TU Delft

Delft University of Technology

Document Version
Final published version

Licence
Dutch Copyright Act (Article 25fa)

Citation (APA)
Goyal, M., Singh, A., & Jamali-Rad, H. (2026). Unified Control for Inference-Time Guidance of Denoising Diffusion Models. In C. Ceballos (Ed.), *Proceedings of the 2026 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (pp. 5437-5446). (Proceedings - 2026 IEEE/CVF Winter Conference on Applications of Computer Vision, WACV 2026). IEEE. https://doi.org/10.1109/WACV61042.2026.00527

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

2026 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) | 979-8-3315-5511-5/26/$31.00 ©2026 IEEE | DOI: 10.1109/WACV61042.2026.00527

Unified Control for Inference-Time Guidance of Denoising Diffusion Models

Maurya Goyal[1], Anuj Singh[1,2], Hadi Jamali-Rad[1,2]

[1]Delft University of Technology, The Netherlands
[2]Shell Global Solutions International B.V., Amsterdam, The Netherlands

Abstract

Aligning diffusion model outputs with downstream objectives is essential for improving task-specific performance. Broadly, inference-time training-free approaches for aligning diffusion models can be categorized into two main strategies: ***sampling-based*** *methods, which explore multiple candidate outputs and select those with higher reward signals, and* ***gradient-guided*** *methods, which use differentiable reward approximations to directly steer the generation process. In this work, we propose a universal algorithm,* ***UniCoDe****, which brings together the strengths of sampling and gradient-based guidance into a unified framework.* `UniCoDe` *integrates local gradient signals during sampling, thereby addressing the sampling inefficiency inherent in complex reward-based sampling approaches. By cohesively combining these two paradigms,* `UniCoDe` *enables more efficient sampling while offering better trade-offs between reward alignment and divergence from the diffusion unconditional prior. Empirical results demonstrate that* `UniCoDe` *remains competitive with state-of-the-art baselines across a range of tasks. The code is available at https://github.com/maurya-goyal10/UniCoDe*

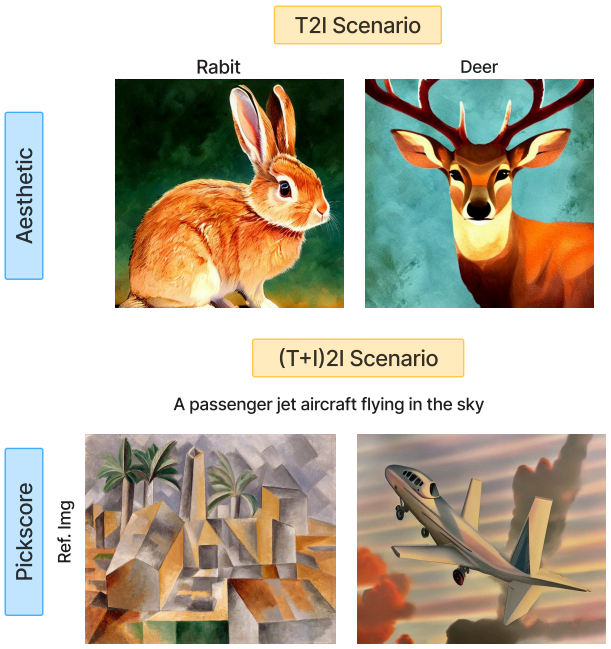


Figure 1. `UniCoDe` generates high-quality images across diverse reward objectives such as aesthetic, pickscore, and in (T+I)2I guided using pickscore, demonstrating adaptable performance.

1. Introduction

Diffusion models [1–4] have illustrated impressive capabilities in sampling complex data distributions, allowing them to generate high-quality images [5–7], videos [8, 9], audio [10, 11], and even biological structures such as proteins and DNA [12, 13]. These models are typically trained on large and diverse datasets, allowing them to capture a wide range of patterns and variations [14]. However, in many practical scenarios, it becomes important to *guide* the generative process towards outputs maximizing specific properties, such as aesthetic appeal or adherence with user-provided prompts or images. Thus, conditioning diffusion models to produce outputs that meet user-defined criteria has emerged as a consequential problem.

There are several ways for conditioning generative models to produce outputs aligned with the user's desired objective. The most straightforward method is fine-tuning [15–18] the unconditional foundational model for each specific task. Fine-tuning has demonstrated strong performance in aligning models with specific tasks, enabling the base model to learn domain-specific nuances, and once fine-tuned, the model can generate aligned outputs rapidly at inference time, requiring no additional computation during sampling. However, fine-tuning is computationally expensive and typically requires substantial labeled data or high-quality reward signals. It also potentially leads to reward over-optimization, a form of reward hacking where the model exploits loopholes in the reward function to produce

2642-9381/26/$31.00 ©2026 IEEE
DOI 10.1109/WACV61042.2026.00527

Authorized licensed use limited to: TU Delft Library. Downloaded on June 05,2026 at 07:53:48 UTC from IEEE Xplore. Restrictions apply.

high-reward outputs that do not align with true user intent. As an alternative, inference-time guidance [4, 6] offers a more flexible solution by enabling training-free conditioning through external reward models during denoising.

In this work, we focus exclusively on inference-time guidance for conditional image generation scenarios, specifically in the setting where the reward model is applied to the final denoised (clean) samples rather than intermediate noisy ones. This allows us to leverage off-the-shelf reward models directly, without modifying them or extra training. The reward models can be any loss function, classifier, or probability estimator that measures how well the output aligns with the given user conditioning signal.

Training-free guidance methods can be extensively categorized into gradient-based [19–23] and sampling-based approaches [24, 25]. Gradient guidance involves perturbing samples during the denoising process in the direction that increases the reward, effectively guiding the generation toward more desirable outcomes. Conversely, sampling-based methods utilize stochasticity by generating multiple candidate samples and preferentially sampling more from those expected to yield higher rewards. While sampling-based approaches often offer a better trade-off between target reward alignment and the divergence from the prior [25], they come at a high computational cost, as the number of required samples increases exponentially with the divergence between the prior distribution and the desired posterior distribution [26]. On the other hand, gradient-based guidance can directly steer the generation toward desirable outputs but faces limitations in flexibility, differentiability, and generalization. In this work, we explore how these two paradigms can be effectively unified to leverage their complementary strengths and aim to answer the key question: **Can sampling-based and gradient-based inference-time guidance be unified to reduce sampling overhead, enhance output quality, and achieve better trade-offs between reward alignment and prior divergence?**

Inspired by Controlled Denoising (`CoDe`) [25] proposed by Singh *et al*., we incorporate blockwise guidance with blockwise sampling to answer this question. This unified framework leverages the strengths of both sampling-based and gradient-based methods: the flexibility of sample-based reward maximization and the efficiency of gradient-driven updates. By using gradient signals to explicitly steer the prior distribution toward the desired posterior, we reduce the number of samples required during the reverse diffusion process, thus improving sampling efficiency while maintaining high output quality and strong alignment with the target conditioning. Furthermore, we also investigate the following set of extensions designed to enhance the overall performance and efficiency of the method, while also improving its generality and adaptability across different tasks and reward structures:

- **Sampling schedule:** Introducing a scheduled sampling strategy that allocates more sampling budget to steps in the denoising process where the reward signal is expected to be most significant.
- **Non-greedy sampling:** Replacing the myopic greedy sampling strategy with more flexible alternatives such as multinomial sampling to better control the exploration-exploitation tradeoff.
- **Clustering-based guidance:** Incorporating clustering techniques to limit the overhead of applying gradient guidance to every sample individually.
- **Non-differentiable rewards:** Also test the method on non-differentiable rewards using zero-order optimization, utilizing the torchopt[27] library.

2. Preliminaries and Related Work

Diffusion Models. Diffusion models are Markov Chains where for the forward process, we iteratively noise a clean input by adding Gaussian noise to it. Thus, as we go from t=0 (clean image) to t= T and for large enough T, we reach Gaussian noise. Using a neural network we learn to reverse this process $p_\theta(x_{t-1}|x_t)$ (denoising) since $q(x_t|x_{t-1})$ is known. Consequently, iteratively applying $p_\theta(x_{t-1}|x_t)$ to a randomly sampled Gaussian noise, we can sample from the target distribution or generate a new image.

Score Based Formulation. Diffusion models have a score-based formulation through SDE [4] to estimate the score function, which is defined as $\nabla_{x_t} \log p(x_t)$ using a neural network. The forward process is defined as:

$$dx = \mathbf{f}(\mathbf{x}, t)dt + g(t)d\mathbf{w}. \tag{1}$$

In the above equation $f(x, t)$ is vectored value function called the *drift* coefficient of $x(t)$ and $g(t)$ is a scalar function known as the *diffusion* coefficient of $x(t)$, both of which are known. The reverse process is defined as:

$$dx = [\mathbf{f}(\mathbf{x}, t)dt - g(t)^2 \nabla_x \log p_t(\mathbf{x})]dt + g(t)d\bar{\mathbf{w}}. \tag{2}$$

Where $\bar{\mathbf{w}}$ is the standard Wiener process when time flows backward from T to 0. The only unknown term is the score function i.e. $\nabla_x \log p_t(x)$ which the neural network learns to predict. Thus after sampling $x(T) \sim p_T$ and reversing the process we can obtain the clean image (sample from p_0).

Alignment Objective. Considering an off-the-shelf reward function $r : \mathcal{X} \to \mathbb{R}$ that quantifies the quality of alignment of generated samples with a given conditioning. Let $\pi(\cdot)$ denote the reward-aligned diffusion model and $p(\cdot)$ the base, unconditional reference diffusion model [25]. From a reinforcement learning perspective, the alignment objective can be formulated as optimizing π to maximize the expected reward under the aligned distribution relative to p, thereby encouraging the generation of samples satisfy-

Authorized licensed use limited to: TU Delft Library. Downloaded on June 05,2026 at 07:53:48 UTC from IEEE Xplore. Restrictions apply.

ing:

$$\pi^*_\lambda = \arg\max_\pi \Bigg[\lambda \, \mathbb{E}_{x_{t-1}\sim\pi(x_{t-1}|x_t)} \Big[\mathbb{E}_{x_0\sim p(x_{t-1}|x_t)}[r(x_0)] \Big] - \mathrm{KL}\big(\pi(x_{t-1}|x_t) \,\|\, p(x_{t-1}|x_t)\big) \Bigg], \quad (3)$$

where $\lambda \in \mathbb{R}^{\geq 0}$ trades off reward for drift from the base diffusion model p.

Gradient Based Guidance. [4, 6] For conditional generation[6], the goal is to sample from $p(x|y)$ instead of $p(x)$. Thus wrt to the score-based formulation we want the $\nabla_{x_t} \log p(x_t|y)$. Using Bayes Rule [28]: ∇ represents ∇_{x_t}

$$\begin{aligned} \nabla \log p(x_t|y) &= \nabla \log \left(\frac{p(x_t)p(y|x_t)}{p(y)} \right) \\ &= \nabla \log p(x_t) + \nabla \log p(y|x_t) - \nabla \log p(y) \\ &= \underbrace{\nabla \log p(x_t)}_{\text{unconditional score}} + \underbrace{\nabla \log p(y|x_t)}_{\text{adversarial gradient}}. \end{aligned} \quad (4)$$

In Eq. 4, we know the first-term unconditional score from the foundational diffusion model and we get the second term using an off-the-shelf reward model.

Sampling Based Guidance (CoDe). The idea of using sampling for guidance builds on the Best-of-N (BoN) approach [29]. When performing conditional sampling from the posterior $p(x|y)$ given the prior $p(x)$, we can draw multiple independent samples from the prior and at the end of the denoising select the one that best aligns with the posterior. A more aggressive way is to estimate the expected alignment with the posterior and then explore more in the direction expected to give higher rewards [24, 25].

Further details on the background and a more comprehensive review of the related literature are provided in the Appendix (refer A.'Related Works' and B.'Background').

3. Methodology

We employ blockwise gradient guidance (Eq. 4) to compute and add gradients at the respective timestep for every parallel stream before performing sampling. This setup allows us to maintain multiple candidate trajectories (streams) that explore different directions in the sample space. Among these, the method preferentially explores directions that are expected to yield higher rewards, while the gradient guidance simultaneously nudges each stream closer to the posterior distribution. The methodology is detailed in Algorithm 1 and illustrated schematically in Fig. 2. Specifically, blockwise gradient guidance (Appendix Algorithm 1) is applied independently across streams (lines 7–10), ensuring localized posterior alignment, after which blockwise sampling (Appendix Algorithm 2) (lines 13–16) selects and emphasizes high-reward trajectories. In this way, our unified framework (UniCoDe) integrates exploration across multiple streams with posterior alignment via gradient guidance, distinguishing it from prior approaches (CoDe,BoN).

Algorithm 1 UniCoDe

Require: Timesteps: T, Samples: N, Blocksizes: B_g, B_s, Temp. τ, p_θ, Reward fn. r, guidance scale: γ

1: Sample initial noise: $\{z_T^{(n)}\}_{n=1}^N \sim \mathcal{N}(0, I)$
2: Initialize counter: $s \leftarrow 1$
3: **for** $t = T, \ldots, 1$ **do**
4: $\quad \{z_{t-1}^{(n)}\}_{n=1}^N \leftarrow p_\theta(\{z_t^{(n)}\}_{n=1}^N, t)$
5: $\quad$ **if** $\mathrm{mod}(s, B_g) = 0$ **then**
6: $\quad\quad$ **if** gradient condition satisfied **then**
7: $\quad\quad\quad$ **for** $k = 1$ to N **do**
8: $\quad\quad\quad\quad g_k \leftarrow \textsc{Grad}(z_{t-1}^{(k)}, t-1)$
9: $\quad\quad\quad\quad z_{t-1}^{(k)} \leftarrow z_{t-1}^{(k)} + \gamma \cdot g_k$
10: $\quad\quad\quad$ **end for**
11: $\quad\quad$ **end if**
12: $\quad$ **end if**
13: $\quad$ **if** $\mathrm{mod}(s, B_s) = 0$ **then**
14: $\quad\quad \{\hat{z}_0^{(n)}\} \leftarrow \mathbb{E}_{p_\theta}(\{z_0^{(n)}\} | \{z_{t-1}^{(n)}\}, t-1)$
15: $\quad\quad \{z_{t-1}^{(n)}\} \leftarrow \textsc{Sample}(\{z_{t-1}^{(n)}\}, r(D(\{\hat{z}_0^{(n)}\})), \tau)$
16: $\quad$ **end if**
17: $\quad s \leftarrow s + 1$
18: **end for**
19: **return** z_0

To make the sampling more general we also change the greedy sampling to temperature-based multinomial sampling, this allows the model to consider paths that can give higher rewards later in the denoising and using the temperature we can adjust the aggressiveness of the sampling. Thus it allows the sampling to explore more and can adjust the tradeoff between exploring and exploiting by changing the temperature parameter.

We also experiment with other extensions like having a sampling schedule that allows exploring more in the denoising steps expected to give more rewards. Moreover, we experiment with clustering to decrease the overhead of adding a gradient to all the sample streams and cluster together points in the latent space that are closer, calculate one gradient for all of them, and add it to all the points in the cluster.

4. Experiments

We evaluate the performance of UniCoDe by comparing it against a range of state-of-the-art guidance methods across both Text-to-Image (T2I) and Text-and-Image-to-Image ((T+I)2I) generation tasks. These evaluations are conducted under both differentiable and non-differentiable reward settings to ensure a comprehensive analysis. All experiments are conducted using the pre-trained Stable Diffusion v1.5 model [7], which was trained on the LAION-

Authorized licensed use limited to: TU Delft Library. Downloaded on June 05,2026 at 07:53:48 UTC from IEEE Xplore. Restrictions apply.

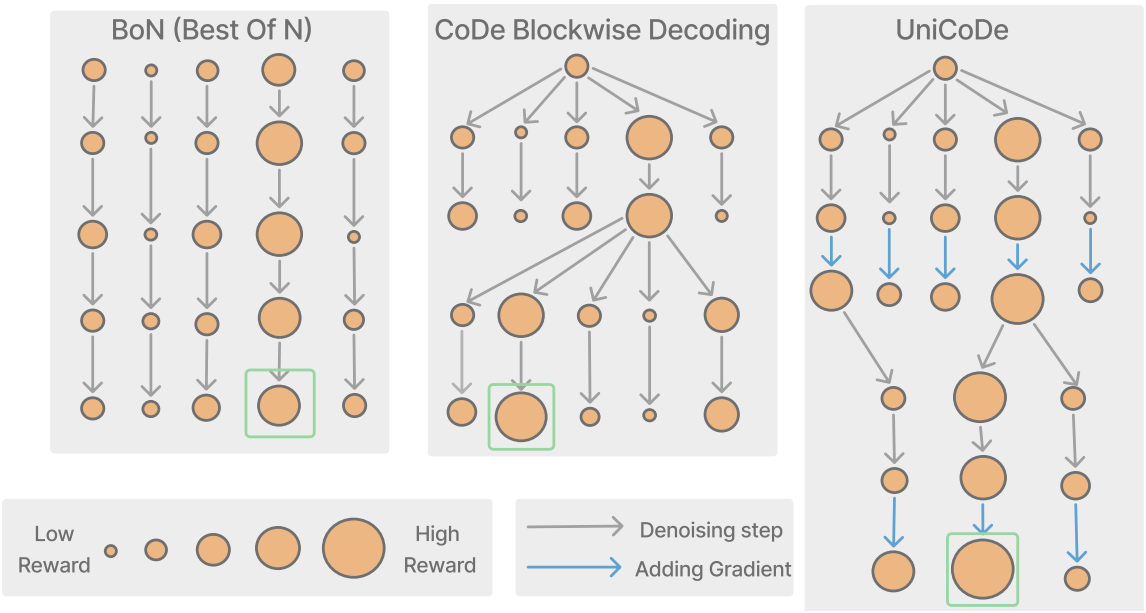


Figure 2. Schematic illustration of sampling-based guidance methods: **BoN** runs parallel diffusion streams and selects the highest-reward sample after full denoising; **CoDe** performs greedy blockwise selection to focus on promising regions early; **UniCoDe** enhances CoDe with blockwise gradient-based guidance, scheduled particle sizes, and multinomial sampling for better efficiency and performance.



Figure 3. Aesthetic alignment: UniCoDe consistently produces visually pleasing images with strong prompt adherence. While it has higher CMMD (divergence), it avoids reward hacking and tends to yield more aesthetically aligned outputs

Authorized licensed use limited to: TU Delft Library. Downloaded on June 05,2026 at 07:53:48 UTC from IEEE Xplore. Restrictions apply.

400M dataset [30]. Stable Diffusion is a text-conditioned latent diffusion model trained on paired image-text data. It operates in latent space to perform denoising and leverages text-conditioning to generate prompt aligned images.

To ensure fair and reproducible comparisons, we generate 10 images per experimental setting (i.e., for each prompt-reference image pair) using 500 DDPM steps. All experiments are conducted using NVIDIA A40 GPUs, with the random seed fixed at 2024 for reproducibility. Both qualitative and quantitative results are reported across a variety of scenarios to highlight the capabilities of UniCoDe. For all scenarios, the sampling block size is set to 5. The gradient block size is adjusted based on the reward model: 5 for Aesthetic, 4 for pickscore, and 2 for the (T+I)2I setting. Additionally, unless specified otherwise, the guidance scale is fixed at 0.2 across all experiments.

Evaluation Settings and Metrics. We evaluate the performance of all methods using a comprehensive set of metrics designed to capture different aspects of alignment, quality, and divergence from the unconditional model. These include the expected reward, Frechet Inception Distance (FID) [31], CLIP-based Maximum Mean Discrepancy (CMMD) [32], I-Gram [33, 34], and CLIPScore [35], referred to as T-CLIP throughout the paper. FID and CMMD measure the deviation from the prior distribution defined by the unconditioned Stable Diffusion model, providing a sense of how far the generations diverge from typical outputs. I-Gram captures style and texture similarity between the reference and generated images, particularly relevant in the (T+I)2I setting, while T-CLIP assesses semantic alignment with the input prompt. To support diverse evaluation goals, we consider a variety of reward models: the LAION Aesthetic Predictor v2 [30, 36] for image-based aesthetic quality, pickscore [37] for text-image alignment, a multi-reward configuration that combines aesthetic and pickscore, style transfer, and compressibility [38] a non-differentiable reward facilitating compact image representations. For the quantitative tables, all results are normalized with respect to the base Stable Diffusion model.

Baselines. In all experimental settings, we compare UniCoDe against CoDe with both $N{=}40$ ($\texttt{CoDe}_{40}$) and $N{=}4$ ($\texttt{CoDe}_{4}$), where N denotes the number of samples generated. We treat the $N{=}4$ configuration as the primary baseline, as UniCoDe itself operates using only 4 sample streams per block. For the Aesthetic and pickscore reward models, we additionally benchmark against state-of-the-art gradient-based guidance methods, including MPGD [22], FreeDoM [21], and Universal Guidance (UG) [20], which are known to improve upon earlier approaches like DPS [19] by providing more accurate gradient estimates.

Notation We denote different variants of UniCoDe as follows: $\texttt{UniCoDe}_1$ and $\texttt{UniCoDe}_2$ use a fixed sampling size of 4, with greedy and multinomial sampling, respectively. $\texttt{UniCoDe}_3$ and $\texttt{UniCoDe}_4$ employ scheduled sampling sizes, again with greedy and multinomial sampling, respectively. $\texttt{UniCoDe}_5$ extends $\texttt{UniCoDe}_4$ by incorporating K-Means clustering (K=2) during sampling. Appendix H and I provide hyperparameter analyses that illustrate how these parameters can be effectively controlled.

Method	Aesthetic Guidance				
	Rew. (↑)	FID (↓)	CMMD (↓)	T-CLIP (↑)	Time (↓)
SD1.5	1.00	1.000	1.00	1.000	1.0
$\texttt{CoDe}_4$	1.08	1.026	1.36	1.001	21.7
$\texttt{CoDe}_{40}$	1.19	1.037	**1.92**	1.003	179.8
$\texttt{UniCoDe}_1$	1.60	1.059	3.24	1.003	36.5
$\texttt{UniCoDe}_2$	1.60	1.048	3.28	1.003	36.0
$\texttt{UniCoDe}_3$	1.61	1.053	3.17	1.002	40.4
$\texttt{UniCoDe}_4$	**1.61**	**1.027**	3.12	1.003	41.3
$\texttt{UniCoDe}_5$	1.58	1.034	2.97	**1.005**	**29.4**

Table 1. Aesthetic guidance comparison, UniCoDe improves aesthetic alignment and reduces runtime 4–5×, with higher CMMD but no reward hacking and strong text alignment.

4.1. Image Reward Models

We first experiment with reward models just based only on the final generated image such as the **aesthetic scorer**. To guide the diffusion denoising process towards generating aesthetically pleasing images, we use the LAION aesthetic predictor V2 [30], which leverages a multi-layer perceptron (MLP) architecture trained on top of CLIP embeddings. This model's training data consists of 176,000 human image ratings, spanning a range from 1 to 10, with images achieving a score of 10 being considered art pieces. We evaluate them on the evaluation dataset of animals from ImageNet [39] (containing 51 prompts).

The quantitative results are summarized in Table 1, we can see that UniCoDe not only takes roughly $4{-}5$ times less time as compared to CoDe [25] but also gives much better rewards. This comes at the cost of higher divergence from the prior distribution which is shown by the increase in the CMMD, while the FID and the T-CLIP stay similar. We also ablate on the different settings in UniCoDe, as evident from the results the major improvement comes from adding the gradients, and then we can observe little improvements from scheduling samples and making the sampling multinomial. Furthermore, using clustering we can further reduce the time taken at the cost of the rewards.

We further compare the performance of UniCoDe against state-of-the-art gradient-based guidance methods, as shown in Table 2. While UniCoDe achieves higher reward scores, this improvement is accompanied by a greater divergence from the unconditional prior. However, as demonstrated by the qualitative results in Fig. 3, where UniCoDe is evaluated alongside all competing methods, this increased divergence does not come at the cost of semantic or perceptual alignment. The samples suggest that UniCoDe is capable of generating aesthetically pleasing images that re-

Authorized licensed use limited to: TU Delft Library. Downloaded on June 05,2026 at 07:53:48 UTC from IEEE Xplore. Restrictions apply.

main faithful to the given prompts, indicating that its strong reward performance is not the result of reward hacking but rather meaningful alignment with the conditioning inputs.

Method	Aesthetic Guidance			
	Rew (↑)	FID (↓)	CMMD (↓)	T-CLIP (↑)
MPGD	1.15	1.081	**2.325**	0.990
FreeDoM	1.27	1.184	3.138	0.983
UG	1.26	1.742	3.670	0.968
UniCoDe	**1.61**	**1.027**	3.123	**1.003**

Table 2. Aesthetic guidance results comparing gradient guidance methods to UniCoDe. UniCoDe achieves better reward and better prompt alignment (T-CLIP score)

In this section, we guide the model to adhere better to the given text prompt. Although Stable Diffusion v1.5 is trained on text input conditioning, the performance (adherence to prompt and output quality) starts degrading as the prompt complexity increases. Hence, we use **pickscore** [37] as the reward to generate good-quality images adhering to the prompts. Pickscore is a CLIP-based scoring function trained on Pick-a-Pic, a large dataset of text-to-image prompts and real user preferences over generated images. We evaluate the performance of the generated images with prompts taken from the HPDv2 [40] evaluation settings.

The quantitative results are given in Tables 3 and 4, and the qualitative examples are shown in Fig. 4. UniCoDe significantly improves efficiency ($3-4\times$ faster), achieves higher rewards, and maintains less divergence from the prior (lower CMMD). Moreover, UniCoDe outperforms gradient guidance methods, providing better prompt alignment and image quality. We also evaluate this setting on a more capable diffusion backbone SD 2.1, and Appendix E Table 1 shows that the same advantages hold.

The tradeoff curves in Fig. 5 demonstrate that UniCoDe achieves the optimal balance, attaining higher rewards (pickscore) with lower divergence (CMMD) from the prior (unconditional base model) compared to other methods. Further results of this ablation is present in Appendix G.

4.2. Multi-Reward

In this section, we consider an objective that balances multiple rewards. Specifically, we guide the generation using a weighted sum of the aesthetic score and the pickscore: $\gamma_1 \times$ **aesthetic** $+ \gamma_2 \times$ **pickscore** . The aesthetic score motivates aesthetically pleasing outputs, while the pickscore assures alignment with the prompt based on human evaluations. This grants users a more controlled trade-off between visual quality and prompt fidelity. By adjusting the weights γ_1 and γ_2, users can tailor the generation process to their specific preferences.

A plot of aesthetic score versus pickscore is shown in Fig. 6. In comparison to the base diffusion model, our approach allows for controlled guidance by adjusting the parameters γ_1 and γ_2, enabling a flexible trade-off between multiple reward signals such as aesthetic quality and pickscore. The performance obtained by UniCoDe demonstrates significantly improved rewards across multiple reward components. Specifically, UniCoDe significantly improves performance in multi-reward scenarios, boosting the aesthetic reward by about 43% and the pickscore by 7% compared to CoDe on average. Against gradient guidance methods like FreeDoM, UniCoDe still outperforms with increases of 29% (aesthetic) and 6% (pickscore).

Method	Pickscore Guidance (T2I-Based)			
	Rew. (↑)	CMMD (↓)	T-CLIP (↑)	Time (↓)
SD1.5	1.000	1.000	1.000	1.0
CoDe_4	1.054	1.220	1.008	5.7
CoDe_{40}	1.107	1.611	1.016	51.2
UniCoDe_1	1.116	1.469	**1.019**	13.5
UniCoDe_2	1.116	1.448	1.014	13.5
UniCoDe_3	**1.118**	1.439	1.016	13.5
UniCoDe_4	1.114	1.395	1.014	13.5
UniCoDe_5	1.106	**1.334**	1.013	**10.4**

Table 3. T2I-based Pickscore guidance results. UniCoDe improves efficiency (3–4× faster), prompt alignment (↑ Pickscore, ↑ T-CLIP), and prior preservation (↓ CMMD).

Method	Pickscore Guidance		
	Rew (↑)	CMMD (↓)	T-CLIP (↑)
MPGD	1.0289	1.5747	1.0032
FreeDoM	1.0444	1.5207	1.0067
UG	1.0423	1.9716	1.0056
UniCoDe	**1.1183**	**1.4394**	**1.0154**

Table 4. Pickscore guidance results comparing gradient guidance methods to UniCoDe. It deviates less from the prior (lower CMMD) and offers better prompt alignment (higher pickscore and T-CLIP).

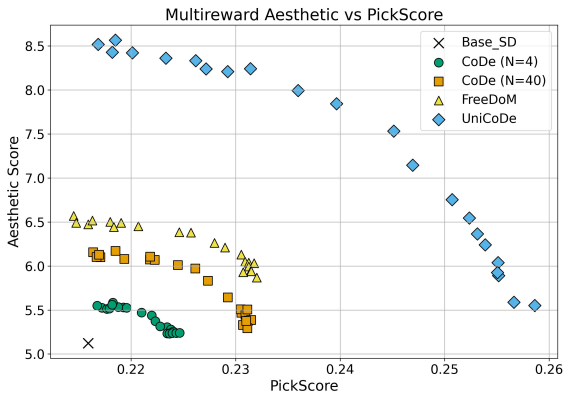


Figure 6. In a multi-reward guidance scenario, UniCoDe consistently achieves higher aesthetic and pickscore, optimizing both rewards compared to other methods.

Authorized licensed use limited to: TU Delft Library. Downloaded on June 05,2026 at 07:53:48 UTC from IEEE Xplore. Restrictions apply.



Figure 4. These qualitative examples illustrate how `UniCoDe` produces images that more accurately and richly reflect the nuances of the input prompts across various scenarios. This is quantitatively supported by higher pickscore and T-CLIP scores (Tables 4, 3).

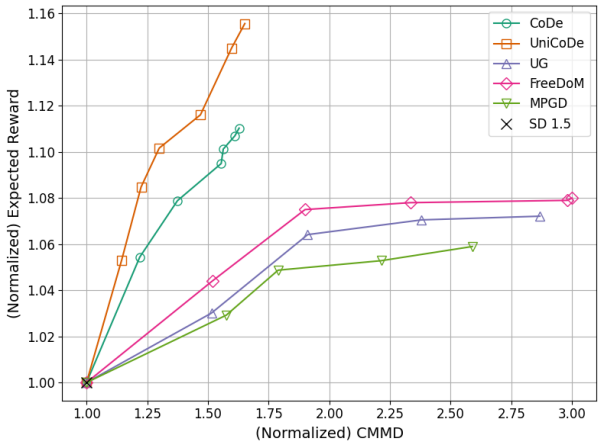


Figure 5. Ablation of rew. (pickscore) vs CMMD (deviation)

4.3. (T+I)2I (Text-and-Image-to-Image)

As we move through the experiments we keep on adding more control. This setting blends an additional reference image, enabling the generation of a new image that not only aligns with a given text prompt but also preserves the stylistic elements of the provided reference. We use `SDEdit` [41] for the style transfer and start the denoising from an intermediary point (controlled by η as $\eta \times T$) by adding noise to the reference image and then during the denoising using the pickscore as the reward to maintain a strong adherence to the prompt.

This offers users more fine-grained control over multiple competing aspects of image generation, specifically balancing adherence to the text prompt with matching the stylistic elements of a reference image. In this case, for prompt adherence, we are guiding the model using the pickscore explicitly while for style transfer it starts the denoising from the noised reference image (η=0.6) [41]. Thus we want the final generated image to preserve the style and structure of the reference image and at the same time be able to adhere to the provided prompt. As seen in Table 5 and Figure 7 `SDEdit` prioritizes style transfer. However due to the absence of guidance its effectiveness in capturing prompt intricacies is limited. In comparison, both `CoDe` and `UniCoDe` demonstrate the capacity to balance both style and prompt adherence. `UniCoDe` achieves a favorable balance between prompt adherence, style transfer, and efficiency, performing slightly better visually while operating at nearly twice the speed of `CoDe`.

4.4. Non-Differentiable Rewards

One of the key advantages of `CoDe` is its ability to operate in settings with non-differentiable rewards, as it does not rely on gradient-based optimization. Motivated by this, we also extended our algorithm to such scenarios. We use zero-order optimization, which enables gradient approximation without explicitly requiring reward differentiability.

[1] UniCoDe$_1$ uses N=35, due to the noisy gradients

Authorized licensed use limited to: TU Delft Library. Downloaded on June 05,2026 at 07:53:48 UTC from IEEE Xplore. Restrictions apply.



Figure 7. (T+I)2I with pickscore: UniCoDe best balances style and prompt fidelity; SDEdit favors style over prompt adherance

Method	(T+I)2I Pickscore Guidance				
	Rew.	I-Gr.	CMMD	T-CLIP	Time
SDEdit	1.000	1.00	1.00	1.000	1.0
CoDe_4	1.046	0.97	1.11	1.025	5.3
CoDe_{40}	1.089	0.94	**1.19**	1.032	47.8
UniCoDe_1	1.090	**0.97**	1.24	**1.033**	18.7
UniCoDe_2	**1.092**	0.97	1.21	1.028	**17.5**

Table 5. UniCoDe improves guidance efficiency (2–3× faster), enhances prompt alignment (Rew., T-CLIP), and better preserves style (I-Gram), with increased prior deviation (CMMD).

Method	Compression Guidance				
	Rew.	FID	CMMD	T-CLIP	Time
SD 1.5	1.000	1.000	1.000	1.000	1.00
CoDe_4	1.676	1.132	1.806	1.004	4.99
CoDe_{40}	3.219	**1.867**	**5.293**	**0.982**	**45.76**
UniCoDe_1 [1]	**3.220**	2.048	5.671	0.980	238.60

Table 6. UniCoDe improves reward and compression performance over CoDe, with slightly reduced T-CLIP alignment and increased compute time.

For the non-differentiable reward, we use image compressibility [38] which measures the size of the image in kilobytes. Therefore we guide the model to create lightweight compressible images. The quantitative results are summarized in Table 6. Approximating gradients in this setting is extremely challenging. Even with a large number of forward passes, the estimated gradients remain blurry due to the high dimensionality of the image space. As a result, the efficiency gained by backpropagating through the reward and denoising pipeline is offset by the high computational cost of noisy gradient estimation. This inefficiency is evident from the results in Table 6. In such scenarios, sampling-based guidance methods like CoDe are more practical and effective.

5. Conclusion and Future Work

We propose a unified framework that efficiently combines blockwise sampling with blockwise gradient-based guidance, reducing sampling overhead while maintaining strong performance trade-offs. This integration makes UniCoDe more effective for differentiable reward settings, where it outperforms purely sampling-based strategies in both efficiency and reward quality. However, with non-differentiable rewards, gradient estimates become highly noisy, requiring many particles for accuracy, thereby leading to degraded performance compared to purely sampling-based methods in such cases.

Despite progress in conditional image generation, slow inference remains a bottleneck, future work should aim to narrow the gap with unconditional diffusion. Additionally, exploring dynamic temperature scheduling and adaptive particle allocation during sampling may improve generalization and efficiency, potentially boosting reward outcomes across a wider range of tasks. Alternatively, for non-differentiable rewards, training-based gradient approximations like surrogate models could help extend efficiency gains to that domain.

Authorized licensed use limited to: TU Delft Library. Downloaded on June 05,2026 at 07:53:48 UTC from IEEE Xplore. Restrictions apply.

References

[1] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pages 2256–2265. pmlr, 2015. 1

[2] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[3] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.

[4] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020. 1, 2, 3

[5] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pages 8162–8171. PMLR, 2021. 1

[6] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021. 2, 3

[7] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 1, 3

[8] Omer Bar-Tal, Hila Chefer, Omer Tov, Charles Herrmann, Roni Paiss, Shiran Zada, Ariel Ephrat, Junhwa Hur, Guanghui Liu, Amit Raj, et al. Lumiere: A space-time diffusion model for video generation. In *SIGGRAPH Asia 2024 Conference Papers*, pages 1–11, 2024. 1

[9] Yuta Oshima, Masahiro Suzuki, Yutaka Matsuo, and Hiroki Furuta. Inference-time text-to-video alignment with diffusion latent beam search. *arXiv preprint arXiv:2501.19252*, 2025. 1

[10] Zhifeng Kong, Wei Ping, Jiaji Huang, Kexin Zhao, and Bryan Catanzaro. Diffwave: A versatile diffusion model for audio synthesis. *arXiv preprint arXiv:2009.09761*, 2020. 1

[11] Rongjie Huang, Jiawei Huang, Dongchao Yang, Yi Ren, Luping Liu, Mingze Li, Zhenhui Ye, Jinglin Liu, Xiang Yin, and Zhou Zhao. Make-an-audio: Text-to-audio generation with prompt-enhanced diffusion models. In *International Conference on Machine Learning*, pages 13916–13932. PMLR, 2023. 1

[12] Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach, Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976):1089–1100, 2023. 1

[13] Bowen Jing, Gabriele Corso, Jeffrey Chang, Regina Barzilay, and Tommi Jaakkola. Torsional diffusion for molecular conformer generation. *Advances in neural information processing systems*, 35:24240–24253, 2022. 1

[14] Andreas Blattmann, Tim Dockhorn, Sumith Kulal, Daniel Mendelevitch, Maciej Kilian, Dominik Lorenz, Yam Levi, Zion English, Vikram Voleti, Adam Letts, et al. Stable video diffusion: Scaling latent video diffusion models to large datasets. *arXiv preprint arXiv:2311.15127*, 2023. 1

[15] Kimin Lee, Hao Liu, Moonkyung Ryu, Olivia Watkins, Yuqing Du, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, and Shixiang Shane Gu. Aligning text-to-image models using human feedback. *arXiv preprint arXiv:2302.12192*, 2023. 1

[16] Mihir Prabhudesai, Anirudh Goyal, Deepak Pathak, and Katerina Fragkiadaki. Aligning text-to-image diffusion models with reward backpropagation. 2023.

[17] Kevin Black, Michael Janner, Yilun Du, Ilya Kostrikov, and Sergey Levine. Training diffusion models with reinforcement learning. *arXiv preprint arXiv:2305.13301*, 2023.

[18] Masatoshi Uehara, Yulai Zhao, Kevin Black, Ehsan Hajiramezanali, Gabriele Scalia, Nathaniel Lee Diamant, Alex M Tseng, Sergey Levine, and Tommaso Biancalani. Feedback efficient online fine-tuning of diffusion models. *arXiv preprint arXiv:2402.16359*, 2024. 1

[19] Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687*, 2022. 2, 5

[20] Arpit Bansal, Hong-Min Chu, Avi Schwarzschild, Soumyadip Sengupta, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Universal guidance for diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 843–852, 2023. 5

[21] Jiwen Yu, Yinhuai Wang, Chen Zhao, Bernard Ghanem, and Jian Zhang. Freedom: Training-free energy-guided conditional diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 23174–23184, 2023. 5

[22] Yutong He, Naoki Murata, Chieh-Hsin Lai, Yuhta Takida, Toshimitsu Uesaka, Dongjun Kim, Wei-Hsiang Liao, Yuki Mitsufuji, J Zico Kolter, Ruslan Salakhutdinov, et al. Manifold preserving guided diffusion. *arXiv preprint arXiv:2311.16424*, 2023. 5

[23] Haotian Ye, Haowei Lin, Jiaqi Han, Minkai Xu, Sheng Liu, Yitao Liang, Jianzhu Ma, James Y Zou, and Stefano Ermon. Tfg: Unified training-free guidance for diffusion models. *Advances in Neural Information Processing Systems*, 37:22370–22417, 2024. 2

[24] Xiner Li, Yulai Zhao, Chenyu Wang, Gabriele Scalia, Gokcen Eraslan, Surag Nair, Tommaso Biancalani, Shuiwang Ji, Aviv Regev, Sergey Levine, et al. Derivative-free guidance in continuous and discrete diffusion models with soft value-based decoding. *arXiv preprint arXiv:2408.08252*, 2024. 2, 3

[25] Anuj Singh, Sayak Mukherjee, Ahmad Beirami, and Hadi Jamali-Rad. Code: Blockwise control for denoising diffusion models. *arXiv preprint arXiv:2502.00968*, 2025. 2, 3, 5

[26] Luhuan Wu, Brian Trippe, Christian Naesseth, David Blei, and John P Cunningham. Practical and asymptotically exact conditional sampling in diffusion models. *Advances in Neural Information Processing Systems*, 36:31372–31403, 2023. 2

Authorized licensed use limited to: TU Delft Library. Downloaded on June 05,2026 at 07:53:48 UTC from IEEE Xplore. Restrictions apply.

[27] Jie Ren, Xidong Feng, Bo Liu, Xuehai Pan, Yao Fu, Luo Mai, and Yaodong Yang. Torchopt: An efficient library for differentiable optimization. *arXiv preprint arXiv:2211.06934*, 2022. 2

[28] Calvin Luo. Understanding diffusion models: A unified perspective. *arXiv preprint arXiv:2208.11970*, 2022. 3

[29] Ahmad Beirami, Alekh Agarwal, Jonathan Berant, Alex D'Amour, Jacob Eisenstein, Chirag Nagpal, and Ananda Theertha Suresh. Theoretical guarantees on the best-of-n alignment policy. 2024. *URL https://api. semanticscholar. org/CorpusID*, 266741736. 3

[30] Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, et al. Laion-5b: An open large-scale dataset for training next generation image-text models. *Advances in neural information processing systems*, 35:25278–25294, 2022. 5

[31] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017. 5

[32] Sadeep Jayasumana, Srikumar Ramalingam, Andreas Veit, Daniel Glasner, Ayan Chakrabarti, and Sanjiv Kumar. Rethinking fid: Towards a better evaluation metric for image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9307–9315, 2024. 5

[33] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2414–2423, 2016. 5

[34] Mao-Chuang Yeh, Shuai Tang, Anand Bhattad, Chuhang Zou, and David Forsyth. Improving style transfer with calibrated metrics. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3160–3168, 2020. 5

[35] Jack Hessel, Ari Holtzman, Maxwell Forbes, Ronan Le Bras, and Yejin Choi. Clipscore: A reference-free evaluation metric for image captioning. *arXiv preprint arXiv:2104.08718*, 2021. 5

[36] Christoph Schuhmann. Laion-aesthetics, 2025. Accessed: 2025-05-01. 5

[37] Yuval Kirstain, Adam Polyak, Uriel Singer, Shahbuland Matiana, Joe Penna, and Omer Levy. Pick-a-pic: An open dataset of user preferences for text-to-image generation. *Advances in Neural Information Processing Systems*, 36:36652–36663, 2023. 5, 6

[38] Ying Fan, Olivia Watkins, Yuqing Du, Hao Liu, Moonkyung Ryu, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, Kangwook Lee, and Kimin Lee. Dpok: Reinforcement learning for fine-tuning text-to-image diffusion models. *Advances in Neural Information Processing Systems*, 36:79858–79885, 2023. 5, 8

[39] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 5

[40] Xiaoshi Wu, Keqiang Sun, Feng Zhu, Rui Zhao, and Hongsheng Li. Human preference score: Better aligning text-to-image models with human preference. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2096–2105, 2023. 6

[41] Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. Sdedit: Guided image synthesis and editing with stochastic differential equations. *arXiv preprint arXiv:2108.01073*, 2021. 7

Authorized licensed use limited to: TU Delft Library. Downloaded on June 05,2026 at 07:53:48 UTC from IEEE Xplore. Restrictions apply.