Rebuttal for CoRL Submission # 597

We thank the reviewers (R1-GGSf,R2-jWEJ,R3-hiiy) for their insightful feedback and positive comments on our experimental performance (R1,R2,R3), dexterity and practicality, and clear writing (R1,R3). We respond to suggestions with discussion [D], clarification [C], or new information [N].

[CN] Baseline fairness and contribution attribution (R2,R3): We clarify that the learned skill alone is not enough to achieve sample-efficient functional dexterity for standard HIL-SERL. To verify this, we add the baseline **Skill-HIL**, which equips standard HIL-SERL with our learned hand skill as suggested. While this reduces the high-dimensional finger exploration/correction burden, two challenges remain: **(1) Sparse outcomes provide ambiguous credit.** The same task reward may arise from different execution errors, making credit between arm alignment and skill activation difficult to distinguish. HiFun addresses this with separated navigation/modulation critics (Sec. 3.3). **(2) Human corrections provide heterogeneous supervision.** Successful interventions can differ in recovery efficiency and consistency: some quickly return to high-value states, while longer/noisier corrections provide weaker supervision. IAW therefore weights each correction by its advantage over the current policy action (Eq. 7–8). We evaluate Skill-HIL together with the other requested controls under matched task settings and training costs across the four Sharpa main-validation tasks. While **Skill-HIL** improves over flat HIL-SERL from 12.5% to 46%, failures mainly involve incorrect contact regions/timing and subsequent collisions, consistent with the arm-skill credit-assignment issue above. On Angle-Spreader Retrieval, even 2× training data improves Skill-HIL only to 35/50 versus HiFun's 50/50, showing that the skill and additional data alone do not close the gap. **Flat-HiFun** directly learns full arm–hand actions and achieves 6%, confirming the exploration benefit of the hand-skill hierarchy. **BC-Skill** (R3), which retains the skill but replaces online RL with BC, reaches 36%, confirming that limited offline coverage is insufficient for deployment-time alignment/timing deviations. Tab. 2 further isolates the contributions of the multi-head critic and IAW.

[CN] Finger-level recovery and generalization (R2,R3): The Sup. Video shows finger-level recovery (0'28"–0'32") and size variation (1'59"–2'12"). We further add qualitative and quantitative evaluations here. **Contact recovery (R3).** Kinesthetic teaching records the reference finger motion and fingertip-force evolution. If the measured contact deviates substantially from the expected state, the policy can automatically refer to the latest contact-consistent reference state and output the residual to adjust finger configuration and force, which can re-establish desired contacts. Fig. 1 and Sup. Video (0'28"–0'32") show this recovery: an unexpected collision reorients the key and breaks thumb contact, which is subsequently restored and stabilized without arm re-alignment. **Controlled finger-level evaluation (R2).** We further isolate

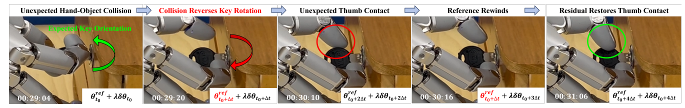

Fig. 1: **Finger-level contact recovery after key reorientation.**

the hand by fixing the arm near the execution region. **(1) Unseen orientations:** Angle-Spreader Retrieval (toolbox yaw 0–30°, tool pitch/roll 0–20°) and Bit Removal (drill yaw $\leq$ 30°) achieve 19/20 and 18/20, with fingers adapting contact configuration. **(2) External contact disturbances:** the fixed-arm Angle-Spreader skill achieves 18/20 under strong wrench perturbations, while a toothpaste-cap task requiring sustained force achieves 20/20. The same hand skill also transfers from a 7.5 cm tool to 6.5/5.5 cm variants (Sup. Video 1'59"–2'12"). These results support the intended scope: task-specific residual skills exhibit local finger-level contact recovery, force regulation, and transfer around the demonstrated contact structure.

[CD] Training cost and efficiency claim (R1,R3): We clarify that "within one hour" refers to the matched online-HIL stage, not end-to-end training. Appendix C.1 already reports per-task kinesthetic, classifier, DMP, and online-HIL costs; adding ∼30 min for hand-skill RL gives ∼109 min/task on average. Reported times include resets and interventions. Most non-HIL costs are one-shot, automated, or shared with HIL-SERL. Thus, the matched online-HIL budget remains the relevant efficiency comparison (Tab. 1).

[D] Novelty (R3): Unlike KineDex's kinesthetic-to-imitation pipeline, HiFun uses only one trajectory as a prior, then acquires the hand skill through real-robot residual RL and reuses it as both the robot-action and human-intervention interface (Sec. 3.2). Prior hierarchies separate hand/arm control, which is not enough to resolve arm-alignment/skill-activation credit under sparse rewards. HiFun addresses this with separate navigation/modulation TD targets and action-subspace stop-gradients (Eq. 5). IAW further differs from outcome/confidence-based intervention use by weighting corrections relative to the current policy value (Eq. 7–8).

[CD] Statistical reliability (R3): We clarify that Tab. 1 is *not* based on a single operator: three operators trained and evaluated the six tasks, with 295/300 successful trials (95% CI: 96.2–99.3%). Final success rates were comparable across operators with around 2-3% differences, although different intervention styles led to slightly different behavior.

[CD] Pipeline details (R1): The arm action is 6D SE(3), and the observed move–stop–act pattern is learned rather than enforced. Some operators stabilize the wrist before skill activation to reduce relative motion/collisions, while Sup. Video (0'08"–0'09") shows simultaneous arm–hand motion. DMP provides systematic, phase-consistent coverage with less operator variation than repeated human demonstrations (App. A). Global BC anchors offline SAC near executed DMP support, and the residual scale is increased progressively (App. C.1).