Proceedings of the Thirty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-26)
Special Track on AI and Robotics

PECHC: Robust Tactile Grasping Stabilization in Vision-Denied Peripersonal Space

Changlin Chen[1], **Sisheng Chen**[1], **Hang Zhang**[1], **Xianglai Zhou**[1], **Zhen Tian**[2], **Weitao Liu**[1], **Feng-Qi Cui**[1], **Erbao Dong**[1*], **Wenjing Chen**[3]
[1]University of Science and Technology of China
[2]School of Computing Science, University of Glasgow, G12 8QQ, UK
[3]Hefei University of Technology
{changlinchen, chensisheng2003, sa25168166, zxl200418, liu_wei_tao, fengqi_cui}@mail.ustc.edu.cn,
tian.zhen@glasgow.ac.uk, ebdong@ustc.edu.cn, 2023211452@mail.hfut.edu.cn

Abstract

In the final "Last-Centimeter" phase of manipulation, where visual occlusion or calibration errors render vision unreliable, robots often suffer high failure rates due to local pose uncertainty and simulation dynamics deviations. To address these issues, this paper proposes the PECHC (Physics-Evolving Cascade Constraint and Human-Correction) algorithm. To rigorously isolate the contribution of tactile feedback in multi-finger coordination, we adopt a decoupled control strategy that focuses on grasp stabilization within the hand's workspace, acting as a fail-safe reflex. The core of our approach is Hybrid Correction Imitation Learning (HCIL), which establishes a "failure-triggered" human-machine mechanism to efficiently resolve the "model gap" via sparse expert corrections. To ensure sample efficiency and baseline performance, we introduce two supporting modules: Cascaded Constraint Scheduling (CCS) addresses the "geometric gap" by enforcing physically plausible behavioral constraints (geometric approach, force closure, and dynamic stability), while Temporal Heterogeneous Distillation (THED) resolves the "physical gap" by enabling implicit system identification from tactile history. Experiments demonstrate that PECHC achieves a 97.3% real-robot success rate on 150 objects from the Visual Dexterity Dataset under fully autonomous testing, where one object is used for one-time HCIL calibration and the remaining 149 objects are evaluated without further intervention. Compared to a standard Sim-to-Real reinforcement learning baseline (Vanilla PPO with Domain Randomization), PECHC delivers a significant performance improvement (+42.8%) and exhibits human-like force modulation capabilities for fragile objects.

1 Introduction

Robust dexterous manipulation is a milestone in embodied intelligence. While vision-based methods excel [Billard and Kragic, 2019; Andrychowicz *et al.*, 2020; Yang *et al.*, 2024], they fail in extreme scenarios like occlusion, darkness, or transparent surfaces. In such *vision-denied* scenarios, tactile feedback serves as the last line of defense, providing the most direct and authentic physical observations of contact states [Liu *et al.*, 2024; Li *et al.*, 2023; Jia *et al.*, 2025; Lin *et al.*, 2025].

Thus, achieving tactile-driven grasp stabilization serves as a robust "Last-Centimeter" reflex policy [Kang *et al.*, 2025]. Rather than replacing global navigation, it complements vision by handling occlusion-induced localization errors within the peripersonal workspace, where tactile feedback becomes the sole reliable modality [Chen *et al.*, 2025b].

In recent years, Sim-to-Real techniques based on reinforcement learning (RL) have shown great potential in dexterous manipulation [Andrychowicz *et al.*, 2020; Kumar *et al.*, 2021].

However, Sim-to-Real for tactile-based systems faces more severe challenges than vision—the "Tactile Reality Gap", which encompasses two core challenges: 1) **Force Adaptation**: Real-world properties (e.g., mass, friction) are invisible and variable, risking slippage or damage; 2) **Simulation Fidelity**: Simulators struggle to model complex contact dynamics (e.g., soft-body deformation), causing policy failure due to distribution shifts. These two factors collectively make policies trained solely in simulation fragile and poorly adaptable when facing complex interactions in the real world [Niu *et al.*, 2024; Chen *et al.*, 2025a].

To address the above core challenges, this paper proposes the PECHC algorithm as an integrated solution to bridge the gaps between simulation and reality.

To bridge the physical gap, we draw inspiration from the privileged distillation paradigm [Kumar *et al.*, 2021]. However, directly applying such methods to blind grasping faces a unique "Geometric Gap": unlike locomotion, a grasper cannot adapt to dynamics if it fails to establish contact due to pose uncertainty. Therefore, PECHC integrates **Cascaded Constraint Scheduling (CCS)** to first guide geometric exploration, serving as a prerequisite for the subsequent dynamic adaptation (THED). We observe that the Tactile Reality Gap comprises two distinct components: 1) Systemic Bias (e.g., global sensor sensitivity, simulation friction scal-

*Corresponding author.

ing offsets), which is consistent across tasks, and 2) Instance-level Dynamics (e.g., specific object mass, compliance). While pure Domain Randomization struggles with the former, blindly adapting to everything risks overfitting. Abandoning the traditional full-time demonstration paradigm, Hybrid Correction Imitation Learning (HCIL) innovatively designs a *failure-triggered* human-machine collaborative mechanism. When the policy fails to adjust grasp force due to object uncertainty (e.g., slippage) or exhibits abnormal actions due to model deviations during real-robot execution, human experts intervene to provide *sparse* residual corrections. This mechanism not only directly corrects model errors (the *model gap*) but, more importantly, injects expert physical intuition for handling uncertainty into the policy, correcting systemic simulation biases via sparse human intervention. This calibration aligns the fundamental physics engine parameters, allowing the THED module to handle instance-level object variations (mass, friction) through online inference without further human guidance.

To maximize the correction efficiency of HCIL and ensure the policy possesses strong initial capabilities, we design two key supporting modules:

1. **Cascaded Constraint Scheduling (CCS)** mitigates the *geometric gap* by decomposing grasping into geometric approach, force closure, and dynamic stability. Physical constraints guide the policy toward plausible tactile primitives and provide a reliable baseline for human corrections.
2. **Temporal Heterogeneous Distillation (THED)** bridges the *physical gap* by learning to infer privileged physical parameters, such as friction and mass, from tactile history. This enables adaptive force modulation under object uncertainty.

Unlike blind manipulation approaches targeting full-scene navigation or object search, our work strictly focuses on grasping stability within the **end-effector's workspace**. PECHC specifically addresses the most intractable challenge in dexterous manipulation: dynamic force adaptation after contact establishment.

Scope and Assumption: It is important to clarify that our work focuses on grasp stabilization rather than global object search. We assume a coarse visual system or heuristics has guided the wrist to the object's vicinity (peripersonal space). PECHC addresses the critical challenge that follows: establishing a stable grasp under local pose uncertainty and dynamic mismatch, where vision is no longer reliable.

2 Related Work

2.1 Tactile and Proprioception-Based Dexterous Grasping

Blind manipulation relying on tactile and proprioceptive feedback offers unique advantages in vision-denied or high-precision force control scenarios. The core challenge lies in deriving adaptive policies from sparse, noisy signals. Suresh et al. [Suresh *et al.*, 2023] addressed this via a Monte Carlo inference framework to recover object geometry from sliding feedback. OpenAI's Dactyl achieved blind object rotation using large-scale RL and domain randomization [Andrychowicz *et al.*, 2020], though this approach requires immense computational resources. To mitigate reliance on randomization, Lin et al. [Lin *et al.*, 2022] developed Tactile Gym 2.0, a high-fidelity simulator that narrows the sim-to-real gap. Regarding feature extraction, Hu et al. [Hu *et al.*, 2025] utilized low-dimensional features for pose tracking, while Guzey et al. [Guzey *et al.*, 2023] introduced self-supervised learning to derive general representations from interaction data. For stability prediction, Zhao et al. [Zhao *et al.*, 2024] modeled spatiotemporal correlations using "pseudo-3D" tactile images and NDConv. Guo et al. [Guo *et al.*, 2025] employed graph convolutional networks to enable skill transfer and verify tactile grasping of soft, slippery objects. Furthermore, Church et al. [Church *et al.*, 2022] achieved millimeter-level tactile servoing via Real-to-Sim optimization. Despite these advances, existing works still lack effective online adaptation mechanisms for significant physical dynamic discrepancies between simulation and reality [Chari *et al.*, 2024; Hang *et al.*, 2024].

2.2 Human–Machine Interaction Correction and Policy Adaptation

Introducing *Human-in-the-Loop* (HITL) is effective for bridging sim-to-real gaps. Early methods like Behavior Cloning (BC) relied on full-time teleoperation, which is constrained by data quality and struggles to correct minor deviations [Mandlekar *et al.*, 2022]. Recently, residual learning has gained attention: Johannink et al. [Johannink *et al.*, 2019] compensated for model errors via learned residual actions, while Losey et al. [Losey *et al.*, 2020] utilized physical contact for real-time correction. In vision domains, TRANSIC [Jiang *et al.*, 2024] adapts via online expert corrections but relies on visual feedback for intervention timing, making it infeasible for *Blind Grasp Stabilization*. Meanwhile, Xue et al. [Xue *et al.*, 2023] focused on preference-based task optimization rather than physical dynamics adaptation. Distinct from these, our Hybrid Correction Imitation Learning (HCIL) is tailored for tactile Blind Grasp Stabilization. Inspired by "Recovery Zones" [Thananjeyan *et al.*, 2021], we replace visual judgment with an automatic failure-triggered mechanism based on Tactile Safety Thresholds. Expert intervention is requested only upon detecting anomalies (e.g., a slip). This *on-demand correction* significantly reduces annotation burden and guides the policy to focus on learning real-world physical dynamics missed by simulators, facilitating efficient transfer [Kojima *et al.*, 2025].

3 Method

The PECHC algorithm aims to solve the problem of general dexterous grasping relying solely on sparse tactile feedback in unstructured environments. As shown in Figure 1, the algorithm consists of three progressive stages: first, learning physically plausible grasp primitives in fully observable simulations via Cascaded Constraint Scheduling (CCS); second, compressing the teacher's privileged physical information into tactile history representations through Temporal Hetero-

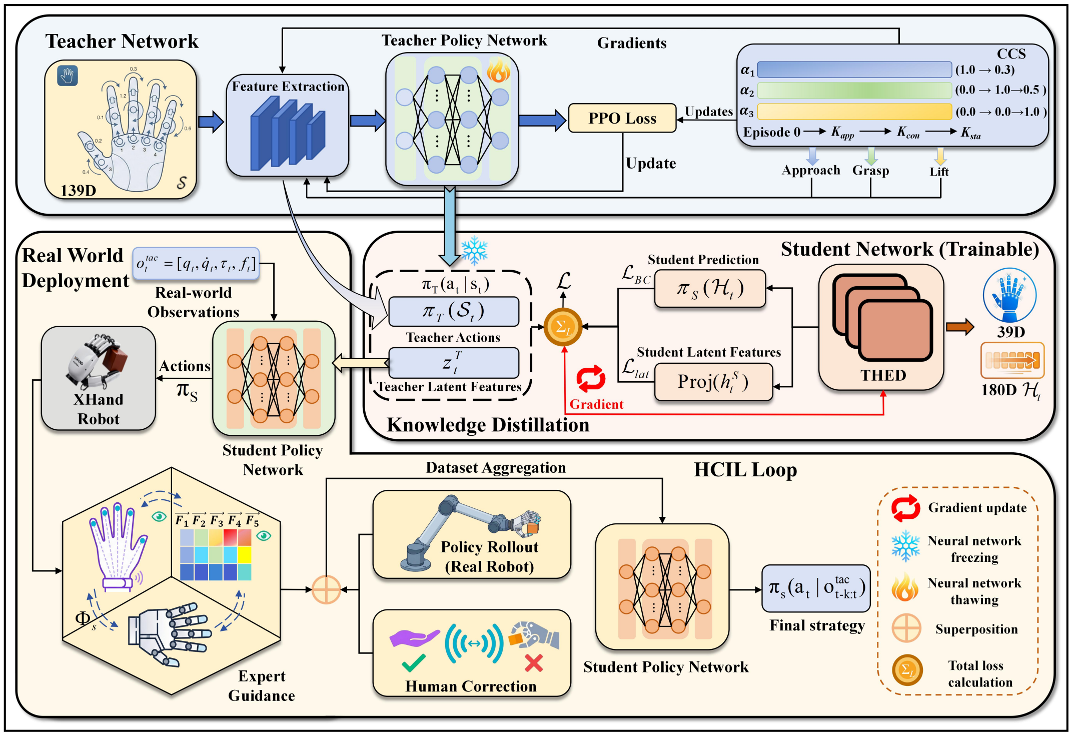

Figure 1: Flowchart of the PECHC algorithm framework. Note the 'Failure-Triggered' loop in HCIL, which activates only upon slip detection, significantly reducing human annotation cost compared to full-time teleoperation.

geneous Distillation (THED); finally, correcting long-tail dynamic errors not covered by simulations via human-machine interaction (HCIL).

3.1 Preliminaries

We model the problem as a Partially Observable Markov Decision Process (POMDP), defined as $\mathcal{M} = (\mathcal{S}, \mathcal{O}^{\text{tac}}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma)$.

Privileged State Space $\mathcal{S} \in \mathbb{R}^{139}$: $\mathbf{s}_t = [\mathbf{q}_t, \dot{\mathbf{q}}_t \in \mathbb{R}^{12}, \mathbf{p}_{obj} \in \mathbb{R}^7, \mathbf{v}_{obj} \in \mathbb{R}^6, \phi_{phy}]$, where $\mathbf{q}_t, \dot{\mathbf{q}}_t$ are joint positions/velocities (12-DoF), $\mathbf{p}_{obj}$ is object pose (3D position + quaternion), $\mathbf{v}_{obj}$ is velocity (linear+angular), and $\phi_{phy} \in \mathbb{R}^{102}$ contains contact geometry, physical properties (friction μ, mass m), and fine-grained contact states.

Tactile Observation Space $\mathcal{O}^{\text{tac}} \in \mathbb{R}^{39}$: $\mathbf{o}_t^{tac} = [\mathbf{q}_t, \dot{\mathbf{q}}_t \in \mathbb{R}^{12}, \mathbf{f}^{\text{tip}} \in \mathbb{R}^{15}]$, where $\mathbf{f}^{\text{tip}}$ represents 5-fingertip 3D forces. Joint torques τ are computed via inverse dynamics. The action space $\mathcal{A} \in \mathbb{R}^{12}$ consists of joint position increments. Goal: learn $\pi_S(\mathbf{a}_t|\mathbf{o}_{t-k:t}^{\text{tac}})$ maximizing $\mathbb{E}[\sum \gamma^t r_t]$ in $\mathcal{M}_{real}$.

3.2 Stage I: Constraint-Aware Teacher Learning via CCS

(Resolving the geometric gap) In the simulation stage ($\mathcal{M}_{sim}$), we train the teacher policy $\pi_T(\mathbf{a}_t|\mathbf{s}_t)$ using privileged information. To address low exploration efficiency in high-dimensional action spaces and "non-physical violent grasping", we design the Cascaded Constraint Scheduling (CCS) mechanism. CCS models the grasping task as a physically constrained optimization problem, dynamically adjusting the reward function structure through state machine logic. Unlike traditional scalar rewards, CCS decouples the grasping process into three hierarchical constraints $\mathcal{K} \in \{K_{app}, K_{con}, K_{sta}\}$ and guides the policy to focus on specific physical sub-goals at different stages via dynamic weights α_i.

(a) Geometric Approach Constraint ($\mathbf{K_{app}}$): Aims to guide the palm to envelope the object with a reasonable configuration. The reward function r_{app} is designed to guide the hand configuration without penalizing contact force, avoiding policy shrinking (fear of contact):

$$r_{app} = -\alpha_1 \lambda_d d_{\min}(\mathbf{x}_{tips}, \mathbf{x}_{obj}) + \lambda_e E(\mathbf{q}) + \mathbb{I}_{\text{contact}} \cdot R_{\text{bonus}} \tag{1}$$

Here, $d_{\min}$ is the fingertip-object distance, $E(\mathbf{q})$ aligns palm orientation, and $\mathbb{I}_{\text{contact}}$ indicates contact; λ_d, λ_e are scaling coefficients. α_1 is high in the early training stage and gradually decreases as the stage progresses (as shown in the upper right of Figure 1), preventing the policy from overfitting to geometric positions while ignoring force control. Note that $d_{\min}$ and object pose $\mathbf{p}_{obj}$ are privileged information available only during simulation training. The Student Policy π_S does not have access to these states and must implicitly infer contact geometry solely from proprioception and tactile history during the approach phase.

(b) Force Closure and Balance Constraint ($\mathbf{K_{con}}$): After geometric contact is established, the goal shifts to achieving mechanically stable grasping. We introduce the minimum singular value $\sigma_{\min}(\mathbf{G})$ of the grasp matrix $\mathbf{G}$ as a force closure quality metric and penalize the variance $\mathrm{Var}(\|\mathbf{f}_{tip}\|)$ of fingertip force distribution to induce human-like compliant multi-finger collaboration instead of relying on single-finger heavy squeezing.

$$r_{con} = \alpha_2 \left(\lambda_c \sigma_{\min}(\mathbf{G}) - \lambda_v \mathrm{Var}(\|\mathbf{f}_{tip}\|)\right) + \lambda_{\text{safe}} \sum_{i=1}^{5} \mathbb{I}(f_{\min} < \|\mathbf{f}_i\| < f_{\max}) \tag{2}$$

Here, $\sigma_{\min}(\mathbf{G})$ quantifies force closure quality, $\mathrm{Var}(\|\mathbf{f}_{tip}\|)$ ensures load distribution among fingers, and the last term prevents motor overload.

The coefficients and thresholds in CCS are physically grounded rather than arbitrarily tuned. For instance, the safety threshold $f_{\max}$ is derived from the sensor's saturation limit, and the geometric weights evolve based on the kinematic workspace of the XHand. We observed that this physically-constrained reward formulation is robust to minor hyperparameter variations compared to pure scalar reward shaping.

(c) Dynamic Stability Constraint ($\mathbf{K_{sta}}$): During the lifting stage, the object must maintain a stable state when disturbed. The policy must keep the object's linear velocity $\|\mathbf{v}_o\|$ and angular velocity $\|\omega_o\|$ below thresholds while the object is subjected to random external force disturbances or being lifted ($h > h_{\text{tgt}}$).

$$r_{sta} = \alpha_3 \left(\mathbb{I}(h > h_{\text{tgt}}) - \lambda_{\text{vel}}(\|\mathbf{v}_o\| + \|\omega_o\|)\right) \tag{3}$$

This term rewards lifting the object to the target height h_{tgt} while penalizing jitter in linear velocity $\mathbf{v}_o$ and angular velocity ω_o during lifting. The weight vector $\alpha = [\alpha_1, \alpha_2, \alpha_3]$ dynamically evolves with learning steps. In the early stage, α_1 dominates to quickly explore contact states; in the later stage, α_2, α_3 dominate to optimize the force control policy, ultimately learning the teacher policy $\pi_T(\mathbf{a}_t|\mathbf{s}_t)$. State switching follows an irreversible logic (unless the task fails and resets), and the policy strictly adheres to the physical causal chain of "geometric approach $\rightarrow$ mechanical balance $\rightarrow$ dynamic stability". The core role of CCS is to use "contact itself" as a geometric sensor in the absence of visual priors, guiding the policy to smoothly transition from "exploratory contact" to "force closure establishment".

3.3 Stage II: Temporal Heterogeneous Distillation (THED)

(Resolving the physical gap) Building upon the success of Rapid Motor Adaptation [Kumar *et al.*, 2021] in locomotion, THED adopts an asymmetric actor-critic structure. While [Kumar *et al.*, 2021] focuses on body dynamics, THED is specifically tailored to reconstruct contact parameters (e.g., friction, compliance) from sparse tactile history.

To transfer the teacher policy π_T relying on privileged information $\mathbf{s}_t$ to the student policy π_S with only $\mathbf{o}_t^{\text{tac}}$, this paper proposes Temporal Heterogeneous Distillation (THED). Its core idea is to implicitly reconstruct environmental physical parameters from tactile temporal history (e.g., sliding history implies friction coefficient information). We explicitly design the student policy as a decoupled tactile reflex layer. In real-robot deployment, the wrist trajectory is handled by a separate kinematic planner, while the student policy assumes full authority over the finger articulations to stabilize the grasp. This decoupling strategy isolates the high-frequency contact stabilization problem from the low-frequency global transport, allowing PECHC to serve as a plug-and-play reflex module compatible with various arm planners.

The teacher encodes $\mathbf{s}_t$ into features $\mathbf{z}_t^T$ via an MLP. The student processes tactile history $\mathcal{H}_t = \{(\mathbf{o}_k^{\text{tac}}, \mathbf{a}_{k-1})\}_{k=t-H}^{t}$ using an LSTM to extract $\mathbf{h}_t^S$, followed by an action MLP. The distillation objective includes two loss functions:

(a) Behavior Cloning Loss ($\mathcal{L}_{BC}$): Collect demonstration trajectories $\mathcal{D}_{demo}$ in simulation via the teacher policy, forcing the student action distribution to fit the teacher's actions:

$$\mathcal{L}_{BC} = \mathbb{E}\left[\|\pi_S(\mathcal{H}_t) - \pi_T(\mathbf{s}_t)\|_2^2\right] \tag{4}$$

(b) Latent Physical Representation Alignment ($\mathcal{L}_{lat}$): Force the projected LSTM latent state $\mathbf{h}_t^S$ of the student network to approximate the teacher's privileged feature $\mathbf{z}_t^T$. This compels the student network to decouple key physical information from tactile history (System Identification). By minimizing the latent $\mathcal{L}_{lat}$, we force the recurrent latent state of the student network to reconstruct the teacher's beliefs about explicit physical parameters. This process goes beyond behavior cloning; it endows the student with online system identification capabilities, enabling it to infer invisible environmental dynamics (e.g., object mass distribution) from contact force history:

$$\mathcal{L}_{lat} = \mathbb{E}\left\|\mathrm{Proj}(\mathbf{h}_t^S) - \mathbf{z}_t^T\right\|_2^2 \tag{5}$$

The total optimization objective is $\mathcal{L} = \mathcal{L}_{BC} + \beta\mathcal{L}_{lat}$, where β is a balance coefficient.

3.4 Stage III: Hybrid Correction Imitation Learning (HCIL)

(Resolving the model gap) Although simulation training endows the policy with basic capabilities, complex dynamic characteristics in the real environment (e.g., cable pulling, nonlinear friction) may still lead to failures. We design a Hybrid Correction Imitation Learning (HCIL) system, introducing human experts as correctors. Our HCIL proceeds in two stages: (1) **Adaptation**: at initial deployment, human experts intervene upon failure detection to provide correction data; (2) **Autonomous deployment**: the fine-tuned π_S^* operates independently on unseen objects without further human intervention.

We empirically observe that a significant portion of the Sim-to-Real gap manifests as **systematic sensor-motor biases** (e.g., global friction scaling, sensor thresholds) rather than purely object-specific dynamics. Therefore, HCIL utilizes a representative object to perform **system-level alignment**. Our experiments suggest that calibrating these funda-

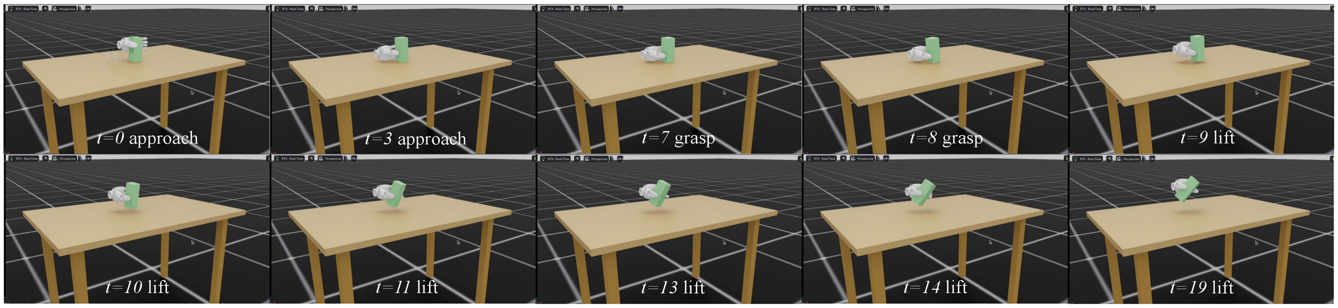

Figure 2: Temporal snapshots of successful grasping and lifting objects in the simulation environment

mental physics engine offsets significantly improves the policy's zero-shot generalization to unseen objects, reducing the need for per-instance fine-tuning.

Unlike full-time teleoperation, HCIL adopts a "failure-triggered" mechanism via tactile safety thresholds $\Phi_s = \{f_{\max} = 15\,\mathrm{N},\ \dot{f}_{\text{slip}} = 50\,\mathrm{N/s},\ \tau_{\text{limit}} = 1.5\,\mathrm{N\,m}\}$, derived from hardware specs (sensor saturation, motor rating) rather than object-specific tuning. When $\|\dot{\mathbf{f}}\| > \dot{f}_{\text{slip}}$ (slip detection) or $\tau > \tau_{\text{limit}}$ (overload), the robot pauses and requests expert correction via Manus Pro.

To prevent catastrophic forgetting, we construct a mixed dataset $\mathcal{D}_{mix}$:

$$\mathcal{D}_{mix} = \mathcal{D}_{\text{exp}} \cup \text{Sample}(\mathcal{D}_{policy}, \rho) \tag{6}$$

where $\mathcal{D}_{\text{exp}}$ is human correction data, $\mathcal{D}_{\text{policy}}$ is self-exploration data, and $\rho = 0.3$ (30% self-exploration + 70% human corrections). By fine-tuning π_S on $\mathcal{D}_{mix}$, the policy can quickly adapt to the physical distribution of the real world while retaining the general force control primitives learned in the simulation stage.

The complete algorithm flow is shown in Algorithm 1.

Algorithm 1 PECHC Training Pipeline

Require: Simulator $\mathcal{M}_{sim}$, Real Robot $\mathcal{M}_{real}$, Expert Human $\mathcal{E}$
Ensure: Robust Tactile Policy π_S^*

▷ *Stage I: Teacher Training*
Initialize π_T, CCS Rewards $\{r_{app}, r_{con}, r_{sta}\}$
while not converged **do**
  Collect trajectories $\zeta \sim \pi_T$ in $\mathcal{M}_{sim}$
  Update π_T using PPO with curriculum state $K(\mathbf{s})$
end while

▷ *Stage II: Distillation*
Collect dataset $\mathcal{D}_{sim} = \{(\mathcal{H}_t, \mathbf{a}_t^T, \mathbf{z}_t^T)\}$ using π_T
Initialize π_S (LSTM-based)
while not converged **do**
  Sample batch $B \sim \mathcal{D}_{sim}$
  Update π_S by minimizing $\mathcal{L}_{BC} + \beta\mathcal{L}_{lat}$
end while

▷ *Stage III: Real-World Adaptation (HCIL)*
$\mathcal{D}_{mix} \leftarrow \varnothing$
for iteration $k = 1$ to N **do**
  Rollout π_S in $\mathcal{M}_{real}$
  if Failure Detected ($\|\dot{\mathbf{f}}\| > \dot{f}_{\text{slip}}$ or $\tau > \tau_{\text{limit}}$) **then**
    Expert $\mathcal{E}$ intervenes, collect correction ζ_{exp}
    $\mathcal{D}_{mix} \leftarrow \mathcal{D}_{mix} \cup \zeta_{\text{exp}}$
  else
    Store successful trajectory ζ_{self}
    $\mathcal{D}_{mix} \leftarrow \mathcal{D}_{mix} \cup \text{Sample}(\zeta_{\text{self}})$
  end if
  Fine-tune π_S on $\mathcal{D}_{mix}$
end for
return π_S^*

4 Experiments

4.1 Experimental Setup and Baseline Description

Experiments use a 12-DoF XHand with 3D force sensors and Manus Pro gloves. Simulations (Isaac Lab, 60 Hz physics) require lifting objects to $h_{\text{tgt}} = 6\,\mathrm{cm}$ for >2 s. Training employs basic geometric shapes, while testing evaluates generalization on 150 complex objects from the Visual Dexterity Dataset and fragile models.

For "Blind Grasp Stabilization", we assume that objects are initialized within a $2\,\mathrm{cm} \times 2\,\mathrm{cm}$ workspace. The wrist follows a fixed trajectory, relying on mechanical compliance and tactile feedback to accommodate positional errors. This definition decouples global navigation from grasping, focusing specifically on assessing policy robustness against geometric and physical uncertainties within peripersonal space.

To strictly quantify the marginal gain of our tactile adaptation modules, we employ **Vanilla PPO with Domain Randomization (DR)** as the primary baseline, consistent with the protocols in Dactyl [Andrychowicz *et al.*, 2020] and RMA [Kumar *et al.*, 2021]. This baseline utilizes the identical LSTM architecture and domain randomization ranges as PECHC but lacks the CCS, THED, and HCIL components. By maintaining architectural parity, we isolate the performance improvements derived specifically from our proposed physics-evolving constraints and human-correction framework rather than general backbone capacity.

4.2 Simulation Performance Analysis

We first evaluate the learning efficiency of the teacher policy in the simulation environment. Figure 2 shows temporal snapshots of successful grasping and lifting objects in the simulation environment. The sequence demonstrates the three stages guided by CCS: approach ($t = 0$–3), where the hand pre-configures and approaches the object; grasping ($t = 7$–8), where stable multi-finger contact is established; lifting ($t = 9$–19), where the object is lifted to the target height while maintaining its posture. The quantitative impact of CCS is shown in Figure 3, which indicates that the CCS mechanism increases the grasping success rate from 3.8% of the unconstrained baseline to 98.9%. Funnel analysis shows that CCS effectively guides the policy to cross the exploration gap from "contact" to "force closure", proving the necessity of physical constraints in long-horizon task learning.

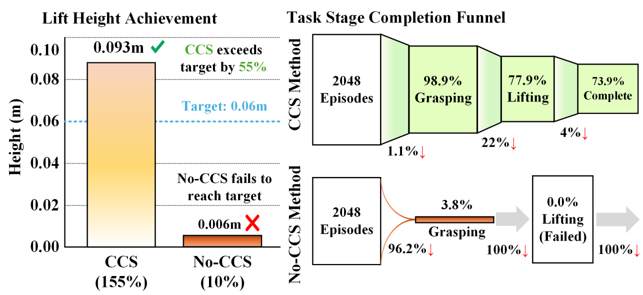

Figure 3: Efficacy analysis of the CCS mechanism during simulation training

4.3 Real-World Evaluation

We set 6 difficulty levels for the Visual Dexterity Dataset [Chen *et al.*, 2023] based on [Lum *et al.*, 2024]. Adopting a 'calibrate-then-generalize' protocol, HCIL adaptation is performed once on a single calibration object from the 150-object evaluation set to align physical gaps. The policy is then frozen and evaluated on the full 150-object set, where the remaining 149 objects are unseen during calibration and no further intervention is allowed. As shown in Table 1, PECHC achieved a 97.3% success rate (146/150), demonstrating excellent post-calibration generalization (see Figure 4 and Figure 5(a)). We benchmark against external vision-based SOTA methods (DemoGrasp [Yuan *et al.*, 2025], DextrAH-G [Lum *et al.*, 2024]) and an internal baseline (Vanilla PPO + DR), representing standard Sim-to-Real RL without our proposed modules.

Table 2 presents results across two distinct evaluation paradigms. Vision-based methods (rows 1–3) achieve 80.0%–99.1% in idealized simulation with perfect visual perception, establishing the theoretical upper bound when geometry is fully observable. These results contextualize task difficulty but are NOT direct competitors since our setting deliberately excludes vision. Our primary contribution lies in the tactile-only comparison (rows 4–5): PECHC achieves 97.3% success on 150 real-world objects using only proprioceptive and tactile feedback—a substantial +42.8% improvement over standard Sim-to-Real RL (Vanilla PPO + DR: 54.5%). This gap demonstrates that domain randomization alone is insufficient for tactile-driven manipulation, validating the necessity of our CCS+THED+HCIL framework.

Figure 4: Difficulty level diagram of the Visual Dexterity Dataset and real-world experimental results

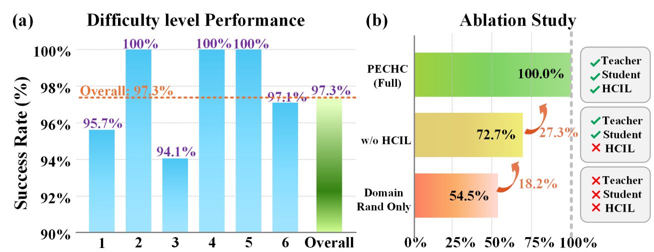

Figure 5: Experimental results on public datasets and ablation study comparison

4.4 Ablation Study and Physical Perception Verification

To quantify the marginal contribution of each module and verify whether the policy truly possesses "physical perception" capabilities, we conduct a detailed ablation study (Figure 5(b)).

1. **HCIL and Systematic Alignment**: Removing HCIL drops the success rate from 97.3% to 72.7%. Notably, performance is restored by calibrating on just a single representative object. This confirms that a dominant component of the Sim-to-Real gap in our setup stems from global systematic biases (e.g., sensor sensitivity, friction scaling) rather than per-object geometric variance. HCIL effectively aligns the student's physical beliefs, facilitating zero-shot transfer to unseen rigid objects.
2. **THED and Implicit Perception**: Without tactile history, the success rate plummets to 54.5%, aligning with the vanilla domain randomization baseline. This 42.8%

Object	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Overall
Success Rate	100.0%	100.0%	100.0%	97.1%	95.7%	94.1%	97.3%
Successful Captures	25	16	12	33	44	16	146
Total Number	25	16	12	34	46	17	150

Table 1: Experimental success rates for each level of the Visual Dexterity Dataset

Method	Perception Modality	Evaluation Domain	Success Rate
DemoGrasp [Yuan *et al.*, 2025]	Vision (RGB-D)	Simulation (Idealized)	99.1%
DextrAH-G [Lum *et al.*, 2024]	Vision (Point Cloud)	Simulation (Idealized)	80.0%
DextrAH-RGB [Singh *et al.*, 2024]	Vision (Stereo)	Simulation (Idealized)	≈89%
Vanilla PPO + DR	Tactile	Real World	54.5%
PECHC (Ours)	Tactile	Real World	97.3%

Note: Vision methods evaluated in idealized simulation; tactile in real hardware.
Direct comparison not applicable due to different evaluation domains.

Table 2: Performance comparison across perception modalities. Vision methods are evaluated in idealized simulation; tactile methods are evaluated on real hardware.

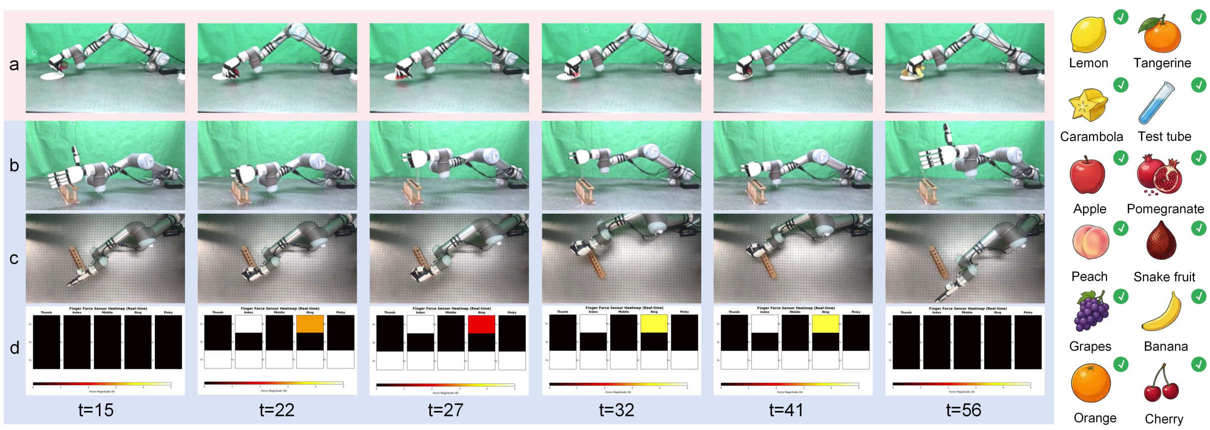

Figure 6: Experimental results on the private dataset

performance gain confirms that THED endows the agent with implicit system identification capabilities. By dynamically inferring invisible environmental parameters (e.g., friction, slip trends) from tactile history, the policy achieves true closed-loop physical adaptation rather than simple open-loop control.

4.5 Fine-Grained Force Control and Failure Analysis

To further evaluate the fine-grained force control capability of the policy, we selected 11 high-fidelity models of fruits with different hardness and shapes (e.g., lemons, red apples, grapes, starfruits) as well as fragile test tubes for testing. Experimental results show that PECHC achieves a 100% success rate (see Figure 6 for details). The policy successfully modulates fingertip force to prevent crushing, a capability often lacking in vision-only policies. Despite the high success rate, we analyzed rare failure cases to understand system limitations:

a. **Geometric outliers**—objects exceeding the hand aperture can fail during the initial "geometric approach" stage, as the blind policy struggles to locate boundaries.

b. **Sensor saturation**—high-speed grasping of heavy objects ($>$ 500 grams) can saturate tactile readings, causing state-estimation loss and slip.

5 Conclusion

The proposed PECHC algorithm addresses the limits of vision-based planning by introducing a tactile-driven "Tactile Reflex" layer that decouples wrist navigation from finger-level interaction. With one-object calibration, PECHC closes the "Last-Centimeter" gap in the visual blind zone and achieves a 97.3% success rate on the 150-object benchmark. HCIL verifies that tactile feedback need not replace vision; instead, a human-calibrated tactile reflex loop with physical perception is sufficient for robust last-centimeter dexterous grasping. Future work will explore "In-Hand Reorientation" and vision-tactile fusion control.

Acknowledgments

The experiments, simulations, and model training were performed on the robotic AI-Scientist platform of the Chinese Academy of Sciences. This work was supported by the Major Project of Anhui Province's Science and Technology Innovation Breakthrough Plan (Grant No. 202423h08050003).

References

[Andrychowicz *et al.*, 2020] Marcin Andrychowicz, Bowen Baker, Maciek Chociej, Rafal Jozefowicz, Bob McGrew, Jakub Pachocki, Arthur Petron, Matthias Plappert, Glenn Powell, Alex Ray, et al. Learning dexterous in-hand manipulation. *The International Journal of Robotics Research*, 39(1):3–20, 2020.

[Billard and Kragic, 2019] Aude Billard and Danica Kragic. Trends and challenges in robot manipulation. *Science*, 364(6446):eaat8414, 2019.

[Chari *et al.*, 2024] Kartik Chari, Niklas Kueper, Su Kyoung Kim, Frank Kirchner, and Elsa Andrea. Inter-hri competition: intrinsic error evaluation during human-robot interaction. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pages 8623–8626, 2024.

[Chen *et al.*, 2023] Tao Chen, Megha Tippur, Siyang Wu, Vikash Kumar, Edward Adelson, and Pulkit Agrawal. Visual dexterity: In-hand reorientation of novel and complex object shapes. *Science Robotics*, 8(84):eadc9244, 2023.

[Chen *et al.*, 2025a] Jiayi Chen, Yubin Ke, Lin Peng, and He Wang. Dexonomy: Synthesizing all dexterous grasp types in a grasp taxonomy. *arXiv preprint arXiv:2504.18829*, 2025.

[Chen *et al.*, 2025b] Zerui Chen, Shizhe Chen, Etienne Arlaud, Ivan Laptev, and Cordelia Schmid. Vividex: Learning vision-based dexterous manipulation from human videos. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3336–3343. IEEE, 2025.

[Church *et al.*, 2022] Alex Church, John Lloyd, Nathan F Lepora, et al. Tactile sim-to-real policy transfer via real-to-sim image translation. In *Conference on Robot Learning*, pages 1645–1654. PMLR, 2022.

[Guo *et al.*, 2025] Ce Guo, Xieyuanli Chen, Zhiwen Zeng, Zirui Guo, Yihong Li, Haoran Xiao, Dewen Hu, and Huimin Lu. Grasp like humans: Learning generalizable multi-fingered grasping from human proprioceptive sensorimotor integration. *IEEE Transactions on Robotics*, 2025.

[Guzey *et al.*, 2023] Irmak Guzey, Ben Evans, Soumith Chintala, and Lerrel Pinto. Dexterity from touch: Self-supervised pre-training of tactile representations with robotic play. In *Proceedings of the Conference on Robot Learning (CoRL)*, pages 3142–3166. PMLR, 2023.

[Hang *et al.*, 2024] Jinglue Hang, Xiangbo Lin, Tianqiang Zhu, Xuanheng Li, Rina Wu, Xiaohong Ma, and Yi Sun. Dexfuncgrasp: A robotic dexterous functional grasp dataset constructed from a cost-effective real-simulation annotation system. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 10306–10313, 2024.

[Hu *et al.*, 2025] Wenbin Hu, Bidan Huang, Wang Wei Lee, Sicheng Yang, Yu Zheng, and Zhibin Li. Dexterous in-hand manipulation of slender cylindrical objects through deep reinforcement learning with tactile sensing. *Robotics and Autonomous Systems*, 186:104904, 2025.

[Jia *et al.*, 2025] Peida Jia, Xuanheng Li, Tianqiang Zhu, Rina Wu, Xiangbo Lin, and Yi Sun. Multi-fingered hand grasps with visuo-tactile fusion via multi-agent deep reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 14594–14601, 2025.

[Jiang *et al.*, 2024] Yunfan Jiang, Chen Wang, Ruohan Zhang, Jiajun Wu, and Li Fei-Fei. TRANSIC: Sim-to-real policy transfer by learning from online correction. In *Proceedings of the Conference on Robot Learning (CoRL)*, pages 1691–1729. PMLR, 2024.

[Johannink *et al.*, 2019] Tobias Johannink, Shikhar Bahl, Ashvin Nair, Jianlan Luo, Avinash Kumar, Matthias Loskyll, Juan Aparicio Ojea, Eugen Solowjow, and Sergey Levine. Residual reinforcement learning for robot control. In *2019 international conference on robotics and automation (ICRA)*, pages 6023–6029. IEEE, 2019.

[Kang *et al.*, 2025] Taewoong Kang, Joonyoung Kim, Seunghwa Oh, Woosung Lim, Junwoo Lee, Seung-Joon Yi, and Jungwon Seo. Dexterous ungrasping manipulation in three dimensions. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 14030–14036. IEEE, 2025.

[Kojima *et al.*, 2025] Takeshi Kojima, Yaonan Zhu, Yusuke Iwasawa, Toshinori Kitamura, Gang Yan, Shu Morikuni, Ryosuke Takanami, Alfredo Solano, Tatsuya Matsushima, Akiko Murakami, and Yutaka Matsuo. A comprehensive survey on physical risk control in the era of foundation model-enabled robotics. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-25)*, pages 10517–10527. International Joint Conferences on Artificial Intelligence Organization, 2025.

[Kumar *et al.*, 2021] Ashish Kumar, Zipeng Fu, Deepak Pathak, and Jitendra Malik. Rapid motor adaptation for legged robots. In *Proceedings of the Conference on Robot Learning (CoRL)*, pages 1342–1352. PMLR, 2021.

[Li *et al.*, 2023] Haoming Li, Xinzhuo Lin, Yang Zhou, Xiang Li, Yuchi Huo, Jiming Chen, and Qi Ye. Contact2grasp: 3d grasp synthesis via hand-object contact constraint. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (IJCAI-23)*, pages 1053–1061. International Joint Conferences on Artificial Intelligence Organization, 2023.

[Lin *et al.*, 2022] Yijiong Lin, John Lloyd, Alex Church, and Nathan F Lepora. Tactile gym 2.0: Sim-to-real deep reinforcement learning for comparing low-cost high-resolution robot touch. *IEEE Robotics and Automation Letters*, 7(4):10754–10761, 2022.

[Lin *et al.*, 2025] Pei Lin, Yuzhe Huang, Wanlin Li, Jianpeng Ma, Chenxi Xiao, and Ziyuan Jiao. Pp-tac: Paper picking using tactile feedback in dexterous robotic hands. *arXiv preprint arXiv:2504.16649*, 2025.

[Liu *et al.*, 2024] Yumeng Liu, Yaxun Yang, Youzhuo Wang, Xiaofei Wu, Jiamin Wang, Yichen Yao, Soeren Schwertfeger, Sibei Yang, Wenping Wang, Jingyi Yu, Xuming He, and Yuexin Ma. Realdex: Towards human-like grasping for robotic dexterous hand. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24)*, pages 6859–6867. International Joint Conferences on Artificial Intelligence Organization, 2024.

[Losey *et al.*, 2020] Dylan P Losey, Krishnan Srinivasan, Ajay Mandlekar, Animesh Garg, and Dorsa Sadigh. Controlling assistive robots with learned latent actions. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 378–384. IEEE, 2020.

[Lum *et al.*, 2024] Tyler Ga Wei Lum, Martin Matak, Viktor Makoviychuk, Ankur Handa, Arthur Allshire, Tucker Hermans, Nathan D. Ratliff, and Karl Van Wyk. DextrAH-G: Pixels-to-action dexterous arm-hand grasping with geometric fabrics. In *Proceedings of the Conference on Robot Learning (CoRL)*, pages 3182–3211. PMLR, 2024.

[Mandlekar *et al.*, 2022] Ajay Mandlekar, Danfei Xu, Josiah Wong, Soroush Nasiriany, Chen Wang, Rohun Kulkarni, Li Fei-Fei, Silvio Savarese, Yuke Zhu, and Roberto Martín-Martín. What matters in learning from offline human demonstrations for robot manipulation. In *Proceedings of the 5th Conference on Robot Learning (CoRL)*, pages 3135–3166. PMLR, 2022.

[Niu *et al.*, 2024] Haoyi Niu, Jianming Hu, Guyue Zhou, and Xianyuan Zhan. A comprehensive survey of cross-domain policy transfer for embodied agents. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24)*, pages 8197–8206. International Joint Conferences on Artificial Intelligence Organization, 2024.

[Singh *et al.*, 2024] Ritvik Singh, Arthur Allshire, Ankur Handa, Nathan Ratliff, and Karl Van Wyk. Dextrah-rgb: Visuomotor policies to grasp anything with dexterous hands. *arXiv preprint arXiv:2412.01791*, 2024.

[Suresh *et al.*, 2023] Sudharshan Suresh, Zilin Si, Stuart Anderson, Michael Kaess, and Mustafa Mukadam. Midastouch: Monte-carlo inference over distributions across sliding touch. In *Conference on Robot Learning*, pages 319–331. PMLR, 2023.

[Thananjeyan *et al.*, 2021] Brijen Thananjeyan, Ashwin Balakrishna, Suraj Nair, Michael Luo, Krishnan Srinivasan, Minho Hwang, Joseph E Gonzalez, Julian Ibarz, Chelsea Finn, and Ken Goldberg. Recovery rl: Safe reinforcement learning with learned recovery zones. *IEEE Robotics and Automation Letters*, 6(3):4915–4922, 2021.

[Xue *et al.*, 2023] Wanqi Xue, Bo An, Shuicheng Yan, and Zhongwen Xu. Reinforcement learning from diverse human preferences. *arXiv preprint arXiv:2301.11774*, 2023.

[Yang *et al.*, 2024] Libing Yang, Yang Li, and Long Chen. Clothppo: A proximal policy optimization enhancing framework for robotic cloth manipulation with observation-aligned action spaces. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24)*, pages 6895–6903. International Joint Conferences on Artificial Intelligence Organization, 2024.

[Yuan *et al.*, 2025] Haoqi Yuan, Ziye Huang, Ye Wang, Chuan Mao, Chaoyi Xu, and Zongqing Lu. Demograsp: Universal dexterous grasping from a single demonstration. *arXiv preprint arXiv:2509.22149*, 2025.

[Zhao *et al.*, 2024] Zhou Zhao, Wenhao He, and Zhenyu Lu. Tactile-based grasping stability prediction based on human grasp demonstration for robot manipulation. *IEEE Robotics and Automation Letters*, 9(3):2646–2653, 2024.