

SwiftExplorer: Training-free Diffusion Model Alignment with Swift Diversity Exploration

Renye Yan*
Peking University
Beijing, China

Jikang Cheng*
Peking University
Beijing, China

You Wu
Nanjing University
Nanjing, China

Baojin Huang
Huazhong Agricultural University
Wuhan, China

Wei Peng
Stanford University
Stanford, California, USA

Zongwei Wang
Peking University
Beijing, China

Ling Liang[†]
Peking University
Beijing, China

Yimao Cai[†]
Peking University
Beijing, China

ABSTRACT

Diffusion models have general generative abilities but struggle to align with specific objectives. Fine-tuning can improve alignment, yet its training cost is often prohibitive. This led to training-free methods that apply objective-guided terms in sampling to bias the generation distribution toward designated regions, e.g., high-reward areas. However, these methods face two issues: (1) the strong directional bias narrows the pretrained distribution and generation diversity, and (2) indiscriminate constant guidance fails to prune redundant signals, hurting both quality and efficiency. To address the above challenges, we propose *SwiftExplorer*, a plugin that mitigates distribution collapse caused by excessive diversity loss and reduces compute costs. First, we adopt an Inheritance-Restart exploration mechanism to avoid early convergence, while exploration also increases the likelihood of high-reward trajectories. Additionally, it balances diversity and fidelity, adding diversity without causing a distribution over-shift. Second, our Quality-Efficiency arbitration mechanism improves guidance by removing incorrect signals, and it reduces computation by dynamically stopping generation when completeness and marginal reward gain are optimal. In an extensive number of experiments and different types of evaluation metrics, the proposed *SwiftExplorer* achieves excellent performance on all metrics, including preference, fidelity, diversity, and richness.

CCS CONCEPTS

• Computing methodologies → Computer vision problems; • Generation Optimization;

KEYWORDS

Language-Vision Conditional Generation, Training-free Guidance

1 INTRODUCTION

Diffusion models [15, 22, 26, 29, 31, 36, 37, 41, 53–55, 60], as one of the most powerful generative frameworks, demonstrate exceptional generative capabilities by training on various large-scale datasets. This is evident in the model’s capability to generate both high-quality images and a diverse range of semantic and stylistic

*Renye Yan and Jikang Cheng contributed equally to this work.

[†]Ling Liang and Yimao Cai are the corresponding authors.

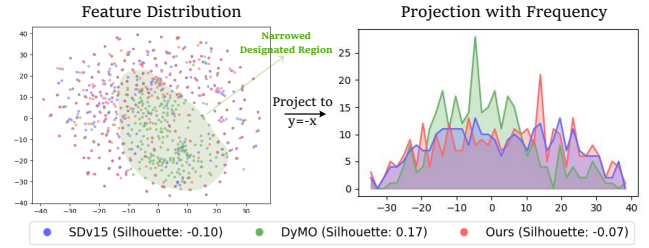


Figure 1: Feature visualization via t-SNE. The features are extracted by a general self-supervised decoder (*i.e.*, DINOv2) from images generated by different methods. With the guidance targets of both Ours and DyMO having achieved comparable scores, the distribution results indicate that existing SOTA guidance methods (*i.e.*, DyMO) yield more compact clustering, narrowing the generated distribution, which in turn reduces the diversity of the generated data. This observation is substantiated by the Silhouette Coefficient [38], a clustering evaluation metric, which demonstrates its convergence and limited diversity. In contrast, our method maintains a broader generative distribution while achieving comparable target reward scores (see Tab. 1).

content. However, in practical applications, it is not only necessary to generate realistic images but also to ensure that the results align with specific target requirements, such as aesthetic preferences. This need for goal alignment is becoming a core research problem, posing the challenge of effectively translating the general capabilities of diffusion models into customized generation.

Early research primarily relied on training strategies [3, 11, 45, 56]. By fine-tuning diffusion models on specific preference objectives, alignment performance could be significantly improved. However, such methods generally face challenges of high training costs and poor transferability [18, 49]. Whenever user requirements or objectives change, diffusion models must be finetuned again, which limits their scalability and transferability in practical applications. To address this, researchers proposed training-free methods [5, 7, 18, 44, 49], which do not modify the model’s weight

but incorporate external signals during generation to achieve goal-guided results. The advantages of such approaches are evident as they require no additional training costs and can quickly adapt to different objective demands. Consequently, training-free guidance has become a crucial diffusion model alignment method. More detailed **Related Works** can be found in Appendix 1.

However, existing training-free methods still suffer from two typical limitations: (1) *overly narrow generative distributions*. To enhance the alignment between generated results and target objectives, training-free methods should incorporate target-guided terms during sampling. These methods gradually shift the generative distribution toward specific regions, such as high-reward subspaces. Although these approaches effectively enhance alignment, their excessive reliance on a consistent guidance direction easily leads to reward hacking [44], which manifests as a rapid narrowing of the pretrained distribution, ultimately reducing image diversity [16, 44]. (2) *unselective, inefficient guidance*. Current methods apply guidance terms indiscriminately throughout the entire generation process, lacking the ability to identify and manage ineffective or redundant signals. This not only results in unnecessary computational waste but may also introduce noise interference due to ineffective guidance, ultimately compromising both alignment quality and generation efficiency.

To address the above challenges, we propose *SwiftExplorer*, an enhanced plugin compatible with existing training-free alignment methods. (1) *SwiftExplorer* introduces an Inheritance-Restart exploration mechanism that prevents premature convergence of generation trajectories to a single mode through probabilistic perturbation and path screening. Simultaneously, we introduce a progress-aware adaptive trade-off to enhance exploration while preserving generation fidelity (see Fig. 1). It promotes early-stage exploration to improve diversity and the discovery of high-reward trajectories. In later stages, it encourages convergence to preserve the pretrained distribution. (2) To improve guidance efficiency, *SwiftExplorer* employs a Quality-Efficiency screening mechanism to filter out ineffective signals, ensuring that each guidance step contributes meaningfully. Additionally, by integrating an adaptive factor and considering the diminishing marginal utility of rewards, our method dynamically cuts guidance in later generations to avoid redundant optimization. The Quality-Efficiency mechanism effectively reduces computational overhead while maintaining alignment quality. Experiments show that while ensuring leading performance on both fidelity and alignment objectives, *SwiftExplorer* effectively improves generative diversity and outperforms existing baselines in computational efficiency. Our contributions are as follows:

- We propose an Inheritance-Restart exploration mechanism to enhance the generation diversity of existing methods.
- We introduce a Quality-Efficiency mechanism to drop invalid signals, thereby improving guidance and reducing costs.
- We achieve two trade-offs: Diversity and Fidelity evaluation during generation progress. Quality and Efficiency adjustment on marginal reward gains.

2 METHOD

SwiftExplorer is a plugin enhancing training-free methods in exploration and efficiency. Sec. 2.1 covers preliminaries. Sec. 2.2 elaborates on the Inheritance-Restart exploration mechanism, achieving an adaptive trade-off between Diversity and Fidelity. Sec. 2.3 introduces a Quality-Efficiency arbitration mechanism guiding early quality and late efficiency. The *SwiftExplorer* framework is shown in Fig. 2, with algorithmic details in Appendix. 4.

2.1 Reward-Guided Diffusion Alignment

2.1.1 Generative Process of Diffusion Models. The discrete-time generative process (or reverse process) aims to produce a sample by denoising a latent variable $\mathbf{x}_t$ iteratively. This process requires two components: a predefined noise schedule $\{\beta_t\}_{t=0}^T$ and a well-trained score function $s_\theta(\mathbf{x}_t, t)$. We define $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$. The generation starts by sampling an initial latent variable from a standard normal distribution: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Then, for each timestep t from T to 0, the sample is iteratively refined using the update rule:

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} (\mathbf{x}_t + \beta_t s_\theta(\mathbf{x}_t, t)) + \sqrt{\beta_t} \mathbf{z}_t, \quad (1)$$

where $\mathbf{z}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is a standard Gaussian noise vector independently sampled at each timestep. The final output after T steps is the generated sample $\mathbf{x}_0$. For convenience, we also define a function $f_\theta(\mathbf{x}_t, t)$ that provides a direct estimate of the clean sample $\mathbf{x}_{0|t}$, from any noisy intermediate $\mathbf{x}_t$ [2, 5, 49]:

$$\mathbf{x}_{0|t} = f_\theta(\mathbf{x}_t, t) = \frac{1}{\sqrt{\bar{\alpha}_t}} (\mathbf{x}_t + (1 - \bar{\alpha}_t) s_\theta(\mathbf{x}_t, t)). \quad (2)$$

2.1.2 Iterative training-free Refinement with Gradient Guidance. To guide the generation process towards a specific objective, we can incorporate the gradient of a reward function $\mathcal{R}$ [24]. The reward function is defined in the data domain (i.e., it operates on the estimated clean sample $\mathbf{x}_{0|t}$ by Eq. 2). The approach of predicting $\mathbf{x}_{0|t}$ using Eq. 2 for reward calculation is widely utilized in the field [2, 49]).

The guidance is injected by adjusting the current sample $\mathbf{x}_t$ in the direction of the reward gradient. Specifically, before applying the update in Eq. 1, we compute a guided sample $\tilde{\mathbf{x}}_t$:

$$\tilde{\mathbf{x}}_t \leftarrow \mathbf{x}_t + w \nabla_{\mathbf{x}_t} \mathcal{R}(f_\theta(\mathbf{x}_t, t)), \quad (3)$$

where w is a guidance scale factor, $\mathcal{R}(\cdot)$ denotes the reward function for inference-time guidance. The above process could be seen as one step of optimization. For simplicity, we can define optimization with m steps of Eq. 3 as:

$$\tilde{\tilde{\mathbf{x}}}_t = g(\mathbf{x}_t, \mathcal{R}, m). \quad (4)$$

Replacing $\mathbf{x}_t$ in Eq. 1 with $\tilde{\tilde{\mathbf{x}}}_t$, the new update rule becomes:

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} (\tilde{\tilde{\mathbf{x}}}_t + \beta_t s_\theta(\tilde{\tilde{\mathbf{x}}}_t, t)) + \sqrt{\beta_t} \mathbf{z}_t. \quad (5)$$

2.2 Inheritance-Restart Exploration mechanism

To mitigate the issue of narrowed generative distribution caused by overly concentrated guidance direction in training-free methods, *SwiftExplorer* equips an Inheritance-Restart exploration mechanism. This mechanism selectively introduces exploratory supplementary

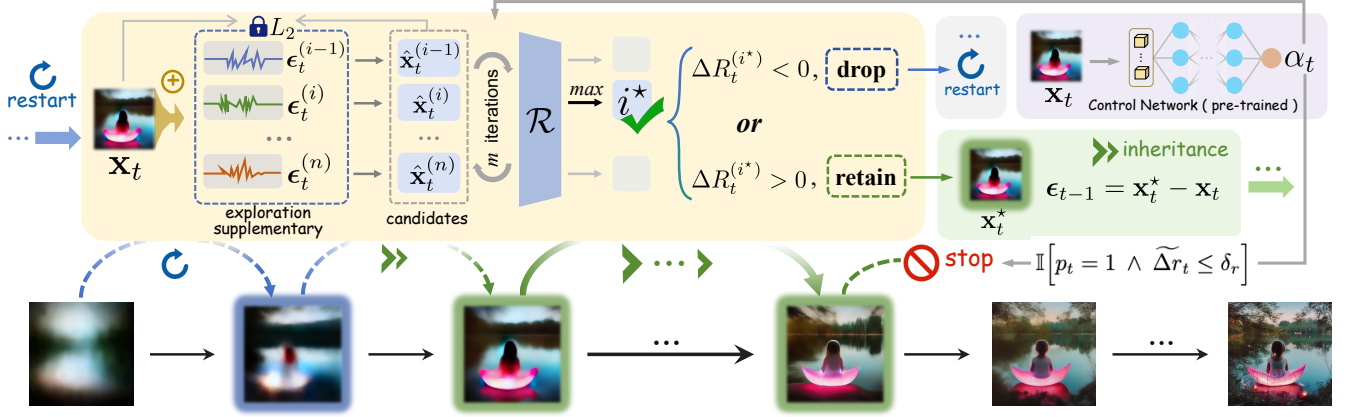


Figure 2: Overall framework of the proposed plug-and-play SwiftExplorer.

terms based on the generation progress, thereby preventing premature convergence of generative trajectories to a single mode and increasing the probability of discovering high-reward paths.

2.2.1 Exploration Supplement Variable. During the generation step t , we introduce an exploration supplement variable $\epsilon_t^{(i)}$, which allows the generated trajectories to maintain a richer distribution. Specifically, we adopt Monte Carlo sampling to obtain n candidates $\epsilon_t^{(i)} \sim \mathcal{N}(0, \sigma^2 I)$, $i = 1, \dots, n$. After incorporating $\epsilon_t^{(i)}$, the original $\mathbf{x}_t$ without exploration is updated into a set $\{\hat{\mathbf{x}}_t^{(i)}\}$:

$$\hat{\mathbf{x}}_t^{(i)} = \mathbf{x}_t + \epsilon_t^{(i)}. \quad (6)$$

Subsequently, each candidate $\hat{\mathbf{x}}_t^{(i)}$ is mapped to its predicted reconstruction $\hat{\mathbf{x}}_{0|t}^{(i)} = f_\theta(\hat{\mathbf{x}}_t^{(i)}, t)$, and is evaluated via the objective reward function $\mathcal{R}(\hat{\mathbf{x}}_{0|t}^{(i)})$. After m rounds based on Eq. 4, we obtain a final set of candidates $\{\tilde{\mathbf{x}}_t^{(i)}\} = g(\{\hat{\mathbf{x}}_t^{(i)}\}, \mathcal{R}, m)$. We then define the optimal generation trajectory policy as:

$$i^* = \arg \max_{i \in \{1, \dots, n\}} \mathcal{R}(\tilde{\mathbf{x}}_{0|t}^{(i)}), \quad \mathbf{x}_t^* = \tilde{\mathbf{x}}_t^{(i^*)}. \quad (7)$$

This process selects the optimal candidate $\mathbf{x}_t^*$ as the sample of generation step t , which is adopted for the following generations. In summary, the injection of the exploration supplementary term expands the local search space. The selection policy picks the optimal generation trajectory with the highest alignment potential, thereby increasing diversity while maximizing alignment rewards.

2.2.2 Inheritance-Restart Technique. In Sec. 2.2.1, the exploration supplementary term expands the generation space. Building upon this, we apply an Inheritance-Restart technique to this term. This technique dynamically regulates exploration behavior based on the target reward improvement, preserving high-quality explorations while eliminating ineffective adjustments. Consequently, the exploration term enhances diversity, while the Inheritance-Restart technique ensures that exploration consistently improves the target reward. Specifically, the reward changing for the supplementary term at generation step t is defined as:

$$\Delta \mathcal{R}_t^{(i^*)} = \mathcal{R}(\mathbf{x}_{0|t}^*) - \mathcal{R}(\mathbf{x}_{0|t}). \quad (8)$$

- (1) If $\Delta \mathcal{R}_t^{(i^*)} > 0$, the exploration still yields a positive gain for the alignment reward. In this case, the latent change from the previous step is carried over to the next step $t-1$, where $\epsilon_{t-1} = \mathbf{x}_t^* - \mathbf{x}_t$. Under this inheritance strategy, the exploration candidate set contains only this inherited signal, i.e., setting $n = 1$ in Eq. 6, and optimization continues via Eq. 4.
- (2) If $\Delta \mathcal{R}_t^{(i^*)} < 0$, the exploration gain becomes saturated or starts to decline. In this case, the restart mechanism is activated: n exploration candidates are resampled in the next step $t-1$ as Eq. 6, and the exploration term is updated through the optimization in Eq. 4 and the selection step in Eq. 7.

The inheritance-restart technique uses reward changes as a criterion to enhance exploration diversity and reduce guidance overhead. Moreover, it ensures the generation alignment with target objective.

2.2.3 Diversity and Fidelity Adaptive Trade-off. Diffusion models face the dual demands of diversity and fidelity. Diversity expands the coverage of potential generation distribution, while fidelity ensures the results are credible and semantically consistent with the conditioning. However, these two objectives often exhibit a trade-off, making their careful balance crucial for achieving high-quality results [8]. To address this, we propose an adaptive coordination technique that dynamically adjusts the emphasis based on the generation progress: it prioritizes exploration in the early stages to foster diversity and gradually increases the strength of regularization in the later stages to ensure fidelity. Unlike approaches that use predefined static parameters, the adaptation technique can flexibly accommodate the dynamic nature of the generation process.

To this end, we construct a Control Network (This network outputs a simple noise scorer, analogous to scoring models such as PickScore and Aesthetic Score, and is therefore used as a **pretrained model**. Experimental support is provided in Tab. 8, with details in Appendix 2.2.), which is designed to assess the denoising progress of the intermediate state in real-time during the generation process. Specifically, the control network is defined as a mapping function: $h_{\theta_p} : \mathbf{x}_t \mapsto p_t$, where $\mathbf{x}_t \in \mathbb{R}^d$ denotes the latent state at generation step t . $p_t \in [0, 1]$ represents the probability of the current latent state close to the clean image $\mathbf{x}_0$, which reflects the degree of the current state's generation progress. The control network is obtained

by minimizing a binary cross-entropy loss as the supervision signal (see Appendix 2.1 for details). We utilize p_t as an adaptive weighting to balance exploration diversity and generation fidelity. In the first optimization round, the combined effect of the exploration and alignment signal gradients is written as:

$$\hat{\mathbf{x}}_t^{(i)} = \mathbf{x}_t + \underbrace{(1 - p_t) \cdot \epsilon^{(i)} - p_t \cdot \nabla_{\mathbf{x}_t} L_2(\hat{\mathbf{x}}_t, \mathbf{x}_t) + \nabla_{\mathbf{x}_t} \mathcal{R}(\mathbf{x}_{0|t})}_{\text{Diversity-Fidelity Trade-off}}, \quad (9)$$

where $\epsilon^{(i)}$ is the exploration supplement variable and $\nabla_{\mathbf{x}_t} \mathcal{R}(\mathbf{x}_{0|t})$ represents the reward-based gradient guidance. We introduce L2 regularization to constrain the exploration scope. Then, based on the iteration defined in Eq. 4, we iteratively introduce guidance to obtain the final state $\tilde{\mathbf{x}}_t^{(i)} = g(\hat{\mathbf{x}}_t^{(i)}, \mathcal{R} - L_2, m)$.

In summary, the adaptive trade-off encourages exploratory diversity in the early generation stages. In the later stages, the continuous increase in the regularization constraint prompts the generation process to transition from exploration to stable convergence, thereby ensuring high-fidelity results. Through this design, *SwiftExplorer* achieves an adaptive trade-off between creativity and stability.

2.3 Quality-Efficiency Arbitration mechanism

2.3.1 Guidance Screening. Existing methods apply guidance signals indiscriminately throughout the entire sampling process, which can lead to ineffective guidance and undermine alignment effectiveness. To address this, *SwiftExplorer* introduces a guidance signal selection technique that dynamically discriminates the signal’s utility. Let the optimal guidance signal generated at step t be $\mathbf{y}_t^{(i^*)}$, which represents the reward gradient corresponding to $\mathbf{x}_t^*$, i.e., the optimal candidate. We then define a selection operator $\mathcal{F}$:

$$\mathcal{F}(\mathbf{y}_t^{(i^*)}) = \{\Delta \mathcal{R}_t^{(i^*)} > 0\}. \quad (10)$$

With the proposed screening process, a guidance signal is retained only if it contributes to a positive reward gain; otherwise, it is discarded, and the original $\mathbf{x}_t$ from the base diffusion model is used. This reward-margin-based selection strategy ensures the quality of the guidance signals for generation.

2.3.2 Quality and Compute Adaptive Trade-off. Unlike existing methods that rigidly couple guidance with the generation steps, *SwiftExplorer* regulates the use of guidance signals through the dual control of target reward marginal gain and generation progress p_t . This approach aims to reduce computational costs while ensuring alignment quality. Guidance is stopped when it is insignificant to reward improvement and the denoising meets the required threshold, thereby avoiding redundant costs and preventing over-guidance. We first denote the step t reward $\mathcal{R}_t^{(i^*)}(\cdot)$ as r_t . Then, the smoothed trend of reward gain is defined as:

$$\widetilde{\Delta r}_t = (1 - \phi) \widetilde{\Delta r}_{t+1} + \phi (r_t - r_{t+1}), \quad (11)$$

where $\phi \in (0, 1]$. If and only if $p_t = 1$ and $\widetilde{\Delta r}_t \leq \delta_r$, the reward return margin has been reached and the early stopping is triggered, where $\delta_r \geq 0$ is the minimal tolerance gain. We define the early-stopping indicator function as:

$$E_{stop}(t) = \mathbb{I}[p_t = 1 \wedge \widetilde{\Delta r}_t \leq \delta_r] \in \{0, 1\}. \quad (12)$$

Exploration and guidance stop when $E_{stop}(t) = 1$ and continue when $E_{stop}(t) = 0$. This criterion ensures triggers early stopping only when both conditions are met: p_t guarantees the state is ‘sufficiently clean’, while $\widetilde{\Delta r}_t$ ensures that ‘further exploration guidance yields no significant reward improvement’. By requiring both conditions simultaneously, this technique eliminates redundant guidance in the later stages of generation without compromising target alignment quality, thereby reducing computational cost.

3 EXPERIMENTS

3.1 Experimental Setup

Datasets. We employ Pick-a-pic [19], HPSv2 [47], GenEval [12], and animals [3] datasets as our basic test bed. The prompts in Pick-a-Pic are relatively complex, abstract, and surreal, whereas those in HPSv2 are more realistic and aligned with the real world. GenEval includes specified counting and composition tasks.

Optimization Objectives and Metrics. We deploy three reward functions as the optimization objectives, that is, PickScore [19] (PS), Aesthetic Score [40] (AES), and ImageReward [50] (IR). To quantitatively evaluate the generative performance as extensively as possible, we introduce 4 evaluating dimensions with 13 distinct metrics, including: *Aesthetic Preference*: AES, PS, IR, and HPSv2 [47]. *Image Fidelity*: ClipScore [35] (Clip), Fr chet Inception Distance (FID) [14], and improved F1 Score (iFS) [21]. *Generative Diversity*: LPIPS [58], TCE [17], and Inception Score [39] (IS). *Compositional Richness*: NIQE [33], BRISQUE (BRI) [32], and Spectral Entropy (SE) [28]. Details of each metric can be found in Appendix. 3.3.

Baselines. For comprehensive comparison, we introduce three types of State-of-The-Art (SoTA) diffusion-based alignment methods, including: *Gradient-based Training*: Diff-DPO [45], Diff-KTO [23], SPO [25], and DRaFT [6]. *RL-based Training*: DDPO [3], DPOK [10]. *Training Free*: DAS [59], TTScale [30], FKSteer [42], DNO [44], and DyMO [49]. All baselines are strictly reproduced based on their official code and settings within our evaluation benchmark.

Backbones. We adopt two types of widely-adopted backbones: *Unet-based* models: SDv14, SDv15, SDv21, SD XL1.0 (XL) and *Flow-matching* models: SD 3.0-medium (SD3), SD3.5-medium (SD3.5).

3.2 Comparisons Against SoTA Methods

To demonstrate the effectiveness of *SwiftExplorer* (SwiftE), we carefully reproduce and evaluate **12 baselines** with PickScore as the objective on SDv15 and pick-a-pic datasets. Since SwiftE is an efficient plugin, this section represents deploying SwiftE to DyMO. Then, we introduce **13 metrics** to assess from four dimensions. As shown in Tab.1, our method achieves the best overall performance in all dimensions, *achieving Top2 performance in 12 metrics*. Moreover, since the prompts in the Pick-a-pic dataset are relatively abstract and surreal, we further conduct an experiment on the photorealistic HPSv2 dataset for comprehensive evaluation. In Tab.2, all methods perform better because realistic objects are more likely to be seen during training and hence easier to generate. *our method still achieves the best results on five metrics and maintains comparable fidelity*. The qualitative results in Appendix Fig. 12 further demonstrate the superiority of our method. In Appendix. 3.4 & Tab. 7, to assess the impact of random seeds on result stability, we compute the coefficient of variation (CoV) of each metric for the



Figure 3: Visual impression on the advanced flow-matching-based models. Our results exhibit finer details, more diverse colors, and improved aesthetic style.

training-free baseline across different seeds.

Comparisons with more advanced SD models. In Appendix

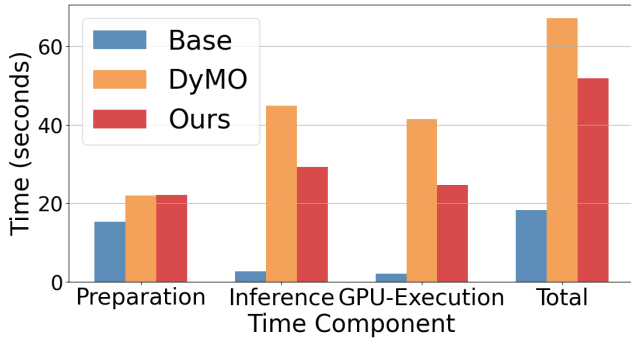


Figure 4: Wall-clock runtime comparison under consistent experimental conditions.

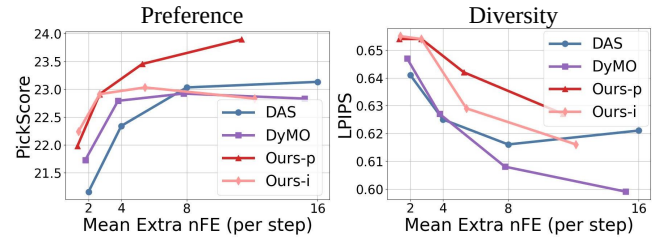


Figure 5: Performance-Computation Trade-off.

Tab. 11, we employ our method and other SoTA on the SD-XL backbone, and introduce the advanced SDv35 as an additional baseline. We also implement it on the advanced SD3 and SD3.5 backbones with PS as the reward and Pick-a-pic as the dataset. Quantitatively, as shown in Tab. 3, our method can still effectively improve the reward score while maintaining the alignment and diversity. Qualitatively, in Fig. 3, it can be clearly observed that our method has achieved improved human preference in aesthetics.

Table 1: Main comparisons with SoTA. All metrics are obtained with SDv15 as backbone, Pick-a-pic as prompt set, and PickScore as reward for guidance. Orange denotes gradient-based training methods. Green denotes reinforcement-learning-based training methods. Blue denotes training-free methods.

Method	Preference				Fidelity			Diversity			Richness			#Top2
	PS↑	AES↑	IR↑	HPS↑	FID↓	CLIP↑	iFS↑	LPIPS↑	IS↑	TCE↑	BRI↓	NIQE↓	SE↑	
SDv15	20.48	5.412	0.181	0.262	-	<u>0.243</u>	-	0.654	23.79	38.05	18.66	5.401	11.26	3
Diff-DPO	20.87	5.551	0.443	0.271	109.1	0.244	0.795	0.639	22.77	39.20	15.12	4.463	11.02	1
Diff-KTO	20.83	5.585	0.599	0.272	101.3	0.240	0.801	0.634	22.70	39.12	26.25	4.361	11.22	0
SPO	20.76	5.613	0.282	0.218	78.76	0.241	0.855	0.649	23.18	39.32	25.67	<u>4.103</u>	11.09	1
DRaFT	22.52	<u>5.697</u>	<u>0.779</u>	0.271	84.47	0.231	0.856	0.572	20.10	37.85	35.64	5.336	10.43	2
DDPO	21.79	5.704	0.196	0.212	147.5	0.242	0.539	0.629	20.01	39.18	12.46	4.328	<u>11.74</u>	1
DPOK	20.97	5.661	0.583	0.272	99.30	0.242	0.820	0.641	22.52	<u>39.35</u>	<u>12.28</u>	4.608	11.23	2
DNO	20.87	5.479	0.425	0.271	79.96	0.239	0.851	0.591	19.94	38.88	17.23	4.757	11.07	0
TTSscale	22.14	5.636	0.619	0.272	66.93	0.240	0.863	0.652	<u>23.73</u>	38.90	15.05	4.447	11.28	2
FKSteer	22.56	5.696	0.539	0.278	73.83	<u>0.243</u>	<u>0.868</u>	0.650	23.76	39.42	13.93	4.092	11.59	2
DAS	22.34	5.658	0.632	0.283	95.42	0.240	0.844	0.625	21.01	39.19	14.36	4.631	11.14	1
DyMO	<u>22.79</u>	5.694	0.709	0.279	90.56	0.239	0.849	0.627	20.97	39.29	16.18	4.545	11.29	1
Ours	22.91	5.707	0.812	<u>0.280</u>	<u>71.36</u>	0.242	0.891	0.654	<u>23.71</u>	39.37	10.92	2.886	12.07	12

Table 2: Results on HPSv2 dataset.

Method	IR	PS	CLIP	LPIPS	TCE	NIQE↓
SDv15	0.231	21.57	0.249	0.643	39.62	4.656
Diff-DPO	0.571	22.03	0.245	0.619	37.97	4.471
SPO	0.300	21.58	0.249	0.595	39.37	4.172
DDPO	0.678	22.26	0.245	0.607	38.96	3.973
DNO	0.446	21.79	0.246	0.575	39.14	4.523
DAS	0.684	23.09	0.246	0.653	39.31	4.133
DyMO	0.805	23.46	0.246	0.610	39.57	4.088
Ours	0.872	23.79	0.247	0.671	39.92	2.595

Table 3: Comparison results with advanced *flow matching* models. Gdt represents the baseline that directly applies the reward gradient to the latent code during inference. Ours is designed based on Gdt with the proposed plugin components.

Method	SD3					SD3.5				
	PS	CLIP	LPIPS	NIQE↓	Top1	PS	CLIP	LPIPS	NIQE↓	Top1
Base	21.62	0.238	0.645	4.312	1	21.90	0.238	0.621	4.553	1
+Gdt	22.17	0.239	0.636	4.304	0	22.95	0.235	0.616	4.174	0
+Ours	22.57	0.241	0.614	2.990	3	23.93	0.238	0.625	3.794	4

3.3 Plug-and-Play Effectiveness

In Fig. 7, we discuss the plugin effectiveness of *SwiftExplorer*. Specifically, we present side-by-side comparisons of results with and without SwiftE. Gdt represents backbones with Gradient guidance.

We choose PS, CLIP, LPIPS, $\frac{1}{\text{NIQE}}$, and image-per-minute as the representations of Preference, Fidelity, Diversity, Richness, and Speed. The results clearly show that SwiftE effectively enhances the diversity of baseline methods and reduces computational cost while improving other quality dimensions.

3.4 Analysis on Inference Efficiency

Comparisons on Wall-clock Runtime. Here, we measure and compare the time consumption of all methods under consistent experimental conditions. To investigate the effectiveness in real-world applications, we conduct comparisons on wall-clock runtime with details. As shown in Fig. 4, we present the time consumption of running the entire script. Firstly, their preparation time are similar and once prepared, the script can conduct arbitrary times of inference. Then, it can be observed that our inference time is much faster than the baseline DyMO, while Base has the smallest inference time since it includes no optimization or guidance.

Performance-Computation Trade-off. To investigate the relation between computational cost and performance, we first deploy *extra mean number of Function Evaluated (nFE)* to quantify the computational cost in a unified manner, which represents the number of extra iterations or sampling in one denoising step. Then, we compare our method with iteration-based DyMO and sampling-based DAS, where the maximum iteration and sampling particle number are set to 2,4,8,16. Notably, *SwiftExplorer* includes both sampling

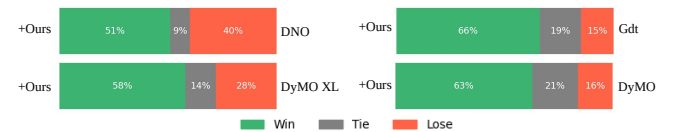


Figure 6: User study for plug-and-play effectiveness.

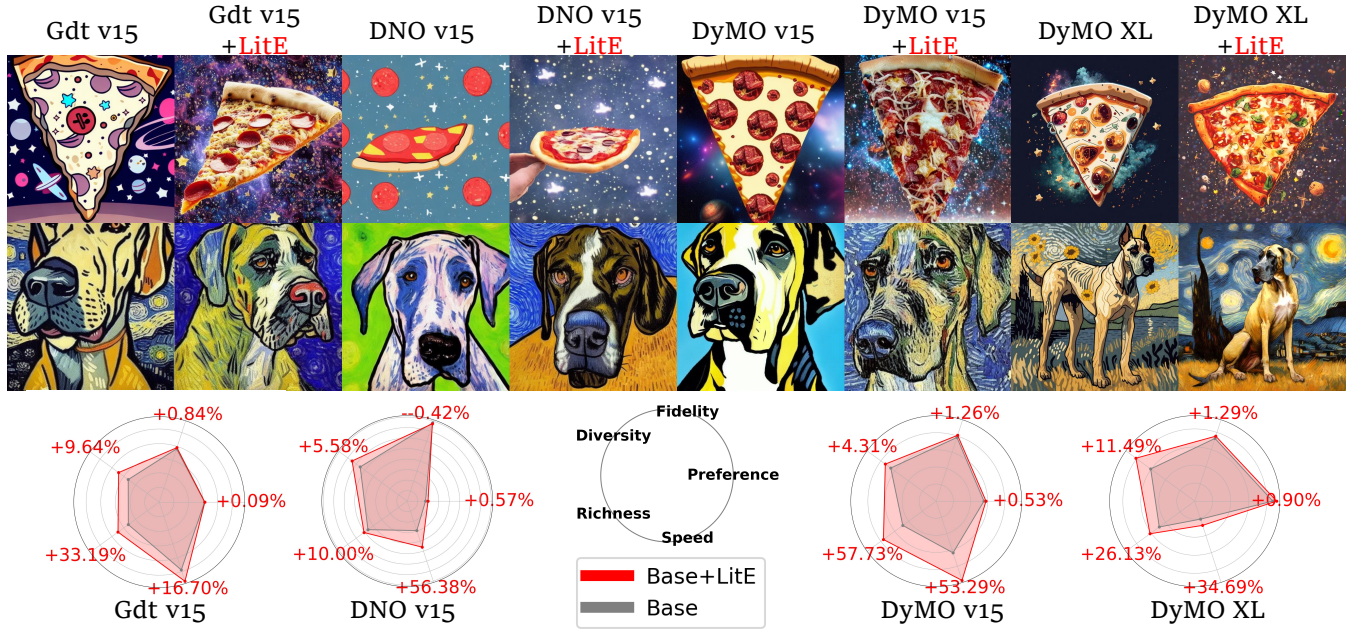


Figure 7: Plug-and-Play effectiveness. While maintaining the Fidelity and Preference, our method significantly enhances Diversity, Richness, and Speed across all cases, such as the “background and pizza filling diversity” in the first row, and “improved Van Gogh style” in the second row.

Table 4: Ablation study. For each component, ESV and M-ESV represent exploration supplement variable and with Monte Carlo sampling. I&R and EarS denote Inheritance-Restart technique and Early Stopping. GuiS denotes Guidance Screening.

Var	Dive		Fide L2	Effi		Qual GuiS	IR	PS	CLIP	LPIPS	TCE	NIQE↓	Time↓	#Top2
	ESV	M-ESV		I&R	EarS									
Base							0.709	22.79	0.239	0.627	39.29	4.454	44.92	0
V1	✓						0.739	22.12	0.236	0.659	39.30	2.897	46.31	1
V2	✓		✓				0.734	22.46	0.240	0.636	39.30	3.379	51.24	0
V3		✓	✓				0.757	22.75	0.240	0.641	39.32	2.939	67.08	0
V4		✓	✓	✓			0.746	22.73	0.241	0.640	39.29	3.039	48.56	0
V5		✓	✓	✓	✓		0.759	22.66	0.242	0.645	39.28	3.120	<u>40.91</u>	2
V6		✓	✓	✓		✓	<u>0.774</u>	22.95	0.241	0.653	<u>39.35</u>	2.847	44.73	4
Ours		✓	✓	✓	✓	✓	0.812	<u>22.91</u>	0.242	<u>0.654</u>	39.37	<u>2.886</u>	37.28	7

and iteration, we decouple these factors and design two variants of our method, that is, Ours-i (iteration) and Ours-p (particles). In Fig. 5, our results can first save nFE because we conduct early stopping. Second, the curves of our results consistently surpass the baselines, indicating the consistent superiority of *SwiftExplorer*.

3.5 Ablation Study

Overall Ablation. *SwiftExplorer* aims to balance the trade-offs between diversity-fidelity and quality-efficiency, and thus can be divided into four categories based on the objectives. Specifically, Diversity (Dive) includes directly adding exploration and adding Monte Carlo exploration. Fidelity (Fide) refers to L2 constraints

for regularization. Efficiency (Effi) includes the I&R and early stopping. Guidance Screening aims to enhance Quality (Qual). Then, we conducted detailed ablations in Tab. 4, where the proposed diversity-fidelity tradeoff effectively improves the diversity of generated images while maintaining consistency. Meanwhile, the I&R mechanism and early stopping greatly reduce computational overhead with almost no impact on performance. GuiS improves inference efficiency and performance by removing invalid guidance. In Appendix. 3.7, we further analyze the mutual correlation among different proposed components based on the ablation results.

Analysis on Generative Diversity. The generative diversity is compared among SD-XL, Diff-DPO, DyMO, and Ours by generating images with 40 seeds. As shown in Fig. 8, our method exhibits improved diversity while maintaining the impressive preference score

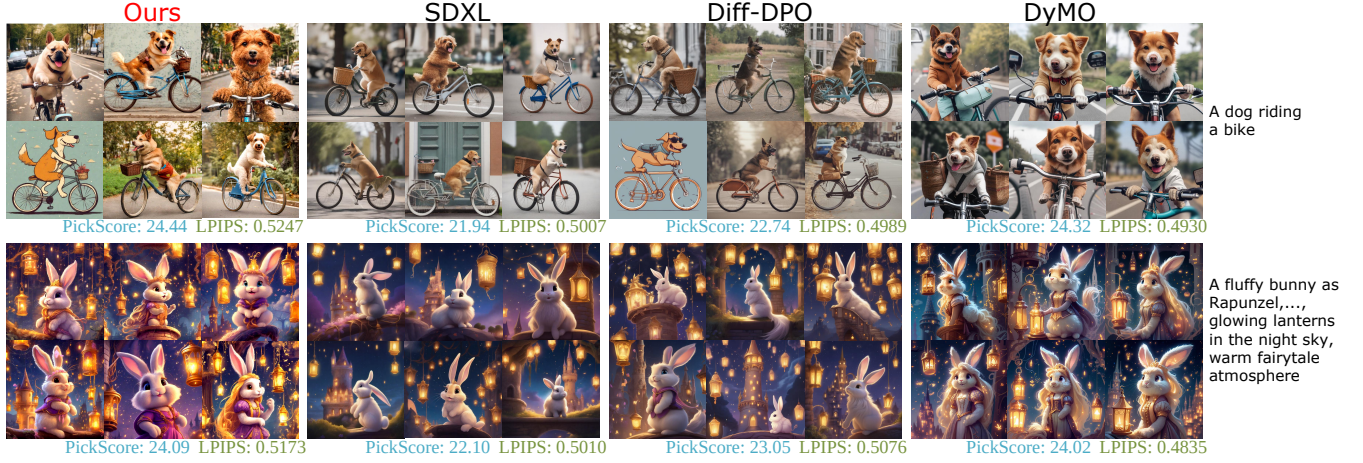


Figure 8: Generative diversity examples with the same prompt and different seeds.

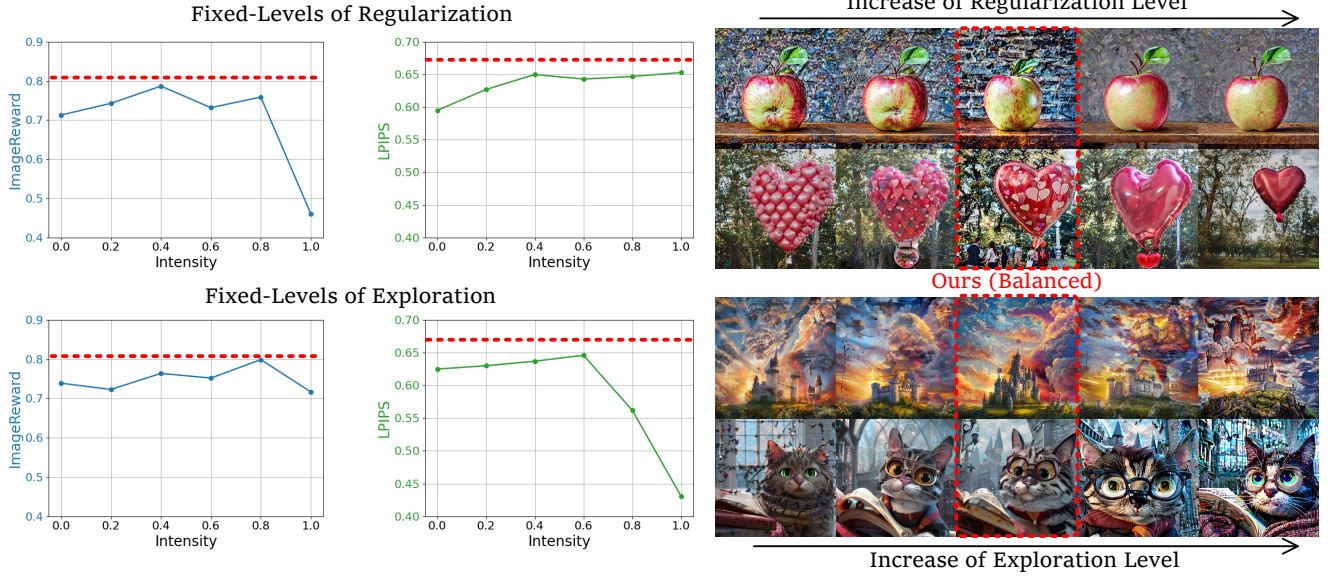


Figure 9: Diversity-fidelity adaptive trade-off (Ours, red dotted lines) with different fixed intensities.

and precise prompt-image alignment. We have also investigated exploration diversity during inference in Appendix 3.6.

Analysis on Diversity-Fidelity Trade-off. In Fig. 9, we examine the Diversity-Fidelity Trade-off by fixing exploration and regularization levels instead of using dynamic p_t . Results show that stronger regularization undermines preference without enhancing diversity, while higher exploration reduces diversity through image distortion. In contrast, our method consistently surpasses baselines and achieves a balanced trade-off between richness and alignment.

3.6 User Study

To subjectively evaluate the generative quality of different methods, we conduct a user study based on the images from the Plug-and-Play experiments in Sec. 3.3. Specifically, we recruit five subjects without prior knowledge to select the preferred image within paired

images with the corresponding prompt as the reference. As shown in Fig. 6, our method consistently performs as an effective plugin that achieves superior quality compared to the implemented original models. User Study details are provided in Appendix 3.10.

4 CONCLUSION

In this work, we address the challenges in aligning diffusion models with specific objectives without incurring prohibitive fine-tuning costs. We introduce *SwiftExplorer*, a plugin that mitigates distribution collapse and enhances efficiency through an Inheritance-Restart exploration mechanism and a Quality-Efficiency arbitration framework. These innovations prevent early convergence, increase exploration for high-reward trajectories, and improve guidance by eliminating irrelevant signals. Experiments show consistent gains in quality, diversity, and efficiency over existing methods.

REFERENCES

- [1] Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bamberick, et al. 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630, 8016 (2024), 493–500.
- [2] Arpit Bansal, Hong-Min Chu, Avi Schwarzschild, Soumyadip Sengupta, Micah Goldblum, Jonas Geiping, and Tom Goldstein. 2023. Universal guidance for diffusion models. In *Computer Vision and Pattern Recognition*. 843–852.
- [3] Kevin Black, Michael Janner, Yilun Du, Ilya Kostrikov, and Sergey Levine. 2023. Training diffusion models with reinforcement learning. *arXiv preprint arXiv:2305.13301* (2023).
- [4] Nanxin Chen, Yu Zhang, Heiga Zen, Ron J Weiss, Mohammad Norouzi, and William Chan. 2020. Wavegrad: Estimating gradients for waveform generation. *arXiv preprint arXiv:2009.00713* (2020).
- [5] Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. 2022. Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687* (2022).
- [6] Kevin Clark, Paul Vicol, Kevin Swersky, and David J Fleet. 2023. Directly fine-tuning diffusion models on differentiable rewards. *arXiv preprint arXiv:2309.17400* (2023).
- [7] Niklas Deckers, Julia Peters, and Martin Potthast. 2024. Manipulating embeddings of stable diffusion prompts. In *International Joint Conference on Artificial Intelligence*. 7636–7644.
- [8] Prafulla Dhariwal and Alexander Nichol. 2021. Diffusion models beat gans on image synthesis. *Advances in Neural Information Processing Systems* 34 (2021), 8780–8794.
- [9] Xin Ding, Lei Yu, Xin Li, Zhijun Tu, Hanjing Chen, Jie Hu, and Zhibo Chen. 2025. RaSS: Improving Denoising Diffusion Samplers with Reinforced Active Sampling Scheduler. In *Computer Vision and Pattern Recognition Conference*. 12923–12933.
- [10] Ying Fan, Olivia Watkins, Yuqing Du, Hao Liu, Moonkyung Ryu, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, Kangwook Lee, and Kimin Lee. 2023. Dpok: Reinforcement learning for fine-tuning text-to-image diffusion models. *Advances in Neural Information Processing Systems* 36 (2023), 79858–79885.
- [11] Ying Fan, Olivia Watkins, Yuqing Du, Hao Liu, Moonkyung Ryu, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, Kangwook Lee, and Kimin Lee. 2024. Reinforcement learning for fine-tuning text-to-image diffusion models. *Advances in Neural Information Processing Systems* 36 (2024).
- [12] Dhruva Ghosh, Hannaneh Hajishirzi, and Ludwig Schmidt. 2023. Geneval: An object-focused framework for evaluating text-to-image alignment. *Advances in Neural Information Processing Systems* 36 (2023), 52132–52152.
- [13] Yi Gu, Zhendong Wang, Yueqin Yin, Yujia Xie, and Mingyuan Zhou. 2024. Diffusion-rpo: Aligning diffusion models through relative preference optimization. *arXiv preprint arXiv:2406.06382* (2024).
- [14] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in Neural Information Processing Systems* 30 (2017).
- [15] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems* 33 (2020), 6840–6851.
- [16] Jonathan Ho and Tim Salimans. 2022. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598* (2022).
- [17] Francisco Ibarrola and Kazjon Grace. 2024. Measuring diversity in co-creative image generation. *arXiv preprint arXiv:2403.13826* (2024).
- [18] Sunwoo Kim, Minkyu Kim, and Dongmin Park. 2025. Test-time alignment of diffusion models without reward over-optimization. *arXiv preprint arXiv:2501.05803* (2025).
- [19] Yuval Kirstain, Adam Polyak, Uriel Singer, Shihbuland Matiana, Joe Penna, and Omer Levy. 2023. Pick-a-pic: An open dataset of user preferences for text-to-image generation. *Advances in Neural Information Processing Systems* (2023).
- [20] Zhifeng Kong, Wei Ping, Jiaji Huang, Kexin Zhao, and Bryan Catanzaro. 2020. Diffwave: A versatile diffusion model for audio synthesis. *arXiv preprint arXiv:2009.09761* (2020).
- [21] Tuomas Kynkäänniemi, Tero Karras, Samuli Laine, Jaakko Lehtinen, and Timo Aila. 2019. Improved precision and recall metric for assessing generative models. *Advances in Neural Information Processing Systems* 32 (2019).
- [22] Haodong Li, Tianfei Ren, Xiaoxiao Ma, Chunmei Qing, Zhen Fang, Sipeng He, Ziyu Guo, Haoyu Wu, Juanxi Tian, Yihang Zou, et al. 2026. VideoCoCo: Code-as-CoT for Physically-Consistent Video Generation via an Agentic Dual-Engine System. *arXiv preprint arXiv:2607.27380* (2026).
- [23] Shufan Li, Konstantinos Kallidromitis, Akash Gokul, Yusuke Kato, and Kazuki Kozuka. 2024. Aligning diffusion models by optimizing human utility. *Advances in Neural Information Processing Systems* 37 (2024), 24897–24925.
- [24] Yuqi Li, Zijie Zhou, Zhiyuan Peng, Junhao Dong, Haochen You, Renye Yan, Shiping Wen, Yingli Tian, and Tingwen Huang. 2025. A preference-driven methodology for efficient code generation. *IEEE Transactions on Artificial Intelligence* (2025).
- [25] Zhanhao Liang, Yuhui Yuan, Shuyang Gu, Bohan Chen, Tiankai Hang, Ji Li, and Liang Zheng. 2024. Step-aware Preference Optimization: Aligning Preference with Denoising Performance at Each Step. *arXiv preprint arXiv:2406.04314* (2024).
- [26] Zeyu Ling, Xinyao Yu, Renye Yan, Jikang Cheng, Zhanke Wang, Qing Shuai, and Changqing Zou. 2026. GenTrack: Physical Alignment for Robot-Native Motion Generation and Zero-Shot Humanoid Tracking. *arXiv preprint arXiv:2608.01410* (2026).
- [27] Jinglin Liu, Chengxi Li, Yi Ren, Feiyang Chen, and Zhou Zhao. 2022. Diffsinger: Singing voice synthesis via shallow diffusion mechanism. In *AAAI conference on artificial intelligence*, Vol. 36. 11020–11028.
- [28] Lixiong Liu, Bao Liu, Hua Huang, and Alan Conrad Bovik. 2014. No-reference image quality assessment based on spatial and spectral entropies. *Signal Processing: Image communication* 29, 8 (2014), 856–863.
- [29] Liangsi Lu, Xuhang Chen, Minzhe Guo, Shichu Li, Jingchao Wang, and Yang Shi. 2026. Chordedit: One-step low-energy transport for image editing. *arXiv preprint arXiv:2602.19083* (2026).
- [30] Nanye Ma, Shangyuan Tong, Haolin Jia, Hexiang Hu, Yu-Chuan Su, Mingda Zhang, Xuan Yang, Yandong Li, Tommi Jaakkola, Xuhui Jia, et al. 2025. Inference-time scaling for diffusion models beyond scaling denoising steps. *arXiv preprint arXiv:2501.09732* (2025).
- [31] Qichao Ma, Jikang Cheng, Ling Liang, Zhaofei Yu, Tiejun Huang, and Renye Yan. 2026. Can We Perform Online RL for Image Editing without Editing Rewards? *arXiv preprint arXiv:2608.22780* (2026).
- [32] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. 2012. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing* 21, 12 (2012), 4695–4708.
- [33] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. 2012. Making a “completely blind” image quality analyzer. *IEEE Signal Processing Letters* 20, 3 (2012), 209–212.
- [34] Zhuoran Qiao, Weili Nie, Arash Vahdat, Thomas F Miller III, and Anima Anandkumar. 2022. Dynamic-backbone protein-ligand structure prediction with multi-scale generative diffusion models. *arXiv preprint arXiv:2209.15171* 1 (2022).
- [35] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. [n. d.]. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*. PmLR.
- [36] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125* 1, 2 (2022), 3.
- [37] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *Computer Vision and Pattern Recognition*. 10684–10695.
- [38] Peter J Rousseeuw. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20 (1987), 53–65.
- [39] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. 2016. Improved techniques for training gans. *Advances in Neural Information Processing Systems* 29 (2016).
- [40] Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, et al. 2022. Laion-5b: An open large-scale dataset for training next generation image-text models. *Advances in Neural Information Processing Systems* (2022).
- [41] Hao Shen, Jikang Cheng, Renye Yan, Zhongyuan Wang, Wei Peng, and Baojin Huang. 2025. When Generative Replay Meets Evolving Deepfakes: Dual Confusion-Aware Regularization for Incremental Face Forgery Detection. *arXiv preprint arXiv:2511.18436* (2025).
- [42] Raghu Singh, Zachary Horvitz, Ryan Teehan, Mengye Ren, Zhou Yu, Kathleen McKeown, and Rajesh Ranganath. 2025. A general framework for inference-time scaling and steering of diffusion models. *arXiv preprint arXiv:2501.06848* (2025).
- [43] Richard S Sutton. 2018. Reinforcement learning: An introduction. *A Bradford Book* (2018).
- [44] Zhiwei Tang, Jiangweizhi Peng, Jiasheng Tang, Mingyi Hong, Fan Wang, and Tsung-Hui Chang. 2024. Inference-Time Alignment of Diffusion Models with Direct Noise Optimization. *arXiv preprint arXiv:2405.18881* (2024).
- [45] Bram Wallace, Meihua Dang, Rafael Rafailov, Linqi Zhou, Aaron Lou, Senthil Purushwalkam, Stefano Ermon, Caiming Xiong, Shafiq Joty, and Nikhil Naik. 2024. Diffusion model alignment using direct preference optimization. In *Computer Vision and Pattern Recognition*. 8228–8238.
- [46] Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach, Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. 2023. De novo design of protein structure and function with RFdiffusion. *Nature* 620, 7976 (2023), 1089–1100.
- [47] Xiaoshi Wu, Yiming Hao, Keqiang Sun, Yixiong Chen, Feng Zhu, Rui Zhao, and Hongsheng Li. 2023. Human preference score v2: A solid benchmark for evaluating human preferences of text-to-image synthesis. *arXiv preprint arXiv:2306.09341* (2023).
- [48] Xiaoshi Wu, Yiming Hao, Manyuan Zhang, Keqiang Sun, Zhaoyang Huang, Guanglu Song, Yu Liu, and Hongsheng Li. 2024. Deep reward supervisions for tuning text-to-image diffusion models. In *European Conference on Computer*

- Vision. Springer, 108–124.
- [49] Xin Xie and Dong Gong. 2025. DyMO: Training-Free Diffusion Model Alignment with Dynamic Multi-Objective Scheduling. In *Computer Vision and Pattern Recognition*. 13220–13230.
 - [50] Jiazheng Xu, Xiao Liu, Yuchen Wu, Yuxuan Tong, Qinkai Li, Ming Ding, Jie Tang, and Yuxiao Dong. 2023. Imagereward: Learning and evaluating human preferences for text-to-image generation. *Advances in Neural Information Processing Systems* (2023).
 - [51] Minkai Xu, Alexander S Powers, Ron O Dror, Stefano Ermon, and Jure Leskovec. 2023. Geometric latent diffusion models for 3d molecule generation. In *International Conference on Machine Learning*. PMLR, 38592–38610.
 - [52] Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. 2022. Geodiff: A geometric diffusion model for molecular conformation generation. *arXiv preprint arXiv:2203.02923* (2022).
 - [53] Renye Yan, Jikang Cheng, Yaozhong Gan, Shikun Sun, You Wu, Yunfan Yang, Liang Ling, Jinlong Lin, Yeshuang Zhu, Jie Zhou, et al. 2025. Entropy-adaptive diffusion policy optimization with dynamic step alignment. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 1924–1934.
 - [54] Renye Yan, Jikang Cheng, Shikun Sun, Yi Sun, You Wu, Wei Peng, Zongwei Wang, Ling Liang, Junliang Xing, and Yimao Cai. 2026. Do Less, Achieve More: Do We Need Every-Step Optimization for RL Fine-tuning of Diffusion Models? *arXiv preprint arXiv:2605.15855* (2026).
 - [55] Renye Yan, Jikang Cheng, You Wu, Ling Liang, Wei Peng, Athanasios V Vasilakos, Qingyu Zhao, Yu Zhang, Ehsan Adeli, Kilian M Pohl, et al. 2026. Pixel-Space Diffusion Transformers. *arXiv preprint arXiv:2607.17585* (2026).
 - [56] Kai Yang, Jian Tao, Jiafei Lyu, Chunjiang Ge, Jiaxin Chen, Weihao Shen, Xiaolong Zhu, and Xiu Li. 2024. Using human feedback to fine-tune diffusion models without any reward model. In *Computer Vision and Pattern Recognition*. 8941–8951.
 - [57] Huizhuo Yuan, Zixiang Chen, Kaixuan Ji, and Quanquan Gu. 2024. Self-play fine-tuning of diffusion models for text-to-image generation. *Advances in Neural Information Processing Systems* 37 (2024), 73366–73398.
 - [58] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Computer Vision and Pattern Recognition*. 586–595.
 - [59] Xiangcheng Zhang, Haowei Lin, Haotian Ye, James Y Zou, Jianzhu Ma, Yitao Liang, and Yilun Du. 2026. Inference-time scaling of diffusion models through classical search. In *International Conference on Learning Representations*, Vol. 2026. 39133–39166.
 - [60] Xueyi Zhang, Chengwei Zhang, Zheng Li, Xiyu Wang, Siqi Cai, Mingrui Lao, Yanming Guo, and Huiping Zhuang. 2026. Rep deep & machine learning: Exemplar-free continual video action recognition via slow-fast collaborative learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 40. 36075–36083.

APPENDIX

1 RELATED WORK

1.1 Diffusion Model

Diffusion models have achieved remarkable progress in generative modeling. Their stability and gradual refinement allow them to approximate complex distributions, producing samples with both high fidelity and diversity. They have shown strong generalization across domains, including text-to-image [9, 23, 57], speech [4, 20, 27], molecular design [51, 52], and protein generation [1, 34, 46]. However, since these models are typically pretrained on broad datasets lacking task-specific signals, they often fail to reflect user-desired distributions in specialized applications. This gap highlights the need for alignment: enabling customized optimization toward target distributions without losing generality.

1.2 Gradient-based Fine-tuning Methods

Direct gradient-based fine-tuning is intuitively feasible for enhancing diffusion models. For examples, DRaFT [6] performs differentiable reward optimization for diffusion models by backpropagating reward gradients through sampling steps, forming a unified gradient-based framework for preference-aligned fine-tuning; DR-Tune [48] stabilizes reward-supervised diffusion training by applying stop-gradient to denoiser inputs and uniformly sampling K timesteps, enabling efficient supervision across all sampling depths; Diffusion-RPO [13] extends relative preference optimization to diffusion models by conducting stepwise preference contrasts with CLIP-based multimodal weighting to learn coherent cross-semantic preference patterns. All these methods improve preference alignment with limited stability, scalability, and robustness under diverse real-world undifferentiated reward signals.

1.3 Fine-tuning Methods

Existing alignment approaches fall into two categories: fine-tuning [3, 10] and training-free methods. The mainstream fine-tuning, for example, relies on reinforcement learning (RL) [43] to adjust diffusion model weights. RL greatly improves alignment to downstream targets but faces challenges: rewards are only available after the full denoising process, making them sparse and prone to reward hacking. This often leads to higher reward scores but reduced diversity and poor distribution coverage. Moreover, RL-based fine-tuning comes with a high computational cost [18].

1.4 Traing-free Methods

Training-free methods [18, 49] provide an alternative: instead of changing diffusion model weights, they bias the sampling process at inference to guide results toward target objectives. For examples, TTScale [30] boosts generation quality by shifting extra inference compute from longer denoising chains to noise-search guided by a lightweight evaluator. FKSteer [42] enables inference-time controllability by using particle-based weighting and resampling to steer diffusion trajectories without fine-tuning. This avoids costly retraining and reduces computation. First, prior work lacks an adaptive mechanism to balance diversity and fidelity [8]. Second, existing training-free methods often overemphasize target guidance, forcing trajectories into narrow regions and reducing coverage of the

pretrained distribution. While less prone to severe reward hacking, this effect still harms diversity. Further, guidance signals are typically applied indiscriminately across all steps [44], without filtering ineffective or saturated signals. This overuse increases inference cost and amplifies noise, sometimes even causing negative guidance effects.

2 METHOD SUPPLEMENT

2.1 Control Network

We assign $y = 0$ to samples close to pure noise $\mathbf{x}_T$, and $y = 1$ to samples close to the target $\mathbf{x}_0$, pretraining the network via the loss defined as follows.

$$\mathcal{L}(\theta) = -\mathbb{E}_{(\mathbf{x}_t, y)} [y \log f_\theta(\mathbf{x}_t) + (1 - y) \log(1 - f_\theta(\mathbf{x}_t))]. \quad (13)$$

2.2 Generalizability of Pre-trained Control Network

Considering training a control network could undermine the claim of “training-free” of the proposed *SwiftExplorer*, it is necessary to demonstrate that the Control Network could also be seen as a pre-trained scorer, which can be pre-trained once and be generalizable for all. Therefore, as shown in Tab. 8, we pre-train the control network on the simple animal set and predict steps on other datasets. The results demonstrate that the network could easily generalize to other datasets with similar effective performance, which is caused by the fixed scheduler having already pre-set the noise level for different steps in a coarse manner.

3 ADDITIONAL EXPERIMENT RESULTS

3.1 Results with alternative optimization targets.

We also conduct experiments by replacing PickScore with Aesthetic score as the optimization target to evaluate the adaptability of *SwiftExplorer*. As shown in Fig. 11, guidance with the aesthetic score demonstrates our overall superiority in aesthetic preference and generative diversity. Then, in the visual results, it could be clearly observed that our generated images contain richer details, better alignment, and improved aesthetic impressions.

3.2 Detailed hyper-parameter setting

The full list of hyper-parameters in our paper is shown in Table 5.

3.3 Metric

We provide a detailed illustration of each introduced metric:

- **Aesthetic Preference**

- **AES (Aesthetic Score)** AES is a metric used for evaluating the aesthetic quality of an image, typically by training a model to predict human aesthetic ratings [40].
- **PS (PickScore)** PS is a method based on human selection preferences for rating the aesthetic quality of images [19].
- **IR (ImageReward)** IR is a metric used for assessing image quality, often in the optimization process of image generation models [50].



Figure 10: Showcase of generated samples. The images from our model achieve state-of-the-art performance.

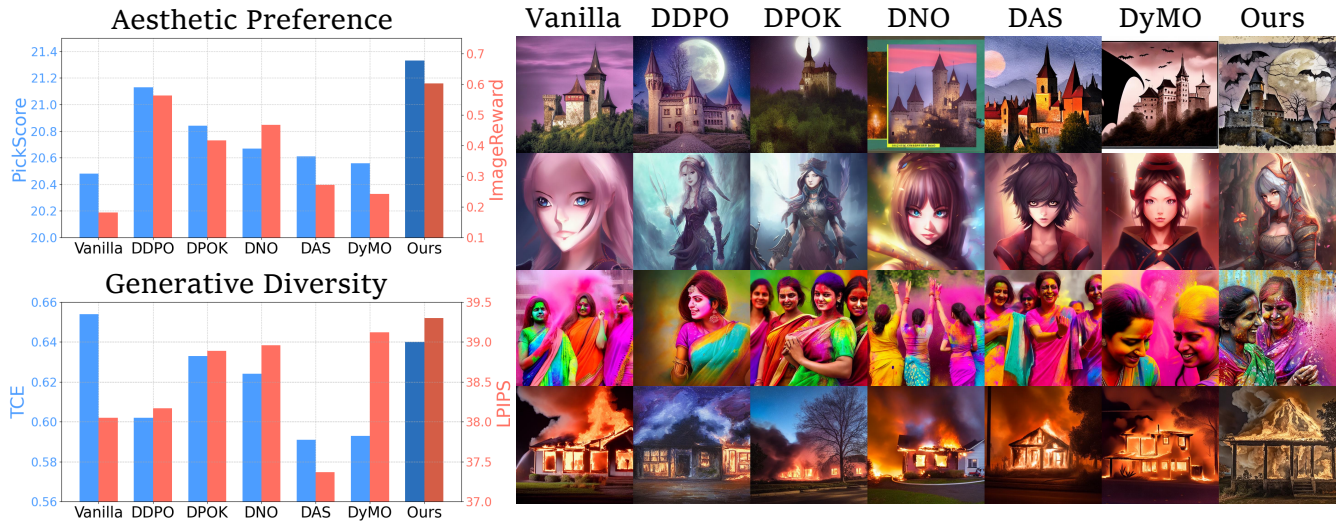


Figure 11: Results with Aesthetic Score as the optimization objective.

- **HPSv2 (Human Preference Score v2)** HPSv2 is a preference-based metric that predicts human preferences for generated images. This model fine-tunes the CLIP model on the HPD v2 dataset. HPSv2 excels in multiple styles, including animation, concept art, paintings, and photographs [47].
- **Image Fidelity**
- **ClipScore** ClipScore is a model-based image scoring metric that evaluates the similarity between a generated image and its textual description. It calculates the cosine similarity between the image and text embeddings. The metric is highly correlated with human judgment and does not require reference text [35].

Table 5: Hyper-parameters and metrics in our experiment. Notably, only n and δ_r are newly introduced by our method; all other hyper-parameters follow DDIM or the baseline methods.

Name	Description	Value
η	eta parameter for the DDIM sampler	1.0
w	classifier-free guidance weight	5.0
m_{dno}	Iteration Number for DNO	20
m_{dyMO}	Iteration Number for DyMO	adaptive
n_{das}	Particle number for DAS	4
mp	mixed precision	fp16
n (<i>SwiftExplorer-unique</i>)	Exploration supplement number	2
δ_r (<i>SwiftExplorer-unique</i>)	Reward enhancement threshold	0.1

- **Fréchet Inception Distance (FID)** FID is a metric used to evaluate the quality of generated images. It compares the mean and covariance of the features extracted from real and generated images using the Inception v3 model. A lower FID value indicates that the generated image is closer to real images [14].
- **Improved F1 Score (iFS)** iFS is an improvement over the traditional F1 score, aiming to provide a better evaluation of generative model performance. It optimizes the balance between precision and recall for more accurate assessments [21].

- **Generative Diversity**

- **LPIPS (Learned Perceptual Image Patch Similarity)** LPIPS is a metric for evaluating perceptual similarity between images. It compares the feature activations of images in pre-trained convolutional neural networks (e.g., VGG). A higher LPIPS value indicates lower perceptual similarity and a higher diversity[58].
- **TCE (Truncated CLIP Entropy)** Truncated CLIP Entropy (TCE) is a measure of semantic diversity of a set of generated images, computed in the joint image–text embedding space of CLIP. It is defined by first mapping each image x_i to a vector $z_i = \text{CLIP}_{\text{img}}(x_i) \in \mathbb{R}^d$, then forming the empirical covariance matrix Σ of these vectors, extracting its top- k eigenvalues $\lambda_1 \geq \dots \geq \lambda_k$, and finally computing

$$\text{TCE}_k = \frac{1}{2} \sum_{i=1}^k \log \lambda_i.$$

TCE gives a tractable proxy for how “spread out” the images are in the CLIP semantic space. [17].

- **Inception Score (IS)** IS is a metric used to evaluate the quality of generated images. It calculates the entropy of

the class distribution of images in a pre-trained Inception v3 model. Higher IS values indicate that the generated images are both clear and diverse [39].

- **Compositional Richness**

- **NIQE (Natural Image Quality Evaluator)** NIQE is a no-reference image quality assessment metric. It evaluates the image quality by measuring the distance between the natural scene statistics of the image and a natural image database of undistorted images [33]. A lower NIQE value indicates better image quality.
- **BRISQUE (Blind/Referenceless Image Spatial Quality Evaluator)** BRISQUE is another no-reference image quality metric. It evaluates image quality by analyzing spatial-domain features. It does not rely on reference images and is suitable for a variety of image quality assessment tasks [32].
- **Spectral Entropy (SE)** SE is a metric for evaluating the spectral properties of images. It calculates the entropy value of an image’s frequency spectrum to assess the complexity and richness of image details. Higher SE values generally indicate more detailed and textured images [28].

Table 7: CoV (%) comparison of methods across four evaluation metrics.

Method	Preference	Fidelity	Diversity	Richness
DNO	0.88	0.52	0.27	1.27
DAS	0.96	0.81	0.30	1.36
DyMO	0.62	0.86	0.28	1.15
Ours	0.65	0.79	0.28	1.41

Table 6: Results of Restart Count and Early Stop across HPSv2 and Pick-a-Pic with SDv15 and SD-XL backbones.

Avg. Value	Base	HPSv2		Pick-a-Pic	
		SDv15	SD-XL	SDv15	SD-XL
Restart Count	50	4.37	8.55	7.32	10.59
Guiding Step	50	37.35	44.21	40.67	44.30

3.4 Supplementary Statistical Significance

Since the generative quality is highly related to the initial noise decided by the seeds, we further conduct experiments with 5 different seeds and calculate their average Coefficient of Variation (CoV). The metrics and evaluation protocol are strictly following Tab. 1. As shown in Tab. 7, it can be observed that all methods exhibit similarly low CoV across various seeds, which indicates that the guidance-based training-free scaling methods perform relatively

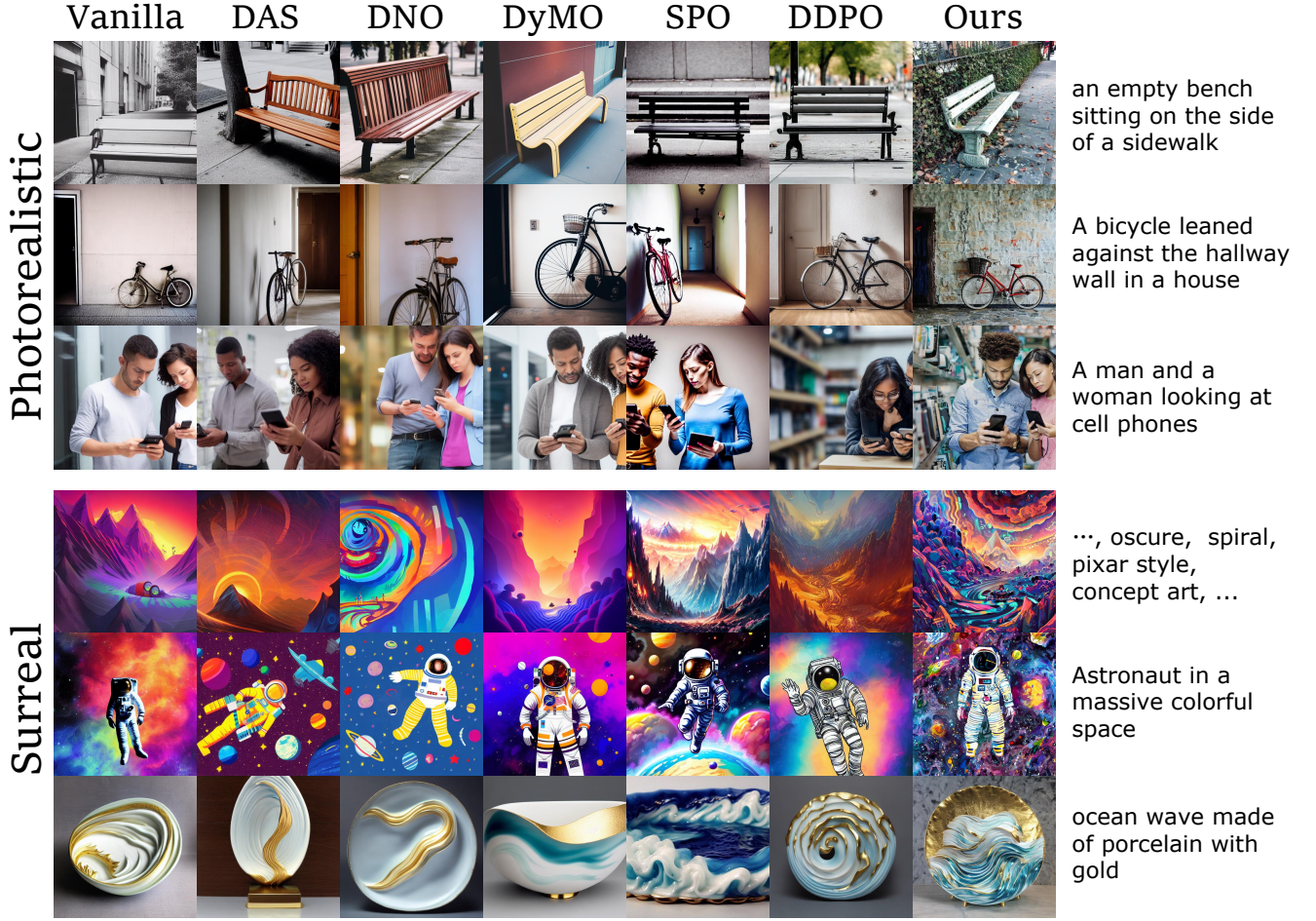


Figure 12: The visual comparisons of different methods on SDv15. For photorealistic prompts from HPSv2, our results contain richer details while maintaining image-prompt alignment. For surreal prompts from Pick-a-Pic, we exhibit much more diverse patterns and colors.

stably. These results further demonstrate that the performance reported in Tab. 1 is statistically significant.

Table 8: Mean step numbers predicted by the control network across different datasets. The control network is trained on the simple animal and directly generalizes to other datasets.

Method	Total	SimpleAnimal	Pick-a-Pic	HPSv2	GenEval
SD15	50	37.3	40.9	39.8	38.1
SDXL	30	24.7	24.9	25.1	24.6

3.5 Supplementary Distribution Comparison with DAS

In Fig. 1, we compare the feature distribution with the iteration-based DyMO, thus demonstrating the diversity of our method. Here, we conduct a similar experiment to evaluate the sampling-based

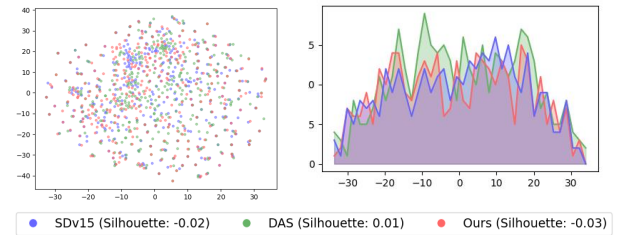


Figure 13: Feature visualization comparison with DAS.

baseline, *i.e.*, DAS. As shown in Fig. 13, although the sampling-based DAS has richer diversity compared to DyMO, it still tightens the distribution range of the original backbone. In contrast, our result exhibits consistent diversity, which is demonstrated by the wider visualized distribution and the lower Silhouette Coefficient value.



Figure 14: The proposed ESV facilitates exploration during the denoising process. Left: ESV enables our method to identify a broader range of patterns and adjust the trajectory toward the “teddy bear lantern” during intermediate stages. Right: Our method exhibits greater posture variation throughout denoising. While this variation is not directly responsible for the superior final generation, it reflects an increased level of exploration.

3.6 Analysis on Exploration.

To demonstrate the exploration effect of introducing ESV, we present the intermediate results during generation. As shown in Fig. 14, our method exhibits superior exploration effect during denoising, implying the improved generative quality.

3.7 Further Analysis on Ablation Study

The proposed *SwiftExplorer* is designed considering two trade-offs, that is, Diversity-Fidelity and Efficiency-Quality trade-offs. Therefore, different components are introduced with paired correlations. Specifically, introducing M-ESV can introduce exploration and thus improve diversity, while L2 can maintain the fidelity to prevent over-exploration that leads to the collapse of the denoising process. Therefore, without ESV that encourages exploration, solely deploying L2 is unnecessary. Then, both I&R and EarS are introduced to enhance the generation quality based on the exploration. Meanwhile, GuiS builds upon the previously introduced exploration, as direct guidance typically provides a straightforward path to higher rewards, making GuiS ineffective in such cases. In contrast, with ESV, guidance is not based on the current state but rather incorporates an exploratory shift, allowing GuiS to effectively filter out harmful optimization. As for the intensity of different components in the trade-offs, we present Fig. 9 to demonstrate the superiority of our adaptive strategy. Overall, the components of our approach are interdependent and mutually reinforcing, collectively contributing to the superior performance of our image generation scheme in terms of inference-time scaling.

3.8 Results on more SD backbones.

In addition to SDv15 and SD-XL presented in the main paper, we provide the results on SDv14 and SDv21-turbo. As shown in Tab. 9 and Tab. 10, our method also leads in all metrics. To further validate robustness on larger and stronger backbones, we additionally report results on advanced models in Tab. 11. The same trend remains clear: our method consistently improves preference-related

quality while preserving fidelity and diversity. Compared with optimization-heavy baselines, it achieves a better balance between visual richness and prompt alignment, indicating that the plugin design remains effective when scaling to stronger pretrained models.

3.9 Exhibited adaptive values for computation reduction.

Here, we provide the exhibited average values of restart counts and early stop positions. Specifically, we calculate the average restart counts and average number of guiding steps on HPSv2 and Pick-a-Pic datasets with SDv15 and SD-XL. As shown in Tab. 6, it can be observed that HPSv2 has smaller values of both variables compared to Pick-a-Pic, while SDv15 has smaller values than SD-XL. This is because HPSv2 has more photorealistic images that are easier than surreal one for denoising, while SDv15’s lesser robustness makes it more amenable to external guidance.

3.10 User Study Details.

In this study, we assessed the plug-and-play effectiveness of our method through a human evaluation, where five evaluators compared our method to four competing models: DNO, DyMO XL, Gdt, and DyMO. The evaluation was method-blind, with participants unaware of which model generated each image. The evaluation criteria is:

- **Win:** Our method was preferred.
- **Tie:** Both methods were considered equally effective, hard to decide distinguishable preference.
- **Lose:** Our method was less preferred.

Each participant compared the generated images based on overall quality, with the results shown in Fig. 6. The preference for our method across different models demonstrates its superior performance in plug-and-play integration.

Table 9: Results with SDv14.

Method	IR	PS	CLIP	LPIPS	TCE	NIQE↓
SDv14	0.744	20.62	0.240	0.671	40.32	4.6312
DDPO	1.056	21.20	0.240	0.609	39.61	4.554
DNO	0.915	21.85	0.239	0.597	39.24	4.875
DAS	1.217	22.58	0.238	0.610	39.56	5.451
DyMO	1.100	23.14	0.235	0.615	39.63	4.550
Ours	1.230	23.39	0.242	0.676	41.10	3.494

Table 11: Results with advanced models.

Method	IR	PS	CLIP	LPIPS	TCE	NIQE↓
SD-XL	0.715	21.45	0.236	0.675	40.18	4.553
Diff-DPO	1.010	22.05	0.236	0.651	39.45	4.440
SPO	1.179	22.81	0.232	0.556	40.42	3.900
SDv35	1.193	21.90	0.238	0.621	39.42	4.873
DNO	0.924	22.58	0.238	0.581	39.34	4.783
DAS	1.171	23.07	0.237	0.594	39.36	5.365
DyMO	1.079	24.34	0.233	0.609	39.55	4.510
Ours	1.201	24.56	0.236	0.679	40.99	3.573

4 ALGORITHM OF THE PROPOSED *SWIFTEXPLORER*

For improved delivery of the proposed *SwiftExplorer*, we show its detailed algorithm in Alg. 1.

5 SHOWCASE PROMPT TABLE

Please refers to Tab. 12, 14, 15, 16, and 13.

Table 10: Results with SDv21.

Method	IR	PS	CLIP	LPIPS	TCE	NIQE↓
SDv21	0.541	20.63	0.280	0.644	39.69	5.672
DDPO	0.794	21.25	0.274	0.622	38.94	4.953
DNO	0.566	20.91	0.277	0.589	39.13	5.551
DAS	0.817	22.29	0.280	0.656	39.55	5.150
DyMO	0.921	22.62	0.279	0.623	39.62	5.176
Ours	0.980	22.91	0.280	0.677	40.07	3.695

Algorithm 1: Algorithm of Swift Diversity Exploration (*SwiftExplorer*)

Input: Total Inference Denoising Step T ; Initialized Latent: $\mathbf{x}_T$; Pre-trained score function $\mathbf{s}_\theta(\cdot)$.
Initialize Inheritance Flag $f_{inh} = \text{False}$;
for timestep $t \in \{T, \dots, 1\}$ **do**
 execute one denoising step
 $\mathbf{x}_t = \frac{1}{\sqrt{\alpha_{t+1}}} (\mathbf{x}_{t+1}^* + \beta_{t+1} \mathbf{s}_\theta(\mathbf{x}_{t+1}^*, t+1)) + \sqrt{\beta_{t+1}} \mathbf{z}_{t+1}$.
 considering early-stopping
 $E_{stop}(t) = \mathbb{I}[p_t = 1 \wedge \widehat{\Delta} r_t \leq \delta_r] \in \{0, 1\}$.
 if $E_{stop}(t) == 0$ **then**
 inheritance-restart mechanism for exploration variable
 if f_{inh} **then**
 inherit previous exploration variable
 $\epsilon_t^{(i)} = \mathbf{x}_{t+1}^* - \mathbf{x}_{t+1}$, $i = 1$
 else
 sample new Monte Carlo exploration set $\{\epsilon_t^{(i)}\}$
 $\epsilon_t^{(i)} \sim \mathcal{N}(0, \sigma_t^2 I)$, $i = 1, \dots, n$
 generate candidates with Diversity-Fidelity trade-off
 $\hat{\mathbf{x}}_t^{(i)} = \underbrace{\mathbf{x}_t + (1 - p_t) \cdot \epsilon_t^{(i)} - p_t \cdot \nabla_{\mathbf{x}_t} L_2(\hat{\mathbf{x}}_t, \mathbf{x}_t) + \nabla_{\mathbf{x}_t} \mathcal{R}(\mathbf{x}_{0|t})}_{\text{Diversity-Fidelity Trade-off}}$
 iteratively optimizing candidates
 $\tilde{\mathbf{x}}_t^{(i)} = g(\hat{\mathbf{x}}_t^{(i)}, \mathcal{R} - L_2, m)$
 obtain the optimal candidates
 $\mathbf{x}_t^* = \tilde{\mathbf{x}}_t^{(i^*)}$, $i^* = \arg \max_{i \in \{1, \dots, n\}} \mathcal{R}(\tilde{\mathbf{x}}_{0|t}^{(i)})$.
 verify optimization condition
 $\Delta \mathcal{R}_t^{(i^*)} = \mathcal{R}(\mathbf{x}_{0|t}^*) - \mathcal{R}(\mathbf{x}_{0|t})$,
 if $\Delta \mathcal{R}_t^{(i^*)} > 0$ **then**
 $f_{inh} = \text{True}$
 else
 $\mathbf{x}_t^* = \mathbf{x}_t$
 $f_{inh} = \text{False}$
 Output: Optimized Inference Result $\mathbf{x}_0$.

Table 12: Detailed prompts used for generated images in Fig. 10 - Part 1

Image	Prompt
Fig. 1, Row 1, Col 1	ballet dancer, insanely detailed, photorealistic, 8k, perfect composition, volumetric lighting, natural complexion, award-winning professional photography, taken with Canon EOS 5D Mark IV, 85mm, mindblowing, masterpiece
Fig. 1, Row 1, Col 2	A fluffy bunny as Rapunzel, with long golden ears flowing down from a tall enchanted tower, glowing lanterns in the night sky, warm fairytale atmosphere
Fig. 1, Row 1, Col 3	A dog in sportswear lifting tiny dumbbells at the gym, determined expression, humorous fitness illustration
Fig. 1, Row 1, Col 4	A boy superhero landing on the ground in classic "hero pose," debris and glowing sparks flying around, comic action shot
Fig. 1, Row 1, Col 5	A phoenix rising up from ashes
Fig. 1, Row 2, Col 1	A young girl standing on a rooftop, blowing dandelions that transform into glowing comets, shooting across the night sky, dreamy fantasy artwork
Fig. 1, Row 2, Col 2	A small hedgehog as the Frog Prince, wearing a tiny crown while sitting on a lilypad, a kind-hearted swan princess leaning close, surrounded by glowing fireflies, magical fairytale illustration
Fig. 1, Row 2, Col 3	a white polar bear cub wearing sunglasses sits in a meadow with flowers.
Fig. 1, Row 2, Col 4	A cat surfing on a giant wave at sunset, wearing cool shades, cinematic sports illustration
Fig. 1, Row 2, Col 5	A sunflower in full bloom under golden sunlight, tiny dewdrops sparkling on its petals, cinematic macro fantasy illustration

Table 14: Detailed prompts used for generated images in Fig. 10 - Part 2

Image	Prompt
Fig. 1, Row 3, Col 1	A warrior standing at the edge of a glowing crater, surrounded by swirling cosmic energy, their silhouette outlined against the birth of a new star, ultimate epic fantasy art
Fig. 1, Row 3, Col 2	A boy lying on the grass in a field, listening to music with glowing headphones, fireflies surrounding him
Fig. 1, Row 3, Col 3	A little girl painting a rainbow bridge from the classroom window into the sky, playful magical fairytale art, hopeful and inspiring
Fig. 1, Row 3, Col 4	A group of playful penguins throwing glowing snowballs at each other, each snowball turning into sparkling stars when it explodes, magical fairytale scene
Fig. 1, Row 3, Col 5	Giant rubber duck floating in the ocean with a small island on its back, surrounded by tropical palm trees and crystal clear water, bright and sunny day, calm seas, vivid colors, cinematic lighting, high detail
Fig. 1, Row 3, Col 6	A cozy library built inside an ancient oak tree, warm lights glowing through round windows, whimsical fairytale healing atmosphere
Fig. 1, Row 3, Col 7	A brave boy carrying a glowing lantern, releasing trails of light that form a golden sunrise, cinematic epic fantasy style

Table 15: Detailed prompts used for generated images in Fig. 11

Image	Prompt
Fig. 11, Row 1	transylvania castle on hiltop and dusk bats and scifi
Fig. 11, Row 2	fantasy character portrait digital painting, anime style, detailed with beautiful emotive lighting suggesting personality, and background that suggests character backstory
Fig. 11, Row 3	Women in Saree playing Holi
Fig. 11, Row 4	a house burning at night

Table 13: Detailed prompts used for generated images in Fig. 9

Image	Prompt
Fig. 9, Row 1	Heart shaped balloon
Fig. 9, Row 2	An apple on a table
Fig. 9, Row 3	A castle in the sky, clouds, sunset, explosion
Fig. 9, Row 4	Harry potter as a cat, pixar style, octane render, HD, high-detail

Table 16: Detailed prompts used for generated images in Fig. 7

Image	Prompt
Fig. 7, Row 1	a slice of pizza floating through space with stars in the background
Fig. 7, Row 2	A Great Dane dog in the style of Vincent Van Gogh