}_{M_0}^{-1} \mathbf{T}_M \mathbf{T}_{\text{robot}}^{\text{cube}} \quad (3)$$

The corresponding arm joint angles $\mathbf{q}_{arm} \in \mathbb{R}^n$ are then derived via inverse kinematics (IK):

$$\mathbf{q}_{arm} = \mathcal{K}^{-1}(\mathbf{T}_{EE}) \quad (4)$$

This marker-based technique provides high robustness to ambient lighting and efficient detection even in cluttered environments

3) Asynchronous Multi-threaded Control and Intervention: We also designed a multi-threaded architecture to handle autonomous execution and human intervention concurrently. The autonomous policy π performs inference at 20Hz, while the human-guided arm and hand teleoperation operate at 30Hz and 90Hz, respectively. The human operator monitors the robot's performance and takes over the execution process of the entire robot system upon detecting imminent task failure.

The resulting dataset is formally represented as:

$$\mathcal{D} = \{(\mathbf{o}_t, \mathbf{q}_{arm,t}, \mathbf{q}_{hand,t}, I_t)\}_{t=1}^T \quad (5)$$

where $I_t \in \{0, 1\}$ is a binary indicator. The control law u_t at time t can be described as:

$$u_t = \begin{cases} \pi(\mathbf{o}_t), & \text{if } I_t = 0 \text{ (Autonomous)} \\ u_{human}, & \text{if } I_t = 1 \text{ (Intervention)} \end{cases} \quad (6)$$

where u_{human} is the command derived from the mapping described in Eqs. (1)-(3).

B. Human-in-the-Loop Post-training pipeline

We propose a Human-in-the-Loop interactive post-training pipeline for VLA models, built upon the teleoperation algorithm and hardware system described in Sec. III-A (Fig. 2). The framework incorporates a policy refinement mechanism and a intervention-aware weighting training strategy that combines both online human-in-the-loop data and offline datasets, enhancing the ability of Human-in-the-Loop to improve VLA models for dexterous manipulation tasks.

1) Intervention-aware Weighting Mechanism: Prior work has demonstrated that human intervention data is crucial for robot learning [14], enabling models to learn error avoidance and recovery from sub-optimal actions. However, such data is often sparse compared to offline datasets. To improve the utilization of these high-value samples, we use a intervention-aware weighting mechanism, which incorporates the weighting term $w(o, a, c)$ into our policy update framework.

Let the aggregated dataset $\mathcal{D}$ contain N samples, where n_c is the number of samples in category c. The original empirical distribution is defined as $P(c) = n_c/N$. The intervention-aware weighting mechanism adjusts this to a target distribution $P^*(c)$ with an increased proportion of interventions. Based on the principle of importance sampling, the sample weight $w(o, a, c)$ for category c is:

$$w(o, a, c) = \frac{P^*(c)}{P(c)} \quad (7)$$

To enable the model to more effectively master contact-rich maneuvers that are otherwise prone to failure, we specifically re-weight the sparse intervention trajectories. In our implementation, we specify $P^*(\text{intervention}) = 0.5$, an empirical equilibrium that further functions as a regularization mechanism, bridging the distribution shift without compromising foundational capabilities. Ultimately, this approach leads to enhanced deployment robustness and superior sample efficiency.

2) Policy Update Mechanism: The policy refinement process is structured into three integrated stages: warm-up initialization, online interactive learning, and strategy-driven data filtering.

a) Warm-up Phase: We begin with a warm-up phase, where we collect an initial offline dataset $\mathcal{D}^0$ via our teleoperation system. The VLA model undergoes full fine-tuning on $\mathcal{D}^0$ to derive the initial robotic policy π_0:

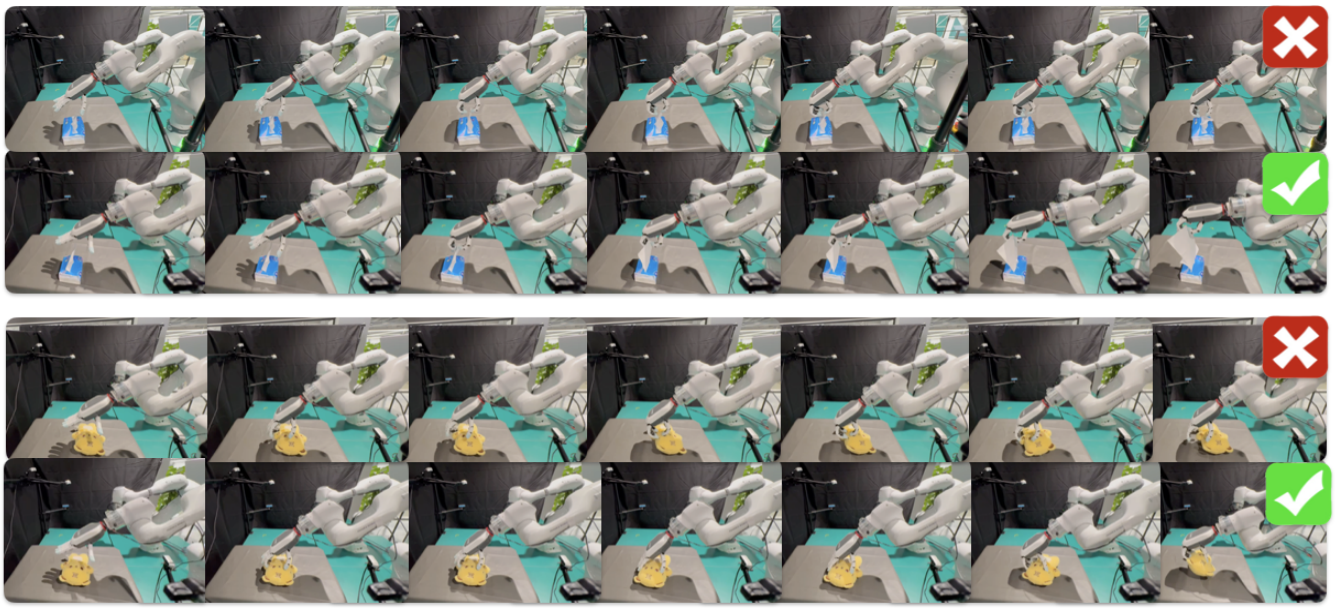

Fig. 3: **Real-world rollouts of dexterous manipulation tasks.** (Up) **Tissue Extraction:** The system achieves precise fingertip alignment and vertical retraction to extract the tissue. (Down) **Plush Toy Grasping:** The controller executes a synchronized multi-joint flexion to securely envelop and lift the deformable object.

$$\theta^{(0)} = \arg\min_{\theta} \mathcal{L}(\theta; \mathcal{D}^0), \qquad \pi_0 = \pi_{\theta^{(0)}}. \tag{8}$$

where $\theta^{(0)} = (\theta_{\text{VLM}}, \theta_{\text{A}})$ denotes the whole parameters comprised with VLM and action head parameters, warmed-up with $\mathcal{D}^0$.

b) Online Training Phase: Upon obtaining π_0, the system enters an online learning loop for n iterations. During each iteration $i \in \{1, 2, \ldots, n\}$, the current policy is deployed on the physical robot under human supervision. When an imminent failure is detected, the human intervenes to provide corrective demonstrations, yielding an online dataset $\mathcal{D}^{i,\prime}$. Following dataset aggregation, we form

$$\mathcal{D}^i \leftarrow \mathcal{D}^{i-1} \cup \mathcal{D}^{i,\prime}, \quad i = 1, 2, \ldots, n \tag{9}$$

We then update the policy by warm-starting from the previous parameters $\theta^{(i-1)}$ and performing supervised post-training on $\mathcal{D}^i$. Concretely, we optimize a weighted imitation objective:

$$\mathcal{L}^{(i)}(\theta; \mathcal{D}^i) = \mathbb{E}_{(o,a,c)\sim\mathcal{D}^i}\Big[w(o, a, c) \cdot \ell_{\text{IL}}(\theta; o, a)\Big], \tag{10}$$

where $w(o, a, c)$ is instantiated by our intervention-aware weighting mechanism to emphasize intervention samples, and ℓ_{IL} denotes the per-sample imitation loss.

In this paper, we instantiate DexHiL with Being-H0.5 [11] as a representative VLA model; since its action head follows a Flow Matching (FM) formulation, we use the FM objective for ℓ_{IL}:

$$\ell_{\text{IL}}(\theta; o, a) = \mathbb{E}_{t,x_t}\left[\|v_\theta(x_t, t, o) - u_t(a \mid x_0)\|_2^2\right], \tag{11}$$

where o represents the multimodal observation, a is the expert action, and $x_t = (1-t)x_0 + ta$ is the probability path interpolated between Gaussian noise x_0 and the target action a. The model v_θ learns to predict the ideal velocity field $u_t(a|x_0) = a - x_0$. The integration of sample weights $w(o, a, c)$ ensures that the VLA model prioritizes gradients derived from high-value intervention data.

The parameters are updated via stochastic gradient descent:

$$\theta^{(i)} \leftarrow \theta^{(i-1)} - \eta \nabla_\theta \mathcal{L}^{(i)}(\theta; \mathcal{D}^i), \qquad \pi_i = \pi_{\theta^{(i)}}, \tag{12}$$

with learning rate η. This iterative human-in-the-loop refinement progressively improves the policy's robustness, especially in failure-prone contact-rich manipulation scenarios.

c) Data Filtering Strategy: To enhance online optimization, we implement a targeted data filtering scheme. For each trial involving human interventions, we record data according to (5), but preserve only the segments from the final intervention to task completion, while discarding all trajectory segments prior to the last takeover.

The rationale behind this filtering is that trajectories with multiple interventions often exhibit significant action incoherence. Furthermore, the model actions preceding the intervention are inherently sub-optimal or erroneous [14]. Training on such inconsistent data can lead to Policy Oscillation or Multimodal Distribution Conflict. By focusing on the terminal recovery segments, the model adopts a Progressive Error Correction paradigm, which effectively facilitates the learning of robust manipulation skills within high-dimensional action spaces.

IV. EXPERIMENTS

In this section, we conduct comprehensive experiments to validate our method from three complementary perspectives:
RQ1: How does DexHiL efficiently improve the performance of the dexterous hand VLA policy? Can the system improve the model's performance over time?
RQ2: To what extent do our hardware-specific algorithmic designs contribute to the overall performance in complex manipulation tasks?

A. Experimental Setup

1) Tasks: To demonstrate the capabilities of our method, we design tasks that emphasize two complementary aspects. First, we evaluate the overall effectiveness of the full system, providing an intuitive validation of the proposed HiL framework. Second, we assess the advantages of our modular design by testing whether it can accomplish fine-grained, dexterous-hand-specific manipulation tasks. We consider:

- **Plush toy grasping.** The dexterous hand grasps and lifts a plush toy from a tabletop. This task primarily reflects the end-to-end policy performance and the hand's grasping capability.
- **Tissue extraction.** The dexterous hand pulls a single tissue from a tissue pack. This is a more challenging task that requires sufficiently precise dexterous retargeting to reliably extract a thin, deformable tissue.

The execution sequences for both tasks are depicted as sequential real-world rollouts in Fig. 3.

2) Implementation Details: Our robotic platform comprises a Franka Research 3 arm equipped with a DexHand021 dexterous hand. Initially, we use 60 offline trajectories for model initialization.In subsequent rounds, we expand the training set by adding 10 additional trajectories per task per round. We consider two conditions: (i) trajectories newly collected in-the-loop via our Dex-HiL procedure, and (ii) a non-HiL baseline where we instead add 10 trajectories per round from the remaining offline pool, matching the data volume while removing the benefit of interactive collection. This protocol isolates the effect of HiL-based post-training under the same data budget.

For training, we perform full training of **Being-H0.5** on 8 NVIDIA H100 GPUs for 60k training iterations, and then fine-tune on the human-interaction data using a single H100 GPU.

B. Baselines and Evaluation Protocol

To evaluate the efficacy of our framework, we design a *data-budget-matched* protocol for baseline comparison. During the fine-tuning of **Being-H0.5**, specific language instructions are utilized: "*Grab that yellow plushie*" for grasping and "*Pull out a napkin*" for extraction. All baselines are trained via full-parameter fine-tuning to ensure consistency.

Offline-40 (Data-matched R1): Employs 40 offline trajectories per task (80 total), matching the data budget of our HiL method at the first iteration round.

Fig. 4: **Visualization of retargeting results for four representative gestures.** We show the input **human hand** poses and the corresponding configurations generated by **Dex-retargeting [35]**, **GeoRT [24]**, and **our method**.Compared to other methods, our retargeting algorithm generates more accurate, smooth, and coordinated hand poses.

Offline-50 (Data-matched R2): Employs 50 offline trajectories per task (100 total), matching the data budget of our HiL method at the second iteration round.

Offline-60 (Data-matched R3): Employs 60 offline trajectories per task (120 total), matching the data budget of our HiL method at the third iteration round.

Evaluation Protocol: We evaluate the policies on two challenging dexterous tasks with rigorous success criteria. For *Tissue extraction*, a trial is successful if the extracted length exceeds half of the napkin's length. For *Plush toy grasping*, success is defined as the object being lifted entirely off the tabletop. We conduct 20 independent trials for each task on the physical robot and report the **Success Rate** as the primary metric. All experiments are initialized with the same pre-trained weights to ensure a fair comparison.

C. Experiment Results

Across both tasks, DexHiL consistently outperforms the baselines as training rounds progress (Tables I). In *Tissue Extraction*, DexHiL achieves a 95% success rate by Round 3, surpassing DAgger* (80%) and the offline Baseline (75%). Similarly, in *Plush Toy Grasping*, DexHiL reaches 65% success, whereas DAgger* and the Baseline struggle to scale, concluding at only 20% and 35%, respectively. These

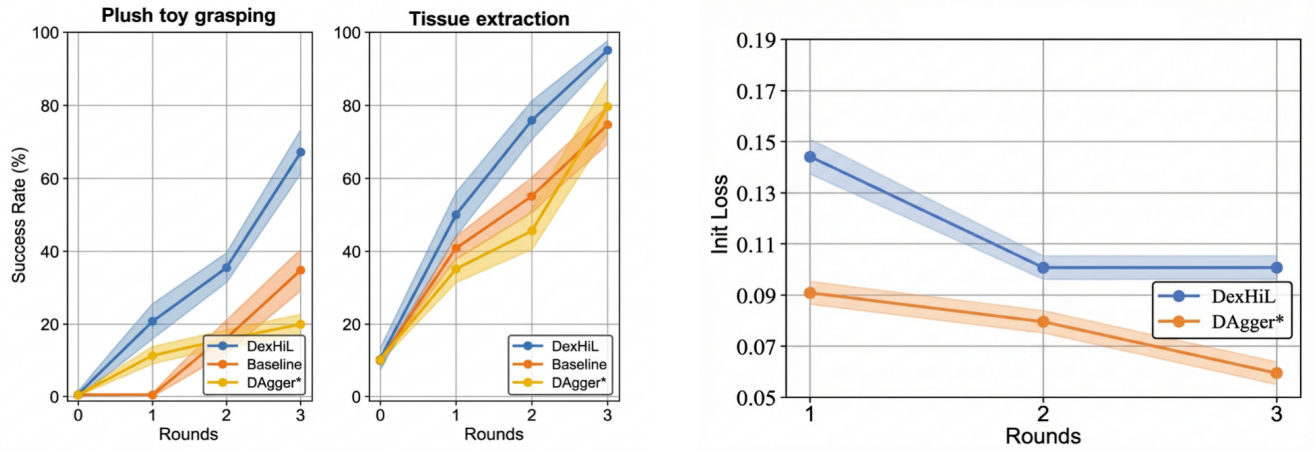

Fig. 5: **Quantitative performance and training loss analysis.** (Left) **Success rates** across three consecutive training rounds for both Tissue Extraction and Plush Toy Grasping tasks. (Right) **Initial training loss** at step 10 for both DexHiL and DAgger*. While the loss previously plateaued between 0.002 and 0.008, it significantly increases after incorporating human corrective trajectories.

results demonstrate the efficacy of our intervention-aware mechanism in facilitating policy refinement through human corrections. Beyond performance, DexHiL exhibits higher efficiency: each intervention segment takes only 3s (vs. 10s for offline), yielding a 35% reduction in total human labor (13 min vs. 20 min) by Round 3. This proves that our intervention-aware mechanism effectively prioritizes high-value corrective signals over redundant data.

TABLE I: **Success Rates** ($n/20$)

Method	**Warm-up**	**Round 1**	**Round 2**	**Round 3**
Task 1: Tissue Extraction				
DexHiL (Ours)	2/20	**10/20**	**15/20**	**19/20**
DAgger*[1]	2/20	7/20	9/20	16/20
Baseline[2]	2/20	8/20	11/20	15/20
Task 2: Plush Toy Grasping				
DexHiL (Ours)	0/20	**4/20**	**6/20**	**13/20**
DAgger*	0/20	2/20	3/20	4/20
Baseline	0/20	0/20	3/20	7/20

[1] **DAgger***: Online training without intervention-aware mechanism.
[2] **Baseline**: Control group with same number of trajectories.

In embodied dexterous control, intervention timing is pivotal. Our experiments identify two dominant failure modes that bottleneck performance: (1) **suboptimal contact-rich maneuvers**, such as failing to pinch a tissue edge, and (2) **arm-hand discoordination**, characterized by premature grasping before the wrist reaches the target. DexHiL addresses these by triggering human corrections at *impending failures* and preserving the subsequent recovery trajectories. This strategy captures essential state transitions required for error recovery, leading to significant performance gains. For instance, in *Tissue Extraction*, mastering the precise pinching maneuver enables the success rate to scale rapidly, nearly reaching a perfect ceiling by Round 3. Similarly, in the coordination-intensive *Plush Toy Grasping* task, DexHiL maintains a strong upward trajectory, substantially outperforming DAgger* and the unweighted baseline, which both struggle to overcome the complexity of the coordination bottleneck.

Beyond performance ceilings, DexHiL significantly improves sample efficiency. As illustrated in the training dynamics (Fig. 5, right), the updated policy π_k exhibits distinct loss spikes at the onset of each iteration ($k = 1, 2, 3$). These spikes signify a substantial *distribution shift*, where human corrections introduce critical out-of-distribution states and compounding errors that the nominal policy initially fails to handle. While standard methods like DAgger* dilute learning capacity across the entire dataset, our intervention-aware weighting mechanism prioritizes optimization on these high-value recovery states. By extracting more informative gradients per demonstration, DexHiL enables the policy to converge to expert-level success rates in fewer interaction rounds, demonstrating superior convergence speed compared to standard intervention paradigms.

To answer RQ2, we compare our method with two representative dexterous-hand teleoperation approaches discussed in the paper. As shown in Fig. 4, the optimization-based Dex-Retargeting [35] is constrained by hard, threshold-like control behavior: it tends to initiate finger opposition before the fingertips have established contact, leading to discontinuous adjustments around contact and making it difficult to smoothly regulate the hand configuration during grasp formation. In contrast, the learning-based GeoRT [24] struggles to form stable grasp postures: except for the thumb, the other four fingers often fail to fully flex to the intended closure, which makes palmar grasps that rely on coordinated

contact between the palm and the four fingers difficult to execute reliably. By overcoming these limitations in control continuity and grasp stability, it is precisely our algorithmic design that enables the high-precision finger control required for tasks like Tissue Extraction, where DexHiL achieves a 95% success rate.

V. CONCLUSION AND FUTURE WORK

We introduced DexHiL, an arm–hand integrated human-in-the-loop post-training framework for dexterous VLA. By combining a lightweight teleoperation interface with intervention-aware sampling that emphasizes corrective segments, DexHiL turns online interventions into high-value post-training data for contact-rich, high-DOF manipulation. Real-robot experiments demonstrate consistent gains in task success, establishing DexHiL as an effective and practical post-training recipe for improving dexterous VLA policies. After completing ongoing hardware and teleoperation optimizations, we will study dexterous-hand representation in VLA—particularly hand tokenizers and integrate them with our post-training pipeline to further improve performance and generalization.

REFERENCES

[1] 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.

[2] B. Zitkovich, T. Yu, S. Xu, P. Xu, T. Xiao, F. Xia, J. Wu, P. Wohlhart, S. Welker, A. Wahid *et al.*, "Rt-2: Vision-language-action models transfer web knowledge to robotic control," in *Proceedings of The 7th Conference on Robot Learning*. PMLR, 2023, pp. 2165–2183.

[3] M. J. Kim, K. Pertsch, S. Karamcheti, T. Xiao, A. Balakrishna, S. Nair, R. Rafailov, E. Foster, G. Lam, P. Sanketi *et al.*, "Openvla: An open-source vision-language-action model," in *Proceedings of The 8th Conference on Robot Learning*. PMLR, 2024, pp. 2679–2713.

[4] K. Black, N. Brown, D. Driess, A. Esmail, M. Equi, C. Finn, N. Fusai, L. Groom, K. Hausman, B. Ichter *et al.*, "π_0: A vision-language-action flow model for general robot control," *arXiv preprint arXiv:2410.24164*, 2024.

[5] L. Liu, W. Wang, Y. Han, Z. Xie, P. Yi, J. Li, Y. Qin, and W. Lian, "Foam: Foresight-augmented multi-task imitation policy for robotic manipulation," *arXiv preprint arXiv:2409.19528*, 2024.

[6] Octo Model Team, D. Ghosh, H. Walke, K. Pertsch, K. Black, O. Mees, S. Dasari, J. Hejna, T. Kreiman, C. Xu *et al.*, "Octo: An open-source generalist robot policy," in *Proceedings of Robotics: Science and Systems*, 2024.

[7] X. Li, M. Liu, H. Zhang, C. Yu, J. Xu, H. Wu, C. Cheang, Y. Jing, W. Zhang, H. Liu, H. Li, and T. Kong, "Vision-language foundation models as effective robot imitators," in *Proceedings of the 12th International Conference on Learning Representations*, 2024.

[8] T. Z. Zhao, V. Kumar, S. Levine, and C. Finn, "Learning fine-grained bimanual manipulation with low-cost hardware," in *Proceedings of Robotics: Science and Systems*, 2023.

[9] P. Wu, Y. Shentu, Z. Yi, X. Lin, and P. Abbeel, "Gello: A general, low-cost, and intuitive teleoperation framework for robot manipulators," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2024, pp. 12 156–12 163.

[10] S. Yang, M. Liu, Y. Qin, R. Ding, J. Li, X. Cheng, R. Yang, S. Yi, and X. Wang, "Ace: A cross-platform visual-exoskeletons system for low-cost dexterous teleoperation," in *Proceedings of The 8th Conference on Robot Learning*. PMLR, 2024.

[11] H. Luo, Y. Wang, W. Zhang, S. Zheng, Z. Xi, C. Xu, H. Xu, H. Yuan, C. Zhang, Y. Wang *et al.*, "Being-h0. 5: Scaling human-centric robot learning for cross-embodiment generalization," *arXiv preprint arXiv:2601.12993*, 2026.

[12] S. Ross, G. Gordon, and D. Bagnell, "A reduction of imitation learning and structured prediction to no-regret online learning," in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 2011, pp. 627–635.

[13] M. Kelly, C. Sidrane, K. Driggs-Campbell, and M. J. Kochenderfer, "Hg-dagger: Interactive imitation learning with human experts," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 8077–8083.

[14] H. Liu, S. Nasiriany, L. Zhang, Z. Bao, and Y. Zhu, "Robot learning on the job: Human-in-the-loop autonomy and learning during deployment," *The International Journal of Robotics Research*, vol. 44, no. 10-11, pp. 1727–1742, 2025.

[15] C. Pan, C. Wang, H. Qi, Z. Liu, H. Bharadhwaj, A. Sharma, T. Wu, G. Shi, J. Malik, and F. Hogan, "Spider: Scalable physics-informed dexterous retargeting," *arXiv preprint arXiv:2511.09484*, 2025.

[16] K. Li, P. Li, T. Liu, Y. Li, and S. Huang, "Maniptrans: Efficient dexterous bimanual manipulation transfer via residual learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025, pp. 6991–7003.

[17] S. Xu, Y.-W. Chao, L. Bian, A. Mousavian, Y.-X. Wang, L. Gui, and W. Yang, "Dexplore: Scalable neural control for dexterous manipulation from reference scoped exploration," in *Conference on Robot Learning*. PMLR, 2025, pp. 2184–2199.

[18] Z. Mandi, Y. Hou, D. Fox, Y. Narang, A. Mandlekar, and S. Song, "Dexmachina: Functional retargeting for bimanual dexterous manipulation," *arXiv preprint arXiv:2505.24853*, 2025.

[19] C. Wang, H. Shi, W. Wang, R. Zhang, L. Fei-Fei, and C. K. Liu, "Dexcap: Scalable and portable mocap data collection system for dexterous manipulation," *arXiv preprint arXiv:2403.07788*, 2024.

[20] Q. Ben, F. Jia, J. Zeng, J. Dong, D. Lin, and J. Pang, "Homie: Humanoid loco-manipulation with isomorphic exoskeleton cockpit," in *RSS 2025 Workshop on Whole-body Control and Bimanual Manipulation: Applications in Humanoids and Beyond*.

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

[22] R. Wen, J. Zhang, G. Chen, Z. Cui, M. Du, Y. Gou, Z. Han, J. Hu, L. Huang, H. Niu *et al.*, "Dexterous teleoperation of 20-dof bytedexter hand via human motion retargeting," *arXiv preprint arXiv:2507.03227*, 2025.

[23] J. Du, J. Ren, Q. Yu, N. Zhang, Y. Deng, X. Wei, Y. Liu, G. Gu, and X. Zhu, "Mile: A mechanically isomorphic exoskeleton data collection system with fingertip visuotactile sensing for dexterous manipulation," *arXiv preprint arXiv:2512.00324*, 2025.

[24] Z.-H. Yin, C. Wang, L. Pineda, K. Bodduluri, T. Wu, P. Abbeel, and M. Mukadam, "Geometric retargeting: A principled, ultrafast neural hand retargeting algorithm," in *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2025, pp. 17 376–17 382.

[25] E. Chong, L. Zhang, and V. J. Santos, "A learning-based harmonic mapping: Framework, assessment, and case study of human-to-robot hand pose mapping," *The International Journal of Robotics Research*, vol. 40, no. 2-3, pp. 534–557, 2021.

[26] Q. Li, Y. Deng, Y. Liang, L. Luo, L. Zhou, C. Yao, L. Zeng, Z. Feng, H. Liang, S. Xu *et al.*, "Scalable vision-language-action model pre-training for robotic manipulation with real-life human activity videos," *arXiv preprint arXiv:2510.21571*, 2025.

[27] Y. Zhong, X. Huang, R. Li, C. Zhang, Z. Chen, T. Guan, F. Zeng, K. N. Lui, Y. Ye, Y. Liang *et al.*, "Dexgraspvla: A vision-language-action framework towards general dexterous grasping," *arXiv preprint arXiv:2502.20900*, 2025.

[28] J. He, D. Li, X. Yu, Z. Qi, W. Zhang, J. Chen, Z. Zhang, Z. Zhang, L. Yi, and H. Wang, "Dexvlg: Dexterous vision-language-grasp model at scale," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025, pp. 14 248–14 258.

[29] Y. Chen, S. Tian, S. Liu, Y. Zhou, H. Li, and D. Zhao, "Conrft: A reinforced fine-tuning method for vla models via consistency policy," *arXiv preprint arXiv:2502.05450*, 2025.

[30] X. Zhang, M. Chang, P. Kumar, and S. Gupta, "Diffusion meets dagger: Supercharging eye-in-hand imitation learning," in *Robotics science and systems*. Robotics science and systems, 2024.

[31] X. Sun, S. Yang, M. Zhou, K. Liu, and R. Mangharam, "Mega-dagger: Imitation learning with multiple imperfect experts," *arXiv preprint arXiv:2303.00638*, 2023.

[32] K. Menda, K. Driggs-Campbell, and M. J. Kochenderfer, "Ensembledagger: A bayesian approach to safe imitation learning," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 5041–5048.

[33] R. Hoque, A. Balakrishna, E. Novoseller, A. Wilcox, D. S. Brown, and K. Goldberg, "Thriftydagger: Budget-aware novelty and risk gating for interactive imitation learning," *arXiv preprint arXiv:2109.08273*, 2021.

[34] Y. Cui, Y. Zhang, L. Tao, Y. Li, X. Yi, and Z. Li, "End-to-end dexterous arm-hand vla policies via shared autonomy: Vr teleoperation augmented by autonomous hand vla policy for efficient data collection," *arXiv preprint arXiv:2511.00139*, 2025.

[35] 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," in *Robotics: Science and Systems*, 2023.