Probability-Conserving Flow Guidance

Parsa Esmati[*,1] **Junha Hyung**[*,2] **Amirhossein Dadashzadeh**[1]
Jaegul Choo[2] **Majid Mirmehdi**[1]

[1] University of Bristol [2] KAIST

[*] Equal contribution



Figure 1: **AdaMaG vs. CFG.** CFG (top) vs AdaMaG (bottom), same prompts and seeds. CFG shows saturation and hallucinated artefacts that grow with guidance scale; AdaMaG limits departure from the learned density path at the same scale, yielding cleaner generations with no inference overhead.

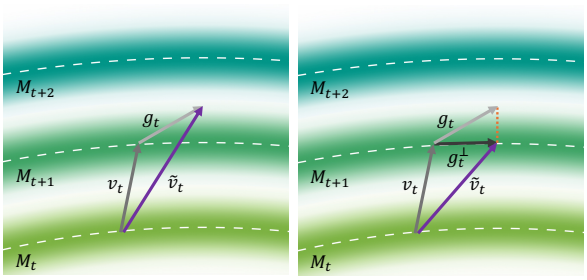
Abstract

Diffusion and flow-based generative models dominate visual synthesis, with guidance aligning samples to user input and improving perceptual quality. However, Classifier-Free Guidance (CFG) and extrapolation-based methods are heuristic linear combinations of velocities/scores that ignore the generative manifold geometry, breaking probability conservation and driving samples off the learned manifold under strong guidance. We analyse guidance through the continuity equation and show its effect decomposes into a divergence term and a score-parallel term defined invariantly across parameterizations. We prove the divergence term blows up structurally as sampling approaches the data manifold, motivating a time-dependent schedule alongside score-parallel attenuation. The resulting plug-and-play rule, Adaptive Manifold Guidance (AdaMaG), targets both terms at no additional inference cost. Finally, we show that several existing saturation and manifold preserving heuristics can be read as acting on one of these two terms. On image generation benchmarks, AdaMaG improves realism, reduces visible artefacts, and induces controlled desaturation under strong guidance. Code is available here.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

1 Introduction

We consider the problem of guided sampling from high-dimensional data distributions with diffusion and flow-matching models, where the generative process is realised by integrating learned ordinary differential equations (ODEs) or stochastic differential equations (SDEs) defined by a time-dependent velocity (or score) field [Song et al., 2021, Lipman et al., 2022, Rombach et al., 2022, Albergo and Vanden-Eijnden, 2023, Esser et al., 2024, Wan et al., 2025].



(a) CFG Guidance (b) AdaMaG (Ours)

Figure 2: **Conceptual overview.** Unlike CFG (a), which extrapolates from the unconditional field and drifts off the manifold M_t, AdaMaG (b) attenuates the score-parallel component of guidance and applies a time-dependent schedule, reducing off-manifold drift at no extra cost.

Classifier-free guidance (CFG) [Ho and Salimans, 2021] and its variants have become the de facto standard for steering samples toward the high support conditional regions by linearly extrapolating the conditional prediction away from the unconditional using a guidance scale [Ho and Salimans, 2021, Chung et al., 2024]. Such extrapolation has been shown to improve perceptual quality, producing sharper and more detailed samples with stronger conditional alignment. Yet, applying CFG, and related autoguidance methods [Hong et al., 2023, Karras et al., 2024, Hong, 2024, Hyung et al., 2025, Ifriqi et al., 2025], violates the learnt probability conservation during sampling, leading to artifacts and oversaturation that grow more severe as the guidance scale increases.

We first revisit guidance from the perspective of conservation of probability and the model's learned density path. In the ideal continuous-time formulation, the unconditional and conditional velocity fields are trained to satisfy a continuity equation that approximately preserves probability mass along trajectories, but the linear combinations used in current guidance mechanisms break this structure at sampling time, introducing unconstrained divergences and off-manifold drifts. We formalise the contribution of a guidance term to the continuity equation and show that it decomposes into two components: (i) a divergence term that locally creates or removes probability mass, and (ii) a score-parallel flux that transports mass across iso-density surfaces. When these contributions cancel, guidance is conservative and offers little fine-grained control; when their sum grows unchecked, trajectories depart the paths realised during training.

Motivated by this analysis and to reduce departures from the base density path, we introduce *Adaptive Manifold Guidance (AdaMaG)*, a plug-and-play modification to the sampler that treats the conditional velocity as a primary, approximately probability-preserving flow, and restricts guidance to act predominantly along its orthogonal directions. At each step, we decompose the guidance term into components parallel and orthogonal to the primary score direction, preserve the orthogonal updates, and attenuate score parallel guided corrections. This construction strengthens conditional alignment while regularising the geometry of the update and mitigating drift away from the model's learned manifold without additional function evaluations. The divergence term, however, cannot be constrained as directly. Its exact evaluation is intractable at inference time, and suppressing it indiscriminately risks compromising the guidance needed to steer samples. Empirically, however, we observe that magnitude of divergence spikes as sampling approaches the clean-data endpoint.

To characterise this behaviour along the generative path, we measure the divergence of the primary manifold velocities and the residual guidance field. We observe that, across most denoising steps, the primary velocity divergence dominates by orders of magnitude, with guidance divergence only becoming comparable in the final sampling iterations as samples approach the data manifold. We further prove that this late-time divergence spike has a structural origin, blowing up as sampling approaches the data manifold at a rate determined by the conditional/unconditional posterior covariance gap. Consequently, we introduce the second component of AdaMaG, a time-dependent guidance schedule that preserves strong early conditioning while attenuating guidance near the endpoint to suppress this late-time divergence spike.

In summary, our contributions are: **(i) Conservation-based view of guidance.** We analyse guidance as a source term in the continuity equation and identify two components, a divergence and a score-parallel flux, that violate probability conservation, and show that a number of prior manifold-preserving methods can be interpreted under this framework. **(ii) Adaptive Manifold Guidance.** A plug-and-play sampler that attenuates the score-parallel term and schedules guidance to suppress late-stage divergence, with no extra function evaluations. **(iii) Empirical evaluation.** On SD3, SD3.5, and Flux, we conduct comprehensive quantitative and qualitative evaluations and show that AdaMaG improves quality and realism relative to CFG in every evaluated model and guidance setting, with Inception Score matched or improved, and ablations isolating each component's contribution.

2 Related works

We review guidance methods in diffusion and flow models, and then consider recent manifold-aware sampling rules most relevant to our approach.

Guidance in diffusion and flows. Diffusion and flow-matching models generate samples by integrating a learned time-dependent score or velocity field from noise to data [Sohl-Dickstein et al., 2015, Ho et al., 2020, Song et al., 2021, Lipman et al., 2022, Albergo and Vanden-Eijnden, 2023]. To steer these models toward desired conditions, guidance terms were added at sampling time. Early work used classifier-based guidance, augmenting the score with the gradient of a separately trained classifier [Dhariwal and Nichol, 2021, Nichol et al., 2021]. CFG [Ho and Salimans, 2021] replaced the external classifier with conditional and unconditional predictions and has become the de facto standard in text-to-image and text-to-video frameworks [Ho and Salimans, 2021, Rombach et al., 2022, Balaji et al., 2022, Esser et al., 2024, Wan et al., 2025].

Beyond classifier-based and CFG-style methods, guidance can also be derived from CLIP, energy, or reward models, as well as training-free surrogates that approximate CFG-like behaviour without dedicated conditional/unconditional training [Nichol et al., 2021, Yu et al., 2023, Song et al., 2023, Lu et al., 2023, Sadat et al., 2025, Jang et al., 2025]. However, CFG and its variants induce large mismatches between the unconditional and conditional fields, increasing the curvature of the sampling trajectory [Chung et al., 2024, Hyung et al., 2025] and pushing samples off the model's learned manifold, which in turn reduces diversity and yields distorted or oversaturated images [Sadat et al., 2024].

Manifold preservation. Most recently, direct manifold preservation during sampling has drawn attention. Manifold-Preserving Gradient Descent (MPGD) [He et al., 2024] for instance enforces manifold consistency by applying guidance on the denoised clean estimate and restricting the update to the autoencoder's image manifold. Their method is however developed for training-free, loss-based conditioning at inference time, and thus does not directly address extrapolation based guidance methods. CFG++ [Chung et al., 2024] instead targets classifier-free guidance in text-conditional diffusion models, and mitigates off-manifold drift by a minimal sampling-rule change. It specifically forms the guided denoised estimate, but keeps the renoising/noise term unconditional, effectively favoring interpolation over extrapolative CFG at high scales. Although the interpolation view keeps updates bounded, it is fundamentally tied to diffusion samplers where the update decomposes into denoising and renoising terms. In the same spirit, Rectified-CFG++ [Saini et al., 2025] extends this idea to rectified-flow models via a predictor–corrector scheme at the cost of roughly doubling the number of function evaluations.

Characteristic [Zheng and Lan, 2023] gives a geometric view of the score-induced Fokker-Planck dynamics and shows that linearly combining conditional and unconditional scores generally violates this non-linear PDE; they therefore add a corrective term. However, this term is obtained through an iterative approach which raises computational cost. APG [Sadat et al., 2024] further reports, empirically, that the tangential component of guidance (defined with respect to the conditional approximate posterior) is a primary driver of saturation and artifacts, and proposes both tangential downscaling and an update bound to curb overshooting.

In contrast, we view the generative manifold as the geometry learned by the model under the continuity equation, and show that guidance decomposes into two conservation-violating components which motivate our update for reducing off-manifold drift. This decomposition offers a common perspective on related guidance methods, including APG.

3 Method

We aim to improve generation quality and mitigate saturation by reducing off-manifold effects of guidance in conditional normalizing flow (CNF) models. We start by preliminaries, then analyse guidance through the lens of probability conservation, and finally introduce the AdaMaG update rule that follows from this analysis.

3.1 Preliminaries

Rectified flow. Let $x_t \in \mathbb{R}^D$ denote the latent state at time $t \in [0, 1]$, and let y be the condition (with $\varnothing$ denoting the empty prompt). We model the probability-flow dynamics with a velocity field $v_t(\cdot, y)$ parameterised by a neural network $v_\theta(x, t, y)$, and define the unconditional and conditional fields as $v_t^u(x) := v_\theta(x, t, \varnothing)$, and $v_t^c(x) := v_\theta(x, t, y)$ respectively. Samples are then obtained by solving the probability-flow ODE,

$$\frac{dx_t}{dt} = v_\theta(x_t, t, y), \qquad t \in [0, 1], \tag{1}$$

initialized from $x_0 \sim p_0$. Following standard flow convention [Lipman et al., 2022] , we parameterize the marginal x_t as a linear mixture of a source sample x_0 and a target (or data) sample x_1, such that

$$x_t = \alpha_t x_1 + \sigma_t x_0, \tag{2}$$

where (α_t, σ_t) is a scalar schedule with $\alpha_0 = 0, \sigma_0 = 1$ and $\alpha_1 = 1, \sigma_1 = 0$. These conditions can be reversed to follow the convention used by large-scale models such as SD3 [Esser et al., 2024], and WAN [Wan et al., 2025].

Classifier-free guidance. Classifier-free guidance forms the guided velocity by linearly combining $v_t^c(x)$ and $v_t^u(x)$ as

$$v_t^{\text{cfg}}(x) = v_t^u(x) + \omega\big(v_t^c(x) - v_t^u(x)\big), \tag{3}$$

where $\omega > 1$ is the guidance scale. Larger ω typically improves generation quality and text alignment, but also amplifies approximation errors and drives trajectories into artifact-prone regions, often accompanied by increased saturation. Throughout this work we denote the raw guidance residual as $g_t(x) := v_t^c(x) - v_t^u(x)$.

3.2 Characterising off-manifold flows

Our goal is to improve sample quality while retaining semantic conditioning and reducing *off-manifold drift*. At each time t, we use the learned manifold M_t to denote the time-dependent density geometry induced by a selected primary flow v_t and its marginal density p_t (Fig. 2). In this sense, we define off-manifold drift as a guidance-induced departure from this base density path through local volume change or transport across its iso-density surfaces. Probability conservation is therefore defined here only relative to the base path. We call a guidance field conserving when the same p_t remains a solution of the continuity equation after a guidance field g_t is added. Our analysis decomposes departure from this condition into divergence and score-parallel contributions, motivating a practical update that attenuates, rather than eliminates, these contributions while retaining conditioning.

Guidance and probability conservation We now characterise when a guidance field perturbs the model's density and when it does not. The analysis applies identically to the conditional and unconditional flows; we therefore drop superscripts (u, c) until the guidance formulation is fixed in Sec. 3.4.

Let p_t denote the density of x_t induced by a primary velocity field v_t. Probability conservation is expressed by the continuity equation [Chen et al., 2018]

$$\partial_t p_t + \nabla\cdot(p_t\, v_t) \;=\; 0. \tag{4}$$

Adding a guidance field g_t to form $\tilde{v}_t \;:=\; v_t + \omega g_t$ preserves p_t as a solution of (4) if and only if $\nabla\cdot(p_t\, g_t) = 0$. Expanding this constraint with the identity $\nabla p_t = p_t\, \nabla \log p_t$ and using $p_t > 0$ on the support yields the following equivalent condition.

Proposition 3.1 (Conservation under guidance)**.** *The guided velocity $\tilde{v}_t = v_t + \omega g_t$ preserves p_t under the continuity equation* (4) *if and only if (see Appendix B.1)*

$$\underbrace{\nabla\cdot g_t(x)}_{\textit{(i) divergence}} \;+\; \underbrace{g_t(x)^\top\, \nabla \log p_t(x)}_{\textit{(ii) score-parallel flux}} \;=\; 0. \tag{5}$$

Table 1: **Robustness under optimal and strong guidance.** Each entry reports the five-seed mean ± standard deviation across seeds. APG denotes the original repository configuration ($\eta = 0.1, r = 20, \beta = -0.75$); APG* denotes the tuned configuration ($\eta = 0, r = 7.5$) without momentum, see Appendix E. **Bold**: best; underlined: second-best.

	Optimal scale								
	SD3 (ω=7)			SD3.5 (ω=7)			Flux (ω=2)		
Method	FID↓	IS↑	SAT↓	FID↓	IS↑	SAT↓	FID↓	IS↑	SAT↓
CFG	32.62 ± 0.19	**32.94** ± 0.45	0.506 ± 0.002	35.61 ± 0.17	29.09 ± 0.18	0.515 ± 0.002	36.31 ± 0.22	34.82 ± 0.30	0.391 ± 0.001
Rect-CFG++	30.41 ± 0.10	31.81 ± 0.59	0.514 ± 0.002	34.43 ± 0.17	28.46 ± 0.36	**0.482** ± 0.001	35.85 ± 0.16	35.60 ± 0.30	0.374 ± 0.001
TAG	32.62 ± 0.16	32.72 ± 0.26	0.513 ± 0.002	35.25 ± 0.21	29.13 ± 0.35	0.525 ± 0.002	36.14 ± 0.17	35.06 ± 0.53	0.381 ± 0.001
APG	34.51 ± 0.17	28.21 ± 0.35	0.542 ± 0.002	43.41 ± 0.30	24.01 ± 0.41	0.518 ± 0.002	35.04 ± 0.17	35.62 ± 0.52	0.348 ± 0.001
APG*	32.71 ± 0.16	29.44 ± 0.56	0.514 ± 0.002	42.25 ± 0.23	24.26 ± 0.32	0.490 ± 0.003	**34.72** ± 0.25	35.65 ± 0.33	**0.346** ± 0.001
Ours	**30.40** ± 0.16	32.78 ± 0.43	**0.501** ± 0.002	**32.18** ± 0.20	**30.99** ± 0.44	**0.482** ± 0.002	35.57 ± 0.14	**35.87** ± 0.29	0.365 ± 0.001

	High-guidance scale								
	SD3 (ω=15)			SD3.5 (ω=15)			Flux (ω=3)		
Method	FID↓	IS↑	SAT↓	FID↓	IS↑	SAT↓	FID↓	IS↑	SAT↓
CFG	39.19 ± 0.19	27.23 ± 0.14	0.583 ± 0.003	52.65 ± 0.57	21.10 ± 0.31	0.644 ± 0.003	37.85 ± 0.15	31.78 ± 0.31	0.420 ± 0.001
Rect-CFG++	33.36 ± 0.20	31.08 ± 0.43	0.561 ± 0.003	44.46 ± 0.61	24.02 ± 0.28	0.601 ± 0.003	36.39 ± 0.27	34.40 ± 0.51	0.395 ± 0.001
TAG	39.28 ± 0.13	27.36 ± 0.31	0.587 ± 0.003	50.83 ± 0.08	21.73 ± 0.20	0.648 ± 0.003	37.08 ± 0.14	32.69 ± 0.26	0.409 ± 0.001
APG	41.61 ± 0.31	26.19 ± 0.44	0.639 ± 0.003	51.06 ± 0.44	21.50 ± 0.39	0.644 ± 0.001	35.16 ± 0.21	35.38 ± 0.28	0.352 ± 0.001
APG*	34.80 ± 0.04	29.66 ± 0.36	**0.508** ± 0.001	**37.31** ± 0.24	26.77 ± 0.36	**0.505** ± 0.002	35.04 ± 0.24	35.28 ± 0.52	**0.347** ± 0.001
Ours	**32.36** ± 0.20	**32.59** ± 0.44	0.542 ± 0.003	40.78 ± 0.27	**26.86** ± 0.33	0.581 ± 0.004	**35.01** ± 0.20	**35.39** ± 0.31	0.381 ± 0.001

The two terms have distinct geometric meaning: (i) is the local volume change induced by g_t, and (ii) is the flux of probability mass across level sets of p_t. Only their *sum* is constrained – the terms may therefore cancel or both vanish for g_t to be conservative.

A computable surrogate. Score-based models provide direct access to $s_t(x) := \nabla \log p_t(x)$, but the divergence $\nabla \cdot g_t$ is intractable for high-dimensional neural fields at inference time. Consequently, here we first focus on the score parallel term and defer the study and treatment of the divergence term to Sec. 3.3. Hence, temporarily neglecting the divergence term in (5) yields:

$$g_t(x)^\top \nabla \log p_t(x) \;\approx\; 0, \tag{6}$$

which states that guidance direction must be tangent to the instantaneous level sets of p_t.

Relating score and rectified-flow velocity. For rectified flow parameterisations of the form described in subsection 3.1, corresponding probability-flow velocity v_t admits a closed-form representation in terms of the score

$$v_t(x) = \frac{\dot{\alpha}_t}{\alpha_t}\, x - \frac{\dot{\sigma}_t \sigma_t \alpha_t - \dot{\alpha}_t \sigma_t^2}{\alpha_t}\, \nabla_x \log p_t(x), \tag{7}$$

where dots denote derivatives with respect to t (derivation in Appendix B.2). Defining scalar coefficients

$$a_t := \frac{\dot{\alpha}_t}{\alpha_t}, \qquad b_t := \frac{\dot{\sigma}_t \sigma_t \alpha_t - \dot{\alpha}_t \sigma_t^2}{\alpha_t}, \tag{8}$$

we can rewrite (7) as

$$v_t(x) = a_t x - b_t\, \nabla_x \log p_t(x). \tag{9}$$

Consequently, the score-parallel term demonstrated in (5) can be obtained from

$$g_t(x)^\top \nabla_x \log p_t(x) = \frac{1}{b_t}\, g_t(x)^\top \underbrace{\big(a_t x - v_t(x)\big)}_{=:\, n_t(x)}. \tag{10}$$

Thus, in rectified flows, we find that guidance that is orthogonal to $a_t x - v_t(x)$, keeps samples primarily within iso-density surfaces, reducing the deviations from the base conservation equation.

3.3 Divergence and guidance schedule

To investigate the significance of the intractable divergence term in (5), we approximate $\nabla_x \cdot g_t(x)$ using Hutchinson's unbiased estimator. This approach utilises random probe vectors ξ satisfying

$\mathbb{E}[\xi\xi^\top] = I$ and is implemented via Jacobian–vector products. While this procedure is too memory-intensive for deployment during sampling, it serves as a critical analytical tool. As detailed in Figure 3, we observe that the divergence of the guidance field remains small relative to the reference velocity divergences for most of the trajectory but exhibits a sharp increase near the end of sampling. Consequently, we rely on the orthogonality approximation $g_t(x)^\top \nabla_x \log p_t(x) \approx 0$ for earlier timesteps, while introducing a power-law decay schedule to attenuate guidance intensity during later stages, where the measured divergence increases and continuity equation does not hold. Starting from a reference scale ω_{ref}, we modulate the guidance as $t \to 1$, bounded by a minimum strength $\omega_{\min}$,

$$\omega(t) = \max\big(\omega_{\min},\, \omega_{\text{ref}}\,(1-t)^\gamma\big), \qquad t \in [0,1]. \tag{11}$$

Here, ω_{ref} corresponds to the standard CFG guidance strength, while $\omega_{\min}$ and γ govern the rate of late-time attenuation. Systematic ablation studies confirm that this schedule improves generation quality across a wide range of hyperparameters.

3.4 Manifold-aware guidance update

We now formalise the AdaMaG algorithm. We adopt the standard CFG framework (3). Here we assign the conditional velocity $v_t^c(x)$ as base flow and conditioning is introduced via a guidance field $g_t(x)$. Our goal is to suppress the score-parallel flux. We first calculate the score-proportional direction $n_t(x)$ as $n_t(x) = a_t x - v_t^c(x)$ (Eq. (10)).

We use the conditional flow v_t^c for this approximation rather than v_t^u because it empirically provides larger gains in our experiments, particularly at low guidance scales. This choice is also consistent with Sadat et al. [2024], who define their projection relative to the conditional model prediction. We then decompose the raw guidance field $g_t(x)$ into components parallel and orthogonal to this normal direction $n_t(x)$.

The parallel component is given by projecting g_t onto n_t:

$$g_t^{\parallel}(x) := \frac{\langle g_t(x), n_t(x)\rangle}{\|n_t(x)\|^2}\, n_t(x), \tag{12}$$

yielding the orthogonal component $g_t^{\perp}(x) := g_t(x) - g_t^{\parallel}(x)$. Finally, combining these components with the schedule $\omega(t)$ from (11), we define the refined guidance field as:

$$\tilde{g}_t(x) := \omega(t) g_t^{\perp}(x) + \beta g_t^{\parallel}(x), \tag{13}$$

which yields a sampling velocity of $\tilde{v}_t(x) := v_t^u(x) + \tilde{g}_t(x)$. Here, β is the fixed coefficient of the score-parallel component, while $\omega(t)$ scales the score-orthogonal component. Relative to CFG at scale $\omega(t)$, the parallel component is attenuated when $0 \le \beta < \omega(t)$. By prioritizing $g_t^{\perp}$, this update biases the guidance direction toward approximate iso-density surfaces while retaining conditioning. We apply stronger guidance early in the process, where departures from base-path conservation are small and guidance is most effective. As sampling progresses ($t \to 1$), we attenuate its intensity; empirically, this reduces the saturation artifacts associated with the late-stage increase in guidance divergence.

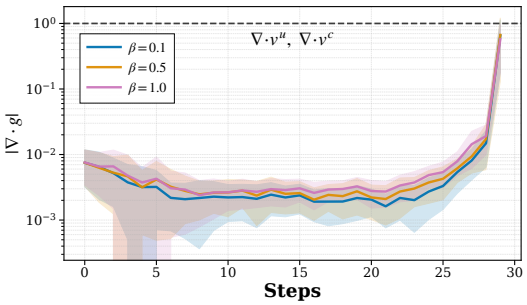


Figure 3: **Divergence magnitude (normalized by dimensionality) along the sampling trajectory.** Black curves show the divergence of the conditional and unconditional velocities as references; colored curves show the divergence of guidance residual under varying score-parallel damping β.

Although the analysis in this section is exact, the design choice to address those are built on two assumptions: (i) generality of late stage divergence blow up, and (ii) independence of the two terms in Eq. (5). For these we provide detailed theoretical and empirical evidence in Appendix A.

4 Experiments

We evaluate our method on text-to-image generation against standard classifier-free guidance (CFG) and recent baselines that explicitly target manifold preservation and saturation. We include Rect-CFG++ [Saini et al., 2025], TAG [Cho et al., 2025], and APG [Sadat et al., 2024]. We also compare

Model	Method	Color ↑	Shape ↑	Texture ↑	Spatial ↑
SD3	CFG	0.7374	0.5789	0.7131	0.3211
	+ Ours	**0.8134**	**0.5811**	**0.7790**	**0.3241**
SD3.5	CFG	0.7415	0.5743	0.7514	0.2851
	+ Ours	**0.8026**	**0.5997**	**0.7821**	**0.2970**
Flux	CFG	0.5943	0.4114	0.5428	**0.2409**
	+ Ours	**0.6118**	**0.4231**	**0.6429**	0.2376

Table 2: **T2I-CompBench results.** AdaMaG improves over CFG across attribute-binding categories and most spatial settings. **Bold**: best per row pair.

Model	Method	FID ↓	IR ↑	PS ↑	HPSv2 ↑	SAT ↓
SD3	CFG	23.89	0.98	0.441	0.275	0.51
	+Ours	**22.34**	**1.043**	**0.557**	**0.288**	**0.48**
SD3.5	CFG	20.29	1.04	0.492	0.281	0.49
	+Ours	**19.18**	**1.118**	**0.594**	**0.293**	**0.46**
Flux	CFG	21.47	1.08	0.512	0.285	0.37
	+Ours	**20.83**	**1.14**	**0.601**	**0.29**	**0.34**

Table 3: **High-resolution evaluation at** 1024×1024**.** IR: ImageReward, PS: PickScore, SAT: Saturation.

with interval guidance, but found it difficult to tune for the models studied here. We therefore report these results separately in Appendix E. Implementation details and hyperparameter settings for all baselines can also be found in the Appendix E.

Experiments are conducted on three large-scale models: Stable Diffusion 3 (SD3) [Esser et al., 2024], Stable Diffusion 3.5 (SD3.5), and Flux. SD3 and SD3.5 use standard classifier-free guidance, which our method directly targets, while Flux applies CFG through negative prompting, testing the proposed update rule in a guidance-distilled setting. Following the standard text-to-image evaluation protocol, we generate 5,000 images from COCO validation-set prompts at 256×256, and 1024×1024, reporting Fréchet Inception Distance (FID), Precision, Recall, Inception Score (IS), ImageReward [Xu et al., 2023], PickScore [Kirstain et al., 2023], HPSv2 [Wu et al., 2023], and a saturation metric capturing the desaturation behaviour characteristic of strong guidance following APG [Sadat et al., 2024]. The Precision here is implemented following [Kynkäänniemi et al., 2019] as an off-manifold indicator, whereas we use the Recall to monitor the sample diversity. We further evaluate compositional alignment on T2I-CompBench [Huang et al., 2023] across color, shape, texture, and spatial attributes. Qualitative results use 150 custom prompts at 1024×1024 (40 of which presented in Appendix F). All images are generated with Euler integration and 40 solver steps on a single H100 GPU.

Divergence study We quantify the divergence term of Eq. 5 along the sampling trajectory using the procedure of Subsection 3.3, focusing on its magnitude in early steps versus its behaviour as $t \to 1$. Figure 3 reports the divergence magnitude (normalised by dimensionality) for the conditional, unconditional, and guidance velocities; the learned velocities serve as a natural manifold baseline since the model is trained to approximately satisfy the continuity equation.

Quantitative comparisons

To benchmark AdaMaG against CFG and related baselines, Table 1 reports FID , IS, and saturation at both moderate and high guidance scales. Since the SD family and guidance-distilled Flux exhibit different guidance behaviour, we use model-specific scales as indicated in the table. Across all models, AdaMaG improves generation quality and reduces saturation relative to CFG. Particularly on SD3 and SD3.5 at $\omega = 7$, AdaMaG achieves an average 8.22% reduction in FID and a 3.70% reduction in saturation, while remaining competitive with other baselines. A further observation is that AdaMaG remains competitive without retuning hyperparameters. In contrast, APG, the second strongest baseline in high guidance scale ranges, is sensitive to its momentum coefficient and radial bound, which can introduce instabilities, particularly at low guidance scales. We study parameter sensitivity for AdaMaG in Section 5.

Beyond standard quality metrics, AdaMaG also strengthens compositional alignment and preference-aligned quality at high resolution (Tables 2, 3). On T2I-CompBench, AdaMaG consistently improves over CFG on attribute-binding categories (Color, Shape, Texture) across all three models, with smaller but generally positive gains on Spatial layout. At 1024×1024, AdaMaG yields concurrent gains in FID, ImageReward, PickScore, HPSv2, and saturation across SD3, SD3.5, and Flux, indicating the manifold-preservation gains transfer cleanly to high-resolution and human aligned outputs.

Qualitative comparisons. Fig. 4 presents qualitative comparisons across the considered baselines, highlighting generation quality and artifact frequency. All samples are generated using each method's optimal guidance scale as selected from Fig. 5. Overall, AdaMaG produces higher-quality generations with noticeably fewer artifacts. Since Fig. 4 uses scales

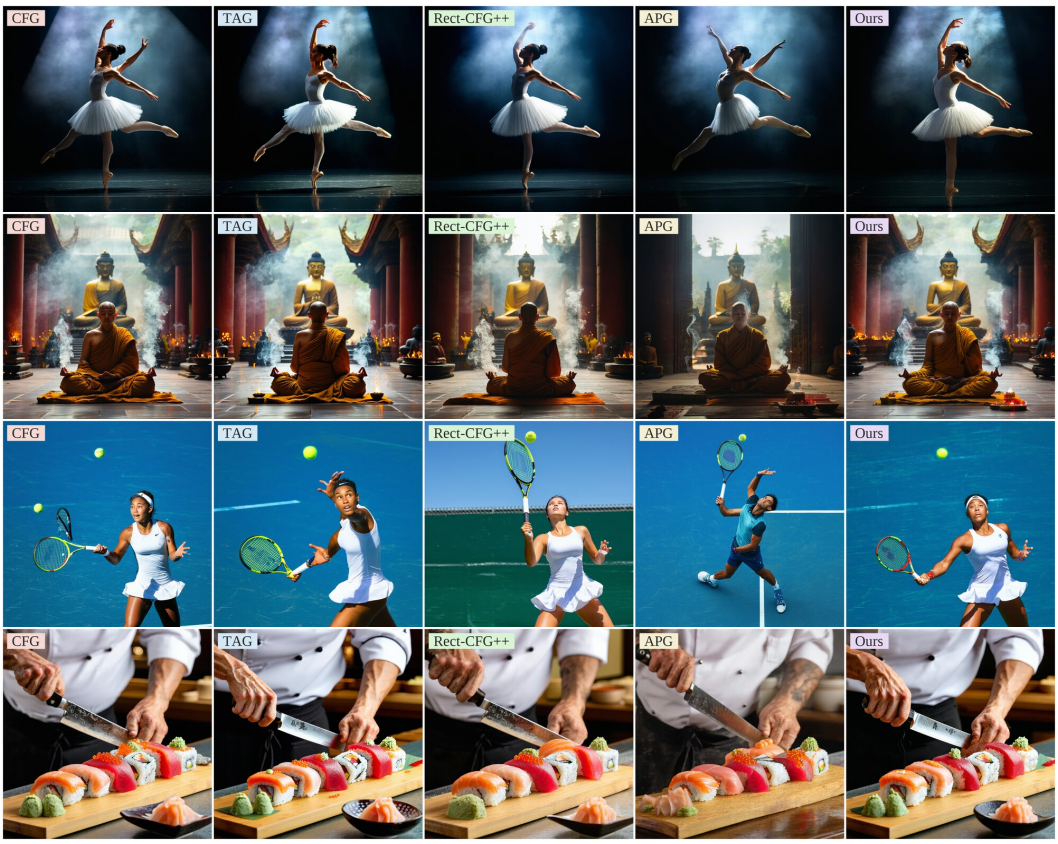


Figure 4: Qualitative comparison between AdaMaG, and other baselines at their optimal setting.

optimised jointly for FID and saturation (Fig. 5), Fig. 6 provides a focused comparison under heavy guidance between CFG, APG (a desaturation-focused method), and AdaMaG.

To complement the automated metrics, we conducted a pairwise preference study with 30 evaluators on 30 prompts using SD3, with each method run at its optimal guidance scale. For every prompt, evaluators chose between AdaMaG and a baseline along three axes (text alignment, image quality, and overall preference) or marked the pair as a tie. Figure 7 reports the aggregated preferences. AdaMaG is preferred over CFG, TAG, and APG across all three criteria, with the strongest margins against APG (63.3% vs. 13.3% on text alignment) and CFG (60.0% vs. 10.0% on overall preference). The lead over TAG is narrower but consistent across all three axes (46.7–53.3% in our favour vs. 13.3–20.0%), indicating that the manifold-preservation gains translate into perceptually meaningful improvements beyond FID and IS.

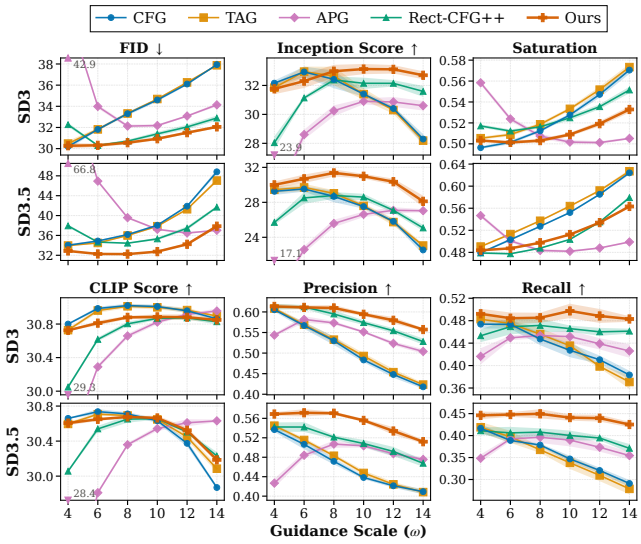


Figure 5: **Sampling metrics with respect to guidance scale.** Guidance-scale sweeps for FID, Inception Score, saturation, CLIP score, precision, and recall on SD3 and SD3.5. Curves show five-seed means, and shaded regions indicate the standard deviation across seeds.

Table 4: **Score-parallel damping ablation.** Effect of varying β on generation quality.

Setting	FID ↓	IS ↑	SAT ↓	Recall ↑
CFG	32.64	33.04 ± 1.44	0.52	0.68
$\beta = 0.0$	31.47	32.67 ± 2.41	0.50	0.70
$\beta = 0.1$	31.39	33.17 ± 1.74	0.50	0.70
$\beta = 1.0$	31.55	33.12 ± 1.05	0.50	0.69
$\beta = 5.0$	32.10	33.13 ± 1.24	0.51	0.70

Table 5: **Guidance schedule exponent ablation.** Effect of varying γ at fixed $\beta = 0.1$.

Setting	FID ↓	IS ↑	SAT ↓	Recall ↑
Ours $\beta = 0.1$	31.39	33.17 ± 1.74	0.50	0.70
$\gamma = 0.1$	31.15	32.79 ± 1.32	0.497	0.71
$\gamma = 1.0$	29.97	33.14 ± 0.88	0.493	0.71
$\gamma = 2.0$	29.55	33.32 ± 1.55	0.491	0.71
$\gamma = 4.0$	29.29	33.06 ± 1.69	0.489	0.72

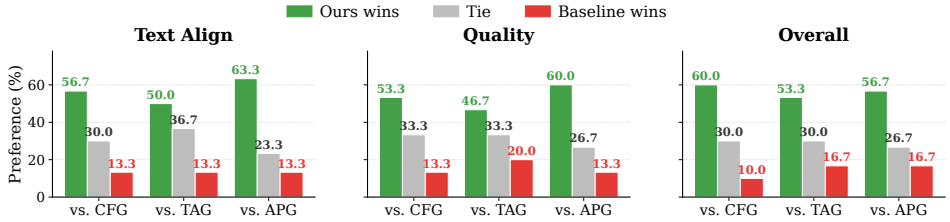


Figure 7: **Human evaluation.** AdaMaG is preferred over CFG, TAG, and APG on text alignment, image quality, and overall preference.

5 Ablations

Score-parallel damping. We first ablate the score-parallel term attenuation in AdaMaG by varying the score-parallel scaling factor β, which controls the magnitude of the score-parallel component in our update. Table 4 reports the resulting trade-off in fidelity and artifact suppression, isolating the effect of minimizing the second term in Eq. 5. This table shows that score-parallel component damping consistently improves fidelity over CFG, reducing FID, and marginally saturation while improving IS. Recall remains essentially unchanged across β, suggesting that scaling down the score-parallel component does not harm the coverage.

Guidance schedule. The first term in (5) captures the divergence of the guidance field, which we mitigate using the power-law guidance schedule introduced in Section 3.3. Table 5 reports an ablation over the exponent γ, showing how the schedule strength affects FID, IS, and saturation. Presented results indicate that increasing γ consistently reduces both FID and saturation, while IS exhibits an optimum at an intermediate setting. Importantly, the schedule never underperforms the no-schedule baseline, and although larger γ continues to improve fidelity and desaturation, we use $\gamma = 4.0$ throughout the study to maintain a competitive IS.

6 Discussion

We provide an additional interpretation of our framework next and defer the discussion of the theoretical connection to other works to Appendix C.

Practical Interpretation. From the decomposition in Eq. 5, guidance splits into an score-parallel component and a divergence component. Both terms violate probability conservation, but they play qualitatively different roles. The divergence term introduces the source/sink mechanism that provides the actual conditioning signal, whereas the score-parallel term induces additional off-manifold drift, which manifests as hallucination and oversaturation.

Consequently, using a larger guidance scale early in sampling strengthens the source/sink effect and helps establish global semantics and structure under



Figure 6: Qualitative saturation comparison $\omega = 15$.

stronger conditioning. As $t \to 1$, we damp guidance to avoid excessive late-stage forcing, which can otherwise produce overly sharp, cartoonish details. In practice, this can be achieved either by reducing the guidance scale $\omega(t)$ or by bounding the magnitude of the guidance residual via a radial norm constraint. For CFG, the applied residual is $\omega(t)g_t = \omega(t)(v_t^c - v_t^u)$, so either mechanism can limit its magnitude.

7 Conclusion

We introduced AdaMaG, a guidance update rule that leverages probability conservation as a proxy for the model's generative manifold, and explicitly limits deviations from this constraint during sampling. Our analysis identified two primary sources of violation, (a) the score-parallel component of the guidance update, and (b) divergence of the guided field. We addressed these effects via score-parallel attenuation and a simple time-dependent guidance schedule, respectively. Through text-guided image generation experiments, we showed that AdaMaG consistently improves upon standard CFG and other baselines, with performance that is largely insensitive to the choice of hyperparameters.

Limitations. A primary limitation of our framework is that it does not explicitly calculate the divergence term during the sampling process due to computational and memory constraints. Addressing this requires future research into generation with cheap, instance-specific likelihood evaluation.

Acknowledgments and Disclosure of Funding

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant EP/W524414/1. This work was also supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). The authors acknowledge the use of resources provided by the Isambard-AI National AI Research Resource (AIRR). Isambard-AI is operated by the University of Bristol and is funded by the UK Government's Department for Science, Innovation and Technology (DSIT) via UK Research and Innovation; and the Science and Technology Facilities Council [ST/AIRR/I-A-I/1023]. This work was supported by Institute for Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (RS-2019-II190075, Artificial Intelligence Graduate School Program(KAIST)). This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2025-00555621) This work was supported by Institute of Information & Communication Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. RS-2025-02263841, Development of a Real-time Multimodal Framework for Comprehensive Deepfake Detection Incorporating Common Sense Error Analysis)

The authors declare no competing interests.

References

Michael Samuel Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *The Eleventh International Conference on Learning Representations*, 2023.

Yogesh Balaji, Seungjun Nah, Xun Huang, Arash Vahdat, Jiaming Song, Qinsheng Zhang, Karsten Kreis, Miika Aittala, Timo Aila, Samuli Laine, et al. ediff-i: Text-to-image diffusion models with an ensemble of expert denoisers. *arXiv preprint arXiv:2211.01324*, 2022.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

Hyunmin Cho, Donghoon Ahn, Susung Hong, Jee Eun Kim, Seungryong Kim, and Kyong Hwan Jin. Tag: Tangential amplifying guidance for hallucination-resistant diffusion sampling. *arXiv preprint arXiv:2510.04533*, 2025.

Hyungjin Chung, Jeongsol Kim, Geon Yeong Park, Hyelin Nam, and Jong Chul Ye. Cfg++: Manifold-constrained classifier free guidance for diffusion models. *arXiv preprint arXiv:2406.08070*, 2024.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 8780–8794. Curran Associates, Inc., 2021.

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, et al. Scaling rectified flow transformers for high-resolution image synthesis. In *Forty-first international conference on machine learning*, 2024.

Yutong He, Naoki Murata, Chieh-Hsin Lai, Yuhta Takida, Toshimitsu Uesaka, Dongjun Kim, Wei-Hsiang Liao, Yuki Mitsufuji, J Zico Kolter, Ruslan Salakhutdinov, and Stefano Ermon. Manifold preserving guided diffusion. In *The Twelfth International Conference on Learning Representations*, 2024.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, 2021.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851. Curran Associates, Inc., 2020.

Susung Hong. Smoothed energy guidance: Guiding diffusion models with reduced energy curvature of attention. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

Susung Hong, Gyuseong Lee, Wooseok Jang, and Seungryong Kim. Improving sample quality of diffusion models using self-attention guidance. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7462–7471, October 2023.

Kaiyi Huang, Kaiyue Sun, Enze Xie, Zhenguo Li, and Xihui Liu. T2i-compbench: A comprehensive benchmark for open-world compositional text-to-image generation. *Advances in Neural Information Processing Systems*, 36:78723–78747, 2023.

Junha Hyung, Kinam Kim, Susung Hong, Min-Jung Kim, and Jaegul Choo. Spatiotemporal skip guidance for enhanced video diffusion sampling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11006–11015, June 2025.

Tariq Berrada Ifriqi, Adriana Romero-Soriano, Michal Drozdzal, Jakob Verbeek, and Karteek Alahari. Entropy rectifying guidance for diffusion and flow models. In *NeurIPS 2025-Thirty-ninth Conference on Neural Information Processing Systems*, 2025.

Sangwon Jang, Taekyung Ki, Jaehyeong Jo, Jaehong Yoon, Soo Ye Kim, Zhe Lin, and Sung Ju Hwang. Frame guidance: Training-free guidance for frame-level control in video diffusion models. *arXiv preprint arXiv:2506.07177*, 2025.

Tero Karras, Miika Aittala, Tuomas Kynkäänniemi, Jaakko Lehtinen, Timo Aila, and Samuli Laine. Guiding a diffusion model with a bad version of itself. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

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. *Advances in neural information processing systems*, 36:36652–36663, 2023.

Tuomas Kynkäänniemi, Tero Karras, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Improved precision and recall metric for assessing generative models. *Advances in neural information processing systems*, 32, 2019.

Tuomas Kynkäänniemi, Miika Aittala, Tero Karras, Samuli Laine, Timo Aila, and Jaakko Lehtinen. Applying guidance in a limited interval improves sample and distribution quality in diffusion models. *Advances in Neural Information Processing Systems*, 37:122458–122483, 2024.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

Yaron Lipman, Marton Havasi, Peter Holderrieth, Neta Shaul, Matt Le, Brian Karrer, Ricky TQ Chen, David Lopez-Paz, Heli Ben-Hamu, and Itai Gat. Flow matching guide and code. *arXiv preprint arXiv:2412.06264*, 2024.

Cheng Lu, Huayu Chen, Jianfei Chen, Hang Su, Chongxuan Li, and Jun Zhu. Contrastive energy prediction for exact energy-guided diffusion sampling in offline reinforcement learning. In *International Conference on Machine Learning*, pages 22825–22855. PMLR, 2023.

Alex Nichol, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob McGrew, Ilya Sutskever, and Mark Chen. Glide: Towards photorealistic image generation and editing with text-guided diffusion models. *arXiv preprint arXiv:2112.10741*, 2021.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10684–10695, June 2022.

Seyedmorteza Sadat, Otmar Hilliges, and Romann M Weber. Eliminating oversaturation and artifacts of high guidance scales in diffusion models. In *The Thirteenth International Conference on Learning Representations*, 2024.

Seyedmorteza Sadat, Manuel Kansy, Otmar Hilliges, and Romann M. Weber. No training, no problem: Rethinking classifier-free guidance for diffusion models. In *The Thirteenth International Conference on Learning Representations*, 2025.

Shreshth Saini, Shashank Gupta, and Alan C Bovik. Rectified-cfg++ for flow based models. *arXiv preprint arXiv:2510.07631*, 2025.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 2256–2265, Lille, France, 07–09 Jul 2015. PMLR.

Jiaming Song, Qinsheng Zhang, Hongxu Yin, Morteza Mardani, Ming-Yu Liu, Jan Kautz, Yongxin Chen, and Arash Vahdat. Loss-guided diffusion models for plug-and-play controllable generation. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 32483–32498. PMLR, 23–29 Jul 2023.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021.

Team Wan, Ang Wang, Baole Ai, Bin Wen, Chaojie Mao, Chen-Wei Xie, Di Chen, Feiwu Yu, Haiming Zhao, Jianxiao Yang, et al. Wan: Open and advanced large-scale video generative models. *arXiv preprint arXiv:2503.20314*, 2025.

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*, 2023.

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. *Advances in Neural Information Processing Systems*, 36:15903–15935, 2023.

Jiwen Yu, Yinhuai Wang, Chen Zhao, Bernard Ghanem, and Jian Zhang. Freedom: Training-free energy-guided conditional diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 23174–23184, 2023.

Candi Zheng and Yuan Lan. Characteristic guidance: Non-linear correction for diffusion model at large guidance scale. *arXiv preprint arXiv:2312.07586*, 2023.

A Trajectory dynamics of guidance

This section studies the structural dynamics of the guidance field along the sampling trajectory and motivates the two design choices in AdaMaG: the time-dependent guidance schedule and the approximately decoupled control provided by the score-parallel damping parameter β. We establish two findings:

- **A sufficient condition for late-stage divergence.** Under the Lipman linear schedule, an exact identity relates $|\nabla_x \cdot g_t|$ to the conditional–unconditional posterior-covariance trace gap. Divergence follows when that gap decays more slowly than σ_t^3. The evaluated models exhibit a late-time increase consistent with this condition. The identity and its sufficient condition are detailed in Appendix A.1.
- **Guidance dynamics exhibit two distinct regimes.** The decomposition of g_t into score-parallel and score-orthogonal components reveals a regime transition along the trajectory: early steps are parallel-dominant, while late steps are orthogonal-dominant and exhibit the divergence increase. Along trajectories sampled with $\beta \in [0.1, 20]$, the divergence of g_t and of its components varies by less than about 20% for all but the last three steps, and grows with β only there, at the orthogonal-dominant end near the clean sample (Figure 8, Appendix A.2).

All proofs are deferred to Appendix B.

A.1 Late-stage divergence spike

The empirical spike in $|\nabla_x \cdot g_t|$ as $t \to 1$ (Figure 3) admits a clean structural explanation. The flow parameterization gives an exact identity relating this divergence to a posterior-covariance gap of the clean data. The identity predicts a spike whenever that gap fails to vanish at the required dimensional rate.

Proposition A.1 (Late-stage divergence behavior)**.** *Let $x_t = \alpha_t x_1 + \sigma_t x_0$ under the Lipman linear schedule with $x_0 \sim \mathcal{N}(0, I)$ and $x_1 \sim p_{\mathrm{data}}$. Let $g_t := v_t^c - v_t^u$ denote the guidance residual, and define the trace gap*

$$\Delta_t(x) \;:=\; \operatorname{tr}\operatorname{Cov}^u[X_1 \mid X_t = x] \;-\; \operatorname{tr}\operatorname{Cov}^c[X_1 \mid X_t = x],$$

between the unconditional and conditional posterior covariances of the clean data X_1 given the noisy state $X_t = x$. Then for every $t \in (0, 1)$,

$$\big|\nabla_x \cdot\, g_t(x)\big| \;=\; \frac{\alpha_t}{\sigma_t^3}\,\big|\Delta_t(x)\big|. \tag{14}$$

In particular, $|\nabla_x \cdot g_t(x)| \to \infty$ as $t \to 1$ whenever $|\Delta_t(x)|$ vanishes strictly slower than σ_t^3. Proof in Appendix B.

Through the exact identity above, the sharp rise of $|\nabla_x \cdot g_t|$ near $t = 1$ in Figure 3 is consistent with $|\Delta_t(x)|$ decaying more slowly than σ_t^3 over the evaluated late-time regime. The finite empirical measurements do not, by themselves, establish the asymptotic decay rate as $t \to 1$.

A.2 Approximate empirical decoupling of the two controls

AdaMaG introduces two hyperparameters: β damping the score-parallel component, and $(\gamma, \omega_{\min})$ shaping the guidance schedule $\omega(t)$. The flux contribution scales exactly with β since $g_t^{\perp}$ is orthogonal to the score,

$$\tilde{g}_t^{\top} \nabla_x \log p_t^c = \beta\, g_t^{\top} \nabla_x \log p_t^c. \tag{15}$$

We next examine the approximate empirical decoupling of these controls and the different guidance regimes along the trajectory.

Empirical sensitivity to β. We measure the divergence of the guidance field and its score-parallel and score-orthogonal components along the sampling trajectory across $\beta \in [0.1, 20]$. Figure 8 reports $|\nabla \cdot g_t|$, $|\nabla \cdot g_t^{\parallel}|$, and $|\nabla \cdot g_t^{\perp}|$ at every denoising step. Despite the resulting change in sampling trajectories, the total divergence remains largely stable across two orders of magnitude in β through

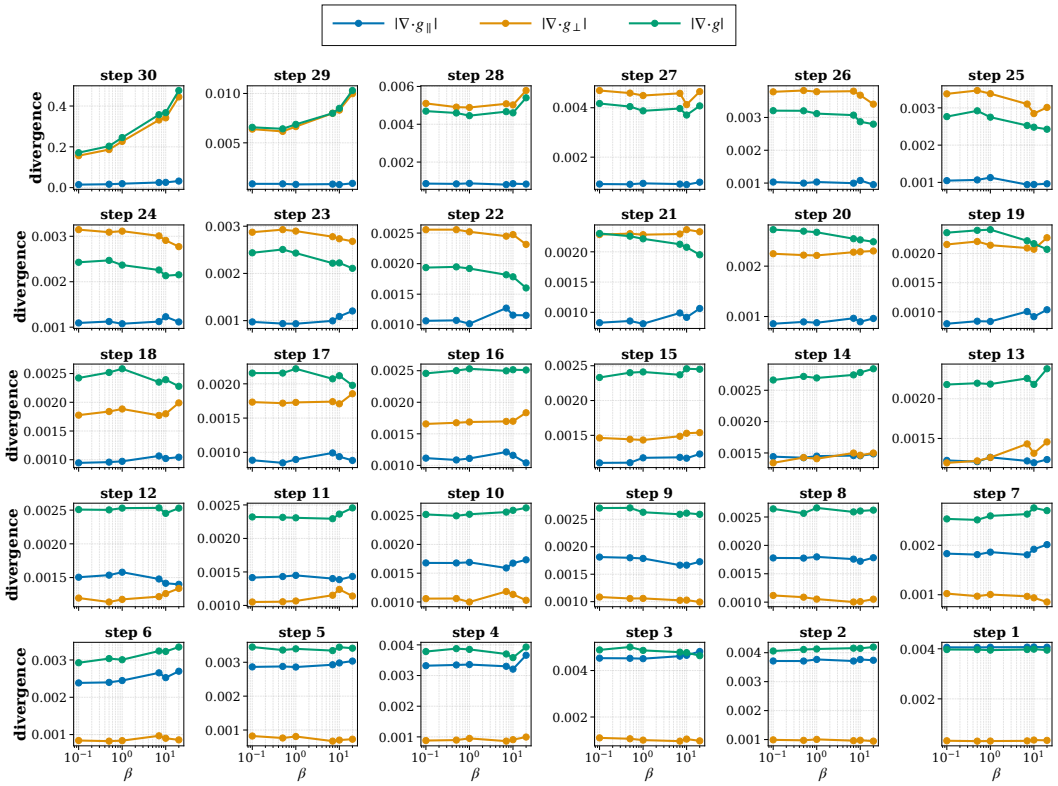


Figure 8: **Divergence components vs. β along the trajectory.** Each panel shows $|\nabla \cdot g_t|$, $|\nabla \cdot g_t^{\parallel}|$, and $|\nabla \cdot g_t^{\perp}|$ as a function of $\beta \in [0.1, 20]$ at a fixed denoising step. The total divergence $|\nabla \cdot g_t|$ (green) remains largely stable across β for most of the trajectory, with greater sensitivity near the endpoint. Early steps exhibit a parallel-dominant regime ($|\nabla \cdot g_t^{\parallel}| > |\nabla \cdot g_t^{\perp}|$), while late steps exhibit an orthogonal-dominant regime in which the divergence increase concentrates in $g_t^{\perp}$.

most steps, while the relative contributions of the parallel and orthogonal components vary over the course of sampling. Near the final denoising steps, the divergence becomes more sensitive to β, with larger values increasing both the total and orthogonal divergence magnitudes. This is the regime in which our schedule attenuates the orthogonal term.

Two regimes along the trajectory. Beyond supporting this approximate decoupling, the per-step decomposition reveals an interesting structural feature of the guidance field. The dominance between the parallel and orthogonal divergence components shifts systematically along the trajectory:

- **Parallel-dominant regime** (early steps, t near 0): $|\nabla \cdot g_t^{\parallel}| > |\nabla \cdot g_t^{\perp}|$, with $|\nabla \cdot g_t| \approx |\nabla \cdot g_t^{\parallel}|$ in the earliest steps. The divergence contribution is therefore dominated by the score-parallel component.
- **Orthogonal-dominant regime** (late steps, t near 1): $|\nabla \cdot g_t^{\perp}|$ dominates and exhibits the late-stage spike of Proposition A.1, while $|\nabla \cdot g_t^{\parallel}|$ remains small.

This regime structure has a concrete design implication: because the late-stage divergence increase is concentrated in $g_t^{\perp}$, β-damping does not act on it directly, although changing the trajectory introduces some endpoint sensitivity to β. The separate schedule $\omega(t)$ attenuates the score-orthogonal component in this regime and is therefore not redundant with β. The factorial ablation over (β, γ) in Section 5 indicates that their improvements to FID and saturation are approximately additive under the evaluated settings.

B Proofs

Here we provide proofs and derivations for the propositions presented in this work.

B.1 Conservation under guidance

Proposition (Conservation under guidance)**.** *Let $p_t \in C^1(\mathbb{R}^D; \mathbb{R}_{>0})$ satisfy $\partial_t p_t + \nabla\cdot(p_t v_t) = 0$ for $v_t \in C^1(\mathbb{R}^D; \mathbb{R}^D)$, and let $g_t \in C^1(\mathbb{R}^D; \mathbb{R}^D)$ and $\omega > 0$. Then p_t satisfies the continuity equation under the guided velocity $\tilde{v}_t := v_t + \omega g_t$ if and only if*

$$\nabla\cdot g_t \;+\; g_t^\top \nabla \log p_t \;=\; 0 \quad \textit{on } \operatorname{supp}(p_t). \tag{16}$$

Proof. By linearity of the divergence, $\nabla\cdot(p_t\tilde{v}_t) = \nabla\cdot(p_t v_t) + \omega\nabla\cdot(p_t g_t)$, so $\partial_t p_t + \nabla\cdot(p_t\tilde{v}_t) = 0$ holds if and only if $\nabla\cdot(p_t g_t) = 0$. The product rule gives

$$\nabla\cdot(p_t g_t) \;=\; p_t \nabla\cdot g_t \;+\; g_t^\top \nabla p_t \;=\; p_t\big(\nabla\cdot g_t \;+\; g_t^\top \nabla \log p_t\big), \tag{17}$$

where the second equality uses $\nabla \log p_t = \nabla p_t / p_t$, valid since $p_t > 0$. As p_t does not vanish on its support, $\nabla\cdot(p_t g_t) = 0$ on $\operatorname{supp}(p_t)$ is equivalent to (16). □

B.2 Velocity-score conversion

We provide a self-contained derivation of the conversion between the marginal score $s_t(x) := \nabla_x \log p_t(x)$ and the marginal velocity field $v_t(x)$ for the Gaussian probability path

$$X_t \;=\; \alpha_t X_1 + \sigma_t X_0, \qquad t \in [0,1], \tag{18}$$

where $X_0 \sim \mathcal{N}(0, I)$ is independent of X_1. Let p_t denote the marginal density of X_t, and let $p_{t|1}(x \mid x_1)$ denote the conditional density of X_t given $X_1 = x_1$.

Step 1: Conditional score. From (18), we have

$$p_{t|1}(x \mid x_1) \;=\; \mathcal{N}\big(x \mid \alpha_t x_1,\, \sigma_t^2 I\big) \;\propto\; \exp\!\left(-\frac{1}{2\sigma_t^2}\,\|x - \alpha_t x_1\|_2^2\right). \tag{19}$$

Taking logs and collecting all terms independent of x into a constant $C(t)$ yields

$$\log p_{t|1}(x \mid x_1) = C(t) - \frac{1}{2\sigma_t^2}\,\|x - \alpha_t x_1\|_2^2. \tag{20}$$

Differentiating (20) with respect to x gives the conditional score

$$\nabla_x \log p_{t|1}(x \mid x_1) = -\frac{1}{\sigma_t^2}\,(x - \alpha_t x_1). \tag{21}$$

Step 2: Fisher's identity and the marginal score. Using $p_t(x) = \int p_{t|1}(x \mid x_1)\, q(x_1)\, dx_1$ and differentiating under the integral sign,

$$\begin{aligned}\nabla_x p_t(x) &= \int \nabla_x p_{t|1}(x \mid x_1)\, q(x_1)\, dx_1 \\ &= \int p_{t|1}(x \mid x_1)\, \nabla_x \log p_{t|1}(x \mid x_1)\, q(x_1)\, dx_1.\end{aligned} \tag{22}$$

Dividing by $p_t(x)$ and recognising the posterior $p(x_1 \mid x) = \frac{p_{t|1}(x|x_1)q(x_1)}{p_t(x)}$ yields Fisher's identity,

$$\nabla_x \log p_t(x) = \mathbb{E}\big[\nabla_x \log p_{t|1}(X_t \mid X_1) \;\big|\; X_t = x\big]. \tag{23}$$

Substituting (21) into (23) gives

$$s_t(x) = \mathbb{E}\left[-\frac{1}{\sigma_t^2}\big(X_t - \alpha_t X_1\big) \;\middle|\; X_t = x\right]. \tag{24}$$

By (18), we have the identity $X_t - \alpha_t X_1 = \sigma_t X_0$, hence

$$s_t(x) = -\frac{1}{\sigma_t^2}\,\mathbb{E}[\sigma_t X_0 \mid X_t = x] = -\frac{1}{\sigma_t}\,\mathbb{E}[X_0 \mid X_t = x]. \tag{25}$$

Defining $x_{0|t}(x) := \mathbb{E}[X_0 \mid X_t = x]$, we obtain the score–$x_0$ conversion

$$x_{0|t}(x) \;=\; -\sigma_t\, s_t(x). \tag{26}$$

Step 3: Velocity in terms of conditional expectations. We define the marginal velocity field

$$v_t(x) \;:=\; \mathbb{E}[\dot{X}_t \mid X_t = x]. \tag{27}$$

Differentiating (18) with respect to t yields $\dot{X}_t = \dot{\alpha}_t X_1 + \dot{\sigma}_t X_0$, and therefore

$$v_t(x) = \dot{\alpha}_t\, \mathbb{E}[X_1 \mid X_t = x] + \dot{\sigma}_t\, \mathbb{E}[X_0 \mid X_t = x] = \dot{\alpha}_t\, x_{1|t}(x) + \dot{\sigma}_t\, x_{0|t}(x), \tag{28}$$

where $x_{1|t}(x) := \mathbb{E}[X_1 \mid X_t = x]$.

Step 4: Eliminating $x_{1|t}$ and expressing v_t via the score. Rearranging (18) gives $X_1 = \alpha_t^{-1}(X_t - \sigma_t X_0)$ (for $\alpha_t \neq 0$), hence taking conditional expectations yields

$$x_{1|t}(x) = \frac{1}{\alpha_t}\, x - \frac{\sigma_t}{\alpha_t}\, x_{0|t}(x). \tag{29}$$

Substituting (29) into (28) gives

$$v_t(x) = \frac{\dot{\alpha}_t}{\alpha_t}\, x + \Big(\dot{\sigma}_t - \sigma_t \frac{\dot{\alpha}_t}{\alpha_t}\Big)\, x_{0|t}(x). \tag{30}$$

Finally, using (26) to eliminate $x_{0|t}(x)$ yields the score–velocity conversion

$$v_t(x) = \frac{\dot{\alpha}_t}{\alpha_t}\, x - \Big(\sigma_t\dot{\sigma}_t - \sigma_t^2 \frac{\dot{\alpha}_t}{\alpha_t}\Big)\, s_t(x). \tag{31}$$

B.3 Laplacian of the log-density via Tweedie

First we obtain an expression of the laplacian log density required for proposition A.1. We take Tweedie's formula as the starting point: for the Gaussian convolution model $X_t = \alpha_t X_1 + \sigma_t X_0$ with $X_0 \sim \mathcal{N}(0, I)$ independent of $X_1 \sim p^k$ ($k \in \{c, u\}$), the marginal density p_t^k satisfies, at every x with $p_t^k(x) > 0$,

$$\nabla_x \log p_t^k(x) \;=\; \frac{\alpha_t\, \mathbb{E}^k[X_1 \mid X_t = x] - x}{\sigma_t^2}. \tag{32}$$

Lemma B.1 (Laplacian of the log-density)**.** *Under the setup above,*

$$\Delta_x \log p_t^k(x) \;=\; \frac{\alpha_t^2\, \mathrm{tr}\, \mathrm{Cov}^k[X_1 \mid X_t = x] - D\sigma_t^2}{\sigma_t^4}. \tag{33}$$

Proof. Differentiating (32) with respect to x gives the Hessian of the log-marginal:

$$\nabla_x^2 \log p_t^k(x) \;=\; \frac{\alpha_t\, \nabla_x \mathbb{E}^k[X_1 \mid X_t = x] - I}{\sigma_t^2}.$$

From Tweedie's equation, it follows that

$$\nabla_x \mathbb{E}^k[X_1 \mid X_t = x] \;=\; \frac{\alpha_t}{\sigma_t^2}\, \mathrm{Cov}^k[X_1 \mid X_t = x],$$

so substituting yields the matrix identity

$$\nabla_x^2 \log p_t^k(x) \;=\; \frac{\alpha_t^2\, \mathrm{Cov}^k[X_1 \mid X_t = x]}{\sigma_t^4} \;-\; \frac{I}{\sigma_t^2}.$$

Taking the trace of both sides yields (33) where D is the dimension, and Δ_x is the Laplacian. □

B.4 Proof of Proposition A.1

Proposition A.1 (Late-stage divergence behavior). *Let $X_t = \alpha_t X_1 + \sigma_t X_0$ with $X_0 \sim \mathcal{N}(0, I)$ independent of $X_1 \sim p_{\text{data}}$, under the Lipman linear schedule $\alpha_t = t,\ \sigma_t = 1 - t$. Let $g_t := v_t^c - v_t^u$ denote the guidance residual, and define the trace gap*

$$\Delta_t(x) \ := \ \operatorname{tr} \operatorname{Cov}^u[X_1 \mid X_t = x] \ - \ \operatorname{tr} \operatorname{Cov}^c[X_1 \mid X_t = x],$$

where the conditional and unconditional posterior covariances are taken under $p_{\text{data}}(\cdot \mid y)$ and p_{data} respectively. Then for every $t \in (0, 1)$ and every x in the joint support of p_t^c and p_t^u,

$$\big|\nabla_x \cdot g_t(x)\big| \ = \ \frac{\alpha_t}{\sigma_t^3} \, \big|\Delta_t(x)\big|.$$

In particular, $|\nabla_x \cdot g_t(x)| \to \infty$ as $t \to 1$ whenever $|\Delta_t(x)|$ vanishes strictly slower than σ_t^3.

Proof. The proof has three steps: reduce the divergence to a Laplacian gap of log-densities (Step 1), apply Lemma B.1 to express this gap in terms of posterior covariances (Step 2), and combine with the schedule-specific value of b_t (Step 3).

Step 1: Reduce divergence to a Laplacian gap. From the velocity–score relation (9), $v_t^k(x) = a_t x - b_t \nabla_x \log p_t^k(x)$ for $k \in \{c, u\}$. The $a_t x$ terms are identical and cancel under subtraction, giving

$$g_t(x) \ = \ -b_t \big[\nabla_x \log p_t^c(x) - \nabla_x \log p_t^u(x)\big].$$

Since b_t depends only on t, the divergence commutes with the scalar factor:

$$\nabla_x \cdot g_t(x) \ = \ -b_t \big[\Delta_x \log p_t^c(x) - \Delta_x \log p_t^u(x)\big]. \tag{34}$$

Step 2: Express the Laplacian gap via posterior covariances. Applying Lemma B.1 to both $k = c$ and $k = u$ gives

$$\Delta_x \log p_t^k(x) \ = \ \frac{\alpha_t^2 \operatorname{tr} \operatorname{Cov}^k[X_1 \mid X_t = x] - D\sigma_t^2}{\sigma_t^4}.$$

The dimensional terms $-D\sigma_t^2/\sigma_t^4$ are identical for c and u and cancel under subtraction:

$$\Delta_x \log p_t^c(x) - \Delta_x \log p_t^u(x) \ = \ -\frac{\alpha_t^2}{\sigma_t^4} \, \Delta_t(x). \tag{35}$$

Step 3: Compute b_t and combine. Under the linear Lipman schedule $\alpha_t = t$, $\sigma_t = 1 - t$, a direct computation gives

$$b_t \ = \ \frac{\dot\sigma_t \sigma_t \alpha_t - \dot\alpha_t \sigma_t^2}{\alpha_t} \ = \ -\frac{\sigma_t}{\alpha_t}.$$

Substituting (35) and this value of b_t into (34),

$$\nabla_x \cdot g_t(x) \ = \ -\Big(-\frac{\sigma_t}{\alpha_t}\Big) \cdot \Big(-\frac{\alpha_t^2}{\sigma_t^4} \, \Delta_t(x)\Big) \ = \ -\frac{\alpha_t}{\sigma_t^3} \, \Delta_t(x).$$

Taking absolute values yields the claimed identity. □

Late-stage blow-up. The final claim follows directly from the identity above. As $t \to 1$, $\alpha_t \to 1$ and $\sigma_t \to 0$, so $\alpha_t/\sigma_t^3 \to \infty$ at rate $\Theta(1/(1-t)^3)$. If $|\Delta_t(x)|$ vanishes strictly slower than σ_t^3, then $|\nabla_x \cdot g_t(x)|$ diverges.

Empirically from 3 and across seeds we observe the same pattern as demonstrated in proposition.

C Connections to related work

Connection to APG. We use $n_t(x) = a_t x - v_t^c(x)$ to decompose the guidance field $g_t(x)$ and attenuate its component along the score, reducing the flux across iso-density surfaces in Eq. 5. This is related to APG [Sadat et al., 2024], which reduces oversaturation by attenuating the component parallel to the conditional denoised prediction $\hat{x}_0^{\text{cond}}$. However, the denoised prediction depends on both the score and the current state. Its direction therefore generally differs from $n_t(x)$, and the two methods suppress different components of guidance. Both methods use projection-based attenuation, but APG references the predicted clean endpoint and AdaMaG the predicted noise endpoint of the same flow interpolation. APG further combines this projection with norm clipping and reverse momentum. Our analysis also identifies a contribution from the divergence of the guidance field, which grows near the clean endpoint. We address this through $\omega(t)$, reducing guidance strength as sampling approaches this regime.

Connection to Rectified-CFG++. Rectified-CFG++ [Saini et al., 2025] adopts a predictor-corrector sampling strategy in which each guided (predictor) update is followed by a corrective step that re-applies the base generative dynamics. Interpreted through our lens, this correction acts as an implicit manifold update inserted between successive guided updates. Consequently, the method approximately enforces our deviation constraint in an alternating manner such that after each guided step that may introduce off-manifold drift, the subsequent corrector step reduces the accumulated deviation by steering the iterate back toward the model's learned trajectory.

D Implementation and reproducibility

Paired per-image seeding. All generation comparisons use paired initial conditions. A single random-number generator would produce distinct noise for each image, but the result for image i would depend on every preceding draw. Restarting a failed job, distributing prompts across jobs or GPUs, changing the batching, or drawing additional random values (e.g., Hutchinson probes) could therefore change the subsequent initial conditions. Instead, we assign each prompt its own seed. For run seed s and global prompt index $i \in \{0, \dots, N-1\}$, we use

$$s_i = 1{,}000{,}003\, s + i. \tag{36}$$

Each image is sampled by a separate generator initialized with s_i. Its noise $x_{1,i}$ therefore depends only on (s, i), allowing cluster jobs to resume at any prompt index, work to be partitioned freely, and every method and guidance scale to receive the same paired input. The multiplier exceeds the number of evaluated prompts ($N = 5{,}000$), so distinct run seeds occupy non-overlapping seed blocks. Here, the subscript in $x_{1,i}$ denotes the Diffusers noise level $\sigma = 1$.

At 256×256, $x_{1,i}$ has shape $16\times32\times32$ and is the only stochastic input to the Euler trajectory. SD3 and SD3.5 use identical FP16 initial latents for a given (s, i); Flux draws in bfloat16, so pairing is guaranteed within each model rather than bitwise across model families.

Algorithm 1 Paired image seeding

Input: ordered prompts $\{y_i\}_{i=0}^{N-1}$; run seed s; latent shape S; precision d.

1. Set $q \leftarrow 1{,}000{,}003$.
2. For each global prompt index i:
   (a) $s_i \leftarrow qs + i$.
   (b) Initialize an independent generator G_i with s_i.
   (c) Draw $x_{1,i} \sim \mathcal{N}(0, I)$ with shape S and precision d using G_i.
3. Evaluate every method and guidance scale using the same paired input $(y_i, x_{1,i})$.

Output: batch-invariant paired initial latents $\{x_{1,i}\}$.

Algorithm 2 AdaMaG sampling update

Input: $x_{\sigma_0} \equiv x_{1,i} \sim \mathcal{N}(0, I)$; decreasing scheduler $\{\sigma_k\}_{k=0}^{K}$; condition y; ω_{ref}, $\omega_{\min}$, γ, β; numerical constant ε.

1. For $k = 0, \dots, K-1$:
   (a) $\tau_k \leftarrow 1 - \sigma_k$.
   (b) Evaluate $v_k^c \leftarrow v_\theta(x_{\sigma_k}, \sigma_k, y)$ and $v_k^u \leftarrow v_\theta(x_{\sigma_k}, \sigma_k, \varnothing)$.
   (c) $g_k \leftarrow v_k^c - v_k^u$ and $n_k \leftarrow x_{\sigma_k} + \tau_k v_k^c$.
   (d) $g_k^{\parallel} \leftarrow \dfrac{\langle g_k, n_k \rangle}{\|n_k\|_2^2 + \varepsilon} n_k$; $g_k^{\perp} \leftarrow g_k - g_k^{\parallel}$.
   (e) $\omega_k \leftarrow \max\{\omega_{\min}, \omega_{\text{ref}} \sigma_k^{\gamma}\}$.
   (f) $\tilde{v}_k \leftarrow v_k^u + \beta g_k^{\parallel} + \omega_k g_k^{\perp}$.
   (g) $x_{\sigma_{k+1}} \leftarrow x_{\sigma_k} + (\sigma_{k+1} - \sigma_k)\tilde{v}_k$.

Output: generated latent x_{σ_K}.

Table 6: **Direct comparison with APG.** Matched component comparison using the same reference guidance scale, sampler, NFE budget, seeds, and APG attenuation convention. APG clipping uses the paper reported value; the indented row adds momentum to that setting.

Setting	FID↓	IS↑	SAT↓	Prec.↑	Rec.↑	CLIP↑
CFG	38.75	27.92	0.576	0.39	0.39	30.97
APG, attenuation only	39.28	27.10	0.641	0.41	0.36	30.83
APG, attenuation + clipping	32.51	30.40	0.587	**0.61**	0.41	30.92
APG, attenuation + clipping + momentum	35.53	27.20	0.610	–	0.39	30.28
AdaMaG, attenuation only	37.93	27.41	0.572	0.41	0.37	**31.05**
AdaMaG, attenuation + schedule	**31.96**	**33.45**	**0.534**	0.54	**0.49**	31.02

AdaMaG implementation. The analysis in this work follows the standard CNF convention [Lipman et al., 2022, 2024], with time increasing from source noise to data. For implementation, we reverse the time parameterization to match open-source packages such as Diffusers, whose noise level σ decreases from 1 to 0. Thus, the presented time in the main body is $\tau = 1 - \sigma$. For a linear rectified flow, the Diffusers velocity consequently has the opposite sign to the Lipman et al. [2022] velocity. Accordingly,

$$x + \tau v_\sigma^c = \tau\big(a_\tau x - v_\tau^c\big), \tag{37}$$

so $x + \tau v_\sigma^c$ is a finite, scale-equivalent implementation of the score-normal direction in Eq. (10). The pseudocode below is written directly in the scheduler noise level σ_k. For the FlowMatch scheduler used in our experiments, this is equivalent to the implementation convention $\sigma_k = t_k/1000$.

Unless stated otherwise, we use $\beta = 0.1$ and $\gamma = 4$ throughout. For SD3 and SD3.5, $\omega_{\min} = 4$ and ω_{ref} is the reference scale specified by each experiment. For Flux, we use $\omega_{\text{ref}} = 3$, $\omega_{\min} = 1.2$, and $\gamma = 4$.

E Baselines

Rect-CFG++. Rect-CFG++ is implemented as a predictor–corrector scheme and therefore requires two network evaluations per integration update (one prediction and one correction), i.e., roughly $2\times$ NFE compared to single-evaluation baselines. To keep comparisons compute-matched, we fix the total NFE budget across methods and split it evenly between predictor and corrector steps. Concretely, with a 40-NFE budget we run 20predictor updates and 20 corrector updates, yielding the same total number of model calls as the 40-step Euler baselines.

APG adaptation to rectified flows. APG was originally introduced for diffusion models with an explicit x_0 (clean-image) prediction. We adapt APG to rectified flows by first forming a one-step estimate of x_0 from the current state x_t and the model velocity at the same noise level. We then apply the standard APG update in x_0-space using the original coefficients from the diffusion formulation. Finally, we map the updated $\hat{x}_0$ back to an equivalent rectified-flow update by taking a single Euler step that returns to the same noise level t, and define the corresponding velocity that realizes this update.

APG hyperparameter sensitivity. We first isolate APG's momentum term in the matched comparison of Table 6. Adding momentum to attenuation and clipping increases FID from 32.51 to 35.53, lowers IS, recall, and CLIP score, and increases saturation. We therefore disable momentum in the subsequent tuning. With momentum removed, the radial clipping threshold is the key remaining hyperparameter. Figure 9 shows that its value materially changes APG's behavior across metrics and guidance scales. The hyperparameters reported for APG do not transfer uniformly to SD3 and SD3.5, so based on tuning across both models we use $\eta = 0$ and $r = 7.5$ without momentum for APG*. For transparency, Table 1 reports both the published configuration (APG) and the tuned configuration (APG*).

For a direct comparison in Table 6, we adopt APG's attenuation convention. APG scales the extrapolation from the conditional prediction, whereas our update is written relative to v_t^u. We

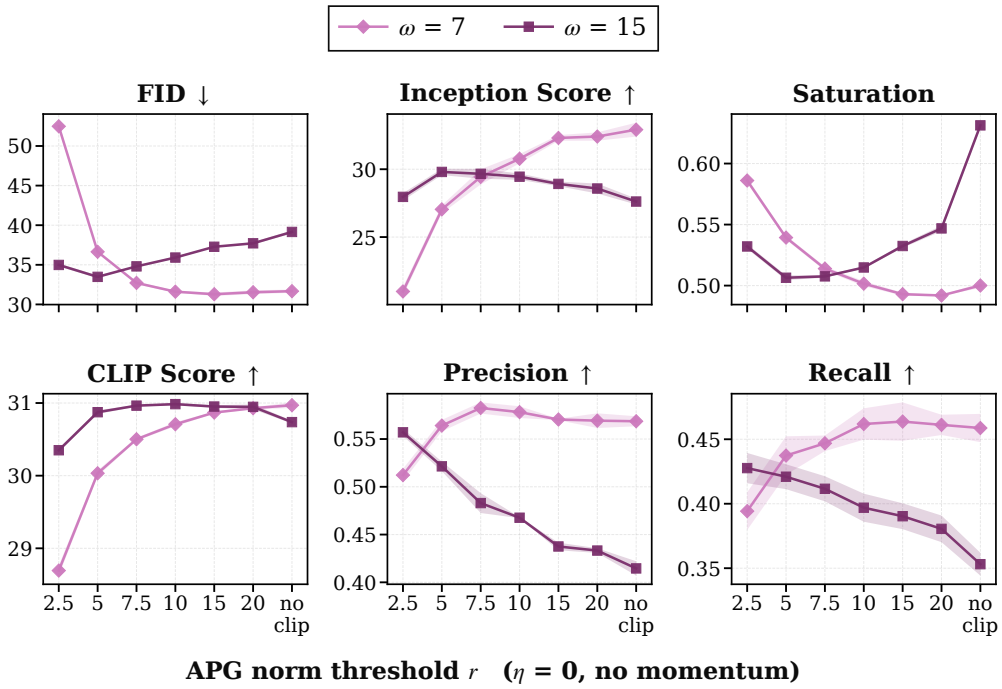


Figure 9: **Sensitivity of APG to the radial norm threshold.** Results on SD3 at guidance scales $\omega = 7$ and $\omega = 15$ with $\eta = 0$ and momentum disabled. Curves show five-seed means, shaded regions indicate the standard deviation across seeds, and "no clip" disables radial norm clipping.

therefore replace the coefficient β of $g_t^{\parallel}$ with $1 + [\omega(t) - 1]\beta$, giving

$$\begin{aligned}\tilde{v}_t &= v_t^u + \omega(t) g_t^{\perp} + \big(1 + [\omega(t) - 1]\beta\big) g_t^{\parallel} \\ &= v_t^c + [\omega(t) - 1]\big(g_t^{\perp} + \beta g_t^{\parallel}\big).\end{aligned}$$

Here, β denotes the shared attenuation parameter, and the conversion uses the effective guidance scale $\omega(t)$ at each step. Both methods thus apply the same attenuation convention, with projections defined relative to their respective reference directions.

Under this convention, attenuation alone gives FID improvement for AdaMaG, while adding the schedule reduces FID from 37.93 to 31.96. APG benefits substantially from norm clipping, reaching 32.51. AdaMaG also achieves higher IS and recall and lower saturation, although APG retains higher precision.

Interval guidance. Interval guidance [Kynkäänniemi et al., 2024] observes empirically that guidance at high noise levels reduces sample diversity, while guidance at low noise levels is largely unnecessary. Based on these observations, it restricts guidance to an interval in the middle of the sampling trajectory and switches it off near both ends.

In our investigation we find that switching guidance off near the noisy end of the trajectory consistently and substantially degrades sample quality for SD3, SD3.5 and FLUX, so we omit this variant from the main table. For completeness, we instead tune the interval limit in the low-noise regime, following the procedure of the original paper, and compare the result with AdaMaG. These results are reported in Table 7.

Table 7: Limited-interval guidance with tuned limits versus AdaMaG (single seed, 5,000 COCO captions, 256×256, 40 steps). Following Kynkäänniemi et al. [2024], the upper limit is tuned first; the best setting leaves only the first (pure-noise) step unguided, and k is the number of final steps run without guidance.

Model	Method	FID↓	IS↑	CLIP↑	Sat.	Prec.↑	Rec.↑
SD3, $\omega = 7$	CFG	32.49	32.97	31.00	0.505	0.548	0.464
	Interval, $k = 0$	32.14	32.97	30.99	0.508	0.545	0.456
	Interval, $k = 4$	32.35	32.70	30.96	0.508	0.538	0.440
	Interval, $k = 8$	32.32	32.36	30.88	0.507	0.539	0.441
	Interval, $k = 12$	31.82	31.89	30.79	0.507	0.549	0.447
	Interval, $k = 16$	31.41	31.68	30.67	0.507	0.556	0.461
	AdaMaG (ours)	30.31	32.88	30.85	0.499	0.612	0.484
SD3, $\omega = 15$	CFG	38.92	27.40	30.75	0.579	0.392	0.374
	Interval, $k = 0$	39.95	26.59	30.65	0.619	0.396	0.371
	Interval, $k = 4$	41.30	25.77	30.62	0.619	0.393	0.359
	Interval, $k = 8$	41.54	25.67	30.57	0.618	0.399	0.351
	Interval, $k = 12$	41.02	25.75	30.51	0.617	0.420	0.362
	Interval, $k = 16$	39.81	26.26	30.41	0.615	0.430	0.402
	AdaMaG (ours)	32.01	32.88	30.83	0.537	0.548	0.500
SD3.5, $\omega = 7$	CFG	35.75	29.12	30.70	0.517	0.483	0.386
	Interval, $k = 0$	35.75	28.74	30.67	0.520	0.492	0.389
	Interval, $k = 4$	36.46	28.14	30.66	0.519	0.491	0.383
	Interval, $k = 8$	35.47	28.77	30.64	0.518	0.504	0.387
	Interval, $k = 12$	34.91	29.56	30.58	0.518	0.510	0.411
	Interval, $k = 16$	34.80	29.14	30.45	0.517	0.515	0.411
	AdaMaG (ours)	32.18	31.52	30.65	0.494	0.575	0.455
SD3.5, $\omega = 15$	CFG	52.70	20.84	29.66	0.642	0.398	0.263
	Interval, $k = 0$	51.05	20.80	29.79	0.650	0.401	0.288
	Interval, $k = 4$	53.50	19.93	29.77	0.648	0.395	0.257
	Interval, $k = 8$	53.68	20.21	29.74	0.646	0.404	0.280
	Interval, $k = 12$	53.10	20.08	29.70	0.645	0.413	0.286
	Interval, $k = 16$	52.42	20.62	29.62	0.642	0.426	0.295
	AdaMaG (ours)	40.43	27.02	29.99	0.587	0.507	0.410
FLUX, $\omega = 2$	CFG	36.22	34.80	31.04	0.392	0.596	0.450
	Interval, $k = 0$	36.24	35.59	31.00	0.388	0.596	0.459
	Interval, $k = 4$	36.31	35.51	31.00	0.387	0.593	0.455
	Interval, $k = 8$	36.38	35.43	31.00	0.387	0.595	0.456
	Interval, $k = 12$	36.41	35.33	30.98	0.387	0.591	0.455
	Interval, $k = 16$	36.47	35.31	30.96	0.386	0.594	0.454
	AdaMaG (ours)	35.56	35.58	30.53	0.365	0.632	0.469
FLUX, $\omega = 3$	CFG	37.77	32.15	31.17	0.419	0.524	0.427
	Interval, $k = 0$	37.34	32.37	31.14	0.421	0.533	0.418
	Interval, $k = 4$	37.51	31.98	31.15	0.420	0.535	0.421
	Interval, $k = 8$	37.67	31.82	31.15	0.419	0.536	0.421
	Interval, $k = 12$	37.78	31.67	31.14	0.418	0.534	0.418
	Interval, $k = 16$	37.88	31.65	31.12	0.417	0.535	0.420
	AdaMaG (ours)	36.43	35.50	30.71	0.382	0.607	0.482

F Custom prompts

Prompt list (In order of appearance)

P1: A ballerina mid-pirouette in a flowing white tutu with motion blur on the fabric and sharp focus on her concentrated expression

P2: Monk meditating in a serene temple with incense smoke curling and golden Buddha statue behind

P3: Tennis player serving with ball tossed high and body arched in powerful motion

P4: A master sushi chef slicing fresh tuna with a long knife at a traditional counter while perfect nigiri pieces are arranged on a wooden board with wasabi and pickled ginger

P5: Steam rising from a copper Turkish coffee pot on an ornate brass tray with ceramic cups painted in cobalt blue patterns

P6: A white horse galloping through shallow surf at sunrise with spray catching pink and gold light and mane flowing

P7: A cormorant drying its wings on a wooden post with water droplets and soft morning backlight

P8: A vintage sailboat with weathered wooden hull and cream canvas sails gliding across turquoise Mediterranean waters at golden hour

P9: A woman with freckles and copper hair laughing in golden hour light with wind catching loose strands

P10: A woman in a white linen shirt reading on a sun-dappled balcony with coffee steaming beside her

P11: A ballerina tying her pointe shoe ribbons with focus and determination visible in her expression

P12: A florist arranging peonies with green-stained fingers and apron covered in petals

P13: Close-up of a woman applying red lipstick in a vintage compact mirror

P14: A barista creating latte art with intense focus and steam rising around her

P15: A woman wrapped in cashmere reading by firelight with snow falling outside the window

P16: A dancer stretching at a barre with morning sun streaming through tall windows

P17: A solitary cypress tree on a Tuscan hilltop at golden hour with rolling green hills and distant farmhouses

P18: An elderly craftsman's hands shaping wet clay on a pottery wheel with earth tones and soft window light

P19: A woman in white linen walking barefoot through shallow tide pools with reflections of clouds and blue sky

P20: A gondolier in striped shirt steering through a narrow Venice canal with laundry hanging above and golden walls

P21: A woman hiking at sunrise with wind in her hair and mountains behind her

P22: A woman floating in a turquoise cenote with white dress billowing underwater

P23: Close-up of hands holding a worn leather journal and fountain pen

P24: An archer drawing her bow with intense focus and forest background

P25: A woman running through a sunflower field with arms outstretched and joy on her face

P26: A traditional wooden rowboat half submerged in clear lake water with autumn trees reflected around it

P27: A woman with wind-blown dark hair on a cliff overlooking a turbulent sea in a camel wool coat

P28: A woman with auburn hair in a rust-colored sweater holding a steaming mug on a foggy morning porch

P29: A sommelier in crisp white shirt decanting wine by candlelight

P30: A woman running through a sunflower field with arms outstretched and joy on her face

P31: Antique maps spread on a table with a brass compass magnifying glass and leather case

P32: A barn owl in flight at dusk with wings spread and soft feathers and golden field below

P33: A falconer with hawk landing on her gloved hand against stormy sky

P34: A toucan with massive bill in orange yellow and black tossing a berry in the air against rainforest backdrop

P35: A Wilson's bird-of-paradise displaying with turquoise crown curled tail feathers and cape of iridescent green

P36: A Bengal cat with marbled coat lounging on a velvet cushion in deep teal with dust motes in sunlight

P37: A hyacinth macaw pair preening each other with cobalt blue feathers and bright yellow eye rings against jungle green

P38: A stop sign saying all way underneath it

P39: A white horse galloping through shallow surf at splashing water as running

P40: A giraffe drinking water by the lake

G Additional results

We provide further qualitative comparisons between AdaMaG and standard CFG at their respective optimal guidance scales. Each pair is generated under the same prompt and seed, illustrating how AdaMaG affects visual fidelity and artifact frequency across a larger set of examples.

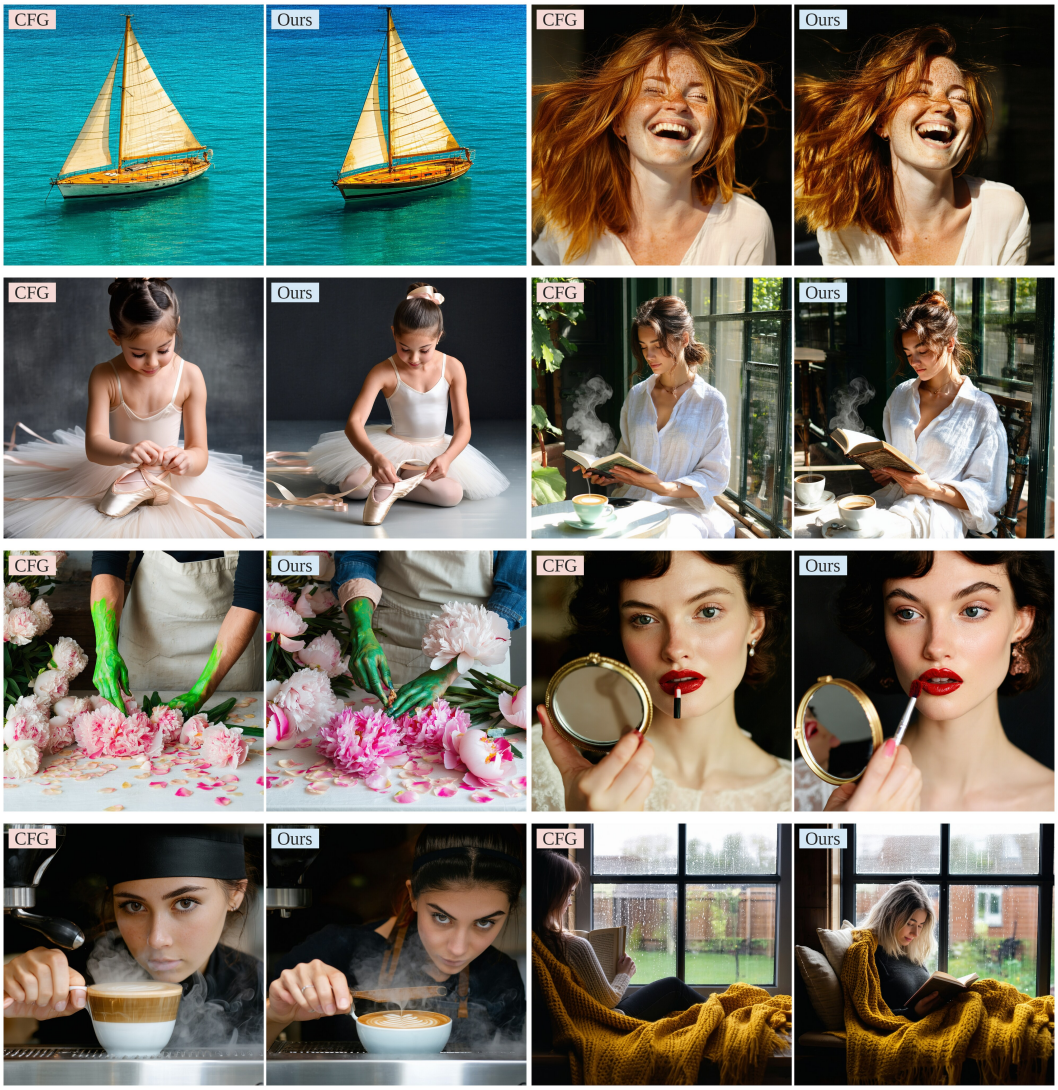


Figure 10: Additional qualitative comparisons between AdaMaG and CFG at their respective optimal guidance scales (part 1/4).

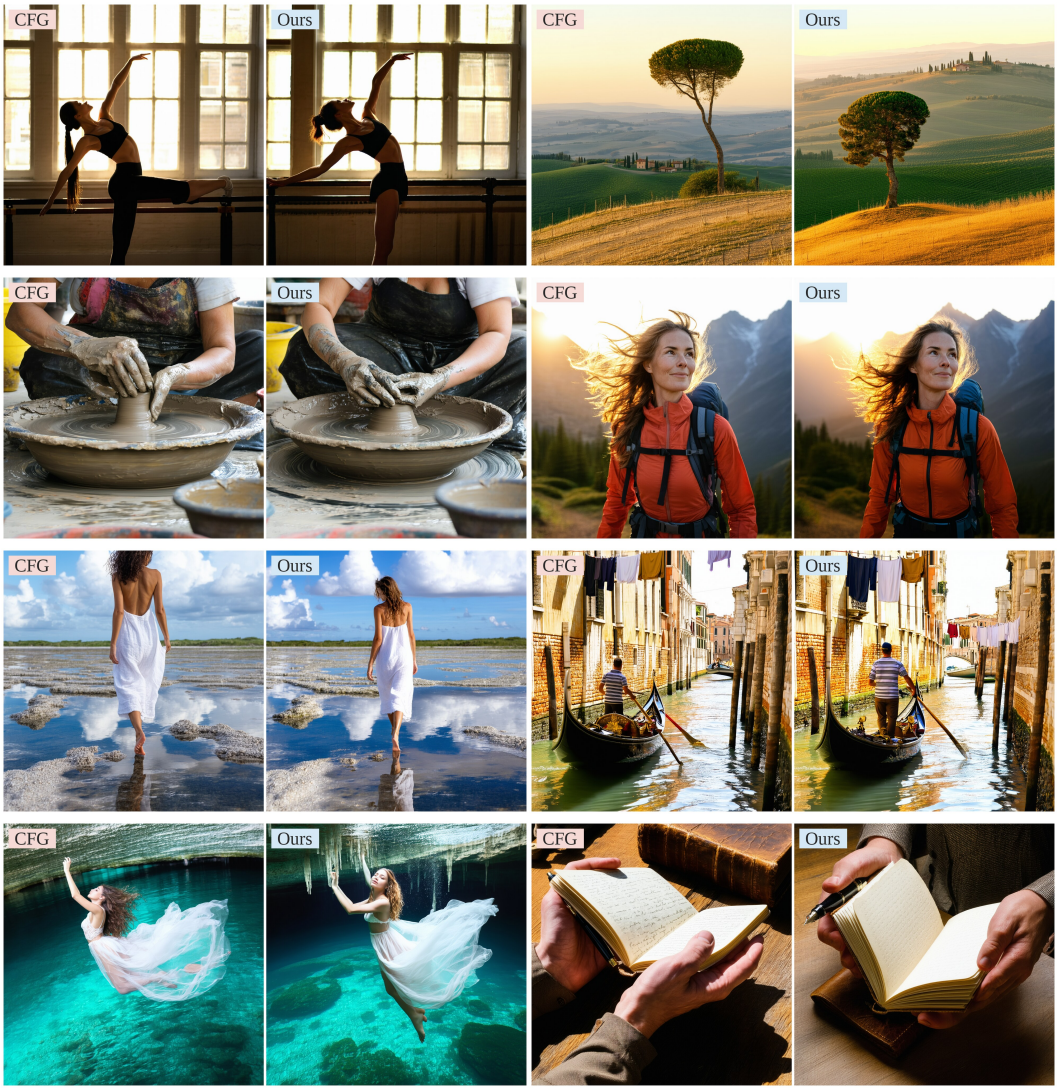


Figure 11: Additional qualitative comparisons between AdaMaG and CFG at their respective optimal guidance scales (part 2/4).

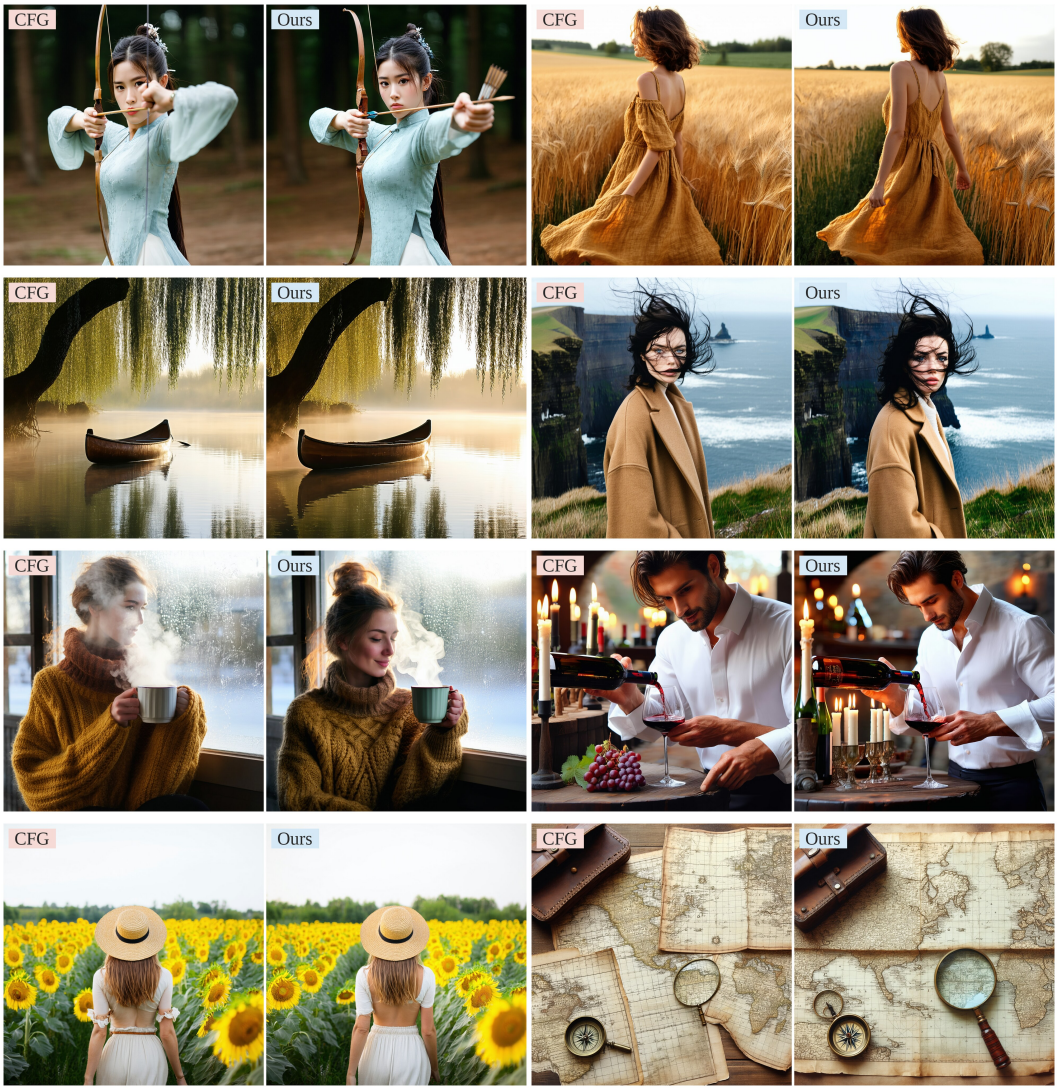


Figure 12: Additional qualitative comparisons between AdaMaG and CFG at their respective optimal guidance scales (part 3/4).

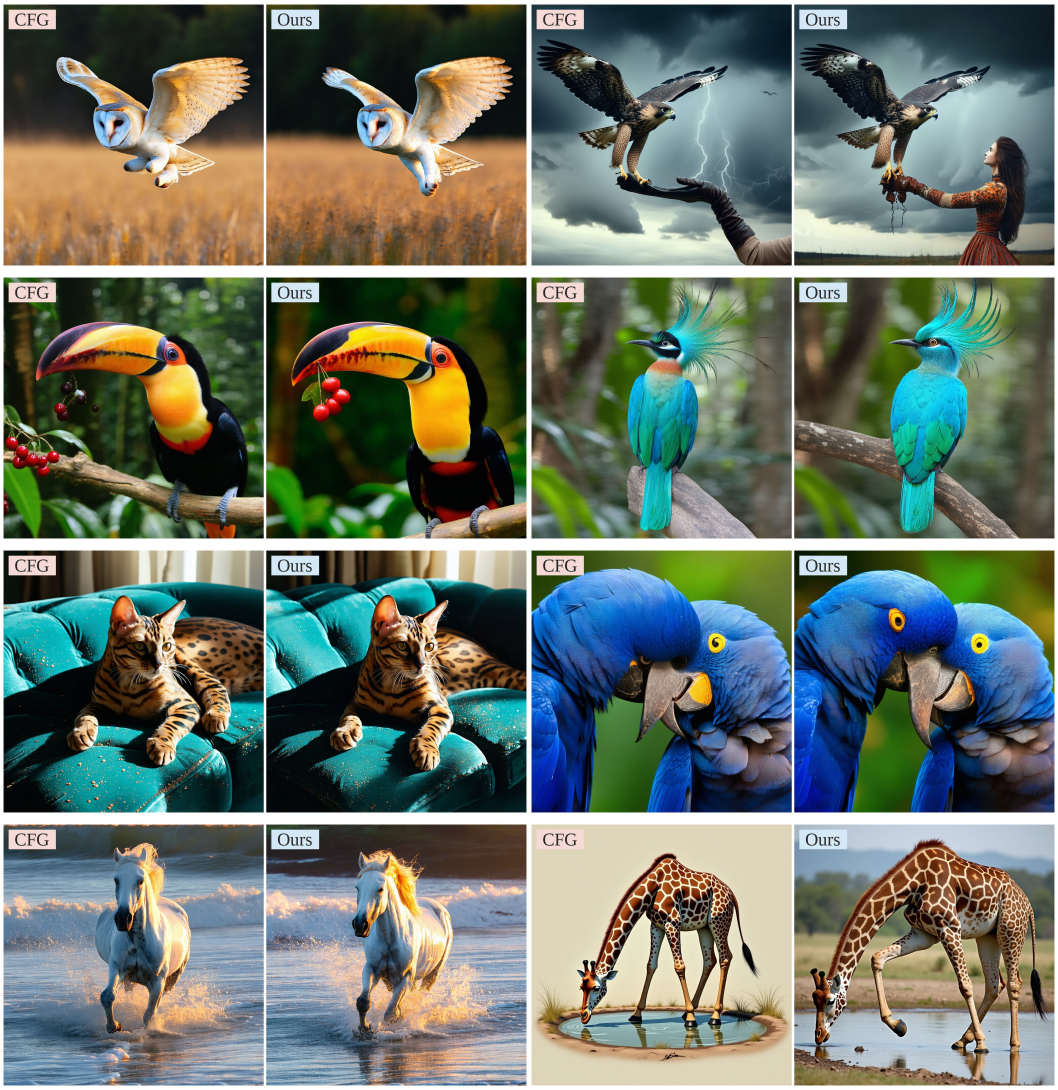


Figure 13: Additional qualitative comparisons between AdaMaG and CFG at their respective optimal guidance scales (part 4/4).

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The paper presents a simple yet Informative theoretical contribution with the title of conservation-based view of guidance. Built on this Adaptive Manifold Guidance is introduced with additional theoretical justifications. The guidance method is then applied on a various prompts and benchmarks and evaluated extensively all of which present in the paper.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The limitation is presented in the conclusion of the paper after summarising the contributions and findings.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: Paper provides one proposition in the main body, and an additional theoretical result in the appendix all of which provided with proofs, and the set of assumptions.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: Paper provides detailed explanation of the methodology and how experiments were done. Codebase is also linked in the abstract.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The code is released with the paper.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Experimental setup, benchmarks, and metrics are standard and presented with explanation in the experiments section. Additionally the baseline setup is also explained in the Appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Paper includes multi-seed runs averaged across seeds to increase the reproducibility and the accuracy of results.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Paper provides the compute resources in the experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: Paper fully conforms with NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: The contribution of the work is on the core science and although there are societal impacts, they are mostly related to the fundamental text-to-image generation framework and not a direct impact of improvements on those models.

    Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: Main models used in this work are deployed from Diffusers where safeguards are in place for the text-to-image generation misuse consequently these are not required in our work.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: Papers and codebase of the adopted assets have been cited and clearly distinguished in the paper.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.

- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The code is released with the paper.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [Yes]

    Justification: Paper provides the detail of the human-study presented and describes how the procedure was conducted.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [No] Justification: The human study consisted of a low-risk pairwise preference task in which evaluators viewed pairs of generated images and selected which they preferred along three axes.

    Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: Usage of LLM has been limited to grammar check.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.