

A Method Details

Trajectory augmentation via Cartesian DMP. To provide offline data for coordination pre-training, we generate rollouts from a single reference demonstration instead of collecting full arm-hand demonstrations for every rollout. The reference contains an arm end-effector trajectory $\{\mathbf{p}_t^{ref}, \mathbf{q}_t^{ref}\}_{t=0}^T$ and a hand-skill activation schedule $\{\alpha_t^{ref}\}_{t=0}^T$ collected with a Gello interface [1]. We fit Cartesian Dynamic Movement Primitives (DMP) [2] to the demonstrated arm trajectory. The translational dynamics follow a second-order spring-damper system with a learned forcing term $\Phi_p(s)$, while the rotational dynamics evolve on the quaternion manifold:

$$\lambda^2 \ddot{\mathbf{p}} = K_p(\mathbf{g}_p - \mathbf{p}) - \lambda D_p \dot{\mathbf{p}} + \Phi_p(s), \quad (1)$$

$$\lambda \dot{\boldsymbol{\eta}} = K_o(2 \log(\mathbf{g}_o * \bar{\mathbf{q}})) - D_o \boldsymbol{\eta} + \Phi_o(s), \quad (2)$$

$$\lambda \dot{\mathbf{q}} = \frac{1}{2} \boldsymbol{\eta} * \mathbf{q}. \quad (3)$$

Here λ is the temporal scaling factor, s is a canonical phase variable, $\mathbf{g}_{\{p,o\}}$ are goal configurations, and $K_{\{p,o\}}, D_{\{p,o\}}$ are stiffness and damping gains. The forcing terms $\Phi_{p,o}(s)$ encode the demonstrated motion profile. To augment the data, we fix the goal and randomize the initial pose $[\mathbf{p}_0, \mathbf{q}_0]$ within the task workspace before integrating Eq. (1)–(3). Since DMP modifies only the arm motion, the hand activation follows the demonstration by phase alignment, $\alpha_t = \alpha^{ref}(s_t)$.

We execute the generated trajectories on the physical robot to collect state-action transitions. The arm action label is the clipped displacement between the next DMP waypoint x_{t+1}^* and the current robot state x_t :

$$\Delta x_t = \text{clip}(x_{t+1}^* - x_t, -1, 1). \quad (4)$$

When $\alpha_t > \tau$, the frozen hand-skill policy π_{skill} generates finger commands following Eq. (??); otherwise the hand remains at θ^{reset} . The collected transitions form $\mathcal{D}_{\text{dmp}}$, which is used for offline pre-training and as seed data in the online replay buffer.

DMP rollout coverage analysis. Figure. 1 compares human demonstrations with DMP-augmented rollouts for coordination pre-training. Human demonstrations capture realistic executions but are slow to collect and vary with operator timing. Starting from one reference demonstration, DMP randomizes the initial arm pose while preserving the demonstrated target pose and skill-activation progress, producing systematic approach-state coverage. This provides a cheaper and more consistent warm start for online fine-tuning than repeatedly collecting full arm-hand demonstrations.

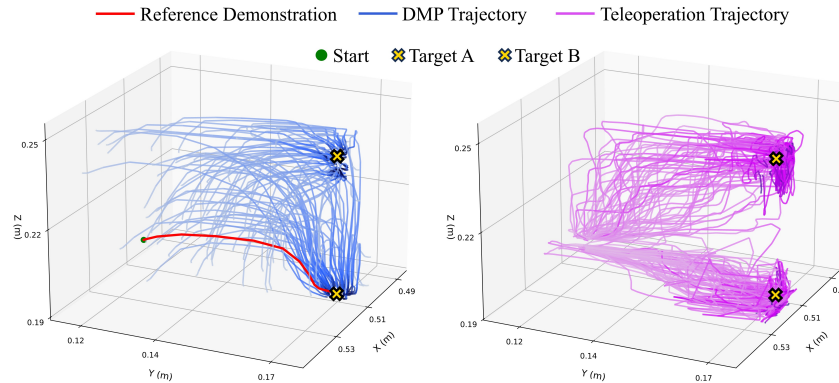


Figure 1: Trajectory distribution from DMP rollouts and human demonstrations.

26 B Experiment Details

27 B.1 Hardware Setup

28 Each task uses a Franka Research 3 arm equipped with either a Sharpa hand or an XHand. The
 29 setup includes wrist RGB observations from an Intel RealSense D405 and a side-view D415 when
 30 needed. Figure 2 summarizes the task setup and camera views across all six tasks.

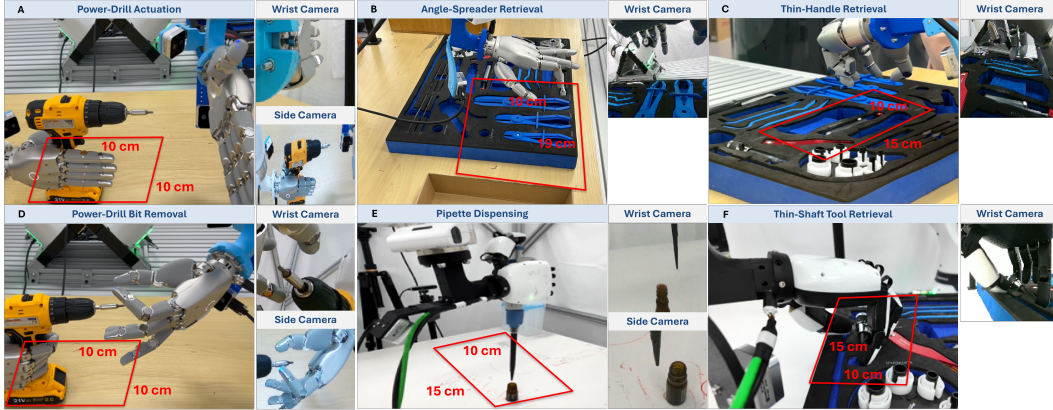


Figure 2: **Task setup and camera views.** The figure shows the six task setups (A–F), the workspace regions used for training and evaluation, and the corresponding wrist and side camera observations.

31 B.2 Task Descriptions

32 The descriptions below define the task sequence, success condition, and main difficulty for each
 33 task, with (A)–(D) serving as Sharpa main validation tasks and (E)–(F) as XHand portability tasks.

34 **(A) Power-Drill Actuation** (Precise Tool-Use) requires maintaining a stable functional grasp on the
 35 drill, aligning the screwdriver bit with the target screw, and actuating the trigger to tighten the screw
 36 into the fixture. Success requires the target screw to be tightened into the fixture. The main difficulty
 37 is that axial alignment must be established before activation, since premature trigger engagement or
 38 excessive contact force can shift the hand pose and cause misalignment.

39 **(B) Angle-Spreader Retrieval** (Constrained-Space Manipulation) requires extracting the target tool
 40 from a constrained tray where surrounding slots limit direct grasping. Success requires the target
 41 tool to be fully extracted. The main difficulty is creating enough fingertip contact to expose or lift
 42 the tool before forming a stable grasp.

43 **(C) Thin-Handle Retrieval** (Constrained-Space Manipulation) requires extracting a thin-handle
 44 tool from a narrow opening where limited clearance prevents direct grasping. Success requires the
 45 target tool to be fully extracted. The main difficulty is executing non-prehensile levering to create
 46 access, followed by a precision grasp while avoiding collisions with the surrounding tray.

47 **(D) Power-Drill Bit Removal** (Contact-Rich Rotation) requires grasping the drill chuck, applying
 48 sustained rotational motion to loosen it, and removing the drill bit. Success requires the drill bit to
 49 be removed from the chuck. The main difficulty is maintaining stable multi-finger contact under
 50 friction and changing contact forces during rotation.

51 **(E) Pipette Dispensing** (Precise Tool-Use) requires aligning the pipette tip with a vial and depress-
 52 ing the plunger. Success requires dispensing liquid into the vial without collision or leakage. The
 53 main difficulty is applying thumb force while preserving grasp stability and keeping the tip aligned
 54 with the vial opening.

55 **(F) Thin-Shaft Tool Retrieval** (Constrained-Space Manipulation) requires extracting a thin-shaft
 56 tool embedded in foam with limited clearance. Success requires the target tool to be fully extracted.

By using a contact sequence similar to Thin-Handle Retrieval with the XHand embodiment, this task tests cross-hand deployment for constrained-space retrieval.

B.3 Evaluation Protocol

Main Evaluation. For main evaluation, each task is tested over 50 trials. For every task, object poses are randomized within the red-box workspace shown in Figure. 2, and robot reset poses follow the train-time reset protocol. This evaluates repeatable functional execution within the trained workspace rather than zero-shot workspace generalization. Full success requires the task-specific functional outcome to be completed: tightening the target screw, fully extracting the target tool, removing the drill bit, or dispensing into the vial. Reported success rate is the mean success credit over 50 trials: full success receives credit 1 and failure receives credit 0. For rare baseline trials that reach the task-specific final contact stage but terminate at the evaluation horizon before completing the functional outcome, we assign credit 0.5. This distinguishes horizon-limited near-completion from failures that never establish the required contact.

Perturbation Evaluation. The robustness evaluation uses 20 perturbation trials per task. The perturbations include randomized object poses across tasks, approach-phase displacements for alignment-sensitive tasks, and drill-chuck re-tightening for *Power-Drill Bit Removal*. Because these perturbations place the policy in off-nominal contact states, success requires recovery through re-alignment or delayed hand-skill activation rather than replaying a nominal trajectory.

C Implementation Details

C.1 Policy Training Details

Across Sharpa and XHand, coordination policies share the same high-level action interface and RL pipeline. Task-dependent differences are the active camera streams and online interaction required during fine-tuning, with *Thin-Shaft Tool Retrieval* using the wrist camera and *Pipette Dispensing* using both wrist and side cameras.

Table 1 summarizes common policy hyperparameters, including control frequency, episode length, network size, optimizer, discount factor, and image augmentation.

Parameter	Value	Parameter	Value
Observation space	wrist and/or side, tcp_pose, tcp_vel, tcp_ft/hand_pose, hand_ft	Environment update frequency	20 Hz
Action space	6D twist, skill modulation weight	Proprio encoder size	64
Reward function	Binary classifier	Policy MLP size	256×256
Classifier views	wrist and/or side	Discount factor	0.98
Classifier accuracy	95%	Optimizer	Adam
Initial offline demonstrations	100	Learning rate	3e-4
Offline global BC weight λ_{gbc}	0.5	IAW imitation-loss weight λ_I	0.1
Reset method	Scripted/Human reset	Image augmentation	Random crop
Max episode length	2000 environment steps		

Table 1: Common policy training hyperparameters shared across tasks.

Hand skill acquisition begins with one kinesthetic reference trajectory, which usually takes less than five minutes to record per task. Because the hand skill reward uses fingertip force feedback, we avoid rapid or jerky contact motions during teaching, which can introduce noisy force and contact measurements. In practice, we teach the contact sequence slowly to obtain consistent tactile signals, and temporally resample the recorded prior when faster reference execution is desired. During early residual-RL training, we initialize the residual action near zero and use a small residual action scale so the policy stays close to the kinesthetic prior while collecting initial contact data. The residual scale is then increased to allow reward-driven adaptation to real contact dynamics.

For arm-hand coordination learning, Table 2 reports the wall-clock time spent on reward-classifier data collection, automated DMP rollout collection, and online coordination-policy fine-tuning. For

the reward-classifier column, the reported time includes both positive and negative frame collection. Across the six task datasets, collection takes 4–13 minutes per task and includes 10–35 positive frames with roughly 1.9k–7.9k negative frames. The imbalance mainly comes from sampling failure frames from rollout videos, while positive frames are short clips of completed outcomes. In practice, classifier quality depends on whether the wrist or side camera can observe the task-completion state and distinguish it from near-success failures. When this visibility condition holds, classifier data collection is inexpensive.

Task	Reward Classifier Data Collection (mins)	DMP Data Collection (mins)	Coordination Policy Training (mins)
Power-Drill Actuation	4	34	60
Angle-Spreader Retrieval	13	13	55
Thin-Handle Retrieval	11	17	60
Power-Drill Bit Removal	7	21	22
Pipette Dispensing	10	14	60
Thin-Shaft Tool Retrieval	7.5	14	20

Table 2: **Per-task time cost for coordination learning.** All times are wall-clock minutes. DMP data collection corresponds to automated rollout collection, and coordination policy training corresponds to online HIL fine-tuning.

C.2 Baseline Details

HIL-SERL variants. Both HIL-SERL baselines use the same RLPD backbone and offline DMP buffer as HiFun but keep a flat policy that directly outputs full arm-hand actions. *HIL-SERL (Glove+Gello)* provides full arm-hand corrections using AnyTeleop [3] for hand-mapped teleoperation and Gello [1] for arm control. *HIL-SERL (Ours Interv.)* uses the same learned hand-skill intervention interface as HiFun while retaining the flat full-DoF policy. This separates the effect of the intervention interface from the effect of the hierarchical architecture.

Behavior Cloning. BC is trained on the same 200 human demonstrations and uses the same frozen pre-trained ResNet-10 visual encoder as our method for fair comparison. Visual features from each camera view and proprioceptive inputs are concatenated and processed by an MLP that outputs a diagonal Gaussian policy. Training minimizes the negative log-likelihood loss $\mathcal{L}_{BC} = -\mathbb{E}[\log \pi_{\theta}(a | o)]$ using Adam with learning rate 3×10^{-4} and batch size 256 for 20,000 steps. Random cropping is applied as image augmentation, and zero-action transitions are filtered from demonstrations.

Diffusion Policy. DP [4] is trained on the same demonstrations as BC and uses a 1D temporal U-Net noise prediction network, with fully fine-tuned ViT-B/16 [5] encoders for each camera view without weight sharing. The observation and action prediction horizons are set to 2 and 16, respectively. Training uses DDIM with 50 diffusion steps and a squared cosine noise schedule, optimizing the MSE loss on predicted noise with AdamW (learning rate 1×10^{-4} , batch size 256) for 100 epochs. EMA with decay 0.995 is applied.

Additional IL training protocol. The main BC/DP comparison tests whether offline imitation from successful human demonstrations is sufficient for functional dexterous tasks. To separate data coverage from the learning paradigm, we additionally train BC and DP on successful trajectories collected during HiFun training, including DMP rollouts, online policy rollouts, and task-completing human-intervention segments. The results for these same-source IL policies are reported in Sec. D.2.

D Additional Experimental Results

D.1 Online Training Curves and Autonomy

Figure 3 reports intervention rate and cycle time during online training across the six tasks for HiFun and HIL-SERL (Ours Interv.). Since each successful human correction completes the task



Figure 3: **Online training curves across six tasks.** For each task, intervention rate (upper) and cycle time (lower) are shown as running averages over 20 episodes, comparing our method (Ours) against the HIL-SERL (Ours Interv.) baseline. Lower intervention rate indicates more autonomous execution, while lower cycle time indicates more efficient completion.

during training, success rate alone does not reveal whether the learned policy becomes autonomous. The intervention rate therefore measures how often human guidance is still needed, and cycle time measures execution efficiency. For HiFun, the intervention rate decreases toward zero within one hour of online training on every task (and in as little as ~ 15 minutes on the fastest tasks), and its cycle time at convergence is consistently lower than HIL-SERL (Ours Interv.). In contrast, the baseline keeps a high intervention rate and never becomes autonomous within the same one-hour training budget, despite using the same intervention interface. This supports the role of HiFun in efficient online learning.

D.2 Additional IL Baseline Results

Table 3 shows that replacing human demonstrations with HiFun training trajectories does not materially improve offline imitation: BC changes from 27.5% to 26.0% average success, and DP changes from 41.5% to 42.5%. Policies trained on HiFun data often approach objects faster and hesitate less near intermediate states, consistent with exposure to rollouts and intervention segments from more varied contact states. However, they still fail to reliably complete precise alignment and contact recovery. BC and DP are trained by supervised action prediction over the collected trajectories, which does not provide a value-based mechanism for selecting which recovery-like actions should

Task	BC Success Rate (%)		DP Success Rate (%)	
	Human Demos	HiFun Data	Human Demos	HiFun Data
Power-Drill Actuation	15	12	30	26
Angle-Spreader Retrieval	30	32	60	64
Thin-Handle Retrieval	35	32	40	40
Power-Drill Bit Removal	30	28	36	40
Average	27.5	26.0	41.5	42.5

Table 3: **Data-source comparison for IL baselines.** Success rates follow the main evaluation protocol on the Sharpa tasks. Human Demos uses the original 200 teleoperated demonstrations, while HiFun Data uses successful trajectories collected during policy training.

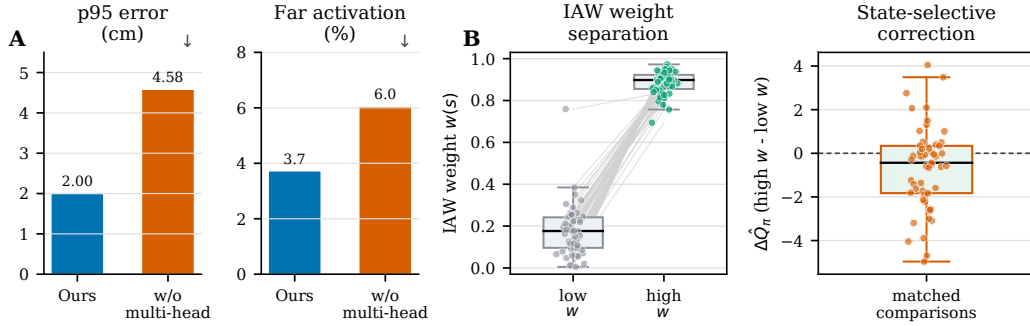


Figure 4: **Coordination component analysis.** (A) Activation localization on *Angle-Spreader Retrieval*: the x-axis compares HiFun with the one-head-critic variant, and the y-axes report p95 activation error and far-state activation rate computed from TCP-to-target error e_t with a 2 cm pre-contact band. (B) IAW weighting: left shows imitation weight $w(s)$ for low- and high-weight interventions; right shows $\Delta\hat{Q}_\pi = \hat{Q}(s_{\text{high-}w}, \bar{a}_\pi(s_{\text{high-}w})) - \hat{Q}(s_{\text{low-}w}, \bar{a}_\pi(s_{\text{low-}w}))$, where negative values indicate high-weight corrections at lower-value policy states.

144 be reinforced for downstream task completion. This explains why recovery-rich offline data alone
145 does not yield reliable recovery behavior without value-guided online RL.

146 D.3 Coordination Component Analysis

147 Figure 4 visualizes the mechanisms behind the coordination ablations in the main paper. In Fig-
148 ure. 4A, the multi-head critic reduces the 95th-percentile activation error from 4.58 cm to 2.00 cm
149 and reduces activations outside the 2 cm pre-contact band from 6.0% to 3.7%. This indicates that
150 separating arm-alignment and skill-activation value learning localizes hand-skill activation closer to
151 the target region, supporting better temporal coordination between arm approach and hand activa-
152 tion. In Figure. 4B, IAW assigns clearly separated imitation weights to weak and strong corrections
153 (0.175 vs. 0.884 mean). The matched policy-value comparison is negative, showing that high-
154 weight corrections occur at states where the current policy action has lower estimated value. These
155 results indicate that IAW avoids uniform cloning by focusing imitation on useful corrections while
156 limiting low-advantage corrections that could conflict with critic optimization.

157 D.4 Hand Motion Control Failure Cases

158 Figure 5 compares hand-control methods on three contact-rich tasks: *Thin-Shaft Tool Retrieval* with
159 XHand, *Angle-Spreader Retrieval* with Sharpa, and *Thin-Handle Retrieval* with Sharpa. After plac-
160 ing the arm near the task execution region and focusing on the task-relevant fingers, the failures
161 mainly reflect functional finger execution rather than arm placement or full-hand teleoperation noise.
162 Kinesthetic replay preserves the demonstrated motion pattern but executes it open-loop, losing the



Figure 5: **Failure analysis for hand motion control.** Representative hand-control outcomes compare the residual hand skill, kinesthetic replay, and DexPilot* across contact-rich manipulation cases.

163 contact-force information needed to lift or lever objects with the task-relevant fingers. DexPilot*,
 164 despite fingertip-based retargeting and per-task tuning, can produce coarse task-relevant contacts but
 165 remains sensitive to the human-robot morphology gap in constrained settings: in *Angle-Spreader Re-*
 166 *trieval*, it can lift the tool but forms an unstable grasp, while in *Thin-Shaft Tool Retrieval* and *Thin-*
 167 *Handle Retrieval*, it struggles with precise pinch alignment and little-finger levering. In contrast,
 168 our residual hand skill adapts the kinesthetic prior with real contact feedback, enabling closed-loop
 169 force modulation for reliable functional hand execution.

170 **Supplementary video.** The supplementary video provides task rollouts, perturbation recovery ex-
 171 amples, and representative intervention behavior for the six real-world tasks.

References

- [1] Philipp Wu, Yide Shentu, Zhongke Yi, Xingyu Lin, and Pieter 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.
- [2] Aleš Ude, Bojan Nemec, Tadej Petrić, and Jun Morimoto. Orientation in cartesian space dynamic movement primitives. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2997–3004. IEEE, 2014.
- [3] Yuzhe Qin, Wei Yang, Binghao Huang, Karl Van Wyk, Hao Su, Xiaolong Wang, Yu-Wei Chao, and Dieter Fox. AnyTeleop: A General Vision-Based Dexterous Robot Arm-Hand Teleoperation System, May 2024.
- [4] Cheng Chi, Zhenjia Xu, Siyuan Feng, Eric Cousineau, Yilun Du, Benjamin Burchfiel, Russ Tedrake, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. *The International Journal of Robotics Research*, 44(10-11):1684–1704, 2025.
- [5] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *ArXiv*, abs/2010.11929, 2020.