EXPLORE BEFORE YOU OPTIMIZE: ROBUST INITIALIZATION FOR REWARD-GUIDED GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Reward-guided test-time optimization improves generation quality but often reduces diversity by concentrating samples around a limited set of high-reward modes. Recent explore-then-refine methods encourage early exploration before reward-guided refinement, yet this process can be inefficient and the resulting diversity may still degrade during subsequent optimization. We address these limitations by constructing a robust, coverage-oriented initialization population. Building on guidance potential, we introduce a local robustness objective that favors latents located in broad low-potential neighborhoods, where the low-contraction property is more stable under reward-driven perturbations. We further propose a population-level semantic coverage objective that reduces redundancy in the predicted representation space, allowing a limited candidate set to span a broader range of semantic modes. Experiments across multiple benchmarks show that our method achieves competitive generation quality while consistently improving output diversity.

1 INTRODUCTION

Maintaining output diversity while improving generation quality has become a central challenge in modern generative models. Recent studies have explored reward-guided test-time optimization, where generation quality is improved by refining either latent variables Eyring et al. (2024); Tang et al. (2024) or denoising trajectories during inference (Ramesh & Mardani, 2026; Singhal et al., 2025b; Kim et al., 2025b; Zhang et al., 2026) according to external rewards, such as human preference scores (Wu et al., 2023b). However, continuously optimizing toward high-reward regions may gradually contract the generation distribution around a limited number of preferred modes, resulting in reduced diversity and a tendency toward mode collapse (Zhang et al., 2024; Jena et al., 2025). To alleviate this issue, existing approaches commonly incorporate diversity control into reward-guided optimization through mechanisms such as particle repulsion (Corso et al., 2023; Zameshina et al., 2023), diversity-aware sampling (Ramesh & Mardani, 2026; Guimaraes & Perona, 2026a; Singhal et al., 2025b), and reward-diversity balancing (Sadat et al., 2024; Kim et al., 2025b). In these methods, diversity preservation remains coupled with reward-based selection or refinement, making early diversity exploration susceptible to unreliable reward signals.

Recent studies have increasingly adopted an explore-then-refine strategy, emphasizing early exploration before reward-guided refinement (Chae et al., 2025; Kasaei et al., 2026). This design aligns with the temporal structure of diffusion sampling: early denoising steps largely determine the global semantic structure of generated samples, providing a critical window for mode exploration (Savani et al., 2026), while reward estimates remain relatively unreliable at this stage (Kang et al., 2026; Yan et al., 2026). However, such exploration mechanisms still face two key challenges: (1) exploration is inefficient, often requiring many independent random samples to identify promising candidates for subsequent optimization; and (2) the resulting diversity is fragile, since distances among initial latent variables do not necessarily translate into semantic diversity in the final generated images, and any diversity established at initialization may further diminish during subsequent reward optimization.

This leads to our central question: how should early exploration be optimized to remain efficient while making the established diversity robust to subsequent reward-driven updates? Recent work offers a geometric perspective through guidance potential(Jeon et al., 2025; Li et al., 2026), showing that high-potential regions induce stronger contraction in the generative flow and tend to drive

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.

distinct trajectories toward similar dominant modes, whereas low-potential regions better preserve trajectory separation. Motivated by this geometry, the guidance-potential posterior reweights the standard prior toward low-potential regions, providing a more efficient way to explore diverse generation modes.

However, under reward-driven test-time optimization, reward gradients may move low-potential candidates into neighboring high-potential regions, weakening their low-contraction property and diversity advantage. To address this issue, a local robustness objective is introduced to extend guidance potential to the local neighborhood, favoring latents supported by broader low-potential regions and improving robustness to subsequent reward-driven updates. Building on this efficient exploration, we further optimize population-level semantic coverage under a finite candidate budget. A favorable initialization distribution promotes mode discovery, but finite sampling can still yield semantically redundant candidates. A population-level semantic coverage mechanism is introduced to reduce redundancy in the predicted representation space, enabling broader coverage under a limited candidate budget. Together, local robustness and population-level semantic coverage provide an efficient way to construct an initialization set tailored for subsequent reward optimization.

Our contributions are summarized as follows:

- We reveal that efficient initialization-based diversity exploration alone is insufficient for reward-guided generation, as the diversity established at initialization may degrade under subsequent reward-driven optimization.
- Based on guidance potential, we propose a local robustness objective that analyzes neighborhood stability and selects latent regions with stable low guidance potential to preserve diversity under subsequent reward-driven updates.
- A population-level semantic coverage objective is proposed to reduce candidate redundancy and improve mode coverage under a limited sampling budget, achieving competitive performance across multiple benchmarks.

2 Related Work

2.1 Training-Free Reward-Guided Optimization

Training-free alignment of pretrained diffusion models can be achieved through reward optimization at inference time (Eyring et al., 2024; Tang et al., 2024; Kim et al., 2025b). A common strategy leverages Tweedie's formula(Efron, 2011; Chung et al., 2022) to estimate clean samples from noisy states and differentiates reward functions through these estimates to guide denoising trajectorie. However, reward guidance based on highly noisy intermediate states can be inaccurate, making early-stage optimization less reliable (Yan et al., 2026). An alternative line formulates reward alignment as a sampling problem, where the generation process is viewed as sampling from a reward-modified target distribution. Based on SMC, these methods progressively reweight and resample samples to approximate this distribution, thereby allocating more sampling computation to promising regions (Trippe et al., 2022; Wu et al., 2023a; Dou & Song, 2024; Kim et al., 2025b). However, maintaining and propagating multiple particles introduces additional inference cost. Overall, these approaches primarily target reward alignment or sampling efficiency, while population-level generation diversity is typically not optimized as an explicit objective.

2.2 Inference-Time Scaling and Exploration

Recent inference-time scaling methods(Li et al., 2025; Ramesh & Mardani, 2026; Zhang et al., 2026) improve generation by allocating additional computation across multiple candidate trajectories, with Best-of-N sampling serving as a simple baseline. More adaptive strategies vary the amount of computation allocated during inference. For example, adaptive-N(Guimaraes & Perona, 2026b) particle filters dynamically adjust the number of particles based on online predictive statistics. PSP(Elvira et al., 2021) applies a similar adaptive-allocation principle to diffusion inference, front-loading exploration with a larger initial seed pool and progressively pruning less promising trajectories under a fixed inference budget. Another adaptive sampler also allocates inference computation non-uniformly over time, but follows the opposite temporal strategy by conserving computation at early stages and increasing resampling later(Kim et al., 2026b).

Although these methods improve inference-time exploration, candidate retention remains largely reward-driven. DivGen(Parmar et al., 2026) jointly considers quality and diversity by combining unary quality scores with pairwise diversity terms to select a high-quality and diverse subset from a larger candidate pool. However, coupling diversity selection with reward-based quality evaluation also makes the process sensitive to inaccurate reward estimates at highly noisy stages. In contrast, our work decouples early diversity exploration from reward evaluation by constructing an initial population with broad semantic coverage and robustness to subsequent reward-guided updates.

3 PRELIMINARIES

Diffusion models learn powerful generative priors from large-scale data, but their pretrained distributions may not fully align with downstream objectives or human preferences. Reward-guided generation addresses this mismatch by adapting the sampling distribution toward regions favored by a task-specific reward function. A common formulation optimizes a trade-off between reward maximization and deviation from the pretrained distribution(Gao et al., 2023; Uehara et al., 2024):

$$p_{\mathrm{tar}} = \arg\max_{p} \left\{ \mathbb{E}_{\mathbf{x}\sim p}[R(\mathbf{x})] - \alpha D_{\mathrm{KL}}(p \| p_{\mathrm{pre}}) \right\}, \tag{1}$$

where p_{pre} denotes the pretrained diffusion distribution and α controls the strength of prior preservation. Although this objective improves alignment with the reward model, it can shift probability mass toward a limited subset of high-reward regions. Such mode-seeking behavior may reduce the coverage of alternative semantic modes, leading to insufficient diversity in the generated population. Therefore, recent test-time optimization methods incorporate explicit diversity objectives to balance reward improvement and population coverage(Parmar et al., 2026):

$$p_{\mathrm{tar}} = \arg\max_{p} \left\{ \mathbb{E}_{\mathbf{x}\sim p}[R(\mathbf{x})] - \alpha D_{\mathrm{KL}}(p \| p_{\mathrm{pre}}) + \lambda \mathcal{D}(p) \right\}, \tag{2}$$

where $\mathcal{D}(p)$ denotes a diversity objective. Existing studies suggest a fundamental limitation of current diversity control strategies: diversity is either optimized jointly with reward guidance, making its establishment during early sampling sensitive to inaccurate reward estimates, or constructed before reward optimization without explicitly accounting for whether the resulting initialization diversity can translate into and persist as semantic diversity in the final outputs.

In this work, we study the initialization problem from the perspective of preserving diversity throughout reward optimization. Specifically, we identify two properties required for such initialization: *local robustness* and *semantic coverage*. Local robustness characterizes whether the diversity-preserving property of an initialization remains stable under reward-induced latent perturbations, while semantic coverage measures whether the initialization set spans diverse and non-redundant semantic regions.

3.1 LOCALLY ROBUST INITIALIZATION FOR DIVERSITY PRESERVATION

Why initialization diversity can vanish. Diversity in the initial noise space does not necessarily translate into semantic diversity in the generated outputs. Let $\Psi_c(z) = h(\mathcal{G}_c(z))$ denote the semantic representation of the sample generated from latent z under condition c. For a local perturbation Δ,

$$\|\Psi_c(z+\Delta) - \Psi_c(z)\|_2^2 = \Delta^\top M_c(z)\Delta + o(\|\Delta\|_2^2), \tag{3}$$

where $M_c(z) = J_{\Psi_c}(z)^\top J_{\Psi_c}(z)$ is the pullback metric that measures how local perturbations in the initial latent space translate into semantic variations after generation. Thus, semantic variation is governed by the local geometry induced by Ψ_c, rather than by Euclidean distance in the initial noise space. In particular, latent differences aligned with directions associated with small eigenvalues of $M_c(z)$ can be strongly attenuated, yielding little semantic variation despite substantial Euclidean separation. This agrees with prior observations that raw diffusion noise lacks a directly meaningful semantic geometry (Preechakul et al., 2022; Park et al., 2023), motivating recent methods to construct diversity in denoised prediction space instead (Kim et al., 2026a).

Recent analyses of diffusion sampling show that regions with higher conditional sharpness exhibit stronger contraction under guided generation, which can cause initially distinct trajectories to concentrate toward dominant modes. To quantify this contraction tendency, Li et al. (2026) introduces

the guidance potential $U_c(z) = \|\widehat{x}_0^c(z) - \widehat{x}_0^\emptyset(z)\|_2^2$ as a tractable proxy for guidance-induced contraction at initialization. However, initialized latent variables continue to evolve through subsequent denoising and reward-guided refinement, making the preservation of this property dependent on the surrounding potential landscape. Specifically, if U_c is locally L-smooth,

$$\sup_{\|\Delta\|_2 \le \rho} U_c(z+\Delta) \ge U_c(z) + \rho\|\nabla U_c(z)\|_2 - \frac{L\rho^2}{2}. \tag{4}$$

This indicates that a seed located in a rapidly varying potential landscape can enter a more contractive region under small downstream updates. Therefore, initialization should favor latent regions where the low-contraction property remains stable throughout subsequent generation dynamics.

Neighborhood-robust potential. Subsequent denoising and reward-guided refinement perturb the initialized latent states. To characterize the sensitivity of the contraction tendency to such local perturbations, we define a neighborhood-robust potential using the entropic risk:

$$\mathcal{R}_{\rho,\lambda}(z) = \frac{1}{\lambda}\log \mathbb{E}_{\delta\sim q_\rho}\left[\exp\left(\lambda \widetilde{U}_c(z+\delta)\right)\right], \tag{5}$$

where $\widetilde{U}_c$ denotes the normalized guidance potential and q_ρ is a local Gaussian perturbation distribution, parameterized as $\delta = \rho\epsilon/\sqrt{d}$ with $\epsilon \sim \mathcal{N}(0, I_d)$ and latent dimension d. The parameter ρ controls the overall perturbation scale, while the exponential aggregation assigns greater weight to perturbations with larger potential. The risk parameter λ determines this sensitivity: as $\lambda \to 0$, $\mathcal{R}_{\rho,\lambda}$ approaches the expected potential under local perturbations, whereas larger values increasingly emphasize the upper tail of the local potential distribution. Minimizing $\mathcal{R}_{\rho,\lambda}(z)$ therefore favors initialization regions in which low guidance potential is preserved under local latent perturbations.

Sampling robust initializations. The neighborhood-robust potential provides an energy function for evaluating the stability of each initialization under local latent evolution. We use it to define a robustness-weighted initialization distribution

$$p_{\mathrm{rob}}(z \mid c) \propto p(z)\exp\left(-\tau\mathcal{R}_{\rho,\lambda}(z)\right), \tag{6}$$

where $p(z)$ is the standard diffusion prior and τ controls the strength of robustness reweighting. This distribution preserves the original Gaussian prior while assigning higher probability to latent regions with stable low-contraction characteristics. Direct sampling from $p_{\mathrm{rob}}(z \mid c)$ is intractable, so we employ Langevin dynamics to iteratively move samples from the Gaussian prior toward the robustness-weighted distribution. Its score is

$$\nabla_z \log p_{\mathrm{rob}}(z \mid c) = -z - \tau\nabla_z\mathcal{R}_{\rho,\lambda}(z), \tag{7}$$

where we use $\nabla_z \log p(z) = -z$ for the standard normal prior. The resulting update at iteration k is

$$z^{(k+1)} = z^{(k)} - \eta\left[z^{(k)} + \tau\nabla_z\mathcal{R}_{\rho,\lambda}\left(z^{(k)}\right)\right] + \sqrt{2\eta}\,\xi^{(k)}, \qquad \xi^{(k)} \sim \mathcal{N}(0, I), \tag{8}$$

where η denotes the Langevin step size. The robust potential term guides the latent toward regions with stable low contraction, the Gaussian prior regularizes its deviation from the original initialization distribution, and the injected noise maintains stochastic exploration during optimization. Further details and analysis of the neighborhood-robust formulation are provided in Appendix A.

3.2 Semantic coverage across robust candidates

Under a limited candidate budget, a desirable initialization should not only favor robust regions, but also promote population-level diversity by spreading candidates across distinct representational directions. For each candidate z, we extract a normalized representation from its predicted clean sample,

$$f_c(z) = \frac{P(z_0^*)}{\|P(z_0^*)\|_2}, \tag{9}$$

where z_0^* denotes the predicted clean sample under condition c, and $P(\cdot)$ is the corresponding feature mapping. To characterize the distribution of candidate samples across semantic representation directions, we aggregate the normalized candidate representations as

$$C_c(\mathcal{Z}) = \frac{1}{M}\sum_{i=1}^{M} f_c(z_i)f_c(z_i)^\top. \tag{10}$$

Algorithm 1 Locally Robust and Coverage-Aware Initialization

Require: condition c; pool size N; output size M; Langevin steps K; $\eta, \rho, \lambda, \tau$
Ensure: initialization population $\mathcal{Z}$

Sample $\{z_i^{(0)}\}_{i=1}^N \sim \mathcal{N}(0, I)$
Stage I: Locally Robust Initialization
for $k = 0, \ldots, K-1$ **do**
  for $i = 1, \ldots, N$ **do**
    Estimate $\mathcal{R}_{\rho,\lambda}(z_i^{(k)})$ using Eq. equation 5
    Update $z_i^{(k+1)}$ according to Eq. equation 8
  end for
end for
$\mathcal{P} \leftarrow \{z_i^{(K)}\}_{i=1}^N$
Stage II: Semantic Coverage
Compute $f_c(z)$ for all $z \in \mathcal{P}$ using Eq. equation 9
$\mathcal{Z} \leftarrow \emptyset$
while $|\mathcal{Z}| < M$ **do**
  $z^\star \leftarrow \arg\max_{z \in \mathcal{P} \setminus \mathcal{Z}} D_{\mathrm{cov}}(\mathcal{Z} \cup \{z\})$ using Eq. equation 11
  $\mathcal{Z} \leftarrow \mathcal{Z} \cup \{z^\star\}$
end while
return $\mathcal{Z}$

Each outer product $f_c(z_i)f_c(z_i)^\top$ captures the directional contribution of an individual candidate, while $C_c(\mathcal{Z})$ summarizes how representation energy is distributed across the population. Since all representations are normalized, the eigenvalues λ_k of $C_c(\mathcal{Z})$ satisfy $\sum_k \lambda_k = \mathrm{tr}(C_c(\mathcal{Z})) = 1$. The spectrum of $C_c(\mathcal{Z})$ characterizes how broadly the candidate population spans the semantic representation space, which motivates the following effective representation dimension:

$$D_{\mathrm{cov}}(\mathcal{Z}) = \frac{1}{\mathrm{tr}(C_c(\mathcal{Z})^2)} = \frac{M}{1 + (M-1)\mathcal{R}_c(\mathcal{Z})}, \tag{11}$$

where

$$\mathcal{R}_c(\mathcal{Z}) = \frac{1}{M(M-1)} \sum_{i \neq j} \langle f_c(z_i), f_c(z_j) \rangle^2 \tag{12}$$

denotes the average representational redundancy of the population. Therefore, maximizing D_{cov} is equivalent to minimizing $\mathcal{R}_c$. Although early predicted clean samples may be insufficient for reliable fine-grained reward evaluation, they can still provide structural information for distinguishing different candidate modes. Further analysis are provided in Appendix B.

4 Experiments

4.1 Experimental Setup

Benchmarks Experiments are conducted on three text-to-image benchmarks: DrawBench(Saharia et al., 2022) (200 prompts), which evaluates generation under diverse and challenging descriptions involving attributes, relationships, and complex concepts; GenEval(Ghosh et al., 2023) (553 prompts), which measures object-level compositional accuracy through object presence, counting, colors, and spatial relations; and EvalMuse(Han et al., 2024) (200 prompts), which assesses fine-grained text–image alignment between generated content and textual instructions.

Baselines Representative baselines from reward-guided optimization, inference-time exploration, and diversity enhancement are considered. Their performance is evaluated in terms of both generation quality and diversity. FK-Steering(Singhal et al., 2025a), RBF(Kim et al., 2025a), and NTS(Ramesh & Mardani, 2025) perform reward-guided refinement of latent variables or sampling trajectories during inference to improve generation quality. PSP(Guimaraes & Perona, 2026b) adopts an explore-then-refine strategy by constructing and pruning candidate populations under a fixed inference budget. DivIn(Li et al., 2026), CADS(Sadat et al., 2024), and DivGen(Parmar et al., 2026) improve generation diversity through diversity-oriented sampling or conditioning mechanisms.

Table 1: Comparison on GenEval. Results are averaged over three runs using public implementations under the same sampling budget (N=4) and ImageReward guidance. We report ImageReward (IR), PickScore, HPSv2, GenEval, and DINOv2. **Bold** and underlined values indicate the best and second-best results, respectively.

	SDXL					SD3.5				
Method	**IR ↑**	**PickScore ↑**	**HPSv2 ↑**	**GenEval ↑**	**DINOv2 ↑**	**IR ↑**	**PickScore ↑**	**HPSv2 ↑**	**GenEval ↑**	**DINOv2 ↑**
BoN	1.101	22.482	0.289	0.634	0.589	1.334	23.176	0.283	0.748	0.613
FKS	1.187	21.851	0.296	0.626	0.560	1.291	22.421	0.305	0.741	0.585
RBF	1.135	22.287	0.301	0.619	0.563	1.255	**23.243**	0.297	0.739	0.592
NTS	0.969	21.574	0.297	0.581	0.572	1.083	21.945	0.294	0.698	0.607
PSP	**1.226**	**22.641**	0.293	0.646	0.607	**1.386**	23.049	0.309	0.746	0.626
DivIn	1.178	21.921	0.291	0.643	0.739	1.325	22.544	0.300	0.754	0.692
CADS	1.152	22.143	0.295	0.647	0.706	1.303	22.675	0.302	0.757	0.709
DivGen	1.144	21.731	0.290	0.638	**0.786**	1.309	22.224	0.299	0.753	0.738
Ours	1.209	22.608	**0.307**	**0.682**	0.784	1.354	22.752	**0.314**	**0.781**	**0.755**

Table 2: Comparison on EvalMuse. Results are averaged over three runs using public implementations under the same sampling budget. We report the corresponding generation quality and diversity metrics. **Bold** and underlined values indicate the best and second-best results, respectively.

Method	SDXL					SD3.5				
	IR	**PickScore**	**HPS**	**EvalMuse**	**DINOv2**	**IR**	**PickScore**	**HPS**	**EvalMuse**	**DINOv2**
BoN	1.098	22.65	0.290	3.014	0.588	1.336	23.25	0.304	3.218	0.562
FKS	1.189	22.42	0.284	3.003	0.568	1.294	23.08	0.284	3.207	0.548
RBF	**1.235**	22.38	0.302	2.994	0.574	1.253	23.15	0.296	3.211	0.555
NTS	1.157	22.30	0.298	2.972	0.596	1.086	23.00	0.295	3.168	0.574
PSP	1.224	22.55	0.294	3.109	0.607	1.380	23.31	0.306	3.237	0.587
DivIn	1.075	22.46	0.296	3.015	0.645	1.305	23.24	0.303	3.224	0.629
CADS	1.010	22.24	0.289	2.978	0.641	1.215	23.04	0.293	3.186	0.618
DivGen	1.090	22.40	0.293	3.012	0.639	1.278	23.18	0.300	3.205	**0.633**
Ours	1.133	**22.78**	**0.306**	**3.150**	**0.648**	**1.445**	**23.47**	**0.313**	**3.354**	0.631

Evaluation Metrics Evaluation is conducted from two main perspectives: generation quality and sample diversity. The former considers both preference alignment and text–image consistency, while the latter measures semantic variation across generated samples. HPSv2(Wu et al., 2023b), ImageReward(Xu et al., 2023), and PickScore(Kirstain et al., 2023) are used to evaluate preference-aligned generation quality. GenEval(Ghosh et al., 2023) measures compositional correctness, while EvalMuse(Han et al., 2024) assesses fine-grained text–image alignment. Generation diversity is evaluated using the average pairwise DINOv2(Oquab et al., 2024) similarity between generated samples.

4.2 Main Results

Across all three benchmarks, our method achieves competitive reward scores while maintaining broader output diversity, as shown in Tables 1 and 2 and Figure 1. Compared with reward-driven optimization methods, it substantially alleviates the diversity degradation caused by reward optimization while preserving comparable generation quality. Compared with diversity-focused methods, our approach achieves a better balance between diversity preservation and reward alignment, maintaining broad semantic coverage without sacrificing generation quality. Additional qualitative results are provided in Figures 2 and 3.

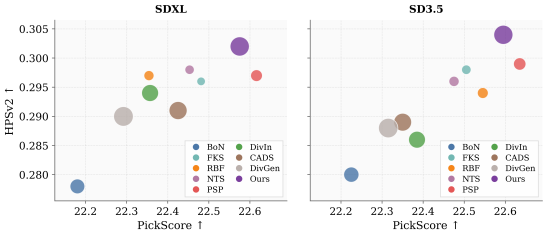


Figure 1: Quality-diversity comparison on DrawBench. Marker size indicates output diversity, with larger markers corresponding to higher diversity.



Figure 2: Qualitative results on EvalMuse with SDXL demonstrate that our strategy produces samples with higher visual quality.

The experimental results further highlight the importance of establishing diversity at initialization, reflecting a different perspective from prior approaches that optimize diversity jointly with reward. At early denoising stages, reward estimates derived from highly noisy samples can be unreliable; coupling diversity control with such signals may therefore bias exploration before meaningful generation modes are fully established. Our method instead decouples early diversity construction from reward optimization, using the initialization stage to form a robust and semantically well-covered candidate population before reward-guided refinement. From this perspective, diversity is not merely a regularizer that counteracts reward-induced mode contraction, but an exploration foundation that determines which generation modes remain accessible to subsequent optimization.

4.3 Computational Overhead

We further evaluate the computational overhead introduced by our initialization procedure. Specifically, the first stage initializes twice the target sampling budget, resulting in eight candidate latents when the final budget is $N = 4$, after which four candidates are retained for subsequent generation. Since the proposed objectives are applied only during this initialization stage, the subsequent reward-guided generation follows the same optimization and denoising procedure as the corresponding baseline. We report the total generation FLOPs under the same sampling budget. The additional cost mainly comes from the local-robustness evaluation and population-level semantic coverage optimization over the initial candidate pool. As shown in Figure 4, our method introduces only moderate additional computation compared with standard reward-guided optimization, while consistently achieving a better trade-off between generation quality and output diversity. Compared with diversity-oriented methods, the diversity established during initialization is less susceptible to degradation under subsequent sampling updates, thereby preserving higher output diversity while leaving more room for further improvements in generation quality.



Figure 3: Qualitative results on GenEval. Compared with BoN, our method generates samples with greater output diversity while maintaining high visual quality.

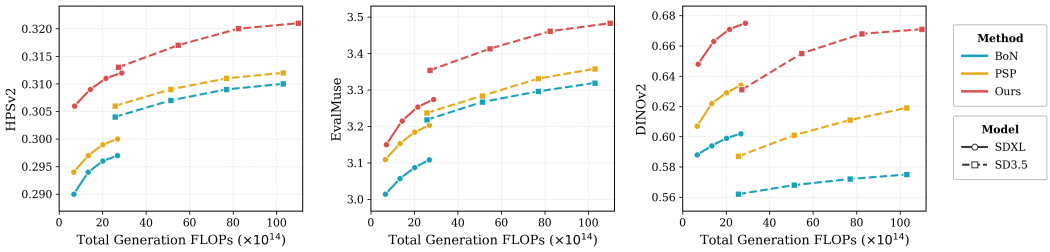


Figure 4: Scaling behavior of generation quality and diversity with generation FLOPs.

Table 3: Ablation study of local robustness(Stage 1) and population-level semantic coverage(Stage 2) on SDXL and SD3.5.

Stage 1	Stage 2	SDXL			SD3.5		
		HPS ↑	EvalMuse ↑	DINOv2 ↑	HPS ↑	EvalMuse ↑	DINOv2 ↑
×	×	0.290	3.014	0.588	0.304	3.218	0.562
×	✓	0.294	3.079	0.627	0.306	3.273	0.609
✓	×	0.301	3.132	0.635	0.309	3.316	0.614
✓	✓	**0.306**	**3.150**	**0.648**	**0.313**	**3.354**	**0.631**

4.4 Ablation Study

The two components of the initialization objective, local robustness and population-level semantic coverage, are evaluated separately on SDXL and SD3.5, as shown in Table 3. Introducing either component consistently improves all metrics over the baseline, demonstrating the effectiveness of both objectives. Local robustness yields a particularly clear improvement when applied alone, suggesting that constructing initialization regions that are less sensitive to subsequent sampling and reward-driven updates is important for preserving the benefits of early exploration. Adding semantic coverage further improves the results by reducing redundancy across the initialized population and encouraging broader exploration. Combining both components achieves the strongest performance across all metrics on both backbones, indicating that robust mode preservation and population-level coverage provide complementary benefits for quality and diversity.

4.5 Orthogonality to Reward Optimization

To assess whether the proposed initialization is complementary to downstream reward optimization, it is combined with PSP by replacing its original initialization while keeping the subsequent optimization procedure and inference budget unchanged. As shown in Table 4, Ours Init + PSP consistently improves both reward and diversity over the original PSP on SDXL and SD3.5. Compared with the proposed initialization alone, incorporating PSP further improves HPS and EvalMuse scores, while retaining most of the diversity gain established at initialization. These results indicate that the proposed initialization is not tied to a specific reward optimization procedure and can provide a stronger starting population for subsequent reward-guided refinement.

4.6 Hyperparameter Analysis

To assess the sensitivity of the proposed initialization strategy to its key hyperparameters, ρ, λ, and τ are varied independently around the default configuration $(0.2, 0.5, 0.5)$, with the results summarized in Table 5. The overall trends are consistent across SDXL and SD3.5. Increasing ρ improves HPS and EvalMuse up to $\rho = 0.2$, whereas DINOv2 attains its maximum at $\rho = 0.1$. A larger neighborhood radius leads to a consistent decline across all metrics, indicating that excessively broad perturbation neighborhoods provide a less discriminative estimate of local robustness. For λ, the highest HPS and EvalMuse are achieved at $\lambda = 0.5$, while $\lambda = 1.0$ yields the strongest DINOv2 performance, reflecting the expected trade-off induced by increased sensitivity to high-potential perturbations. The parameter τ exhibits a clearer concentration effect: HPS and EvalMuse

Table 4: Orthogonality to reward-guided optimization on EvalMuse. We replace the original initialization of PSP with our coverage-oriented initialization while keeping the reward optimization procedure and inference budget unchanged.

	SDXL			SD3.5		
Method	**HPS ↑**	**EvalMuse ↑**	**DINOv2 ↑**	**HPS ↑**	**EvalMuse ↑**	**DINOv2 ↑**
PSP	0.294	3.109	0.607	0.306	3.237	0.587
Ours	0.306	3.150	**0.648**	0.313	3.354	**0.631**
Ours Init + PSP	**0.308**	**3.154**	0.632	**0.317**	**3.359**	0.625

Table 5: Hyperparameter sensitivity analysis of the proposed robust initialization on SDXL and SD3.5. Unless otherwise specified, we use $\rho = 0.2$, $\lambda = 0.5$, and $\tau = 0.5$.

Hyperparameter	**Value**	SDXL			SD3.5		
		HPS ↑	**EvalMuse ↑**	**DINOv2 ↑**	**HPS ↑**	**EvalMuse ↑**	**DINOv2 ↑**
ρ	0.05	0.302	3.119	0.645	0.308	3.316	0.628
	0.10	0.304	3.138	**0.651**	0.311	3.338	**0.635**
	0.20	**0.306**	**3.150**	0.648	**0.313**	**3.354**	0.631
	0.40	0.305	3.143	0.640	0.312	3.345	0.622
λ	0.25	0.304	3.138	0.642	0.311	3.339	0.625
	0.50	**0.306**	**3.150**	0.648	**0.313**	**3.354**	0.631
	1.00	0.305	3.146	**0.653**	0.312	3.349	**0.636**
	2.00	0.303	3.132	0.649	0.310	3.334	0.633
τ	0.25	0.304	3.136	**0.654**	0.311	3.337	**0.638**
	0.50	**0.306**	**3.150**	0.648	**0.313**	**3.354**	0.631
	1.00	0.305	3.145	0.642	0.312	3.348	0.625
	2.00	0.302	3.124	0.634	0.309	3.326	0.617

are maximized at $\tau = 0.5$, whereas DINOv2 decreases monotonically as τ increases, consistent with stronger reweighting toward low-risk regions narrowing the effective support of the initialization distribution. Overall, the default configuration yields the most favorable balance between reward-aligned quality and output diversity across both backbone models.

The remaining hyperparameters are fixed throughout all experiments, with the step size set to $\eta = 0.1$, the number of neighborhood samples set to 3, and the number of initialization update steps set to 3. These settings are selected to balance estimation reliability and computational efficiency, as further increasing the number of neighborhood evaluations or optimization steps introduces additional computational overhead while yielding only limited empirical gains.

5 Conclusion

We study diversity preservation in reward-guided test-time optimization from the perspective of initialization. Rather than treating diversity as an auxiliary objective during reward optimization, we focus on constructing an initial candidate population whose diversity is both efficient to establish and robust to subsequent reward-driven updates. To this end, we introduce a local robustness objective that favors broad low-potential regions and a population-level semantic coverage objective that reduces redundancy under a limited sampling budget. Experiments across multiple benchmarks and generative backbones show that the proposed method consistently improves output diversity while maintaining competitive generation quality, and remains complementary to existing reward-based optimization strategies. These results highlight initialization as an important design dimension for balancing exploration and reward alignment in test-time generative optimization.

AI USE STATEMENT

In this work, we used generative AI tools to assist with broad literature exploration and the identification of potentially relevant prior work, as well as to cross-check the correctness of selected mathematical derivations in the development of our method. All literature identified with AI assistance was independently verified against the original sources, and all AI-assisted derivations were manually reviewed and validated by the authors.

REFERENCES

Daewon Chae, June Suk Choi, Jinkyu Kim, and Kimin Lee. Diffexp: Efficient exploration in reward fine-tuning for text-to-image diffusion models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 15696–15703, 2025.

Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687*, 2022.

Gabriele Corso, Yilun Xu, Valentin de Bortoli, Regina Barzilay, and Tommi Jaakkola. Particle guidance: non-i.i.d. diverse sampling with diffusion models. *arXiv*, 2023.

Zehao Dou and Yang Song. Diffusion posterior sampling for linear inverse problem solving: A filtering perspective. In *International Conference on Learning Representations*, volume 2024, pp. 12069–12095, 2024.

Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011.

Víctor Elvira, Joaquín Miguez, and Petar M Djurić. On the performance of particle filters with adaptive number of particles. *Statistics and Computing*, 31(6):81, 2021.

Luca Eyring, Shyamgopal Karthik, Karsten Roth, Alexey Dosovitskiy, and Zeynep Akata. Reno: Enhancing one-step text-to-image models through reward-based noise optimization. *Neural Information Processing Systems (NeurIPS)*, 2024.

Leo Gao, John Schulman, and Jacob Hilton. Scaling laws for reward model overoptimization. In *International conference on machine learning*, pp. 10835–10866. PMLR, 2023.

Dhruba Ghosh, Hanna Hajishirzi, and Ludwig Schmidt. Geneval: An object-focused framework for evaluating text-to-image alignment. *ArXiv*, abs/2310.11513, 2023. URL https://api.semanticscholar.org/CorpusID:264288728.

Rogerio Guimaraes and Pietro Perona. Inference-time scaling of diffusion models via progressive seed pruning. *arXiv preprint arXiv:2607.21591*, 2026a.

Rogerio Guimaraes and Pietro Perona. Inference-time scaling of diffusion models via progressive seed pruning. *arXiv preprint arXiv:2607.21591*, 2026b.

Shuhao Han, Haotian Fan, Jiachen Fu, Liang Li, Tao Li, Junhui Cui, Yunqiu Wang, Yang Tai, Jingwei Sun, Chunle Guo, and Chongyi Li. Evalmuse-40k: A reliable and fine-grained benchmark with comprehensive human annotations for text-to-image generation model evaluation, 2024. URL https://arxiv.org/abs/2412.18150.

Rohit Jena, Ali Taghibakhshi, Sahil Jain, Gerald Shen, Nima Tajbakhsh, and Arash Vahdat. Elucidating optimal reward-diversity tradeoffs in text-to-image diffusion models. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 232–242. IEEE, 2025.

Dongjae Jeon, Dueun Kim, and Albert No. Understanding and mitigating memorization in generative models via sharpness of probability landscapes. In *Forty-second International Conference on Machine Learning*, 2025.

Hyeongyu Kang, Jaewoo Lee, Woocheol Shin, Kiyoung Om, and Jinkyoo Park. Diffusion finetuning via reparameterized policy gradient of the soft q-function. In *International Conference on Learning Representations*, volume 2026, pp. 78690–78725, 2026.

Seyed Amir Kasaei, Ali Aghayari, Arash Marioriyad, Niki Sepasian, Shayan Baghayi Nejad, MohammadAmin Fazli, Mahdieh Soleymani Baghshah, and Mohammad Hossein Rohban. CARINOX: Inference-time scaling with category-aware reward-based initial noise optimization and exploration. *Transactions on Machine Learning Research*, 2026. ISSN 2835-8856. URL https://openreview.net/forum?id=XB1cwXHV0c.

Byungjun Kim, Soobin Um, and Jong Chul Ye. Diverse text-to-image generation via contrastive noise optimization. In *The Fourteenth International Conference on Learning Representations*, 2026a. URL https://openreview.net/forum?id=EVRMnAREc3.

Jaihoon Kim, TaeHoon Yoon, Jisung Hwang, and Minhyuk Sung. Inference-time scaling for flow models via stochastic generation and rollover budget forcing. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025a. URL https://openreview.net/forum?id=quY3zRPalR.

Jaihoon Kim, Taehoon Yoon, Jisung Hwang, and Minhyuk Sung. Inference-time scaling for flow models via stochastic generation and rollover budget forcing. *Advances in Neural Information Processing Systems*, 38:30830–30864, 2026b.

Sunwoo Kim, Minkyu Kim, and Dongmin Park. Test-time alignment of diffusion models without reward over-optimization. In *The Thirteenth International Conference on Learning Representations*, 2025b. URL https://openreview.net/forum?id=vi3DjUhFVm.

Yuval Kirstain, Adam Polyak, Uriel Singer, Shahbuland Matiana, Joe Penna, and Omer Levy. Pick-a-pic: An open dataset of user preferences for text-to-image generation. 2023.

Xiang Li, Dianbo Liu, and Kenji Kawaguchi. Initialization is half the battle: Generating diverse images from a guidance potential posterior. *arXiv preprint arXiv:2606.02453*, 2026.

Xiner Li, Masatoshi Uehara, Xingyu Su, Gabriele Scalia, Tommaso Biancalani, Aviv Regev, Sergey Levine, and Shuiwang Ji. Dynamic search for inference-time alignment in diffusion models. *arXiv preprint arXiv:2503.02039*, 2025.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel HAZIZA, Francisco Massa, Alaaeldin El-Nouby, Mido Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Herve Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=a68SUt6zFt. Featured Certification.

Yong-Hyun Park, Mingi Kwon, Jaewoong Choi, Junghyo Jo, and Youngjung Uh. Understanding the latent space of diffusion models through the lens of riemannian geometry. *Advances in Neural Information Processing Systems*, 36:24129–24142, 2023.

Gaurav Parmar, Or Patashnik, Daniil Ostashev, Kuan-Chieh Wang, Kfir Aberman, Srinivasa Narasimhan, and Jun-Yan Zhu. Scaling group inference for diverse and high-quality generation. In *International Conference on Learning Representations*, volume 2026, pp. 89763–89790, 2026.

Konpat Preechakul, Nattanat Chatthee, Suttisak Wizadwongsa, and Supasorn Suwajanakorn. Diffusion autoencoders: Toward a meaningful and decodable representation. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10609–10619. IEEE, 2022.

Vignav Ramesh and Morteza Mardani. Test-time scaling of diffusion models via noise trajectory search. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=fnkkY7UGRl.

Vignav Ramesh and Morteza Mardani. Test-time scaling of diffusion models via noise trajectory search. *Advances in Neural Information Processing Systems*, 38:87284–87317, 2026.

Seyedmorteza Sadat, Jakob Buhmann, Derek Bradley, Otmar Hilliges, and Romann Weber. Cads: Unleashing the diversity of diffusion models through condition-annealed sampling. In *International Conference on Learning Representations*, volume 2024, pp. 23723–23755, 2024.

Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al. Photorealistic text-to-image diffusion models with deep language understanding. *Advances in neural information processing systems*, 35:36479–36494, 2022.

Yash Savani, Branislav Kveton, Yuchen Liu, Yilin Wang, Jing Shi, Subhojyoti Mukherjee, Nikos Vlassis, and Krishna Kumar Singh. Stepwise credit assignment for grpo on flow-matching models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.

Raghav Singhal, Zachary Horvitz, Ryan Teehan, Mengye Ren, Zhou Yu, Kathleen McKeown, and Rajesh Ranganath. A general framework for inference-time scaling and steering of diffusion models. In *Forty-second International Conference on Machine Learning*, 2025a. URL https://openreview.net/forum?id=Jp988ELppQ.

Raghav Singhal, Zachary Horvitz, Ryan Teehan, Mengye Ren, Zhou Yu, Kathleen McKeown, and Rajesh Ranganath. A general framework for inference-time scaling and steering of diffusion models, 2025b. URL https://arxiv.org/abs/2501.06848.

Zhiwei Tang, Jiangweizhi Peng, Jiasheng Tang, Mingyi Hong, Fan Wang, and Tsung-Hui Chang. Inference-time alignment of diffusion models with direct noise optimization. *arXiv preprint arXiv:2405.18881*, 2024.

Brian L Trippe, Jason Yim, Doug Tischer, David Baker, Tamara Broderick, Regina Barzilay, and Tommi Jaakkola. Diffusion probabilistic modeling of protein backbones in 3d for the motif-scaffolding problem. *arXiv preprint arXiv:2206.04119*, 2022.

Masatoshi Uehara, Yulai Zhao, Kevin Black, Ehsan Hajiramezanali, Gabriele Scalia, Nathaniel Lee Diamant, Alex M Tseng, Tommaso Biancalani, and Sergey Levine. Fine-tuning of continuous-time diffusion models as entropy-regularized control. *arXiv preprint arXiv:2402.15194*, 2024.

Luhuan Wu, Brian Trippe, Christian Naesseth, David Blei, and John P Cunningham. Practical and asymptotically exact conditional sampling in diffusion models. *Advances in Neural Information Processing Systems*, 36:31372–31403, 2023a.

Xiaoshi Wu, Yiming Hao, Keqiang Sun, Yixiong Chen, Feng Zhu, Rui Zhao, and Hongsheng Li. Human preference score v2: A solid benchmark for evaluating human preferences of text-to-image synthesis. *arXiv preprint arXiv:2306.09341*, 2023b.

Jiazheng Xu, Xiao Liu, Yuchen Wu, Yuxuan Tong, Qinkai Li, Ming Ding, Jie Tang, and Yuxiao Dong. Imagereward: learning and evaluating human preferences for text-to-image generation. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, pp. 15903–15935, 2023.

Renye Yan, Jikang Cheng, Shikun Sun, Yi Sun, You Wu, Wei Peng, Zongwei Wang, Ling Liang, Junliang Xing, and Yimao Cai. Do less, achieve more: Do we need every-step optimization for rl fine-tuning of diffusion models? *arXiv preprint arXiv:2605.15855*, 2026.

Mariia Zameshina, Olivier Teytaud, and Laurent Najman. Diverse diffusion: Enhancing image diversity in text-to-image generation. *arXiv preprint arXiv:2310.12583*, 2023.

Xiangcheng Zhang, Haowei Lin, Haotian Ye, James Y Zou, Jianzhu Ma, Yitao Liang, and Yilun Du. Inference-time scaling of diffusion models through classical search. In *International Conference on Learning Representations*, volume 2026, pp. 39133–39166, 2026.

Ziyi Zhang, Sen Zhang, Yibing Zhan, Yong Luo, Yonggang Wen, and Dacheng Tao. Confronting reward overoptimization for diffusion models: A perspective of inductive and primacy biases. In *Forty-first International Conference on Machine Learning*, 2024.

A NEIGHBORHOOD-ROBUST POTENTIAL

We model plausible downstream latent variation by

$$\delta = \frac{\rho}{\sqrt{d}}\epsilon, \qquad \epsilon \sim \mathcal{N}(0, I_d), \tag{13}$$

where d is the flattened latent dimension. Since $\mathbb{E}\|\delta\|_2^2 = \rho^2$, ρ directly controls the overall perturbation magnitude independently of the latent dimensionality. We denote the resulting local perturbation distribution by q_ρ. To place the guidance potential on a fixed numerical scale for the subsequent robustness objective, we estimate global calibration statistics from a fixed calibration set $\mathcal{C}_{\mathrm{cal}}$. Specifically, we define

$$m = \mathbb{E}_{c\sim\mathcal{C}_{\mathrm{cal}},\, z\sim p_0,\, \delta\sim q_\rho}\left[U_c(z+\delta)\right], \tag{14}$$

and

$$s^2 = \mathrm{Var}_{c\sim\mathcal{C}_{\mathrm{cal}},\, z\sim p_0,\, \delta\sim q_\rho}\left[U_c(z+\delta)\right], \tag{15}$$

where $p_0 = \mathcal{N}(0, I)$ is the standard diffusion initialization prior. In practice, these expectations are estimated once using finite samples and the resulting statistics are kept fixed for all prompts under the same model and inference configuration. Using these global statistics, we define the normalized guidance potential as

$$\widetilde{U}_c(x) = \frac{U_c(x) - m}{s + \varepsilon}, \tag{16}$$

where $\varepsilon > 0$ is a small constant for numerical stability. Since the same affine transformation is applied to every condition, this normalization does not remove prompt-dependent differences in U_c; it only provides a common numerical reference scale for the robustness objective. In particular, m and s are fixed during initialization optimization and therefore introduce no additional gradient terms or prompt-specific calibration at evaluation time.

The proposed neighborhood-robust potential is

$$\mathcal{R}_{\rho,\lambda}(z) = \frac{1}{\lambda}\log \mathbb{E}_{\delta\sim q_\rho}\left[\exp\left(\lambda \widetilde{U}_c(z+\delta)\right)\right], \qquad \lambda > 0. \tag{17}$$

For $\lambda \to 0^+$,

$$\mathcal{R}_{\rho,\lambda}(z) = \mathbb{E}_{q_\rho}\left[\widetilde{U}_c(z+\delta)\right] + \frac{\lambda}{2}\,\mathrm{Var}_{q_\rho}\left[\widetilde{U}_c(z+\delta)\right] + O(\lambda^2). \tag{18}$$

Thus, the robust potential reduces to the average neighborhood potential in the small-λ limit, while positive λ additionally penalizes variation and places greater emphasis on unfavorable perturbations. This choice also admits a distributionally robust interpretation. By the variational representation of the log moment-generating function,

$$\boxed{\mathcal{R}_{\rho,\lambda}(z) = \sup_{\nu \ll q_\rho}\left\{\mathbb{E}_{\delta\sim\nu}\left[\widetilde{U}_c(z+\delta)\right] - \frac{1}{\lambda}D_{\mathrm{KL}}\left(\nu\|q_\rho\right)\right\}.} \tag{19}$$

Therefore, for any local perturbation law $\nu_R \ll q_\rho$,

$$\mathbb{E}_{\delta\sim\nu_R}\left[\widetilde{U}_c(z+\delta)\right] \le \mathcal{R}_{\rho,\lambda}(z) + \frac{1}{\lambda}D_{\mathrm{KL}}\left(\nu_R\|q_\rho\right). \tag{20}$$

Equation equation 20 provides the main justification for our robust objective. The downstream reward-induced displacement is unavailable when initialization is constructed and may vary across optimization steps. Rather than anticipating a particular reward direction, minimizing $\mathcal{R}_{\rho,\lambda}$ controls the expected potential over a family of local perturbation distributions around the nominal kernel q_ρ. For a finite set of perturbations $\{\delta_j\}_{j=1}^J$, the empirical objective is

$$\widehat{\mathcal{R}}_{\rho,\lambda}(z) = \frac{1}{\lambda}\log\left[\frac{1}{J}\sum_{j=1}^{J}\exp\left(\lambda\widetilde{U}_c(z+\delta_j)\right)\right]. \tag{21}$$

The log-sum-exp form smoothly interpolates between average neighborhood behavior and high-potential perturbations, avoiding the non-smooth optimization induced by a hard local maximum.

B Derivation of the Semantic Coverage Objective

The connection between the effective representation dimension in Eq. equation 11 and the population redundancy objective in Eq. equation 12 can be derived directly from the representation second-moment matrix. For simplicity, let

$$f_i = f_c(z_i), \qquad C = C_c(\mathcal{Z}) = \frac{1}{M}\sum_{i=1}^{M} f_i f_i^\top,$$

where $\|f_i\|_2 = 1$ for all i. Expanding the squared second-moment matrix gives

$$\operatorname{tr}(C^2) = \frac{1}{M^2}\operatorname{tr}\left(\sum_{i=1}^{M}\sum_{j=1}^{M} f_i f_i^\top f_j f_j^\top\right) \tag{22}$$

$$= \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\operatorname{tr}\left(f_i f_i^\top f_j f_j^\top\right). \tag{23}$$

Using

$$\operatorname{tr}\left(f_i f_i^\top f_j f_j^\top\right) = (f_i^\top f_j)^2, \tag{24}$$

it follows that

$$\operatorname{tr}(C^2) = \frac{1}{M^2}\sum_{i,j}(f_i^\top f_j)^2 \tag{25}$$

$$= \frac{1}{M^2}\left[\sum_{i=1}^{M}\|f_i\|_2^4 + \sum_{i\neq j}(f_i^\top f_j)^2\right]. \tag{26}$$

Since the representations are normalized, $\|f_i\|_2 = 1$, yielding

$$\operatorname{tr}(C^2) = \frac{1}{M^2}\left[M + \sum_{i\neq j}(f_i^\top f_j)^2\right] \tag{27}$$

$$= \frac{1}{M}\left[1 + (M-1)\mathcal{R}_c(\mathcal{Z})\right], \tag{28}$$

where

$$\mathcal{R}_c(\mathcal{Z}) = \frac{1}{M(M-1)}\sum_{i\neq j}(f_i^\top f_j)^2.$$

Therefore,

$$D_{\text{cov}}(\mathcal{Z}) = \frac{1}{\operatorname{tr}(C^2)} = \frac{M}{1 + (M-1)\mathcal{R}_c(\mathcal{Z})}. \tag{29}$$

For a fixed population size M, D_{cov} is strictly decreasing with $\mathcal{R}_c$. Maximizing the effective representation dimension is therefore equivalent to minimizing the average squared similarity among candidate representations.

Let $\{\lambda_k\}$ denote the eigenvalues of C. Since $\operatorname{tr}(C) = 1$,

$$D_{\text{cov}} = \frac{1}{\sum_k \lambda_k^2}. \tag{30}$$

This quantity measures the effective number of representation directions occupied by the candidate population. If all representation energy is concentrated along a single direction, then $\lambda_1 = 1$ and $D_{\text{cov}} = 1$. In contrast, if the energy is evenly distributed across r orthogonal directions, such that $\lambda_1 = \cdots = \lambda_r = 1/r$, then

$$D_{\text{cov}} = r. \tag{31}$$

Hence, reducing pairwise representational redundancy encourages the candidate population to distribute its representation energy across a larger number of effective directions rather than repeatedly occupying the same representational subspace.

C EXPERIMENTAL DETAILS

For all experiments, we adopt a fixed sampling configuration for each diffusion backbone and keep it unchanged across different inference-time optimization methods to ensure a fair comparison. For SDXL, images are generated at a resolution of 1024×1024 using 50 denoising steps with a classifier-free guidance (CFG) scale of 5.0. For SD3.5, images are generated at the same resolution of 1024×1024 using 28 denoising steps with a CFG scale of 3.5.

For each prompt, all methods ultimately generate $N = 4$ output samples. The size of the initial candidate population follows the default setting of each method. For example, PSP initializes the sampling procedure with $2N = 8$ candidates, and our method adopts the same initial population size of $2N = 8$ candidates before subsequent refinement and selection.