Curriculum-Aware Diffusion Distillation: Representing Contact Phases for Long-Horizon Door Manipulation

Anonymous Submission

***Abstract*— Learning continuous control for long-horizon, contact-rich tasks—such as manipulating a heavy articulated door—remains a key challenge in robot learning. While Reinforcement Learning (RL) with privileged state information can solve these tasks in simulation, distilling such policies into vision-based Student networks via Behavioral Cloning often results in catastrophic failure due to state ambiguity and compounding errors. In this paper, we propose Curriculum-Aware Diffusion Distillation (CADD), a framework that bridges this gap through representation design and phase-conditioned generation. CADD explicitly masks privileged states to ensure sim-to-real readiness, maps absolute poses to a continuous Residual 9D action space, and converts hard-coded RL curriculum stages into continuous soft labels. By training a lightweight classification head to predict the current contact phase, CADD modulates a Diffusion Policy via Feature-wise Linear Modulation. Evaluated on the end-to-end door opening task with a challenging bottleneck initialization, CADD achieves a 42.0% success rate compared to 0% for Vanilla Diffusion Policies. Qualitative and quantitative evaluations demonstrate that CADD effectively mitigates both the state-ambiguity-induced denoising failure of generative baselines and the spatial compounding errors of deterministic policies at contact boundaries.**

I. INTRODUCTION

Contact-rich robotic manipulation is a critical step toward general embodied intelligence. Tasks such as opening a heavy hinged door are inherently multi-stage, involving free-space motion (approach), rigid contact establishment (grasping), constrained motion (pulling), and whole-body coordination (circular repositioning).

A prevalent pipeline for such tasks employs a Teacher-Student architecture: a Teacher policy is trained using Reinforcement Learning (RL) with privileged exact state information [1], [2]; subsequently, a vision-based Student policy learns to mimic the Teacher. However, distilling these multi-stage tasks is notoriously difficult due to the partial observability induced by privileged state removal, leaving the Student without crucial contact context. While Diffusion Policies (DPs) [3], [4] have recently emerged to successfully mitigate standard Behavioral Cloning (BC) mode collapse, they still struggle with long-horizon manipulation. When observation histories near transition boundaries are insufficient to disambiguate dynamic contact states, the unconditioned diffusion denoising process may generate incoherent action chunks that violate kinematic constraints.

To address these challenges, we propose the Curriculum-Aware Diffusion Distillation (CADD) framework. CADD moves beyond naive end-to-end mapping in favor of an architecture that elevates the Teacher's discrete "curriculum phase" into a continuous, dynamically predicted contact representation. By training a lightweight head to predict this ongoing physical phase and injecting it as an explicit conditioning signal into the diffusion process, CADD effectively disambiguates conflicting state histories and stabilizes long-horizon generation.

The main contributions of this paper are:

- **Phase-Conditioned Generative Architecture:** We propose CADD, a framework that converts the Teacher's discrete curriculum stages into continuous soft phase labels and injects them into a Diffusion U-Net via FiLM modulation. We show that this explicit phase conditioning significantly improves action coherence at contact-phase boundaries.
- **Representation Design for Sim-to-Real Readiness:** We adopt strict masking of privileged states and a continuous Residual 9D action space as design principles to prevent information leakage and ensure continuous spatial generation, aiming to facilitate future sim-to-real transfer.

II. RELATED WORK

A. Contact-Rich Manipulation and Representation

Analytical methods utilizing complementarity formulations [5] or explicit contact mode scheduling [6] provide formal physical guarantees but struggle to scale. While data-driven approaches [7] improve prediction accuracy, they focus on local contact modeling rather than long-horizon policy generation. Conversely, end-to-end RL [1], [2] scales efficiently in simulation but frequently suffers from the sim-to-real gap. This gap often stems from privileged information leakage during distillation and the discontinuous nature of rigid contacts. To bridge this divide, CADD conceptualizes Teacher curriculum stages as a soft, differentiable analogue to classical hybrid contact modes, employs strict privileged state masking, and maps visual-proprioceptive features to a continuous Residual 9D action space using 6D rotations [8].

B. Conditioned Generative Policies in Robotics

Diffusion Models [4] and Action Chunking with Transformers (ACT) [9] have substantially improved visuomotor policies by effectively expressing multi-modal distributions [3]. While recent advancements explore goal-conditioned diffusion, these conditioning signals typically encode static spatial intents. Standard Diffusion Policies remain implicitly conditioned regarding the ongoing physical state, frequently experiencing multi-modal denoising collapse at abrupt contact boundaries. Conceptually akin to temporal abstraction in the options framework [10], [11],

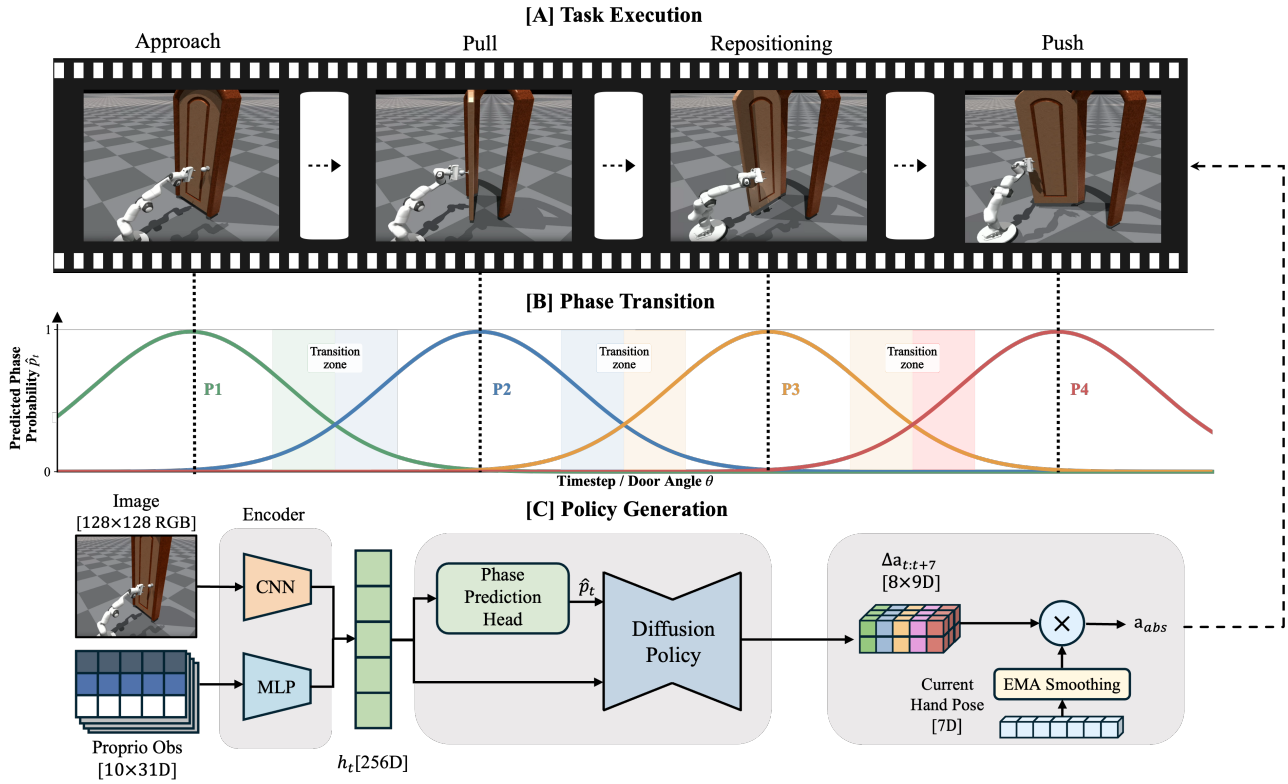

Fig. 1. **CADD Architecture Overview. (A)** The door-opening task is decomposed into four physical phases (Approach, Pull, Reposition, Push). **(B)** Discrete curriculum labels are converted into soft phase probabilities $\mathbf{p}_t$, which vary continuously across phase boundaries. **(C)** A dual-stream encoder processes visual and proprioceptive inputs into a shared latent $\mathbf{h}_t$. A phase prediction head estimates $\hat{\mathbf{p}}_t$, which modulates the diffusion U-Net via FiLM (Eq. 3–4) to generate Residual 9D action chunks. Privileged states (door handle position, hinge angle) are masked at inference to facilitate sim-to-real transfer.

CADD utilizes Feature-wise Linear Modulation (FiLM) [12] to modulate the diffusion U-Net directly on the predicted dynamic contact phase, addressing action ambiguity during kinematic bottlenecks.

III. METHODOLOGY

CADD consists of a state-action representation for sim-to-real transfer, a soft label generation module, and a phase-conditioned diffusion decoder.

A. State-Action Representation for Sim-to-Real

To prevent the Student policy from relying on information unavailable in the real world, we strictly mask privileged environmental states—specifically, the door handle position, hinge angle, and door body orientation—from the 40D proprioceptive observation vector, retaining only robot joint states, end-effector kinematics, and historical actions. The visual input is provided by a fixed third-person camera, encoded via a 4-layer Convolutional Neural Network (CNN) into a 128D feature vector.

To ensure continuous spatial generation, CADD outputs a **Residual 9D action** $\Delta a_t = [\Delta p, \Delta R_{6D}]$. During closed-loop inference, the 6D components are strictly orthogonalized into $SO(3)$ rotation matrices via Gram-Schmidt. The training loss operates directly on the raw 6D representation to preserve gradient smoothness. The executed trajectories are then smoothed using temporal ensembling across overlapping action chunks to prevent hardware jitter.

B. Continuous Contact Phase Representation

Real-world physical contact transitions occur continuously. To train the Student policy, we convert discrete, hard-coded Teacher curriculum labels $y_t \in \{1, 2, 3, 4\}$ into a smooth soft distribution $\mathbf{p}_t \in \mathbb{R}^4$. To handle challenging initializations without state ambiguity, the smoothing logic is hierarchically anchored to the Teacher's current phase, utilizing distinct metrics for different boundaries.

For instance, the Approach-to-Pull transition is governed by a temporal sigmoid over a stable grasp duration counter. Conversely, kinematic transitions are governed by the door opening angle θ. The Pull-to-Reposition transition is formulated as:

$$p_{t,3} = \frac{1}{1 + \exp(-(\theta - \theta_{th})/\tau)}, \tag{1}$$

$$p_{t,2} = 1 - p_{t,3}. \tag{2}$$

This design establishes a piecewise sliding window across the curriculum. During any specific transition boundary, non-adjacent phases are explicitly masked ($p_{t,1} = p_{t,4} = 0$) to ensure $\mathbf{p}_t$ remains a valid probability simplex. Outside these narrow transition windows, $\mathbf{p}_t$ strictly collapses to a

one-hot vector. Subsequent transitions (e.g., Reposition-to-Push) employ similarly structured kinematic functions but with distinct task-specific parameters (e.g., $\theta_{th} = 95°$ instead of $77°$).

C. Phase-Conditioned Diffusion Architecture

Given a two-step observation horizon O_t (with each step comprising a single visual frame and 5-frame stacked proprioceptive vectors), the encoded features are concatenated to form a shared latent representation $\mathbf{h}_t$. From this representation, a lightweight classification head predicts the current contact phase distribution $\hat{\mathbf{p}}_t$.

Action generation is performed by a conditional 1D U-Net ϵ_θ, which predicts an action chunk of $T_a = 8$ steps. Within each residual block, the observation representation $\mathbf{h}_t$ and the diffusion timestep embedding e_k are injected additively. Subsequently, the predicted phase probability strictly modulates these observation-and-timestep-conditioned features via FiLM [12]:

$$\gamma = W_\gamma \hat{\mathbf{p}}_t, \quad \beta = W_\beta \hat{\mathbf{p}}_t, \tag{3}$$

$$\mathbf{f}_{modulated} = \gamma \odot (\mathbf{f} + W_c[\mathbf{h}_t \oplus e_k]) + \beta. \tag{4}$$

This architecture allows the phase condition to dynamically scale the foundational physical states and denoising subspaces simultaneously.

During training, the framework is optimized jointly. The classification head is supervised via KL divergence. The diffusion U-Net is optimized via standard Mean Squared Error (MSE), utilizing teacher forcing on the ground-truth phase probabilities to ensure stable gradient flow. Gradients from the diffusion loss are explicitly stopped from propagating into the classification head to prevent coupled optimization instability.

During inference, the U-Net is conditioned on the predicted $\hat{\mathbf{p}}_t$. While training via teacher forcing theoretically introduces exposure bias, we explicitly design the phase predictor as a lightweight classification head. This architecture allows the network to robustly anchor its predictions to continuous visual-proprioceptive features, aiming to mitigate the train-test distribution shift during closed-loop execution.

IV. EXPERIMENTS

A. Experimental Setup

Experiments are conducted in IsaacGym [13] using a Franka arm opening a heavy hinged door. Our simulation environment is built upon the UniDoorManip framework [14]. The expert Teacher policy is pre-trained via Proximal Policy Optimization (PPO) [15] using privileged exact environment states. The dataset comprises over 182K frames (464 episodes) of expert trajectories collected at 20 Hz within this environment.

To isolate and evaluate policy behavior at the most critical phase transition, we adopt a controlled stress test strategy. The door is initialized at a fixed opening angle of $65°$, representing a typical large-angle kinematic bottleneck where the Pull-to-Reposition transition demands the most drastic change in arm posture. Importantly, the policies are still required to execute the task end-to-end from this bottleneck: the robot must approach the handle, establish a stable grasp, resolve the phase boundary to release, and reposition to push the door open. This setup ensures the policies are evaluated exactly at the region of highest state ambiguity. Evaluating robustness across diverse initialization angles remains an important direction for future work.

B. Quantitative Baseline Comparisons

We compare CADD against a Naive BC baseline parameterized by a Multi-Layer Perceptron (MLP) and a Vanilla DP. All methods share the same visual encoder and Residual 9D action space. To evaluate task completion, a trajectory is considered successful *only if* the door is opened beyond a $110°$ clearance threshold *and* the robot correctly transitions to the inside of the door (programmatically verified by the end-effector crossing the door's interior plane).

TABLE I
TASK SUCCESS AND PHASE SURVIVAL RATES (50 EPISODES)

Method	Avg Angle	P_3	P_4	Success
Naive BC	$52.4° \pm 32.8°$	**82%**	0%	0%
Vanilla DP	$60.1° \pm 22.1°$	0%	0%	0%
CADD (Ours)	**89.5°± 22.2°**	56%	**50%**	**42.0%**

As shown in Table I, the transition-critical initialization clearly exposes the vulnerabilities of standard paradigms (0% success).

Vanilla DP suffers from severe **state-ambiguity-induced denoising failure**. At the phase boundary, without explicit phase conditioning, the unconditioned denoising process generates physically incoherent action chunks. Consequently, the policy fails to even trigger the Phase 3 transition (0% P_3).

Conversely, Naive BC demonstrates **deterministic compounding errors**. BC achieves a relatively high 82% survival rate for initiating P_3, reflecting the simplicity of the local "release" decision. However, lacking the generative capacity to model spatially diverse reposition trajectories, it suffers complete attrition (82% $\rightarrow$ 0%) from P_3 to P_4.

In contrast, CADD explicitly mitigates state ambiguity, achieving a stable 56% P_3 transition. CADD exhibits high **phase-conversion efficiency**, captured by the $P_3 \rightarrow P_4$ retention rate: CADD retains 89% (25/28) of the trajectories that enter P_3, while Naive BC retains 0% (0/41). This demonstrates that CADD effectively sustains coherent behavior through the kinematic bottleneck. Future work will include ablations isolating the individual contributions of soft phase labels and FiLM to further contextualize these gains.

C. Qualitative Behavior Analysis

To further understand the performance gap, we visually inspect the rollout behaviors in Fig. 2.

Vanilla DP (Fig. 2a) exhibits immediate mechanical failure. The policy outputs erratic end-effector poses, causing the gripper to exhibit severe angular deviations that conflict

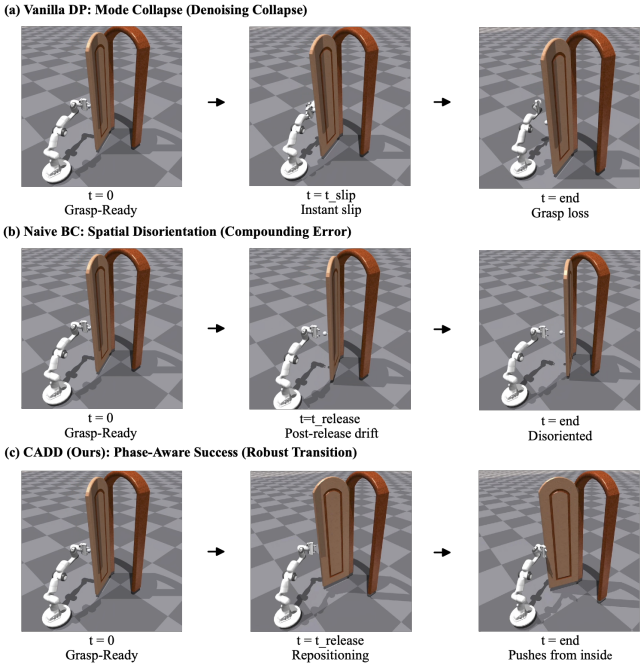

Fig. 2. **Qualitative Analysis of Pull-to-Reposition transition boundary.** (a) Vanilla DP suffers from denoising collapse, resulting in instant grasp loss. (b) Naive BC successfully releases the handle but drifts aimlessly, failing to navigate to the inner side. (c) CADD explicitly identifies the phase boundary, smoothly releasing the handle and navigating to the inside.

with the rigid hinge constraints, breaking physical contact instantly.

Naive BC (Fig. 2b) correctly opens the gripper but completely loses spatial awareness during the free-space retreat. The deterministic policy averages diverse training trajectories into an unviable path; the end-effector drifts aimlessly or collides with the door's outer edge, leaving the robot stranded.

CADD (Fig. 2c) demonstrates improved physical engagement. Anchored by the predicted phase, the policy executes a smooth "Pull-to-Reposition" sequence, retreating along a collision-free spatial arc to execute the push. While CADD significantly improves transition success, remaining failure cases (58%) primarily manifest as residual spatial drift during the final P_4 push execution, indicating that robust spatial reasoning through this phase remains an open challenge.

V. CONCLUSIONS

We introduced CADD, a phase-conditioned distillation framework that successfully integrates the generative capacity of Diffusion Policies with the explicit structural awareness of curriculum phases. By combining state-action representations with soft phase modulation, CADD achieves improved long-horizon manipulation capabilities. Our evaluations demonstrate that explicit phase conditioning significantly mitigates both the state-ambiguity-induced denoising failure of standard generative models and the spatial compounding errors of deterministic baselines, offering a structured approach to phase-aware distillation for contact-rich tasks.

REFERENCES

[1] D. Chen, B. Zhou, V. Koltun, and P. Krähenbühl, "Learning by cheating," in *Proc. Conf. Robot Learn. (CoRL)*, 2020, pp. 66–75.
[2] L. Pinto, M. Andrychowicz, P. Welinder, W. Zaremba, and P. Abbeel, "Asymmetric actor critic for image-based robot learning," in *Proc. Robotics: Sci. Syst. (RSS)*, 2018.
[3] C. Chi, et al., "Diffusion policy: Visuomotor policy learning via action diffusion," in *Proc. Robotics: Sci. Syst. (RSS)*, 2023.
[4] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2020.
[5] M. Posa, C. Cantu, and R. Tedrake, "A direct method for trajectory optimization of rigid bodies through contact," *Int. J. Robot. Res. (IJRR)*, vol. 33, no. 1, pp. 69–81, 2014.
[6] F. R. Hogan and A. Rodriguez, "Reactive planar manipulation with convex hybrid MPC," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2020, pp. 247–253.
[7] N. Fazeli, et al., "Learning data-efficient rigid-body contact models: Case study of planar impact," in *Proc. Conf. Robot Learn. (CoRL)*, 2017.
[8] Y. Zhou, C. Barnes, J. Lu, J. Yang, and H. Li, "On the continuity of rotation representations in neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 5745–5753.
[9] T. Z. Zhao, et al., "Learning fine-grained bimanual manipulation with low-cost hardware," in *Proc. Robotics: Sci. Syst. (RSS)*, 2023.
[10] R. S. Sutton, D. Precup, and S. Singh, "Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning," *Artif. Intell.*, vol. 112, no. 1-2, pp. 181–211, 1999.
[11] P.-L. Bacon, J. Harb, and D. Precup, "The option-critic architecture," in *Proc. AAAI Conf. Artif. Intell.*, 2017.
[12] E. Perez, et al., "FiLM: Visual reasoning with a general conditioning layer," in *Proc. AAAI Conf. Artif. Intell.*, 2018.
[13] V. Makoviychuk, et al., "Isaac Gym: High performance GPU-based physics simulation for robot learning," in *Proc. NeurIPS Datasets Benchmarks Track*, 2021.
[14] Y. Li, et al., "UniDoorManip: Learning universal door manipulation policy over large-scale and diverse door manipulation environments," *arXiv preprint arXiv:2403.02604*, 2024.
[15] J. Schulman, et al., "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.