TACO: TActile World Model as a Self-COrrector for Scalable VLA Post-Training

Shengbang Liu[1,3*], Yueru Jia[1,2*], Yuyang Yan[1*], Jiaming Liu[1*†], Xinran Zhang[1,2], Qiuxuan Feng[1], Yandong Guo[2], Shiji Zhou[4], Boxin Shi[1], Shanghang Zhang[1✉]

[1]State Key Laboratory of Multimedia Information Processing, School of Computer Science, Peking University; [2]AI[2] Robotics; [3]Sun Yat-sen University; [4]Beihang University

*Equal Contribution, †Project Lead, ✉Corresponding Author

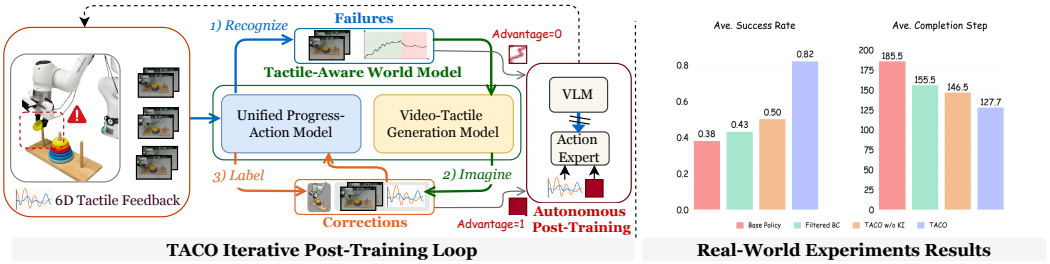


Figure 1: **Overview.** TACO is a tactile-aware world-model-driven framework for scalable VLA post-training. Given real-world rollouts, TACO follows a *Recognize–Imagine–Label* loop to identify failure-adjacent states, generate local visuo-tactile corrections, and label corrective actions. These corrections are used for advantage-conditioned post-training with knowledge-insulated tactile adaptation, improving contact recovery without repeated human intervention.

Abstract: Vision-Language-Action (VLA) models have shown promising generalization in robotic manipulation, but they still struggle with contact-rich tasks, where minor contact perturbations can cause unrecoverable failures that are hard to detect from vision alone. Since these failures are localized rather than task-level semantic errors, tactile-aware corrective post-training offers an efficient way to improve recovery. However, scaling such supervision through human intervention is costly. Recent works have explored world models to synthesize imagined rollouts for policy improvement, but vision-only world models may produce visually plausible yet contact-inconsistent trajectories. We therefore introduce **TACO**, a tactile-aware world-model-driven framework for scalable VLA post-training in contact-rich manipulation. Given real robot rollouts, TACO follows a *Recognize–Imagine–Label* loop with a tactile-aware world model: a unified progress-action model recognizes failure-adjacent states using progress estimates, a visuo-tactile generation model imagines local correction segments, and the progress-action model labels them with executable corrective actions. To incorporate tactile corrective supervision into VLA post-training, TACO combines *knowledge-insulated tactile adaptation* with advantage-conditioned training, enabling the policy to learn from imagined corrections without degrading pretrained visual-language priors. These components enable TACO to convert real-world failures into imagined visuo-tactile corrections for iterative VLA post-training. Experiments on real-world contact-rich manipulation tasks show that TACO achieves 44% absolute success rate improvement over the base policy and 32% over the policy without knowledge-insulated tactile adaptation.

Keywords: Robotic Manipulation, Tactile World Model

1 Introduction

Vision-Language-Action (VLA) models have recently shown promising progress in robotic manipulation by transferring vision-language priors to robot action generation [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13], but remain fragile in contact-rich tasks, where minor contact perturbations can cause unrecoverable failures. In real-world rollouts, such failures often occur near contact transitions, where visual observations change only slightly while tactile signals shift significantly due to slippage, insufficient pressure, or abnormal torque. For example, in *Wipe Whiteboard*, the eraser may cover the target mark without enough force to remove it; in *Twist Bottle Cap*, the gripper may align with the cap but fail to generate effective twisting torque. These failures are localized rather than semantic: the policy knows what to do, but cannot recover when contact shifts unexpectedly. This suggests that tactile-aware corrective post-training can improve recovery from localized contact failures by focusing supervision on contact-sensitive stages rather than entire trajectories.

However, collecting corrective demonstrations through human intervention is difficult to scale, as it requires repeated monitoring and manual recovery at failure states [14, 15, 16, 17, 18, 19, 20, 21], making large-scale post-training expensive. Recent works therefore explore world models to synthesize additional imagined rollouts for policy improvement [22, 23, 24, 25, 26, 27, 28, 29]. Yet vision-only world models remain unreliable in contact-rich settings, as visually plausible rollouts may still contain inconsistent contact dynamics. This motivates a tactile-aware world model that aligns visual evolution with tactile dynamics when imagining local corrective segments.

Motivated by these observations, we propose **TACO**, a tactile-aware world-model-driven framework for scalable VLA post-training in contact-rich manipulation. Given real robot rollouts, TACO uses a tactile-aware world model for a *Recognize–Imagine–Label* procedure. First, a unified progress-action model *recognizes* failure-adjacent states by estimating progress from visual and tactile signals. Then a visuo-tactile generation model *imagines* correction segments by jointly denoising future video and force sequences, with temporal RoPE aligning video and force tokens within self-attention. Finally, the same progress-action model *labels* the imagined segments with executable corrective actions, producing tactile-aware supervision for post-training.

To incorporate tactile corrective supervision into VLA post-training, a key challenge is that naively fine-tuning the full model may erode pretrained visual-language priors. TACO therefore adopts *knowledge-insulated tactile adaptation*, which applies stop-gradient to isolate the pretrained VLM backbone while routing tactile learning to the action expert [30]. To leverage failures and avoid overfitting to successful trajectories, TACO further adopts *advantage-conditioned training* as an offline reinforcement learning objective [31, 32], with binary advantage labels separating corrective segments from failures. These designs together enable an autonomous correction loop that converts real-world failures into visuo-tactile corrections for iterative VLA post-training.

We evaluate TACO on six real-world manipulation tasks. TACO achieves 44% absolute success rate improvement over the base policy and 32% over the policy without knowledge-insulated tactile adaptation. Ablation studies further validate the generation and effectiveness of the imagined correction data. In summary, our main contributions are as follows:

1) We develop a tactile-aware world model with a visuo-tactile generation model that jointly denoises video and force sequences via temporal RoPE alignment, and a unified progress-action model that predicts progress and corrective actions from visual and tactile signals.

2) We propose an iterative *Recognize–Imagine–Label* framework for supervision generation, which recognizes failure-adjacent contact states, imagines visuo-tactile corrections, and labels the actions needed for recovery.

3) We introduce knowledge-insulated tactile adaptation for VLA post-training, which incorporates tactile corrective supervision without eroding pretrained visual-language priors, together with advantage-conditioned training to separate corrective segments from failures.

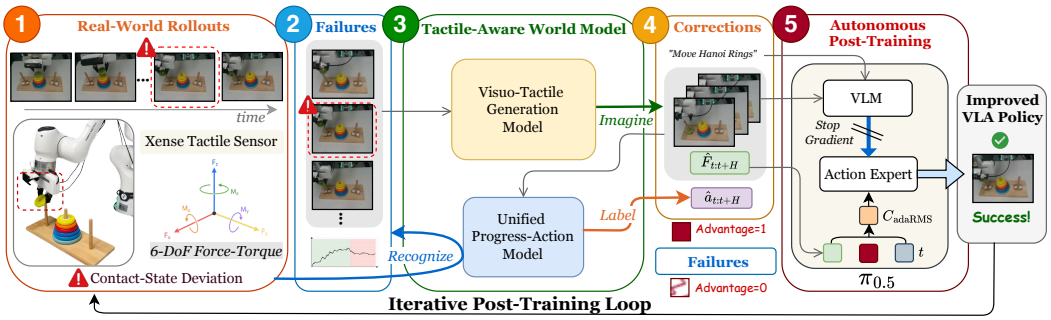


Figure 2: **TACO framework.** TACO follows an iterative *Recognize–Imagine–Label* loop over real-world rollouts: it identifies failure-adjacent states, imagines visuo-tactile recovery segments with a tactile-aware world model, and labels the corresponding actions. The resulting supervision is used for advantage-conditioned post-training through knowledge-insulated tactile adaptation.

2 Related Work

World Models for Robot Learning. World models predict future observations or visual states, and recent robot-learning methods use them to generate visual trajectories of task execution [22, 23, 33]. To turn imagined futures into policy-improvement signals, inverse dynamics models infer actions from generated videos [34, 35, 36], while reward models provide dense feedback for failure localization and rollout selection [37, 38, 39, 40, 41]. Recent works use world models as simulators for RL-based policy refinement and policy post-training [42, 43, 44, 27, 25, 24, 45, 26, 46]. Together, these components and frameworks close the loop between imagined experience and policy improvement. Unlike prior methods that rely on long-horizon prediction or online human correction [14, 15, 16, 17, 18, 19, 20, 21], we use real rollouts to recognize failure states, imagine segments with tactile estimation, and relabel them into corrective action data for iterative VLA post-training.

Tactile-Aware Robot Learning. Vision-Language-Action (VLA) and World Action Models (WAM) have shown strong semantic grounding and task generalization, yet remain limited in contact-rich manipulation due to the weak observability of physical interactions from vision alone. Recent tactile robot policies incorporate contact feedback such as force–torque measurements and visuo-tactile images to improve contact awareness, force-sensitive control, and physical grounding [47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60]. Beyond policy conditioning, recent visuo-tactile world models further show that tactile signals can improve the physical fidelity of imagined rollouts, suggesting that touch should be modeled as part of the environment dynamics rather than used only as an auxiliary policy input [61, 62]. However, naively adding tactile inputs can impair pre-contact perception and grounding [63]. We address this by using tactile feedback for both tactile-aware action prediction and force-conditioned world-model imagination. Our training recipe further protects pretrained VLM representations from tactile-action losses [30], preserving VLA priors during tactile corrective supervision.

3 Method

TACO is a tactile-aware world-model-driven framework for scalable VLA post-training in contact-rich manipulation. Given real robot rollouts, TACO converts real-world failures into visuo-tactile corrective supervision through a *Recognize–Imagine–Label* loop: recognizing failure-adjacent states, imagining local correction segments, and labeling executable corrective actions. The tactile-aware world model and iterative correction paradigm are described in Sections 3.1 and 3.2, followed by knowledge-insulated tactile adaptation for VLA post-training in Section 3.3.

3.1 Tactile-Aware World Model

The tactile-aware world model converts failure-adjacent states into local visuo-tactile corrections for VLA post-training. It combines a visuo-tactile generation model for joint video-force denoising with a unified progress-action model for progress estimation and corrective action prediction.

Visuo-Tactile Generation Model. The visuo-tactile generation model produces correction segments from failure-adjacent states via *joint video-force denoising*. Built on Wan2.2-TI2V-5B [64], it is first fine-tuned on broad robot trajectories for visual fidelity and robot-scene consistency, then adapted to contact-rich demonstrations with sliding windows.

Given video latent tokens $X^v \in \mathbb{R}^{B\times N_v\times d}$ and a force sequence $F \in \mathbb{R}^{B\times T\times 12}$, where B, N_v, T, and d denote the batch size, number of video tokens, force length, and hidden dimension, respectively, we tokenize force signals as $X^f = T_\eta(F) \in \mathbb{R}^{B\times T\times d}$. The 12 force dimensions correspond to concatenated left/right 6-DoF force-torque readings. We concatenate video and force tokens as $X = [X^v; X^f] \in \mathbb{R}^{B\times(N_v+T)\times d}$, enabling bidirectional video-force interaction within DiT self-attention. After denoising, the outputs are decoded into future video and 12-D force trajectories. For training, video and force share the same sampled denoising timestep. Let (ξ_1^v, ξ_1^f) denote clean video-latent and force segments, and (ξ_0^v, ξ_0^f) denote the corresponding Gaussian noise. The model predicts video and force flow fields u_ψ^v and u_ψ^f with the joint flow-matching loss:

$$\mathcal{L}_{\text{joint}} = \left\|u_\psi^v - (\xi_1^v - \xi_0^v)\right\|_2^2 + \lambda_f \left\|u_\psi^f - (\xi_1^f - \xi_0^f)\right\|_2^2,$$

where λ_f balances the force-denoising term.

We introduce temporal RoPE alignment and first-frame force anchoring for stable video-force joint denoising. Since Wan2.2 applies RoPE over a 3D video latent grid, while force tokens only contain temporal force-torque information, we align each force token to the video latent temporal axis. Given force-token length T and video latent temporal length f, the i-th force token is assigned

$$\rho(i) = \text{round}\left(\frac{i}{T-1}(f-1)\right), \quad i = 0, \ldots, T-1.$$

Each force token uses temporal RoPE at $\rho(i)$ with spatial RoPE set to $1 + 0j$; we keep $F_0 \in \mathbb{R}^{12}$ clean as a first-frame anchor to reduce contact-state ambiguity and stabilize prediction.

Unified Progress-Action Model. Given an RGB frame I_t and force-tactile signal $F_t \in \mathbb{R}^{12}$, the model predicts a corrective action and task progress as $(\hat{a}_t, \hat{p}_t) = U_\phi(I_t, F_t)$, where $\hat{a}_t \in \mathbb{R}^7$ and $\hat{p}_t \in [0, 1]$. It uses a DINOv2 [65]-based visual pathway with a direction-aware decoder for spatial grounding, and an MLP-based tactile pathway for normalized 12-D force-torque encoding. The fused embedding $[z_t^v; z_t^f]$ is passed to an action head $\hat{a}_t = h_a([z_t^v; z_t^f])$ and a progress head $\hat{p}_t = \sigma(h_p([z_t^v; z_t^f]))$. This joint design allows contact cues to guide both corrective action generation and progress estimation. We train it with a joint action-progress objective:

$$\mathcal{L}_{\text{UPA}} = \text{SmoothL1}(\hat{a}_t, a_t) + m_t\|\hat{p}_t - p_t\|_2^2,$$

where m_t indicates whether a valid progress label is available. The unified progress-action model therefore labels imagined correction segments with both corrective actions and progress values for advantage-conditioned VLA post-training.

3.2 TACO Iterative Correction Framework

Recognize Failure-Adjacent States. At iteration k, we deploy the current policy $\pi_\theta^{(k)}$ to collect robot rollouts. Rather than treating rollout states uniformly, TACO recognizes failure-adjacent states where task progress stalls or decreases. Given a rollout τ, the unified progress-action model predicts a dense progress score p_t at each timestep. We then select correction anchors as $\mathcal{S}_{\text{anchor}}^{(k)} = \{(\tau, t) \mid \tau \in \mathcal{D}_{\text{roll}}^{(k)},\ p_{t+\Delta} - p_t < \epsilon\}$, where Δ is a short window and ϵ a progress threshold.

Imagine Visuo-Tactile Corrections. Starting from each anchor state, TACO imagines a local correction over $T = 49$ timesteps. Specifically, the visuo-tactile generation model denoises future video

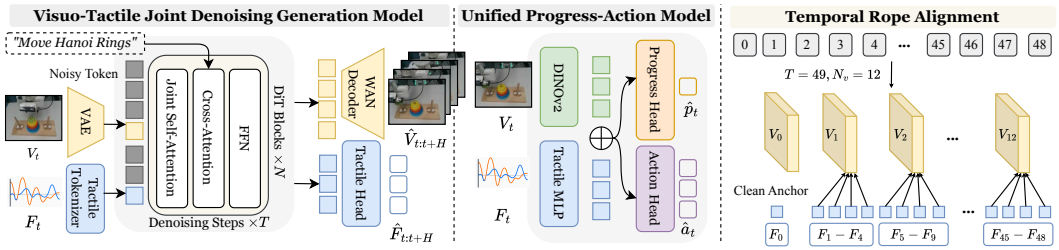


Figure 3: **Tactile-Aware World Model architecture.** The world model imagines visuo-tactile correction segments through joint denoising, aligns force tokens with video latent tokens via temporal RoPE, and converts the imagined rollouts into corrective actions.

and force sequences conditioned on the current visual observation, force, and language instruction, producing a locally plausible correction segment $(\hat{I}_{t:t+T}, \hat{F}_{t:t+T}) \sim G_\psi(\cdot \mid I_t, F_t, l)$ that captures both visual evolution and contact-force dynamics around the failure-adjacent state.

Label Actions. Finally, TACO labels each imagined visuo-tactile segment to turn it into executable supervision. Given the imagined visual and tactile signals, the unified progress-action model predicts actions and dense progress scores as $(\hat{a}_{t:t+T}, \hat{p}_{t:t+T}) = U_\phi(\hat{I}_{t:t+T}, \hat{F}_{t:t+T})$. We assign a binary advantage label $y_t \in \{0, 1\}$ to distinguish recovery-oriented corrections from initial failed segments, where $y_t = 1$ denotes a valid correction and $y_t = 0$ denotes an initial failure. These labels serve as recovery supervision for advantage-conditioned VLA post-training.

3.3 Knowledge-Insulated Tactile Adaptation

At each iteration, we post-train the VLA with original demonstrations, real rollouts, and imagined corrections. Directly optimizing the entire VLA on tactile-heavy correction data can degrade the pretrained visual-language knowledge needed for pre-contact perception and grounding. To avoid this, TACO blocks tactile-action gradients from updating the pretrained VLM backbone and confines tactile-aware learning to the action expert. Specifically, image, language, and state tokens are encoded as VLM prefix tokens, while force history and advantage are injected only into the action expert through adaRMSNorm conditioning. Only the tactile encoder, adaptation layers, and action expert are optimized during post-training. After post-training, the updated policy $\pi_\theta^{(k+1)}$ is deployed again to collect new rollouts, forming a closed real-to-imagine-to-real loop that progressively reduces contact-sensitive failures.

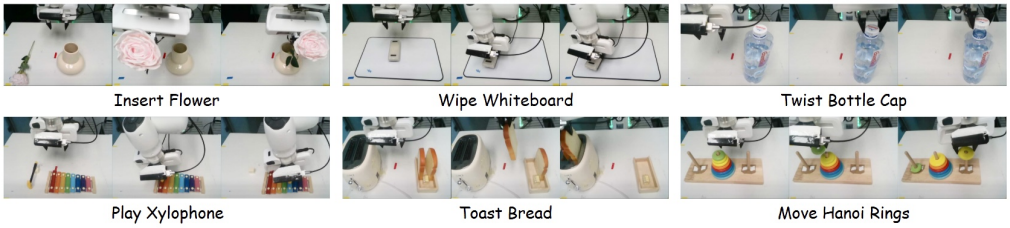
Advantage-Conditioned Post-Training. We train the action expert with force and advantage conditioning to distinguish high-progress corrections from stalled or ambiguous behaviors. Given a noisy action chunk $x_\sigma = \sigma\epsilon + (1-\sigma)a_t$, the action expert predicts the flow velocity under classifier-free guidance:

$$\mathcal{L}_\pi = \mathbb{E}\Big[\|u_\theta(x_\sigma, \sigma \mid z_t, \tilde{c}_{\text{adaRMS}}) - (\epsilon - a_t)\|_2^2\Big], \qquad c_{\text{adaRMS}} = c_t + \lambda_f c_f + \lambda_a c_a.$$

Here, z_t denotes the VLM prefix representation, c_t is the flow timestep condition, and c_f and c_a are force and advantage conditions encoded from force history and segment advantage, respectively. During training, $\tilde{c}_{\text{adaRMS}}$ is obtained by randomly replacing c_{adaRMS} with a null condition, allowing the action expert to learn both conditional and unconditional predictions. At inference time, we use the positive advantage condition to encourage high-progress tactile recovery behaviors.

4 Experiments

Section 4.1 introduces the experimental setup. Section 4.2 reports results on real-world tasks through two iterations of post-training, while Section 4.3 presents ablation studies of key components. Section 4.4 analyzes action distributions and OOD generalization.



Method	*Insert Flower*		*Wipe Whiteboard*		*Twist Bottle Cap*		*Play Xylophone*		*Toast Bread*		*Move Hanoi Rings*		Ave	
	SR	CS	SR	CS	SR	CS	SR	CS	SR	CS	SR	CS	SR	CS
Base Policy	0.50	250	0.51	151	0.45	131	0.46	132	0.30	183	0.08	266	0.38	185.5
Iteration 1														
Filtered BC	0.55	274	0.54	120	0.50	62	0.49	128	0.32	189	0.07	120	0.41	148.8
TACO (w/o KI)	0.55	233	0.33	100	0.55	58	0.58	144	0.48	193	0.42	201	0.49	154.8
TACO	0.70	207	0.55	95	0.85	56	0.63	115	0.70	184	0.51	194	0.66	141.8
Iteration 2														
Filtered BC	0.52	289	0.57	133	0.48	79	0.51	125	0.36	177	0.11	130	0.43	155.5
TACO (w/o KI)	0.62	223	0.35	98	0.65	**51**	0.52	120	0.51	191	0.37	196	0.50	146.5
TACO	**0.93**	**169**	**0.65**	**87**	**0.98**	52	**0.78**	**97**	**0.81**	**177**	**0.79**	**184**	**0.82**	**127.7**

Table 1: **Quantitative Results on Real-World Tasks.** We report success rate (SR) and completion steps (CS) for six contact-rich manipulation tasks. Iteration 1 and Iteration 2 denote the first and second rounds of post-training, respectively. Ave reports the mean SR and CS across all tasks.

4.1 Experiment Setup

Data Collection. We evaluate TACO on real-world contact-rich manipulation tasks using a single-arm Franka Research 3 robot, equipped with a front-view camera and two Xense tactile sensors mounted on the gripper. We perform six tasks: 1) *Insert Flower*, 2) *Wipe Whiteboard*, 3) *Twist Bottle Cap*, 4) *Play Xylophone*, 5) *Toast Bread* and 6) *Move Hanoi Rings*. Each task includes 50 demonstration trajectories collected via SpaceMouse teleoperation.

Training and Evaluation Details. For each task, we first warm-start $\pi_{0.5}$ [7] on the collected demonstrations as the base policy. Then, we deploy the current policy to collect rollouts, generate imagined correction data with the tactile-aware world model, and post-train on the demonstrations, experience data, and imagined corrections. We run two iterations per task. Each method is evaluated over 40 independent episodes per task with randomized tabletop object positions.

4.2 Main Results

Visualization of Imagined Correction Data. As shown in Figure 4, we compare real-world failure rollouts (top) with the imagined corrections generated by the tactile-aware world model (bottom). After recognizing the failure-adjacent states, the tactile-aware world model generates locally consistent imagined correction data that predict the correct trajectory toward task completion. We take two representative tasks as examples. In *Move Hanoi Rings*, the real rollout misaligns the ring with the peg and fails to insert it, whereas the imagined correction adjusts the contact to align the ring and seats it onto the peg; in *Twist Bottle Cap*, the real rollout contacts the cap but slips without effective twisting torque, whereas the imagined correction maintains firm contact and completes the twist. These corrections recover contact transitions that are weakly observable from RGB alone, yielding the corrective supervision used for post-training.

Quantitative and Qualitative Results. As shown in Table 1, TACO consistently improves over the base policy and baselines across two post-training iterations. After the second iteration, TACO improves the average success rate by 44% over the base policy, 39% over *Filtered BC*, and 32% over

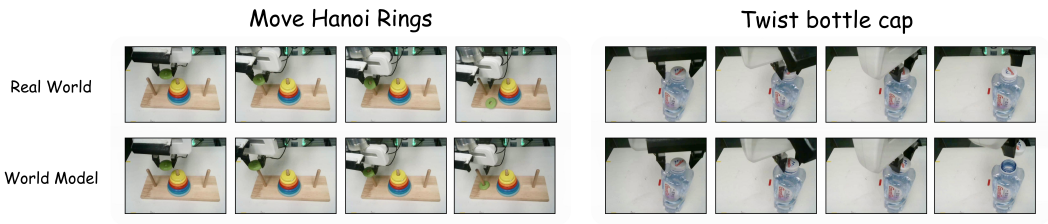


Figure 4: **Visualization of imagined correction data.** The tactile-aware world model generates locally consistent imagined corrections (bottom) that recover failed contact interactions (top).

Setting	Visuo-Tactile Generation		Progress-Action Model		Val. Loss		Progress Eval.		Real SR ↑
	Input	Output	Input	Output	$F\downarrow$	$A\downarrow$	VOC ↑	FL ↑	
w/o tactile generation	V	V	V	$A+R+F$	0.004	0.025	0.78	0.87	0.28
w/o tactile labeling	$V+F$	$V+F$	V	$A+R$	**0.002**	0.038	0.88	0.90	0.65
TACO	$\mathbf{V+F}$	$\mathbf{V+F}$	$\mathbf{V+F}$	$\mathbf{A+R}$	**0.002**	**0.019**	**0.94**	**0.95**	**0.82**

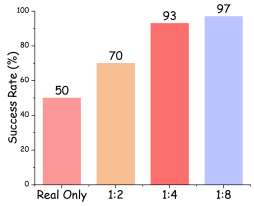


Figure 5: Ablation Study. Left: Generation of Imagined Correction Data. V, F, A, and R denote video, force, action, and progress, respectively. Validation losses are reported for force prediction (F) and action prediction (A). VOC denotes Video frame-wise progress rank correlation, FL denotes failure localization accuracy, and Real SR denotes the real-world success rate. **Right:** Scaling of Imagined Correction Data.

the policy without knowledge-insulated tactile adaptation. *Filtered BC* yields only a modest gain, as its filtered successful rollouts contain no recovery behavior at failure-adjacent contact states and keep reinforcing the narrow demonstration manifold, so its improvement saturates across iterations. In contrast, TACO preserves the pretrained visual-language priors through knowledge-insulated tactile adaptation, while using imagined visuo-tactile corrections to supervise failure-adjacent contact transitions. This enables the policy to both approach the target reliably and adjust contact force more effectively, leading to higher success rates on contact-rich tasks. TACO also achieves fewer average completion steps on most tasks, indicating smoother execution with fewer pauses, redundant motions, and indecisive contact transitions.

4.3 Ablation Study

Generation of Imagined Correction Data. As shown in Figure 5, we first validate the necessity of tactile feedback in the tactile-aware world model. We compare our full model against two alternatives: (a) a visual-only world generation setting, where the world model predicts future visual observations without generating tactile trajectories; and (b) a tactile-free labeling setting, where tactile trajectories are generated during imagination but are excluded from the progress-action labeling stage. For Real-World Experiment, removing tactile generation causes the success rate to drop to 28%, indicating that visual imagination alone struggles to capture contact-state transitions that are weakly observable from RGB observations. Excluding tactile feedback from the labeling stage reduces the performance to 65%, suggesting that tactile signals are most effective when they directly participate in corrective action and progress prediction rather than serving only as auxiliary imagined observations. These results show that tactile feedback plays complementary roles in both imagined dynamics generation and corrective labeling.

Scaling of Imagined Correction Data. As shown in Figure 5, we further evaluate the effectiveness of imagined correction data in *Insert Flower* under the same training recipe. Compared with the Base Policy, incorporating imagined correction data consistently improves the average success rate. When increasing the real-to-imagined data ratio from 1:2 to 1:4 and 1:8, the success rate increases from 70% to 93% and 97%, respectively. Notably, the 1:8 setting further outperforms 1:4, indicating that larger-scale imagined corrections provide broader coverage of failure-adjacent contact states.

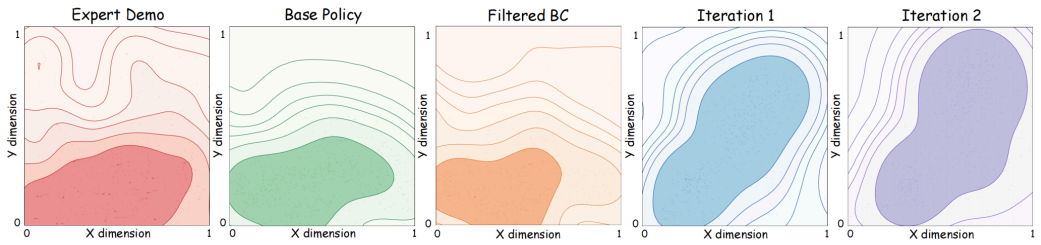


Figure 6: **Action Distribution Analysis.** We project the end-effector poses (X-Y dimensions) from 40 successful rollouts under different configurations in *Insert Flower* task.

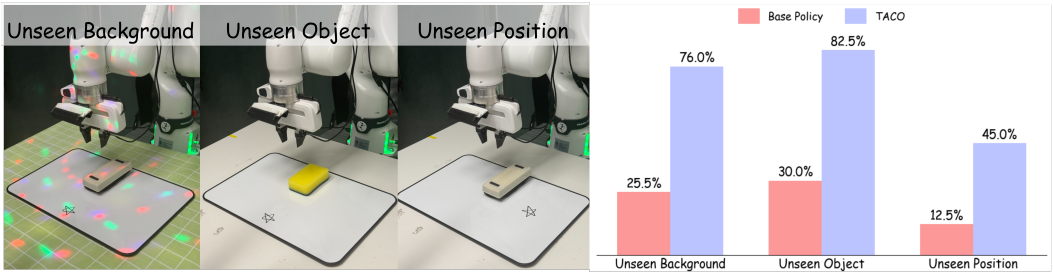


Figure 7: **Generalization Performance.** Generalization experiments under unseen backgrounds, unseen objects, and unseen object positions.

These results show that imagined data effectively improves policy post-training while substantially reducing the cost of correction data collection.

4.4 Analysis

Action Distribution Analysis. To further analyze how TACO improves action learning, we visualize the action distributions on the *Insert Flower* task. During evaluation, the initial positions of the flower and vase are fixed. Following the task setup, we project the end-effector (EE) poses along the world-frame z-axis and compare their distributions in the x-y plane. For each policy, we collect 40 successful rollouts and record the EE poses. We compare five configurations: (a) expert demonstrations, (b) base policy, (c) Filtered BC (d) TACO Iteration 1, and (e) TACO Iteration 2. As shown in Figure 6, the base policy (b) exhibits a narrowly concentrated action distribution around the demonstration manifold, making it sensitive to accumulated execution errors. *Filtered BC* reinforces the trajectories the base policy already covers therefore cannot expand beyond the original manifold, whereas our method progressively yields a broader distribution across iterations (d, e), exposing the policy to more diverse successful adjustment behaviors. These results suggest that TACO broadens the action space of the policy beyond the narrow demonstration manifold, enabling it to recover and complete the task even in scenarios unseen in the expert data.

Generalization Experiment. To evaluate whether TACO supports fast adaptation to unseen scenarios, we assess the models under three distinct paradigms: *Unseen Backgrounds*, *Unseen Objects*, and *Unseen Positions*. We compare the base policy, trained only on in-domain expert demonstrations, with the policy updated by one round of TACO using out-of-domain (OOD) imagined correction data. As shown in Figure 7, the base policy suffers clear degradation across all three shifts. In contrast, TACO achieves consistently higher success rates after only one adaptation iteration: by generating imagined correction data around OOD states, it exposes the policy to diverse successful recovery behaviors without requiring additional expert demonstrations in the target scenarios. Notably, these OOD settings lie outside the world model's training distribution in vision, tactile, and action, yet it still generates effective corrections, indicating that the tactile-aware world model generalizes robustly beyond its training data. These results show that TACO not only improves in-

domain contact-rich execution but also offers an efficient mechanism for adapting VLA policies to novel positions, objects, and visually perturbed environments.

5 Conclusion and Limitations

We presented TACO, a tactile-aware world-model-driven framework for scalable VLA post-training in contact-rich manipulation. Following a Recognize–Imagine–Label loop, TACO converts real-world failures into imagined corrections without repeated human intervention: a tactile-aware world model jointly denoises future video and force sequences, while a unified progress-action model recognizes failure-adjacent states and labels imagined segments with corrective actions. To incorporate this supervision without eroding pretrained priors, we introduced knowledge-insulated tactile adaptation with advantage-conditioned training. Experiments on real-world contact-rich tasks show that TACO improves the average success rate by 44% over the base policy and generalizes to unseen backgrounds, objects, and positions. One limitation is that imagined corrections are generated offline rather than online during deployment, while our main insight is to use tactile-aware imagination to recover localized contact failures. Future work will explore online correction generation and tighter coupling between the world model and policy.

References

[1] A. Brohan, N. Brown, J. Carbajal, Y. Chebotar, J. Dabis, C. Finn, K. Gopalakrishnan, K. Hausman, A. Herzog, J. Hsu, et al. Rt-1: Robotics transformer for real-world control at scale. *arXiv preprint arXiv:2212.06817*, 2022.

[2] B. Zitkovich, T. Yu, S. Xu, P. Xu, T. Xiao, F. Xia, J. Wu, P. Wohlhart, S. Welker, A. Wahid, et al. Rt-2: Vision-language-action models transfer web knowledge to robotic control. In *Conference on Robot Learning*, pages 2165–2183. PMLR, 2023.

[3] A. O'Neill, A. Rehman, A. Maddukuri, A. Gupta, A. Padalkar, A. Lee, A. Pooley, A. Gupta, A. Mandlekar, A. Jain, et al. Open x-embodiment: Robotic learning datasets and rt-x models: Open x-embodiment collaboration 0. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6892–6903. IEEE, 2024.

[4] O. M. Team, D. Ghosh, H. Walke, K. Pertsch, K. Black, O. Mees, S. Dasari, J. Hejna, T. Kreiman, C. Xu, et al. Octo: An open-source generalist robot policy. *arXiv preprint arXiv:2405.12213*, 2024.

[5] M. J. Kim, K. Pertsch, S. Karamcheti, T. Xiao, A. Balakrishna, S. Nair, R. Rafailov, E. Foster, G. Lam, P. Sanketi, et al. Openvla: An open-source vision-language-action model. *arXiv preprint arXiv:2406.09246*, 2024.

[6] K. Black, N. Brown, D. Driess, A. Esmail, M. Equi, C. Finn, N. Fusai, L. Groom, K. Hausman, B. Ichter, et al. π_0 : A vision-language-action flow model for general robot control. *arXiv preprint arXiv:2410.24164*, 2024.

[7] P. Intelligence, K. Black, N. Brown, J. Darpinian, K. Dhabalia, D. Driess, A. Esmail, M. Equi, C. Finn, N. Fusai, et al. $\pi_{0.5}$: a vision-language-action model with open-world generalization. *arXiv preprint arXiv:2504.16054*, 2025.

[8] P. Intelligence, B. Ai, A. Amin, R. Aniceto, A. Balakrishna, G. Balke, K. Black, G. Bokinsky, S. Cao, T. Charbonnier, et al. $\pi_{0.7}$: a steerable generalist robotic foundation model with emergent capabilities. *arXiv preprint arXiv:2604.15483*, 2026.

[9] Q. Li, Y. Liang, Z. Wang, L. Luo, X. Chen, M. Liao, F. Wei, Y. Deng, S. Xu, Y. Zhang, et al. Cogact: A foundational vision-language-action model for synergizing cognition and action in robotic manipulation. *arXiv preprint arXiv:2411.19650*, 2024.

[10] S. Liu, L. Wu, B. Li, H. Tan, H. Chen, Z. Wang, K. Xu, H. Su, and J. Zhu. Rdt-1b: a diffusion foundation model for bimanual manipulation. In *International Conference on Learning Representations*, volume 2025, pages 29982–30009, 2025.

[11] J. Wen, Y. Zhu, M. Zhu, Z. Tang, J. Li, Z. Zhou, X. Liu, C. Shen, Y. Peng, and F. Feng. Diffusionvla: Scaling robot foundation models via unified diffusion and autoregression. In *Forty-second International Conference on Machine Learning*, 2025.

[12] Y. Jia, J. Liu, S. Liu, R. Zhou, W. Yu, Y. Yan, X. Chi, Y. Guo, B. Shi, and S. Zhang. Video2act: A dual-system video diffusion policy with robotic spatio-motional modeling. *arXiv preprint arXiv:2512.03044*, 2025.

[13] Z. Liu, J. Liu, H. Chen, J. Yu, Z. Guo, C. Hou, C. Gu, X. Mi, R. Zhang, K. Wu, et al. Last $_{\{0\}}$: Latent spatio-temporal chain-of-thought for robotic vision-language-action model. *arXiv preprint arXiv:2601.05248*, 2026.

[14] S. Ross, G. Gordon, and D. Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 627–635. JMLR Workshop and Conference Proceedings, 2011.

[15] R. Hoque, A. Mandlekar, C. Garrett, K. Goldberg, and D. Fox. Intervengen: Interventional data generation for robust and data-efficient robot imitation learning. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2840–2846. IEEE, 2024.

[16] Y. Korkmaz and E. Bıyık. Mile: Model-based intervention learning. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 15673–15679. IEEE, 2025.

[17] X. Xu, Y. Hou, Z. Liu, and S. Song. Compliant residual dagger: Improving real-world contact-rich manipulation with human corrections. *Advances in Neural Information Processing Systems*, 38:139559–139581, 2026.

[18] Y. Wang, R. Syed, F. Wu, M. Zhang, A. Onol, J. Barreiros, H. Nayyeri, T. Dear, H. Zhang, and Y. Li. Interactive world simulator for robot policy training and evaluation. *arXiv preprint arXiv:2603.08546*, 2026.

[19] Y. Li, Z. Zhou, Y. Chen, Y. Guo, J. Liu, S. Zhang, J. Chen, and Y. Zhu. Hi-wm: Human-in-the-world-model for scalable robot post-training. *arXiv preprint arXiv:2604.21741*, 2026.

[20] Q. Xu, J. Liu, R. Zhou, S. Shi, N. Han, Z. Liu, C. Gu, S. Gu, Y. Yue, G. Huang, et al. Twinrl-vla: Digital twin-driven reinforcement learning for real-world robotic manipulation. *arXiv preprint arXiv:2602.09023*, 2026.

[21] W. Yu, J. Lv, Z. Ying, Y. Jin, C. Wen, and C. Lu. Armada: Autonomous online failure detection and human shared control empower scalable real-world deployment and adaptation. *arXiv preprint arXiv:2510.02298*, 2025.

[22] S. Zhou, Y. Du, J. Chen, Y. Li, D.-Y. Yeung, and C. Gan. Robodreamer: Learning compositional world models for robot imagination. *arXiv preprint arXiv:2404.12377*, 2024.

[23] Y. Li, X. Wei, X. Chi, Y. Li, Z. Zhao, H. Wang, N. Ma, M. Lu, and S. Zhang. Manipdreamer: Boosting robotic manipulation world model with action tree and visual guidance. In *ICASSP 2026-2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 12027–12031. IEEE, 2026.

[24] Y. Guo, T. Lee, L. X. Shi, J. Chen, P. Liang, and C. Finn. Vlaw: Iterative co-improvement of vision-language-action policy and world model. *arXiv preprint arXiv:2602.12063*, 2026.

[25] F. Zhu, Z. Yan, Z. Hong, Q. Shou, X. Ma, and S. Guo. Wmpo: World model-based policy optimization for vision-language-action models. *arXiv preprint arXiv:2511.09515*, 2025.

[26] J. Yang, K. Lin, J. Li, W. Zhang, T. Lin, L. Wu, Z. Su, H. Zhao, Y.-Q. Zhang, L. Chen, et al. Rise: Self-improving robot policy with compositional world model. *arXiv preprint arXiv:2602.11075*, 2026.

[27] Z. Jiang, S. Zhou, Y. Jiang, Z. Huang, M. Wei, Y. Chen, T. Zhou, Z. Guo, H. Lin, Q. Zhang, et al. Wovr: World models as reliable simulators for post-training vla policies with rl. *arXiv preprint arXiv:2602.13977*, 2026.

[28] J. Jang, S. Ye, Z. Lin, J. Xiang, J. Bjorck, Y. Fang, F. Hu, S. Huang, K. Kundalia, Y.-C. Lin, et al. Dreamgen: Unlocking generalization in robot learning through video world models. *arXiv preprint arXiv:2505.12705*, 2025.

[29] A. Yu, Z. Chen, P. Song, Z. Hong, H. Wang, D. Zhang, T. He, Y. Ding, and D. Zhang. Wm-dagger: Enabling efficient data aggregation for imitation learning with world models. *arXiv preprint arXiv:2604.11351*, 2026.

[30] D. Driess, J. Springenberg, B. Ichter, L. Yu, A. Li-Bell, K. Pertsch, A. Ren, H. Walke, Q. Vuong, L. X. Shi, et al. Knowledge insulating vision-language-action models: Train fast, run fast, generalize better. *Advances in Neural Information Processing Systems*, 38:102867–102888, 2026.

[31] K. Frans, S. Park, P. Abbeel, and S. Levine. Diffusion guidance is a controllable policy improvement operator. *arXiv preprint arXiv:2505.23458*, 2025.

[32] P. Intelligence, A. Amin, R. Aniceto, A. Balakrishna, K. Black, K. Conley, G. Connors, J. Darpinian, K. Dhabalia, J. DiCarlo, et al. $\pi_{0.6}$: a vla that learns from experience. *arXiv preprint arXiv:2511.14759*, 2025.

[33] Q. Feng, J. Yu, J. Liu, Y. Jia, Z. Wu, H. Chen, Z. Qian, S. Gu, P. Jia, S. Ma, et al. Harmowam: Harmonizing generalizable and precise manipulation via adaptive world action models. *arXiv preprint arXiv:2605.10942*, 2026.

[34] H. Tan, Y. Feng, X. Mao, S. Huang, G. Liu, Z. Hao, H. Su, and J. Zhu. Anypos: Automated task-agnostic actions for bimanual manipulation. *arXiv preprint arXiv:2507.12768*, 2025.

[35] W. Mi, Y. Bao, X. Chi, X. Ju, Z. Qin, K. Ge, K. Tang, P. Jia, S. Zhang, and J. Tang. Tc-idm: Grounding video generation for executable zero-shot robot motion. *arXiv preprint arXiv:2601.18323*, 2026.

[36] K. Li, Z. Jing, X. Wang, Z. Zhu, Y. Zhou, G. Huang, D. Li, Q. Yang, and H. Huang. Stableidm: Stabilizing inverse dynamics model against manipulator truncation via spatio-temporal refinement. *arXiv preprint arXiv:2604.17887*, 2026.

[37] Y. J. Ma, V. Kumar, A. Zhang, O. Bastani, and D. Jayaraman. Liv: Language-image representations and rewards for robotic control. In *International Conference on Machine Learning*, pages 23301–23320. PMLR, 2023.

[38] T. Lee, A. Wagenmaker, K. Pertsch, P. Liang, S. Levine, and C. Finn. Roboreward: General-purpose vision-language reward models for robotics. *arXiv preprint arXiv:2601.00675*, 2026.

[39] J. Lv, H. Li, J. Li, Y. Nie, F. Kong, Y. Wang, X. Wang, Z. Zhu, C. Ni, Q. Deng, et al. Viva: A video-generative value model for robot reinforcement learning. *arXiv preprint arXiv:2604.08168*, 2026.

[40] H. Tan, S. Chen, Y. Xu, Z. Wang, Y. Ji, C. Chi, Y. Lyu, Z. Zhao, X. Chen, P. Co, et al. Robodopamine: General process reward modeling for high-precision robotic manipulation. *arXiv preprint arXiv:2512.23703*, 2025.

[41] A. Liang, Y. Korkmaz, J. Zhang, M. Hwang, A. Anwar, S. Kaushik, A. Shah, A. S. Huang, L. Zettlemoyer, D. Fox, et al. Robometer: Scaling general-purpose robotic reward models via trajectory comparisons. *arXiv preprint arXiv:2603.02115*, 2026.

[42] Z. Jiang, K. Liu, Y. Qin, S. Tian, Y. Zheng, M. Zhou, C. Yu, H. Li, and D. Zhao. World4rl: Diffusion world models for policy refinement with reinforcement learning for robotic manipulation. *arXiv preprint arXiv:2509.19080*, 2025.

[43] J. Xiao, Y. Yang, X. Chang, R. Chen, F. Xiong, M. Xu, W.-S. Zheng, and Q. Zhang. Worldenv: Leveraging world model as a virtual environment for vla post-training. *arXiv preprint arXiv:2509.24948*, 2025.

[44] A. K. Sharma, Y. Sun, N. Lu, Y. Zhang, J. Liu, and S. Yang. World-gymnast: Training robots with reinforcement learning in a world model. *arXiv preprint arXiv:2602.02454*, 2026.

[45] X. Liu, Z. Bai, H. Ci, K. Y. Ma, and M. Z. Shou. World-vla-loop: Closed-loop learning of video world model and vla policy. *arXiv preprint arXiv:2602.06508*, 2026.

[46] G. Team, B. Wang, B. Li, C. Ni, G. Huang, G. Zhao, H. Li, J. Li, J. Lv, J. Liu, et al. Gigabrain-0.5 m*: a vla that learns from world model-based reinforcement learning. *arXiv preprint arXiv:2602.12099*, 2026.

[47] Z. Liu, J. Liu, J. Xu, N. Han, C. Gu, H. Chen, K. Zhou, R. Zhang, K. C. Hsieh, K. Wu, et al. Mla: A multisensory language-action model for multimodal understanding and forecasting in robotic manipulation. *arXiv preprint arXiv:2509.26642*, 2025.

[48] J. Yu, H. Liu, Q. Yu, J. Ren, C. Hao, H. Ding, G. Huang, G. Huang, Y. Song, P. Cai, et al. Forcevla: Enhancing vla models with a force-aware moe for contact-rich manipulation. *Advances in Neural Information Processing Systems*, 38:93409–93439, 2026.

[49] Y. Li, H. Jiang, J. Xia, H. Zhang, J. Du, Y. Zhou, J. Zeng, C. Hao, J. Ren, Q. Yu, et al. Forcevla2: Unleashing hybrid force-position control with force awareness for contact-rich manipulation. *arXiv preprint arXiv:2603.15169*, 2026.

[50] J. Huang, S. Wang, F. Lin, Y. Hu, C. Wen, and Y. Gao. Tactile-vla: unlocking vision-language-action model's physical knowledge for tactile generalization. *arXiv preprint arXiv:2507.09160*, 2025.

[51] Z. Cheng, Y. Zhang, W. Zhang, H. Li, K. Wang, L. Song, and H. Zhang. Omnivtla: Vision-tactile-language-action model with semantic-aligned tactile sensing. *arXiv preprint arXiv:2508.08706*, 2025.

[52] C. Morissette, A. Abyaneh, W.-D. Chang, A. Houssaini, D. Meger, H.-C. Lin, J. Tremblay, and G. Dudek. Tactile modality fusion for vision-language-action models. *arXiv preprint arXiv:2603.14604*, 2026.

[53] K. Zhang, H. Zhang, Z. Xu, Z. Zhang, M. R. I. Prince, X. Li, X. Han, Y. Zhou, A. Ajoudani, and Y. She. Tacvla: Contact-aware tactile fusion for robust vision-language-action manipulation. *arXiv preprint arXiv:2603.12665*, 2026.

[54] Y. Huang, P. Lin, W. Li, D. Li, J. Li, J. Jiang, C. Xiao, and Z. Jiao. Tactile-force alignment in vision-language-action models for force-aware manipulation. *arXiv preprint arXiv:2601.20321*, 2026.

[55] J. Bi, K. Y. Ma, C. Hao, M. S. Zheng, and H. Soh. Vla-touch: Enhancing vision-language-action model with dual-level tactile feedback. *IEEE Robotics and Automation Letters*, 2026.

[56] S. Yu, K. Lin, A. Xiao, J. Duan, and H. Soh. Octopi: Object property reasoning with large tactile-language models. *arXiv preprint arXiv:2405.02794*, 2024.

[57] Z. Wang, Y. Wang, M. Ren, P. Li, Y. Liu, Y. Nie, L. Long, Y. Ye, X. Wang, Z. Zhu, et al. Tacmamba: A tactile history compression adapter bridging fast reflexes and slow vla reasoning. *arXiv preprint arXiv:2603.01700*, 2026.

[58] B. Huang, Y. Wang, X. Yang, Y. Luo, and Y. Li. 3d-vitac: Learning fine-grained manipulation with visuo-tactile sensing. *arXiv preprint arXiv:2410.24091*, 2024.

[59] K. Gubernatorov, M. Sannikov, I. Mikhalchuk, E. Kuznetsov, M. Artemov, O. F. Ouwatobi, M. Fernando, A. Asanov, Z. Guo, and D. Tsetserukou. Hapticvla: Contact-rich manipulation via vision-language-action model without inference-time tactile sensing. *arXiv preprint arXiv:2603.15257*, 2026.

[60] H. Xue, J. Ren, W. Chen, G. Zhang, Y. Fang, G. Gu, H. Xu, and C. Lu. Reactive diffusion policy: Slow-fast visual-tactile policy learning for contact-rich manipulation. *arXiv preprint arXiv:2503.02881*, 2025.

[61] Y. Zheng, S. Gu, W. Li, Y. Zheng, Y. Zang, S. Tian, X. Li, C. Hao, C. Gao, S. Liu, et al. Omnivta: Visuo-tactile world modeling for contact-rich robotic manipulation. *arXiv preprint arXiv:2603.19201*, 2026.

[62] C. Higuera, S. Arnaud, B. Boots, M. Mukadam, F. R. Hogan, and F. Meier. Visuo-tactile world models. *arXiv preprint arXiv:2602.06001*, 2026.

[63] X. Li, M. Cai, J. Xu, J. Zhu, H. Fan, Y. Shen, G. Ren, and H. Dong. At-vla: Adaptive tactile injection for enhanced feedback reaction in vision-language-action models. *arXiv preprint arXiv:2605.07308*, 2026.

[64] T. Wan, A. Wang, B. Ai, B. Wen, C. Mao, C.-W. Xie, D. Chen, F. Yu, H. Zhao, J. Yang, et al. Wan: Open and advanced large-scale video generative models. *arXiv preprint arXiv:2503.20314*, 2025.

[65] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.

[66] A. Khazatsky, K. Pertsch, S. Nair, A. Balakrishna, S. Dasari, S. Karamcheti, S. Nasiriany, M. K. Srirama, L. Y. Chen, K. Ellis, et al. Droid: A large-scale in-the-wild robot manipulation dataset. *arXiv preprint arXiv:2403.12945*, 2024.

[67] Q. Bu, J. Cai, L. Chen, X. Cui, Y. Ding, S. Feng, S. Gao, X. He, X. Hu, X. Huang, et al. Agibot world colosseo: A large-scale manipulation platform for scalable and intelligent embodied systems. *arXiv preprint arXiv:2503.06669*, 2025.

[68] K. Wu, C. Hou, J. Liu, Z. Che, X. Ju, Z. Yang, M. Li, Y. Zhao, Z. Xu, G. Yang, et al. Robomind: Benchmark on multi-embodiment intelligence normative data for robot manipulation. *arXiv preprint arXiv:2412.13877*, 2024.

TACO: TActile World Model as a Self-COrrector for Scalable VLA Post-Training

Supplementary Material

We provide additional details, as well as quantitative and qualitative results of TACO in this supplementary material. The outline is shown below.

- **A. Real-World Setup (Appendix A)**
  - Real-World Robot Configuration
  - Control and Data Collection
  - Out-of-Domain (OOD) Scenario Construction
- **B. Additional Method Details (Appendix B)**
  - Baselines
  - Algorithm Pseudocode
  - Visuo-Tactile Generation Model Details
  - Unified Progress-Action Model Details
  - TACO Iterative Correction Framework Details
  - Training Recipe Details
- **C. Additional Ablation Study (Appendix C)**
  - Evaluation Metric Details
  - Scaling of Imagined Correction Data
  - Effect of Advantage-Conditioned Training
  - Effect of Failure-Adjacent Anchor Selection
- **D. Additional Analysis (Appendix D)**
  - Failure Case Analysis
  - Additional Generalization Experiment
- **E. Additional Visualization (Appendix E)**
  - Additional Imagined Corrections Visualizations

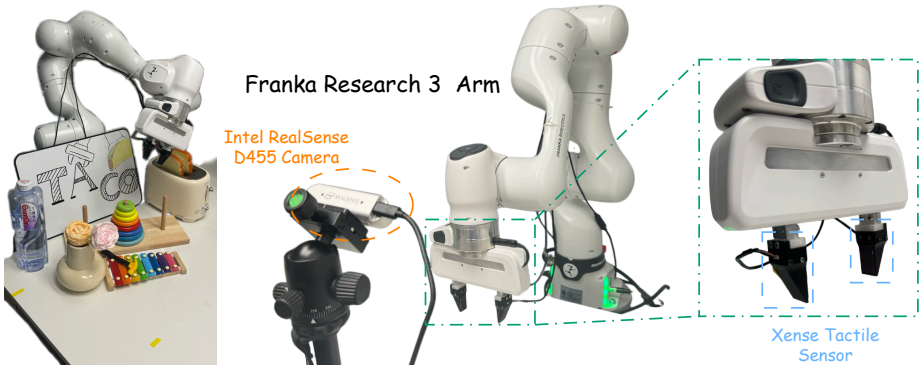
A Real-World Setup

A.1 Real-World Robot Configuration

As shown in Figure 8, our single-arm platform is built on a 7-DoF Franka Research 3 (FR3) manipulator equipped with a parallel-jaw gripper whose fingertips integrate Xense tactile sensors, enabling 6D force/torque sensing for contact-rich precision manipulation tasks. The perception system employs a single front-view Intel RealSense D455 camera, which provides a global view of the workspace using RGB images at a resolution of 640×480 pixels. At each timestep, the policy observation consists of the RGB image, the 6D force/torque readings from the Xense tactile sensors, and the robot's proprioceptive state.

A.2 Control and Data Collection

The end-effector pose, gripper state, 6D force/torque readings from the Xense tactile sensors, and other proprioceptive signals are synchronized with camera observations through a real-time communication interface. All demonstrations are collected via teleoperation using a SpaceMouse 3D



(a) Tasks overview. (b) Setup of Franka robot and tactile sensor.

Figure 8: **Real-world robot setup and experimental assets.** We use a single-arm Franka Research 3 (FR3) platform equipped with a parallel-jaw gripper and fingertip-mounted Xense tactile sensors for 6D force/torque sensing. A front-view Intel RealSense D455 camera provides global RGB observations of the workspace, together with the object assets used across our real-world contact-rich manipulation tasks.

controller, with the operator receiving real-time feedback from the camera view. As shown in Figure 9, we visualize the execution progress for the six real-world tasks from the front camera view. Detailed stage decomposition and success criteria are as follows:

1. Insert Flower. The robot picks up a flower from a random location on the table and inserts it into a vase. S1: insert the flower into the vase. S1 is considered successful if the robot securely inserts the flower into the vase without dropping it or knocking over the vase.

2. Wipe Whiteboard. The robot picks up an eraser and wipes off a star drawn at a random location on a whiteboard. S1: grasp the eraser; S2: wipe off the star. S1 is considered successful if the robot securely grasps and lifts the eraser. S2 is considered successful if the robot completely erases the star from the whiteboard.

3. Twist Bottle Cap. The robot grasps the cap of a water bottle, applies a twisting motion to loosen it, and then lifts the cap away from the bottle. S1: grasp the bottle cap; S2: twist open and lift the cap. S1 is considered successful if the robot securely grips the cap with stable contact and without displacing the bottle. S2 is considered successful if the robot rotates the cap open and lifts it upward, fully separating it from the bottle without dropping the cap.

4. Play Xylophone. The robot picks up a mallet and sequentially strikes the 1st, 3rd, 5th, and 8th keys of an eight-key xylophone in order. S1: strike the 1st key; S2: strike the 3rd key; S3: strike the 5th key; S4: strike the 8th key. Each stage is considered successful if the robot accurately strikes the designated key with the mallet tip, while following the specified order and avoiding unintended strikes on adjacent keys.

5. Toast Bread. The robot sequentially picks up two slices of bread from a box and inserts them into a toaster. S1: grasp the first slice; S2: insert the first slice into the toaster; S3: grasp the second slice; S4: insert the second slice. Each stage is considered successful if the corresponding action is completed without dropping the bread or misaligning the insertion.

6. Move Hanoi Rings. The robot moves the top ring from the middle peg of a Hanoi tower (where all rings are initially stacked in size order) to the left empty peg, then moves the next ring to the right empty peg. S1: grasp the top ring; S2: place the top ring on the left peg; S3: grasp the second ring; S4: place the second ring on the right peg. S1 is considered successful if the robot securely grasps and lifts the top ring. S2 is considered successful if the robot places the top ring on the left peg without dropping it. S3 is considered successful if the robot securely grasps and lifts the

second ring. S4 is considered successful if the robot places the second ring on the right peg without dropping it or disturbing the first ring.

A.3 Out-of-Domain (OOD) Scenario Construction

Unseen Background. To evaluate robustness against visual context shifts, we construct background OOD scenes by altering the tabletop appearance and ambient lighting. Specifically, a green tablecloth with white stripes is spread over the workspace, and a red-green-blue disco light ball is introduced to illuminate the tabletop, producing dynamically varying color reflections and highlights not encountered during training. This setting tests the model's ability to maintain reliable perception and stable manipulation under unfamiliar and visually distracting background conditions.

Unseen Object. To evaluate robustness to object-level semantic, geometric, and tactile variations, target objects are replaced with substitutes that differ significantly from the training objects in both visual appearance and physical properties. For example, the standard whiteboard eraser is replaced with a sponge, which differs not only in color and texture but also in compliance and friction characteristics, requiring the model to adapt both its visual recognition and its grasp force and interaction strategy. This setting evaluates the model's ability to generalize to novel objects when both perception and tactile dynamics deviate from training distributions.

Unseen Position. To evaluate spatial generalization, target items are placed in workspace regions outside the spatial coverage of training demonstration trajectories. For instance, the star on the whiteboard is drawn at a rear position far from the locations seen during data collection. This setting evaluates the model's capability to handle unseen spatial coordinates while maintaining accurate reaching and manipulation.

B Additional Method Details

B.1 Baselines

All methods share the same warm-started base policy. They differ in what supervision is used and how it is incorporated into the policy, as summarized in Table 3.

Base Policy warm-starts $\pi_{0.5}$ [7] on the 50 expert demonstrations per task, serving as the common initialization.

Filtered BC filters successful trajectories from real-world rollouts and performs supervised fine-tuning on these trajectories together with original demonstrations.

TACO (w/o KI) uses the same imagined corrections and advantage-conditioned objective as TACO, but fine-tunes the full VLA models instead of insulating the pretrained backbone.

B.2 Algorithm Pseudocode

Algorithm 1 summarizes the full TACO procedure as an iterative *real-to-imagine-to-real* loop. At each iteration, the current policy is deployed to collect real-world rollouts, the tactile-aware world model converts failure-adjacent states into local imagined corrections through the *Recognize–Imagine–Label* steps, and the policy is updated via knowledge-insulated tactile adaptation. The updated policy is then redeployed to collect new rollouts, progressively reducing contact-sensitive failures without repeated human intervention.

B.3 Visuo-Tactile Generation Model Details

The visuo-tactile generation model produces multimodal correction segments from failure-adjacent states. Rather than treating force feedback as a passive condition for video generation, TACO formulates correction generation as a *joint denoising* problem over both video and tactile sequences. This design is important for contact-rich manipulation: a generated video can look plausible while

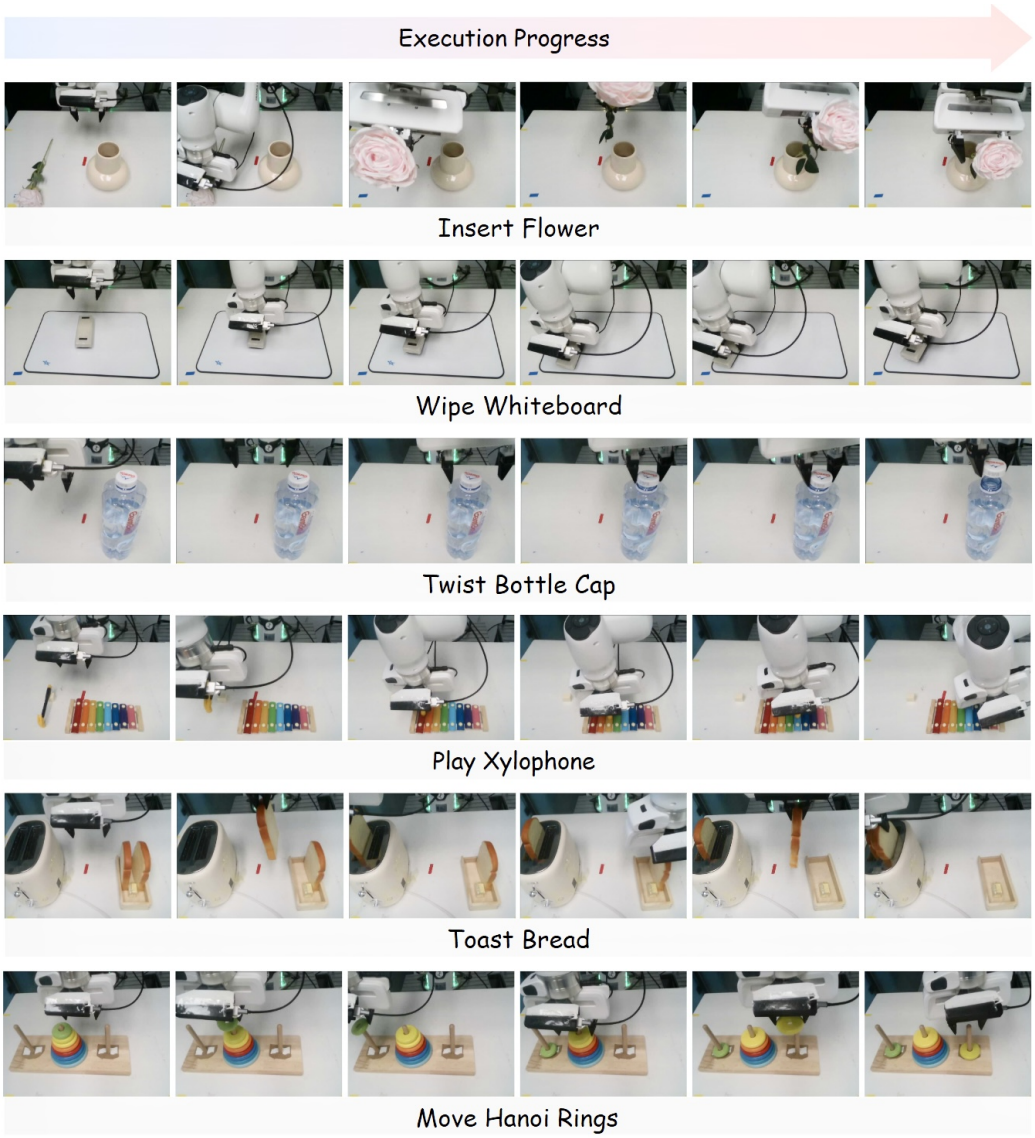


Figure 9: **Robot execution progress in real-world tasks.** We visualize key frames of the robot's execution process from the front camera view in real-world tasks.

Dataset	Robot Arm / Platform	Number of Trajectories
DROID	Franka Panda	201,119
AgiBot	AgiBot G1	3,017
RoboMIND	Franka / UR / Ark / Agilex / TienKung	1,721,985

Table 2: Statistics of the public robot datasets used for Wan2.2-TI2V-5B pretraining.

the underlying contact force is physically inconsistent. By denoising visual and force trajectories together, the model learns contact-aware futures whose visual and tactile dynamics are temporally aligned.

Our model is initialized from a Wan2.2-TI2V-5B [64] backbone pretrained on broad robot datasets, including publicly available datasets such as DROID [66], AgiBot [67], and RoboMIND [68], together with closed-source robot data. Detailed statistics of the public datasets are provided in Ta-

Algorithm 1 TACO: Tactile-Aware World Model as a Self-Corrector for VLA Post-Training

Require: Base VLA policy π_θ; visuo-tactile generation model G_ψ; unified progress-action model U_ϕ; demonstrations $\mathcal{D}_{\text{demo}}$; iterations K_{iter}; correction length T; progress window Δ; progress threshold ϵ; conditioning weights λ_f, λ_a

Ensure: Post-trained VLA policy π_θ^* for contact-rich recovery

Warm-start base policy π_θ on $\mathcal{D}_{\text{demo}}$; set $T = 49$; set $y = 1$ for all expert demonstrations ▷ demos are positive

for $k = 1$ **to** K_{iter} **do**

  (1) Real-world rollouts

  Deploy π_θ in the real world to collect rollouts $\mathcal{D}_{\text{roll}}$

  (2) Recognize failure-adjacent states and assign advantage

  $\mathcal{S}_{\text{anchor}} \leftarrow \varnothing$

  for each rollout $\tau \in \mathcal{D}_{\text{roll}}$ **do**

    $\{p_t\}_t \leftarrow U_\phi(\tau)$ ▷ dense progress

    $\mathcal{A}_\tau \leftarrow \{(\tau, t) \mid p_{t+\Delta} - p_t < \epsilon\}$ ▷ failure-adjacent anchors

    $t^* \leftarrow \min\{t : (\tau, t) \in \mathcal{A}_\tau\}$ ▷ first anchor = failure onset

    $y_t \leftarrow 1$ for $t < t^*$, $y_t \leftarrow 0$ for $t \geq t^*$ ▷ real-rollout advantage

    $\mathcal{S}_{\text{anchor}} \leftarrow \mathcal{S}_{\text{anchor}} \cup \mathcal{A}_\tau$

  end for

  (3) Imagine visuo-tactile corrections and label actions

  $\mathcal{D}_{\text{corr}} \leftarrow \varnothing$

  for each anchor $(\tau, t) \in \mathcal{S}_{\text{anchor}}$ **do**

    $(\hat{I}_{t:t+T}, \hat{F}_{t:t+T}) \sim G_\psi(\cdot \mid I_t, F_t, l)$ ▷ joint video–force denoising

    $(\hat{a}_{t:t+T}) \leftarrow U_\phi(\hat{I}_{t:t+T}, \hat{F}_{t:t+T})$ ▷ label corrective actions

    $y \leftarrow 1$ ▷ imagined corrections are positive recovery supervision

    $\mathcal{D}_{\text{corr}} \leftarrow \mathcal{D}_{\text{corr}} \cup \{(\hat{I}, \hat{F}, \hat{a}, y)\}$

  end for

  (4) Knowledge-insulated tactile adaptation

  for each minibatch from $\mathcal{D}_{\text{demo}} \cup \mathcal{D}_{\text{roll}} \cup \mathcal{D}_{\text{corr}}$ **do**

    $z_t \leftarrow \text{sg}[\,\text{VLM}(\text{image}, \text{lang}, \text{state})\,]$ ▷ insulate VLM backbone

    $c_{\text{adaRMS}} \leftarrow c_t + \lambda_f c_f + \lambda_a c_a$ ▷ force & advantage conditioning

    $\tilde{c}_{\text{adaRMS}} \leftarrow$ randomly drop c_{adaRMS} to $\varnothing$ ▷ CFG dropout

    Update action expert, tactile encoder, adaptation layers by $\min \mathcal{L}_\pi$

  end for

end for

return $\pi_\theta^* \leftarrow \pi_\theta$

ble 2. This pretraining provides strong priors for visual realism and robot-scene consistency. We then perform full fine-tuning on our real-world demonstration dataset with a Flow Matching objective, adapting the model to downstream contact-rich tasks through video-force joint denoising.

Given video latent tokens $X^v \in \mathbb{R}^{B \times N_v \times d}$ and a force sequence $F \in \mathbb{R}^{B \times T \times 12}$, we encode the force sequence with a tactile tokenizer as $X^f = T_\eta(F) \in \mathbb{R}^{B \times T \times d}$. In our implementation, we use $N_v = 12$ video latent tokens and align the force sequence at a higher temporal resolution with $T = 4N_v + 1 = 49$ steps. The DiT hidden dimension is $d = 3072$. Here, the 12 force dimensions correspond to left/right 6-DoF force-torque readings. The DiT input is formed by concatenating video and force tokens along the token dimension,

$$X = [X^v; X^f] \in \mathbb{R}^{B \times (N_v + T) \times d}.$$

This token-level formulation allows video and force tokens to interact bidirectionally within the same DiT self-attention layers, rather than treating force merely as an external condition. After DiT denoising, the output is split back into video and force tokens. The video tokens are decoded by the original Wan unpatchify pathway, while the force-token output $Y^f \in \mathbb{R}^{B \times T \times d}$ is projected by a tactile head as

$$\hat{u}^f = H_\eta(Y^f) \in \mathbb{R}^{B \times T \times 12}.$$

The model therefore predicts video and force velocity fields in a single forward pass.

For training, video and force share the same sampled denoising timestep. Let ξ_1^v, ξ_1^f denote the clean video-latent and force segments, and ξ_0^v, ξ_0^f denote the corresponding Gaussian noise. The model predicts video and force flow fields u_ψ^v and u_ψ^f with the joint flow-matching loss:

$$\mathcal{L}_{\text{joint}} = \left\| u_\psi^v - (\xi_1^v - \xi_0^v) \right\|_2^2 + \lambda_f \left\| u_\psi^f - (\xi_1^f - \xi_0^f) \right\|_2^2 .$$

B.4 Unified Progress-Action Model Details

The unified progress-action model consists of a visual encoder, a tactile encoder, and two prediction heads. The visual pathway uses a DINOv2-with-Registers backbone to extract 37×37 patch features with a feature dimension of 768. These features are further processed by a direction-aware decoder composed of four dilated convolution branches and angle-sensitive pooling, producing a 1024-dimensional visual embedding. The tactile pathway encodes the normalized 12-dimensional force-torque signal using a two-layer MLP with a hidden dimension of 128 and an output dimension of 256, yielding a 256-dimensional tactile embedding. The visual and tactile embeddings are concatenated into a shared 1280-dimensional representation. The action head contains a 512-dimensional hidden layer and predicts a 7-DoF end-effector action, while the progress head contains a 256-dimensional hidden layer followed by a sigmoid activation function to estimate task progress in the range [0,1].

B.5 TACO Iterative Correction Framework Details

Recognize. At each iteration we deploy the current policy to collect real rollouts and run the unified progress-action model over every rollout to obtain a dense progress score p_t. From each failed trajectory we select correction anchors at the timesteps where progress stalls or decreases. In practice we annotate up to 10 failure-adjacent anchors per failed trajectory, which keeps the anchors concentrated on the contact-sensitive stages rather than spread uniformly over the trajectory. We assign the binary advantage label in this stage. For each rollout, we take the first recognized anchor as the failure onset: timesteps before this anchor are still on track and receive advantage $y = 1$, whereas timesteps from the first anchor onward correspond to the failed segment and receive $y = 0$. For teleoperated expert demonstrations, which complete the task correctly, we set $y = 1$ over the entire trajectory.

Imagine and Label. From each anchor state, the visuo-tactile generation model imagines a local correction of $T = 49$ timesteps, denoising future video and 12-D force sequences conditioned on the current observation, force history, and language instruction. The unified progress-action model then labels each imagined segment with corrective actions. Since imagined segments are generated to recover from failure-adjacent states toward task completion, we assign them advantage $y = 1$, so that advantage-conditioned post-training treats them as positive recovery supervision alongside the expert demonstrations.

Data Ratio. Because each failed trajectory yields up to 10 failure-adjacent anchors and each anchor produces one imagined correction segment, the real-to-imagined data ratio across the six tasks typically falls in the range 1:4 to 1:5.

B.6 Training Recipe Details

Policy architecture. We build on $\pi_{0.5}$ [7], a flow-matching VLA with a PaliGemma (2B) vision-language backbone and a 300M action expert. Images, language, and the discretized state are encoded as VLM prefix tokens, and the action expert predicts a 7-DoF end-effector action chunk of horizon 30 (zero-padded to an action dimension of 32) by denoising under flow matching. All policies are initialized from the publicly released $\pi_{0.5}$-DROID checkpoint.

Tactile and advantage conditioning. TACO and its tactile variants extend the action expert with force and advantage conditioning injected through the adaRMSNorm path, alongside the flow-matching timestep. The force input is the concatenated left/right 6-DoF force-torque reading (12

Method	Real Succ.	Imag. Corr.	KI	Purpose
Base Policy	✗	✗	✗	Warm-start on 50 demos only
Filtered BC	✓	✗	✗	Test if real successes alone suffice
TACO (w/o KI)	✓	✓	✗	Test if dropping KI erodes VLM priors
TACO	✓	✓	✓	Full method

Table 3: **Comparison of baselines.** *Real Succ.* denotes fine-tuning on successful real-world rollouts; *Imag. Corr.* denotes imagined corrections; *KI* denotes knowledge-insulated tactile adaptation.

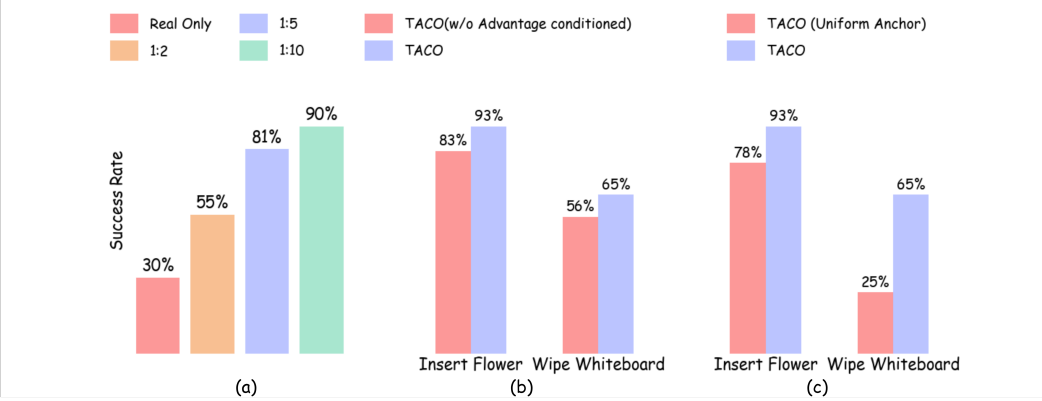


Figure 10: **Additional ablation studies. (a)** Scaling of imagined correction data on *Toast Bread*: success rate improves as the real-to-imagined ratio increases. **(b)** Effect of advantage-conditioned training. **(c)** Effect of failure-adjacent anchor selection. Results in (b) and (c) are reported on *Insert Flower* and *Wipe Whiteboard*.

dimensions) over a history window of length 8. A force encoder (LayerNorm $\rightarrow$ Linear $\rightarrow$ SiLU $\rightarrow$ Linear $\rightarrow$ LayerNorm, hidden dimension 256) maps this history to a conditioning vector matching the action-expert width, and the segment advantage is encoded by an analogous advantage encoder (hidden dimension 256). The final adaRMS condition is the sum of the timestep, force, and advantage embeddings. We further adopt classifier-free guidance: during training the condition is randomly replaced by a learned null embedding with probability 0.1, and at inference we condition on the positive advantage to elicit high-progress recovery behaviors.

Knowledge-insulated tactile adaptation. TACO applies knowledge insulation by computing the VLM prefix once and applying a stop-gradient to its representation before the action expert attends to it, so that the tactile-action loss cannot propagate into the PaliGemma vision-language backbone. Only the action expert and the force and advantage encoders are updated; the vision encoder and the language backbone remain frozen.

Optimization. All models are trained for 30,000 steps with a batch size of 32 using AdamW (gradient norm clipped to 1.0) and an exponential moving average of 0.999. The learning rate follows a cosine schedule with 300 warmup steps and a peak of 5×10^{-5}. Training uses fully-sharded data parallelism across 8 GPUs.

C Additional Ablation Studies

C.1 Evaluation Metric Details

We detail the evaluation protocol and metrics in Section 4.3. We assess each setting along four complementary axes that together cover both the quality of the imagined corrections and their usefulness as corrective supervision.

Action validation loss (A). We report the validation loss of the unified progress-action model on held-out corrective segments, measuring how well the model predicts corrective actions from the imagined visual and tactile signals. Lower values indicate more accurate action labeling.

Force prediction loss (F). To evaluate the tactile quality produced by the visuo-tactile generation model and consumed by the progress-action model, we construct a held-out test set of real contact-rich segments with ground-truth 12-D force-torque readings. We report the prediction error against these ground-truth forces, so that F reflects how faithfully the generated and labeled tactile signals match real contact dynamics.

Video-frame progress rank correlation (VOC). We measure whether the progress estimates produced by the unified progress-action model are correctly ordered along the true temporal progression of a trajectory. For each of the six tasks, we hold out a set of trajectories, shuffle their frames, and compute the rank correlation between the predicted progress and the ground-truth chronological order. Given two frames $t_A < t_B$, a correct prediction requires the estimated progress at t_A to be lower than at t_B. The VOC score lies in $[-1, 1]$, with higher values indicating a more faithful understanding of task progression.

Failure-localization accuracy (FL). To verify the Recognize step directly, we manually annotate failure-adjacent frames in held-out failed rollouts where contact begins to break down. We then check whether the progress-action model, by detecting where progress stalls or decreases, flags these annotated frames as failure-adjacent. FL reports the accuracy of this localization across the six tasks, so a higher FL means the Recognize step more reliably identifies the contact transitions that need correction.

C.2 Scaling of Imagined Correction Data

To verify the scaling trend of imagined correction data, we repeat the scaling study on *Toast Bread* under the same training recipe. As shown in Figure 10 (a), incorporating imagined correction data consistently improves the success rate over the Base Policy. When increasing the real-to-imagined data ratio from 1:2 to 1:5 and 1:10, the success rate increases from 55% to 81% and 90%, respectively. Consistent with the main-text results, the 1:10 setting further outperforms 1:5, indicating that larger-scale imagined corrections provide broader coverage of failure-adjacent contact states. These results confirm that imagined data effectively improves policy post-training while substantially reducing the cost of correction data collection.

C.3 Effect of Advantage-Conditioned Training

We further isolate the contribution of advantage-conditioned training, which is the component that allows TACO to learn from failed rollouts rather than only from successful ones. We compare TACO against a variant, TACO (w/o Advantage-Conditioned), that removes the failed rollouts from each iteration's training set and retains only expert demonstrations, successful rollouts, and imagined corrections. This variant then fine-tunes the policy on all retained data with standard supervised learning, without the binary advantage labels that separate corrective segments from failures. As a result, it treats imagined corrections and ordinary successful trajectories identically, and discards the failure rollouts that mark which contact states require recovery. We evaluate both methods on *Insert Flower* and *Wipe Whiteboard*. As shown in Figure 10 (b), removing advantage-conditioned training lowers the success rate from 93% to 83% on *Insert Flower* and from 65% to 56% on *Wipe Whiteboard*. Although the corrective supervision is identical in both settings, discarding the failure rollouts removes the negative signal that tells the policy which contact transitions to avoid, and supervising all retained data uniformly causes the policy to imitate successful and corrective trajectories without distinguishing recovery behaviors from the failures they are meant to correct. In contrast, advantage-conditioned training keeps the failure rollouts and assigns them a negative advantage, so the policy is explicitly steered toward high-progress corrections at inference time through the positive advantage condition. These results show that the gains of TACO come not only from

imagined corrections but also from treating correction as an offline reinforcement learning objective that learns from failures rather than filtering them away.

C.4 Effect of Failure-Adjacent Anchor Selection

We further examine the Recognize step, which selects where to imagine corrections. TACO uses the unified progress-action model to predict a dense progress score along each failed rollout, and places correction anchors at failure-adjacent states where progress stalls or decreases. To isolate the value of this progress-guided selection, we compare TACO against TACO (Uniform Anchor), which draws the same number of anchors uniformly at random from the failed rollouts instead of locating them at progress-stalling states. We evaluate both on *Insert Flower* and *Wipe Whiteboard*.

As shown in Figure 10 (c), uniform anchor selection lowers the success rate from 93% to 78% on *Insert Flower* and from 65% to 25% on *Wipe Whiteboard*. Uniform sampling spends much of the imagination budget on states far from the contact transition, where the rollout has either not yet reached the failure or has already drifted into an unrecoverable configuration, so the resulting corrections carry little recovery-relevant signal. By contrast, anchoring at progress-stalling states concentrates imagination on the contact transitions where recovery is both necessary and still feasible, yielding corrective supervision that is better matched to the failures it is meant to fix. These results show that recognizing failure-adjacent states is an essential part of TACO, and that the quality of the imagined corrections depends on where they are anchored.

D Additional Analysis

D.1 Failure Case Analysis

Figure 11 shows representative rollouts of Filtered BC, TACO (w/o KI), and TACO on three contact-rich tasks. The two baselines fail in distinct but complementary ways, and comparing them illustrates why both imagined visuo-tactile corrections and knowledge-insulated tactile adaptation are necessary.

Filtered BC: stalls at contact transitions. Trained only on its own successful rollouts, Filtered BC inherits the narrow action distribution of the base policy and provides no recovery behavior at failure-adjacent contact states. As a result, it approaches the target but stalls once contact becomes critical: in *Wipe Whiteboard* it presses the eraser onto the board without applying enough force to remove the mark, in *Move Hanoi Rings* it hovers around the peg but fails to align and seat the ring, and in *Twist Bottle Cap* it grips the cap without generating effective twisting torque. Because these contact-transition failures rarely self-correct within a rollout, they never enter the filtered successful set, so Filtered BC keeps reproducing the same incomplete behavior.

TACO (w/o KI): degraded pre-contact perception. TACO (w/o KI) learns from the same imagined corrections as TACO but fine-tunes the full VLA end-to-end, allowing tactile-action gradients to erode the pretrained visual-language priors. Without these priors, pre-contact perception and spatial grounding degrade: the policy reaches the contact stage but with less accurate approach and alignment, leading to imprecise contact (e.g., off-target wiping, misaligned ring insertion, or unstable cap engagement). It thus recovers some contact behaviors from the tactile corrections but loses the visual grounding needed to position itself reliably before contact.

TACO: reliable approach and contact recovery. TACO combines imagined visuo-tactile corrections with knowledge-insulated tactile adaptation, preserving the pretrained visual-language priors for accurate pre-contact approach while learning failure-adjacent recovery from the tactile corrections. It therefore both reaches the target reliably and adjusts contact force effectively, completing all three tasks: it wipes the mark clean, seats the ring onto the peg, and twists the cap open. These cases show that the two components address complementary failure modes, and only their combination yields robust contact-rich execution.

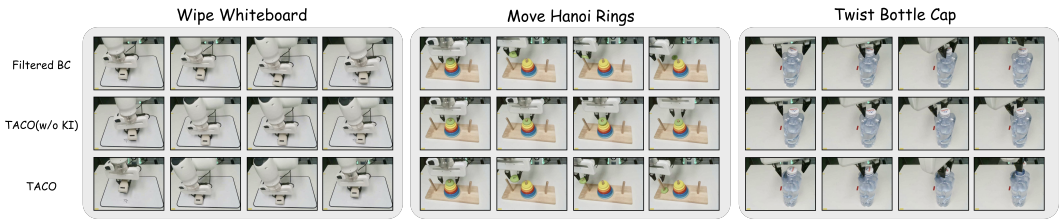


Figure 11: **Failure case analysis.** Representative rollouts of Filtered BC, TACO (w/o KI), and TACO on *Wipe Whiteboard*, *Move Hanoi Rings*, and *Twist Bottle Cap*. The two baselines fail at contact transitions in complementary ways, whereas TACO completes all three tasks.

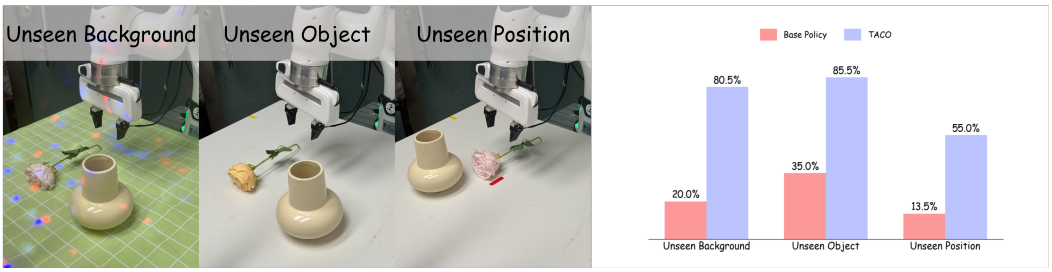


Figure 12: **Additional generalization experiment.** Generalization experiments under unseen backgrounds, unseen objects, and unseen object positions for *Insert Flower*.

D.2 Additional Generalization Experiment

To further verify the generalization gains of TACO, we conduct evaluation on *Insert Flower* under the three paradigms: *Unseen Backgrounds*, *Unseen Objects*, and *Unseen Positions*. We compare the base policy, trained only on in-domain expert demonstrations, with the policy updated by one round of TACO using out-of-domain (OOD) imagined correction data. As shown in Figure 12, the base policy degrades from 50.0% to 20.0%, 35.0%, and 13.5% under the three shifts, respectively, whereas TACO recovers to 80.5%, 85.5%, and 55.0% after only one adaptation iteration. Consistent with the main-text results, by generating imagined correction data around OOD states, TACO exposes the policy to diverse successful recovery behaviors without requiring additional expert demonstrations in the target scenarios. These results further support that TACO offers an efficient mechanism for adapting policies to novel positions, objects, and visually perturbed environments.

E Additional Visualization

E.1 Additional Imagined Corrections Visualizations

Beyond the two representative tasks shown in the main text, we further visualize imagined corrections for the remaining four contact-rich tasks in Figure 13, again pairing real-world failure rollouts (top) with the imagined corrections produced by the tactile-aware world model (bottom). Across all tasks, the world model localizes the failure-adjacent contact transition and imagines a segment that restores the intended contact behavior toward task completion. In *Insert Flower*, the real rollout brings the flower toward the vase but misaligns the stem with the narrow opening and fails to seat it, whereas the imagined correction realigns the approach and guides the stem into the vase. In *Wipe Whiteboard*, the real rollout places the eraser over the target mark but applies insufficient contact force to remove it, whereas the imagined correction increases downward pressure and wipes the mark clean. In *Play Xylophone*, the real rollout swings toward the target bar but under-contacts the strike, whereas the imagined correction adjusts the contact so the mallet lands firmly on the intended bar. In *Toast Bread*, the real rollout grasps the bread but misaligns its placement into the slot, whereas the imagined correction corrects the alignment and seats the slice in place. Consistent

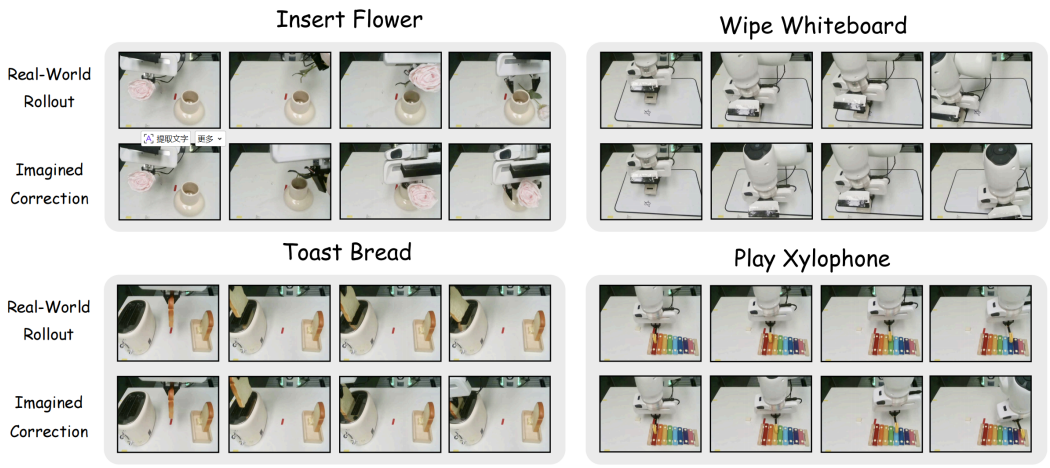


Figure 13: **Visualization of imagined correction data.** The tactile-aware world model generates locally consistent imagined corrections (bottom) that recover failed contact interactions (top).

with the main-text examples, these corrections recover contact transitions that are weakly observable from RGB alone, yielding the corrective supervision used for post-training.