ANCHORED MANIFOLD ALIGNMENT STABILIZES BILEVEL REFINEMENT FOR LONG-TAILED DIFFUSION

Anonymous authors
Paper under double-blind review

ABSTRACT

Diffusion models often suffer from degraded tail category generation when processing long-tail distribution data due to extreme sample imbalance. Although Bilevel Optimization has recently been introduced for downstream task alignment in diffusion models, existing frameworks are predominantly designed for training from scratch. This approach readily induces calibration drift when fine-tuning pre-calibrated models, undermining established tail category semantics. This paper proposes a phased decoupled fine-tuning framework specifically designed for long-tail diffusion models. The core innovation lies in introducing the Static Pool Resampling (SPR) mechanism: during upper-layer objective evaluation, resampling from a fixed anchor sample pool generated in the pre-calibration stage (Stage 1) transforms unconstrained sampling optimization into manifold-constrained distribution feature alignment. Theoretical and experimental evidence demonstrates that this "resampled proxy evaluation" maintains statistical consistency with hyper-gradient properties even under extreme long-tail distributions, ensuring both task-specific alignment and distributional integrity. This discovery provides crucial technical support for diffusion model fine-tuning. Results across multiple benchmark long-tail datasets demonstrate that our method enhances task performance while effectively mitigating calibration bias, validating its robustness and superiority under complex long-tail distributions. The code is provided in the supplementary material.

1 INTRODUCTION

Recent work has formulated downstream alignment of diffusion models, such as reward optimization and aesthetic refinement, as a generative bilevel optimization problem (Xiao et al., 2025; Black et al., 2024; Fan et al., 2023). Existing bilevel frameworks, however, are mainly designed for training or alignment from scratch. When directly applied to diffusion models already calibrated on long-tailed data, such as those trained with Oriented Calibration (OC-LT) Zhang et al. (2024), the lower-level updates may disrupt the calibrated tail-class semantics. We refer to this failure mode as *Calibration Drift*: task-driven refinement improves the upper-level objective while gradually moving the model away from the minority-class manifold established during pre-calibration.

This issue is closely related to the decoupled training principle in long-tailed recognition Kang et al. (2020); Sun et al. (2025). A reliable second-stage adaptation should operate on top of a stable and balanced representation rather than freely reshape the learned manifold. For long-tailed diffusion models, this means that bilevel fine-tuning must be anchored to the pre-calibrated generative manifold; otherwise, online upper-level evaluation can be dominated by head-class samples and induce biased hyper-gradients. Therefore, the key challenge is to enable task-specific trajectory refinement while preserving distributional integrity.

To address this challenge, we propose *Decoupled Generative Refinement* (DGR), a two-stage fine-tuning framework for pre-calibrated long-tailed diffusion models. The core component of DGR is *Static Pool Resampling* (SPR). Instead of evaluating the upper-level objective on unconstrained online samples, we first generate a static anchor pool of 50,000 samples from the Stage-1 calibrated model Zhang et al. (2024). During Stage-2 bilevel refinement, the upper-level objective is evaluated by resampling from this fixed pool. In this way, SPR transforms stochastic online alignment into anchored manifold alignment, making the hyper-gradient estimation less sensitive to sampling

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

volatility and more faithful to the calibrated long-tailed distribution. This idea is also consistent with distributional anchoring, where a stable reference distribution is used to regularize fine-tuning Uehara et al. (2024).

Our main contributions are summarized as follows:

❶ We identify *Calibration Drift* as a key obstacle when applying generative bilevel optimization to pre-calibrated long-tailed diffusion models.

❷ We propose DGR, a two-stage decoupled refinement framework that uses OC-LT for manifold foundation and anchored bilevel optimization for task-level trajectory refinement.

❸ We introduce SPR, which evaluates upper-level objectives on a fixed Stage-1 anchor pool, stabilizing hyper-gradient estimation while preserving long-tailed distributional integrity.

2 Background and Foundations

2.1 Long-tailed Diffusion and Decoupled Refinement

Denoising diffusion models and score-based generative models have become standard generative frameworks (Sohl-Dickstein et al., 2015; Ho et al., 2020; Song et al., 2021b;a; Nichol & Dhariwal, 2021; Dhariwal & Nichol, 2021; Ho & Salimans, 2022). They learn a reverse denoising process by predicting the injected noise along a diffusion trajectory. Under long-tailed data, where $\mathbf{N}_1 \gg \mathbf{N}_K$, the learned score field is dominated by head classes, resulting in degraded tail-class synthesis. Recent long-tailed diffusion methods, including CBDM (Qin et al., 2023) and OC-LT (Zhang et al., 2024), alleviate this issue through class-balanced training or oriented head-to-tail calibration, thereby improving manifold recovery for minority categories.

However, Stage-1 distributional calibration alone does not address task-driven refinement. This motivation is consistent with long-tailed recognition studies on re-balancing, margin adjustment, logit calibration, and decoupled classifier adaptation (Cui et al., 2019; Cao et al., 2019; Kang et al., 2020; Ren et al., 2020; Menon et al., 2021; Zhou et al., 2020; Zhong et al., 2021; Wang et al., 2021; Sun et al., 2025). Inspired by this decoupled training paradigm, we view long-tailed diffusion refinement as a two-stage process: Stage 1 builds a calibrated generative manifold, while Stage 2 aligns the sampling trajectory with downstream objectives. Existing generative refinement methods are mostly limited to training-free selection (Kim et al., 2025) or post-hoc heuristics, lacking a parameter-level mechanism for controlled task alignment. This motivates a refinement framework that improves task-specific generation while preserving the calibrated long-tailed manifold.

2.2 Bilevel Optimization and Calibration Drift

Bilevel optimization has been widely used in hyperparameter optimization, meta-learning, and differentiable learning (Maclaurin et al., 2015; Pedregosa, 2016; Franceschi et al., 2018; Grazzi et al., 2020). Task-driven refinement can be formulated as a generative bilevel optimization problem (Xiao et al., 2025):

$$\min_{\phi} \mathcal{F}(\theta^*(\phi), \phi), \quad \theta^*(\phi) = \arg\min_{\theta} \mathcal{L}_{\text{train}}(\theta, \phi), \tag{1}$$

where θ denotes generative model parameters and ϕ denotes structural or sampling hyper-parameters. Existing bilevel or reward-alignment methods, such as AlignProp (Prabhudesai et al., 2024), usually estimate the upper-level gradient with online samples generated during training. In long-tailed scenarios, this online estimate is biased toward head-class signals because tail-class samples are statistically sparse.

This bias causes *calibration drift*: the refinement process gradually moves the model away from the Stage-1 calibrated tail manifold, eroding minority-class semantics. Therefore, upper-level evaluation should not rely solely on stochastic online samples. Instead, it should be anchored to a stable approximation of the calibrated long-tailed distribution. This motivates our Static Pool Resampling (SPR) mechanism, which reuses a fixed anchor pool generated by the Stage-1 model to decouple hyper-gradient estimation from online sampling volatility.

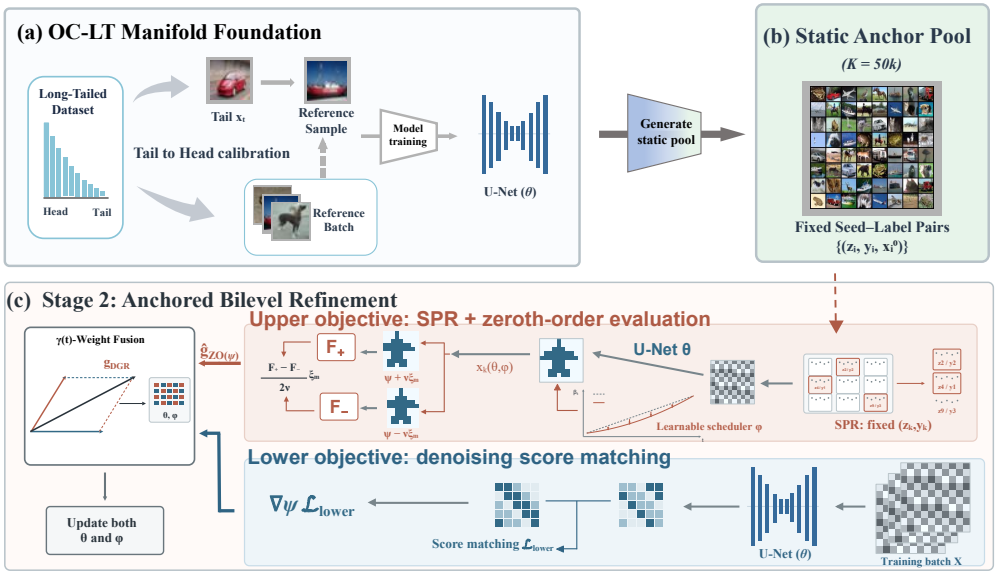


Figure 1: Overview of DGR. Stage 1 constructs a calibrated Static Anchor Pool via OC-LT. Stage 2 reuses fixed anchor seed-label pairs for SPR-based upper-level evaluation and fuses the resulting zeroth-order hyper-gradient with the lower-level score-matching gradient through a $\gamma(t)$-weighted mechanism.

3 Methodology

As illustrated in Fig. 1, we propose the Decoupled Generative Refinement (DGR) framework, a two-stage paradigm designed to align pre-trained long-tail diffusion models with downstream tasks without sacrificing distributional integrity. The framework consists of (1) Manifold Foundation via Oriented Calibration, and (2) Anchored Trajectory Refinement via Bilevel Optimization with Static Pool Resampling.

3.1 Problem Formulation and Calibration Drift

Consider a diffusion model f_θ pre-trained on a long-tailed dataset D_{LT}. Generative bilevel optimization aims to refine structural or sampling hyper-parameters ϕ by minimizing an upper-level task objective,

$$\min_{\phi} \mathcal{F}(\theta^*(\phi), \phi), \quad \theta^*(\phi) = \arg\min_{\theta} \mathcal{L}_{\text{lower}}(\theta, \phi; D_{\text{LT}}), \tag{2}$$

where $\mathcal{F}$ can be a distribution-level metric such as FID (Seitzer, 2020) or a reward-based objective. In long-tailed settings, however, the lower-level gradient is dominated by head-class samples. As a result, unconstrained bilevel refinement shifts the score field toward high-density head regions and gradually erodes the calibrated tail semantics learned in Stage 1. We refer to this failure mode as *calibration drift*. DGR addresses this issue by anchoring the upper-level evaluation to a fixed approximation of the Stage-1 calibrated manifold.

3.2 Stage 1: Pre-Calibration Foundation via Oriented Calibration (T2H)

As illustrated in the top portion of Fig. 1, Stage 1 aims to establish a robust generative foundation by rectifying the biased manifold. To achieve this, we adopt the Oriented Calibration (OC-LT) paradigm (Zhang et al., 2024). OC-LT rectifies the biased score estimation by shifting the denoising target ϵ of a tail sample x_t toward semantic-rich head targets ϵ_z.

For a noisy tail sample x_t, the calibrated denoising target $\hat{\epsilon}_i$ is estimated by:

$$\hat{\epsilon}_i = \epsilon_z, \quad \text{where } z \sim p_{\text{sel}}(z) = \frac{\exp\left(-\frac{\|x_z - x_t\|^2}{2\sigma_t^2}\right)}{\sum_{j \in B} \exp\left(-\frac{\|x_j - x_t\|^2}{2\sigma_t^2}\right)} \tag{3}$$

where B is a balanced reference batch obtained via batch re-sampling. This stage yields a pre-calibrated model f_{θ_0} whose score field is enriched with head-class priors for minority categories. Based on this calibrated model, we construct a static anchor tuple pool

$$\mathcal{P}_{\text{anchor}} = \{(z_k, y_k, x_k^0)\}_{k=1}^K, \qquad x_k^0 = G_{\theta_0,\phi_0}(z_k, y_k), \tag{4}$$

where z_k is the fixed sampling seed, y_k is the class label, and x_k^0 is the Stage-1 generated anchor sample. In our implementation, $K = 50{,}000$. During Stage 2, DGR reuses the fixed seed-label pairs (z_k, y_k) for upper-level evaluation, while x_k^0 serves as a non-parametric approximation of the calibrated Stage-1 manifold.

3.3 Stage 2: Manifold-Anchored Bilevel Optimization

In Stage 2, we refine the sampling trajectory while preserving the calibrated manifold established in Stage 1. Unlike standard BLO (Xiao et al., 2025), which estimates the upper-level objective from online generated samples, DGR anchors upper-level evaluation to the fixed pool $\mathcal{P}_{\text{anchor}}$. The resulting bilevel objective is written as

$$\min_{\phi} \mathcal{F}_{\text{SPR}}\big(\theta^*(\phi), \phi; \mathcal{P}_{\text{anchor}}\big), \qquad \theta^*(\phi) = \arg\min_{\theta} \mathcal{L}_{\text{lower}}(\theta, \phi; D_{\text{LT}}). \tag{5}$$

Here, $\mathcal{F}_{\text{SPR}}$ denotes the upper-level objective evaluated on the Stage-1 anchor pool, while $\mathcal{L}_{\text{lower}}$ is the denoising score-matching objective used to preserve the diffusion manifold.

Importantly, the anchoring effect is not implemented as an explicit KL penalty. Instead, it arises from evaluating the upper-level objective on the fixed Stage-1 pool and fusing the resulting SPR-guided zeroth-order hyper-gradient with the lower-level score-matching gradient. Therefore, Stage 2 converts unconstrained online bilevel refinement into anchored manifold alignment without requiring explicit distribution matching.

3.4 Static Pool Resampling

Static Pool Resampling is the core mechanism of DGR. In standard online BLO (Xiao et al., 2025), the upper-level gradient is estimated from samples generated on the fly. Under long-tailed data, such samples are biased toward head classes, causing tail-class alignment signals to vanish. SPR avoids this by evaluating the upper-level objective on the fixed anchor tuple pool:

$$\hat{g}_{\text{SPR}} = \frac{1}{K} \sum_{k=1}^{K} w_k \nabla_\phi R(G_{\theta,\phi}(z_k, y_k)), \tag{6}$$

where $G_{\theta,\phi}$ denotes the sampling trajectory, $R(\cdot)$ is the upper-level reward or evaluation signal, and (z_k, y_k) is reused from $\mathcal{P}_{\text{anchor}}$. In practice, we construct the anchor pool with class-balanced seed-label pairs and set $w_k = 1$ in the main experiments; the weighted form provides a general importance-sampling interpretation. By fixing the upper-level evaluation inputs, SPR removes sampling volatility and computes hyper-gradients relative to the Stage-1 calibrated manifold rather than transient online batches.

3.5 Hyper-gradient Consistency and Stability

SPR provides a theoretical justification for anchoring Stage 2 refinement to a fixed pool. In standard online BLO, the upper-level signal is estimated from stochastic samples,

$$\hat{g}_{\text{online}} = \frac{1}{B} \sum_{i=1}^{B} \nabla_\phi R(x_i). \tag{7}$$

Under long-tailed data, this estimator is dominated by head-class gradients because tail samples rarely appear in mini-batches, creating a hyper-gradient noise trap.

SPR instead evaluates the upper-level objective on a fixed anchor distribution, which is related to importance sampling and off-policy evaluation (Precup, 2000; Owen, 2013; Swaminathan & Joachims, 2015; Dudík et al., 2011). Let p_{bal} denote the desired balanced distribution and q denote the Stage-1 surrogate distribution. Under the importance-weighted surrogate objective, the SPR estimator can be written as

$$\hat{g}_{\mathrm{SPR}} = \frac{1}{K}\sum_{k=1}^{K} \frac{p_{\mathrm{bal}}(x_k^0)}{q(x_k^0)} \nabla_\phi R(G_{\theta,\phi}(z_k, y_k)) \,. \tag{8}$$

Assuming bounded reward gradients, bounded density ratios, and non-degenerate Stage-1 support, SPR consistently estimates the anchored balanced objective, with MSE decreasing at $O(1/K)$. Its variance is bounded by

$$\mathrm{Var}_q(\hat{g}_{\mathrm{SPR}}) \le \frac{C_R^2}{K}\left(\chi^2(p_{\mathrm{bal}} \| q) + 1\right), \tag{9}$$

where $\chi^2(p_{\mathrm{bal}} \| q)$ measures the gap between the calibrated Stage-1 distribution and the desired balanced distribution. Thus, SPR realizes the decoupling advantage and anchoring effect by decoupling hyper-gradient evaluation from transient online samples. Proofs and generalization analysis are provided in Appendices B.1–B.2.

3.6 Bilevel Trajectory Refinement via ZO-Gradient Fusion

To optimize distribution-level objectives over the SPR anchors, we adopt a Zeroth-Order (ZO) gradient estimator (Xiao et al., 2025; Ghadimi & Lan, 2013; Nesterov & Spokoiny, 2017; Liu et al., 2020). Compared with first-order reward-alignment methods such as AlignProp (Prabhudesai et al., 2024), ZO estimation is more suitable for our setting for two reasons. First, it avoids backpropagating through the full T-step denoising chain, reducing memory overhead from storing intermediate activations. Second, it can optimize non-differentiable or highly non-linear distribution metrics such as FID (Seitzer, 2020) through directional perturbations.

Let ψ denote the trainable parameters refined in Stage 2. In our main experiments, ψ includes the learnable scheduler parameters ϕ and the trainable U-Net parameters θ enabled during refinement. Given perturbation directions $\xi_m \sim \mathcal{N}(0, I)$ and perturbation scale ν, we estimate the upper-level gradient on the fixed anchor pool $\mathcal{P}_{\mathrm{anchor}}$ as

$$\hat{g}_{\mathrm{ZO}}(\psi) = \frac{1}{M}\sum_{m=1}^{M} \frac{\mathcal{F}(\psi + \nu\xi_m; \mathcal{P}_{\mathrm{anchor}}) - \mathcal{F}(\psi - \nu\xi_m; \mathcal{P}_{\mathrm{anchor}})}{2\nu} \xi_m, \tag{10}$$

where $\mathcal{F}$ denotes the upper-level evaluation metric and we set $\nu = 0.001$ in our experiments. Since both perturbed evaluations are performed on the same anchor pool, the estimated gradient reflects the effect of parameter perturbations rather than stochastic sampling fluctuations.

γ-weighted gradient fusion. The final update direction combines the lower-level manifold-preserving gradient with the SPR-guided upper-level gradient:

$$g_{\mathrm{DGR}} = \nabla_\psi \mathcal{L}_{\mathrm{lower}} + \gamma(t)^{-1} \hat{g}_{\mathrm{ZO}}(\psi), \tag{11}$$

where $\gamma(t)$ is an SNR-aware balancing factor. This fusion allows DGR to improve task-level fidelity while keeping the refinement trajectory anchored to the Stage-1 calibrated long-tailed manifold. As a result, DGR performs trajectory refinement without eroding the tail-class semantics established by OC-LT (Zhang et al., 2024) and distributional anchoring (Uehara et al., 2024).

4 Experiments

In this section, we conduct a series of experiments to evaluate the effectiveness of our proposed phased decoupled fine-tuning framework. We specifically aim to demonstrate how the Static Pool Resampling (SPR) mechanism alleviates the "Calibration Drift" and improves the generation quality of tail categories.

Table 1: Comparison with other long-tailed generation methods on CIFAR-10LT and CIFAR-100LT under different imbalance ratios.

Method	CIFAR10LT				CIFAR100LT	
	Imbalance Ratio = 0.01		Imbalance Ratio = 0.001		Imbalance Ratio = 0.01	
	FID (↓)	IS (↑)	FID (↓)	IS (↑)	FID (↓)	IS (↑)
CBGAN (Rangwani et al., 2021)	37.23	6.01	46.61	5.77	33.01	7.04
gSR-GAN (Rangwani et al., 2022)	12.86	8.56	38.71	6.89	13.96	10.01
DDPM (Ho et al., 2020)	10.72	9.37	15.00	9.16	10.25	12.96
CBDM (Qin et al., 2023)	7.27	9.37	12.71	9.01	7.82	12.40
CORAL (Rodriguez et al., 2025)	5.32	9.69	11.03	9.13	5.37	13.53
DDPM+T2H (Zhang et al., 2024)	7.62	9.61	7.88	9.70	6.82	12.19
Ours without SPR	9.67	9.23	10.05	9.58	9.50	11.29
Ours (Bilevel Refinement)	**6.18**	9.67	**6.86**	9.72	**5.73**	12.25

4.1 Experimental Setup

Datasets and Metrics. Following standard long-tailed generation settings (Kang et al., 2020; Zhang et al., 2024), we evaluate DGR on CIFAR-10-LT with $\rho \in \{0.01, 0.001\}$ and CIFAR-100-LT with $\rho = 0.01$. We report FID and IS on 50,000 generated samples (Salimans et al., 2016; Heusel et al., 2017; Seitzer, 2020), with complementary fidelity-diversity diagnostics following prior evaluation protocols (Bińkowski et al., 2018; Kynkäänniemi et al., 2019; Naeem et al., 2020).

Implementation, Reproducibility, and Compute. Our framework follows a two-stage decoupled paradigm. Stage 1 trains a conditional 128-channel U-Net for 200k steps with batch size 128, learning rate 2×10^{-4}, 5k warmup steps, EMA decay 0.9999, CFG training, and T2H transfer calibration. Stage 2 loads the Stage 1 checkpoint and the 50k static anchor pool, then performs bilevel trajectory refinement for 3 epochs with batch size 128, inner-loop steps $(2, 1)$, $\gamma = 5$, a linear learnable scheduler initialized from 10^{-4} to 0.02, scheduler learning rates 5×10^{-5}, U-Net learning rates 5×10^{-6} and 10^{-5}, and ZO perturbation $\delta = 0.001$. We use AdamW with OneCycleLR, a 100-step deterministic DDIM sampler, and CFG scale $\omega = 1.5$ for training and inference. Experiments are run on a single NVIDIA V100-32GB GPU with 6 vCPUs and 25GB RAM. DGR does not require external classifiers during refinement; the alignment signal is obtained from SPR-guided internal hyper-gradient estimation.

4.2 Main Results on Long-Tailed Synthesis

Table 1 reports the main quantitative results. On CIFAR-10-LT with IF=100 ($\rho = 0.01$), DGR reduces the FID of DDPM+T2H from 7.62 to 6.18, a relative improvement of approximately 19%, while increasing IS from 9.61 to 9.67. This shows that Stage-2 trajectory refinement further improves the pre-calibrated model rather than merely preserving its Stage-1 performance. Although CORAL obtains the best reported result in this setting, DGR achieves the second-best FID and IS while directly improving its DDPM+T2H initialization.

Under the more challenging IF=1000 ($\rho = 0.001$) setting, we construct a stronger DDPM+T2H baseline by continuing training from the IF=100 checkpoint, obtaining an FID of 7.88 and an IS of 9.70. DGR further improves these results to 6.86 and 9.72, respectively, achieving the best performance in Table 1. It also improves over the reported CORAL result by 4.17 FID points and 0.59 IS. These results indicate that anchored refinement remains effective even when the upper-level signal becomes highly sensitive to scarce tail samples.

The comparison with the unanchored variant confirms the importance of SPR. Without SPR, refinement degrades the DDPM+T2H FIDs from 7.62, 7.88, and 6.82 to 9.67, 10.05, and 9.50. Reintroducing SPR reduces them to 6.18, 6.86, and 5.73, corresponding to relative improvements of 36.1%, 31.7%, and 39.7% over the unanchored variant. This consistent gap provides direct evidence that unconstrained upper-level evaluation causes calibration drift, whereas SPR preserves the calibrated manifold.

On CIFAR-100-LT with $\rho = 0.01$, DGR improves DDPM+T2H from 6.82 to 5.73 FID, an improvement of approximately 16%, while increasing IS from 12.19 to 12.25. Although CORAL reports the

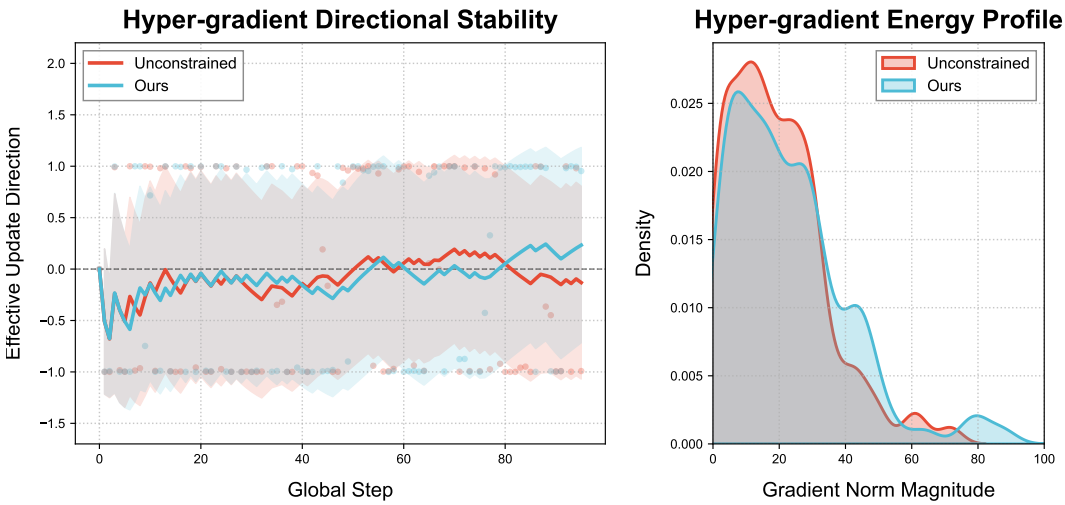


Figure 2: Hyper-gradient Directional Stability and Energy Profile. Left: Directional stability of hypergradients across global training steps; Right: Energy profile (probability density) of hypergradient norm magnitudes.

best result in this setting, DGR remains competitive and substantially outperforms both its Stage-1 initialization and the variant without SPR. Additional class-wise results are provided in Appendix A.2.

To examine scalability beyond CIFAR-scale benchmarks, we additionally evaluate DGR on TinyImageNet200LT in Table 2. DGR achieves an FID of 16.90 and an IS of 18.16, improving the DDPM+T2H baseline of 17.81 and 18.12, respectively. Although Base DDPM obtains a slightly higher IS of 18.20, its substantially worse FID of 19.24 indicates a less favorable fidelity–diversity trade-off. This compact experiment suggests that SPR-based refinement remains applicable to higher-resolution, larger-label-space long-tailed synthesis.

4.3 Mechanistic Analysis: Hyper-gradient Stability and Optimization Geometry

We further analyze why SPR mitigates calibration drift by examining hyper-gradient dynamics and the geometry of the upper-level optimization landscape. A theoretical curvature concentration analysis connecting the empirical landscape to Hessian spectral fluctuations is provided in Appendix B.3.

Table 2: Results on TinyImageNet200LT with diffusion baselines.

Method	FID	IS
Base DDPM (Ho et al., 2020)	19.24	**18.20**
CBDM (Qin et al., 2023)	18.07	18.01
DDPM+T2H (Zhang et al., 2024)	<u>17.81</u>	18.12
Ours (Bilevel Refinement)	**16.90**	<u>18.16</u>

Hyper-gradient stability. Figure 2 compares online bilevel refinement with SPR from two perspectives: directional stability and gradient-energy distribution. The unconstrained online baseline exhibits large directional oscillations and a broad variance band, indicating that its hyper-gradient is highly sensitive to stochastic mini-batch sampling. Its energy profile is also concentrated in the low-norm region, suggesting that effective update signals are weak and easily overwhelmed by noise. This supports the gradient noise trap hypothesis: in long-tailed data, scarce tail-class samples are rarely captured by online batches, causing the upper-level update to be dominated by unstable head-class signals.

In contrast, SPR produces more stable update directions and a more structured energy distribution. By fixing the anchor set generated from the Stage-1 calibrated model, SPR converts transient online evaluation into a deterministic proxy over a globally calibrated manifold. As a result, the hypergradient becomes both more directionally consistent and more informative, enabling scheduler updates to align sampling trajectories without drifting away from tail-class semantics.

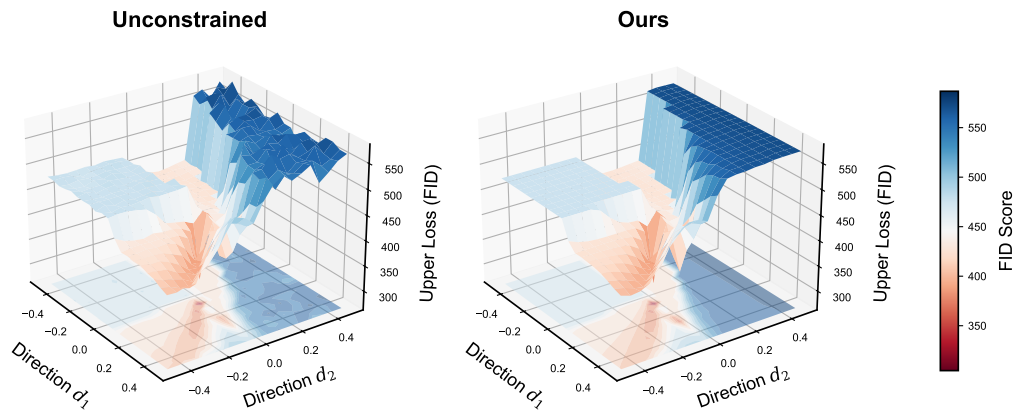
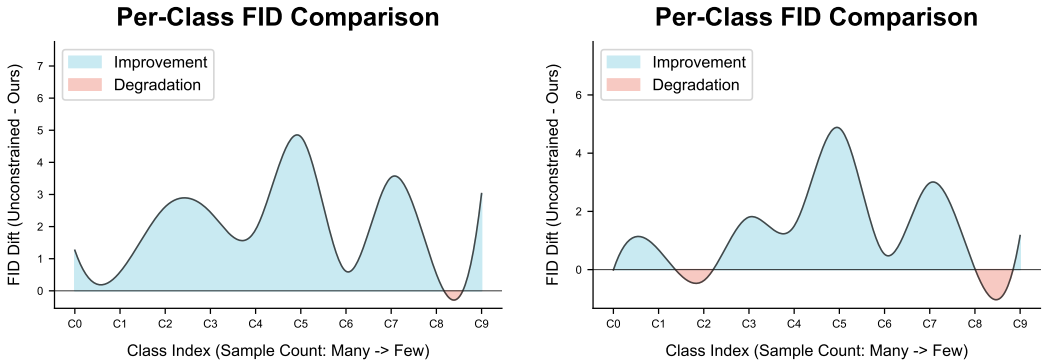


Figure 3: Topological smoothing and flat minima of optimization geometry. Left: Rugged optimization landscape with sharp minima without SPR; right: Smooth landscape with a flatter basin under SPR.



(a) Out-of-pool FID gain under $\rho = 0.01$ (mainstream long-tailed setting).

(b) Out-of-pool FID gain under $\rho = 0.001$ (extreme long-tailed setting).

Figure 4: Out-of-pool per-class FID gain on CIFAR-10-LT. Left: $\rho = 0.01$; right: $\rho = 0.001$. All measurements use re-uniform sampling with seeds unseen during Stage 2 refinement.

Optimization geometry. Figure 3 visualizes the upper-level loss landscape along two orthogonal perturbation directions in parameter space. Without SPR, the landscape is highly rugged and contains sharp local minima, implying that small scheduler perturbations can cause large fluctuations in the upper-level objective. Such sensitivity makes refinement vulnerable to calibration drift under noisy online gradients.

With SPR, the landscape becomes noticeably smoother and forms a broader basin. This suggests that the static anchor pool acts as a non-parametric manifold buffer: it filters out sampling-induced fluctuations and regularizes the upper-level objective around the Stage-1 calibrated distribution. The resulting flatter geometry provides a more robust optimization region, allowing DGR to improve fidelity while remaining anchored to the balanced semantic manifold established in Stage 1.

4.4 Out-of-pool Generalization and Monotonic Gain Analysis

We next examine whether DGR generalizes beyond the static anchor pool used during Stage 2 refinement. To provide a complementary view of semantic trajectories and manifold anchoring, we include PCA visualizations in Appendix A.1. Here, we focus on out-of-pool evaluation to verify that SPR mitigates calibration drift without simply memorizing anchor samples.

To test this, all FID gains in Figure 4 are computed using *re-uniform sampling*: we generate 50,000 samples from random noise seeds that are never used during Stage 2 refinement. This protocol directly evaluates whether the refined sampling trajectory generalizes to unseen latent seeds.

As shown in Figure 4a, under $\rho = 0.01$, DGR achieves consistent per-class FID improvements on out-of-pool seeds. This indicates that SPR does not merely overfit the fixed anchor pool, but instead induces a global refinement of the sampling trajectory. The result is consistent with the hyper-gradient generalization analysis in Appendix A.3.

Under the more extreme $\rho = 0.001$ setting, Figure 4b shows that DGR still maintains broadly non-negative gains across categories. This is important because unconstrained refinement may improve head-class rewards while degrading minority classes. In contrast, SPR stabilizes the refinement process by anchoring upper-level evaluation to the Stage-1 calibrated manifold, thereby reducing the risk of tail-class collapse and calibration drift.

Note: Complementary semantic-fidelity and classification-certainty results using an external oracle evaluator are provided in Appendix A.4.

4.5 Semantic Fidelity Analysis

Beyond perceptual quality measured by FID, we further evaluate whether DGR preserves the semantic correctness of generated samples. This analysis aims to verify that the improvement brought by SPR does not come from diversity collapse or class-boundary distortion, but from better distributional integrity.

Oracle-based semantic fidelity. We use a pre-trained ResNet-56 oracle classifier to evaluate 50,000 generated samples. As shown in Figure 5, classes are grouped into Head, Medium, and Tail categories according to their sample frequencies, and we report the average classification accuracy for each group.

The results show that DGR improves oracle accuracy most clearly on the Tail group. This indicates that the refined model does not merely improve low-level visual texture, but also better preserves discriminative semantics for scarce categories. Therefore, SPR-guided refinement improves generation quality while maintaining class-level semantic fidelity. Complementary analyses of class-wise recall evolution and distributional integrity during optimization are provided in Appendix A.3.

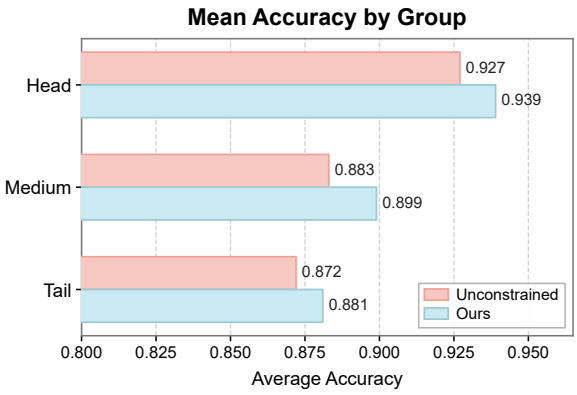


Figure 5: Semantic fidelity on CIFAR-10-LT ($\rho = 0.01$). Mean oracle accuracy over 50,000 generated samples shows that SPR-guided refinement consistently outperforms unconstrained optimization across the Head, Medium, and Tail groups.

5 Conclusion and Discussion

We identify **Calibration Drift** as a failure mode in long-tailed diffusion refinement, where unconstrained bilevel updates disrupt Stage 1 tail semantics. DGR mitigates this drift through Static Pool Resampling, anchoring upper-level evaluation to samples from the calibrated model. Experiments show improved synthesis across imbalance settings and unseen seeds, with more stable hyper-gradients and smoother optimization.

Limitations and broader impacts. DGR depends on anchor-pool coverage and incurs additional ZO evaluations. It may benefit underrepresented categories but retains image-misuse risks. Evaluation is limited to low-resolution public benchmarks, and no general-purpose generator is released. Future work will study adaptive anchor pools and larger diffusion backbones.(Rombach et al., 2022; Peebles & Xie, 2023).

AI USE STATEMENT

Generative AI tools were used to assist with editing and polishing the writing of this manuscript, including improving its clarity, grammar, and readability. All AI-assisted text was carefully reviewed, verified, and revised by the authors. The authors take full responsibility for the final content of the paper.

REFERENCES

Mikołaj Bińkowski, Dougal J. Sutherland, Michael Arbel, and Arthur Gretton. Demystifying MMD GANs. In *Proceedings of the International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=r1lUOzWCW.

Kevin Black, Michael Janner, Yilun Du, Ilya Kostrikov, and Sergey Levine. Training diffusion models with reinforcement learning. In *Proceedings of the International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=YCWjhGrJFD.

Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Aréchiga, and Tengyu Ma. Learning imbalanced datasets with label-distribution-aware margin loss. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/621461af90cadfdaf0e8d4cc25129f91-Abstract.html.

Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. Class-balanced loss based on effective number of samples. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9268–9277, 2019. URL https://arxiv.org/abs/1901.05555.

Prafulla Dhariwal and Alexander Quinn Nichol. Diffusion models beat GANs on image synthesis. In *Advances in Neural Information Processing Systems*, volume 34, pp. 8780–8794, 2021. URL https://papers.nips.cc/paper/2021/hash/49ad23d1ec9fa4bd8d77d02681df5cfa-Abstract.html.

Miroslav Dudík, John Langford, and Lihong Li. Doubly robust policy evaluation and learning. In *Proceedings of the 28th International Conference on Machine Learning*, pp. 1097–1104, 2011. URL https://icml.cc/2011/papers/554_icmlpaper.pdf.

Ying Fan, Olivia Watkins, Yuqing Du, Hao Liu, Moonkyung Ryu, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, Kangwook Lee, and Kimin Lee. DPOK: Reinforcement learning for fine-tuning text-to-image diffusion models. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://arxiv.org/abs/2305.16381.

Luca Franceschi, Paolo Frasconi, Saverio Salzo, Riccardo Grazzi, and Massimiliano Pontil. Bilevel programming for hyperparameter optimization and meta-learning. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 1568–1577. PMLR, 2018. URL https://proceedings.mlr.press/v80/franceschi18a.html.

Saeed Ghadimi and Guanghui Lan. Stochastic first- and zeroth-order methods for nonconvex stochastic programming. *SIAM Journal on Optimization*, 23(4):2341–2368, 2013. doi: 10.1137/120880811. URL https://doi.org/10.1137/120880811.

Riccardo Grazzi, Luca Franceschi, Massimiliano Pontil, and Saverio Salzo. On the iteration complexity of hypergradient computation. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 3748–3758. PMLR, 2020. URL https://proceedings.mlr.press/v119/grazzi20a.html.

M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter. GANs trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 30, 2017.

J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pp. 6840–6851, 2020.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance, 2022. URL https://arxiv.org/abs/2207.12598.

B. Kang, S. Xie, M. Rohrbach, Z. Yan, A. Gordo, J. Feng, and Y. Kalantidis. Decoupling representation and classifier for long-tailed recognition. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2020.

Sunwoo Kim, Minkyu Kim, and Dongmin Park. Test-time alignment of diffusion models without reward over-optimization. In *Proceedings of the International Conference on Learning Representations*, 2025. doi: 10.48550/arXiv.2501.05803. URL https://openreview.net/forum?id=vi3DjUhFVm. Spotlight.

Tuomas Kynkäänniemi, Tero Karras, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Improved precision and recall metric for assessing generative models. In *Advances in Neural Information Processing Systems*, volume 32, 2019. URL https://proceedings.neurips.cc/paper/2019/hash/0234c510bc6d908b28c70ff313743079-Abstract.html.

Sijia Liu, Jie Chen, Pin-Yu Chen, and Alfred O. Hero. A primer on zeroth-order optimization in signal processing and machine learning: Principals, recent advances, and applications. *IEEE Signal Processing Magazine*, 37 (5):43–54, 2020. URL https://arxiv.org/abs/2006.06224.

Dougal Maclaurin, David Duvenaud, and Ryan P. Adams. Gradient-based hyperparameter optimization through reversible learning. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2113–2122. PMLR, 2015. URL https://proceedings.mlr.press/v37/maclaurin15.html.

Aditya Krishna Menon, Sadeep Jayasumana, Ankit Singh Rawat, Himanshu Jain, Andreas Veit, and Sanjiv Kumar. Long-tail learning via logit adjustment. In *Proceedings of the International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=37nvvqkCo5.

Muhammad Ferjad Naeem, Seong Joon Oh, Youngjung Uh, Yunjey Choi, and Jaejun Yoo. Reliable fidelity and diversity metrics for generative models. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 7176–7185. PMLR, 2020. URL https://proceedings.mlr.press/v119/naeem20a.html.

Yurii Nesterov and Vladimir Spokoiny. Random gradient-free minimization of convex functions. *Foundations of Computational Mathematics*, 17(2):527–566, 2017. doi: 10.1007/s10208-015-9296-2. URL https://doi.org/10.1007/s10208-015-9296-2.

Alex Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8162–8171. PMLR, 2021. URL https://proceedings.mlr.press/v139/nichol21a.html.

Art B. Owen. *Monte Carlo Theory, Methods and Examples*. Self-published, 2013. URL https://artowen.su.domains/mc/.

Fabian Pedregosa. Hyperparameter optimization with approximate gradient. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 737–746. PMLR, 2016. URL https://proceedings.mlr.press/v48/pedregosa16.html.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4195–4205, 2023. doi: 10.1109/ICCV51070.2023.00387. URL https://openaccess.thecvf.com/content/ICCV2023/html/Peebles_Scalable_Diffusion_Models_With_Transformers_ICCV_2023_paper.html.

M. Prabhudesai, A. Goyal, D. Pathak, and K. Fragkiadaki. Aligning text-to-image diffusion models with reward backpropagation. In *International Conference on Learning Representations*, 2024.

D. Precup. Eligibility traces for off-policy policy evaluation. In *Proceedings of the 17th International Conference on Machine Learning (ICML)*, pp. 759–766, 2000.

Y. Qin, H. Zheng, J. Yao, M. Zhou, and Y. Zhang. Class-balancing diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18434–18443, 2023.

H. Rangwani, KR. Mopuri, and RV. Babu. Class balancing gan with a classifier in the loop. In *Uncertainty in Artificial Intelligence (UAI)*, pp. 1618–1627. PMLR, 2021.

H. Rangwani, N. Jaswani, T. Karmali, V. Jampani, and RV. Babu. Improving gans for long-tailed data through group spectral regularization. In *European Conference on Computer Vision (ECCV)*, pp. 426–442. Springer, 2022.

Jiawei Ren, Cunjun Yu, Xiao Ma, Haiyu Zhao, Shuai Yi, and Hongsheng Li. Balanced meta-softmax for long-tailed visual recognition. In *Advances in Neural Information Processing Systems*, volume 33, pp. 4175–4186, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/2ba61cc3a8f44143e1f2f13b2b729ab3-Abstract.html.

E. Rodriguez, M. Welfert, N. Stromberg, J. A. Camarena, S. McDowell, and L. Sankar. Coral: Disentangling latent representations in long-tailed diffusion. In *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*, 2025.

R. Römbach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10684–10695, 2022.

Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training GANs. In *Advances in Neural Information Processing Systems*, volume 29, 2016. URL https://proceedings.neurips.cc/paper/2016/hash/8a3363abe792db2d8761d6403605aeb7-Abstract.html.

M. Seitzer. pytorch-fid: FID score for PyTorch. GitHub Repository, 2020.

Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 2256–2265. PMLR, 2015. URL https://proceedings.mlr.press/v37/sohl-dickstein15.html.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *Proceedings of the International Conference on Learning Representations*, 2021a. URL https://openreview.net/forum?id=St1giarCHLP.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *Proceedings of the International Conference on Learning Representations*, 2021b. URL https://openreview.net/forum?id=PxTIG12RRHS.

S. Sun et al. Rethinking classifier re-training in long-tailed recognition: Label over-smooth can balance. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2025.

Adith Swaminathan and Thorsten Joachims. Counterfactual risk minimization: Learning from logged bandit feedback. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 814–823. PMLR, 2015. URL https://proceedings.mlr.press/v37/swaminathan15.html.

M. Uehara, Y. Zhao, K. Black, E. Hajiramezanali, G. Scalia, N. Diamant, A. Tseng, T. Biancalani, and S. Levine. Fine-tuning of continuous-time diffusion models as entropy-regularized control. *arXiv preprint arXiv:2402.15194*, 2024.

Xudong Wang, Long Lian, Zhongqi Miao, Ziwei Liu, and Stella X. Yu. RIDE: Long-tailed recognition by routing diverse distribution-aware experts. In *Proceedings of the International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=D9I3drBz4UC.

Quan Xiao, Hui Yuan, A F M Saif, Gaowen Liu, Ramana Rao Kompella, Mengdi Wang, and Tianyi Chen. A first-order generative bilevel optimization framework for diffusion models. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 68535–68558. PMLR, 2025. URL https://proceedings.mlr.press/v267/xiao25i.html.

T. Zhang, H. Zheng, J. Yao, X. Wang, M. Zhou, Y. Zhang, and Y. Wang. Long-tailed diffusion models with oriented calibration. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2024.

Zhisheng Zhong, Jiequan Cui, Shu Liu, and Jiaya Jia. Improving calibration for long-tailed recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16489–16498, 2021. URL https://openaccess.thecvf.com/content/CVPR2021/html/Zhong_Improving_Calibration_for_Long-Tailed_Recognition_CVPR_2021_paper.html.

Boyan Zhou, Quan Cui, Xiu-Shen Wei, and Zhao-Min Chen. BBN: Bilateral-branch network with cumulative learning for long-tailed visual recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9719–9728, 2020. URL https://openaccess.thecvf.com/content_CVPR_2020/html/Zhou_BBN_Bilateral-Branch_Network_With_Cumulative_Learning_for_Long-Tailed_Visual_Recognition_CVPR_2020_paper.html.

Appendix

Anchored Manifold Alignment Stabilizes Bilevel Refinement for Long-Tailed Diffusion

A Supplementary Empirical Results and Scalability Analysis

A.1 Multi-dimensional Evaluation and Distributional Anchoring Mechanism for Mitigating Calibration Drift

Figure 6 systematically reveals the physical mechanism of Static Pool Resampling (SPR) in mitigating Calibration Drift by integrating fine-grained PRDC gain heatmaps and FID evolution curves across sparse scales. As observed in 6(a), SPR exhibits significant positive gains (dark blue regions) across all classes relative to Standard bilevel optimization in two core dimensions (Density and Precision), peaking in the tail classes (C8–C9) with extremely scarce samples.

The mechanistic origin of this phenomenon lies in transforming the originally dynamically unconstrained evolutionary optimization process into a Constrained manifold alignment task. Within the traditional bilevel optimization framework, the estimation of Hyper-gradient ($\nabla_\lambda F$) is highly dependent on dynamic random samples in the training loop. Under long-tailed distributions, this randomness causes gradient signals to be dominated by head classes, thereby eroding the detailed semantics established in Stage 1 (OC-LT) under Stable feature representation.

The core mechanism by which SPR blocks this drift process is the implementation of Distributional Anchoring. As shown in 6(b) and 6(c), even under the extremely sparse condition of $\rho = 0.001$, the FID gains of tail classes remain favorable. This is because our Two-stage Decoupled Fine-tuning strategy caches a representative Anchor Set from Stage 1, providing a stable reference frame anchored to the "Gold Standard" long-tail distribution established in Stage 1 for Stage 2 bilevel refinement.

Mathematically, this design achieves statistical consistency analogous to Off-policy Evaluation (OPE), ensuring that upper-level feedback signals remain anchored within the Manifold Buffer regardless of how skewed the Lower-level mini-batch data may be. Consequently, SPR effectively filters out random interference caused by sampling variance, enforcing the model to precisely lock onto the Distributional Integrity of Stage 1 while pursuing Task-specific Alignment, fundamentally resolving the Calibration Drift issue faced by existing frameworks when fine-tuning Pre-calibrated models.

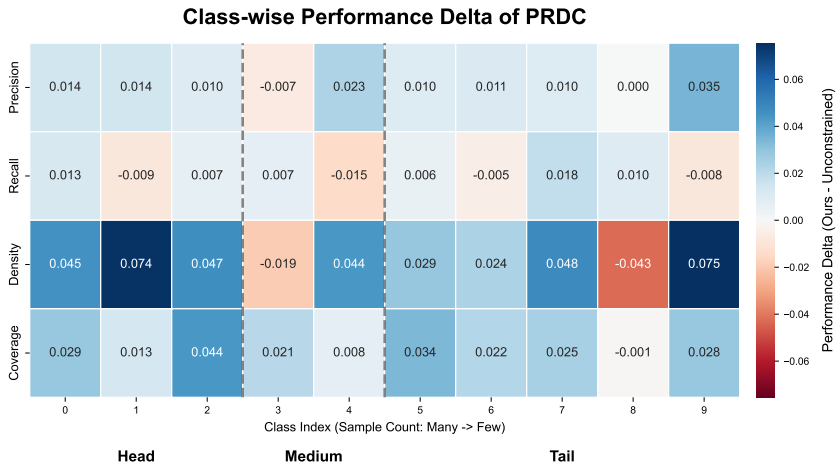
To qualitatively analyze the anchoring effect of SPR, Figure 7(a) visualizes the evolutionary trend of generated samples in the semantic space using PCA. This serves as direct evidence for the theoretical stability claims in Section 4.3. As shown in Figure 7(a), we visualize the evolutionary trend of generated samples in the semantic space using Principal Component Analysis (PCA):

- **Directional Pull-back Effect**: The long red arrows in the figure represent the centroid shift of features for Tail Classes. It can be observed that the fine-tuning process does not randomly expand the sampling boundary, but precisely "pulls back" the deviated tail semantics to the high-density center of the real data distribution. This intuitively demonstrates the guiding role of SPR anchors.
- **Manifold Stability**: The displacement of the blue point set representing head classes is minimal, verifying that during fine-tuning, SPR acts as a "Manifold Buffer" to protect the established mature distribution from being eroded by task-alignment gradients, achieving local and precise repair of the imbalanced manifold.

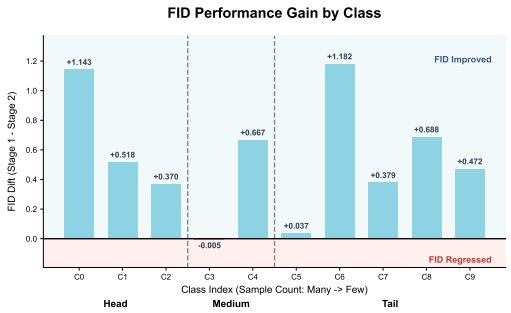
A.2 Scalability and Manifold Coverage Validation in Dense Semantic Spaces

To verify the universality and scalability of the proposed framework in complex multi-class environments, Figure 8 presents the performance of Stage 2 Bilevel Refinement on CIFAR-100-LT ($\rho = 0.01$), which contains 100 classes. As shown in 8(a), the statistical distribution of Recall reveals that after refinement, the mean Recall across all classes increases from 0.360 to 0.403, with the KDE distribution curve shifting significantly rightward, accompanied by a sharp contraction of the long-tailed low-score region on the left.

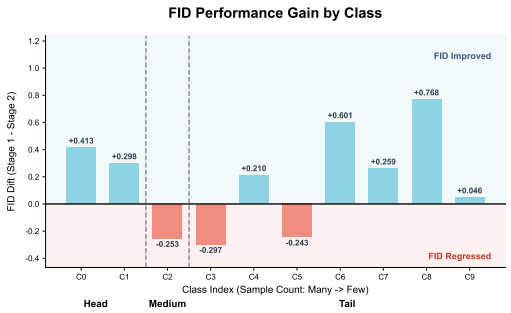
The mechanistic origin of this phenomenon lies in the fact that, in dense semantic spaces, Standard Unconstrained Bilevel Optimization (BLO) faces severe Hyper-gradient Variance Explosion due to



(a) Fine-grained PRDC heatmap ($\text{Score}_{\text{SPR}} - \text{Score}_{\text{Unconstrained}}$, blue = SPR superiority).



(b) Per-class FID gain under $\rho = 0.01$ (consistent positive gains).



(c) Per-class FID gain under $\rho = 0.001$ (stable gains for head/tail classes).

Figure 6: Comprehensive Multi-metric Evaluation of SPR Strategy.

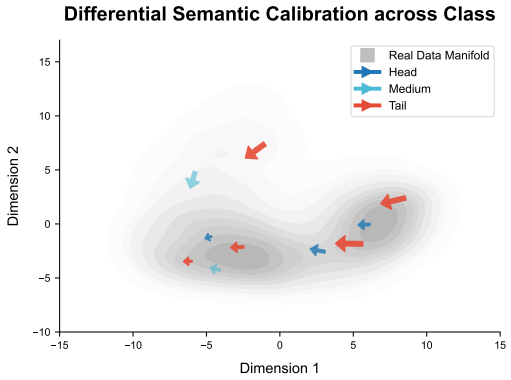
the extreme sparsity of tail samples. Mini-batches sampled randomly can hardly cover the global features of all 100 classes in a single evaluation, causing the upper-level objective to easily ignore the boundary weights of minority classes and thus trigger catastrophic Calibration Drift.

Further, combined with the annular Recall contour in 8(b), it can be observed that in the yellow tail-class regions where semantic erosion is most likely to occur, the performance bars of Stage 2 closely align with and even exceed the red dashed baseline established by Stage 1 (OC-LT). This fully demonstrates that our Two-stage Decoupled Fine-tuning strategy successfully transforms the disordered training caused by a surge in categories into a Constrained Manifold Alignment Task.

Its implementation mechanism is to use the 50,000 calibrated samples from Stage 1 as a static "Manifold Buffer", which acts as a non-parametric prior to regularize the fine-tuning process. This design logically mimics Off-policy Evaluation, ensuring that each fine-grained manifold receives consistent and stable hyper-gradient guidance amid the complex dynamics of concurrent optimization across 100 classes. Ultimately, this strategy not only maintains high fidelity in head classes but also blocks semantic forgetting in tail regions, validating the robustness of the fine-tuning paradigm anchored to the "Gold Standard" long-tail distribution when handling ultra-large-scale imbalanced datasets.

A.3 Evolutionary Dynamics and Cross-generational Quality Leap Analysis

Figure 9 provides in-depth insights into the stability and final quality gains of the Two-stage Decoupled Fine-tuning strategy during dynamic optimization. As shown in 9(a), it records the evolution curves of the upper-level objective (FID) and diversity metric (Inception Score, IS) during Stage 2 Anchored



(a) Semantic Spacing and Latent PCA Correction. The long red arrows represent correction vectors of tail classes, and blue point clusters denote the distribution of head classes from Stage 1 to Stage 2.

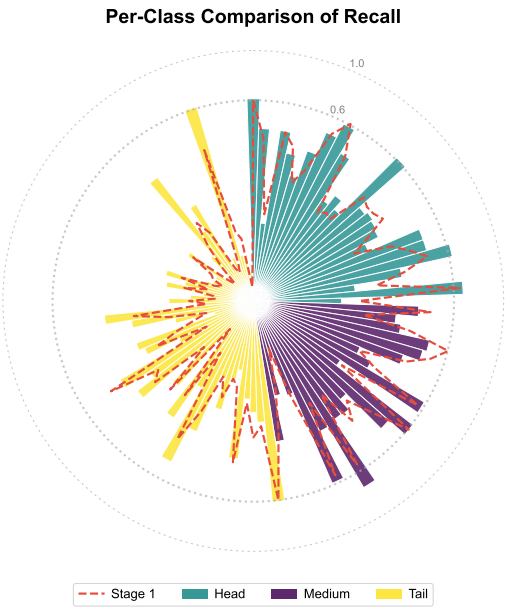


(b) Recall Evolution and Distributional Integrity of Tail Classes. Recall scores of tail classes across training epochs during fine-tuning, demonstrating monotonic growth without mode collapse.

Figure 7: Evolutionary Dynamics & Quality Leap (SPR-induced Hypergradient Stability to Quality Jump). Left : SPR provides hypergradient stability, leading to monotonic convergence with reduced variance during Stage 2 fine-tuning; Right: This stability translates to a collective "generational leap" in model quality. Blue dots from Stage 2 shift right from red dots from Stage 1 for almost all classes, showing that SPR prevents Calibration Drift and achieves precise semantic manifold alignment without destroying the original distribution structure.



(a) KDE, Boxplot & Scatter Plot



(b) Annular Recall Plot

Figure 8: Scalability to Large-scale Label Space on CIFAR-100-LT. Left: Recall distribution shifts right with reduced extreme low scores; Right: Tail class performance is preserved in few-shot regions.

Bilevel Optimization. Notably, FID exhibits an exceptionally smooth and monotonic convergence behavior.

In traditional generative Bilevel Optimization (BLO), the high variance of dynamic sampling evaluations at each iteration typically causes the upper-level objective to fluctuate, trapping the optimizer in

unstable local minima and eroding the high-quality foundation established in Stage 1. However, the stability observed in 9(a) is a direct manifestation of the hyper-gradient consistency argued in the main text. By introducing Static Pool Resampling (SPR), we constrain the originally unconstrained, noise-ridden hyper-gradient estimates onto a stable manifold supported by the anchor set generated from Stage 1 calibration

This dynamic stability provided by SPR directly gives rise to the cross-generational quality leap illustrated in 9(b) and 9(c). The per-class KID scatter plots clearly depict the performance jump across all classes from Stage 1 (calibrated manifold) to Stage 2 (anchored bilevel optimization). All classes, especially the extremely scarce tail classes C8 and C9, achieve a collective rightward shift, indicating reduced distribution discrepancy.

This phenomenon dissects the underlying logic of our method to mitigate Calibration Drift: as a Non-parametric Prior, SPR acts as a "Manifold Buffer" during each hyper-gradient update $\nabla_\lambda F$, forcibly anchoring the fine-tuning direction to the "Gold Standard" long-tail distribution. This mechanism ensures that task alignment is achieved through refined path search while preserving Distributional Integrity, rather than sacrificing minority-class semantics for local gains.

Beyond standard fidelity and stability metrics, we further examine the cross-generational integrity of our model. Compared with the FID convergence shown in Figure 9(a), the recall progression in Figure 7(b) reveals a synergetic relationship: as the model aligns with the task objective, it simultaneously expands its coverage of tail-class semantics.

A common criticism of generative optimization is the potential sacrifice of sampling diversity for lower FID scores, which often leads to mode collapse. To address this concern, Figure 7(b) records the recall evolution curve of tail classes during the fine-tuning process, where the horizontal axis represents training epochs and the vertical axis represents recall scores.

Experimental evidence strongly refutes this conjecture: as bilevel optimization progresses, the Recall of tail classes shows a significant **"monotonic growth"** trend. This indicates that our method actually expands the coverage of generated samples and explores more diverse semantic patterns while improving image quality. This "win-win" situation of quality and diversity benefits from the stable gradient flow across the entire manifold provided by SPR, which prevents the model from falling into narrow plateaus of head classes during optimization and instead guides it to explore broader tail manifolds.

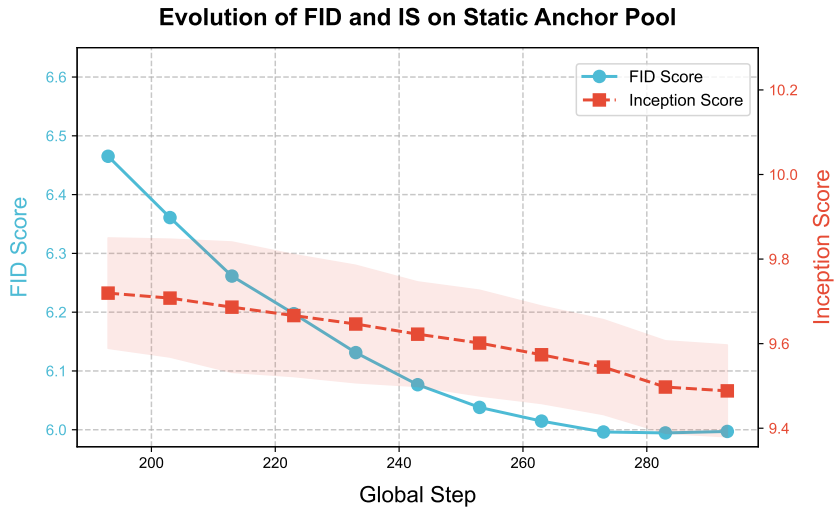
From the perspective of multimedia generation, the protection of **"Distributional Integrity"** is of high value. It means that our model can not only synthesize high-fidelity single images but also restore the original diversity landscape of the entire long-tailed dataset. This robust generative capability is the core competitiveness of using it as a data augmentation tool or a generative base model.

A.4 Sampling Robustness and Semantic Certainty Analysis

Figure 10 provides an in-depth analysis of the anchoring effect of the Two-stage Decoupled Fine-tuning strategy on the manifold structure of the U-Net diffusion model through cross-sampling distribution experiments and external semantic consistency evaluation. As shown in 10(a), we present the per-class FID comparison of U-Net after re-uniform sampling across the entire space, independent of the specific cached pool used during training. The results indicate that the scatter points of the vast majority of classes are steadily located below the diagonal, which strongly proves that the performance gains obtained by Static Pool Resampling (SPR) are not derived from overfitting to a specific anchor set, but essentially refine the global generative manifold quality of U-Net by stabilizing the hyper-gradient ($\nabla_\lambda F$).

Without SPR constraints, standard bilevel optimization often causes U-Net parameters to experience uncontrollable manifold deviation when pursuing task rewards; however, SPR ensures that we preserve the global statistical distribution characteristics endowed to U-Net by Stage 1 (OC-LT) while achieving Task-specific Alignment. This stability of the manifold is further verified by the semantic evaluation in 10(b). We introduce an independent External Oracle Evaluator pre-trained on the full dataset to detect the semantic certainty of U-Net generated samples.

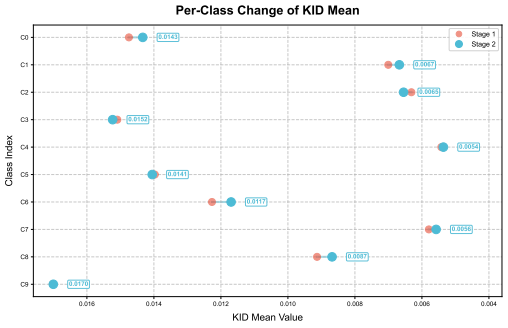
Experiments reveal the deep mechanism of Calibration Drift: in traditional unconstrained optimization, the underlying updates of U-Net gradually erode the fragile tail-semantic boundaries established



(a) Training Trajectory ($\rho = 0.01$). Evolution of FID during Stage 2 fine-tuning, showing monotonic convergence and reduced variance (shaded region). This reflects the hypergradient stability provided by SPR.



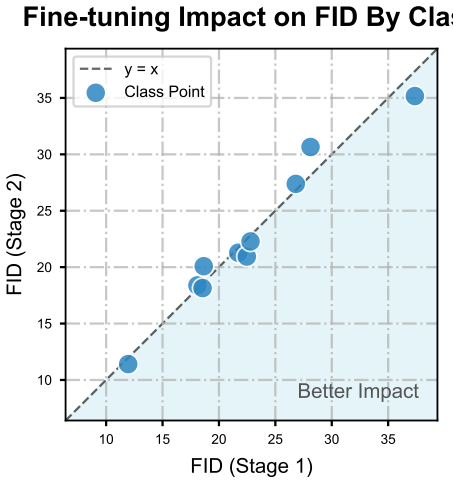
(b) Per-class KID ($\rho = 0.01$). Per-class KID comparison under $\rho = 0.01$ imbalance ratio. Blue dots (Stage 2) are located to the right of red dots (Stage 1) for almost all classes, indicating performance improvement.



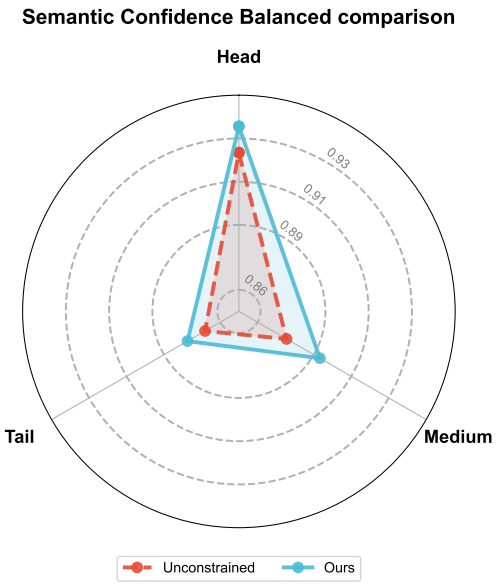
(c) Per-class KID ($\rho = 0.001$). Per-class KID comparison under $\rho = 0.001$ imbalance ratio.Blue dots (Stage 2) are located to the right of red dots (Stage 1) for almost all classes, showing consistent performance improvement across all classes.

Figure 9: Evolutionary Dynamics & Quality Leap (SPR-induced Hypergradient Stability to Quality Jump). Top (Plot 2): SPR provides hypergradient stability, leading to monotonic FID convergence with reduced variance during Stage 2 fine-tuning; Bottom (Plot 3a/3b): This stability translates to a collective "generational leap" in model quality. Blue dots from Stage 2 shift right from red dots from Stage 1 for almost all classes under both $\rho = 0.01$ and $\rho = 0.001$, showing that SPR prevents Calibration Drift and achieves precise semantic manifold alignment without destroying the original distribution structure.

in the Stage 1 Manifold Foundation, resulting in generated images that satisfy the reward objective but lose classification certainty in the eyes of external judges. As clearly depicted in 10(b), our method successfully suppresses this drift. The underlying logic lies in SPR's Non-parametric Prior acting as a "Manifold Buffer" for this stage. In Stage 2 Anchored Bilevel Optimization, SPR forces every optimization step of U-Net to update in reference to the "Gold Standard" distribution established in Stage 1. Through this anchoring mechanism, we effectively filter out gradient noise caused by random sampling, ensuring that the pursuit of alignment objectives does not come at the expense of Distributional Integrity, thereby constructing a solid semantic defense line on the evolutionary trajectory of U-Net.



(a) Sampling Robustness Under Uniform Sampling (Per-class FID Scatter Plot).Per-class FID scatter plot comparing Stage 1 (X-axis) and Stage 2 (Y-axis) under uniform sampling settings. Most points lie below the $y = x$ diagonal, indicating consistent performance improvement across classes.



(b) Sampling Robustness and Semantic Certainty (Oracle Confidence Radar Chart). Radar chart of mean classification confidence scores from a pre-trained Oracle classifier across Head, Medium, and Tail groups. The blue solid line (Stage 2) expands outward compared to the red dashed line (baseline).

Figure 10: Multi-dimensional Reliability Analysis of SPR Strategy. Left (Plot 7): Uniform sampling FID scatter plot: per-class FID of Stage 1 (X-axis) vs. Stage 2 (Y-axis) under uniform sampling, where most points are below the $y = x$ diagonal, proving consistent performance improvement across all classes; Right (Plot 13): Oracle semantic confidence radar chart: mean classification confidence across scores from a pre-trained Oracle classifier across Head, Medium, and Tail groups, showing the blue solid line (Stage 2 with SPR) expands outward in all dimensions compared to the red dashed line (baseline without SPR), demonstrating enhanced semantic certainty and manifold alignment.

B Sampling Robustness and Semantic Certainty Analysis

B.1 First-order Consistency and Unbiasedness of SPR Gradient

Theorem B.1 (Unbiased Estimation of the Anchored Surrogate)**.** *Let* $\mathcal{F}(\phi;\phi_1) = \mathbb{E}_{\epsilon\sim p(\epsilon)}[w_{\phi_1}(\epsilon)R(G_{\theta,\phi}(\epsilon))]$ *be the anchored surrogate objective with* $w_{\phi_1}(\epsilon) = \frac{p_{bal}(G_{\theta,\phi_1}(\epsilon))}{q_{\phi_1}(G_{\theta,\phi_1}(\epsilon))}$*. The SPR estimator* $\hat{g}_{SPR}$ *is an unbiased estimator of* $\nabla_\phi\mathcal{F}(\phi;\phi_1)$*. Furthermore, at the anchor point* $\phi = \phi_1$*, the surrogate gradient is consistent with the true alignment gradient* $g^* = \nabla_\phi\mathbb{E}_{x\sim p_{\theta,\phi}}[R(x)]$*.*

Proof. The SPR algorithm optimizes the model parameters ϕ on a fixed set of seeds $\Xi = \{\epsilon_k\}_{k=1}^K$ using weights w_{ϕ_1} computed from the Stage 1 manifold. The objective $\mathcal{F}(\phi;\phi_1)$ is defined in the constant noise space $\mathcal{E}$ with a fixed density $p(\epsilon)$. The expectation of the estimator is:

$$\begin{aligned}\mathbb{E}_\epsilon[\hat{g}_{\text{SPR}}] &= \mathbb{E}_\epsilon\left[\frac{1}{K}\sum_{k=1}^K w_{\phi_1}(\epsilon_k)\nabla_\phi R(G_{\theta,\phi}(\epsilon_k))\right] \\ &= \nabla_\phi\int_{\mathcal{E}} p(\epsilon)w_{\phi_1}(\epsilon) \\ &\quad\cdot R(G_{\theta,\phi}(\epsilon))d\epsilon,\end{aligned}$$

where the Leibniz integral rule applies under the assumption that $R\circ G$ is continuously differentiable and $\int\|w\nabla R\|p(\epsilon)d\epsilon < \infty$. This proves $\mathbb{E}_\epsilon[\hat{g}_{\text{SPR}}] = \nabla_\phi\mathcal{F}(\phi;\phi_1)$.

To analyze the relationship with the true objective

$$L(\phi) = \mathbb{E}_{x\sim p_{\theta,\phi}}[R(x)]$$

, we evaluate the gradient at the anchor point $\phi = \phi_1$. Using the Law of the Unconscious Statistician and changing variables from ϵ to $x \sim q_{\phi_1}$:

$$\begin{aligned}\nabla_\phi \mathcal{F}(\phi;\phi_1)|_{\phi=\phi_1} &= \int_{\mathcal{X}} q_{\phi_1}(x)\left(\frac{p_{\text{bal}}(x)}{q_{\phi_1}(x)}\right)\nabla_\phi R(x)dx \\ &= \int_{\mathcal{X}} p_{\text{bal}}(x)\nabla_\phi R(x)dx.\end{aligned}$$

According to the score-path gradient identity for diffusion models (or reparameterization trick), the true gradient of $L(\phi)$ at $\phi = \phi_1$ is precisely $\int p_{\text{bal}}(x)\nabla_\phi R(x)dx$.

While a drift $\Delta(\phi) = \|\nabla_\phi \mathcal{F} - \nabla_\phi L\|$ may emerge as ϕ moves away from ϕ_1, the SPR provides an exact first-order consensus at the manifold established by Stage 1. Given that our Decoupled strategy performs local refinement (e.g., via LoRA), this local consistency ensures a valid descent direction toward the balanced distribution. □

B.2 Generalization Bound via Rademacher Complexity in Seed Space

Theorem B.2 (Uniform Convergence with Explicit Constraints)**.** *Assume: (1) Density ratio is bounded* $\sup w_{\phi_1}(\epsilon) \le W$*; (2) Sample gradients are bounded* $\|\nabla_\phi R\| \le C$ *and L-Lipschitz w.r.t.* ϕ*; (3)* Φ *is convex and compact with diameter D. Then for any* $\delta > 0$*, with probability* $\ge 1-\delta$*:*

$$\sup_{\phi\in\Phi} \|\hat{g}_{SPR}(\phi) - g^*(\phi)\| \le 2\mathcal{R}_K(\mathcal{H}) + WC\sqrt{\frac{2\ln(2/\delta)}{K}}$$

Proof. The function class $\mathcal{H} = \{\epsilon \mapsto w_{\phi_1}(\epsilon)\nabla_\phi R(G_{\theta,\phi}(\epsilon))\}$ is defined over the fixed seed space $\mathcal{E}$. By the bounded difference property, replacing one $\epsilon_k \in \Xi$ changes the empirical mean by at most $2WC/K$. Applying McDiarmid's inequality yields the concentration term $WC\sqrt{2\ln(2/\delta)/K}$.

While $\mathcal{H}$ contains vector-valued functions, we utilize the Maurer-vector contraction principle for Rademacher complexity. Given that w_{ϕ_1} is fixed and independent of ϕ, and $\phi \mapsto \nabla_\phi R(G_{\theta,\phi}(\epsilon))$ is L-Lipschitz:

$$\mathcal{R}_K(\mathcal{H}) \le C_{\text{v}} \cdot W \cdot L \cdot \mathcal{R}_K(\Phi),$$

where C_{v} is a constant specific to the vector-valued contraction for d-dimensional gradients.

For a D-bounded parameter space $\Phi \subset \mathbb{R}^d$, the complexity $\mathcal{R}_K(\Phi) \le D\sqrt{d/K}$. Combining these terms, the generalization gap is:

$$\|\hat{g}_{\text{SPR}}(\phi) - g^*(\phi)\| \le \mathcal{O}\left(\frac{WLD\sqrt{d}}{\sqrt{K}}\right) + \mathcal{O}\left(\frac{WC\sqrt{\ln(1/\delta)}}{\sqrt{K}}\right).$$

Here, the density ratio W reflects the "unbalance degree" of the Stage 1 manifold. In long-tailed scenarios (e.g., $\rho = 0.01$), Stage 1 (OC-LT) significantly reduces W compared to the initial pre-trained model, thereby ensuring that the generalization gap vanishes at a rate of $O(1/\sqrt{K})$. With $K = 50,000$, the anchored surrogate surface $\hat{\mathcal{F}}_K$ provides a high-fidelity approximation of the population manifold $F^*(\phi)$, effectively suppressing overfitting in the noise space. □

B.3 Stability and Curvature Concentration Analysis

Theorem B.3 (Reduction of Multi-source Variance and Curvature Concentration)**.** *Let* $f(\phi,\epsilon)$ *be the reward on a latent seed* ϵ*. The SPR-anchored gradient* $\hat{g}_{SPR}$ *achieves a reduction in total variance by eliminating the inter-step functional drift* σ^2_{drift}*. Furthermore, the empirical Hessian* $\nabla^2\hat{F}_K(\phi)$ *converges to the smooth population Hessian* $\nabla^2 F^*(\phi)$ *at a spectral rate of* $O(\sqrt{d/K})$*, effectively suppressing local high-curvature spikes.*

Proof. **Part I: Reduction of Optimization Stochasticity (Relative to Plot 9)**

We distinguish between two sources of randomness: the perturbation direction $\xi \sim \mathcal{U}(\mathbb{S}^{d-1})$ and the generative noise $\epsilon \sim q$. Let $\hat{g}(\phi, \epsilon, \xi)$ be the ZO estimator. In the unconstrained baseline, where ϵ is resampled at each iteration t, the total variance across the optimization path is decomposed as:

$$\text{Var}_{\text{total}} = \underbrace{\mathbb{E}_\epsilon[\text{Var}_\xi(\hat{g}|\epsilon)]}_{\text{Intra-step noise}} + \underbrace{\text{Var}_\epsilon(\mathbb{E}_\xi[\hat{g}|\epsilon])}_{\text{Inter-step drift}}.$$

Here we substitute the standard variance bound for the ZO estimator (using B directions): $\mathbb{E}_\xi[\|\hat{g} - \nabla f\|^2|\epsilon] \leq \mathcal{O}(\frac{d}{B\nu^2}\sigma_f^2)$. In the unconstrained regime, the second term $\text{Var}_\epsilon(\nabla_\phi f(\phi, \epsilon))$ captures the functional jitter caused by switching generative paths at each step.

Subtracting the baseline from our framework, since SPR fixes the anchor pool $\Xi = \{\epsilon_k\}_{k=1}^K$ for all $t \geq 0$, the drift term $\text{Var}_{\epsilon,t}(\cdot)$ vanishes algebraically. The optimizer follows a deterministic surrogate function $\hat{F}_K(\phi)$, where the only variance source is the perturbation directions ξ_t. This explains the high cosine similarity and directional stability observed in Fig. 9.

Part II: Hessian Concentration and Landscape Smoothing (Relative to Plot 12)

The "ruggedness" in the unconstrained landscape (Fig. 12 Left) corresponds to high-frequency local variations in the curvature of individual $f(\phi, \epsilon_k)$. The smoothness of the SPR landscape (Fig. 12 Right) is defined by the spectral stability of the empirical Hessian $\mathbf{H}_K = \frac{1}{K}\sum_{k=1}^K \nabla^2 f(\phi, \epsilon_k)$.

Here we assume that the population reward surface $F^*(\phi) = \mathbb{E}_\epsilon[f(\phi, \epsilon)]$ is L-Lipschitz smooth with a bounded Hessian $\|\nabla^2 F^*(\phi)\|_2 \leq \mu$. This assumption is justified by the Stage 1 Diffusion training, where the score-matching objective implicitly regularizes F^* via Gaussian kernel convolution (the denoising process).

By applying the matrix Bernstein inequality to the sum of independent Hessians $\mathbf{H}_k$, we obtain:

$$\text{Prob}\left(\|\mathbf{H}_K - \mathbf{H}^*\|_2 \geq \tau\right) \leq 2d \exp\left(-\frac{K\tau^2/2}{\bar{\sigma}^2 + M\tau/3}\right),$$

where $\bar{\sigma}^2$ is the variance of the individual curvatures. Solving for τ, the spectral deviation is bounded by:

$$\|\mathbf{H}_K - \mathbf{H}^*\|_2 \leq \bar{\sigma}\sqrt{\frac{2\ln(2d)}{K}} + \frac{M\ln(2d)}{3K}.$$

As K increases to $50,000$, the specific curvature "spikes" of individual samples (which may have high spectral norm M) are suppressed by the $O(1/\sqrt{K})$ factor. The empirical landscape $\hat{F}_K(\phi)$ consequently inherits the high-fidelity smoothness of the population manifold F^*, manifesting as the flat basin observed in Fig. 12.

Taking all parts together, the SPR mechanism achieves a stable optimization path by eliminating functional drift and provides a well-behaved landscape by concentrating the empirical curvature toward the smooth Stage 1 expectation. □