

ConTrack: Constrained Hand Motion Tracking with Adaptive Trade-off Control

Yutong Liang, Quanquan Peng, Ri-Zhao Qiu, and Xiaolong Wang

University of California San Diego, La Jolla, CA, USA

Abstract. Human demonstrations provide strong priors for robot manipulation, yet it is non-trivial to transfer them to execute on real robots due to the kinematic gap. In dexterous manipulation, it remains challenging to track long-horizon, contact-rich sequences even in simulators: a reference-tracking policy must keep objects on their target trajectories while preserving demonstrated joint motion and contact timing. Existing approaches often rely on hand-crafted reward tuning that require per-sequence tuning and break under limited interaction budgets. We introduce ConTrack, a reinforcement learning (RL) framework that scales with tracking data. ConTrack treats object tracking as a constraint and allocates remaining control authority to motion fidelity, which allows it to adapt task-style trade-offs online using a dual-variable update. In addition, ConTrack also stabilizes long-horizon learning with an adaptive mid-trajectory reset library that reuses policy-reachable simulator states. Our qualitative and quantitative results in simulation tracking and real robot demonstrate that ConTrack improves success and object pose accuracy significantly over prior arts while preserving joint and contact fidelity. Website: <https://www.lyt0112.com/projects/ConTrack>.

Keywords: Dexterous manipulation · In-hand manipulation · Reference tracking · Multi-objective reinforcement learning

1 Introduction

Recent progress in robot manipulation has been driven by scaling robot data and training general-purpose policies [2, 16, 28, 43]. A parallel trend aims to learn from humans, motivated by the abundance of human demonstrations and their potential to supervise dexterous behavior at scale [4, 22, 23, 31, 32, 38, 47, 53]. Yet many of these pipelines extract high-level intent or coarse motion cues from human observations and still rely on an additional stage to calibrate end-effector interaction for contact-rich manipulation [4, 53].

This extra stage matters because dexterous manipulation is dominated by contact, and contact is where human-to-robot transfer is least forgiving. Compared to collecting robot-native demonstrations through teleoperation interfaces [8, 9, 48], human data introduces a systematic mismatch between the demonstrated hand motion and what a robot hand can physically realize. The root cause is morphological and actuation differences between human hands and robot



Fig. 1: ConTrack enables long-horizon, contact-rich dexterous hand tracking. With adaptive task-style trade-off control, ConTrack yields physically plausible object motion balancing hand-object contacts and geometric tracking. This aligns the embodiment gap between human hands and dexterous hands while transferring human dexterous play data to real robots (Tasks from top to bottom: functional tool usage, articulated object interaction, and in-hand rotation).

end effectors, which manifest as a kinematic gap when one attempts to execute human motion directly.

One intuitive way to address the kinematic gaps is retargeting. Optimization-based retargeting aligns a human hand trajectory to a robot hand by minimizing geometric errors, supported by rigid-body dynamics toolkits [5] and vision-based teleoperation systems that infer targets [14, 37]. Neural retargeting further improves scalability by learning fast mappings from human to robot kinematics [50]. Purely geometric objectives tend to be fragile once contact schedules matter. Recent work incorporates physical feasibility and object-centric goals, either through physics-informed retargeting that reasons about contact, or through functional retargeting that prioritizes object motion over joint-level imitation [33, 34]. Learned tracking controllers further reduce per-clip optimization by training neural policies to follow time-indexed references [29]. These approaches still leave an open question: when the reference is not jointly realizable, how should a system choose a trade-off between task success and motion style without per-sequence manual tuning.

In other domains, tracking-centric formulations have reduced reliance on manual reward engineering and improved scalability with diverse motion data [27, 32, 35, 36, 44, 51]. Inspired by this shift, we propose **ConTrack**, a reference-tracking reinforcement learning framework for contact-rich hand-object interaction in physics simulation. ConTrack introduces an online task-style mixing mechanism that adapts the trade-off during training via a dual-variable update,

and an adaptive mid-trajectory reset library that stabilizes long-horizon learning by restarting from policy-reachable simulator states near the current failure boundary. The resulting policy can deviate from the reference when required by physics while remaining anchored to the demonstrated motion and contact cues.

We present quantitative and qualitative evaluations on GRAB [42], ARCTIC [11], and DexterHand [26]. ConTrack improves success and object pose accuracy while preserving joint and contact fidelity. We also include a real-world feasibility study on a bimanual xArm7+xHand platform to verify that the learned trajectories are executable on real hardware.

In sum, our contributions are:

- An online adaptive task-style mixing mechanism based on a dual-variable update.
- An adaptive mid-trajectory reset library that reuses policy-reachable simulator states.
- A framework that transforms physically-infeasible reference motion into executable robotic trajectories, evaluated across three benchmark tiers of increasing contact complexity and validated on real-world bimanual manipulation setup.

2 Related Work

Dexterous manipulation from human demonstrations Human hand-object recordings are used as supervision for robotic dexterity, with simulation and policy learning bridging the embodiment gap from captured kinematics to robot actuation [3, 6, 11, 26, 42]. The supervision signal spans larger motion-capture datasets, touch sensing, and video or teleoperation sources [8, 9, 12–15, 18–21, 24, 30, 37, 41, 48, 52]. To make references executable, methods range from object-centric conditioning [7] and residual reinforcement learning on top of imitation [25] to tracking-centric controllers [29] and functional retargeting that prioritizes object outcomes [33]; across settings, task success and motion fidelity can disagree, so a learner must allocate limited interaction budget to resolve the trade-off.

Reference tracking in physics simulation Reference tracking has a long history in physics-based character control, where motion capture clips specify the desired behavior and reinforcement learning discovers feedback control that is robust to perturbations and missing information [27, 32, 35, 36, 44, 51]. Long-horizon training often benefits from curricula that emphasize later states and from exploration strategies that expand the reachable set of states over time [10, 39, 49]. These ideas motivate mid-trajectory training, but dexterous hand-object interaction introduces a sharp complication. Directly resetting to the reference can be physically inconsistent because contacts and object configurations are coupled through discontinuous dynamics. This mismatch motivates ConTrack’s emphasis on policy-reachable mid-trajectory states, which preserves the efficiency benefits

of mid-clip training while avoiding unrecoverable starts induced by infeasible reference configurations.

Constrained and multi-objective reinforcement learning Balancing task tracking against imitation is often implemented through a single shaped reward that mixes task terms and style terms with fixed weights [25, 29, 35, 36, 44]. In clip-conditioned tracking with contact discontinuities, fixed mixtures can require retuning as task difficulty and signal scale shift across clips and across phases of a clip [25, 29, 33]. Constrained reinforcement learning instead treats task success as a requirement and optimizes a secondary objective within the feasible set, commonly via primal–dual updates, reward-constrained policy optimization, and Lagrangian relaxations [1, 17, 45]. Closely related work studies how to learn stylistic behavior from imperfect demonstrations under a task optimality constraint [46]. ConTrack adopts this constrained viewpoint, using the normalized task constraint and an online dual update to adjust the task–style allocation during training.

3 Method

ConTrack models object tracking problem as a constraint and allocates remaining optimization pressure to style fidelity. Three components make this practical under a fixed interaction budget: an online dual controller (Sec. 3.2) that adjusts the task–style allocation, a reset library (Sec. 3.3) that restarts rollouts from policy-reachable states near the current failure boundary, and contact priors (Sec. 3.4) that anchor the style objective to reference interaction patterns.

3.1 Preliminaries

In our setup, each reference clip induces a finite-horizon MDP with horizon T and discrete time index $t \in \{0, \dots, T-1\}$. The simulator has physical state $x_t \in \mathcal{X}$, including robot joints, object poses, and their velocities. The reference provides time-indexed targets. It includes a joint trajectory $q_t^{\text{ref}} \in \mathbb{R}^D$ over D actuated joints and object pose targets for O objects, given by translations $p_{t,o}^{\text{ref}} \in \mathbb{R}^3$ and unit quaternions $\bar{q}_{t,o}^{\text{ref}} \in \mathbb{S}^3$ for $o \in \{1, \dots, O\}$. Additionally, we provide link-level contact annotations for L hand links, given by binary contact events $c_{t,o,\ell}^{\text{ref}} \in \{0, 1\}$ and object-local contact points $y_{t,o,\ell}^{\text{ref}} \in \mathbb{R}^3$ for $\ell \in \{1, \dots, L\}$. The joint reference $q_{0:T-1}^{\text{ref}}$ is obtained by retargeting captured human hand motion to the robot embodiment [5, 34, 50].

We treat the reference index as part of the state and define $s_t = (x_t, t)$. A stochastic policy $\pi(a_t | s_t)$ selects an action $a_t \in \mathbb{R}^D$. The agent receives a reward that we decompose as

$$r(s_t, a_t) = r_g(s_t, a_t) + r_s(s_t, a_t) + r_p(s_t, a_t), \quad (1)$$

where r_g measures task success through object pose tracking, r_s measures style fidelity through hand kinematics and contacts, and r_p penalizes high-frequency motion.

The policy controls the robot through residual joint targets around the reference. At frame t , the policy outputs a residual joint displacement a_t and we set the joint position target as: $q_t^{\text{tar}} = q_t^{\text{ref}} + a_t$.

This parameterization keeps exploration near the reference while allowing corrective deviations that compensate for kinematic and dynamic mismatch.

We evaluate task and style with discounted returns using a discount factor $\gamma \in (0, 1]$

$$J_g(\pi) = \mathbb{E}_\pi \left[\sum_{t=0}^{T-1} \gamma^t r_g(s_t, a_t) \right], \quad J_s(\pi) = \mathbb{E}_\pi \left[\sum_{t=0}^{T-1} \gamma^t r_s(s_t, a_t) \right]. \quad (2)$$

ConTrack maximizes style while meeting a minimum level of task success:

$$\max_{\pi} J_s(\pi) \quad \text{s.t.} \quad J_g(\pi) \geq \alpha J_g^*, \quad (3)$$

where $\alpha \in (0, 1]$ is a target task ratio and J_g^* is a clip-specific normalizer defined as a running maximum of a moving estimate of J_g during training. The ratio J_g/J_g^* is dimensionless and makes the constraint comparable across clips with different task scales.

3.2 Adaptive Task–Style Mixing

Fixed reward weights are brittle in contact-rich tracking because difficulty varies across clips and within a single clip. ConTrack therefore adapts the task–style trade-off online using a scalar controller driven by the normalized task return.

We connect Eq. 3 to an online weight controller through a Lagrangian relaxation [1, 45, 46]

$$L(\pi, \lambda) = J_s(\pi) - \lambda \left(\alpha - \frac{J_g(\pi)}{J_g^*} \right), \quad (4)$$

where $\lambda \in \mathbb{R}$ is a scalar controller state. Larger λ increases pressure to satisfy the task constraint, while smaller λ allocates more optimization capacity to style.

PPO [40] uses an advantage estimate to update the policy. We compute three advantage estimates, A_g from the task reward r_g , A_s from the style reward r_s , and A_p from the penalty reward r_p , and update the policy using a mixed advantage

$$A_{\text{mix}} = w_{\text{task}} A_g + (1 - w_{\text{task}}) A_s + A_p, \quad (5)$$

where $w_{\text{task}} \in [0, 1]$ is shared across all parallel environments of the clip. We map this scalar state to a convex mixing weight

$$w_{\text{task}} = \sigma(\lambda), \quad (6)$$

where σ is the logistic sigmoid.

Algorithm 1: CMDP-style task constraint with an online dual controller

Input: target ratio α , step size η
Data: dual state λ , running task estimate $\hat{J}_g$, running maximum J_g^*
1 while *training* **do**
2 collect rollouts under the current reset distribution // $\triangleright$ Sec. 3.3
3 $G_g \leftarrow$ discounted episodic return from r_g
4 $\hat{J}_g \leftarrow$ runningmean(G_g); $J_g^* \leftarrow \max\{J_g^*, \hat{J}_g\}$
5 $\lambda \leftarrow \lambda + \eta \left(\alpha - \hat{J}_g / J_g^* \right)$ // $\triangleright$ Eq. 7
6 $w_{\text{task}} \leftarrow \sigma(\lambda)$
7 compute A_g, A_s, A_p
8 $A_{\text{mix}} \leftarrow w_{\text{task}} A_g + (1 - w_{\text{task}}) A_s + A_p$
9 update π using A_{mix}

We update λ from an online estimate of the task return J_g . Let $\hat{J}_g$ be a running estimate of J_g under the current policy, and let J_g^* be its running maximum. The update is

$$\lambda \leftarrow \lambda + \eta \left(\alpha - \frac{\hat{J}_g}{J_g^*} \right), \quad (7)$$

where $\eta > 0$ is a step size. The ratio $\hat{J}_g / J_g^*$ is scale-free across clips. When it drops below α , Eq. 7 increases w_{task} and shifts optimization toward task tracking. When it stays above α , the update decreases w_{task} and allocates more capacity to style.

3.3 Adaptive Mid-Trajectory Reset Library

Long-horizon tracking is inefficient when training always starts from the first reference frame, and curriculum strategies that restart from later states can substantially improve coverage [10, 39]. Early failures dominate the rollout distribution, and later contact phases receive few learning updates. ConTrack addresses this with a mid-trajectory reset library indexed by reference frame. For each frame $k \in \{0, \dots, T-1\}$, the library stores a simulator state that contains robot joints and velocities and object poses and velocities. Resetting to a stored state initializes the simulator consistently and sets the current reference index to k .

Online refresh. In human motion tracking without object interaction, resets are often initialized by setting the simulator directly to the reference motion at a chosen frame [35, 36, 44]. In hand-object interaction, such resets can be physically inconsistent because contact configurations may not be realizable under the simulator dynamics. Resetting by copying the reference state can therefore produce unrecoverable starts. ConTrack instead refreshes the reset library from

Algorithm 2: Adaptive mid-trajectory reset library

Input: clip length T , temperature τ , smoothing coefficient β
Data: for all $k \in [0, T - 1]$: cached simulator states $B[k]$, EMA continuation length $\bar{\ell}_k$, best length $\ell_k^{\max}$

```

1 foreach completed episode do
2    $t_s \leftarrow$  start frame;  $t_e \leftarrow$  end frame
3   for  $k \leftarrow t_s$  to  $t_e$  do
4      $\ell \leftarrow t_e - k + 1$ ;  $\bar{\ell}_k \leftarrow (1 - \beta)\bar{\ell}_k + \beta\ell$ 
5     if  $\ell > \ell_k^{\max}$  then
6        $B[k] \leftarrow$  simulator state at frame  $k$ ;
7        $\ell_k^{\max} \leftarrow \ell$ 
8    $u_k \leftarrow \bar{\ell}_k / (T - k)$  for all  $k$  // ▷ survival ratio, Eq. 9
9   sample  $k \propto \exp(-u_k / \tau)$  // ▷ Eq. 10
10
11 reset simulator to  $B[k]$  and set reference index to  $k$ 

```

empirically reachable states collected from rollouts of the current policy. Consider a completed episode that starts at frame t_{start} and terminates at frame t_{end} . For each visited frame $k \in [t_{\text{start}}, t_{\text{end}}]$ we define the continuation length

$$\ell_k = t_{\text{end}} - k + 1. \quad (8)$$

We maintain an exponential moving average $\bar{\ell}_k$ of ℓ_k and a running maximum $\ell_k^{\max}$. If a rollout achieves $\ell_k > \ell_k^{\max}$, we overwrite the stored state at frame k with the visited simulator state and set $\ell_k^{\max} = \ell_k$. This keeps the library aligned with policy-reachable states that support longer continuations.

Difficulty-aware sampling. We define a survival ratio

$$u_k = \frac{\bar{\ell}_k}{T - k}, \quad (9)$$

and sample reset frames with

$$p(k) \propto \exp\left(-\frac{u_k}{\tau}\right), \quad (10)$$

where $\tau > 0$ is a temperature parameter. This concentrates training on segments where the policy tends to fail quickly relative to the remaining horizon. This produces rollouts that repeatedly practice entering difficult contact phases, and the reset distribution moves earlier as tracking improves.

3.4 Contact Priors

We use reference contact annotations as an auxiliary style prior. Let $c_{t,o,\ell} \in \{0, 1\}$ denote the contact event between link ℓ and object o observed in simulation at



Fig. 2: Contact priors from data. Red dots mark reference object contact points. During reinforcement learning, the style objective encourages each contacting link to match its target contact.

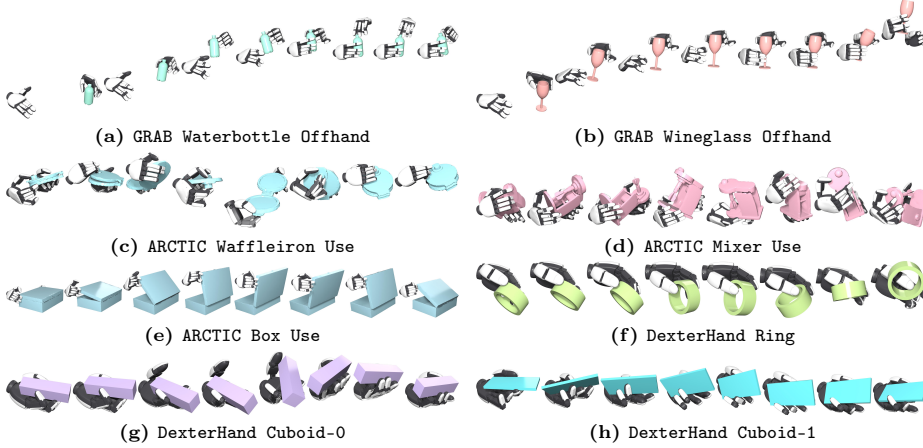


Fig. 3: Simulation snapshots.

frame t , and let $y_{t,o,\ell} \in \mathbb{R}^3$ denote the corresponding object-local contact point. The style reward includes a contact event overlap term that rewards agreement with the reference contact events $c_{t,o,\ell}^{\text{ref}}$. When both the simulator and reference indicate contact, we additionally penalize the distance between $y_{t,o,\ell}$ and the annotated contact point $y_{t,o,\ell}^{\text{ref}}$.

We map dataset contact annotations to the robot link set and use them as link-level contact targets. Fig. 2 visualizes the targets. We evaluate contact fidelity with explicit event and point metrics and ablate contact priors in Sec. 4. See appendix for annotation processing.

4 Experiments

4.1 Benchmarks

We evaluate ConTrack on reference clips spanning bimanual interaction and dexterous in-hand manipulation. Each clip defines an independent reference-tracking task with horizon T , and we train one policy per clip. We group clips into three tiers of increasing contact complexity. **GRAB** [42] focuses on bimanual rigid-object interaction. **ARCTIC** [11] adds articulated objects and

Table 1: Benchmark clips and reference-only statistics. Max disp: max object AABB diagonal displacement. End2End rot: end-to-end rotation between first and last frame. Contact occ.: fraction of frames with any hand-object contact.

Dataset	Clip	Time(s)	Hand	Articulated	Max disp(m)	End2End rot(rad)	Contact occ.
GRAB	cubemedium_offhand	3.0	bimanual	False	0.581	1.426	0.819
	cylindermedium_offhand	3.0	bimanual	False	0.805	2.675	0.942
	hammer_use	6.0	bimanual	False	0.304	0.001	0.878
	waterbottle_offhand	4.0	bimanual	False	0.470	0.439	0.808
	wineglass_offhand	3.0	bimanual	False	0.595	0.751	0.967
ARCTIC	box_use	4.1	bimanual	True	0.140	0.171	1.000
	mixer_use	10.0	bimanual	True	0.271	1.785	1.000
	notebook_use	3.0	bimanual	True	0.272	2.962	1.000
	waffleiron_use	4.0	bimanual	True	0.231	0.358	1.000
DexterHand	Cuboid_00	3.0	single-hand	False	0.098	3.082	1.000
	Cuboid_01	2.0	single-hand	False	0.060	2.849	1.000
	Cylinder	5.0	single-hand	False	0.113	2.477	1.000
	Ring	5.0	single-hand	False	0.105	2.415	1.000

multi-object contact. **DexterHand** [26] targets continuous single-hand in-hand rotation. Table 1 summarizes clip statistics computed from the reference signals. Fig. 3 shows representative simulated rollouts, and additional rollouts are shown in appendix.

4.2 Training and Evaluation Protocol

All learning-based methods are trained for 5000 PPO [40] updates per clip under a fixed simulator-step budget. Mid-trajectory resets redistribute interaction along the clip without increasing total steps. During training each method uses its own reset strategy. For evaluation we always reset from the first reference frame so that progress reflects end-to-end tracking.

We evaluate with fixed pose-break thresholds of 0.10 m translation error and 1.00 rad rotation error. Let $p_{t,o} \in \mathbb{R}^3$ and $\bar{q}_{t,o} \in \mathbb{S}^3$ denote the simulated translation and unit quaternion of object o at frame t , and let $p_{t,o}^{\text{ref}}$ and $\bar{q}_{t,o}^{\text{ref}}$ be the reference targets. Define $\Delta p_{t,o} = \|p_{t,o} - p_{t,o}^{\text{ref}}\|_2$ and $\Delta \theta_{t,o} = 2 \arccos(|\bar{q}_{t,o}^\top \bar{q}_{t,o}^{\text{ref}}|)$. An episode terminates at the first index τ such that some object violates $\Delta p_{\tau,o} > 0.10$ or $\Delta \theta_{\tau,o} > 1.00$, or at $\tau = T - 1$ if it reaches the end. We report success rate SR as the fraction of evaluation episodes with $\tau = T - 1$, and *progress* as the normalized termination index $\text{progress} = \frac{\tau}{T-1}$, where τ is the termination frame index.

4.3 Metrics

We report both task tracking and style fidelity. Task metrics are mean object translation error in meters and mean object rotation error in radians with respect to the reference object tracks. For an episode with termination index τ , we compute

$$E_{\text{pos}} = \frac{1}{(\tau + 1)O} \sum_{t=0}^{\tau} \sum_{o=1}^O \Delta p_{t,o}, \quad E_{\text{rot}} = \frac{1}{(\tau + 1)O} \sum_{t=0}^{\tau} \sum_{o=1}^O \Delta \theta_{t,o}. \quad (11)$$

Table 2: Main experiment: baseline comparison under the fixed 5000-update budget. We report mean and standard deviation across clips.

Method	Progress $\uparrow$	Obj pos (m) $\downarrow$	Obj rot (rad) $\downarrow$	Finger err (rad) $\downarrow$	Contact F1 $\uparrow$	Contact pt (m) $\downarrow$
Ours	0.899 $\pm$ 0.195	0.026 $\pm$ 0.006	0.272 $\pm$ 0.105	0.163 $\pm$ 0.014	0.784 $\pm$ 0.072	0.018 $\pm$ 0.005
ManipTrans	0.743 $\pm$ 0.292	0.012 $\pm$ 0.009	0.207 $\pm$ 0.078	0.277 $\pm$ 0.089	0.620 $\pm$ 0.068	0.030 $\pm$ 0.010
DexMachina	0.246 $\pm$ 0.052	0.038 $\pm$ 0.018	0.348 $\pm$ 0.121	0.147 $\pm$ 0.016	0.708 $\pm$ 0.041	0.024 $\pm$ 0.003
SPIDER	0.444 $\pm$ 0.341	0.201 $\pm$ 0.113	1.104 $\pm$ 0.599	0.157 $\pm$ 0.019	0.191 $\pm$ 0.225	0.036 $\pm$ 0.011

Style metrics include mean arm and finger joint position errors in radians with respect to the reference joint trajectories. Let $q_t \in \mathbb{R}^D$ denote the robot joint configuration and q_t^{ref} the reference target at frame t . We report the mean absolute joint error, aggregated over frames up to τ , and separately averaged over arm and finger joints.

To quantify contact fidelity, we aggregate contact event classification over time and over annotated link-object pairs. Let $c_{t,o,\ell} \in \{0, 1\}$ denote whether link ℓ contacts object o at frame t in simulation, and let $c_{t,o,\ell}^{\text{ref}}$ denote the reference contact annotation. We compute precision, recall, and F1 from the aggregated confusion matrix of $c_{t,o,\ell}$ against $c_{t,o,\ell}^{\text{ref}}$. We also report contact point error in meters in object-local coordinates on matched contact events. Since early termination shortens the effective horizon, we always report progress alongside error metrics.

4.4 Baselines and Ablations

We compare ConTrack against three baselines that resolve the task-style conflict in different ways: ManipTrans [25]¹, DexMachina [33]², and SPIDER [34]. ManipTrans starts from an imitative controller and uses residual reinforcement learning to correct execution errors under contact dynamics. DexMachina targets functional retargeting with object-centric goals. SPIDER constructs a physically feasible retargeted trajectory without policy learning. We evaluate all methods under the same termination rule and metrics.

We study three ablations that isolate the mechanisms responsible for stability and trade-off control. The first ablation replaces adaptive task-style mixing in Sec. 3.2 with fixed mixing weights. The second ablation replaces the reset library in Sec. 3.3 with simpler reset strategies that do not adapt to the current failure boundary. The third ablation removes contact priors from the style objective. Unless stated otherwise, all variants use the same observations, action space, and termination rule.

4.5 Main Results

Table 2 reports our main experiment: a baseline comparison under the fixed 5000-update budget. Table 3 shows ConTrack results per clip across tiers. Con-

¹ We reimplement ManipTrans to match our environment and evaluation setup.

² DexMachina’s code is not fully open source, so we integrated their Virtual Object Controllers into our codebase.

Table 3: Ours: tracking metrics on the benchmark set.

Dataset	Clip	Progress $\uparrow$	Obj pos (m) $\downarrow$	Obj rot (rad) $\downarrow$	Finger err (rad) $\downarrow$	Contact F1 $\uparrow$	Contact pt (m) $\downarrow$
GRAB	cubemedium_offhand	0.994	0.023	0.226	0.166	0.765	0.015
	cylindermedium_offhand	1.000	0.043	0.349	0.186	0.693	0.016
	hammer_use	0.984	0.026	0.175	0.195	0.875	0.014
	waterbottle_offhand	1.000	0.018	0.141	0.159	0.745	0.012
	wineglass_offhand	1.000	0.029	0.200	0.170	0.915	0.011
ARCTIC	box_use	1.000	0.020	0.110	0.152	0.865	0.021
	mixer_use	0.853	0.026	0.231	0.167	0.653	0.023
	notebook_use	0.993	0.030	0.234	0.160	0.731	0.023
	waffleiron_use	0.931	0.031	0.262	0.148	0.766	0.020
DexterHand	Cuboid_00	0.745	0.029	0.381	0.152	0.766	0.020
	Cuboid_01	0.987	0.025	0.400	0.147	0.781	0.014
	Cylinder	0.928	0.018	0.349	0.150	0.856	0.021
	Ring	0.272	0.024	0.474	0.169	0.789	0.025
GRAB	Avg	0.996	0.028	0.218	0.175	0.798	0.013
ARCTIC	Avg	0.944	0.027	0.209	0.157	0.754	0.022
DexterHand	Avg	0.733	0.024	0.401	0.155	0.798	0.020
Overall	Avg	0.899	0.026	0.272	0.163	0.784	0.018

Table 4: Ablation for adaptive vs. fixed task–style mixing.

Variant	Progress $\uparrow$	Obj pos (m) $\downarrow$	Obj rot (rad) $\downarrow$	Finger err (rad) $\downarrow$	Contact F1 $\uparrow$	Contact pt (m) $\downarrow$
Adaptive (ours)	0.899 $\pm$ 0.195	0.026 $\pm$ 0.006	0.272 $\pm$ 0.105	0.163 $\pm$ 0.014	0.784 $\pm$ 0.072	0.018 $\pm$ 0.005
Fixed task	0.764 $\pm$ 0.172	0.026 $\pm$ 0.004	0.250 $\pm$ 0.031	0.157 $\pm$ 0.021	0.679 $\pm$ 0.131	0.022 $\pm$ 0.004
Fixed 1:1	0.868 $\pm$ 0.107	0.029 $\pm$ 0.008	0.297 $\pm$ 0.029	0.165 $\pm$ 0.028	0.701 $\pm$ 0.106	0.023 $\pm$ 0.002

Track reaches near-complete progress on GRAB and strong progress on ARCTIC, while DexterHand remains challenging within the same budget. ManipTrans attains the lowest object pose error but drifts in finger motion and contacts. DexMachina preserves finger motion and contact fidelity more closely but makes limited progress under the same budget. SPIDER stays close in kinematics but struggles to maintain object tracking once contact dynamics dominate. ConTrack narrows this gap by adapting the task–style balance online and by stabilizing long-horizon learning through mid-trajectory resets.

Failure analysis for DexterHand Ring. DexterHand Ring is the hardest clip under our fixed budget and failures concentrate in a rotation-dominant phase rather than emerging from gradual drift. With longer training the clip becomes feasible under the same thresholds. See appendix for extended-budget results and termination statistics.

4.6 Ablations

Adaptive vs. Fixed Task–Style Mixing Table 4 shows that adaptive mixing improves progress while preserving contact fidelity compared to fixed mixing variants. Fig. 4 instantiates this trade-off and plots each trained policy as a point in the plane of progress and contact F1. For a policy π , let $g(\pi) \in [0, 1]$ denote its mean progress and let $f(\pi) \in [0, 1]$ denote its mean contact F1 against the reference contact annotations. We say π_1 dominates π_2 if $g(\pi_1) \geq g(\pi_2)$ and $f(\pi_1) \geq f(\pi_2)$ with at least one strict inequality. Sweeping the target ratio α traces an empirical Pareto frontier across the tested range. In our sweep,

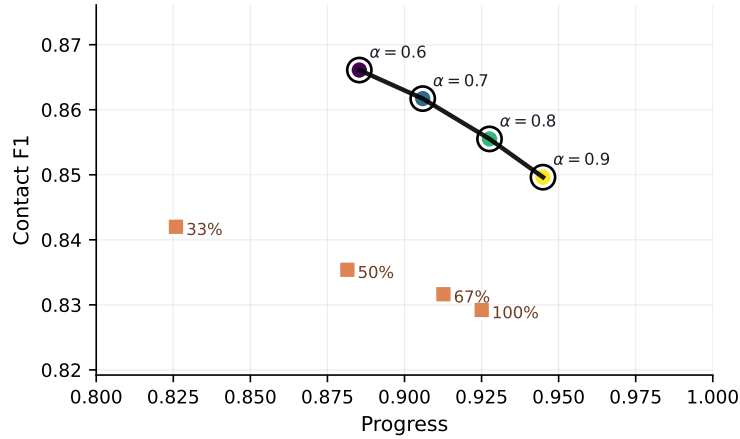


Fig. 4: Task-style trade-off controlled by the target ratio α . Each round point is a policy trained with a fixed α , and each square point is a policy trained with a fixed task-style mixing weight, with task weight annotated. Sweeping α traces an empirical Pareto-optimal frontier between progress and contact fidelity, and fixed reward mixing baselines fall inside this frontier.

Table 5: Ablation for reset library trick.

Variant	Progress $\uparrow$	Obj pos (m) $\downarrow$	Obj rot (rad) $\downarrow$	Finger err (rad) $\downarrow$	Contact F1 $\uparrow$	Contact pt (m) $\downarrow$
Adaptive library (Ours)	0.899 $\pm$ 0.195	0.026 $\pm$ 0.006	0.272 $\pm$ 0.105	0.163 $\pm$ 0.014	0.784 $\pm$ 0.072	0.018 $\pm$ 0.005
Reset from start	0.700 $\pm$ 0.270	0.039 $\pm$ 0.010	0.370 $\pm$ 0.045	0.168 $\pm$ 0.011	0.739 $\pm$ 0.088	0.017 $\pm$ 0.002
Reset uniform	0.727 $\pm$ 0.348	0.033 $\pm$ 0.010	0.298 $\pm$ 0.080	0.152 $\pm$ 0.016	0.714 $\pm$ 0.119	0.021 $\pm$ 0.004

fixed reward mixing policies are dominated by points on this frontier, which means that for some achieved progress level, the adaptive update attains higher naturalness as measured by contact fidelity under the same budget.

Reset Library Ablation Table 5 shows that the reset library improves progress under the same budget compared to always resetting from the first frame and to a uniform mid-clip reset. Fig. 5 tracks the distribution of sampled reset frames throughout training. Early updates concentrate resets near the end of the clip where short-horizon tracking is feasible. As learning expands the reachable suffix, resets move earlier, so training continually targets the current transition point between success and failure.

Contact Prior Reward Ablation Table 6 indicates that contact priors provide complementary supervision for contact-rich tracking. Removing either contact term reduces progress or contact accuracy, suggesting that the priors help resolve contact dynamics that are not fully specified by object pose errors alone.

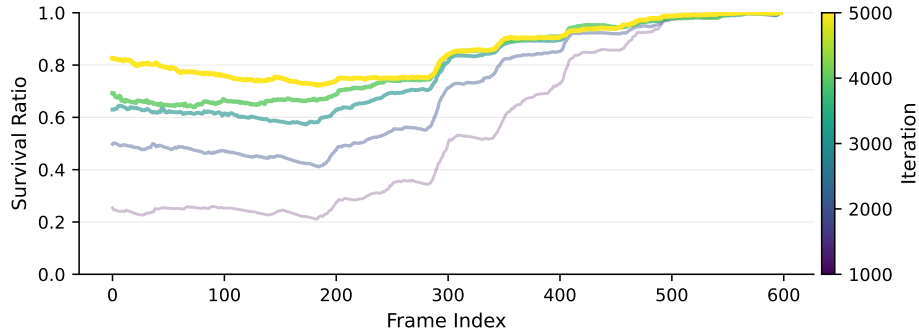


Fig. 5: Reset library dynamics over the 5000 training updates. As the policy learns to track further into the clip, the set of reachable mid-trajectory states expands toward earlier frames, allowing resets to move backward and concentrate learning on the remaining difficult segments.

Table 6: Ablation for contact prior reward.

Variant	Progress $\uparrow$	Obj pos (m) $\downarrow$	Obj rot (rad) $\downarrow$	Finger err (rad) $\downarrow$	Contact F1 $\uparrow$	Contact pt (m) $\downarrow$
Full (Ours)	0.899 $\pm$ 0.195	0.026 $\pm$ 0.006	0.272 $\pm$ 0.105	0.163 $\pm$ 0.014	0.784 $\pm$ 0.072	0.018 $\pm$ 0.005
w/o contact reward	0.861 $\pm$ 0.173	0.032 $\pm$ 0.010	0.320 $\pm$ 0.044	0.168 $\pm$ 0.019	0.699 $\pm$ 0.092	0.020 $\pm$ 0.005
w/o contact distance reward	0.868 $\pm$ 0.057	0.032 $\pm$ 0.011	0.288 $\pm$ 0.071	0.149 $\pm$ 0.016	0.753 $\pm$ 0.089	0.023 $\pm$ 0.005

4.7 Real-world Feasibility

We study real-world feasibility on a tabletop bimanual platform with two xArm7 arms and two xHands. For each clip, we stream the policy-predicted joint references from the simulator to a separate real-time controller over TCP. Each side tracks 7 arm joints and 12 hand joints, with outer reference updates at 0.15 s and an internal arm control loop running at 250 Hz. Fig. 6 shows representative executions, and additional experiments are shown in appendix.

5 Conclusion

Limitations. Our constrained formulation uses an online normalization based on a running maximum of task return estimates, which provides stable scaling in practice but does not strictly guarantee satisfaction of the constraint. Contact priors rely on the availability and quality of contact annotations, and tracking remains challenging for the hardest in-hand manipulation clips under a limited interaction budget. Finally, our hardware study currently focuses on executing joint commands on the real arm-hand platform. Extending the system with richer perception and tighter sim-to-real alignment could further broaden robustness in real-world deployment.

In summary, we presented ConTrack, a constrained reference-tracking reinforcement learning framework for long-horizon, contact-rich hand-object interaction in physics simulation. ConTrack separates task success from style fidelity

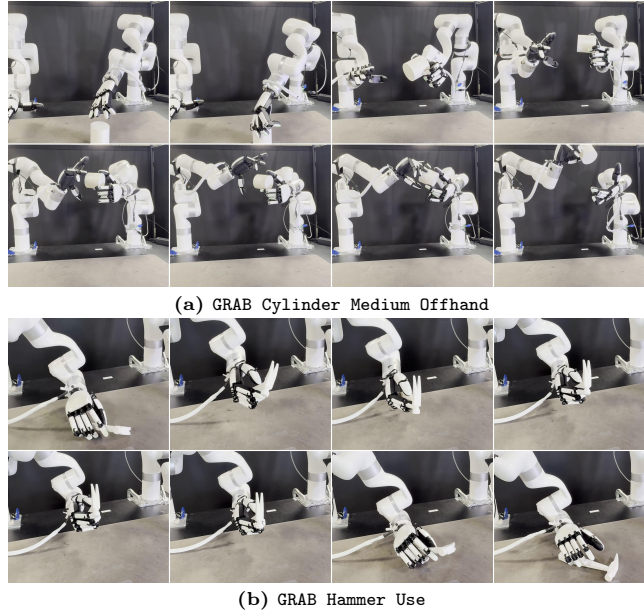


Fig. 6: Real-world experiment snapshots on the bimanual xArm7+xHand platform.

and controls their trade-off online through a scalar dual controller. To address the instability of learning long clips from the first frame, ConTrack maintains an adaptive mid-trajectory reset library that refreshes entries from policy-reachable simulator states, and samples start frames to focus learning on difficult segments. Across bimanual interaction and in-hand manipulation benchmarks under a fixed training budget, ConTrack improves progress and contact fidelity while retaining accurate object tracking, and exposes an interpretable target ratio that traces a smooth task-style trade-off frontier.

References

1. Achiam, J., Held, D., Tamar, A., Abbeel, P.: Constrained policy optimization. In: Precup, D., Teh, Y.W. (eds.) *Proceedings of the 34th International Conference on Machine Learning*. *Proceedings of Machine Learning Research*, vol. 70, pp. 22–31. PMLR (06–11 Aug 2017)
2. Barreiros, J., Beaulieu, A., Bhat, A., Cory, R., Cousineau, E., Dai, H., Fang, C.H., Hashimoto, K., Irshad, M.Z., Itkina, M., et al.: A careful examination of large behavior models for multitask dexterous manipulation. *arXiv preprint arXiv:2507.05331* (2025)
3. Brahmbhatt, S., Tang, C., Twigg, C.D., Kemp, C.C., Hays, J.: ContactPose: A dataset of grasps with object contact and hand pose. In: *The European Conference on Computer Vision (ECCV)* (August 2020)
4. Cai, X., Qiu, R.Z., Chen, G., Wei, L., Liu, I., Huang, T., Cheng, X., Wang, X.: In-n-on: Scaling egocentric manipulation with in-the-wild and on-task data. *arXiv preprint arXiv:2511.15704* (2025)
5. Carpentier, J., Saurel, G., Buondonno, G., Mirabel, J., Lamiroux, F., Stasse, O., Mansard, N.: The pinocchio c++ library – a fast and flexible implementation of rigid body dynamics algorithms and their analytical derivatives. In: *IEEE International Symposium on System Integrations (SII)* (2019)
6. Chao, Y.W., Yang, W., Xiang, Y., Molchanov, P., Handa, A., Tremblay, J., Narang, Y.S., Van Wyk, K., Iqbal, U., Birchfield, S., Kautz, J., Fox, D.: DexYCB: A benchmark for capturing hand grasping of objects. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021)
7. Chen, Y., Wang, C., Yang, Y., Liu, C.K.: Object-centric dexterous manipulation from human motion data. In: *CoRL* (2024)
8. Cheng, X., Li, J., Yang, S., Yang, G., Wang, X.: Open-television: Teleoperation with immersive active visual feedback. *arXiv preprint arXiv:2407.01512* (2024)
9. Chi, C., Xu, Z., Pan, C., Cousineau, E., Burchfiel, B., Feng, S., Tedrake, R., Song, S.: Universal manipulation interface: In-the-wild robot teaching without in-the-wild robots. *arXiv preprint arXiv:2402.10329* (2024)
10. Ecoffet, A., Huizinga, J., Lehman, J., Stanley, K.O., Clune, J.: Go-explore: a new approach for hard-exploration problems. *arXiv preprint arXiv:1901.10995* (2019)
11. Fan, Z., Taheri, O., Tzionas, D., Kocabas, M., Kaufmann, M., Black, M.J., Hilliges, O.: ARCTIC: A dataset for dexterous bimanual hand-object manipulation. In: *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2023)
12. Fu, R., Zhang, D., Jiang, A., Fu, W., Funk, A., Ritchie, D., Sridhar, S.: Gigahands: A massive annotated dataset of bimanual hand activities. In: *Proceedings of the Computer Vision and Pattern Recognition Conference*. pp. 17461–17474 (2025)
13. Geng, H., Wang, F., Wei, S., Li, Y., Wang, B., An, B., Cheng, C.T., Lou, H., Li, P., Wang, Y.J., Liang, Y., Goetting, D., Xu, C., Chen, H., Qian, Y., Geng, Y., Mao, J., Wan, W., Zhang, M., Lyu, J., Zhao, S., Zhang, J., Zhang, J., Zhao, C., Lu, H., Ding, Y., Gong, R., Wang, Y., Kuang, Y., Wu, R., Jia, B., Sferrazza, C., Dong, H., Huang, S., Wang, Y., Malik, J., Abbeel, P.: Roboverse: Towards a unified platform, dataset and benchmark for scalable and generalizable robot learning. *arXiv preprint arXiv:2504.18904* (2025)
14. Handa, A., Van Wyk, K., Yang, W., Liang, J., Chao, Y.W., Wan, Q., Birchfield, S., Ratliff, N., Fox, D.: Dexpilot: Vision-based teleoperation of dexterous robotic hand-arm system. In: *ICRA* (2020)

15. Hsieh, J., Tu, K.H., Hung, K.H., Ke, T.W.: Dexman: Learning bimanual dexterous manipulation from human and generated videos. arXiv preprint arXiv:2510.08475 (2025)
16. Intelligence, P., Black, K., Brown, N., Darpinian, J., Dhabalia, K., Driess, D., Esmail, A., Equi, M., Finn, C., Fusai, N., et al.: pi0.5: a vision-language-action model with open-world generalization. arXiv preprint arXiv:2504.16054 (2025)
17. Ishihara, Y., Takasugi, N., Kawakami, K., Kinoshita, M., Aoyama, K.: Constraints as rewards: Reinforcement learning for robots without reward functions. arXiv preprint arXiv:2501.04228 (2025)
18. Jian, J., Liu, X., Li, M., Hu, R., Liu, J.: Affordpose: A large-scale dataset of hand-object interactions with affordance-driven hand pose. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 14713–14724 (October 2023)
19. Jiang, G., Chang, H., Qiu, R.Z., Liang, Y., Ji, M., Zhu, J., Dong, Z., Zou, X., Wang, X.: Gsworld: Closed-loop photo-realistic simulation suite for robotic manipulation. arXiv preprint arXiv:2510.20813 (2025)
20. Jiang, G., Liang, Y., Ye, J., Huang, J.Y., Jing, C., Duan, R., Abbeel, P., Wang, X., Zou, X.: Cross-hand latent representation for vision-language-action models. arXiv preprint arXiv:2603.10158 (2026)
21. Jiang, N., Liu, T., Cao, Z., Cui, J., Chen, Y., Wang, H., Zhu, Y., Huang, S.: Full-body articulated human-object interaction. In: ICCV (2023)
22. Kareer, S., Patel, D., Punamiya, R., Mathur, P., Cheng, S., Wang, C., Hoffman, J., Xu, D.: Egomimic: Scaling imitation learning via egocentric video. In: ICRA. IEEE (2025)
23. Kareer, S., Pertsch, K., Darpinian, J., Hoffman, J., Xu, D., Levine, S., Finn, C., Nair, S.: Emergence of human to robot transfer in vision-language-action models. arXiv preprint arXiv:2512.22414 (2025)
24. Kim, J., Kim, J., Na, J., Joo, H.: ParaHome: Parameterizing Everyday Home Activities Towards 3D Generative Modeling of Human-Object Interactions . In: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1816–1828. IEEE Computer Society (2025)
25. Li, K., Li, P., Liu, T., Li, Y., Huang, S.: Maniptrans: Efficient dexterous bimanual manipulation transfer via residual learning. In: CVPR (2025)
26. Liang, Y., Xu, S., Zhang, Y., Zhan, B., Zhang, H., Liu, L.: DexterCap: An affordable and automated system for capturing dexterous hand-object manipulation. arXiv preprint arXiv:2601.05844 (2026)
27. Liao, Q., Truong, T.E., Huang, X., Gao, Y., Tevet, G., Sreenath, K., Liu, C.K.: Beyondmimic: From motion tracking to versatile humanoid control via guided diffusion. arXiv preprint arXiv:2508.08241 (2025)
28. Liu, S., Wu, L., Li, B., Tan, H., Chen, H., Wang, Z., Xu, K., Su, H., Zhu, J.: Rdt-1b: a diffusion foundation model for bimanual manipulation. arXiv preprint arXiv:2410.07864 (2024)
29. Liu, X., Adalibieke, J., Han, Q., Qin, Y., Yi, L.: Dextrack: Towards generalizable neural tracking control for dexterous manipulation from human references. In: ICLR (2025)
30. Lu, J., Huang, C.H.P., Bhattacharya, U., Huang, Q., Zhou, Y.: Humoto: A 4d dataset of mocap human object interactions. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 10886–10897 (October 2025)

31. Luo, H., Wang, Y., Zhang, W., Zheng, S., Xi, Z., Xu, C., Xu, H., Yuan, H., Zhang, C., Wang, Y., et al.: Being-h0. 5: Scaling human-centric robot learning for cross-embodiment generalization. arXiv preprint arXiv:2601.12993 (2026)
32. Luo, Z., Yuan, Y., Wang, T., Li, C., Chen, S., Castaneda, F., Cao, Z.A., Li, J., Minor, D., Ben, Q., et al.: Sonic: Supersizing motion tracking for natural humanoid whole-body control. arXiv preprint arXiv:2511.07820 (2025)
33. Mandi, Z., Hou, Y., Fox, D., Narang, Y., Mandlekar, A., Song, S.: Dexmachina: Functional retargeting for bimanual dexterous manipulation. arXiv preprint arXiv:2505.24853 (2025)
34. Pan, C., Wang, C., Qi, H., Liu, Z., Bharadhwaj, H., Sharma, A., Wu, T., Shi, G., Malik, J., Hogan, F.: Spider: Scalable physics-informed dexterous retargeting. arXiv preprint arXiv:2511.09484 (2025)
35. Peng, X.B., Abbeel, P., Levine, S., van de Panne, M.: Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Trans. Graph.* **37**(4), 143:1–143:14 (Jul 2018)
36. Peng, X.B., Guo, Y., Halper, L., Levine, S., Fidler, S.: Ase: Large-scale reusable adversarial skill embeddings for physically simulated characters. *ACM Trans. Graph.* **41**(4) (Jul 2022)
37. Qin, Y., Yang, W., Huang, B., Van Wyk, K., Su, H., Wang, X., Chao, Y.W., Fox, D.: Anyteleop: A general vision-based dexterous robot arm-hand teleoperation system. arXiv preprint arXiv:2307.04577 (2023)
38. Qiu, R.Z., Yang, S., Cheng, X., Chawla, C., Li, J., He, T., Yan, G., Yoon, D.J., Hoque, R., Paulsen, L., et al.: Humanoid policy~ human policy. In: CoRL (2025)
39. Resnick, C., Raileanu, R., Kapoor, S., Peysakhovich, A., Cho, K., Bruna, J.: Back-play: "man muss immer umkehren". arXiv preprint arXiv:1807.06919 (2018)
40. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O.: Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347 (2017)
41. Song, Y.R., Li, J., Fu, R., Murphy, D., Zhou, K., Shiv, R., Li, Y., Xiong, H., Owens, C.E., Du, Y., et al.: Opentouch: Bringing full-hand touch to real-world interaction. arXiv preprint arXiv:2512.16842 (2025)
42. Taheri, O., Ghorbani, N., Black, M.J., Tzionas, D.: GRAB: A dataset of whole-body human grasping of objects. In: European Conference on Computer Vision (ECCV) (2020)
43. Team, S.A.: Spirit-v1.5: Clean data is the enemy of great robot foundation models. Spirit AI Blog (2026)
44. Tessler, C., Guo, Y., Nabati, O., Chechik, G., Peng, X.B.: Maskedmimic: Unified physics-based character control through masked motion inpainting. *ACM Transactions on Graphics (TOG)* (2024)
45. Tessler, C., Mankowitz, D.J., Mannor, S.: Reward constrained policy optimization. arXiv preprint arXiv:1805.11074 (2018)
46. Wen, K., Li, C., He, J., Hutter, M.: Constrained style learning from imperfect demonstrations under task optimality. arXiv preprint arXiv:2507.09371 (2025)
47. Wen, R., Chen, G., Cui, Z., Du, M., Gou, Y., Han, Z., Huang, L., Lei, M., Li, Y., Li, Z., et al.: Gr-dexter technical report. arXiv preprint arXiv:2512.24210 (2025)
48. Xu, M., Zhang, H., Hou, Y., Xu, Z., Fan, L., Veloso, M., Song, S.: Dexumi: Using human hand as the universal manipulation interface for dexterous manipulation. arXiv preprint arXiv:2505.21864 (2025)
49. Xu, S., Chao, Y.W., Bian, L., Mousavian, A., Wang, Y.X., Gui, L.Y., Yang, W.: Dexplore: Scalable neural control for dexterous manipulation from reference-scoped exploration. arXiv preprint arXiv:2509.09671 (2025)

- 50. Yin, Z.H., Wang, C., Pineda, L., Bodduluri, K., Wu, T., Abbeel, P., Mukadam, M.: Geometric retargeting: A principled, ultrafast neural hand retargeting algorithm. In: IROS (2025)
- 51. Ze, Y., Chen, Z., Araujo, J.P., Cao, Z.a., Peng, X.B., Wu, J., Liu, K.: Twist: Teleoperated whole-body imitation system. In: Lim, J., Song, S., Park, H.W. (eds.) Proceedings of The 9th Conference on Robot Learning. Proceedings of Machine Learning Research, vol. 305, pp. 2143–2154. PMLR (27–30 Sep 2025)
- 52. Zhao, S., Zhu, X., Chen, Y., Li, C., Xie, L., Zhang, X., Ding, M., Tomizuka, M.: Dexh2r: Task-oriented dexterous manipulation from human to robots. arXiv preprint arXiv:2411.04428 (2026)
- 53. Zheng, R., Niu, D., Xie, Y., Wang, J., Xu, M., Jiang, Y., Castañeda, F., Hu, F., Tan, Y.L., Fu, L., et al.: Egoscale: Scaling dexterous manipulation with diverse egocentric human data. arXiv preprint arXiv:2602.16710 (2026)