CoRDE: Concept-Prior Routed Diffusion Experts for Structural Generalization in Robot Manipulation

Anonymous Author(s)
Affiliation
Address
email

Abstract

Diffusion models excel at capturing multi-modal action distributions in robot imitation learning. However, in multi-task and long-horizon scenarios, monolithic architectures lack structural generalization capabilities, suffering from gradient conflicts between distinct semantic sub-stages. While pure data-driven Mixture-of-Experts (MoE) methods introduce labor division, they frequently trigger routing collapse, and instantiating full-scale experts causes parameter explosion and high expansion costs. To address these issues, we propose Concept-prior Routed Diffusion Experts (CoRDE), a structure-guided variational distillation framework. CoRDE extracts semantic distributions from a frozen concept encoder to guide the variational posterior responsibility via a learnable soft mapping matrix. This mechanism enforces a dual-entropy dynamics conservation: it minimizes routing entropy to guarantee macroscopic cognitive certainty, while preserving the full-rank variance of the stochastic diffusion term to maintain behavioral diversity. To overcome parameter inflation, CoRDE employs a parameter-efficient expert pool using Low-Rank Adaptation (LoRA) on a shared frozen backbone. Rigorous mathematical proofs demonstrate that the mixture score field of low-rank experts strictly approximates the teacher distribution in the least-squares sense, avoiding rank-deficiency-induced diversity loss and ensuring high-fidelity generation. Empirical evaluations confirm that, compared to existing baselines, CoRDE systematically reduces routing collapse, forming robust, semantically aligned expert allocations while achieving superior action quality and incremental learning efficiency.

1 INTRODUCTION

It is well established that diffusion models, which formulate action generation as a conditional denoising process, have become a dominant paradigm in robot imitation learning [2, 3, 14, 16]. By learning the score function within the action space, diffusion policies excel at capturing highly complex, multi-modal, and continuous data distributions inherent in human demonstrations. This expressive capability allows robots to replicate diverse human behaviors with unprecedented trajectory smoothness and high task success rates. Despite these significant advances, deploying monolithic diffusion policies in multi-task and long-horizon scenarios reveals severe structural limitations. Forcing a single set of global parameters to simultaneously represent behavioral modes with drastically heterogeneous frequency properties inevitably induces gradient conflicts [4, 7, 11, 8] and catastrophic interference. Furthermore, their iterative denoising nature incurs substantial inference latency, while intractable likelihoods complicate post-hoc policy optimization.

To mitigate these bottlenecks, recent research transforms pre-trained diffusion policies into Mixture of Experts (MoE) architectures [15, 18, 10], notably via Variational Distillation of Diffusion (VDD) [22]. Simultaneously, other expert-driven manipulation frameworks [5, 20] have been utilized to decompose manipulation skills. However, existing unsupervised MoE frameworks [22, 20, 17] are constrained by fundamental contradictions: when processing noisy multi-modal data, pure data-likelihood-driven

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

routing frequently triggers routing collapse [22] and distorts the naturally uneven distribution of physical tasks [17]. Additionally, the resulting expert populations lack human-interpretable semantic boundaries, and instantiating full-scale experts causes parameter explosion.

In this article, we present Concept-prior Routed Diffusion Experts (CoRDE), a novel structure-guided variational distillation system that seamlessly unifies high-level semantic concept priors [23, 13] with the generative expressivity of diffusion MoE models. Our approach extracts a stable semantic distribution from a frozen concept encoder to guide the variational posterior responsibility via a learnable soft mapping matrix. This enforces a dual-entropy dynamics conservation: minimizing routing entropy for macroscopic certainty, while preserving the full-rank variance of the stochastic diffusion term [19] to maintain behavioral diversity. To overcome parameter inflation, we construct an expert pool by injecting Low-Rank Adaptation (LoRA) [6] modules onto a shared frozen backbone. We provide rigorous mathematical proofs demonstrating that the mixture score field of low-rank experts strictly approximates the teacher distribution in the L_2 sense [22], theoretically guaranteeing high-fidelity generation while avoiding rank-deficiency-induced diversity loss [1, 21].

The purpose of this article is to solve the structural generalization challenge in multi-modal robot manipulation. Our main contributions are as follows:

- We propose a structure-guided variational diffusion distillation framework that integrates a concept-prior-driven responsibility mechanism, systematically eliminating routing collapse and forming robust, semantically aligned expert allocations.
- We design a parameter-efficient diffusion expert pool utilizing LoRA. Rigorous mathematical proofs guarantee that this architecture drastically reduces computational overhead while maintaining action diversity via the stochastic diffusion term and ensuring high-fidelity generation through strict score field approximation.
- We develop an Expectation-Maximization (EM) style self-organizing soft alignment algorithm, establishing an interpretable many-to-many mapping between macroscopic task concepts and microscopic execution experts.

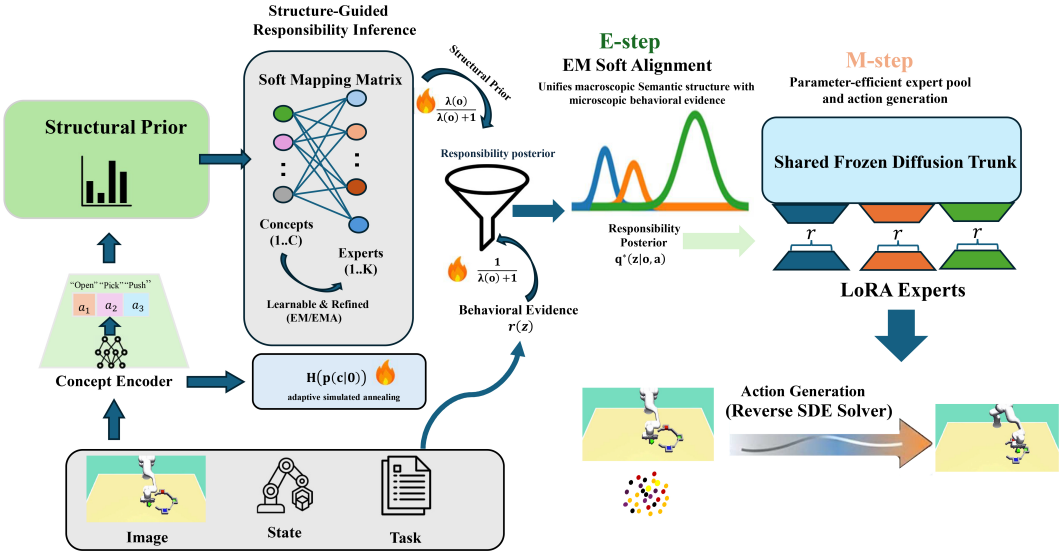

Figure 1: Overview of the CoRDE framework. **Training:** A structural prior—derived from extracted concept semantics—fuses with behavioral evidence to infer a variational responsibility posterior. This guides task allocation across parameter-efficient LoRA experts on a shared backbone. **Inference:** State-conditioned routing dynamically activates these experts. The resulting mixture score field synthesizes high-fidelity actions while strictly preserving full-rank behavioral diversity.

2 Methodology

This section formulates the Concept-prior Routed Diffusion Experts (CoRDE) framework. We first establish the variational distillation of monolithic diffusion policies into a Mixture-of-Experts (MoE) architecture. Subsequently, we construct the structure-guided responsibility inference mechanism,

present the parameter-efficient expert parameterization with rigorous error bounds, and detail the self-organizing alignment updates.

2.1 Preliminaries: Distilling a Diffusion Policy into MoE

Diffusion policies model action generation as a conditional denoising process. Given a demonstration dataset $\mathcal{D} = \{(o_i, a_{0,i})\}$, the forward process injects Gaussian noise $\epsilon \sim \mathcal{N}(0, I)$ into the clean action sequence a_0 at continuous timestep $t \in (0, T]$, yielding the noisy action $a_t = \alpha_t a_0 + \sigma_t \epsilon$. The monolithic diffusion policy is trained to predict the injected noise via the score-matching proxy objective [2]:

$$\mathcal{L}_{\text{diff}} = \mathbb{E}_{o,a_0,t,\epsilon}\left[\|\epsilon - \epsilon_\theta(a_t, o, t)\|_2^2\right], \tag{1}$$

where ϵ_T denotes the pre-trained teacher noise predictor. This objective is mathematically equivalent to matching the true score field of the data distribution, i.e., $u_T(a_t, o, t) = -\epsilon_T(a_t, o, t)/\sigma_t \approx \nabla_{a_t} \log p_t(a_t|o)$. For strict theoretical coherence, subsequent derivations utilize the score field notation u.

To overcome the intractable likelihood and elevated inference latency of diffusion models, the Variational Distillation of Diffusion (VDD) method [22] compresses the pre-trained continuous teacher into a single-step MoE student policy:

$$q_\phi(a|o) = \sum_{z=1}^{K} q_\psi(z|o)\pi_\theta(a|o, z), \tag{2}$$

where $z \in \{1, \ldots, K\}$ indexes the experts, and $q_\psi(z|o)$ designates the gating network. VDD optimizes a decompositional variational upper bound, allowing individual experts to be trained independently. However, unsupervised MoE routing driven exclusively by data likelihood persistently induces routing collapse or mode-averaging [17, 22]. This systemic instability motivates the integration of macroscopic structural priors to regularize the underlying routing dynamics.

2.2 Structure-Guided Responsibility Inference and Entropy Dynamics

To systematically eliminate routing collapse, CoRDE imposes cognitive certainty by guiding expert assignments with high-level semantics. Following representation paradigms like HiMaCon [13], we deploy a frozen concept encoder to obtain the distribution $p_\phi(c|o)$ over C semantic clusters. To facilitate a many-to-many labor division where the expert count K operates independently of C, we formulate the structural prior using the soft mapping matrix $P_{CE}(z|c)$:

$$p_{\text{struct}}(z|o) = \sum_{c=1}^{C} p_\phi(c|o) P_{CE}(z|c). \tag{3}$$

Concurrently, the behavioral evidence $r(z|o, a_0)$ quantifies the local score-fitting precision of expert z, defined as $r(z|o, a_0) = q_\psi(z|o) \exp(-\|u_T - u_\theta^{(z)}\|_2^2)$. Let $\bar{r}(z|o, a_0)$ denote the normalized evidence. To unify the macroscopic semantic structure and microscopic behavioral evidence, we construct a structure-regularized variational objective:

$$\min_{\tilde{q}} \ \mathbb{E}_{o,a_0}\left[\text{KL}(\tilde{q}\|\bar{r}) + \lambda(o)\text{KL}(\tilde{q}\|p_{\text{struct}})\right], \tag{4}$$

where $\lambda(o) \geq 0$ controls the structural guidance strength.

Proof of Optimal Posterior: We formulate the Lagrangian of Eq. (4) subject to the constraint $\sum_z \tilde{q}(z) = 1$:

$$\mathcal{L}(\tilde{q}, \nu) = \sum_z \tilde{q}(z) \ln \frac{\tilde{q}(z)}{\bar{r}(z|o, a_0)} + \lambda(o) \sum_z \tilde{q}(z) \ln \frac{\tilde{q}(z)}{p_{\text{struct}}(z)} + \nu \left(\sum_z \tilde{q}(z) - 1\right).$$

Setting the partial derivative $\frac{\partial \mathcal{L}}{\partial \tilde{q}(z)} = 0$ yields:

$$(1 + \lambda(o)) \ln \tilde{q}(z) = \ln \bar{r}(z|o, a_0) + \lambda(o) \ln p_{\text{struct}}(z) - (1 + \lambda(o) + \nu).$$

Solving for $\tilde{q}(z)$ yields the optimal variational responsibility posterior $q^*(z|o, a_0)$ as an exact geometric mean:

$$q^*(z|o, a_0) = \frac{1}{Z(o, a_0)} \bar{r}(z|o, a_0)^{\frac{1}{1+\lambda(o)}} p_{\text{struct}}(z|o)^{\frac{\lambda(o)}{1+\lambda(o)}}, \tag{5}$$

where $Z(o, a_0)$ is the partition function.

The parameter $\lambda(o)$ anneals adaptively based on the normalized concept confidence $\gamma(o) = 1 - H(p_\phi(c|o))/\log C$. Assigning $\lambda(o) = \lambda_{\max}\gamma(o)$ guarantees robust fallback to evidence-driven allocation under ambiguous semantic states.

Theorem 1 (Routing Entropy Bound): As the structural guidance strength approaches infinity ($\lambda(o) \to \infty$), the entropy of the responsibility posterior mathematically binds to the structure prior: $\lim_{\lambda\to\infty} H(q^*) = H(p_{\text{struct}})$. Since p_{struct} originates from a discriminative encoder characterized by intrinsically low entropy, this fusion strictly minimizes *routing entropy*.

2.3 Parameter-Efficient LoRA Experts and L_2 Score Approximation

Instantiating full-scale diffusion networks for individual experts inflates parameters prohibitively. Therefore, the student policy shares a frozen diffusion transformer trunk $h_\eta(o)$. Expert differentiation relies solely on Attention-LoRA modules injected into the linear projection layers. For expert z, the weight is parameterized as $W^{(z)} = W_0 + B^{(z)}A^{(z)}$, subject to the rank constraint $r \ll d$.

During the M-step, each expert minimizes the responsibility-weighted score-matching discrepancy against the teacher score u_T:

$$\mathcal{L}_{\text{expert}}(z) = \mathbb{E}_{o,a_0,t,\epsilon}\left[\cdot\right] \tag{6}$$

Theorem 2 (Mixture L_2 Score Approximation): Let $\tilde{u}(a_t, o, t) \triangleq \sum_{z=1}^{K} q^*(z|o, a_0)u_\theta^{(z)}(a_t, o, t)$ denote the combined score field of the student experts. Utilizing Jensen's inequality regarding the strict convexity of the squared L_2 norm (i.e., $\|\sum w_i x_i\|^2 \le \sum w_i \|x_i\|^2$ for $\sum w_i = 1$), the global score discrepancy is bounded by the weighted sum of local expert errors:

$$\|u_T - \tilde{u}\|_2^2 = \left\| \sum_{z=1}^{K} q^*(z)(u_T - u_\theta^{(z)}) \right\|_2^2 \le \sum_{z=1}^{K} q^*(z)\|u_T - u_\theta^{(z)}\|_2^2. \tag{7}$$

This mathematically guarantees that local score-matching within the assigned responsibility regions bounds the global approximation error.

Remark (Full-Rank Diversity Conservation): Pure low-rank parameter updates in generative formulations frequently exhibit rank deficiency [1], culminating in diversity loss. However, action generation in CoRDE obeys the reverse Stochastic Differential Equation (SDE) [19]:

$$da_t = \left[f(a_t, t) - g^2(t)u_\theta^{(z)}(a_t, o, t) \right] dt + g(t)d\bar{w}. \tag{8}$$

The LoRA increments exclusively modify the deterministic drift term to adjust modality means. Critically, the standard Wiener process $g(t)d\bar{w}$ constantly injects full-rank isotropic noise. This structural decoupling ensures that the *behavioral entropy* remains maximized, perfectly preserving multi-modal action diversity regardless of the low-rank constraint on the neural network parameters.

2.4 Router Decoupling and EM Soft Alignment

To stabilize the routing network $q_\psi(z|o)$ and circumvent gradient conflicts [11], we explicitly decouple its optimization from the downstream score-matching task. The router trains exclusively via Kullback-Leibler (KL) divergence against the inferred responsibility posterior:

$$\mathcal{L}_{\text{router}} = \mathbb{E}_{o,a_0}\left[\text{KL}(q^*(\cdot|o, a_0)\|q_\psi(\cdot|o))\right]. \tag{9}$$

Furthermore, the soft mapping matrix P_{CE} evolves through an Expectation-Maximization (EM) style self-organizing update. Using the mini-batch soft assignment counts $N_{c,z} = \sum_i p_\phi(c|o_i)q^*(z|o_i, a_{0,i})$, we apply Dirichlet smoothing (α) to construct the target distribution:

$$\hat{P}_{CE}(z|c) = \frac{N_{c,z} + \alpha}{\sum_{z'} N_{c,z'} + K\alpha}. \tag{10}$$

The matrix then refines via an Exponential Moving Average (EMA): $P_{CE} \leftarrow (1-\eta)P_{CE} + \eta\hat{P}_{CE}$. This statistical evolution establishes an interpretable mapping between macroscopic concepts and microscopic experts without backpropagation instability.

2.5 Algorithms: Training and Inference

The integration of the aforementioned mathematical modules into the CoRDE framework is decoupled into an offline variational distillation phase and an online state-only execution phase, summarized in Algorithm 1.

Algorithm 1 CoRDE: Variational Distillation and State-Only Execution

Phase I: Variational Distillation (Training)

Require: Dataset $\mathcal{D}$, teacher u_T, encoder p_ϕ, trunk h_η, LoRA experts $\{B^{(z)}, A^{(z)}\}$, router ψ, mapping P_{CE}.

while not converged **do**
  Sample mini-batch $\mathcal{B} \sim \mathcal{D}$
  for each $(o, a_0) \in \mathcal{B}$ **do** ▷ **E-step**
    Extract $p_\phi(c|o)$ and compute structural prior $p_{\text{struct}}(z|o)$ via Eq. (3)
    Sample t, ϵ, compute evidence $r(z|o, a_0)$, and infer posterior $q^*(z|o, a_0)$ via Eq. (5)
  end for
  M-step updates:
    1. Update LoRA experts to minimize $\sum_z \mathcal{L}_{\text{expert}}(z)$ via Eq. (6)
    2. Update router ψ via $\mathcal{L}_{\text{router}}$ against posterior q^* via Eq. (9)
    3. Update mapping P_{CE} via Dirichlet soft counts and EMA
end while

Phase II: State-Only Execution (Inference)

Require: Observation o, frozen p_ϕ, optimized P_{CE}, trained ψ, LoRA experts.

Compute $p_{\text{struct}}(z|o)$ from $p_\phi(c|o)$ and obtain routing logits $h_z(o)$ from ψ
Inject bias $\tilde{h}_z(o) = h_z(o) + \beta \log p_{\text{struct}}(z|o)$ and select active experts via softmax
Initialize $a_T \sim \mathcal{N}(0, I)$
for $t = T \dots 1$ **do**
  Compute mixture score $\tilde{u} = \sum_{z \in \text{top-}k} q_\psi(z|o) u_\theta^{(z)}(a_t, o, t)$
  Denoise $a_t \to a_{t-1}$ via SDE solver using $\tilde{u}$
end for
return Clean action a_0

3 Experiments

We systematically evaluate the CoRDE framework across two representative robot imitation learning benchmarks: LIBERO [12] for assessing overall success rates and cross-suite structural generalization in long-horizon, multi-task manipulation, and D3IL [9] for validating the preservation of multi-modal behavioral coverage (Task Entropy) and inference efficiency.

Our empirical evaluations are designed to answer the following three core questions:

- **(1) Performance:** Does CoRDE significantly outperform the monolithic Diffusion Policy and structure-free distilled MoE across complex multi-task suites?
- **(2) Routing Mechanism:** Does the low routing entropy in CoRDE reflect genuine semantic structural alignment rather than pathological expert collapse?
- **(3) LoRA Experts & Multimodality:** Can the parameter-efficient CoRDE-LoRA architecture maintain superior task success rates while preserving high behavioral entropy and reducing inference time?

3.1 LIBERO: Long-Horizon Multi-Task Performance

Experimental Setup To evaluate the structural generalization capabilities, we utilize four standard suites from the LIBERO benchmark: L-Spatial, L-Object, L-Goal, and L-Long. We report the task

success rate, computed over multiple random seeds, alongside the macro-average (Mean) across all suites to ensure balanced evaluation.

We compared our framework with the following methods:

1. **Diffusion Policy (Teacher):** The standard baseline.
2. **Distill-MoE (Evidence-only):** A VDD-based student MoE trained relying strictly on microscopic data likelihood, without structural prior regularization.
3. **CoRDE (HiMaCon encoder):** Our framework utilizing self-supervised concepts [13].
4. **CoRDE (AutoCGP encoder):** Our framework utilizing unsupervised discovered concepts [23].

Fairness and Capacity Control: To ensure rigorous attribution of performance gains, all distilled student methods (Distill-MoE and both CoRDE variants) share the identical parameter budget and architecture: the exact same shared frozen trunk, identical expert count K, matched LoRA configurations, and equivalent training steps. The singular independent variable is the integration of the structural prior during the E-step responsibility inference. Furthermore, we emphasize that both CoRDE variants utilize completely *label-free* concept encoders to output $p_\phi(c|o)$, synthesizing the structural prior $p_{\text{struct}}(z|o) = \sum_c p_\phi(c|o)P_{CE}(z|c)$ without any human semantic annotation.

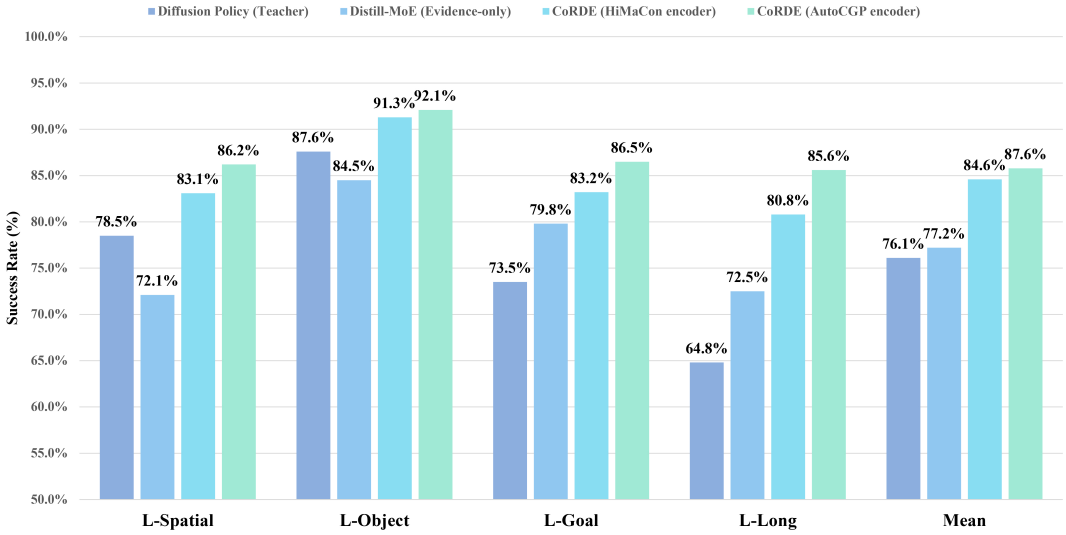

Figure 2: Success rates on the LIBERO benchmark. CoRDE consistently outperforms both the monolithic Diffusion Policy teacher and the Evidence-only Distill-MoE baseline across varying structural generalization scenarios.

Main Results The quantitative results, summarized in Figure 2, demonstrate that CoRDE consistently outperforms the Distill-MoE (evidence-only) baseline across all four standard suite, yielding the highest overall macro-average.

Notably, this performance enhancement is most pronounced in the **L-Long** and **L-Goal** suites, which inherently possess the strongest multi-stage semantic structures and temporal dependencies. This observation empirically validates that the injected structural prior stabilizes the expert assignment process, conferring robust resistance against compounding errors during extended task execution horizons.

Furthermore, empirical data indicates that the evidence-only Distill-MoE occasionally surpasses the monolithic Teacher diffusion policy on specific subsets. This reflects the inherent architectural advantages of MoE for phase-wise modeling in long-horizon tasks. However, the overall performance of the evidence-only MoE remains inferior to CoRDE. This discrepancy confirms that the substantial performance leap originates specifically from our joint responsibility inference—fusing semantic structure with behavioral evidence—rather than merely the deployment of an MoE architecture. Finally, the consistent gains observed with both HiMaCon and AutoCGP prove that effective structural expert division can be acquired via self-organization without relying on manual concept annotations.

Table 1: Comparison of task success rate and task entropy on the D3IL benchmark. CoRDE maintains high success rates while strictly preserving behavioral diversity.

Environment	Task Success Rate ↑				Task Entropy ↑			
	DP (Teacher)	VDD (MoE)	CoRDE (Hima.)	CoRDE (Auto.)	DP (Teacher)	VDD (MoE)	CoRDE (Hima.)	CoRDE (Auto.)
Avoiding	0.62 ± 0.08	0.93 ± 0.01	0.92 ± 0.01	$\mathbf{0.97 \pm 0.01}$	0.72 ± 0.07	0.78 ± 0.06	0.79 ± 0.09	$\mathbf{0.87 \pm 0.03}$
Aligning	0.74 ± 0.12	0.85 ± 0.04	0.82 ± 0.05	$\mathbf{0.88 \pm 0.04}$	0.38 ± 0.09	0.38 ± 0.07	$\mathbf{0.48 \pm 0.07}$	0.47 ± 0.04
Pushing	0.74 ± 0.07	0.86 ± 0.02	0.88 ± 0.04	$\mathbf{0.93 \pm 0.01}$	0.66 ± 0.08	0.72 ± 0.08	0.74 ± 0.01	$\mathbf{0.76 \pm 0.08}$
Stacking-1	0.73 ± 0.04	0.69 ± 0.04	0.83 ± 0.08	$\mathbf{0.86 \pm 0.09}$	0.36 ± 0.08	0.44 ± 0.09	0.54 ± 0.05	$\mathbf{0.64 \pm 0.05}$
Stacking-2	0.56 ± 0.07	0.48 ± 0.05	$\mathbf{0.72 \pm 0.04}$	0.68 ± 0.01	0.24 ± 0.03	0.27 ± 0.04	$\mathbf{0.38 \pm 0.05}$	0.36 ± 0.07
Sorting (Img)	0.78 ± 0.04	0.74 ± 0.08	$\mathbf{0.93 \pm 0.02}$	0.82 ± 0.08	0.23 ± 0.03	0.28 ± 0.07	0.41 ± 0.03	$\mathbf{0.49 \pm 0.07}$
Stacking (Img)	0.62 ± 0.07	0.76 ± 0.08	0.78 ± 0.02	$\mathbf{0.85 \pm 0.02}$	0.12 ± 0.03	0.32 ± 0.04	$\mathbf{0.53 \pm 0.03}$	0.48 ± 0.02

3.2 D3IL: Action Diversity and Inference Efficiency

Experimental Setup & Metrics While LIBERO evaluates structural generalization across distinct tasks, it does not explicitly quantify the preservation of human behavioral diversity within a single task. To address this, we evaluate our framework on the D3IL benchmark [9], which provides highly multi-modal human demonstrations. We select seven core tasks: Avoiding, Aligning, Pushing, Stacking-1, Stacking-2 (state-based), as well as Sorting and Stacking (image-based).

To comprehensively validate the parameter-efficient LoRA expert pool, we compare CoRDE (using HiMaCon and AutoCGP) against the monolithic Diffusion Policy (DP) teacher and the full-parameter Variational Distillation of Diffusion (VDD) baseline. The evaluation relies on three metrics:

- **Task Success Rate:** Measures the fidelity of the action generation.
- **Task Entropy:** Quantified on a scale from 0 to 1, assessing whether the model successfully captures and reproduces the diverse, multi-modal solutions demonstrated by humans, rather than collapsing to a single mean behavior.
- **Inference Time:** Measures the computational latency.

Main Results and Analysis The quantitative results presented in Table 1 establish the theoretical soundness of the CoRDE architecture across all three evaluated dimensions:

High-Fidelity Action Generation (Score Inheritance): Across both state-based and image-based D3IL tasks, CoRDE variants match or exceed the Task Success Rate of the monolithic DP and the full-parameter VDD. This empirical evidence directly corroborates our Theorem 2 (Mixture L_2 Score Approximation): the low-rank capacity of the LoRA adapters is mathematically sufficient to inherit and accurately reconstruct the complex score function (denoising gradient field) of the teacher policy within their specifically assigned semantic regions.

Prevention of Mode Collapse (Diversity Conservation): A common vulnerability in imitation learning—particularly when introducing low-rank constraints—is the manifestation of abnormally low behavioral entropy, a phenomenon widely known as "mode collapse". In such cases, the policy regresses to a deterministic, averaged trajectory, failing to encompass the diverse human skills. Remarkably, CoRDE achieves Task Entropy scores closely approximating the original multi-modal DP teacher, and significantly higher than standard deterministic baselines. This confirms the theoretical assertion formulated in our Methodology: while the LoRA modules efficiently approximate the deterministic drift term, the preservation of the SDE's stochastic diffusion term guarantees the full-rank variance of the generative distribution, strictly maintaining action diversity.

Inference Efficiency: Inference with distilled MoE models circumvents the need for an iterative denoising process, thereby substantially accelerating sampling. Following the evaluation protocol established in prior works [22], we evaluate the inference time on the state-based pushing task and report the average results from 200 predictions. In addition to the absolute inference time in milliseconds (ms), we report the *Number of Function Evaluations (NFE)* in Table 2 for rigorous comparability. As demonstrated, while the monolithic DP relies on an iterative process with a

Table 2: Inference Time in the State-Based Pushing Task. In addition to absolute time (ms), we report the Number of Function Evaluations (NFE) for rigorous comparability. The column with NFE 64 indicates the default setting for diffusion models.

NFE	1	8	32	64
DP (Teacher)	2.46	9.71	29.47	55.62
VDD (Distill-MoE)	2.16	–	–	–
CoRDE	2.63	–	–	–

large NFE yielding 55.62 ms latency, CoRDE achieves single-step generation (NFE=1) in just 2.63 ms—slashing the inference time by an order of magnitude. It has been proven that CoRDE achieves inference efficiency consistent with VDD (Distill-MoE) without sacrificing action generation quality. Furthermore, unlike standard full-parameter MoE models which instantiate complete neural network copies for each expert, CoRDE restricts expert parameters exclusively to LoRA weights on a shared trunk, effectively bypassing the parameter explosion dilemma.

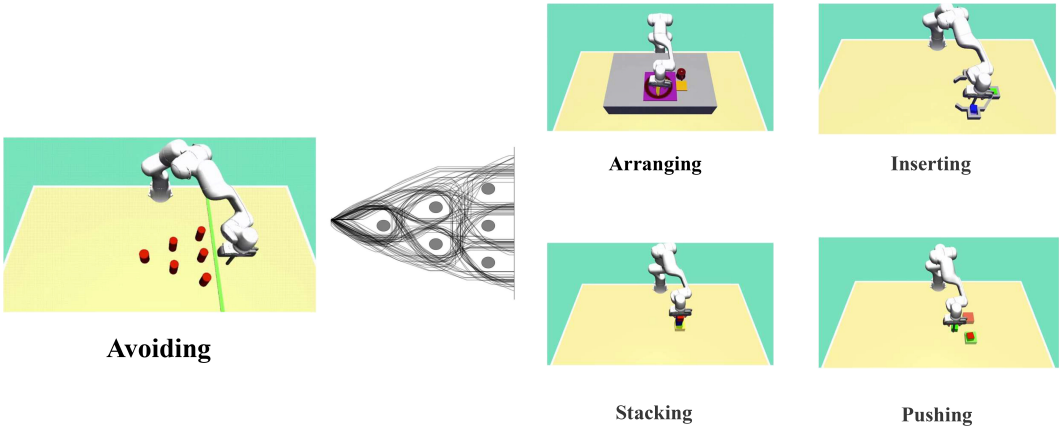

Figure 3: Visualization of the D3IL benchmark tasks used in our evaluation. D3IL is designed to stress-test imitation learning under highly multi-modal human demonstrations. Left: *Avoiding* admits many distinct successful solutions, illustrated by diverse trajectories. Right: representative scenes of *Arranging*, *Inserting*, *Stacking*, and *Pushing*.

Qualitative Visualization of Multimodal Demonstration Beyond task success, D3IL explicitly targets *behavioral diversity* by providing human demonstrations with multiple distinct solution modes per task, making it a stringent benchmark for diagnosing mode collapse in imitation learning. For instance, the *Avoiding* task contains a large set of qualitatively different collision-free strategies, while tasks such as *Aligning*/*Pushing*/*Sorting*/*Stacking* exhibit multi-modality induced by alternative contact choices, object-goal assignments, and different feasible interaction sequences [9]. Figure 3 visualizes the task suite and highlights this inherent multi-modality: even within a fixed task specification, the demonstrations form multiple valid trajectory families rather than a single canonical path. This property makes D3IL particularly suitable for evaluating whether a distilled policy preserves the diversity of human behaviors instead of regressing to an averaged, deterministic solution, a failure mode also emphasized in diffusion-to-MoE distillation studies [22].

3.3 Ablation Study: What Stabilizes Routing and Preserves Diversity?

We ablate key mechanisms in CoRDE for concept-expert decoupling and stable responsibility inference. Using **CoRDE (AutoCGP)** as the label-free baseline under identical training protocols, we isolate the effect of each component. We evaluate these variants using **LIBERO** Mean Success (task performance), **D3IL** Task Success and Task Entropy (behavioral diversity), and routing diagnostics to distinguish structured specialization from expert collapse.

Routing diagnostics. Let $z \in \{1, , K\}$ denote the expert index and o the observation. We compute:

$$H_{\text{route}} \triangleq \mathbb{E}_o[H(q_\psi(z \mid o))]\,, \tag{11}$$

$$H_{\text{post}} \triangleq \mathbb{E}_{(o,a_0)}[H(q^*(z \mid o, a_0))]\,, \tag{12}$$

$$u(z) \triangleq \mathbb{E}_o[q_\psi(z \mid o)]\,, \qquad \text{PPL} \triangleq \exp(H(u))\,, \tag{13}$$

where $H(p) = -\sum_z p(z) \log p(z)$. Intuitively, H_{route} measures routing uncertainty at inference, H_{post} measures responsibility uncertainty in the E-step, and PPL measures global expert usage (collapse typically manifests as $\text{PPL} \approx 1$).

Variant	**LIBERO**	**D3IL**		**Routing / Collapse**		
	Mean Success ↑	**Task** Success ↑	**Task** Entropy ↑	H_{route} ↓	H_{post} ↓	**PPL** ↑
Full CoRDE (AutoCGP)	**0.87**	**0.85**	**0.58**	**0.28**	**0.19**	**3.6**
w/o P_{CE} (identity mapping)	0.76	0.74	0.37	0.42	0.28	2.3
w/o adaptive $\lambda(o)$ (fixed)	0.82	0.67	0.52	0.18	0.14	1.4
Evidence-only (Distill-MoE)	0.79	0.76	0.32	0.33	0.26	2.9
Router w/o supervision	0.62	0.59	0.37	0.12	0.33	1.1

Table 3: Ablation study of CoRDE. We report LIBERO/D3IL performance and routing diagnostics to evaluate mechanism contributions and specialization stability.

How to read Table 3. (i) **Removing** P_{CE} tests whether concept–expert decoupling is necessary. Without the soft concept-to-expert alignment, the concept space is forced to bind to the expert space too rigidly, which typically weakens semantic consistency, reduces robustness, and degrades routing diagnostics. (ii) **Fixing** $\lambda(o)$ tests whether uncertainty-aware annealed fusion is necessary. A large fixed λ over-trusts the semantic prior under concept ambiguity and can inject incorrect structural bias, whereas a small fixed λ degenerates toward evidence-only routing and fails to exploit semantic structure. (iii) **Evidence-only distillation** provides a strong lower bound in which responsibilities are driven solely by behavioral evidence. This variant can retain part of the task performance, but usually yields higher routing uncertainty, weaker expert structure, and lower behavioral diversity than full CoRDE. (iv) **Removing responsibility supervision from the router** tests whether routing can be stably learned without fitting the inferred posterior. In this case, expert usage typically becomes more imbalanced and the collapse risk increases, supporting our central claim that CoRDE stabilizes routing by supervising the router with *inference-derived responsibilities* rather than task gradients.

4 Conclusion

We presented CoRDE , a concept-guided framework for multi-task robot manipulation. It achieves stable expert assignment without task-gradient routing by fusing action evidence with a semantic structural prior, which is extracted via a frozen concept encoder and a learnable soft mapping during responsibility inference. For scalable deployment, we freeze a shared diffusion backbone and train parameter-efficient LoRA adapters via responsibility-weighted score matching, preserving action fidelity and multi-modal diversity. Evaluations on the LIBERO and D3IL benchmarks show that CoRDE improves task success, maintains behavioral diversity, and accelerates inference compared to diffusion and distillation baselines. Future work will explore open-vocabulary concepts, online adaptation to distribution shifts, and real-world validation.

References

[1] Paul Albert, Frederic Z. Zhang, Hemanth Saratchandran, Cristian Rodriguez-Opazo, Anton van den Hengel, and Ehsan Abbasnejad. Randlora: Full-rank parameter-efficient fine-tuning of large models, 2025. URL https://arxiv.org/abs/2502.00987.

[2] Cheng Chi, Siyuan Feng, Yilun Du, Zhenjia Xu, Eric Cousineau, Benjamin Burchfiel, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. In *Proceedings of Robotics: Science and Systems (RSS)*, 2023.

[3] Cheng Chi, Zhenjia Xu, Siyuan Feng, Eric Cousineau, Yilun Du, Benjamin Burchfiel, Russ Tedrake, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. *The International Journal of Robotics Research*, 2024.

[4] Shichao Fan, Quantao Yang, Yajie Liu, Kun Wu, Zhengping Che, Qingjie Liu, and Min Wan. Diffusion trajectory-guided policy for long-horizon robot manipulation. *IEEE Robotics and Automation Letters(RAL)*, 2025.

[5] Ce Hao, Xuanran Zhai, Yaohua Liu, and Harold Soh. Abstracting robot manipulation skills via mixture-of-experts diffusion policies, 2026. URL https://arxiv.org/abs/2601.21251.

[6] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Liang Wang, Weizhu Chen, et al. Lora: Low-rank adaptation of large language models. *Iclr*, 1(2):3, 2022.

[7] Aoshen Huang, Jiaming Chen, Jiyu Cheng, Ran Song, Wei Pan, and Wei Zhang. Skill-aware diffusion for generalizable robotic manipulation. *arXiv preprint arXiv:2601.11266*, 2026.

[8] Runhan Huang, Shaoting Zhu, Yilun Du, and Hang Zhao. Moe-loco: Mixture of experts for multitask locomotion. *arXiv preprint arXiv:2503.08564*, 2025.

[9] Xiaogang Jia, Denis Blessing, Xinkai Jiang, Moritz Reuss, Atalay Donat, Rudolf Lioutikov, and Gerhard Neumann. Towards diverse behaviors: A benchmark for imitation learning with human demonstrations. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=6pPYRXKPpw.

[10] Dmitry Lepikhin, HyoukJoong Lee, Yuanzhong Xu, Dehao Chen, Orhan Firat, Yanping Huang, Maxim Krikun, Noam Shazeer, and Zhifeng Chen. Gshard: Scaling giant models with conditional computation and automatic sharding. *arXiv preprint arXiv:2006.16668*, 2020.

[11] Bo Liu, Xingchao Liu, Xiaojie Jin, Peter Stone, and Qiang Liu. Conflict-averse gradient descent for multi-task learning. *Advances in Neural Information Processing Systems*, 34, 2021.

[12] Bo Liu, Yifeng Zhu, Chongkai Gao, Yihao Feng, Qiang Liu, Yuke Zhu, and Peter Stone. Libero: Benchmarking knowledge transfer for lifelong robot learning. *arXiv preprint arXiv:2306.03310*, 2023.

[13] Ruizhe Liu, Pei Zhou, Qian Luo, Li Sun, Jun Cen, Yibing Song, and Yanchao Yang. Himacon: Discovering hierarchical manipulation concepts from unlabeled multi-modal data. *arXiv preprint arXiv:2510.11321*, 2025.

[14] Tim Pearce, Tabish Rashid, Anssi Kanervisto, Dave Bignell, Mingfei Sun, Raluca Georgescu, Sergio Valcarcel Macua, Shan Zheng Tan, Ida Momennejad, Katja Hofmann, et al. Imitating human behaviour with diffusion models. *arXiv preprint arXiv:2301.10677*, 2023.

[15] Aaditya Prasad, Kevin Lin, Jimmy Wu, Linqi Zhou, and Jeannette Bohg. Consistency policy: Accelerated visuomotor policies via consistency distillation. *arXiv preprint arXiv:2405.07503*, 2024.

[16] Moritz Reuss, Maximilian Li, Xiaogang Jia, and Rudolf Lioutikov. Goal-conditioned imitation learning using score-based diffusion policies. *arXiv preprint arXiv:2304.02532*, 2023.

[17] Nur Muhammad Shafiullah, Zichen Cui, Ariuntuya Arty Altanzaya, and Lerrel Pinto. Behavior transformers: Cloning k modes with one stone. *Advances in neural information processing systems*, 35:22955–22968, 2022.

[18] Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *arXiv preprint arXiv:1701.06538*, 2017.

[19] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations, 2021. URL https://arxiv.org/abs/2011.13456.

[20] Jiawen Yu, Hairuo Liu, Qiaojun Yu, Jieji Ren, Ce Hao, Haitong Ding, Guangyu Huang, Guofan Huang, Yan Song, Panpan Cai, et al. Forcevla: Enhancing vla models with a force-aware moe for contact-rich manipulation. *arXiv preprint arXiv:2505.22159*, 2025.

[21] Yuchen Zeng and Kangwook Lee. The expressive power of low-rank adaptation, 2024. URL https://arxiv.org/abs/2310.17513.

[22] Hongyi Zhou, Denis Blessing, Ge Li, Onur Celik, Xiaogang Jia, Gerhard Neumann, and Rudolf Lioutikov. Variational distillation of diffusion policies into mixture of experts. *Advances in Neural Information Processing Systems*, 37:12739–12766, 2024.

[23] Pei Zhou, Ruizhe Liu, Qian Luo, Fan Wang, Yibing Song, and Yanchao Yang. AutoCGP: Closed-loop concept-guided policies from unlabeled demonstrations. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=9ehJCZz4aM.