Characteristic Guidance:
Non-linear Correction for Diffusion Model at Large Guidance Scale

Candi Zheng [* 1 2]   **Yuan Lan** [*]

Abstract

Popular guidance for denoising diffusion probabilistic model (DDPM) linearly combines distinct conditional models together to provide enhanced control over samples. However, this approach overlooks nonlinear effects that become significant when guidance scale is large. To address this issue, we propose characteristic guidance, a guidance method that provides first-principle nonlinear correction for classifier-free guidance. Such correction forces the guided DDPMs to respect the Fokker-Planck (FP) equation of diffusion process, in a way that is training-free and compatible with existing sampling methods. Experiments show that characteristic guidance enhances semantic characteristics of prompts and mitigate irregularities in image generation, proving effective in diverse applications ranging from simulating magnet phase transitions to latent space sampling.

1. Introduction

The diffusion model (Sohl-Dickstein et al., 2015; Song & Ermon, 2019; Song et al., 2020b) is a family of generative models that produce high-quality samples utilizing the diffusion processes of molecules. Denoising diffusion probabilistic model (DDPM) (Ho et al., 2020) is one of the most popular diffusion models whose diffusion process can be viewed as a sequence of denoising steps. It is the core of several large text-to-image models that have been widely adopted in AI art production, such as the latent diffusion model (stable diffusion) (Rombach et al., 2021) and DALL·E (Betker et al., 2023). DDPMs control the generation of samples by learning conditional distributions that are specified by sam-

ple contexts. However, such conditional DDPMs has weak control over context hence do not work well to generate desired content (Luo, 2022).

Guidance techniques, notably classifier guidance (Song et al., 2020b; Dhariwal & Nichol, 2021) and classifier-free guidance (Ho, 2022), provide enhanced control at the cost of sample diversity. Classifier guidance, requiring an additional classifier, faces implementation challenges in non-classification tasks like text-to-image generation. Classifier-free guidance circumvents this by linearly combining conditional and unconditional DDPM weighted by a guidance scale parameter. However, large guidance scale often leads to overly saturated and unnatural images (Ho, 2022). Though techniques like dynamic thresholding (Saharia et al., 2022) can handle color issues by clipping out-of-range pixel values, a systematic solution for general sampling tasks, including latent space diffusion (Rombach et al., 2021) and those beyond image generation, remains absent.

This paper proposes characteristic guidance as a systematic solution to the large guidance scale issue by addressing the non-linearity neglicted in classifier-free guidance. The name *characteristic* serves a dual purpose: Technically, characteristic guidance utilizes *the method of characteristics* in solving the score Fokker-Planck (FP) equation (Lai et al., 2022) for diffusion process, addressing its non-linearity to which the classifier-free guidance fails to adhere. It demonstrates clear theoretical advantages on Gaussian models in removing large guidance scale irregularities. Experimentally, characteristic guidance notably enhances the *semantic characteristics* of images, ensuring closer alignment with specified conditions. This is evident in our image generation experiments on CIFAR-10, ImageNet 256, and Stable diffusion (Fig.1). Notably, it also mitigates color and exposure issues common at large-scale guidance. Furthermore, characteristic guidance is compatible with any continuous data type and is robust across continuous diverse applications, ranging from simulating magnet phase transitions to latent space sampling.

---
[*]Equal contribution [1]Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong SAR, China [2]Department of Mechanics and Aerospace Engineering, Southern University of Science and Technology, Shenzhen, China. Correspondence to: Candi Zheng <czhengac@connect.ust.hk>.
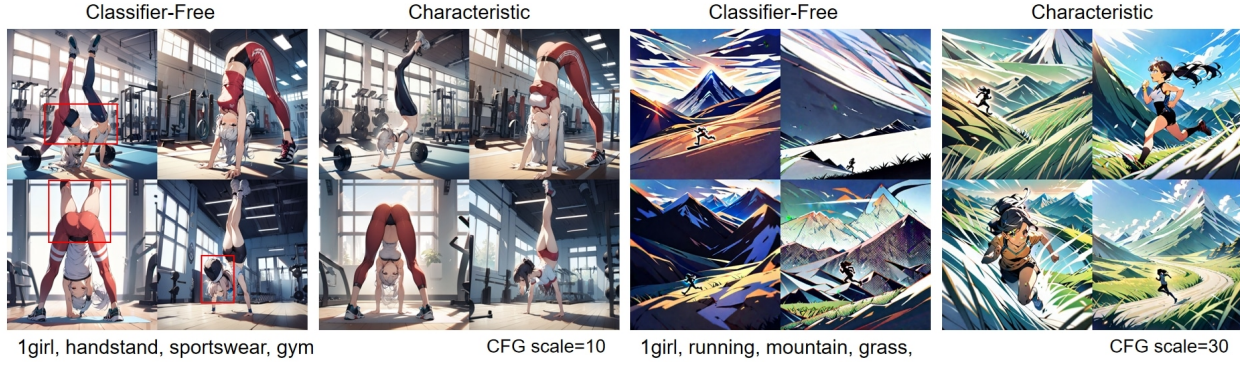
Figure 1. Comparative visualization of images sampled from Stable diffusion XL (Podell et al., 2023) between Classifier Free Guidance and Characteristic Guidance (Model name: animagineXL 3.0 (CagliostroLab, 2023), Seeds: 0,1,2,3). By addressing the non-linear effects of the FP equation, characteristic guidance demonstrates ability to mitigate irregularity in color, exposure and anatomy and enhancing prompt's semantic characteristics (e.g., "grass" depicted in the right-hand example).

2. Related Works

2.1. Diffusion Models and Guidance Techniques

Diffusion models, notably Denoising Diffusion Probabilistic Models (DDPMs) (Ho et al., 2020), have become pivotal in generating high-quality AI art (Sohl-Dickstein et al., 2015; Song & Ermon, 2019; Song et al., 2020b). Various guidance methods, including classifier-based (Song et al., 2020b; Dhariwal & Nichol, 2021) and classifier-free techniques (Ho, 2022), have been developed. Recent efforts aim to refine control styles (Bansal et al., 2023; Yu et al., 2023) and enhance sampling quality (Hong et al., 2022; Kim et al., 2022). However, current guidance methods suffer from quality degradation and color saturation issues at high guidance scales. Dynamic thresholding (Saharia et al., 2022) attempts to address this by normalizing quantiles of pixels, but its effectiveness is limited to color-based image pixels, falling short in general tasks like latent space sampling. To suppress artifacts, community later had to use "scale mimic" that mixes dynamic thresholded latents with low guidance scale latents (Mcmonkeyprojects, 2024). Instead of these workarounds, our research aims to address quality issues of classifier-free guidance theoretically and systematically for general tasks, by tackling the non-linear aspects of DDPMs.

2.2. Fast Sampling Methods

The significant challenge posed by the slow sampling speed of DDPMs, requiring numerous model evaluations, has spurred the development of accelerated sampling strategies. These strategies range from diffusion process approximations (Song et al., 2020b; Liu et al., 2022; Song et al., 2020a; Zhao et al., 2023) to employing advanced solvers like Runge–Kutta and predictor-corrector methods (Karras et al., 2022; Lu et al., 2022; Zhao et al., 2023). These methods primarily utilize scores or predicted noises for de-noising. Our research focuses on designing a high guidance scale

correction that seamlessly integrates with these fast sampling methods, primarily through the modification of scores or predicted noises.

3. Background

3.1. DDPM and the FP Equation

DDPM models distributions of images $p(\mathbf{x})$ by recovering an original image $\mathbf{x}_0 \sim p(\mathbf{x})$ from one of its noise contaminated versions $\mathbf{x}_i$. The noise contaminated image is a linear combination of the original image and a gaussian noise

$$\mathbf{x}_i = \sqrt{\bar{\alpha}_i}\mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_i}\bar{\boldsymbol{\epsilon}}_i; \quad 1 \le i \le n, \qquad (1)$$

where n is the total diffusion steps, $\bar{\alpha}_i$ is the contamination weight at a time $t_i \in [0, T]$ in the forward diffusion process, and $\bar{\boldsymbol{\epsilon}}_i$ is a standard Gaussian random noise. More detailed information can be found in Appendix A.

The DDPM trains a denoising neural network $\boldsymbol{\epsilon}_\theta(\mathbf{x}, t_i)$ to predict and remove the noise $\bar{\boldsymbol{\epsilon}}_i$ from $\mathbf{x}_i$. It minimizes the denoising objective (Ho et al., 2020):

$$L(\boldsymbol{\epsilon}_\theta) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}_{\mathbf{x}_0 \sim p(\mathbf{x}), \bar{\boldsymbol{\epsilon}}_i \sim \mathcal{N}(\mathbf{0},I)}\|\bar{\boldsymbol{\epsilon}}_i - \boldsymbol{\epsilon}_\theta(\mathbf{x}_i, t_i)\|_2^2. \ (2)$$

This neural network is later used in the backward diffusion process to generate image samples from pure Gaussian noises. In the ideal scenario where diffusion time steps among t_i are **infinitesimal**, the optimal solution for (2) is defined by:

$$\boldsymbol{\epsilon}(\mathbf{x}, t) = \operatorname*{argmin}_{\boldsymbol{\epsilon}_\theta} L(\boldsymbol{\epsilon}_\theta). \qquad (3)$$

The forward process (1) places stringent constraints on its permissible forms. First, $\boldsymbol{\epsilon}(\mathbf{x}, t_i)$ is proportional to the score function $\mathbf{s}(\mathbf{x}, t_i) = \nabla_{\mathbf{x}_i} \log p(\mathbf{x}_i)$ (Yang et al., 2022) (Appendix A). Second, the score function is a solution of the

score FP equation (Lai et al., 2022), which could be rewritten (Appendix C) in terms of $\boldsymbol{\epsilon}(\mathbf{x}, t)$ as

$$\frac{\partial \boldsymbol{\epsilon}}{\partial t} = \frac{1}{2} \left(\mathcal{L}\boldsymbol{\epsilon} - \frac{1}{\sigma(t)} \nabla_{\mathbf{x}} \|\boldsymbol{\epsilon}\|_2^2 \right), \quad (4)$$

in which $\mathcal{L}\boldsymbol{\epsilon} = \nabla_{\mathbf{x}}(\boldsymbol{\epsilon} \cdot \mathbf{x}) + \nabla_{\mathbf{x}}^2 \boldsymbol{\epsilon} + \frac{1-\sigma(t)^2}{\sigma(t)^2} \boldsymbol{\epsilon}$ and $\sigma(t) = \sqrt{1 - e^{-t}}$. These constraints lay the foundation for the duality between forward and backward diffusion processes, which is essential for successful sampling from $p(\mathbf{x})$.

3.2. Conditional DDPM and Classifier-Free Guidance

Conditional DDPMs, which generate images based on a given condition $\mathbf{c}$, model the conditional distribution $p(\mathbf{x}|\mathbf{c})$ with a denoising neural network represented as $\boldsymbol{\epsilon}_\theta(\mathbf{x}|\mathbf{c}, t_i)$. However, the training data in practice might only have weak or noised information about condition $\mathbf{c}$, therefore we need a way to enhance the control strength in this situation.

Guidance (Song et al., 2020b; Dhariwal & Nichol, 2021; Ho, 2022) is a technique for conditional image generation that trades off control strength and image diversity. It generally aims to sample from the distribution:

$$p(\mathbf{x}|\mathbf{c}, \omega) \propto p(\mathbf{x}|\mathbf{c})^{1+\omega} p(\mathbf{x})^{-\omega} \propto p(\mathbf{c}|\mathbf{x})^{1+\omega} p(\mathbf{x}), \quad (5)$$

where $\omega > 0$ is the guidance scale. When ω is large, this distribution concentrates on samples that have the highest conditional likelihood $p(\mathbf{c}|\mathbf{x})$.

Classifier free guidance (Ho, 2022) use the following guided denoising neural network $\boldsymbol{\epsilon}_{CF}$, deduced from (5) using $\boldsymbol{\epsilon} \propto \nabla \log p$, to approximately sample from $p(\mathbf{x}|\mathbf{c}, \omega)$:

$$\boldsymbol{\epsilon}_{CF}(\mathbf{x}|\mathbf{c}, t_i, \omega) = (1 + \omega)\, \boldsymbol{\epsilon}_\theta(\mathbf{x}|\mathbf{c}, t_i) - \omega\, \boldsymbol{\epsilon}_\theta(\mathbf{x}, t_i). \quad (6)$$

It exactly computes $\boldsymbol{\epsilon}(\mathbf{x}|\mathbf{c}, t_i, \omega)$, the de-noising neural network of $p(\mathbf{x}|\mathbf{c}, \omega)$, at time $t_i = 0$ (Appendix B). However, it is not a good approximation for $t_i > 0$ as it deviates from the FP equation (4) when ω is large.

3.3. Guidance's Deviation from the FP Equation

We measure the deviation of a guidance method from the FP equation using the **mixing error**:

$$\mathbf{e}_m(\boldsymbol{\epsilon}, \mathbf{x}, t) = \frac{\partial \boldsymbol{\epsilon}}{\partial t} - \frac{1}{2} \left(\mathcal{L}\boldsymbol{\epsilon} - \frac{\nabla_{\mathbf{x}} \|\boldsymbol{\epsilon}\|_2^2}{\sigma(t)} \right). \quad (7)$$

in which $\boldsymbol{\epsilon}$ is the guided denoising neural network under **infinitesimal** diffusion time steps.

For classifier-free guidance, the mixing error comes from the non-linear term of (4) as

$$\mathbf{e}_m(\boldsymbol{\epsilon}_{CF}, \mathbf{x}, t) = \frac{1}{\sigma(t)} \nabla_{\mathbf{x}}(\|\boldsymbol{\epsilon}_{CF}(\mathbf{x}|\mathbf{c}, t_i, \omega)\|_2^2 - \quad (8)$$
$$(1+\omega)\|\boldsymbol{\epsilon}_\theta(\mathbf{x}|\mathbf{c}, t_i)\|_2^2 + \omega\|\boldsymbol{\epsilon}_\theta(\mathbf{x}, t_i)\|_2^2),$$

whose amplitude escalates with an increase in ω and further intensifies as $\sigma(t)$ decreases.

3.4. The Method of Characteristics

The method of characteristics (Evans, 2022) provides an analytical approach to solving certain kind of non-linear partial differential equations (PDE). It employs characteristic lines, a family of time-dependent trajectories defined by $\mathbf{x}_t = \mathbf{x}(t)$, along which the equation

$$\boldsymbol{\epsilon}(\mathbf{x}_t, t) = \mathbf{F}_t(\boldsymbol{\epsilon}(\mathbf{x}_0, 0)) \quad (9)$$

holds, where $\boldsymbol{\epsilon}$ is a solution of the PDE and $\mathbf{F}_t$ is a deterministic and invertible function. This implies that the value of $\boldsymbol{\epsilon}(\mathbf{x}_t, t)$ can be inferred from its initial condition $\boldsymbol{\epsilon}(\mathbf{x}_0, 0)$ at a specific point $\mathbf{x}_0$, determined by the characteristic line itself.

The method of characteristics is applicable to (4) under certain assumptions, whose characteristic lines are detailed in (50) of Appendix D. Particularly, the function $\mathbf{F}_t$ for (4) is a linear function.

4. Problem Description

Our method emphasizes the importance of adhering to the Fokker-Planck (FP) equation for two main reasons: Theoretically, the deviation from the FP equation leads to mixing errors, disrupting the equivalence between forward and backward diffusion processes and resulting in sample generation irregularities. Empirically, as we will shown in Sec.6.1, aligning with the FP equation helps to remove irregularities of classifier-free guidance.

We focus on a guided DDPM that generates samples from $p(\mathbf{x}|\mathbf{c}, \omega)$ (5), constructing its denoising neural network $\boldsymbol{\epsilon}_{CH}(\mathbf{x}|\mathbf{c}, t_i, \omega)$ from two known networks: $\boldsymbol{\epsilon}_\theta(\mathbf{x}|\mathbf{c}, t_i)$ and $\boldsymbol{\epsilon}_\theta(\mathbf{x}, t_i)$. The desired properties of $\boldsymbol{\epsilon}_{CH}$ include:

- $\boldsymbol{\epsilon}_{CH}$ has no mixing error (satisfying FP equation);

- $\boldsymbol{\epsilon}_{CH}$ requires no training.

Constructing such $\boldsymbol{\epsilon}_{CH}$ is highly non-trivial as it requires tackling the non-linear term in the Fokker-Planck equation.

5. Methodology

5.1. The Characteristic Guidance

We propose the characteristic guidance as non-linear corrected classifier-free guidance:

$$\boldsymbol{\epsilon}_{CH}(\mathbf{x}|\mathbf{c}, t_i, \omega) = (1 + \omega)\, \boldsymbol{\epsilon}_\theta(\mathbf{x}_1|\mathbf{c}, t_i) - \omega\, \boldsymbol{\epsilon}_\theta(\mathbf{x}_2, t_i), \quad (10)$$

in which $\mathbf{x}_1 = \mathbf{x} + \omega\Delta\mathbf{x}$, $\mathbf{x}_2 = \mathbf{x} + (1+\omega)\Delta\mathbf{x}$, and $\Delta\mathbf{x}$ is a non-linear correction term. It is evident that when $\Delta\mathbf{x} = 0$, the characteristic guidance is equivalent to the classifier-free guidance (6).

The correction $\Delta\mathbf{x}$ is determined from the training-free non-linear relation:

$$\Delta\mathbf{x} = \mathbf{P} \circ (\boldsymbol{\epsilon}_\theta(\mathbf{x}_2, t_i) - \boldsymbol{\epsilon}_\theta(\mathbf{x}_1|\mathbf{c}, t_i))\, \sigma_i, \qquad (11)$$

where $\sigma_i = \sqrt{1 - \bar{\alpha}_i}$ is a scale parameter and the operator $\mathbf{P}$ could be an arbitrary orthogonal projection operator.

Equation (11) can be solved using the fixed-point iteration method (Evans, 2022) (Appendix E). For practical efficiency, we propose accelerated fixed-point iteration algorithms, Alg.2 and Alg.3, to minimize the iterations needed for convergence.

The projection operator $\mathbf{P}$ should be theoretically the identity, but we found that orthogonal projections that serve as regularization greatly accelerates the computation. In practice, $\mathbf{P}$ acts channel-wisely and is specified by a vector $\mathbf{g}$:

$$\mathbf{P}_\mathbf{g} \circ \mathbf{v} = \frac{\mathbf{g} \cdot \mathbf{v}}{\mathbf{g} \cdot \mathbf{g}}\mathbf{g} \qquad (12)$$

For pixel space diffusion model, we suggest the operator $\mathbf{g} = \mathbf{1}$ as projection to the channel-wise mean. For latent space diffusion model, we suggest $\mathbf{g} = (\boldsymbol{\epsilon}_\theta(\mathbf{x}, t_i) - \boldsymbol{\epsilon}_\theta(\mathbf{x}|\mathbf{c}, t_i))\, \sigma_i$. For low dimensional cases that are not images, we suggest the operator $\mathbf{P}$ to be identity.

5.2. Theoretical Foundation

This section derives the characteristic guidance (10)-(11) by solving the FP equation (4) under the Harmonic ansatz. We propose the Harmonic ansatz assuming that the optimal DDPM solution $\epsilon(\mathbf{x}, t)$ of (3) is harmonic:

Ansatz 5.1. *The Harmonic Ansatz: The optimal solution $\epsilon(\mathbf{x}, t)$ is harmoic*

$$\nabla_\mathbf{x}^2 \boldsymbol{\epsilon}(\mathbf{x}, t) = 0. \qquad (13)$$

This ansatz is inspired by and holds exactly when the distribution of original images $p(\mathbf{x})$ is a Gaussian distribution, and we will show that it works for other cases as well in experiments 6.1 and 6.2.

Characteristic guidance adopts the same formulation as classifier-free guidance at $t = 0$

$$\boldsymbol{\epsilon}_{CH}(\mathbf{x}_0|\mathbf{c}, 0, \omega) = (1 + \omega)\, \boldsymbol{\epsilon}(\mathbf{x}_0|\mathbf{c}, 0) - \omega\, \boldsymbol{\epsilon}(\mathbf{x}_0, 0), \quad (14)$$

in which parameters θ are omitted since we are considering optimal solutions with infinitesimal time steps. This guarantees that $\boldsymbol{\epsilon}_{CH}$ samples from $p(\mathbf{x}|\mathbf{c}, \omega)$, which aligns with the target of classifier-free guidance.

Next, our aim is to construct the unknown $\boldsymbol{\epsilon}_{CH}$ for any time $t > 0$ using the known $\boldsymbol{\epsilon}(\mathbf{x}|\mathbf{c}, t)$ and $\boldsymbol{\epsilon}(\mathbf{x}, t)$, by solving the Fokker-Planck equation (4) analytically.

The harmonic ansatz 5.1 simplifies the Fokker-Planck equation (4) into a quasi-linear equation, which can be solved by the method of characteristics in Sec.3.4. Assuming $\boldsymbol{\epsilon}_{CH}(\mathbf{x}|\mathbf{c}, t, \omega)$, $\boldsymbol{\epsilon}(\mathbf{x}|\mathbf{c}, t)$, and $\boldsymbol{\epsilon}(\mathbf{x}, t)$ are three solutions of (4), there exist three distinct characteristic lines $\mathbf{x}(t)$, $\mathbf{x}^{(1)}(t)$, and $\mathbf{x}^{(2)}(t)$ passing through the point $\mathbf{x}_0$ at $t = 0$, satisfying

$$\begin{aligned} \boldsymbol{\epsilon}_{CH}(\mathbf{x}_0|\mathbf{c}, 0, \omega) &= \mathbf{F}_t^{-1}\left[\boldsymbol{\epsilon}_{CH}(\mathbf{x}_t|\mathbf{c}, t, \omega)\right], \\ \boldsymbol{\epsilon}(\mathbf{x}_0|\mathbf{c}, 0) &= \mathbf{F}_t^{-1}\left[\boldsymbol{\epsilon}(\mathbf{x}_t^{(1)}|\mathbf{c}, t)\right], \\ \boldsymbol{\epsilon}(\mathbf{x}_0, 0) &= \mathbf{F}_t^{-1}\left[\boldsymbol{\epsilon}(\mathbf{x}_t^{(2)}, t)\right], \end{aligned} \qquad (15)$$

in which the function $\mathbf{F}_t$ for characteristic lines of (4) is a linear function. Consequently, these characteristic lines and the initial condition (14) yield an elegant analytical solution of $\boldsymbol{\epsilon}_{CH}$:

$$\boldsymbol{\epsilon}_{CH}(\mathbf{x}_t|\mathbf{c}, t, \omega) = (1 + \omega)\, \boldsymbol{\epsilon}(\mathbf{x}_t^{(1)}|\mathbf{c}, t) - \omega\, \boldsymbol{\epsilon}(\mathbf{x}_t^{(2)}, t). \quad (16)$$

This result establishes the possibility of constructing $\boldsymbol{\epsilon}_{CH}$ using $\boldsymbol{\epsilon}_\theta(\mathbf{x}|\mathbf{c}, t_i)$ and $\boldsymbol{\epsilon}_\theta(\mathbf{x}, t_i)$. Its precise form is described in the following lemma:

Lemma 5.2. *Let $\boldsymbol{\epsilon}(\mathbf{x}, t)$, $\boldsymbol{\epsilon}_1(\mathbf{x}, t)$, and $\boldsymbol{\epsilon}_2(\mathbf{x}, t)$ be three distinct solutions of the FP equation (4), satisfying the Harmonic Ansatz 5.1. Moreover, their initial condition satisfies*

$$\boldsymbol{\epsilon}(\mathbf{x}, 0) = (1 + \omega)\boldsymbol{\epsilon}_1(\mathbf{x}, 0) - \omega\boldsymbol{\epsilon}_2(\mathbf{x}, 0), \qquad (17)$$

Then, we have the relation

$$\boldsymbol{\epsilon}(\mathbf{x}, t) = (1 + \omega)\boldsymbol{\epsilon}_1(\mathbf{x} + \omega\Delta\mathbf{x}, t) - \omega\boldsymbol{\epsilon}_2(\mathbf{x} + (1 + \omega)\Delta\mathbf{x}, t), \qquad (18)$$

where $\Delta\mathbf{x}$ is given by

$$\Delta\mathbf{x} = (\boldsymbol{\epsilon}_2(\mathbf{x} + (1 + \omega)\Delta\mathbf{x}, t) - \boldsymbol{\epsilon}_1(\mathbf{x} + \omega\Delta\mathbf{x}, t))\, \sigma(t), \qquad (19)$$

in which $\sigma(t) = \sqrt{1 - e^{-t}}$.

The proof of lemma 5.2 is established in Appendix D. Replacing the continuous functions $\sigma(t)$, $\boldsymbol{\epsilon}$, $\boldsymbol{\epsilon}_1$, and $\boldsymbol{\epsilon}_2$ with their discretized version σ_i, $\boldsymbol{\epsilon}_{CH}(\mathbf{x}|\mathbf{c}, t_i, \omega)$, $\boldsymbol{\epsilon}_\theta(\mathbf{x}|\mathbf{c}, t_i)$, and $\boldsymbol{\epsilon}_\theta(\mathbf{x}, t_i)$ gives us the characteristic guidance. Besides, a direct consequence of lemma 5.2 is the following Theorem

Theorem 5.3. *The characteristic guidance has **no mixing error** when the Harmonic ansatz applies and diffusion time steps are infinitesimal:*

$$\mathbf{e}_m(\boldsymbol{\epsilon}_{CH}, \mathbf{x}, t) = 0. \qquad (20)$$

In summary, $\boldsymbol{\epsilon}_{CH}$ is an exact solution to the FP equation (4) with initial condition (14) under the harmonic ansatz, eliminating the mixing error caused by non-linearity of (4).

6. Experiments

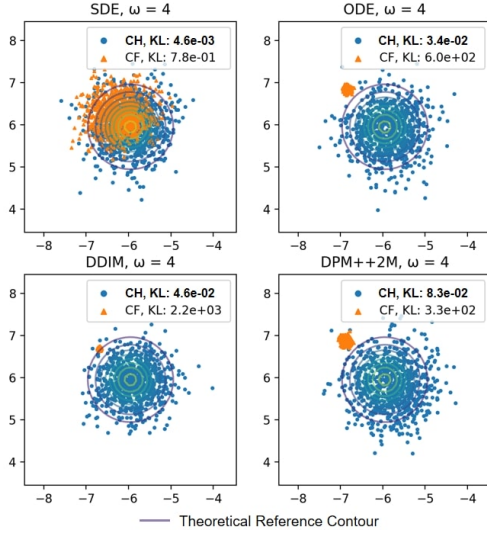## 6.1. Gaussians: Where the Harmonic Ansatz Holds

Figure 2. Samples and the KL divergence from characteristic guidance (CH) and classifier free guidance (CF) guided DDPM modeling conditional Gaussian distribution. The contours corresponds to the theoretical reference (ground truth) distribution of the guided DDPM (5).

This section demonstrates the theoretical advantage of characteristic guidance on Gaussian models, where we can compare samples against analytical solutions. We conducted experiments to compare classifier-free guidance and characteristic guidance under two scenarios: sampling from (1) conditional 2D Gaussian distributions, where the harmonic ansatz is exact; (2) mixture of Gaussian distributions, where the harmonic ansatz holds for most of the region except where the components overlap significantly. The guided diffusion model (39) aims to sample from the distribution (5) whose contour is plotted in Fig.2 and Fig.3.

In the conditional Gaussian scenario, we analytically set two DDPMs for the conditional distribution $p(\mathbf{x}|\mathbf{c})$ and the unconditional distribution $p(\mathbf{x})$ as follows:

$$
\begin{aligned}
p(\mathbf{x}|\mathbf{c}) &= \mathcal{N}\left(x_1, x_2|(c_1, c_2)^T, I\right) \\
p(\mathbf{x}) &= \mathcal{N}\left(x_1, x_2|(0,0)^T, 5I\right).
\end{aligned}
\tag{21}
$$

where $c_1 = -5, c_2 = 5$ in the experiment. In the mixture of Gaussian, we trained conditional and unconditional DDPMs to learn the distributions $p(\mathbf{x}|\mathbf{c}) = \prod_{i=0}^{2} \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_i, I)^{c_i}$ and $p(\mathbf{x}) = \frac{1}{3}\sum_{\{\mathbf{e}_i, i=0,1,2\}} p(\mathbf{x}|\mathbf{e}_i)$, where $\mathbf{c} = (c_0, c_1, c_2)^T$ represents a three dimensional one-hot vector, $\boldsymbol{\mu}_0 = (-1, -1/\sqrt{3})^T$, $\boldsymbol{\mu}_1 = (1, -1/\sqrt{3})^T$, $\boldsymbol{\mu}_2 = (0, \sqrt{3} - 1/\sqrt{3})^T$, $\{\mathbf{e}_i, i = 0, 1, 2\}$ represents all one-hot vectors in three dimension space.
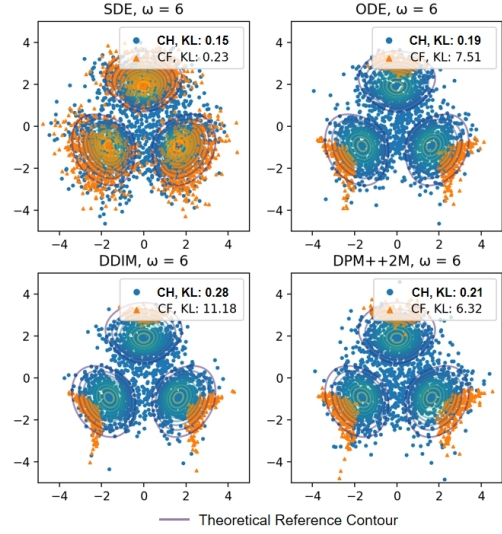

Figure 3. Comparison between CH (ours) and CF guided DDPM on modeling mixture of Gaussian distribution. The contours corresponds to the theoretical reference distribution of the guided DDPM (5). Samples from characteristic guidance shows better KL divergence than those from classifier-free guidance.

In both experiments, we evaluated four different sampling methods: SDE (35) (1000 steps), probabilistic ODE (Song et al., 2020b) (1000 steps), DDIM (Song et al., 2020a) (20 steps), and DPM++2M (Lu et al., 2022) (20 steps). The projection operator $\mathbf{P}$ for characteristic guidance was set to identity. Samples were compared against the theoretical reference using Kullback–Leibler (KL) divergence.

In the conditional Gaussian scenario, Fig.2 shows that characteristic guidance achieves better KL divergence than classifier-free guidance for every sampling method. It's worth noting that classifier-free guidance suffers from severe bias and catastrophic loss of diversity when ODE based sampling methods (ODE, DDIM, and DPM++2M) are used. Contrarily, characteristic guidance always yields correct sampling with KL divergence better than classifier-free guidance's best SDE samples. In the Mixture of Gaussian scenario, Fig.3 shows that characteristic guidance again showed less bias and better diversity. However, the overlap regions in the mixture of Gaussian scenario led to less clear component boundaries. This is because the harmonic ansatz is not valid in the overlap regions. Besides, minor artifacts appear because iterating on the trained neural network $\boldsymbol{\epsilon}$ brings in approximation errors. Overall, characteristic guidance outperformed classifier-free guidance in reducing bias, increasing diversity, and narrowing the performance gap between ODE and SDE sampling methods. This was consistently observed in cases where the harmonic ansatz holds exactly or approximately. More results could be found in Fig.8 and Fig.9.

5

6.2. Cooling of the Magnet: Where the Harmonic Ansatz Not Supposed to Hold
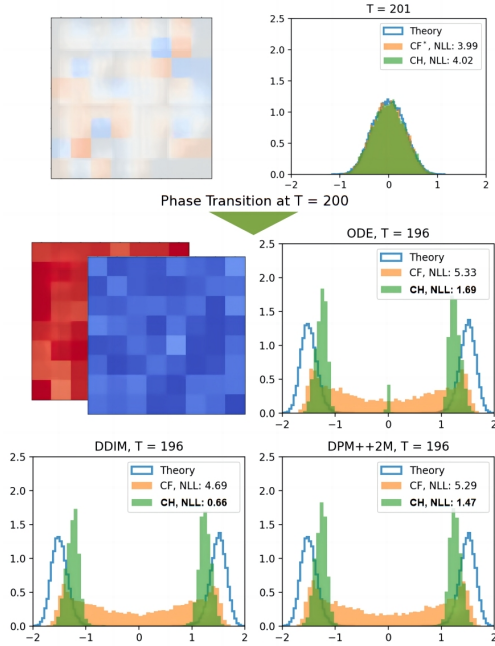

Figure 4. Comparison between CH (ours) and CF guided DDPM on simulating magnet cooling. The images in the upper and middle left (blue and red lattices) depict samples from DDPMs, while the histograms illustrate the distribution of samples' mean value (magnetization) across different temperatures. The contours corresponds to the theoretical reference distribution of magnetization. The characteristic guidance has better NLL and is more capable in capturing peak separation of sample magnetization.

In this section, we explore a simulation of the cooling of a magnet, comparing our method's performance with that of classifier-free guidance. This simulation is not just a representation of a complex real-world physical phenomenon but also a test of our characteristic guidance on a case where the harmonic ansatz does not hold.

We focus on the cooling around the Curie temperature, a critical point where a paramagnetic material undergoes a phase transition to permanent magnetism. This transition, driven by 3rd order terms in the score function, challenges the harmonic ansatz. Despite this, our method successfully captures the most important phenomenon of this transition: the separation of two distinct peaks below the Curie temperature, which classifier-free guidance fails to reproduce. These results emphasize the robustness and effectiveness of our method in modeling intricate physical processes, even in scenarios where the harmonic ansatz does not hold.

Suppose we have a square slice of 2-dimensional paramagnetic material which can be magnetized only in the direction perpendicular to it. It can be further divided uniformly into

8×8 smaller pieces, with each pieces identified by its index (i, j). We use a scalar field $\phi_{i,j}$ to record the magnitude of magnetization of each piece, which can be viewed as a single channel 8×8 image of magnetization.

The probability of observing a particular configuration of ϕ, according to statistical physics, is proportional to the Boltzmann distribution

$$p(\phi; T) \propto e^{-\beta H(\phi; T)}, \qquad (22)$$

where $\beta H(\phi; T)$ is Hamiltonian at temperature T, assigning an energy to each of possible ϕ. At temperature near the Curie temperature T_c, a Landau-Ginzburg model (Kardar, 2007) of magnet uses a Hamiltonian (67) (Appendix H) with 4-th order terms. Such terms are the key to characterise phase transition of magnet, but its gradient (score function) is a third order term that does not respect the harmonic ansatz.

A remarkable property about the Landau-Ginzburg model, bearing surprising similarity with the DDPM guidance (5), is

$$p(\phi; (1+\omega)T_1 - \omega T_0) \propto p(\phi; T_1)^{1+\omega} p(\phi; T_0)^{-\omega}, \quad (23)$$

where ω is the guidance scale, $T_1 > T_0$ are two distinct temperature values. This means if we know the distribution of our magnet at two distinct temperatures T_0 and T_1, we know its distribution at any temperature cooler than T_1 by adjusting the guidance scale ω.

We train a conditional DDPM simulating the cooling of a magnet with the Curie temperature $T_c = 200$. The model is trained at two distinct temperatures $p(\phi|T_0 = 201)$ and $p(\phi|T_1 = 200)$, corresponds to $\omega = -1$ and $\omega = 0$. Samples at lower temperatures $T = 196$ ($\omega = 4$) are later obtained by adjusting the guidance scale. Four distinct sampling methods, identical to those discussed in Sec.6.1, are evaluated. The projection operator $\mathbf{P}$ for characteristic guidance is set to be the channal-wise mean.

Fig.4 demonstrates the phase transition of mean magnetization, which presents the histograms of the mean value of ϕ sampled from the trained DDPM (8192 samples). Above the Curie temperature $T_c = 200$, the histogram of the mean magnetization has one peak centered at 0. Below the Curie temperature, the histogram of the mean magnetization has two peaks with non-zero centers. More details are available in Fig.10. The change in the number of peaks represents a phase change from non-magnet to a permanent magnet.

Characteristic guidance outperforms classifier-free guidance in sample quality below the Curie temperature, showcasing better negative log-likelihood (NLL) and effectively capturing distinct peaks in magnetization of samples. In contrast, classifier-free guidance struggles to produce distinct peaks. Despite its slightly biased peak positions compared with
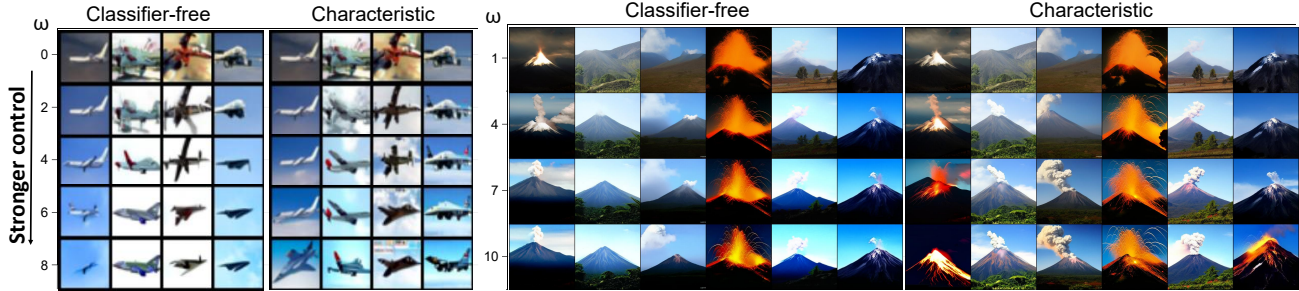
Figure 5. Left: CIFAR-10 aircraft images generated via DDPM highlight the difference between Classifier-free Guidance (CF) and Characteristic Guidance (CH) across various guidance scales (ω). The CF-guided images tend to have dull or even white backgrounds at higher ω, whereas the CH-guided images creates more vibrant scenes with skies and clouds. *Right*: In ImageNet 256, volcano samples generated using latent diffusion models with CF and CH Guidance without cherry-picking. CF images show color cast and underexposure at higher ω, while CH images maintain consistent color and exposure, better highlighting volcanic features, such as smoke and lava. For these visual comparisons, consistent initial noise ensures that both CF and CH guided images maintain similar contexts at lower ω values.

theoretical expectations, characteristic guidance still demonstrates superior performance in describing distinct peaks and modeling phase transitions.

6.3. CIFAR-10: Natural Image Generation

In this study, we evaluate the effectiveness of classifier-free (CF) and characteristic (CH) guidance in generating natural images using a DDPM trained on the CIFAR-10 dataset. We assess the quality of the generated images using Frechet Inception Distance (FID), Inception Score (IS), and visual inspection (Sec 6.6).

Table 1. Comparison of FID and IS for CF and CH guided DDPM on the CIFAR-10 dataset, using DDIM (50 steps) and DPM++2M (50 steps). The operator $\mathbf{P}$ for CH is the projection from the input to its channel-wise mean. Characteristic guidance generally achieves lower FID and comparable IS, particularly at higher guidance scales ω.

	DDIM				DPM++2M			
	FID ↓		IS ↑		FID ↓		IS ↑	
ω	CF	CH	CF	CH	CF	CH	CF	CH
0.3	4.52	**4.46**	9.18	**9.23**	**3.33**	3.35	9.69	9.69
0.6	4.80	**4.64**	9.39	**9.43**	3.51	**3.44**	9.93	9.93
1.0	6.22	**5.86**	9.52	**9.53**	4.75	**4.51**	**10.06**	10.04
1.5	8.56	**7.89**	9.50	**9.57**	6.85	6.37	**10.08**	10.07
2.0	11.02	**10.10**	9.49	**9.51**	9.11	**8.34**	10.04	10.04
4.0	19.85	**18.15**	**9.23**	9.22	17.04	**15.52**	9.75	**9.76**
6.0	27.04	**24.77**	8.85	**8.86**	23.47	**21.46**	9.33	**9.38**

Our experiments, detailed in Table.1, demonstrate that characteristic guidance achieves a lower FID while maintaining a comparable IS to classifier-free guidance, illustrating its effectiveness in balancing control strength and diversity. Unlike the typical trade-off between FID and IS in guidance methods (Ho, 2022), where enhancing diversity marked by improved FID often reduces control strength by a decrease in IS, characteristic guidance improves FID without compromising control strength. These results underscore its

advantage in producing high-fidelity images while preserving both diversity and control.

6.4. ImageNet 256: Correction for Latent Space Model

The characteristic guidance's effect on latent space diffusion models (LDM) (Rombach et al., 2021) is different: it enhances the control strength rather than lowering the FID. We test the characteristic guidance on the LDM for ImageNet256 dataset, with codes and models adopted from the LDM paper (Rombach et al., 2021). The quality of generated images are evaluated by FID, IS and visual inspection (Sec 6.6).

Fig.6 presents sample quality metrics using DDIM (50 steps) and DPM++2M (50 steps). While CH improves IS, it adversely affects FID compared to CF. This IS-FID trade-off complicates performance evaluation. Notably, the IS-FID curves of CH and CF align closely, with CH demonstrating a rightward shift. This means CH guidance attains a certain IS with lower guidance scale, suggesting enhanced control strength.

The difference in CH guidance's functionality between CIFAR-10 and ImageNet256 may relate to the choice of projection operator $\mathbf{P}$. In latent space, we apply channel-wise projection to the residual vector, avoiding channel-wise mean projection inappropriate for RGB color correction in latent space. Despite the unclear interpretation of this approach in latent space, our visual inspection suggest CH guidance achieves semantic characteristics enhancing and exposure correction, as shown in Fig.5 (Right).

6.5. Stable Diffusion: Text/Image to Image Generation

Characteristic guidance, designed for broad compatibility, can be seamlessly integrated with any DDPM that supports classifier-free guidance. We have successfully integrated
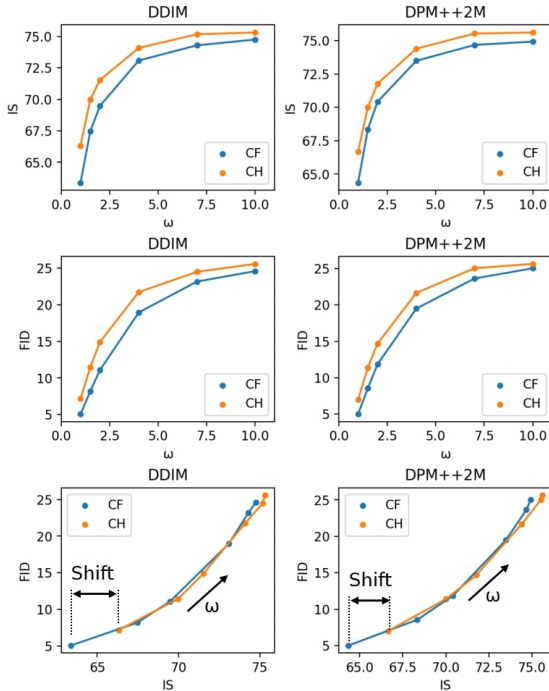
Figure 6. FID and IS comparison between CH and CF guided DDPM on ImageNet 256, using DDIM and DPM++2M sampling methods. CH shows better IS but higher FID than CF, with a rightward shift in its IS-FID curve indicating higher effective guidance scale.

our characteristic guidance into Stable Diffusion WebUI (AUTOMATIC1111, 2023) as public available extension, supporting all provided samplers under both Txt2Img and Img2Img mode. Fig.1 showcases a comparative visualization of images from Stable Diffusion XL (Podell et al., 2023). More visulizations could be found in Appendix.

6.6. Visual Inspection

This section evaluates images generated using classifier-free (CF) and characteristic (CH) guidance from the same initial noise, demonstrating that CH guidance is more capable in enhancing semantic characteristics.

For aircraft images created by the Cifar-10 model, as shown in Fig.5 (Left), CF guidance at higher scales (ω) produces backgrounds that are often dull or white. In contrast, CH guidance results in more vivid and realistic backgrounds, such as skies with clouds that open appears in aircrafts photos. Similarly, with volcano images from ImageNet 256's latent diffusion models (Fig.5, Right), CF-guided images at elevated ω levels exhibit color casting and underexposure. CH guidance, conversely, maintains consistent color and exposure, better accentuating key volcanic elements like smoke and lava, thus emphasizing the semantic characteristics more effectively.

In stable diffusion scenarios, Fig.1 compares CF and CH guidance across various prompts and CFG scales. At a CFG scale of 10 ($\omega = 9$), CH guidance mitigates anatomical irregularities, as seen in the handstand girl images (Fig.1 Left). Increasing the scale to 30 ($\omega = 29$), CF guidance struggles with color cast and exposure issues, while CH guidance produces images that more aptly emphasize prompt-relevant features, such as running girls and lush mountains (Fig.1 Right).

Overall, these comparisons underscore CH guidance's capability in not just adjusting basic elements like color and exposure, but more importantly, in amplifying the semantic characteristics of the subjects in line with the given prompts.

7. Discussion

7.1. The Stopping Criteria for Iteration.

Characteristic guidance applies non-linear correction at each time step by iteratively solving (11). The iteration process is governed by two key stopping criteria: a predefined threshold (tolerance η, detailed in Appendix.E) and a maximum iteration limit set at 10. We investigate the impact of the threshold on evaluation metrics (FID and IS) and the total number of iterations, which is crucial for computational efficiency as each iteration involves calling the de-noising neural network.

As illustrated in Fig.7, we observe that convergence typically occurs at thresholds below $1e^{-3}$. This finding informs the optimal setting of η to balance quality and efficiency. Additionally, the graph also highlights the iteration frequency employed by characteristic guidance across two threshold levels. It is noteworthy that non-linear corrections predominantly occur in the middle stages of the diffusion process, a phase critical for context generation as identified in (Zhang et al., 2023).

7.2. The Difference with Dynamical Thresholding

Dynamical thresholding (Saharia et al., 2022) is a technique designed for RGB images. It predicts the final output image $\mathbf{x}_0$ with (1) at each step, correcting and normalizing any out-of-range pixels back to $[-1, 1]$.

Dynamical threshold is effective in correct color cast when pixel values representing RGB colors. But it falls short in general tasks like latent space sampling and correcting contextual inaccuracies. In contrast, characteristic guidance is a theoretical method that is applicable to any continuous data type of any dimension. In practice, as Fig.20 shows, characteristic guidance is capable of correcting contextual inaccuracies while dynamical thresholding focuses more on color.
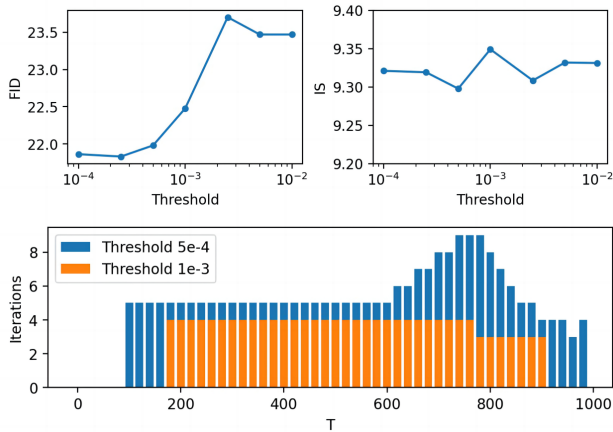
Figure 7. Analysis of convergence and iterations in characteristic guidance on the CIFAR-10 dataset ($\omega = 6$, DPM++2M, 50 steps). The graph depicts the influence of varying threshold levels on FID and IS, and the iteration count across different thresholds relative to the diffusion time step T, ranging from 0 (image) to 1000 (noise).

8. Conclusion

We introduced characteristic guidance, a novel guidance method providing non-linear correction to classifier-free guided DDPMs. It is, as far as we know, the first attempt to lay the theoretical framework for large guidance scale correction. Our comprehensive evaluations of characteristic guidance covered various scenarios, each with a distinct objective: validating its theoretical advantage on Gaussian models, applying it to physics problems like the Landau-Ginzburg model for magnetism, and demonstrating its robustness in image generation across different spaces — CIFAR-10 in pixel-space, ImageNet and Stable Diffusion in latent space. The results consistently highlight the method's effectiveness in enhancing the semantic characteristics of subjects and mitigating generation irregularities.

9. Limitation

FID's Effectiveness. The effectiveness of FID for evaluating guided DDPM is compromised by the difference between the target conditional probabilities $p(\mathbf{x}|\mathbf{c}, \omega)$ and marginal probabilities $p(\mathbf{x})$. FID's focus on the distance from $p(\mathbf{x})$ contrasts with guided DDPM's goal of sampling from $p(\mathbf{x}|\mathbf{c}, \omega)$, suggesting the need for more apt metrics that address this discrepancy.

Speed. The process of resolving the non-linear correction (11) requires iterative computations involving neural networks. This approach, without effective regularization, tends to be slow. Exploring advanced regularization techniques beyond the projection $\mathbf{P}$ could enhance the convergence rate hence accelerate the computation.

Software and Data

Characteristic guidance is available at `https://scraed.github.io/CharacteristicGuidance/` as open source and public extension for Stable Diffusion WebUI (AUTOMATIC1111, 2023), supporting all provided samplers under both Txt2Img and Img2Img mode.

Acknowledgements

We would like to express our gratitude to Professor Yang Wang at the Hong Kong University of Science and Technology for his support of this study.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Anderson, B. D. O. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12:313–326, 1982.

AUTOMATIC1111. Stable diffusion webui, 2023. URL `https://github.com/AUTOMATIC1111/stable-diffusion-webui`.

Bansal, A., Chu, H.-M., Schwarzschild, A., Sengupta, S., Goldblum, M., Geiping, J., and Goldstein, T. Universal guidance for diffusion models. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 843–852, 2023.

Betker, J., Goh, G., Jing, L., Brooks, T., Wang, J., Li, L., Ouyang, L., Zhuang, J., Lee, J., Guo, Y., et al. Improving image generation with better captions, 2023.

CagliostroLab. Animagine xl 3.0, 2023. URL `https://huggingface.co/cagliostrolab/animagine-xl-3.0`.

Coderpiaobozhe. Pytorch implementation of classifier-free diffusion guidance, 2023. URL `https://github.com/coderpiaobozhe/classifier-free-diffusion-guidance-Pytorch`.

Dhariwal, P. and Nichol, A. Diffusion models beat gans on image synthesis. *ArXiv*, abs/2105.05233, 2021.

Evans, L. C. *Partial differential equations*, volume 19. American Mathematical Society, 2022.

Ho, J. Classifier-free diffusion guidance. *ArXiv*, abs/2207.12598, 2022.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. *ArXiv*, abs/2006.11239, 2020.

Hong, S., Lee, G., Jang, W., and Kim, S. W. Improving sample quality of diffusion models using self-attention guidance. *ArXiv*, abs/2210.00939, 2022.

Kardar, M. *Statistical fields*, pp. 19–34. Cambridge University Press, 2007. doi: 10.1017/CBO9780511815881.003.

Karras, T., Aittala, M., Aila, T., and Laine, S. Elucidating the design space of diffusion-based generative models. *ArXiv*, abs/2206.00364, 2022.

Kim, D., Kim, Y., Kang, W., and Moon, I.-C. Refining generative process with discriminator guidance in score-based diffusion models. In *International Conference on Machine Learning*, 2022.

Lai, C.-H., Takida, Y., Murata, N., Uesaka, T., Mitsufuji, Y., and Ermon, S. Fp-diffusion: Improving score-based diffusion models by enforcing the underlying score fokker-planck equation. In *International Conference on Machine Learning*, 2022.

Langley, P. Crafting papers on machine learning. In Langley, P. (ed.), *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.

Liu, L., Ren, Y., Lin, Z., and Zhao, Z. Pseudo numerical methods for diffusion models on manifolds. *ArXiv*, abs/2202.09778, 2022.

Lu, C., Zhou, Y., Bao, F., Chen, J., Li, C., and Zhu, J. Dpm-solver++: Fast solver for guided sampling of diffusion probabilistic models. *ArXiv*, abs/2211.01095, 2022.

Luo, C. Understanding diffusion models: A unified perspective. *ArXiv*, abs/2208.11970, 2022.

Mcmonkeyprojects. Sd-dynamic-thresholding: A dynamic thresholding method for stable diffusion, 2024. URL https://github.com/mcmonkeyprojects/sd-dynamic-thresholding.

Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., Penna, J., and Rombach, R. Sdxl: Improving latent diffusion models for high-resolution image synthesis, 2023.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. High-resolution image synthesis with latent diffusion models. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10674–10685, 2021.

Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., Ghasemipour, S. K. S., Ayan, B. K., Mahdavi, S. S., Lopes, R. G., Salimans, T., Ho, J., Fleet, D. J., and Norouzi, M. Photorealistic text-to-image diffusion models with deep language understanding, 2022.

Sohl-Dickstein, J. N., Weiss, E. A., Maheswaranathan, N., and Ganguli, S. Deep unsupervised learning using nonequilibrium thermodynamics. *ArXiv*, abs/1503.03585, 2015.

Song, J., Meng, C., and Ermon, S. Denoising diffusion implicit models. *ArXiv*, abs/2010.02502, 2020a.

Song, Y. and Ermon, S. Generative modeling by estimating gradients of the data distribution. In *Neural Information Processing Systems*, 2019.

Song, Y., Sohl-Dickstein, J. N., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-based generative modeling through stochastic differential equations. *ArXiv*, abs/2011.13456, 2020b.

Uhlenbeck, G. E. and Ornstein, L. S. On the theory of the brownian motion. *Physical Review*, 36:823–841, 1930.

Vincent, P. A connection between score matching and denoising autoencoders. *Neural Computation*, 23:1661–1674, 2011.

Walker, H. F. and Ni, P. Anderson acceleration for fixed-point iterations. *SIAM J. Numer. Anal.*, 49:1715–1735, 2011.

Yang, L., Zhang, Z., Hong, S., Xu, R., Zhao, Y., Shao, Y., Zhang, W., Yang, M.-H., and Cui, B. Diffusion models: A comprehensive survey of methods and applications. *ACM Computing Surveys*, 2022.

Yu, J., Wang, Y., Zhao, C., Ghanem, B., and Zhang, J. Freedom: Training-free energy-guided conditional diffusion model. *ArXiv*, abs/2303.09833, 2023.

Zhang, Z., Zhao, Z., Yu, J., and Tian, Q. Shiftddpms: Exploring conditional diffusion models by shifting diffusion trajectories. *ArXiv*, abs/2302.02373, 2023.

Zhao, W., Bai, L., Rao, Y., Zhou, J., and Lu, J. Unipc: A unified predictor-corrector framework for fast sampling of diffusion models. *ArXiv*, abs/2302.04867, 2023.

A. The Denoising Diffusion Probabilistic Model (DDPMs)

DDPMs (Ho et al., 2020) are models that generate high-quality images from noise via a sequence of denoising steps. Denoting images as random variable $\mathbf{x}$ of the probabilistic density distribution $p(\mathbf{x})$, the DDPM aims to learn a model distribution that mimics the image distribution $p(\mathbf{x})$ and draw samples from it. The training and sampling of the DDPM utilize two diffusion process: the forward and the backward diffusion process.

The forward diffusion process of the DDPM provides necessary information to train a DDPM. It gradually adds noise to existing images $\mathbf{x}_0 \sim p(x)$ using the Ornstein–Uhlenbeck diffusion process (OU process) (Uhlenbeck & Ornstein, 1930) within a finite time interval $t \in [0, T]$. The OU process is defined by the stochastic differential equation (SDE):

$$d\mathbf{x}_t = -\frac{1}{2}\mathbf{x}_t dt + d\mathbf{W}_t, \tag{24}$$

in which t is the forward time of the diffusion process, $\mathbf{x}_t$ is the noise contaminated image at time t, and $\mathbf{W}_t$ is a standard Brownian motion. The standard Brownian motion formally satisfies $d\mathbf{W}_t = \sqrt{dt}\epsilon$ with ϵ be a standard Gaussian noise. In practice, the OU process is numerically discretized into the variance-preserving (VP) form (Song et al., 2020b):

$$\mathbf{x}_i = \sqrt{1 - \beta_{i-1}}\mathbf{x}_{i-1} + \sqrt{\beta_{i-1}}\epsilon_{i-1}, \tag{25}$$

where $i = 1, \cdots, n$ is the number of the time step, β_i is the step size of each time step, $\mathbf{x}_i$ is image at ith time step with time $t_i = \sum_{j=0}^{i-1}\beta_j$, ϵ_i is standard Gaussian random variable. The time step size usually takes the form $\beta_i = \frac{i(b_2 - b_1)}{n-1} + b_1$ where $b_1 = 10^{-4}$ and $b_2 = 0.02$. Note that our interpretation of β differs from that in (Song et al., 2020b), treating β as a varying time-step size to solve the autonomous SDE (24) instead of a time-dependent SDE. Our interpretation holds as long as every β_i^2 is negligible and greatly simplifies future analysis. The discretized OU process (25) adds a small amount of Gaussian noise to the image at each time step i, gradually contaminating the image until $\mathbf{x}_n \sim \mathcal{N}(\mathbf{0}, I)$.

Training a DDPM aims to recover the original image x_0 from one of its contaminated versions x_i. In this case (25) could be rewritten into the form

$$\mathbf{x}_i = \sqrt{\bar{\alpha}_i}\mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_i}\bar{\epsilon}_i; \quad 1 \le i \le n, \tag{26}$$

where $\bar{\alpha}_i = \prod_{j=0}^{i-1}(1 - \beta_j)$ is the weight of contamination and $\bar{\epsilon}_i$ is a standard Gaussian random noise to be removed. This equation tells us that the distribution of $\mathbf{x}_i$ given $\mathbf{x}_0$ is a Gaussian distribution

$$p(\mathbf{x}_i|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}|\sqrt{\bar{\alpha}_i}\mathbf{x}_0, (1 - \bar{\alpha}_i)I), \tag{27}$$

and the noise $\bar{\epsilon}_i$ is related to the its score function

$$\bar{\epsilon}_i = -\sqrt{1 - \bar{\alpha}_i}\mathbf{s}(\mathbf{x}_i|\mathbf{x}_0, t_i), \tag{28}$$

where $\mathbf{s}(\mathbf{x}_i|\mathbf{x}_0, t_i) = \nabla_{\mathbf{x}_i}\log p(\mathbf{x}_i|\mathbf{x}_0)$ is the score of the density $p(\mathbf{x}_i|\mathbf{x}_0)$ at $\mathbf{x}_i$.

DDPM aims to removes the noise $\bar{\epsilon}_i$ from $\mathbf{x}_i$ by training a denoising neural network $\epsilon_\theta(\mathbf{x}, t_i)$ to predict and remove the noise $\bar{\epsilon}_i$. This means that DDPM minimizes the denoising objective (Ho et al., 2020):

$$L_{denoise}(\epsilon_\theta) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{E}_{\mathbf{x}_0 \sim p(\mathbf{x})}\mathbf{E}_{\bar{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, I)}\|\bar{\epsilon}_i - \epsilon_\theta(\mathbf{x}_i, t_i)\|_2^2. \tag{29}$$

This is equivalent to, with the help of (28) and tricks in (Vincent, 2011), a denoising score matching objective

$$L_{denoise}(\epsilon_\theta) = \frac{1}{n}\sum_{i=1}^{n}(1 - \bar{\alpha}_i)\mathbf{E}_{\mathbf{x}_i \sim p(\mathbf{x}_i)}\|\mathbf{s}(\mathbf{x}_i, t_i) + \frac{\epsilon_\theta(\mathbf{x}_i, t_i)}{\sqrt{1 - \bar{\alpha}_i}}\|_2^2. \tag{30}$$

where $\mathbf{s}(\mathbf{x}, t_i) = \nabla_{\mathbf{x}_i}\log p(\mathbf{x}_i)|_{\mathbf{x}_i = \mathbf{x}}$ is the score function of the density $p(\mathbf{x}_i)$. This objectives says that the denoising neural network $\epsilon_\theta(\mathbf{x}, t_i)$ is trained to approximate a scaled score function $\epsilon(\mathbf{x}, t_i)$ (Yang et al., 2022)

$$\epsilon(\mathbf{x}, t_i) = \underset{\epsilon_\theta}{\operatorname{argmin}}\, L(\epsilon_\theta) = -\sqrt{1 - \bar{\alpha}_i}\mathbf{s}(\mathbf{x}, t_i). \tag{31}$$

Another useful property we shall exploit later is that for **infinitesimal** time steps, the contamination weight $\bar{\alpha}_i$ is the exponential of the diffusion time t_i

$$\lim_{\max_j \beta_j \to 0} \bar{\alpha}_i \to e^{-t_i}. \tag{32}$$

In this case, the discretized OU process (25) is equivalent to the OU process (24), hence the scaled score function $\epsilon(\mathbf{x}, t_i)$ is

$$\epsilon(\mathbf{x}, t) = -\sqrt{1 - e^{-t}}\mathbf{s}(\mathbf{x}, t), \tag{33}$$

where $\mathbf{s}(\mathbf{x}, t)$ is a solution of the score Fokker-Planck equation (4) of the OU process.

The backward diffusion process is used to sample from the DDPM by removing the noise of an image step by step. It is the time reversed version of the OU process, starting at $x_{0'} \sim \mathcal{N}(\mathbf{x}|\mathbf{0}, I)$, using the reverse of the OU process (Anderson, 1982):

$$d\mathbf{x}_{t'} = \left(\frac{1}{2}\mathbf{x}_{t'} + \mathbf{s}(\mathbf{x}, T - t')\right)dt' + d\mathbf{W}_{t'}, \tag{34}$$

in which $t' \in [0, T]$ is the backward time, $\mathbf{s}(\mathbf{x}, T - t') = \nabla_{\mathbf{x}_{T-t'}} \log p(\mathbf{x}_{T-t'})|_{\mathbf{x}}$ is the score function of the density of $\mathbf{x}_{t=T-t'}$ in the forward process. In practice, the backward diffusion process is discretized into

$$\mathbf{x}_{i'+1} = \frac{\mathbf{x}_{i'} + \mathbf{s}(\mathbf{x}_{i'}, T - t'_{i'})\beta_{n-i'}}{\sqrt{1 - \beta_{n-i'}}} + \sqrt{\beta_{n-i'}}\epsilon_{i'}, \tag{35}$$

where $i' = 0, \cdots, n$ is the number of the backward time step, $\mathbf{x}_{i'}$ is image at i'th backward time step with time $t'_{i'} = \sum_{j=0}^{i'-1} \beta_{n-1-j} = T - t_{n-i'}$, and $\mathbf{s}(\mathbf{x}_{i'}, T - t'_{i'})$ is estimated by (31). This discretization is consistent with (34) as long as β_i^2 are negligible.

The forward and backward process forms a dual pair when total diffusion time t_n is large enough and the time step sizes are small enough, at which $p(\mathbf{x}_n) = \mathcal{N}(\mathbf{0}, I)$. In this case the density of $\mathbf{x}_t$ in the forward process and the density of $\mathbf{x}_{t'}$ in the backward process satisfying the relation

$$p(\mathbf{x}_{t'})|_{t'=T-t} = p(\mathbf{x}_t). \tag{36}$$

This relation tells us that the discrete backward diffusion process generate image samples approximately from the image distribution

$$p(\mathbf{x}_{n'}) \approx p(\mathbf{x}_0) = p(\mathbf{x}), \tag{37}$$

despite the discretization error and estimation error of the score function $\mathbf{s}(\mathbf{x}, t')$. An accurate estimation of the score function is one of the key to sample high quality images.

B. Conditional DDPM and Guidance

Conditional DDPMs, which generate images based on a given condition $\mathbf{c}$, model the conditional image distribution $p(\mathbf{x}|\mathbf{c})$. One can introduce the dependency on conditions by extending the denoising neural network to include the condition $\mathbf{c}$, represented as $\epsilon(\mathbf{x}|\mathbf{c}, t_i)$.

Guidance is a technique for conditional image generation that trades off control strength and image diversity. It aims to sample from the distribution (Song et al., 2020b; Dhariwal & Nichol, 2021)

$$p(\mathbf{x}|\mathbf{c}, \omega) \propto p(\mathbf{c}|\mathbf{x})^{1+\omega}p(\mathbf{x}), \tag{38}$$

where $\omega > 0$ is the guidance scale. When ω is large, guidance control the DDPM to produce samples that have the highest classifier likelihood $p(\mathbf{c}|\mathbf{x})$.

Another equivalent guidance without the need of the classifier $p(\mathbf{c}|\mathbf{x})$ is the Classifier-free guidance (Ho, 2022):

$$p(\mathbf{x}|\mathbf{c}, \omega) \propto p(\mathbf{x}|\mathbf{c})^{1+\omega}p(\mathbf{x})^{-\omega}, \tag{39}$$

Sampling from $p(\mathbf{x}|\mathbf{c}, \omega)$ using DDPM requires the corresponding denoising neural network $\epsilon(\mathbf{x}|\mathbf{c}, t_i, \omega)$ that is unknown. Classifier-free guidance provides an approximation of it by linearly combine $\epsilon(\mathbf{x}|\mathbf{c}, t_i)$ and $\epsilon(\mathbf{x}, t_i)$ of conditional and

unconditional DDPMs. The classifier-free guidance is inspired by the fact that $\boldsymbol{\epsilon}$ is proportional to the score $\mathbf{s}$ in (31). At $t = t_0 = 0$, the score of the distribution $p(\mathbf{x}|\mathbf{c}, \omega)$ is the linear combination of scores of $p(\mathbf{x}|\mathbf{c})$ and $p(\mathbf{x})$:

$$\mathbf{s}(\mathbf{x}|\mathbf{c}, t_0, \omega) = (1 + \omega)\,\mathbf{s}(\mathbf{x}|\mathbf{c}, t_0) - \omega\,\mathbf{s}(\mathbf{x}, t_0), \tag{40}$$

where $\mathbf{s}(\mathbf{x}|\mathbf{c}, t_0, \omega) = \nabla_{\mathbf{x}} \log p(\mathbf{x}|\mathbf{c}, \omega)$, $\mathbf{s}(\mathbf{x}|\mathbf{c}, t_0) = \nabla_{\mathbf{x}} \log p(\mathbf{x}|\mathbf{c})$, and $\mathbf{s}(\mathbf{x}, t_0) = \nabla_{\mathbf{x}} \log p(\mathbf{x})$. Since scores $\mathbf{s}$ are proportional to de-noising neural netowrks $\boldsymbol{\epsilon}$ according to equation (31), equation (40) inspires the classifier free guidance to use the following guided denoising neural network $\boldsymbol{\epsilon}_{CF}(\mathbf{x}|\mathbf{c}, t_i, \omega)$:

$$\boldsymbol{\epsilon}_{CF}(\mathbf{x}|\mathbf{c}, t_i, \omega) = (1 + \omega)\,\boldsymbol{\epsilon}_\theta(\mathbf{x}|\mathbf{c}, t_i) - \omega\,\boldsymbol{\epsilon}_\theta(\mathbf{x}, t_i), \tag{41}$$

where i is the number of time step, $t_i = \sum_{j=0}^{i} \beta_j$, $\mathbf{c}$ is the condition, and $\omega > 0$ is the guidance scale. The classifier free guidance exactly computes the de-noising neural network $\boldsymbol{\epsilon}(\mathbf{x}|\mathbf{c}, t_i, \omega)$ of $p(\mathbf{x}|\mathbf{c}, \omega)$ at time $t_0 = 0$ because of the connection between score function and $\boldsymbol{\epsilon}$ (31). However, $\boldsymbol{\epsilon}_{CF}$ is not a good approximation of the theoretical $\boldsymbol{\epsilon}(\mathbf{x}|\mathbf{c}, t_i, \omega)$ for most of time steps $0 < i < n$ when the guidance scale ω is large.

C. Fokker-Plank Equation

The probability density distribution $p(\mathbf{x}_t)$ of the diffusion process (24) is a function of $\mathbf{x}$ and t, governed by the Fokker-Planck equation of the OU process:

$$\frac{\partial p}{\partial t} = \frac{1}{2}\nabla_{\mathbf{x}} \cdot (\mathbf{x}p) + \frac{1}{2}\nabla_{\mathbf{x}}^2 p. \tag{42}$$

The corresponding score function $\mathbf{s} = \nabla_{\mathbf{x}} \log p$ is a vector valued function of $\mathbf{x}$ and t governed by the score Fokker-Planck equation:

$$\frac{\partial \mathbf{s}}{\partial t} = \frac{1}{2}\left(\nabla_{\mathbf{x}}(\mathbf{s} \cdot \mathbf{x}) + \nabla_{\mathbf{x}}^2 \mathbf{s} + \nabla_{\mathbf{x}}\|\mathbf{s}\|_2^2\right) \tag{43}$$

that has been recently studied by (Lai et al., 2022) (our score Fokker-Planck equation differs from theirs slightly by noting that $\nabla_{\mathbf{x}}\nabla_{\mathbf{x}} \cdot \mathbf{s} = \nabla_{\mathbf{x}}^2 \mathbf{s}$ where $\mathbf{s}$ is a gradient). This equation holds for both unconditional, conditional, and guided DDPM because they share the same forward diffusion process. Their corresponding initial conditions at time $t = t_0 = 0$ are $\mathbf{s}(\mathbf{x}, t_0)$, $\mathbf{s}(\mathbf{x}|\mathbf{c}, t_0)$ and $\mathbf{s}(\mathbf{x}|\mathbf{c}, t_0, \omega)$.

However, the score Fokker-Planck equation is a non-linear partial differential equation, which means a linear combination of scores of the unconditional and conditional DDPM is not equivalent to the guided score:

$$\mathbf{s}(\mathbf{x}|\mathbf{c}, t, \omega) \neq (1 + \omega)\,\mathbf{s}(\mathbf{x}|\mathbf{c}, t) - \omega\,\mathbf{s}(\mathbf{x}, t); \quad t_0 < t < t_n, \tag{44}$$

even though their initial conditions at $t = t_0 = 0$ satisfies the linear relation (40). Consequently, the classifier-free guidance $\boldsymbol{\epsilon}_{CF}$ is not a good approximation of $\boldsymbol{\epsilon}$ for $p(\mathbf{x}|\mathbf{c}, \omega)$ at $t_0 < t < t_n$. Furthermore, $\boldsymbol{\epsilon}_{CF}$ does not correspond to a DDPM of any distribution $p(\mathbf{x})$ when $\omega > 0$, because all DDPMs theoretically have corresponding scores satisfying the score Fokker-Planck (4) while $\boldsymbol{\epsilon}_{CF}$ does not.

Note that when the diffusion time steps are infinitesimally small, equation (43) could be rewritten in terms of $\boldsymbol{\epsilon}$ with the help of the relation (33).

$$\frac{\partial \boldsymbol{\epsilon}}{\partial t} = \frac{1}{2}\left(\nabla_{\mathbf{x}}(\boldsymbol{\epsilon} \cdot \mathbf{x}) + \nabla_{\mathbf{x}}^2 \boldsymbol{\epsilon} - \frac{\nabla_{\mathbf{x}}\|\boldsymbol{\epsilon}\|_2^2}{\sqrt{1 - e^{-t}}} + \frac{e^{-t}}{(1 - e^{-t})}\boldsymbol{\epsilon}\right) \tag{45}$$

Our work aims to address this issue by providing non-linear corrections to the classifier-free guidance, making it approximately satisfy the score Fokker-Planck equation. We propose the Harmonic ansatz which says the Laplacian term $\nabla_{\mathbf{x}}^2 \mathbf{s}$ in the score Fokker-Planck equation is negligible. It allow us to use the method of characteristics to handle the non-linear term. Noting from (31) that $\boldsymbol{\epsilon} \propto -\mathbf{s}$, the Harmonic ansatz stands as a good approximation as long as $\|\nabla_{\mathbf{x}}^2 \boldsymbol{\epsilon}\| \ll \|-\nabla_{\mathbf{x}}(\boldsymbol{\epsilon} \cdot \mathbf{x}) + \nabla_{\mathbf{x}}\|\boldsymbol{\epsilon}\|_2^2\|$ along possible diffusion trajectories of DDPM.

D. Deriving the Characteristic Guidance Using the Method of Characteristics

The exactness of the classifier-free guidance revealed by (40) as initial condition at $t_0 = 0$ and its failure at $t_i > t_0$ due to non-linearity (44) are key observations that inspires the characteristic guidance. Characteristic guidance answers the

following problem: Given two known solutions $s_1(\mathbf{x}, t)$ and $s_2(\mathbf{x}, t)$ of the score Fokker-Planck equation (4), compute another solution $s(\mathbf{x}, t)$ of the score Fokker-Planck equation with the following initial condition:

$$s(\mathbf{x}, 0) = (1 + \omega)s_1(\mathbf{x}, 0) - \omega s_2(\mathbf{x}, 0), \tag{46}$$

Note that this initial condition is the same as the condition $\epsilon(\mathbf{x}, 0) = (1 + \omega)\epsilon_1(\mathbf{x}, 0) - \omega\epsilon_2(\mathbf{x}, 0)$ used in lemma 5.2 because score s is proportional to ϵ (31). Similarly, the problem is equivalent to express $\epsilon(\mathbf{x}, t)$ in terms of $\epsilon_1(\mathbf{x}, t)$ and $\epsilon_2(\mathbf{x}, t)$ as stated in lemma 5.2.

It is not easy to express $s(\mathbf{x}, t)$ as a function of $s_1(\mathbf{x}, t)$ and $s_2(\mathbf{x}, t)$ without training or taking complex derivatives in Eq. (4). Fortunately, the harmonic ansatz 5.1 leads us to a viable solution. **The harmonic ansatz** reduces the score Fokker-Planck equation (4) to the following non-linear first-order partial differential equation (PDE)

$$\frac{\partial \mathbf{s}}{\partial t} = \frac{1}{2} \left(\nabla_{\mathbf{x}}(\mathbf{s} \cdot \mathbf{x}) + \nabla_{\mathbf{x}}\|\mathbf{s}\|_2^2 \right), \tag{47}$$

where the Laplacian term is omitted but the non-linear term remains, capturing the non-linear effect in DDPMs. We treat $\mathbf{s}(\mathbf{x}, t)$, $\mathbf{s}_1(\mathbf{x}, t)$, and $\mathbf{s}_2(\mathbf{x}, t)$ as solutions to this PDE and deduce $\mathbf{s}(\mathbf{x}, t)$ using the method of characteristics.

Characteristic Lines The method of characteristics reduce a partial differential equation to a family of ordinary differential equations (ODE) by considering solutions along characteristic lines. In our case, we consider the characteristic lines $\mathbf{x}(t)$ satisfying the ODE:

$$\frac{d\mathbf{x}(t)}{dt} = - \left(\frac{1}{2}\mathbf{x} + \mathbf{s}(\mathbf{x}(t), t) \right), \tag{48}$$

substitute such characteristic lines into the PDE (47) yields the dynamics of the score functions along the lines

$$\frac{d\mathbf{s}(\mathbf{x}(t), t)}{dt} = \frac{1}{2}\mathbf{s}(\mathbf{x}(t), t), \tag{49}$$

The characteristic lines of the simplified score Fokker-Planck equation (47), which are solutions of the ODEs (48) and (49), has the form

$$\mathbf{x}(t) = e^{-\frac{t}{2}}\mathbf{x}_0 - (1 - e^{-t})\mathbf{s}(\mathbf{x}(t), t)$$
$$\mathbf{s}(\mathbf{x}(t), t) = \mathbf{s}(\mathbf{x}_0, 0)e^{\frac{t}{2}}, \tag{50}$$

where $\mathbf{x}_0 = \mathbf{x}(0)$ is the starting point of the line and serves as a constant parameter, $\mathbf{s}(\mathbf{x}_0, 0)$ depends on the initial condition of $\mathbf{s}$ at $t = 0$. Besides, such characteristic lines of $\mathbf{s}$ also induces characteristic lines of ϵ as they are connected by (33).

An important remark on the characteristic lines is that (50) is valid for $\mathbf{s}_1$ and $\mathbf{s}_2$ as well, if we replace every $\mathbf{s}$ with $\mathbf{s}_1$ or $\mathbf{s}_2$. This is because $\mathbf{s}$, $\mathbf{s}_1$, and $\mathbf{s}_2$ are solutions to the same PDE (47). The merit of these characteristic lines lies in their ability to determine the value of $\mathbf{s}(\mathbf{x}(t), t)$ at $t > 0$ using information of $\mathbf{x}_0$ and $\mathbf{s}(\mathbf{x}_0, 0)$ at $t = 0$, thereby allowing us to utilize the initial condition (46).

Deducing $s(\mathbf{x}, t)$ Our next move is deduce the value $\mathbf{s}(\mathbf{x}, t)$ from $\mathbf{s}_1(\mathbf{x}, t)$ and $\mathbf{s}_2(\mathbf{x}, t)$ with the help of characteristic lines. Denote the characteristic lines of the three score functions $\mathbf{s}$, $\mathbf{s}_1$, and $\mathbf{s}_2$ to be $\mathbf{x}(t)$, $\mathbf{x}_1(t)$, $\mathbf{x}_2(t)$. If the three characteristic lines meets at $\mathbf{x}_0$ at $t = 0$

$$\mathbf{x}(0) = \mathbf{x}_1(0) = \mathbf{x}_2(0) = \mathbf{x}_0, \tag{51}$$

then the initial condition (46) and the equation (50) tells that

$$\mathbf{s}(\mathbf{x}(t), t) = (1 + \omega)\mathbf{s}_1(\mathbf{x}_1(t), t) - \omega\mathbf{s}_2(\mathbf{x}_2(t), t), \tag{52}$$

which achieves our goal in expressing the value $\mathbf{s}$ in terms of $\mathbf{s}_1$ and $\mathbf{s}_2$. In summary, the following system of equations determines the value of $\mathbf{s}$ in terms of $\mathbf{s}_1$ and $\mathbf{s}_2$

$$\begin{cases} \mathbf{x}(t)e^{\frac{t}{2}} + (e^{\frac{t}{2}} - e^{-\frac{t}{2}})\mathbf{s}(\mathbf{x}(t), t) &= \mathbf{x}_0 \\ \mathbf{x}_1(t)e^{\frac{t}{2}} + (e^{\frac{t}{2}} - e^{-\frac{t}{2}})\mathbf{s}_1(\mathbf{x}_1(t), t) &= \mathbf{x}_0 \\ \mathbf{x}_2(t)e^{\frac{t}{2}} + (e^{\frac{t}{2}} - e^{-\frac{t}{2}})\mathbf{s}_2(\mathbf{x}_2(t), t) &= \mathbf{x}_0 \\ (1 + \omega)\mathbf{s}_1(\mathbf{x}_1(t), t) - \omega\mathbf{s}_2(\mathbf{x}_2(t), t) &= \mathbf{s}(\mathbf{x}(t), t), \end{cases} \tag{53}$$

14

in which the first three equations are reformulations of (50) and say that the three characteristic lines meets at the same $\mathbf{x}_0$ at $t = 0$. In practice, we wish to compute the value of $\mathbf{s}$ for a given $\mathbf{x}(t)$ at time t, with the functions $\mathbf{s}_1$ and $\mathbf{s}_2$ known in advance. This completes the system with four equations and four unknowns: $\mathbf{x}_0$, $\mathbf{x}_1(t)$, $\mathbf{x}_2(t)$, and $\mathbf{s}(\mathbf{x}(t), t)$.

Since the dependency of $\mathbf{x}$ on t is already characterized by the system of equation, we can simplify the notation by omitting it in the following discussion. Hence, we write $\mathbf{x}$, $\mathbf{x}_1$, $\mathbf{x}_2$ instead of $\mathbf{x}(t)$, $\mathbf{x}_1(t)$, $\mathbf{x}_2(t)$ in the following discussion.

Simplify the System   It is possible to simplify the system of four equations into one. We start from eliminating the unknowns $\mathbf{s}(\mathbf{x}, t)$ and $\mathbf{x}_0$ with some algebra, reducing the system into two equations

$$\begin{cases} \mathbf{x}_1 &= \mathbf{x} + \omega \left(\mathbf{s}_1(\mathbf{x}_1, t) - \mathbf{s}_2(\mathbf{x}_2, t) \right) (1 - e^{-t}) \\ \mathbf{x}_2 &= \mathbf{x} + (1 + \omega) \left(\mathbf{s}_1(\mathbf{x}_1, t) - \mathbf{s}_2(\mathbf{x}_2, t) \right) (1 - e^{-t}). \end{cases} \tag{54}$$

The above equations also indicate that $\mathbf{x}_1$ and $\mathbf{x}_2$ are not independent but linearly correlated as

$$(1 + \omega)\mathbf{x}_1 - \omega \mathbf{x}_2 = \mathbf{x}. \tag{55}$$

This inspires us to define the correction term $\Delta \mathbf{x}$ as

$$\begin{aligned} \mathbf{x}_1 &= \mathbf{x} + \omega \Delta \mathbf{x} \\ \mathbf{x}_2 &= \mathbf{x} + (1 + \omega)\Delta \mathbf{x}, \end{aligned} \tag{56}$$

which is exactly the same notation used in the characteristic guidance (10).

Finally, by combining (54) and (55), we obtain a single equation for $\Delta \mathbf{x}$ as follows:

$$\Delta \mathbf{x} = \left(\mathbf{s}_1(\mathbf{x} + \omega \Delta \mathbf{x}, t) - \mathbf{s}_2(\mathbf{x} + (1 + \omega)\Delta \mathbf{x}, t) \right)(1 - e^{-t}). \tag{57}$$

Once we solve for $\Delta \mathbf{x}$ from the above equation, we can easily compute the desired value of the score function $\mathbf{s}(\mathbf{x}, t)$ by using (56) and (53):

$$\mathbf{s}(\mathbf{x}, t) = (1 + \omega)\mathbf{s}_1(\mathbf{x}_1, t) - \omega \mathbf{s}_2(\mathbf{x}_2, t). \tag{58}$$

Characteristic Guidance, Lemma 5.2, and Theorem 5.3   Now we have nearly completed the derivation of the characteristic guidance. It remains to replace the score functions $\mathbf{s}$ with the de-noising neural networks $\boldsymbol{\epsilon}$ used in DDPM. They are related by (33) if the diffusion time steps in DDPM are small enough. Consequently we rewrite (58) and (57) into

$$\boldsymbol{\epsilon}(\mathbf{x}, t) = (1 + \omega)\boldsymbol{\epsilon}_1(\mathbf{x}_1, t) - \omega \boldsymbol{\epsilon}_2(\mathbf{x}_2, t), \tag{59}$$

and

$$\Delta \mathbf{x} = \left(\boldsymbol{\epsilon}_2(\mathbf{x}_2, t) - \boldsymbol{\epsilon}_1(\mathbf{x}_1, t) \right) \sigma(t), \tag{60}$$

where $\sigma(t) = \sqrt{1 - e^{-t}}$. These equations are exactly the content of the Lemma 5.2.

In practice, DDPMs do not use infinitesimal time steps. Therefore we replace e^{-t_i} with $\bar{\alpha}_i$ everywhere since they are related through (32). This leads to the characteristic guidance (10) and (11).

As for Theorem 5.3, it is a direct consequence of Lemma 5.2. Lemma 5.2 says that $\mathbf{s}(\mathbf{x}, t)$ is a solution of the equation (47), then its corresponding $\boldsymbol{\epsilon}(\mathbf{x}, t)$, automatically satisfies the FP equation (4) except the Laplacian term, which can be further eliminated by the Harmonic ansatz. Therefore $\boldsymbol{\epsilon}(\mathbf{x}, t)$ in (59) has no mixing error (7) under the Harmonic ansatz.

E. Fixed Point Iteration Methods for Solving the Correction Term

We solve $\Delta \mathbf{x}$ from the equation (11) by fixed point iterative methods. The key idea is treating the residual vector $\mathbf{g}$ of the equation (11)

$$\mathbf{g} = \Delta \mathbf{x} - \mathbf{P} \circ \left(\boldsymbol{\epsilon}_\theta(\mathbf{x}_2, t_i) - \boldsymbol{\epsilon}_\theta(\mathbf{x}_1 | \mathbf{c}, t_i) \right) \sigma_i,$$

as the gradient in an optimization problem. Consequently we could apply any off-the-shelf optimization algorithm to solve the fixed point iteration problem.

Algorithm 1 Successive over-relaxation iteration for $\Delta\mathbf{x}$

Require: $\mathbf{x}, \mathbf{c}, t_i, \omega, \epsilon_\theta, \sigma_i, \eta$(tolerance)
Require: $\mathbf{x}, \mathbf{c}, t_i, \omega, \sigma_i$ {Inputs for the DDPM model}
Require: γ (lr) {Standard parameters for gradient descent}
Require: η {The tolerance as stopping criteria}
Ensure: $\Delta\mathbf{x}$
 1: Initialize $\Delta\mathbf{x}^{(0)}$ as a zero vector
 2: Set $k = 0$
 3: **repeat**
 4:    $k \leftarrow k + 1$
 5:    $\mathbf{x}_1^{(k-1)} = \mathbf{x} + \omega\Delta\mathbf{x}^{(k-1)}$
 6:    $\mathbf{x}_2^{(k-1)} = \mathbf{x} + (1 + \omega)\Delta\mathbf{x}^{(k-1)}$
 7:    $\mathbf{g}^{(k)} = \Delta\mathbf{x}^{(k-1)} - \mathbf{P} \circ \left(\epsilon_\theta(\mathbf{x}_2^{(k-1)}, t_i) - \epsilon_\theta(\mathbf{x}_1^{(k-1)}|\mathbf{c}, t_i)\right)\sigma_i,$
 8:    Update $\Delta\mathbf{x}^{(k)} = \Delta\mathbf{x}^{(k-1)} - \gamma\mathbf{g}^{(k)},$
 9: **until** $\|\mathbf{g}\|_2^2 < \eta^2\mathrm{dim}(\mathbf{g})$
10: Return $\Delta\mathbf{x}^{(k)}$ as $\Delta\mathbf{x}$

Algorithm 2 RMSprop iteration for $\Delta\mathbf{x}$

Require: $\mathbf{x}, \mathbf{c}, t_i, \omega, \sigma_i$ {Inputs for the DDPM model}
Require: α, ϵ, γ (lr) {Standard parameters for the RMSprop algorithm}
Require: η {The tolerance as stopping criteria}
Require: D {Learning rate schedule}
Ensure: $\Delta\mathbf{x}$
 1: Initialize $\Delta\mathbf{x}^{(0)}$ as a zero vector
 2: Set $k = 0$
 3: **repeat**
 4:    $k \leftarrow k + 1$
 5:    $\mathbf{x}_1^{(k-1)} = \mathbf{x} + \omega\Delta\mathbf{x}^{(k-1)}$
 6:    $\mathbf{x}_2^{(k-1)} = \mathbf{x} + (1 + \omega)\Delta\mathbf{x}^{(k-1)}$
 7:    $\mathbf{g}^{(k)} = \Delta\mathbf{x}^{(k-1)} - \mathbf{P} \circ \left(\epsilon_\theta(\mathbf{x}_2^{(k-1)}, t_i) - \epsilon_\theta(\mathbf{x}_1^{(k-1)}|\mathbf{c}, t_i)\right)\sigma_i,$
 8:    $\mathbf{v}^{(k)} = \alpha\mathbf{v}^{(k-1)} + (1 - \alpha)(\mathbf{g}^{(k)})^2$
 9:    $\gamma^{(k)} = \gamma/(1 + Dk)$
10:    Update $\Delta\mathbf{x}^{(k)} = \Delta\mathbf{x}^{(k-1)} - \gamma^{(k)}\mathbf{g}^{(k)}/(\sqrt{\mathbf{v}^{(k)}} + \epsilon),$
11: **until** $\|\mathbf{g}\|_2^2 < \eta^2\mathrm{dim}(\mathbf{g})$
12: Return $\Delta\mathbf{x}^{(k)}$ as $\Delta\mathbf{x}$

Algorithm 3 Anderson acceleration (AA) iteration for $\Delta \mathbf{x}$

Require: $\mathbf{x}, \mathbf{c}, t_i, \omega, \sigma_i$ {Inputs for the DDPM model}
Require: $m \geq 2, \gamma$ (lr), $\Delta \mathbf{x}_B = [\,]$, $\mathbf{g}_B = [\,]$ {Parameters and buffers for AA algorithm}
Require: η {The tolerance as stopping criteria}
Ensure: $\Delta \mathbf{x}$
 1: Initialize $\Delta \mathbf{x}^{(0)}$ as a zero vector
 2: Set $k = 0$
 3: **repeat**
 4:    $k \leftarrow k + 1$
 5:    $\mathbf{x}_1^{(k-1)} = \mathbf{x} + \omega \Delta \mathbf{x}^{(k-1)}$
 6:    $\mathbf{x}_2^{(k-1)} = \mathbf{x} + (1 + \omega) \Delta \mathbf{x}^{(k-1)}$
 7:    $\mathbf{g}^{(k)} = \Delta \mathbf{x}^{(k-1)} - \mathbf{P} \circ \left(\boldsymbol{\epsilon}_\theta(\mathbf{x}_2^{(k-1)}, t_i) - \boldsymbol{\epsilon}_\theta(\mathbf{x}_1^{(k-1)}|\mathbf{c}, t_i) \right) \sigma_i,$
 8:    $\Delta \mathbf{x}_B.\text{append}(\Delta \mathbf{x}^{(k-1)}), \ \mathbf{g}_B.\text{append}(\mathbf{g}^{(k)})$
 9:    **if** $\text{len}(\Delta \mathbf{x}_B) \geq 2$ **then**
10:       $\mathbf{g}_B[-2] = \mathbf{g}_B[-1] - \mathbf{g}_B[-2]$
11:       $\Delta \mathbf{x}_B[-2] = \Delta \mathbf{x}_B[-1] - \Delta \mathbf{x}_B[-2]$
12:       **if** $\text{len}(\Delta \mathbf{x}_B) > m$ **then**
13:          delete $\mathbf{g}_B[0]$; delete $\Delta \mathbf{x}_B[0]$;
14:       **end if**
15:       $A_g = \mathbf{g}_B[: -1]; A_x = \Delta \mathbf{x}_B[: -1];$
16:       $\mathbf{b}_g = \mathbf{g}_B[-1]; \mathbf{b}_x = \Delta \mathbf{x}_B[-1];$
17:       $\mathbf{w} = \text{argmin}_{\mathbf{w}} \| A_g \mathbf{w} - \mathbf{b}_g \|_2^2$
18:       $\Delta \mathbf{x}_{AA}^{(k-1)} = \mathbf{b}_x - A_x \mathbf{w}$
19:       $\mathbf{g}_{AA}^{(k)} = \mathbf{b}_g - A_g \mathbf{w}$
20:    **else**
21:       $\Delta \mathbf{x}_{AA}^{(k-1)} = \Delta \mathbf{x}^{(k-1)}$
22:       $\mathbf{g}_{AA}^{(k)} = \mathbf{g}^{(k)}$
23:    **end if**
24:    Update $\Delta \mathbf{x}^{(k)} = \Delta \mathbf{x}_{AA}^{(k-1)} - \gamma \mathbf{g}_{AA}^{(k)},$
25: **until** $\| \mathbf{g} \|_2^2 < \eta^2 \text{dim}(\mathbf{g})$
26: Return $\Delta \mathbf{x}^{(k)}$ as $\Delta \mathbf{x}$

For example, adopting the gradient descent methods in leads us to the well-known successive over-relaxation iteration method for fixed point problem (Alg.1). Similarly, adopting the RMSprop method leads us to Alg.2 with faster convergence.

Another kind of effective iteration method to solve $\Delta \mathbf{x}$ is the Anderson acceleration method (Walker & Ni, 2011) Alg.3, which turns out to be faster than RMSprop when applied to latent diffusion models.

In practice, the number of iterations required to solve $\Delta \mathbf{x}$ depends on both the iteration method and the DDPM model. Different iteration methods may have different convergence rates and stability properties. Therefore, we suggest trying different iteration methods to find the most efficient and effective one for a given DDPM model.

F. Conditional Gaussian

This experiment compares the classifier free guidance and the characteristic guidance on sampling from a conditional 2D Gaussian distributions, which strictly satisfies the harmonic ansatz.

The diffusion model models two distribution: the conditional distribution $p(x_1, x_2 | c_1, c_2)$ and the unconditional distribution $p(x_1, x_2)$. In our case, both of them are 2D Gaussian distributions

$$
\begin{aligned}
p(x_1, x_2 | c_1, c_2) &= \mathcal{N}\left(x_1, x_2 | (c_1, c_2)^T, I\right) \\
p(x_1, x_2) &= \mathcal{N}\left(x_1, x_2 | (0, 0)^T, 5I\right)
\end{aligned}
\tag{61}
$$

The guided diffusion model (39) aims to sample from the distribution

$$
p(x_1, x_2 | c_1, c_2; \omega) = \mathcal{N}\left(x_1, x_2 | (\frac{5c_1(\omega+1)}{4\omega+5}, \frac{5c_2(\omega+1)}{4\omega+5})^T, \frac{5}{4\omega+5}I\right)
\tag{62}
$$

having the properties $p(x_1, x_2 | c_1, c_2; 0) = p(x_1, x_2 | c_1, c_2)$ and $p(x_1, x_2 | c_1, c_2; -1) = p(x_1, x_2)$.

The score function of the unconditional, conditional, and guided model along the forward diffusion process can be solved analytically from the FP equation:

$$
\mathbf{s}(x_1, x_2 | c_1, c_2; t, \omega) = \left(\frac{5c_1(\omega+1)e^{t/2} - (4\omega+5)e^t x_1}{(4\omega+5)e^t - 4\omega}, \frac{5c_2(\omega+1)e^{t/2} - (4\omega+5)e^t x_2}{(4\omega+5)e^t - 4\omega}\right)^T
$$

with $\omega = -1$ and $\omega = 0$ corresponds to the scores of unconditional and conditional distributions. Therefore the denoising neural networks ϵ_θ of unconditional and conditional DDPMs could be analytically calculated as

$$
\begin{aligned}
\epsilon_\theta(x_1, x_2, t) &= -\sqrt{1 - e^{-t}} \mathbf{s}(x_1, x_2; t, -1) \\
\epsilon_\theta(x_1, x_2 | c_1, c_2; t) &= -\sqrt{1 - e^{-t}} \mathbf{s}(x_1, x_2 | c_1, c_2; t, 0)
\end{aligned}
\tag{63}
$$

The corresponding classifier free guided ϵ_{CF} and characteristic guided ϵ_{CH} are computed according to (41) and (10). The characteristic guidance is computed with Alg.2 with $\gamma = 0.01$, $\eta = 0.01$, and $\alpha = 0.9999$.

We sample from guided DDPMs using four different ways: the SDE (35), probabilistic ODE (Song et al., 2020b), DDIM (Song et al., 2020a), and DPM++2M (Lu et al., 2022). We set $c_1 = -5$, $c_2 = 5$, $\beta_1 = 1e - 4$, $\beta_2 = 0.015$ and the total time step to be $n = 1000$ for SDE and ODE, and $n = 20$ for DDIM and DPM++2M. The sample results are compared with the theoretical reference distribution (62) by assuming the samples are Gaussian distributed then computing the KL divergence with the theoretical reference. The results are plotted in Fig.8. Our characteristic guidance outperforms classifier free guidance in all cases with $\omega > 0$.

G. Mixture of Gaussian

This experiment compares the classifier free guidance and the characteristic guidance when the assumption $\nabla_\mathbf{x} \nabla_\mathbf{x} \cdot \mathbf{s} = 0$ is not satisfied. We train conditional and unconditional DDPM to learn two distributions: the conditional distribution $p(\mathbf{x} | \mathbf{c})$ and the unconditional distribution $p(\mathbf{x}) = p(\mathbf{x})$.

$$
\begin{aligned}
p(\mathbf{x} | \mathbf{c}) &= \mathcal{N}\left(\mathbf{x} | (-1, -1/\sqrt{3})^T, I\right)^{c_0} \mathcal{N}\left(\mathbf{x} | (1, -1/\sqrt{3})^T, I\right)^{c_1} \mathcal{N}\left(\mathbf{x} | (0, \sqrt{3} - 1/\sqrt{3})^T, I\right)^{c_2} \\
p(\mathbf{x}) &= \left(p(\mathbf{x} | (1, 0, 0)^T) + p(\mathbf{x} | (0, 1, 0)^T) + p(\mathbf{x} | (0, 0, 1)^T)\right) / 3
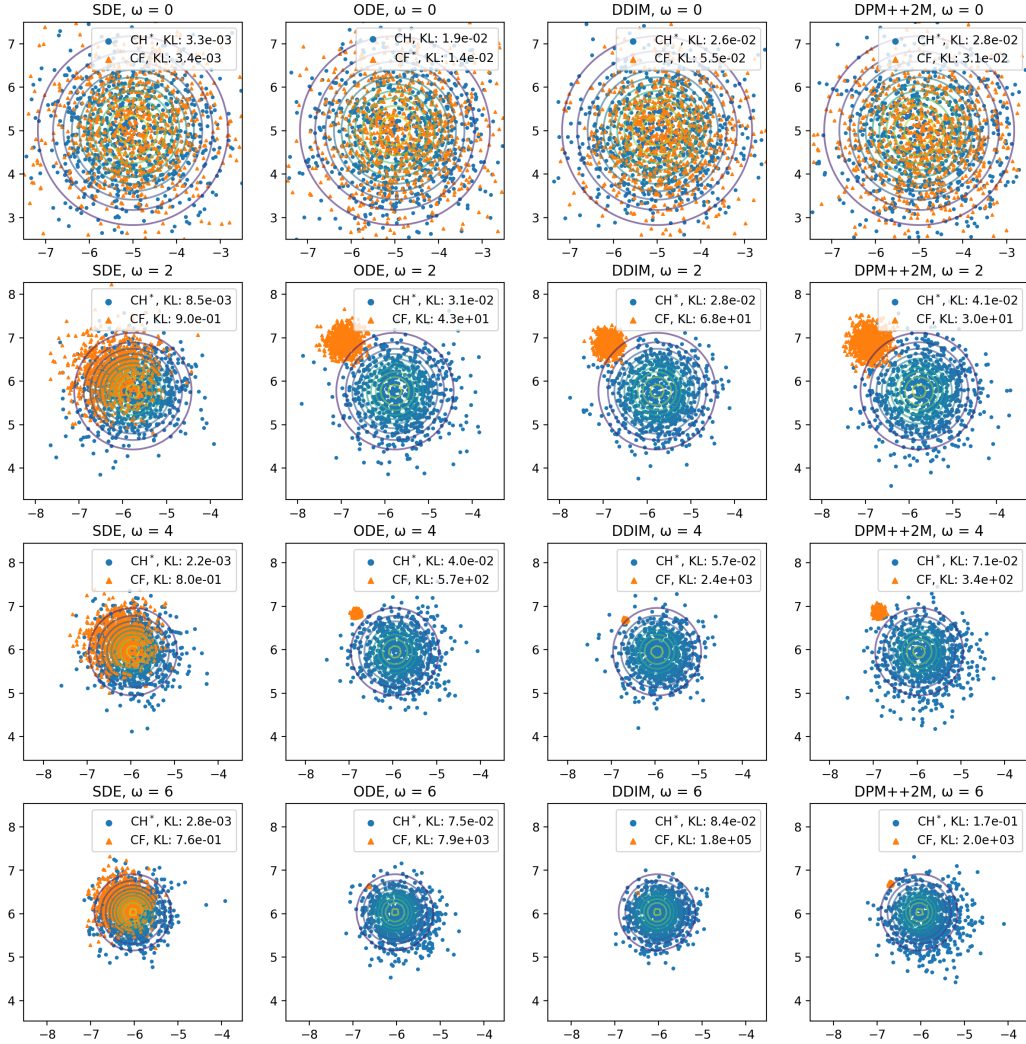\end{aligned}
\tag{64}
$$

Figure 8. Comparison between characteristic guidance (CH) and classifier free guidance (CF) of DDPM on the conditional Gaussian experiment. Samples are drawn with various sample methods: SDE, probabilistic ODE, DDIM and DPM++2M. The contours corresponds to the theoretical reference distribution of guided DDPMs in (62).

where $\mathbf{c}$ is a three dimensional one-hot vector.

The guided diffusion model aims to sample from the distribution

$$p(\mathbf{x}|\mathbf{c};\omega) = \frac{1}{Z(\omega, \mathbf{c})} p(\mathbf{x}|\mathbf{c})^{1+\omega} p(\mathbf{x})^{-\omega} \tag{65}$$

where $Z(\omega, \mathbf{c}) = \int p(\mathbf{x}|\mathbf{c})^{1+\omega} p(\mathbf{x})^{-\omega} d\mathbf{x}$ is the partition function that can be computed with the Monte Carlo method numerically.

We train the denoising neural networks $\boldsymbol{\epsilon}_\theta(\mathbf{x}|\mathbf{c}, t)$ and $\boldsymbol{\epsilon}_\theta(\mathbf{x}, t)$ for conditional and unconditional neural network on a dataset of 20000 pairs of $(\mathbf{x}, \mathbf{c})$. Particularly, The $\boldsymbol{\epsilon}_\theta(\mathbf{x}, t)$ is trained as a special case of conditioned $\boldsymbol{\epsilon}_\theta(\mathbf{x}|\mathbf{c}, t)$ with $\mathbf{c} = \mathbf{0}$. The characteristic guidance is computed by Alg.2 with $\gamma = 0.05$, $\eta = 0.02$, and $\alpha = 0.99$.

We sample from guided DDPMs using four different ways: the SDE (35), probabilistic ODE (Song et al., 2020b), DDIM (Song et al., 2020a), and DPM++2M (Lu et al., 2022). During sampling, we set $\beta_1 = 1e - 4$, $\beta_2 = 0.02$ and the total time step to be $n = 500$ for SDE and ODE, and $n = 20$ for DDIM and DPM++2M. The sample results are compared with the theoretical reference distribution (65) by assuming the samples from DDPM are Gaussian distributed then computing the KL

divergence with the theoretical reference. The results are plotted in Fig.9. Our characteristic guidance outperforms classifier free guidance in all cases with $\omega > 0$.
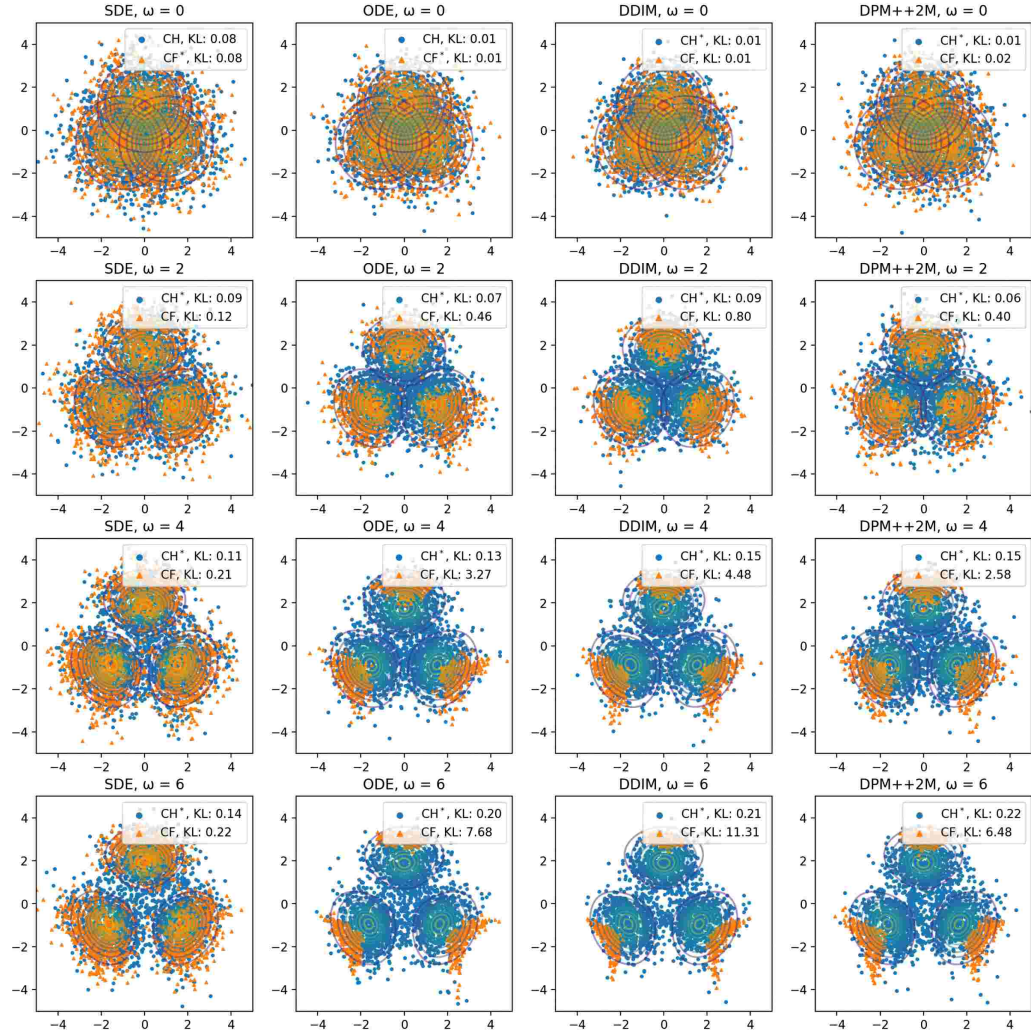

Figure 9. Comparison between characteristic guidance (CH) and classifier free guidance (CF) of DDPM on the mixture of Gaussian experiment. Samples are drawn with various sample methods: SDE, probabilistic ODE, DDIM and DPM++2M. The contours corresponds to the theoretical reference distribution of guided DDPMs in (65).

Fig.9 also shows a significant loss of diversity for classifier free guided ODE samplers when $\omega > 0$. This loss of diversity is corrected by our characteristic guidance, yielding samples that perfectly matches the theoretical reference.

H. Diffusion Model for Science: Cooling of Magnet

Guidance, as described in 39, can be conceptualized as a cooling process with the guidance scale ω representing the temperature drop. We demonstrate that the guidance can simulate the cooling process of a magnet around the Curie temperature, at which a paramagnetic material gains permanent magnetism through phase transition. We also demonstrate that non-linear correction for classifier-free guidance is important for correctly characterize the megnet's phase transition.

The phase transition at the Curie temperature can be qualitatively described by a scalar Landau-Ginzburg model. We consider a thin sheet of 2D magnetic material which can be magnetized only in the direction perpendicular to it. The magnetization of this material at each point is described by a scalar field $\phi(\mathbf{x})$ specifying the magnitude and the direction of magnetization. When ϕ has a higher absolute value, it means that region is more strongly magnetized. Numerically, we discretize $\phi(\mathbf{x})$ into

periodic 8×8 grid, with ith grid node assigned with a float number ϕ_i. It is equivalent to say that the magnetization field $\phi(\mathbf{x})$ is represented by 8×8 picture with 1 channel. At any instance, there is a field ϕ describing the current magnetization of our magnet.

Thermal movements of molecules in our magnet gradually change our magnet's magnetization ϕ from one configuration to another. If we keep track of the field ϕ every second for a long duration, the probability of observing a particular configuration of ϕ is proportional to the Boltzmann distribution

$$p(\phi; T) \propto e^{-\beta H(\phi; T)}, \tag{66}$$

where $\beta H(\phi; T)$ is Hamiltonian at temperature T, assigning an energy to each of possible ϕ. Around the Curie temperature, the Landau-Ginzburg model of our magnet use the following Hamiltonian:

$$\beta H(\phi; T) = K \left(\frac{1}{2} \sum_{\langle i,j \rangle} (\phi_i - \phi_j)^2 + \sum_i \left(\frac{m^2}{2}(T - T_c)\phi_i^2 + \frac{\lambda}{4!}\phi_i^4 \right) \right) \tag{67}$$

where $m^2 = 0.1$, $\lambda = 1.0$, $K = 1$ are parameters and $T_c = 200$ is the Curie temperature. The first term sums over adjacent points i and j on the grid, representing the interaction between neighbors. The second term describes self-interaction of each grid. The Landau-Ginzburg model allow us to control the temperature T of magnet using guidance similar to (40)

$$\boldsymbol{s}(\phi; (1 + \omega)T_1 - \omega T_0) = (1 + \omega)\boldsymbol{s}(\phi; T_1) - \omega \boldsymbol{s}(\phi; T_0)$$

where $\omega > 0$ is the guidance scale, $T_0 > T_1$ are two distinct temperature, and $\boldsymbol{s}(\phi; T) = \nabla \log P(\phi; T)$ is the score of the Boltzmann distribution in (66). This means training DDPMs of the magnetization field ϕ at two distinct temperature $T_0 > T_1$ theoretically allow us to sample ϕ at temperatures below T_1 using guidance, corresponding to the cooling of the magnet.

To simulate the cooling of the magnet, we train a conditional DDPM of ϕ for two distinct temperatures $p(\phi|T_0 = 201)$ and $p(\phi|T_1 = 200)$. The dataset consists of 60000 samples at $T_0 = 201$ and 60000 samples at $T_1 = 200$ generated by the Metropolis-Hastings algorithm. Then we sample from the trained DDPM using classifier-free guidance in (41) and characteristic guidance in (10)

$$\begin{aligned} \epsilon_{CF}(\mathbf{x}|(1 + \omega)T_1 - \omega T_0, t_i) &= (1 + \omega)\,\epsilon_\theta(\mathbf{x}|T_1, t_i) - \omega\,\epsilon_\theta(\mathbf{x}|T_0, t_i) \\ \epsilon_{CH}(\mathbf{x}|(1 + \omega)T_1 - \omega T_0, t_i) &= (1 + \omega)\,\epsilon_\theta(\mathbf{x} + \omega\Delta\mathbf{x}|T_1, t_i) - \omega\,\epsilon_\theta(\mathbf{x} + (1 + \omega)\Delta\mathbf{x}|T_0, t_i). \end{aligned} \tag{68}$$

The characteristic guidance is computed with Alg.2 with $\gamma = 0.01$, $\eta = 0.1$, and $\alpha = 0.999$. Samples from the guided DDPM is treated approximately as samples at temperature $T = (1 + \omega)T_1 - \omega T_0$.

We sample from guided DDPMs using four different ways: the SDE (35), probabilistic ODE (Song et al., 2020b), DDIM (Song et al., 2020a), and DPM++2M (Lu et al., 2022). During sampling, we set $\beta_1 = 1e - 4$, $\beta_2 = 0.015$ and the total time step to be $n = 1000$ for SDE and ODE, and $n = 20$ for DDIM and DPM++2M. For each temperature, we generate theoretical reference samples from the Boltzmann distribution in (66) with the Metropolis-Hastings algorithm. The negative log-likelihood (NLL) of samples of DDPM is computed as their mean Landau-Ginzburg Hamiltonian, subtracting the mean Landau-Ginzburg Hamiltonian of samples of the Metropolis-Hastings algorithm. The histogram of the mean value of the magnetization ϕ field values are plotted in Fig.10.

A phase transition occurs at the Curie temperature $T_c = 200$. Above the Curie temperature, the histogram of the mean magnetization has one peak centered at 0. At a certain instance, the thermal movements of molecules in our magnet may lead to a non-zero net magnetization, but the average magnetization over a long time is still zero, corresponding to a paramagnetic magnet. Below the Curie temperature, the histogram of the mean magnetization has two peaks with non-zero centers. Jumping from one peak to another at this case is difficult because the thermal movements of molecules only slightly change the mean magnetization and are insufficient to jump between peaks. This leads to a non-zero average magnetization over a long time and corresponds to a permanent magnet. Both classifier-free and characteristic guidance generate accurate samples above the Curie temperature where the DDPM is trained. However, the characteristic guidance generates better samples below the Curie temperature and has better NLL. Moreover, samples of characteristic guidance have well-separated peaks while samples of classifier-free guidance are not. This means the characteristic guidance has a better capability to model the phase transition.
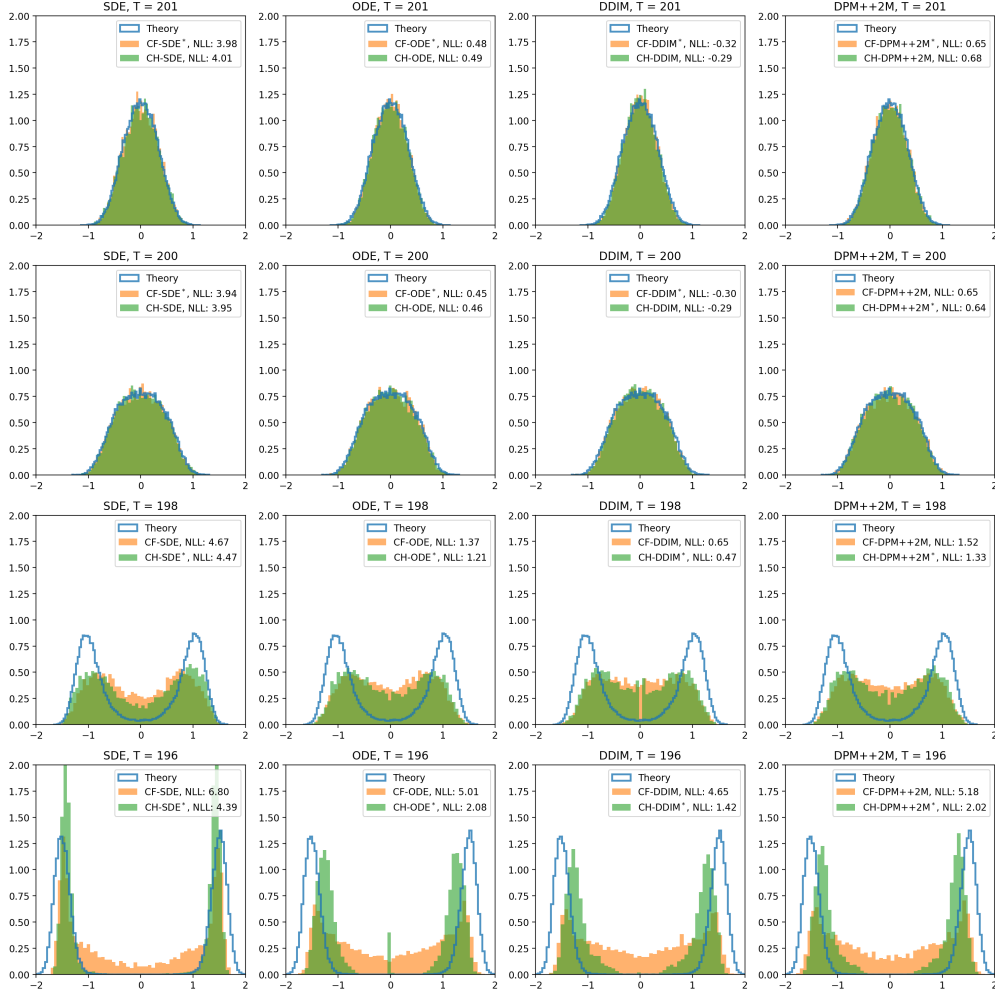
Figure 10. The histogram of the mean magnetization of the Landau-Ginzburg magnet model around the Curie temperature $T = 200$. The magnetization fields are generated by classifier-free and characteristic guidance. The characteristic guidance generates better samples below the Curie temperature $T = 200$ and has better NLL.

I. Experimental Details on Cifar-10 and ImageNet 256

Our CIFAR-10 experiment utilized a modified U-Net model with 39 million parameters, sourced from a classifier-free diffusion guidance implementation (Coderpiaobozhe, 2023). The model featured a channel size of 128 and a conditional embedding dimension of 10, operating over 1000 timesteps. Training involved 400,000 iterations with a batch size of 256, using the AdamW optimizer at a 2×10^{-4} learning rate and a 0.1 dropout rate. Stability was ensured with an EMA decay rate of 0.999, and the diffusion process followed a linear β_t schedule from 10^{-4} to 0.02 over 1000 timesteps. The characteristic guidance is computed by Alg.2 with $\gamma = 0.002$, $\eta = 0.001$, and $\alpha = 0.999$. All CIFAR-10 experiments were conducted on 1 NVIDIA Geforce RTX 3090 GPU.

For the ImageNet 256 experiment, we utilized a pretrained latent diffusion model, as detailed in (Rombach et al., 2021), featuring over 400 million parameters. This model is configured to run with a total of 1000 timesteps and works on latents with the shape of $3 \times 64 \times 64$. The characteristic guidance is computed by Alg.3 with $\gamma = 1$, $m = 2$, $\eta = 0.005$. Experiments were conducted on 6 NVIDIA Geforce RTX 4090 GPUs.

J. Visualization Results on Latent Diffusion Model

Figure 11. Comparative visualization of butterfly (323) images generated from latent diffusion model using Classifier Free Guidance (CF) versus Characteristic Guidance (CH).

Figure 12. Comparative visualization of whale (147) images generated from latent diffusion model using Classifier Free Guidance (CF) versus Characteristic Guidance (CH).
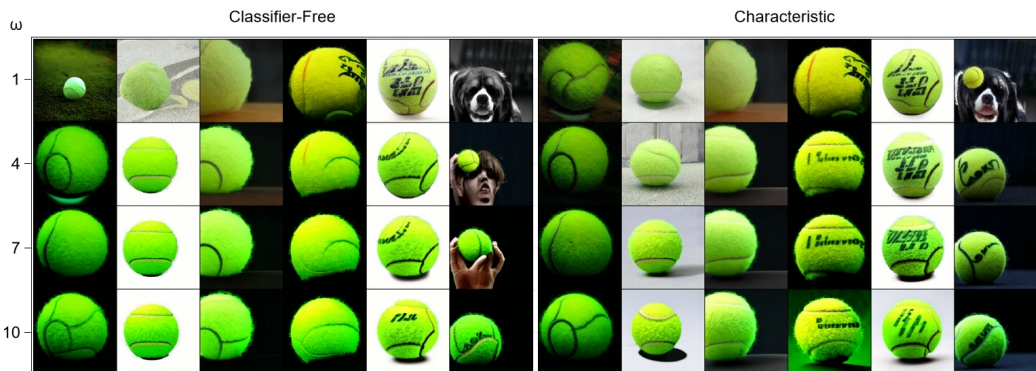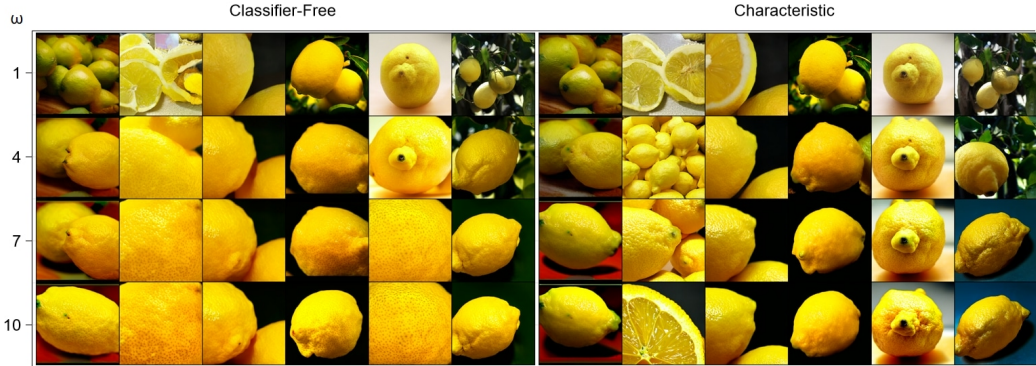

Figure 13. Comparative visualization of butterfly (852) images generated from latent diffusion model using Classifier Free Guidance (CF) versus Characteristic Guidance (CH).

Figure 14. Comparative visualization of lemon (147) images generated from latent diffusion model using Classifier Free Guidance (CF) versus Characteristic Guidance (CH).

Figure 15. Comparative visualization of barn (425) images generated from latent diffusion model using Classifier Free Guidance (CF) versus Characteristic Guidance (CH).
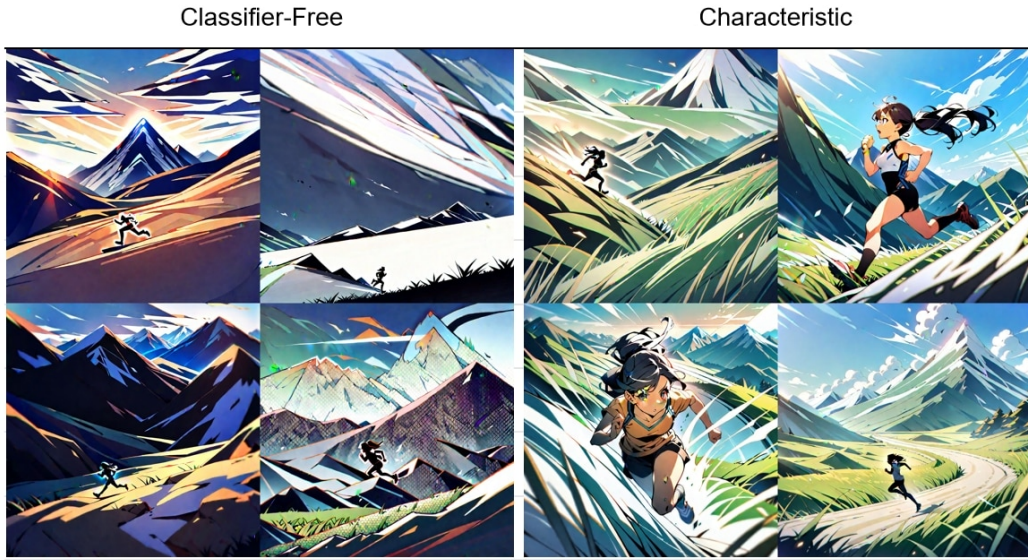
Figure 16. 1girl, running, mountain, grass. Comparative visualization of images generated from Stable Diffusion WebUI using Classifier Free Guidance versus Characteristic Guidance. Infotext of image (batch size 4, with the Characteristic Guidance Web UI extension): *1girl, running, mountain, grass. Negative prompt: low quality, worst quality. Steps: 30, Sampler: UniPC, CFG scale: 30, Seed: 0, Size: 1024x1024, Model hash: 1449e5b0b9, Model: animagineXLV3_v30, CHG: "{RegS: 5, RegR: 1, MaxI: 50, NBasis: 1, Reuse: 0, Tol: -4, IteSS: 1, ASpeed: 0.4, AStrength: 0.5, AADim: 2, CMode: 'More ControlNet'}", Version: v1.7.0*
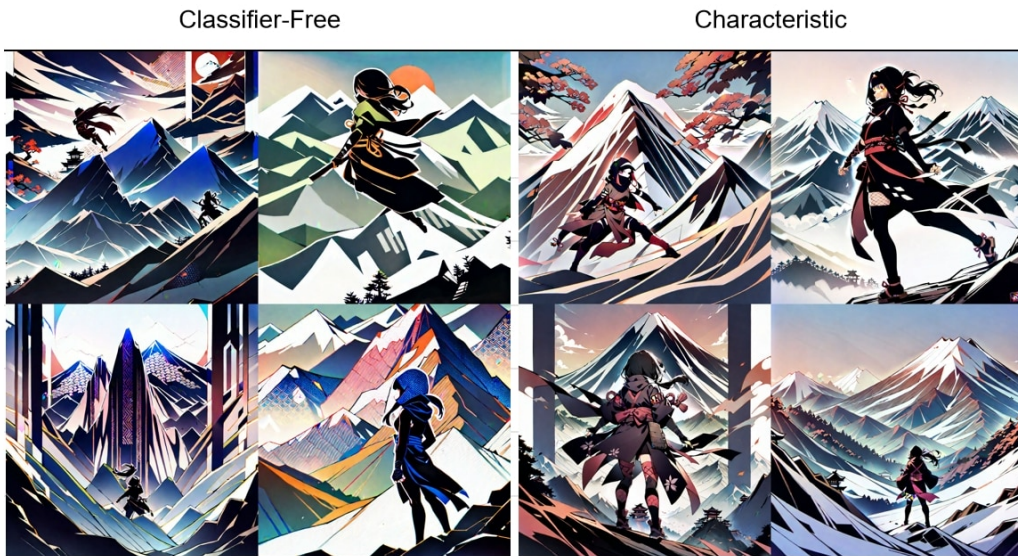

Figure 17. 1girl, ninja, mountain. Comparative visualization of images generated from Stable Diffusion WebUI using Classifier Free Guidance versus Characteristic Guidance. Infotext of image (batch size 4, with the Characteristic Guidance Web UI extension): *1girl, ninja, mountain, Negative prompt: low quality, worst quality, Steps: 30, Sampler: UniPC, CFG scale: 30, Seed: 0, Size: 1024x1024, Model hash: 1449e5b0b9, Model: animagineXLV3_v30, CHG: "{RegS: 5, RegR: 1, MaxI: 50, NBasis: 1, Reuse: 0, Tol: -4, IteSS: 1, ASpeed: 0.4, AStrength: 0.5, AADim: 2, CMode: 'More ControlNet'}", Version: v1.7.0*

Classifier-Free   Characteristic

Figure 18. frog, forest, rain. Comparative visualization of images generated from Stable Diffusion WebUI using Classifier Free Guidance versus Characteristic Guidance. Infotext of image (batch size 4, with the Characteristic Guidance Web UI extension): *frog, forest, rain, Negative prompt: low quality, worst quality, Steps: 30, Sampler: UniPC, CFG scale: 30, Seed: 0, Size: 1024x1024, Model hash: 1449e5b0b9, Model: animagineXLV3_v30, CHG: "{RegS: 5, RegR: 1, MaxI: 50, NBasis: 1, Reuse: 0, Tol: -4, IteSS: 1, ASpeed: 0.4, AStrength: 0.5, AADim: 2, CMode: 'More ControlNet'}", Version: v1.7.0*

Classifier-Free   Characteristic

Figure 19. battleship, space, moon. Comparative visualization of images generated from Stable Diffusion WebUI using Classifier Free Guidance versus Characteristic Guidance. Infotext of image (batch size 4, with the Characteristic Guidance Web UI extension): *battleship, space, moon, Negative prompt: low quality, worst quality, Steps: 30, Sampler: UniPC, CFG scale: 30, Seed: 0, Size: 1024x1024, Model hash: 1449e5b0b9, Model: animagineXLV3_v30, CHG: "{RegS: 5, RegR: 1, MaxI: 50, NBasis: 1, Reuse: 0, Tol: -4, IteSS: 1, ASpeed: 0.4, AStrength: 0.5, AADim: 2, CMode: 'More ControlNet'}", Version: v1.7.0*

26

Figure 20. 1girl, handstand, sportswear, gym. Comparative visualization of images generated from Stable Diffusion WebUI using Classifier Free Guidance, Classifier Free Guidance + Dynamical thresholding + Scale Mimic, and Characteristic Guidance. The "scale mimic" technique is a workaround to suppress artifacts of dynamical thresholding for latent space generation tasks. The seeds used to generate these images are 0,1,2,3 indicating no cherry picking.

Infotext of image (batch size 4): *1girl, handstand, sportswear, gym, Negative prompt: low quality, worst quality, Steps: 30, Sampler: DPM++ 2M Karras, CFG scale: 10, Seed: 0, Size: 1024x1024, Model hash: 1449e5b0b9, Model: animagineXLV3_v30, Version: v1.7.0*

Intotext of dynamic thresholding: *Dynamic thresholding enabled: True, Mimic scale: 7, Separate Feature Channels: True, Scaling Startpoint: MEAN, Variability Measure: AD, Interpolate Phi: 1, Threshold percentile: 100, Mimic mode: Half Cosine Down, Mimic scale minimum: 0,*

Intotext of characteristic guidance: *CHG: "{RegS: 1, RegR: 1, MaxI: 50, NBasis: 1, Reuse: 0, Tol: -4, IteSS: 1, ASpeed: 0.4, AStrength: 0.5, AADim: 2, CMode: 'More ControlNet'}",*