HiFun: a Hierarchical Framework for Efficient Functional Dexterous Manipulation Learning

Anonymous Author(s)

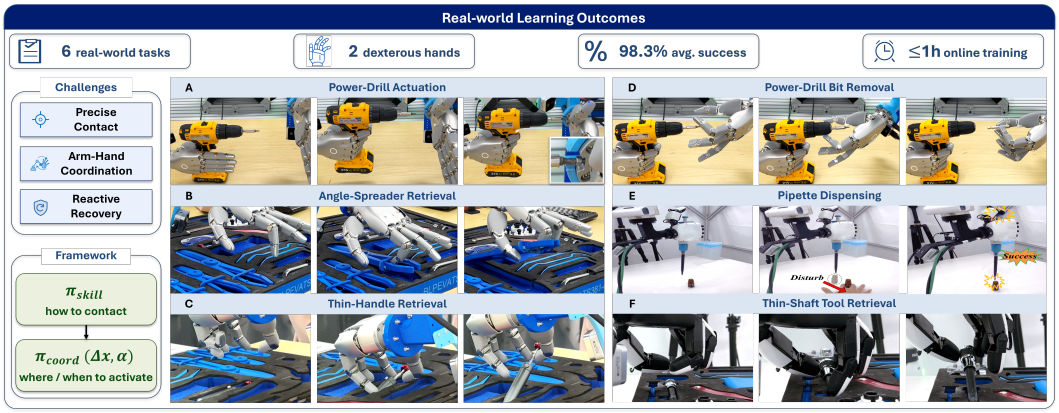

Figure 1: **HiFun is a hierarchical framework for functional dexterous manipulation.** It learns contact-aware skills and arm-hand coordination across precise tool-use (A,E), contact-rich rotational manipulation (D), and constrained-space manipulation (B,C,F), achieving 98.3% average success on six real-world tasks with two dexterous hands within one hour of online training.

Abstract:

While multi-fingered dexterous robot hands have the potential to provide human-level dexterity, using them for functional manipulation remains difficult because task success depends on precise contact, coordinated arm-hand motion, and reactive recovery from execution errors. Precise contact depends on complex contact dynamics that are difficult to model accurately in simulation, motivating real-world learning. However, real-world demonstrations in the full arm-hand control space are costly and noisy because each trajectory must coordinate arm motion with high-DoF multi-finger contact. Small execution deviations can further cause slipped contact or premature force application, making recovery from off-nominal contact states important for robust execution. To deal with these challenges, we propose HiFun, a hierarchical real-world RL framework for functional dexterous manipulation tasks. Our key insight is to decouple the control complexity into a functional hand skill for precise contact execution and an arm-hand coordination policy for arm control and hand-skill activation. This design enables effective real-time human intervention for reactive recovery learning and efficient data collection with Dynamic Movement Primitive (DMP) augmentation. Across six real-world tasks on two dexterous hands, HiFun achieves 98.3% average success within one hour of online training and remains robust to dynamic disturbances.

Keywords: Dexterous Manipulation, Real-world Reinforcement Learning

1 Introduction

Functional dexterous manipulation is central for robot applications from factories to homes. We study this setting through representative tasks in precise tool use, contact-rich rotation, and constrained-space manipulation. For example, Thin-Handle Tool Retrieval (Fig. 1C) requires extracting a foam-embedded handle through a narrow opening where direct grasping is blocked. The

Submitted to the 10th Conference on Robot Learning (CoRL 2026). Do not distribute.

robot must place fingertip contact on the exposed handle, coordinate arm motion with multi-finger levering to create access, and form a precision grasp without losing contact. Small contact errors can cause slipping, trapping, or collisions with the opening. These stages require precise contact, high-DoF arm-hand coordination, and recovery from contact errors, making functional dexterous manipulation challenging to learn and execute.

Specifically, precise contact requires the robot to place fingertips on task-relevant regions and regulate force as local contact changes. High-DoF arm-hand coordination requires arm motion, multi-finger contact, and contact timing to be synchronized within one execution sequence. Because fine-grained contact tasks are sensitive to small deviations, policies must recover from slipped contact, missed insertion, partial progress, or premature force application rather than only reproduce nominal successful trajectories.

Existing learning routes have made substantial progress, but these challenges remain only partially addressed for functional dexterity. Precise contact can be explored at scale in simulation, enabling robust dexterous manipulation performance [1, 2, 3, 4, 5, 6]. However, sim-to-real transfer is difficult for functional manipulation due to its intricate contact dynamics. High-DoF arm-hand behaviors can be represented by imitation learning (IL) and pretrained visuomotor policies from demonstrations and robot datasets [7, 8, 9, 10, 11, 12]. However, collecting real-world demonstrations for functional dexterity remains costly and noisy because each trajectory must jointly specify arm motion, multi-finger contact, and contact timing, which are difficult for an operator to perform at the same time, especially considering the embodiment gap between human and robotic hands [13, 14]. Furthermore, these demonstrations usually focus on successful cases, providing limited evidence for recovery after failures. Recovery from execution errors can be improved by real-world RL from online physical outcomes [15, 16, 17], but direct exploration and human correction in the full arm-hand action space are inefficient for learning reactive recovery.

To address these limitations, we propose HiFun, a hierarchical real-world RL framework for functional dexterous manipulation. For fine-grained finger control and precise contact execution, HiFun adapts kinesthetic hand-motion priors with residual RL and real contact feedback, producing functional hand skills grounded in physical interaction. To improve exploration efficiency in high-DoF arm-hand coordination, HiFun freezes the hand skills as action primitives and trains a coordination policy over arm motion and skill activation, with DMP-augmented rollouts providing initial coverage. During online fine-tuning, human corrections are provided through the hand-skill interface to make interventions more consistent and effective, while a multi-head critic improves temporal coordination and intervention advantage weighting (IAW) prioritizes useful corrections for reactive recovery. Together, this factorization reduces the burden of full arm-hand demonstrations, exploration, and correction, enabling reliable functional dexterous manipulation with efficient online training. We evaluate HiFun on six functional dexterous tasks across two dexterous hands, as shown in Fig. 1. HiFun achieves 98.3% average success within one hour of online training, outperforms IL and flat human-in-the-loop RL baselines, and remains robust to disturbances, while ablations verify the contributions of the hand-skill and coordination components.

In summary, our contributions are: (1) a hierarchical real-world RL framework that decouples high-DoF arm-hand control into functional hand-skill learning and skill-conditioned coordination, enabling efficient real-world learning of functional dexterous tasks; (2) residual-RL hand-skill learning that adapts kinesthetic priors with real contact feedback for fine-grained contact execution; (3) skill-conditioned coordination that uses hand skills as action primitives, DMP-augmented rollouts for efficient initial coverage, a multi-head critic for temporal coordination of arm motion and hand activation, and IAW for sample-efficient online recovery learning.

2 Related Work

Dexterous Manipulation. Model-based control plans stable dexterous grasps from explicit object and contact models [18, 19], but functional manipulation depends on contact-rich dynamics that are difficult to model accurately. Large-scale simulation has enabled robust dexterous grasping [20, 21, 22, 23, 24] and in-hand rotation [1, 2, 3, 4], yet functional contact dynamics involving

deformation, friction variation, and fine-grained force feedback remain difficult to simulate [25, 26], causing precise contact behaviors to transfer unreliably. IL and pretrained visuomotor policies can represent high-DoF arm-hand behaviors from demonstrations via diffusion-based policies [7, 27], transformer policies with action chunking [8, 28], and vision-language-action models [9, 12, 11]. For dexterous arm-hand systems, however, such demonstrations require specialized teleoperation or capture interfaces [13, 14, 29, 30, 31], and human-to-robot retargeting can introduce contact noise under morphology mismatch [32]. Thus, demonstrations for functional dexterity remain costly and noisy, struggle to cover fine-grained finger contacts and dynamic arm-hand coordination, and provide limited recovery evidence. HiFun instead obtains hand-motion priors through kinesthetic teaching [33], adapts them into contact-aware skills with residual RL, and uses these skills to collect downstream coordination data, reducing direct teleoperation of all dexterous-hand DoFs.

Real-World Reinforcement Learning. Real-world RL learns directly from physical interactions, capturing hardware dynamics and contact physics that are difficult to simulate [15, 34, 35]. Human-in-the-loop (HIL) methods improve sample efficiency by incorporating online corrections during training [17, 36, 37], and HIL-SERL achieves near-perfect success on precise gripper tasks with demonstrations and real-time guidance [17]. For functional dexterous manipulation, however, the high-dimensional arm-hand system degrades data precision and efficiency while increasing exploration burden, and coupled finger joints make consistent real-time correction across all DoFs difficult [38]. Hierarchical and skill-based methods reduce exploration complexity by reusing lower-level behaviors or guiding policy search. Prior work chains dexterous sub-policies for long-horizon manipulation [39], reuses learned rotation skills for harder in-hand reorientation [5], and guides RL exploration with simple sub-skill controllers [40]. We adopt a hierarchical framework that decouples hand primitive skills from arm-hand coordination, shifting online learning from full arm-hand exploration to skill-conditioned coordination and recovery.

3 Method

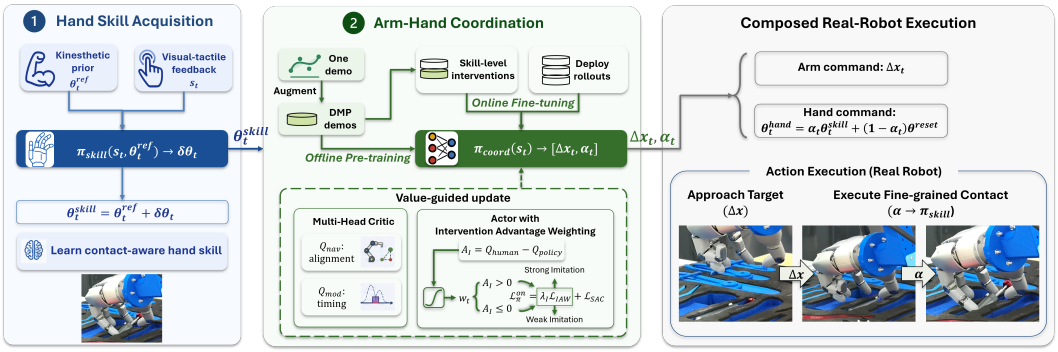

Figure 2: **Overview of HiFun, a hierarchical framework.** It learns (1) a residual-RL hand skill for contact-aware execution and (2) a coordination policy for arm motion and skill activation, initialized by DMP rollouts and refined by value-guided fine-tuning. At execution time, it combines arm motion and skill activation with the frozen hand skill for target approach and fine-grained contact.

To efficiently learn reliable contact-rich dexterous manipulation policies for high-DoF arm-hand systems, HiFun uses a hierarchical real-world RL framework (Fig. 2). It learns (1) a residual-RL hand skill for contact-aware execution and (2) a coordination policy for arm motion and skill activation. Coordination uses DMP rollouts for warm start, a multi-head critic to separate arm-alignment and skill-activation Q-values, and IAW for advantage-weighted intervention imitation. We introduce formulation (Sec. 3.1), hand skill training (Sec. 3.2), and coordination training (Sec. 3.3).

3.1 Problem Formulation

We formulate dexterous arm-hand manipulation as Markov Decision Processes with observation $s_t = (I_t, q_t, f_t)$, where I_t contains wrist and external RGB images, q_t contains arm-hand proprio-

ception, and f_t contains fingertip contact forces. The hand skill policy $\pi_{\text{skill}}(s_t, \theta_t^{\text{ref}})$ takes a reference hand pose from a kinesthetic demonstration and outputs a residual correction $\delta\theta_t \in \mathbb{R}^{12}$:

$$\theta_t^{\text{skill}} = \theta_t^{\text{ref}} + \lambda_{\text{res}}\delta\theta_t, \tag{1}$$

where λ_{res} limits residual exploration around the demonstrated motion. The coordination policy $\pi_{\text{coord}}(s_t)$ outputs $a_t^{\text{coord}} = [\Delta x_t, \alpha_t]$, where $\Delta x_t \in \mathbb{R}^6$ is the arm end-effector increment and $\alpha_t \in [0, 1]$ modulates the hand skill from initial neutral hand configuration θ^{reset}:

$$\theta_t^{\text{hand}} = \alpha_t\theta_t^{\text{skill}} + (1 - \alpha_t)\theta^{\text{reset}}. \tag{2}$$

The robot executes $(\Delta x_t, \theta_t^{\text{hand}})$, so π_{skill} determines how the fingers form contact and π_{coord} determines where and when that contact is used.

3.2 Contact-Aware Hand Skill Learning

The hand-skill policy abstracts high-dimensional finger articulation into robust, contact-aware functional skills. We acquire a kinesthetic prior using the hand-over-hand teaching interface of [33], collecting reference trajectories $\xi_{\text{ref}} = \{\theta_0^{\text{ref}}, \ldots, \theta_T^{\text{ref}}\}$ and per-finger force thresholds $F_{\max}^{(i)}$ for reward clipping. Residual RL then adapts this prior to real contact dynamics: given s_t and θ_t^{ref}, π_{skill} outputs $\delta\theta_t$ and executes Eq. (1). We train the critic on the executed hand command θ_t^{skill} so value estimates correspond to the actual finger configuration.

The hand skill reward combines trajectory tracking, contact incentive, and sparse skill success:

$$r_{\text{skill}} = -\,\omega_{\text{track}}\|\theta_t^{\text{cur}} - \theta_t^{\text{ref}}\|^2 + \omega_{\text{force}}\mathcal{F}(f_t) + \mathbb{I}_{\text{skill-success}}R_{\text{skill}}, \tag{3}$$

where $\mathcal{F}(f_t) = \sum_{i=1}^{N_f} \min(\|f_t^{(i)}\|, F_{\max}^{(i)})$ sums clipped fingertip-force magnitudes. The tracking term keeps exploration near contact geometry, the force term encourages task-relevant contact, and the sparse bonus marks operator-labeled success. We train π_{skill} using SAC [41] on the real robot and initialize replay with kinesthetic demonstrations. Once converged, π_{skill} is frozen as (i) the skill primitive invoked by π_{coord} and (ii) a reliable intervention interface for human operators.

3.3 Skill-Conditioned Arm-Hand Coordination

Given the frozen hand skill, π_{coord} maps s_t to $[\Delta x_t, \alpha_t]$, determining *where to move* the arm and *when to activate* the hand skill. We use RLPD [35] for sample-efficient learning from offline initialization data and online real-world interaction. Task completion is detected by a binary classifier over ResNet-10 visual features, fingertip forces, and joint torques, producing the terminal reward $R_{\text{task}} \in \{0, 1\}$.

Since the sparse reward R_{task} guides approach but provides no feedback on activation timing, the modulation reward r_{mod} penalizes premature activation and rewards successful activation:

$$r_{\text{mod}} = \begin{cases} r_{\text{bonus}} & \text{if } \alpha_t > \tau \text{ and } R_{\text{task}} = 1 \\ r_{\text{penalty}} & \text{if } \alpha_t > \tau \text{ and } R_{\text{task}} = 0 \\ 0 & \text{otherwise,} \end{cases} \qquad (r_{\text{penalty}} < 0 < r_{\text{bonus}}) \tag{4}$$

Multi-head critic. Naively training one Q-function on $R_{\text{task}} + r_{\text{mod}}$ can make the policy avoid activation: early high-α trials often fail, so accumulated penalties dominate the values of high-α actions. To decouple spatial approach from activation timing, we use two value heads trained with separate TD targets: a navigation head Q_ϕ^{nav} whose target y^{nav} is built from R_{task}, and a modulation head Q_ϕ^{mod} whose target y^{mod} is built from r_{mod}:

$$\mathcal{L}_Q(\phi) = \mathbb{E}\big[(Q_\phi^{\text{nav}} - y^{\text{nav}})^2 + (Q_\phi^{\text{mod}} - y^{\text{mod}})^2\big]. \tag{5}$$

Then the actor is trained with the combined value: $\hat{Q}_\phi(s, a) = Q_\phi^{\text{nav}}(s, a) + Q_\phi^{\text{mod}}(s, a)$. During actor optimization, we evaluate Q^{nav} with gradients stopped through α_t, and evaluate Q^{mod} with gradients stopped through Δx_t. This encourages the navigation head to shape arm alignment and

the modulation head to shape activation timing, while both heads score the same physical action. With this multi-head critic, coordination training follows an offline-to-online scheme.

Offline pre-training. To avoid the prohibitive sample complexity of learning from scratch, we bootstrap coordination from one demonstration containing an arm trajectory and skill-activation schedule. Cartesian Dynamic Movement Primitives (DMP) [42] generate diverse arm trajectories by randomizing initial poses while fixing the goal. DMP action labels use clipped waypoint differences for Δx_t and align α_t with the demonstration by task progress. The resulting $\mathcal{D}_{\text{dmp}}$ provides initial-state coverage. We pre-train π_{coord} with SAC plus a global behavior-cloning term:

$$\begin{aligned}\mathcal{L}_\pi^{\text{off}} = \mathbb{E}_{s\sim\mathcal{D}_{\text{dmp}},a\sim\pi}[\eta \log \pi(a|s) - \hat{Q}_\phi(s,a)] \\ + \lambda_{\text{gbc}}\mathbb{E}_{(s,a^{\text{dmp}})\sim\mathcal{D}_{\text{dmp}}}[\|\bar{a}_\pi(s) - a^{\text{dmp}}\|^2],\end{aligned} \tag{6}$$

where a^{dmp} and $\bar{a}_\pi$ denote the DMP label and deterministic policy action. This keeps the policy near generated approach-and-activation behavior while allowing critic improvement.

Online fine-tuning. Online training uses real policy rollouts and skill-level corrections to refine π_{coord} for better target alignment, skill timing, and recovery under real contact dynamics. Arm interventions use relative SE(3) retargeting through Gello [43] for smooth takeover, while hand interventions trigger the frozen π_{skill} by button press rather than teleoperating all hand DoFs. Rollouts and intervention segments are stored for RLPD updates, and IAW weights an intervention-imitation loss by the critic-estimated advantage of each correction. For an intervention action a^I, we estimate its advantage over the current deterministic action using the target critic ensemble:

$$A^I(s) = \min_i \hat{Q}_{\bar{\phi},i}(s,a^I) - \min_i \hat{Q}_{\bar{\phi},i}(s,\bar{a}_\pi(s)). \tag{7}$$

The imitation weight and loss are

$$w(s) = \sigma(\text{Norm}(A^I(s))/\tau_I), \qquad \mathcal{L}_{\text{IAW}} = \mathbb{E}_{(s,a^I)\sim\mathcal{D}_I}[w(s)\|\bar{a}_\pi(s) - a^I\|^2]. \tag{8}$$

The online actor objective augments the standard SAC actor loss with the IAW term, $\mathcal{L}_\pi^{\text{on}} = \mathcal{L}_{\text{SAC}} + \lambda_I\mathcal{L}_{\text{IAW}}$. IAW prevents blind cloning: corrections estimated to improve over the current policy receive larger weights, while low-advantage corrections have limited influence.

4 Experiments

4.1 Experimental Setup

We evaluate HiFun on six real-world tasks spanning Power-Drill Actuation, Angle-Spreader Retrieval, Thin-Handle Retrieval, Power-Drill Bit Removal, Pipette Dispensing, and Thin-Shaft Tool Retrieval, corresponding to Fig. 1A–F in order. These tasks cover precise tool-use, contact-rich rotation, and constrained-space manipulation. A trial is counted as successful only when the task-specific functional outcome is completed: tightening the target screw (A), fully extracting the target tools (B,C,F), removing the drill bit (D), or dispensing into the vial (E). All tasks use Franka Research 3 arms [44] with either a Sharpa hand [45] or an XHand [46], wrist RGB observations from an Intel RealSense D405, and a side-view D415 when needed.

Metrics. For final evaluation, each task randomizes object poses within the training workspace while using the same robot reset strategy as training. We report IL data-collection time for recording 200 demonstrations, RL data-collection time for generating 100 automated DMP rollouts without human labor, RL online training time for HIL fine-tuning of the coordination policy, and success rate, the percentage of evaluation trials that meet the task criterion. Full evaluation protocol details are provided in the Appendix.

Baselines and Protocol. To evaluate HiFun, we compare against HIL-SERL [17], a state-of-the-art HIL-RL method that combines offline data, online RL, and real-time interventions for real-world manipulation. The key difference is that HiFun packages contact execution into a frozen hand skill and trains a hierarchical coordination policy over arm motion and skill activation, whereas HIL-SERL keeps a flat policy architecture that directly outputs full-DoF arm-hand actions.

Task	IL Data Collection (mins)	RL Data Collection (mins)	RL Online Training (mins)	Success Rate (%)				
				HIL-SERL (Glove+Gello)	HIL-SERL (Ours Interv.)	HiFun (Ours)	Behavior Cloning (BC)	Diffusion Policy (DP)
Sharpa Main Validation Tasks								
Power-Drill Actuation	150	34	60	15	$62 — 30^{\ddagger}$	**100**	15	30
Angle-Spreader Retrieval	60	13	55	–	$40 — 0^{\ddagger}$	**100**	30	60
Thin-Handle Retrieval	80	17	60	–	$28 — 0^{\ddagger}$	**100**	35	40
Power-Drill Bit Removal	80	21	22	–	$60 — 20^{\ddagger}$	**100**	30	36
XHand Portability Tasks								
Pipette Dispensing	100	14	60	12	$60 — 32^{\ddagger}$	**90**	12	26
Thin-Shaft Tool Retrieval	100	14	20	–	$30 — 0^{\ddagger}$	**100**	20	58
Average	95.0	18.8	46.2	4.5	13.7	**98.3**	23.7	41.7

Table 1: **Main results.** Success rates are computed over 50 trials. "–" marks HIL-SERL (Glove+Gello) settings where direct arm-hand teleoperation can not consistently provide reliable corrections for contact-rich tasks. $^{\ddagger}a — b$ reports arm approach success (a) and final task success with the required hand contact (b). Averages use final-task success and count "–" as 0.

HIL-SERL (Glove+Gello) uses AnyTeleop [14] for hand-mapped teleoperation and Gello [43] for arm control, testing the benefit of the hand-skill intervention interface over full arm-hand correction. **HIL-SERL (Ours Interv.)** uses the same learned intervention interface as HiFun but keeps a flat full-DoF policy, isolating the benefit of the hierarchical architecture beyond the intervention mechanism. All RL methods are initialized from an offline demo buffer containing 100 DMP-generated trajectories and trained online for up to one hour, a cap that pilot runs found sufficient while limiting variability from operator attention and intervention timing.

For IL baselines, **Behavior Cloning (BC)** [47] and **Diffusion Policy (DP)** [7] are trained with 200 successful teleoperated demonstrations per task to test whether offline imitation from human demonstrations can handle functional dexterous tasks, where small contact or timing deviations often cause failure. To distinguish training-data coverage from the learning paradigm, the Appendix reports additional IL baselines trained on the same successful trajectory sources as HiFun.

4.2 Main Results

We evaluate the performance of HiFun by comparing it against representative baselines that include both RL and IL paradigms. Tab. 1 shows the strongest performance: 98.3% average success across six tasks and two hands with 18.8 minutes of RL data collection and 46.2 minutes of online training. To further validate the autonomy of HiFun, we monitor the intervention rate over time, defined as the ratio of intervention steps to total steps per episode. In the Appendix, intervention-rate curves show human guidance approaching zero by the end of training, indicating autonomous execution.

Despite sharing the same RL backbone and initial data, HiFun substantially outperforms the HIL-SERL baselines, which we attribute to intervention quality and exploration efficiency. Specifically, standard teleoperation methods (Glove+Gello) generate noisy interventions due to morphology mismatches and retargeting errors, leading to inconsistent demonstrations (for instance, wrist shaking when operating a power drill) and a degraded replay buffer. For tool retrieval tasks with complex hand motions, obtaining successful corrections becomes exceedingly challenging, leading to complete failures marked "–" in Tab. 1. Using our hand-skill intervention interface improves correction quality, as HIL-SERL (Ours Interv.) rises to 13.7% average success, but its flat full-DoF policy remains far below HiFun. The staged success results marked with ‡, such as $40 — 0$, $28 — 0$, and $30 — 0$, show that this flat policy often approaches the target but fails to complete reliable hand contact. This indicates that the absence of hierarchical decomposition still limits exploration efficiency, particularly for tasks with complex hand motions. For example, a flat policy for *Thin-Handle Retrieval* must discover sustained non-prehensile levering, precision grasping, and collision avoidance as one multi-stage contact sequence, which requires many full arm-hand interactions, producing many noisy rollouts before the correct contact sequence is found. Overall, the results indicate that our hierarchical structure effectively improves intervention quality by leveraging fine-grained hand skill primitives and enables more efficient learning by simplifying the exploration space.

Data Source Comparison				
Task	**Success Rate (%)**		**Data Collection Time (mins)**	
	(w/ DMP)	**(w/ Human)**	**(w/ DMP)**	**(w/ Human)**
Angle-Spreader Retrieval	100	85	13	60
Power-Drill Bit Removal	100	100	21	80
Component Ablation (Success Rate %)				
Task	**Ours (Full)**	**w/o Multi-head**	**w/o IAW (Uniform BC)**	**w/o Interv. BC**
Angle-Spreader Retrieval	100	85	60	40
Power-Drill Bit Removal	100	75	60	40

Table 2: **Coordination policy analysis.** Top: DMP rollouts vs. human demonstrations for warm-start data. Bottom: ablations of the multi-head critic and intervention advantage weighting (IAW).

Furthermore, even with 200 expert demonstrations per task, IL baselines struggle with precise and contact-rich tasks. BC suffers from compounding error in long-horizon execution, reaching only 15% on *Power-Drill Actuation* and 12% on *Pipette Dispensing*. DP improves by modeling multimodal action distributions, achieving 60% on *Angle-Spreader* and 58% on *Thin-Shaft Tool*, but still fails on the *Bit Removal* (36%) and *Pipette* (26%) tasks. These failures highlight deficiencies in fine-grained alignment and reactive behavior, indicating that successful offline demonstrations weakly cover contact deviations and provide limited recovery signal. To separate data coverage from the learning paradigm, we additionally report IL baselines in the Appendix trained with successful trajectories from HiFun. These policies remain brittle under precise alignment requirements and off-nominal contact states, indicating that purely offline training makes them sensitive to distribution shifts during inference. Our method addresses this issue by incorporating online exploration and human interventions to refine closed-loop behaviors. In summary, the results show that our method achieves both more efficient and effective learning of contact-rich functional dexterous manipulation tasks compared with the baselines.

Robustness Evaluation. We evaluate whether HiFun can recover from off-nominal states using 20 perturbation trials per task. The perturbations cover three failure modes: (1) randomized object poses across all tasks, (2) approach-phase displacements for alignment-sensitive tasks, including *Power-Drill Actuation*, *Tool Retrieval*, and *Pipette Dispensing*, and (3) drill-chuck re-tightening for *Power-Drill Bit Removal*. HiFun maintains near-perfect robustness across both embodiments, achieving 20/20 success on four *Tool Retrieval* tasks, and 18/20 success on *Power-Drill Actuation* and *Pipette Dispensing*. The remaining failures stem from screw-drill angular misalignment or vial motion after alignment. Successful trials show that HiFun re-aligns before contact and suspends hand-skill activation until contact is reachable (Fig. 1), supporting recovery through delayed activation rather than fixed replay. Additional examples are available in the supplementary video.

4.3 Ablation and Analysis

1) Arm-Hand Coordination: Tab. 2 analyzes coordination policy training from two aspects: (i) the effect of warm-start data sources (DMP rollouts vs. human demonstrations) and (ii) the contributions of the multi-head critic and intervention advantage weighting (IAW) through component ablations.

Data Source Analysis. Compared with human demonstrations, DMP rollouts achieve the same or higher final success (100%) with about $4\times$ less collection time (Tab. 2), indicating that automatically collected DMP data can serve as a practical alternative to labor-intensive human demonstrations. Additional trajectory-distribution analysis is provided in the Appendix.

Effect of Multi-head Critic. Tab. 2 shows that the multi-head critic is important for complex arm-skill coordination: removing it drops success from 100% to 85% on *Angle-Spreader* and to 75% on *Bit Removal*. To explain this drop, we measure how close the arm end-effector is to the target when the trained policy triggers the hand skill. The multi-head critic keeps 95% of activations within the $2.00\,\mathrm{cm}$ pre-contact region, while the single-head variant increases the 95th-percentile error to $4.58\,\mathrm{cm}$. This indicates that a single-head critic activates the skill before the arm is sufficiently aligned, because it conflates sparse terminal success rewards with smaller modulation penalties. Simply increasing the penalties does not resolve the conflict, as it can over-suppress activation to

Task	Ours (π_{skill})	Kinesthetic Replay	DexPilot*	AnyTeleop*
Power-Drill Actuation	20/20	0/20	18/20	19/20
Angle-Spreader Retrieval	20/20	0/20	7/20	5/20
Thin-Handle Retrieval	20/20	0/20	0/20	0/20
Power-Drill Bit Removal	20/20	5/20	3/20	3/20
Pipette Dispensing	20/20	0/20	20/20	20/20
Thin-Shaft Tool Retrieval	20/20	0/20	0/20	0/20

Table 3: **Hand Motion Control Comparison.** Success rate over 20 trials.

avoid penalty accumulation. By decoupling arm-alignment and skill-activation value learning, the multi-head critic enables efficient arm movement with better-timed hand-skill activation.

Effect of IAW. During online fine-tuning, intervention actions provide an additional imitation loss for the actor objective. IAW weights each intervention-imitation loss by the critic-estimated advantage in Eq. (7), comparing the human correction with the current policy action. We compare it with two variants: *Uniform Interv. BC*, which keeps the imitation loss but assigns all interventions equal weight, and *w/o Interv. BC*, which removes this imitation loss from the actor objective. *Uniform Interv. BC* drops success to 60% and *w/o Interv. BC* to 40%, while HiFun attains lower cycle time (6.69 s) than both variants (9.21 s and 10.03 s). These results show that intervention imitation helps, but uniform cloning can copy noisy or lower-value corrections and pull the actor away from Q maximization. The Appendix further shows that IAW assigns separated weights to weak and strong corrections (0.175 vs. 0.884 mean), with high-weight corrections occurring where the current policy action has lower estimated value. Thus, IAW concentrates imitation on states where the policy needs guidance while limiting low-advantage corrections that could conflict with the critic objective.

2) Hand Motion Control Comparison: To isolate contact execution, Tab. 3 compares the learned residual hand skill π_{skill} with open-loop kinesthetic replay and retargeted teleoperation, with DexPilot* and AnyTeleop* reported after task-specific tuning. Ours succeeds in all 20 trials on all tasks, while Kinesthetic Replay succeeds only partially on Power-Drill Bit Removal (5/20) and fails elsewhere because geometric trajectories without force feedback cannot sustain precision contact or constrained extraction; retargeted teleoperation reaches 20/20 on Pipette but drops on constrained retrieval and drill-bit removal due to fine-grained thumb-finger retargeting errors. Thus, residual contact modulation is necessary for reliable functional hand skills across both embodiments.

5 Conclusion

Functional dexterous manipulation on real robots is difficult because precise contact, arm-hand coordination, and recovery are tightly coupled in a high-dimensional control space. We present HiFun, a hierarchical real-world RL framework that learns contact-aware hand skills separately and freezes them for coordination and human intervention. The resulting skill-conditioned policy optimizes arm motion, skill activation, and recovery without directly exploring or correcting all arm-hand DoFs. DMP-augmented rollouts warm-start coordination, the multi-head critic improves temporal arm-hand coordination, and IAW prioritizes useful interventions for recovery learning. Across six real-world tasks and two dexterous hands, HiFun achieves 98.3% average success within one hour of online training and remains robust to dynamic disturbances. These results suggest that structuring dexterous learning around functional hand skills is a practical path toward efficient, reliable real-world manipulation with multi-fingered hands.

6 Limitations

HiFun assumes task-relevant hand skills learned from kinesthetic priors and real contact feedback. This improves fine-grained contact execution and reduces full arm-hand exploration, but tasks demanding contact behaviors beyond the learned skill set may fail and require additional priors, online adaptation, or reusable hand-skill libraries. Our method also targets reliable low-level execution rather than open-ended planning or language-conditioned generalization. This scope makes it complementary to VLA policies and world models: HiFun could provide contact-rich low-level controllers and consistent autonomous rollouts, while pretrained models provide task selection and semantic grounding. Combining pretrained reasoning with HiFun controllers remains future work.

References

[1] OpenAI, M. Andrychowicz, B. Baker, M. Chociej, R. Jozefowicz, B. McGrew, J. Pachocki, A. Petron, M. Plappert, G. Powell, A. Ray, J. Schneider, S. Sidor, J. Tobin, P. Welinder, L. Weng, and W. Zaremba. Learning dexterous in-hand manipulation. *arXiv preprint arXiv:1808.00177*, 2019.

[2] H. Qi, A. Kumar, R. Calandra, Y. Ma, and J. Malik. In-hand object rotation via rapid motor adaptation. In *Conference on Robot Learning*, pages 1722–1732. PMLR, 2023.

[3] T. Chen, M. Tippur, S. Wu, V. Kumar, E. Adelson, and P. Agrawal. Visual dexterity: In-hand reorientation of novel and complex object shapes. *Science Robotics*, 2023. doi:10.1126/scirobotics.adc9244. arXiv:2211.11744.

[4] X. Liu, H. Wang, and L. Yi. Dexndm: Closing the reality gap for dexterous in-hand rotation via joint-wise neural dynamics model. *arXiv preprint arXiv:2510.08556*, 2025.

[5] H. Qi, B. Yi, M. Lambeta, Y. Ma, R. Calandra, and J. Malik. From simple to complex skills: The case of in-hand object reorientation. *arXiv preprint arXiv:2501.05439*, 2025.

[6] H. Zhang, S. Christen, Z. Fan, O. Hilliges, and J. Song. GraspXL: Generating grasping motions for diverse objects at scale. In *European Conference on Computer Vision (ECCV)*, 2024.

[7] C. Chi, Z. Xu, S. Feng, E. Cousineau, Y. Du, B. Burchfiel, R. Tedrake, and S. Song. Diffusion policy: Visuomotor policy learning via action diffusion. *The International Journal of Robotics Research*, 44(10-11):1684–1704, 2025.

[8] T. Z. Zhao, V. Kumar, S. Levine, and C. Finn. Learning Fine-Grained Bimanual Manipulation with Low-Cost Hardware, Apr. 2023.

[9] A. Brohan, N. Brown, J. Carbajal, Y. Chebotar, X. Chen, K. Choromanski, T. Ding, D. Driess, A. Dubey, C. Finn, P. Florence, C. Fu, M. G. Arenas, K. Gopalakrishnan, K. Han, K. Hausman, A. Herzog, J. Hsu, B. Ichter, A. Irpan, N. Joshi, R. Julian, D. Kalashnikov, Y. Kuang, I. Leal, L. Lee, T.-W. E. Lee, S. Levine, Y. Lu, H. Michalewski, I. Mordatch, K. Pertsch, K. Rao, K. Reymann, M. Ryoo, G. Salazar, P. Sanketi, P. Sermanet, J. Singh, A. Singh, R. Soricut, H. Tran, V. Vanhoucke, Q. Vuong, A. Wahid, S. Welker, P. Wohlhart, J. Wu, F. Xia, T. Xiao, P. Xu, S. Xu, T. Yu, and B. Zitkovich. Rt-2: Vision-language-action models transfer web knowledge to robotic control, 2023. URL https://arxiv.org/abs/2307.15818.

[10] A. O'Neill, A. Rehman, A. Gupta, A. Maddukuri, A. Gupta, A. Padalkar, A. Lee, A. Pooley, A. Gupta, A. Mandlekar, et al. Open x-embodiment: Robotic learning datasets and rt-x models, 2023. URL https://arxiv.org/abs/2310.08864.

[11] M. J. Kim, K. Pertsch, S. Karamcheti, T. Xiao, A. Balakrishna, S. Nair, R. Rafailov, E. Foster, G. Lam, P. Sanketi, Q. Vuong, T. Kollar, B. Burchfiel, R. Tedrake, D. Sadigh, S. Levine, P. Liang, and C. Finn. Openvla: An open-source vision-language-action model, 2024. URL https://arxiv.org/abs/2406.09246.

[12] K. Black, N. Brown, D. Driess, A. Esmail, M. Equi, C. Finn, N. Fusai, L. Groom, K. Hausman, B. Ichter, S. Jakubczak, T. Jones, L. Ke, S. Levine, A. Li-Bell, M. Mothukuri, S. Nair, K. Pertsch, L. X. Shi, J. Tanner, Q. Vuong, A. Walling, H. Wang, and U. Zhilinsky. π_0: A vision-language-action flow model for general robot control, 2026. URL https://arxiv.org/abs/2410.24164.

[13] A. Handa, K. V. 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, Oct. 2019.

[14] Y. Qin, W. Yang, B. Huang, K. V. Wyk, H. Su, X. Wang, Y.-W. Chao, and D. Fox. AnyTeleop: A General Vision-Based Dexterous Robot Arm-Hand Teleoperation System, May 2024.

[15] S. Levine, C. Finn, T. Darrell, and P. Abbeel. End-to-end training of deep visuomotor policies. *Journal of Machine Learning Research*, 17(39):1–40, 2016.

[16] J. Luo, Z. Hu, C. Xu, Y. L. Tan, J. Berg, A. Sharma, S. Schaal, C. Finn, A. Gupta, and S. Levine. Serl: A software suite for sample-efficient robotic reinforcement learning. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 16961–16969. IEEE, 2024.

[17] J. Luo, C. Xu, J. Wu, and S. Levine. Precise and dexterous robotic manipulation via human-in-the-loop reinforcement learning. *Science Robotics*, 10(105):eads5033, 2025.

[18] A. Miller and P. Allen. Graspit! A versatile simulator for robotic grasping. *IEEE Robotics & Automation Magazine*, 2004.

[19] M. A. Roa and R. Suárez. Grasp quality measures: review and performance. *Auton. Robots*, 2015.

[20] H. Zhang, Z. Wu, L. Huang, S. Christen, and J. Song. Robustdexgrasp: Robust dexterous grasping of general objects. *arXiv preprint arXiv:2504.05287*, 2025.

[21] Z. Chen, Q. Yan, Y. Chen, T. Wu, J. Zhang, Z. Ding, J. Li, Y. Yang, and H. Dong. Clutterdexgrasp: A sim-to-real system for general dexterous grasping in cluttered scenes. In *Conference on Robot Learning*, 2025.

[22] L. Huang, H. Zhang, Z. Wu, S. Christen, and J. Song. Fungrasp: functional grasping for diverse dexterous hands. *IEEE Robotics and Automation Letters*, 2025.

[23] T. G. W. Lum, M. Matak, V. Makoviychuk, A. Handa, A. Allshire, T. Hermans, N. D. Ratliff, and K. V. Wyk. DextrAH-G: Pixels-to-Action Dexterous Arm-Hand Grasping with Geometric Fabrics. In *Conference on Robot Learning (CoRL)*, 2024.

[24] R. Singh, A. Allshire, A. Handa, N. Ratliff, and K. Van Wyk. Dextrah-rgb: Visuomotor policies to grasp anything with dexterous hands. *arXiv preprint arXiv:2412.01791*, 2024.

[25] J. Wang, Y. Yuan, H. Che, H. Qi, Y. Ma, J. Malik, and X. Wang. Lessons from learning to spin" pens". *arXiv preprint arXiv:2407.18902*, 2024.

[26] E. Hsieh, W.-H. Hsieh, Y.-J. Wang, T. Lin, J. Malik, K. Sreenath, and H. Qi. Learning dexterous manipulation skills from imperfect simulations. *arXiv preprint arXiv:2512.02011*, 2025.

[27] Y. Ze, G. Zhang, K. Zhang, C. Hu, M. Wang, and H. Xu. 3D Diffusion Policy: Generalizable Visuomotor Policy Learning via Simple 3D Representations. In *Robotics: Science and Systems (RSS)*, 2024.

[28] T. Buamanee, M. Kobayashi, Y. Uranishi, and H. Takemura. Bi-act: Bilateral control-based imitation learning via action chunking with transformer, 2024. URL `https://arxiv.org/abs/2401.17698`.

[29] C. Wang, H. Shi, W. Wang, R. Zhang, L. Fei-Fei, and K. Liu. DexCap: Scalable and Portable Mocap Data Collection System for Dexterous Manipulation. In *Robotics: Science and Systems XX*. Robotics: Science and Systems Foundation, July 2024. ISBN 979-8-9902848-0-7. doi: 10.15607/RSS.2024.XX.043.

[30] M. Xu, H. Zhang, Y. Hou, Z. Xu, L. Fan, M. Veloso, and S. Song. DexUMI: Using Human Hand as the Universal Manipulation Interface for Dexterous Manipulation, May 2025.

[31] H.-S. Fang, B. Romero, Y. Xie, A. Hu, B.-R. Huang, J. Alvarez, M. Kim, G. Margolis, K. Anbarasu, M. Tomizuka, E. Adelson, and P. Agrawal. Dexop: A device for robotic transfer of dexterous human manipulation. *arXiv preprint arXiv:2509.04441*, 2025.

[32] 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, Mar. 2025.

[33] D. Zhang, C. Yuan, C. Wen, H. Zhang, J. Zhao, and Y. Gao. Kinedex: Learning tactile-informed visuomotor policies via kinesthetic teaching for dexterous manipulation. *arXiv preprint arXiv:2505.01974*, 2025.

[34] J. Luo, O. Sushkov, R. Pevceviciute, W. Lian, C. Su, M. Vecerik, N. Ye, S. Schaal, and J. Scholz. Robust multi-modal policies for industrial assembly via reinforcement learning and demonstrations: A large-scale study. *arXiv preprint arXiv:2103.11512*, 2021.

[35] P. J. Ball, L. Smith, I. Kostrikov, and S. Levine. Efficient online reinforcement learning with offline data. In *International Conference on Machine Learning*, pages 1577–1594. PMLR, 2023.

[36] 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)*, pages 8077–8083. IEEE, 2019.

[37] S. Ross, G. J. Gordon, and J. A. Bagnell. A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning, Mar. 2011.

[38] H. Zhu, J. Yu, A. Gupta, D. Shah, K. Hartikainen, A. Singh, V. Kumar, and S. Levine. The ingredients of real-world robotic reinforcement learning. *arXiv preprint arXiv:2004.12570*, 2020.

[39] Y. Chen, C. Wang, L. Fei-Fei, and C. K. Liu. Sequential Dexterity: Chaining Dexterous Policies for Long-Horizon Manipulation. *Conference on Robot Learning (CoRL)*, 2023.

[40] G. Khandate, C. P. Mehlman, X. Wei, and M. Ciocarlie. Dexterous in-hand manipulation by guiding exploration with simple sub-skill controllers. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6551–6557. IEEE, 2024.

[41] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International conference on machine learning*, pages 1861–1870. Pmlr, 2018.

[42] A. Ude, B. Nemec, T. Petrić, and J. Morimoto. Orientation in cartesian space dynamic movement primitives. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2997–3004. IEEE, 2014.

[43] 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)*, pages 12156–12163. IEEE, 2024.

[44] FRANKA ROBOTICS. FRANKA-RESEARCH-3. https://franka.de/franka-research-3.

[45] SHARPA. SHARPA WAVE. https://www.sharpa.com/pages/wave.

[46] ROBOTERA. XHAND1. https://www.robotera.com/goods/2.html.

[47] D. A. Pomerleau. Alvinn: An autonomous land vehicle in a neural network. *Advances in neural information processing systems*, 1, 1988.