NoiseTilt: Noise-Tilted Reverse Kernels for Diffusion Reward Alignment

Jisung Hwang[1,†], Yunhong Min[1], Jaihoon Kim[1], I-Chao Shen[2,‡], and Minhyuk Sung[1,‡]

[1] KAIST, Daejeon, Republic of Korea
[2] The University of Tokyo, Tokyo, Japan

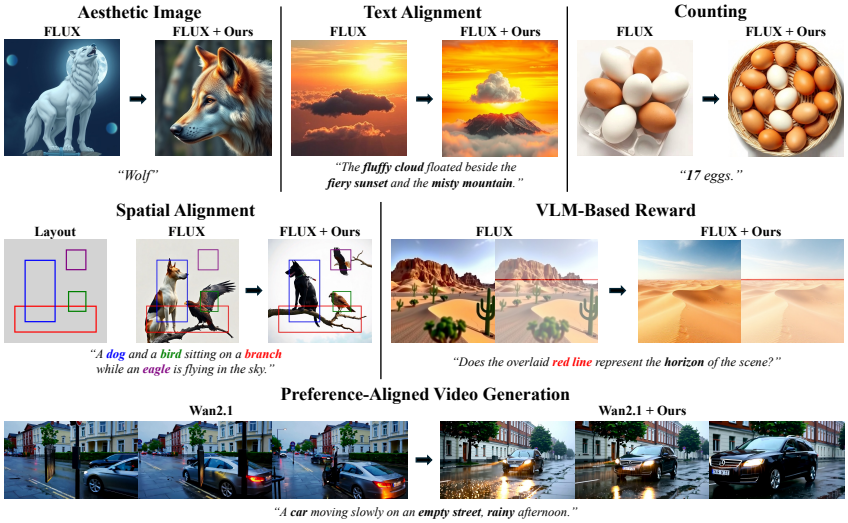

Fig. 1: Inference-time reward alignment results using our Noise-Tilted Reverse Kernels (NTRK) across diverse reward-guided diffusion applications.

Abstract. We introduce the Noise-Tilted Reverse Kernel (NTRK), a reward-guided diffusion sampler that injects reward gradients through the noise term, leaving the pretrained reverse kernel unchanged and requiring only a single sample per step. Reward-guided sampling at inference time has greatly expanded the versatility of pretrained diffusion models. Yet existing methods face a trade-off. Gradient-based guidance shifts the reverse mean, steering generation but pushing intermediate states outside the region that the model was trained on and degrading quality. Search-based methods preserve quality but gain no gradient signal. No prior method achieves both. NTRK resolves this by keeping the reverse mean fixed and biasing the noise term toward high reward. This is enabled by a whitening operator, the central mechanism behind NTRK, which converts reward gradients into noise-compatible perturbations without losing their guiding

[†] This work was done in part while the author was a visiting researcher at The University of Tokyo.
[‡] I-Chao Shen and Minhyuk Sung are co-corresponding authors.

signal. Across various reward alignment tasks, NTRK outperforms recent state-of-the-art baselines without losing sample quality. Remarkably, on aesthetic generation, NTRK surpasses the reward of the best baseline at 500 NFEs using only 25 NFEs, a 20× reduction in compute.

Keywords: Noise-Tilted Reverse Kernel · Whitening Operator · Reward Alignment · Inference-Time Scaling

1 Introduction

Inference-time scaling has become one of the most powerful levers for expanding the capabilities of pretrained generative models, enabling alignment to user- and task-specific preferences without costly retraining. In particular, diffusion models, which underpin state-of-the-art image and video generation [4, 33, 48, 64], are well suited for inference-time scaling thanks to their iterative denoising process. When preferences can be expressed as a reward function, this iterative procedure offers a model-agnostic mechanism to reshape the sampling distribution toward high-reward regions. As a result, reward-guided sampling at inference time has been applied broadly, including deblurring [5, 7, 11, 16, 51, 56, 68], super-resolution [7, 8, 56, 68], aesthetic image generation [28, 41, 69], spatial alignment [1, 69], text-image alignment [27, 28], and controllable video generation [25, 39, 40, 44, 71]. As illustrated in Figure 1, the same sampler can be applied across these diverse reward-guided generation settings.

Behind all these applications lies the same question: how should each denoising step be steered toward high-reward outcomes? Existing methods broadly fall into two paradigms, distinguished by how they modify or utilize the reverse Gaussian kernel. The first, **mean-shifted** gradient guidance methods, pioneered by DPS [7], augment the mean of the reverse Gaussian kernel with reward gradients, steering the trajectory toward high-reward regions. MCMC-based and particle-based extensions further build on this principle [1, 28, 57, 69, 70]. The second, **search-based** methods [27, 36, 41], draw K candidates from the stochastic reverse kernel and select among them by a reward criterion such as argmax or importance sampling, without gradients or kernel modification.

The two paradigms represent opposite ends of a fundamental trade-off between gradient guidance and noise compatibility. Mean-shifted methods gain the former but introduce a critical mismatch: adding reward gradients to the reverse mean displaces sampled states outside the noise-compatible regime, the narrow region where the pretrained model was trained to operate, with the displacement growing as the gradient magnitude increases. This can drive intermediate states out-of-distribution, degrade generation quality, and make optimization prone to reward hacking [15]. Search-based methods preserve the latter but forgo gradient guidance entirely, relying on random sampling to find reward-improving directions. To the best of our knowledge, **no prior method achieves both**, and our work addresses this gap.

To resolve this trade-off, we propose the *Noise-Tilted Reverse Kernel* (NTRK), a drop-in alternative that leaves the reverse-kernel mean entirely unchanged

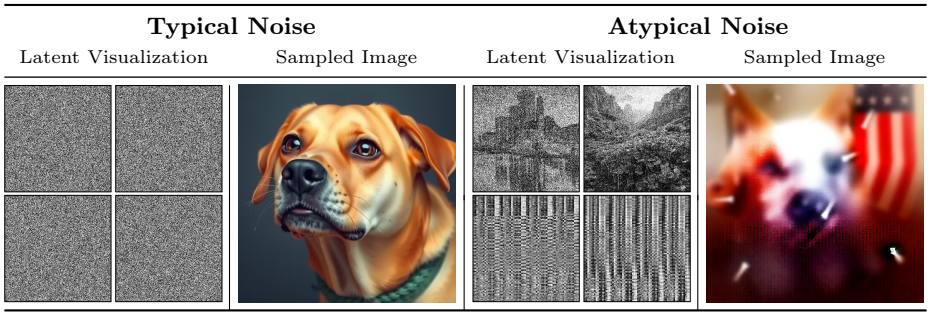

Fig. 2: Typical vs. atypical noise. We show noise vectors (latent visualizations) and the corresponding samples obtained when using them as injected noise. Typical noise produces high-quality outputs, whereas atypical noise induces artifacts.

and routes reward information through the noise term at each step. The central challenge is that any signal injected through the noise term must still be ***typical*** Gaussian noise [22,23], a multi-faceted property encompassing norm concentration, spatial uncorrelatedness, and collective Gaussian statistics. A raw reward gradient is usually highly structured, carrying spatially correlated patterns and localized features rather than the typical statistics of Gaussian noise; injecting it directly through the noise channel produces atypical noise and again drives states out-of-distribution, as illustrated in Figure 2. Inspired by whitening in statistical learning [24, 26], we introduce a *whitening operator* that processes the reward gradient before injection, suppressing structured components incompatible with the pretrained model while preserving the directional information of the gradient. Whitening is therefore not an auxiliary step but the core mechanism of `NTRK`. Accordingly, the effectiveness of reward alignment depends directly on the quality of this whitening operator.

In our experiments, we demonstrate the effectiveness of `NTRK`, which can also be combined with multi-particle strategies such as Best-of-N (BoN) [59]. Across aesthetic image generation, text-aligned image generation, VLM-based reward alignment, and preference-aligned video generation, `NTRK` paired with simple BoN outperforms recent state-of-the-art approaches that employ more complex techniques such as trajectory rollback [70], particle filtering [27, 28, 36], and initial-point search [69]. In particular, for aesthetic image generation, our method surpasses the reward achieved by the state-of-the-art baseline with 500 NFEs using only *25 NFEs* (i.e., 5%), while maintaining image quality.

2 Related Work

We review prior work on reward alignment for diffusion models at inference time. Fine-tuning-based approaches include direct backpropagation [9, 43] and reinforcement learning [3, 14, 45, 63]; we focus on inference-time methods that operate directly on pretrained models without additional training, and can further be applied on top of fine-tuned models [35,38]. We organize these methods into two threads: methods that modify or exploit the per-step reverse kernel, summarized

Table 1: Comparison of reverse-kernel guidance mechanisms. NTRK is the only approach satisfying all three properties simultaneously.

	Base	**Mean-Shifted**	**Search-Based**	**Hybrid**	**Noise-Tilted**
Examples	—	DPS [7]	SVDD [36]	DAS [28]	NTRK
Gradient guidance	✗	✓	✗	✓	✓
Pretrained-compatible	✓	✗	✓	✗	✓
Draws / step	1	1	K	K	1

in Tab. 1, and methods for making reward gradients noise-compatible, which underpin our whitening operator.

2.1 Reverse Kernel Methods

Rather than manipulating attention maps [6, 17, 18, 31, 42], which tend to be task- and model-specific, we focus on methods that modify the reverse kernel using a differentiable reward. A separate line that treats the injected noise as an optimization variable [13, 60] is covered below.

Mean-shifted reverse kernels. A foundational method in this line is DPS [7], which incorporates reward information directly into diffusion sampling by shifting the mean of each reverse kernel using gradients derived from the reward model via Tweedie's formula [47]. By leveraging the pretrained diffusion prior, this approach can be applied to diverse tasks and has become one of the most widely adopted frameworks for inference-time alignment. Numerous extensions have since been proposed to improve sampling efficiency, accuracy, and applicability: FreeDoM [70] incorporates additional Monte Carlo sampling, other works develop more advanced solvers [16,51,56,57,65], and the framework has been extended to latent diffusion models [49, 50] and applied across image and video generation [1, 10, 32, 34, 68]. As shown in Tab. 1, however, shifting the kernel mean pushes samples outside the noise-compatible Gaussian regime that pretrained models are trained on.

Search-based methods. Rather than modifying the kernel mean, search-based methods exploit the stochasticity of the reverse diffusion process by drawing K candidates from the unmodified base kernel and selecting those with higher rewards [41, 46, 55]. These methods trade additional computation for improved sample quality while strictly preserving the native reverse kernel. SVDD [36] is a representative example that selects the highest-reward sample at each denoising step; RBF [27] improves efficiency by dynamically allocating the sampling budget across timesteps. These methods remain noise-compatible but forgo gradient guidance and require K draws per step.

Hybrid methods. A third group combines gradient guidance with multi-particle sampling, gaining the directionality of gradient guidance and the diversity of multiple draws. DAS [28] incorporates reward gradients within a Sequential Monte

Carlo framework [12], and Ψ-Sampler [69] additionally reshapes the initial particle distribution toward the reward-aligned posterior. However, because these methods still shift the kernel mean to inject gradient guidance, they inherit the noise-compatibility issue of mean-shifted kernels. Across all three groups, no existing approach simultaneously achieves gradient guidance and noise compatibility.

2.2 Noise-Compatible Perturbations

Our work addresses this gap with `NTRK`, which routes reward information through the stochastic noise term rather than the mean and is the only approach satisfying all three properties in Tab. 1. The central challenge is that the injected noise must remain noise-compatible, a property more demanding than simply matching a Gaussian norm [22, 23]. Inspired by statistical whitening [24, 26], our whitening operator processes the reward gradient into a noise-compatible direction before injection.

Regularization-based methods. Several prior works have studied how to encourage Gaussian typicality in vectors involved in diffusion inference, each from a different context: earlier approaches apply norm-based regularization to keep optimized noise or latent vectors near the Gaussian concentration shell [2, 13, 52], while DNO [60], MPGR [22], and StressDream [54] additionally regularize higher-order spatial statistics alongside norm, in the respective contexts of noise-space optimization, gradient guidance stabilization, and video world model steering. All of these impose soft constraints for specific properties of Gaussian noise, however, and cannot guarantee those properties.

Projection-based methods. WGNC [23] is the first to reframe noise compatibility as a *projection* problem, directly mapping the reward gradient onto a white Gaussian noise feasible set defined by hard blockwise norm constraints in the Fourier domain. However, its hard equality constraints can distort noise that is already noise-compatible. Our whitening operator upgrades this projection approach: by replacing hard equality constraints with confidence-interval projections, it strongly suppresses atypical structure while leaving genuine Gaussian noise nearly unchanged; a detailed comparison is provided in the **supplementary document**.

3 Overview

We formalize the reward alignment problem and analyze design choices for the per-step reverse kernel. The central question is how reward information can be incorporated at each denoising step without violating the noise-compatible regime that the pretrained model relies on.

3.1 Problem Definition

Given a pretrained diffusion model that maps a source noise distribution, $p_T = \mathcal{N}(\mathbf{0}, \boldsymbol{I})$, to a data distribution p_0, our objective is to generate high-reward

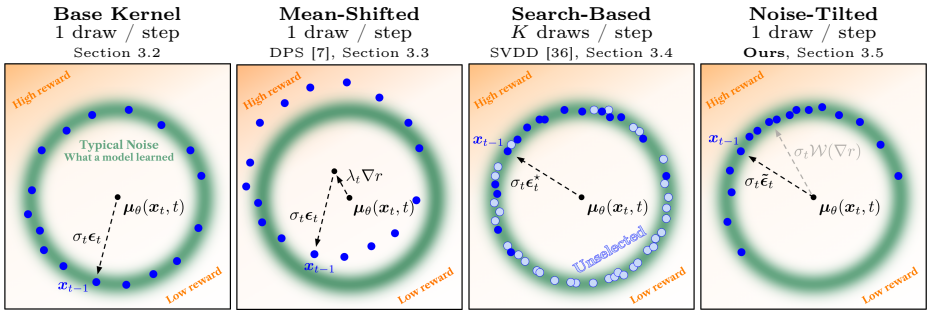

Fig. 3: Reverse-kernel guidance mechanisms. The green annulus indicates the noise-compatible regime; blue dots show the induced sample distribution. Mean-shifted guidance injects reward information by shifting the reverse mean, pushing samples outside the noise-compatible regime. Search-based guidance preserves the base mean by selecting the best among K candidate noise draws. `NTRK` also preserves the base mean but constructs a single reward-tilted noise draw, achieving reward alignment without leaving the noise-compatible regime.

samples $\boldsymbol{x}_0$, a task generally known as reward alignment. Formally, this objective is formulated as finding a target distribution p_0^* [30, 61, 62] such that:

$$p_0^* = \arg\max_q \; \mathbb{E}_{\boldsymbol{x}_0 \sim q}\left[r(\boldsymbol{x}_0)\right] - \beta\, \mathcal{D}_{\mathrm{KL}}\left[q \| p_0\right], \tag{1}$$

which maximizes the expected reward $r(\boldsymbol{x}_0)$ while the KL divergence prevents deviation from the pretrained data distribution. The temperature $\beta > 0$ controls this trade-off: smaller β produces stronger reward tilting, while larger β stays closer to the pretrained distribution.

The optimal reverse kernel $p_\theta^*(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t)$ required to sample from this target distribution can be approximated as follows:

$$p_\theta^*(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t) \propto p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t) \exp\left(\frac{V_{t-1}(\boldsymbol{x}_{t-1})}{\beta}\right), \tag{2}$$

where the exact value function $V_t(\boldsymbol{x}_t)$ is intractable and is approximated using Tweedie's posterior mean [47]: $V_t(\boldsymbol{x}_t) \approx r(\hat{\boldsymbol{x}}_{0|t})$, where $\hat{\boldsymbol{x}}_{0|t} := \mathbb{E}[\boldsymbol{x}_0 \mid \boldsymbol{x}_t]$. The design of the per-step reverse kernel determines how this approximation is realized in practice, and whether the resulting updates remain noise-compatible.

3.2 Diffusion Reverse Kernel

A diffusion model [19, 58] progressively denoises an initial Gaussian noise by sequentially applying a Markovian reverse kernel, modeled as a Gaussian transition from $\boldsymbol{x}_t$ to $\boldsymbol{x}_{t-1}$:

$$p_\theta(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t) = \mathcal{N}\big(\boldsymbol{x}_{t-1}; \boldsymbol{\mu}_\theta(\boldsymbol{x}_t, t), \sigma_t^2 \boldsymbol{I}\big). \tag{3}$$

To generate a sample in practice, $\boldsymbol{x}_{t-1}$ is drawn using the standard reparameterization trick:

$$\boldsymbol{x}_{t-1} = \boldsymbol{\mu}_\theta(\boldsymbol{x}_t, t) + \sigma_t \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}). \tag{4}$$

Since the reverse kernel is Gaussian, any valid transition between consecutive latent states should have a standardized perturbation that behaves as standard Gaussian noise:

$$\boldsymbol{\eta}_t := \frac{\boldsymbol{x}_{t-1} - \boldsymbol{\mu}_\theta(\boldsymbol{x}_t, t)}{\sigma_t}. \tag{5}$$

Under the base reverse process, $\boldsymbol{\eta}_t = \boldsymbol{\epsilon}_t$ holds trivially, so the standardized perturbation is standard Gaussian noise.

3.3 Mean-Shifted Reverse Kernel [7]

For efficient reward alignment, first-order reward information can be injected directly into the reverse transition by adding a reward gradient to the mean:

$$\boldsymbol{x}_{t-1} = \boldsymbol{\mu}_\theta(\boldsymbol{x}_t, t) + \lambda_t \nabla_{\boldsymbol{x}_t} r(\hat{\boldsymbol{x}}_{0|t}) + \sigma_t \boldsymbol{\epsilon}_t, \tag{6}$$

where $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$, $\lambda_t > 0$ is a guidance hyperparameter, and detailed derivations are provided in the **supplementary document**. As illustrated in Figure 3, this injects reward guidance by shifting the reverse-kernel mean while keeping the injected noise term unchanged.

Despite its practical success, a closer look at the standardized perturbation reveals a mismatch with the base reverse dynamics:

$$\boldsymbol{\eta}_t^{\text{mean}} = \boldsymbol{\epsilon}_t + \frac{\lambda_t}{\sigma_t} \nabla_{\boldsymbol{x}_t} r(\hat{\boldsymbol{x}}_{0|t}). \tag{7}$$

The added reward-gradient term in Equation (7) is generally structured, so the standardized perturbation is no longer typical Gaussian noise assumed by the pretrained reverse kernel. Consequently, repeated reverse updates can drift away from the learned intermediate distributions.

3.4 Search-Based Reverse Kernel

To approximate the optimal reverse transition in Equation (2), search-based methods draw K candidates from the base reverse kernel, all centered at the pretrained mean:

$$\boldsymbol{x}_{t-1}^{(i)} = \boldsymbol{\mu}_\theta(\boldsymbol{x}_t, t) + \sigma_t \boldsymbol{\epsilon}_t^{(i)}, \quad \boldsymbol{\epsilon}_t^{(i)} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}), \tag{8}$$

where $i = 1, \ldots, K$. As a representative search-based method, SVDD [36] selects the candidate whose predicted reward score is highest:

$$i^\star = \underset{i \in \{1,\ldots,K\}}{\arg\max}\ r\left(\hat{\boldsymbol{x}}_{0|t-1}^{(i)}\right), \quad \boldsymbol{x}_{t-1} = \boldsymbol{x}_{t-1}^{(i^\star)}. \tag{9}$$

Reward information therefore appears through the selected stochastic perturbation $\boldsymbol{\epsilon}_t^{(i^\star)}$ rather than through a deterministic mean shift. Search-based guidance can thus be interpreted as an implicit form of noise tilting. However, when high-reward samples lie in low-density regions of the pretrained distribution, this strategy becomes inefficient because it requires a prohibitively large number of samples. The goal of `NTRK` is to retain this mean-preserving view of search-based guidance while replacing exhaustive sampling with a single, gradient-informed noise draw.

3.5 Noise-Tilted Reverse Kernel (NTRK)

Our key idea is to preserve the pretrained reverse mean $\boldsymbol{\mu}_\theta(\boldsymbol{x}_t, t)$ and inject reward information only through the noise, thereby avoiding the noise-compatibility mismatch of mean-shifted kernels. Unlike search-based selection, which requires drawing K candidates per step, our approach achieves this with a single noise draw.

The remaining challenge is that a raw reward gradient is not typical Gaussian noise: it often contains spatially correlated and localized structures induced by the reward model and the current sample. Thus, simply moving the gradient from the mean term to the noise term is not sufficient; the reward-gradient signal must first be converted into a noise-compatible form. This requirement is precisely the role of our whitening operator $\mathcal{W}$, which maps the reward gradient to a *whitened reward direction*; we define $\boldsymbol{w}_t := \mathcal{W}(\nabla_{\boldsymbol{x}_t} r(\hat{\boldsymbol{x}}_{0|t}))$.

To incorporate $\boldsymbol{w}_t$ into the noise term while preserving the form of typical Gaussian noise, we use a basic Gaussian mixing identity as a design principle: if $\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ are independent, then for any $\rho \in [0, 1]$,

$$\sqrt{\rho}\,\boldsymbol{\epsilon}_1 + \sqrt{1-\rho}\,\boldsymbol{\epsilon}_2 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}). \tag{10}$$

Since whitening makes $\boldsymbol{w}_t$ noise-compatible, we define the guided injected noise by mixing this whitened reward direction with unbiased Gaussian noise:

$$\tilde{\boldsymbol{\epsilon}}_t = \sqrt{\rho_t}\,\boldsymbol{w}_t + \sqrt{1-\rho_t}\,\boldsymbol{\epsilon}_t, \tag{11}$$

where $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ and $\rho_t \in [0, 1]$ controls the guidance strength. The first term injects reward-gradient information through the whitened reward direction, while the second term preserves unbiased Gaussian noise. Using Equation (11), we replace the base sampling rule in Equation (4), resulting in our noise-tilted reverse kernel:

$$\boldsymbol{x}_{t-1} = \boldsymbol{\mu}_\theta(\boldsymbol{x}_t, t) + \sigma_t \tilde{\boldsymbol{\epsilon}}_t. \tag{12}$$

Most importantly, the standardized perturbation relative to the base mean becomes

$$\boldsymbol{\eta}_t^{\text{noise}} := \frac{\boldsymbol{x}_{t-1} - \boldsymbol{\mu}_\theta(\boldsymbol{x}_t, t)}{\sigma_t} = \tilde{\boldsymbol{\epsilon}}_t. \tag{13}$$

The kernel is therefore *noise-tilted*: reward guidance is expressed through the noise term rather than the mean, while the pretrained reverse-kernel structure is preserved. Specifically, Equation (12) uses the same base mean $\boldsymbol{\mu}_\theta(\boldsymbol{x}_t, t)$ and the same nominal noise scale σ_t as the pretrained reverse kernel in Equation (3). This construction hinges on whitening: without it, the reward signal would still enter the sampler as structured, atypical noise even though it appears in the noise channel. In Sec. 4, we formalize the whitening operator $\mathcal{W}$ that makes noise-tilted guidance practical in high dimensions.

4 Whitening Operator

Our goal is to transform the reward-gradient signal used by NTRK into *typical* standard Gaussian noise. Although $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$ has nonzero density everywhere in $\mathbb{R}^N$,

in high dimensions almost all probability mass concentrates on a narrow *typical* region. Pretrained generative models are trained on trajectories whose injected noise lies in this typical region; consequently, injected noise that drifts away from it can act as an out-of-distribution input to the learned reverse dynamics. (See Figure 2.) We therefore design a whitening operator $\mathcal{W} : \mathbb{R}^N \to \mathbb{R}^N$ that moves the reward gradient toward typical standard Gaussian noise while preserving its guiding direction for the noise-tilted kernel.

The typical set is difficult to characterize exactly in closed form. Instead, we approximate it using a collection of *high-confidence* constraints induced by known statistics of the standard normal distribution. Concretely, we precompute 99.99% confidence bounds and define $\mathcal{W}$ as a sequence of projections onto the corresponding confidence sets. A key building block is a two-level order-statistic projection (**2OS**), which we describe next.

Two-level order statistics. Let $\boldsymbol{x} \in \mathbb{R}^N$ be the vector to whiten and reshape it into a tile matrix $\boldsymbol{Y} \in \mathbb{R}^{M \times D}$ (so $N = MD$). The 2OS statistic is obtained by sorting *within* each tile and then sorting *across* tiles at each within-tile rank:

$$\boldsymbol{Z} = \mathrm{sort}_0\big(\mathrm{sort}_1(\boldsymbol{Y})\big), \tag{14}$$

where sort_1 sorts each row of $\boldsymbol{Y}$ and sort_0 sorts each column. Intuitively, $\boldsymbol{Z}$ captures a "rank-of-rank" summary: $Z_{r,j}$ is the r-th smallest value among the j-th order statistics collected from all tiles. For standard Gaussian noise, each entry $Z_{r,j}$ concentrates sharply, allowing tight confidence bounds for each (r, j).

Confidence bounds for the 2OS statistic. Fix a small probability level α (we use $\alpha = 10^{-4}$) and let Φ denote the cumulative distribution function (CDF) of the standard normal distribution. For each tile rank $r \in \{1, \dots, M\}$ and within-tile rank $j \in \{1, \dots, D\}$, we compute the nested quantile bounds:

$$L_{r,j} = \Phi^{-1}\Big(\mathrm{Beta}^{-1}\Big(\mathrm{Beta}^{-1}\big(\tfrac{\alpha}{2}; r, M - r + 1\big);\, j, D - j + 1\Big)\Big), \tag{15}$$

$$U_{r,j} = \Phi^{-1}\Big(\mathrm{Beta}^{-1}\Big(\mathrm{Beta}^{-1}\big(1 - \tfrac{\alpha}{2}; r, M - r + 1\big);\, j, D - j + 1\Big)\Big). \tag{16}$$

These bounds specify a $1 - \alpha$ confidence set for the doubly-sorted matrix $\boldsymbol{Z}$:

$$\mathcal{C}_{2\mathrm{os}} = \Big\{\boldsymbol{Y} \in \mathbb{R}^{M \times D} : L_{r,j} \le \big(\mathrm{sort}_0(\mathrm{sort}_1(\boldsymbol{Y}))\big)_{r,j} \le U_{r,j}\,,\ \forall r, j\Big\}. \tag{17}$$

2OS projection via sort–clip–unsort. Given $\boldsymbol{Y}$, we compute $\boldsymbol{Z}$ as in Equation (14), clip each element to its confidence interval

$$Z_{r,j} \leftarrow \min\{U_{r,j}\,, \max\{L_{r,j}\,, Z_{r,j}\}\}, \tag{18}$$

and then invert the two sorting permutations to map the clipped $\boldsymbol{Z}$ back to the original tile layout. We prove in the **supplementary document** that this

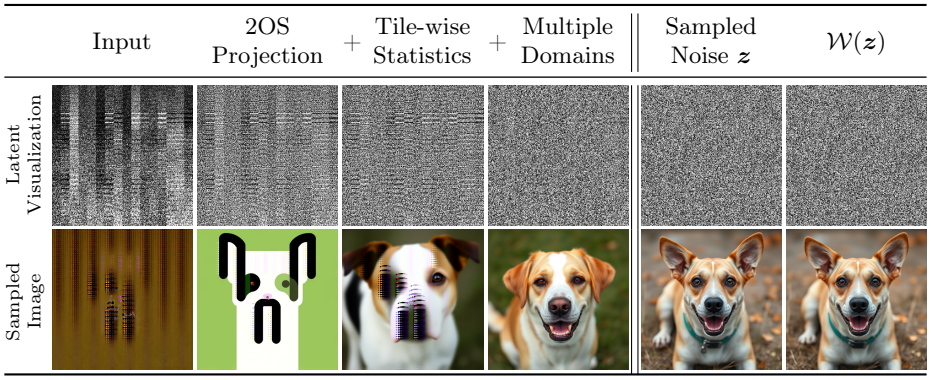

Fig. 4: Effect of $\mathcal{W}$. As we apply the components of $\mathcal{W}$ to a structured latent (left to right), it becomes more similar to typical Gaussian noise with more realistic samples. For typical Gaussian noise, $\mathcal{W}(\boldsymbol{z})$ changes negligibly (two rightmost columns).

sort–clip–unsort operation equals the Euclidean projection onto $\mathcal{C}_{\text{2os}}$, and we therefore call it the *2OS projection*.

The 2OS projection prevents extreme values from concentrating in a few tiles. Because clipping is applied to each rank pair (r, j) of the doubly-sorted statistic, each tile contains a balanced spread of small-to-large values consistent with typical standard Gaussian noise. This is the core mechanism by which the operator suppresses localized structure and pushes the input toward typical Gaussian noise.

The same principle is applied beyond raw vector values. We also apply 2OS projections to tile-wise mean and centered-energy statistics, which prevents a small subset of spatial blocks from carrying unusually large bias or energy. In addition, we repeat the projection pipeline in several orthogonally transformed domains, so that structured correlations that are weakly visible in the value domain can still be exposed and suppressed after a change of basis. These components are complementary: value-level 2OS controls the marginal distribution within and across tiles, moment-level 2OS controls blockwise statistics, and transformed-domain projections capture nonlocal correlation patterns.

The full whitening operator $\mathcal{W}$ is composed of 2OS projections applied to various tile-wise statistics across multiple transformed domains, as described in **supplementary document**. We visualize this progression in Figure 4. Since all constraints are derived from 99.99% confidence intervals under the reference distribution, typical standard Gaussian noise passes through with negligible modification in practice (cosine similarity > 0.99999; see the two rightmost columns in Figure 4), whereas structured inputs are substantially whitened into noise-compatible directions.

5 Experiments

In this section, we present experimental results demonstrating the effectiveness of NTRK compared to prior baselines, with the experimental setup described in Section 5.1. We first evaluate aesthetic image generation and text-aligned image generation in Section 5.2 and Section 5.3, respectively. We then extend the evaluation to preference-aligned video generation in Section 5.4. Finally, we show that our method can also be applied on top of fine-tuned models in Section 5.5. In the **supplementary document**, we provide (i) additional applications, including counting tasks and VLM-based reward alignment, and (ii) alignment results using a different diffusion model [4].

5.1 Experiment Setup

Tasks. We evaluate NTRK across three reward-guided generation settings: aesthetic image generation, text-aligned image generation, and preference-aligned video generation. For aesthetic image generation, we use 45 animal prompts from previous work, DDPO [3]. For text-aligned image generation, we use 100 prompts in complex category of T2I-CompBench++ [20]. For preference-aligned video generation, we use 200 prompts from VBench [21] animal and scenery categories.

For image and video generation applications, we use FLUX [33] and Wan2.1 [64] as the base flow models, respectively. In all experiments, we fix the sampling steps to 25. As a reference, we also include the results of the base models without any guidance method applied.

Baselines. We compare NTRK against a range of inference-time reward-alignment algorithms discussed in Sec. 2, including both gradient-based guidance methods and search-based approaches. Specifically, we consider DPS [7] and FreeDoM [70] as single-particle gradient-based methods, DAS [28] and Ψ-Sampler [69] as multi-particle gradient-based methods, and SVDD [36], RBF [27], and BoN [59] as search-based methods.

Note that multi-particle and search methods utilize multiple samples during sampling, whereas single-particle methods produce only a single trajectory. For fair comparison, we fix the total number of function evaluations (NFE) across all methods. In particular, for single-particle sampling methods such as DPS and FreeDoM, we augment sampling with Best-of-N (BoN) [59], which runs multiple independent sampling processes and selects the highest-reward output, ensuring that the overall NFE is comparable across methods. [3] For all quantitative results, methods augmented with BoN are marked with †. For NTRK, we report results both with and without BoN to isolate the effect of the proposed method.

5.2 Aesthetic Image Generation

Evaluation Metrics. In this work, we refer to the reward used for inference-time optimization as the *target reward*, and to rewards not observed during optimization

[3] FreeDoM [70] incorporates additional MCMC sampling, which increases the computational cost.

Table 2: Quantitative comparison on image reward alignment. Left: aesthetic image generation (target: Aesthetic Score [53]). Right: text-aligned image generation (target: PickScore [29]). For single-particle methods we augment sampling with Best-of-N to match the total NFE, denoted with †. Dark green cells indicate the best result for each metric, light green the second best.

Method	NFE	Aesthetic Image Generation					Text-Aligned Image Generation				
		Target Reward	Held-Out Reward				Target Reward	Held-Out Reward			
		Aesthetic Score ↑	Pick-Score ↑	HPSv2 ↑	Image Reward ↑	VQA Score ↑	Pick-Score ↑	Aesthetic Score ↑	HPSv2 ↑	Image Reward ↑	VQA Score ↑
Base [33]	25	6.0282	0.2144	0.2759	1.0538	0.9644	0.2054	5.4664	0.2316	0.1710	0.8011
BoN [59]	500	6.7310	0.2197	0.2890	1.1419	0.9597	0.2146	5.8582	0.2619	0.6883	0.8021
DPS† [7]	500	6.7647	0.2191	0.2861	1.0639	0.9624	0.2147	5.8073	0.2622	0.6310	0.8028
FreeDoM† [70]	533	6.8406	0.2185	0.2853	0.9941	0.9635	0.2133	5.8492	0.2572	0.5354	0.7990
SVDD [36]	500	7.1363	0.2177	0.2814	1.0256	0.9510	0.2204	5.8743	0.2699	0.7592	0.8201
RBF [27]	500	6.9900	0.2183	0.2826	1.0761	0.9689	0.2202	5.8618	0.2682	0.7583	0.8149
DAS [28]	500	6.9384	0.2183	0.2860	1.0568	0.9706	0.2139	5.8385	0.2568	0.5226	0.7990
Ψ-Sampler [69]	500	7.0116	0.2188	0.2847	1.1235	0.9737	0.2120	5.7329	0.2551	0.4590	0.8145
NTRK **(Ours)**	25	7.4510	0.2200	0.2928	1.2565	0.9728	0.2224	5.7720	0.2601	0.5257	0.8210
NTRK† **(Ours)**	500	7.9656	0.2197	0.2932	1.1669	0.9609	0.2327	5.9020	0.2817	0.7370	0.8017

as *held-out rewards*. In this task, the target reward is Aesthetic Score [53]. As held-out rewards, we evaluate image quality using ImageReward [67] and HPSv2 [66], and text-image alignment using PickScore [29] and VQA Score [37].

Results. The quantitative and qualitative results are presented in Tab. 2 and Figure 5, respectively. Overall, NTRK achieves the best target reward performance across all methods, and even outperforms all baselines with only 1/20 of NFE. On held-out rewards, NTRK also delivers the best image quality for both ImageReward and HPSv2, while remaining comparable on text-image alignment metrics. The qualitative examples in Figure 5 further support these trends. Across different prompts, NTRK produces more visually appealing samples than the baselines, achieving the highest rewards [53].

5.3 Text-Aligned Image Generation

Evaluation Metrics. The target reward used to align text-image is PickScore [29]. For held-out rewards, we evaluate text-image alignment using VQA Score [37], and image quality using Aesthetic Score [53], ImageReward [67], and HPSv2 [66].

Results. The quantitative and qualitative results are presented in Tab. 2 and Figure 5, respectively. As in the aesthetic image generation task, NTRK achieves the best target reward performance across all methods, with the same efficiency trend that the 25-NFE setting already outperforms all baselines. These gains also transfer to held-out rewards: NTRK with 500 NFE achieves the best Aesthetic Score and HPSv2, while NTRK with 25 NFE attains the best VQA Score. Qualitatively, NTRK produces samples that better align with the text prompts such as spatial and logical relations than the baselines.

5.4 Preference-Aligned Video Generation

As done in the image generation task, we use 25 sampling steps for all methods except FreeDoM [70], for which we use 13 steps due to its additional MCMC

Fig. 5: Qualitative comparison on image reward alignment. Top: aesthetic image generation (target: Aesthetic Score [53]). Bottom: text-aligned image generation (target: PickScore [29]). Scores shown in italics below each image. For single-particle methods we augment sampling with Best-of-N to match the total NFE, denoted with †.

sampling, ensuring that the total NFE remains comparable. For video generation task, we test the baselines and our method with a single particle.

Evaluation Metrics. In this task, the target reward is VideoReward [39], which provides three component scores for Motion Quality (MQ), Visual Quality (VQ), and Text Alignment (TA). We use the sum of these components (MQ + VQ + TA) as the target reward. For held-out evaluation, we report metrics from VBench [21], grouped into six categories: Subject Consistency and Background Consistency for text alignment, Motion Smoothness and Dynamic Degree for motion quality, and Aesthetic Quality and Imaging Quality for visual quality.

Table 3: Quantitative comparison on preference-aligned video generation. The target reward is VideoReward [39], and the held out rewards are the metrics proposed in VBench [21]. Dark green cells indicate the best result for each metric across all runs, while light green cells denote the second best.

Method	NFE	Target Reward	Held-Out Reward					
			Motion Quality		Visual Quality		Text Alignment	
		VideoReward ↑	Smooth.↑	Dynamic↑	Aesthetic↑	Imaging.↑	Subject↑	Back.↑
Base [64]	25	-0.399	0.9629	0.9300	0.6104	0.6791	0.9589	0.9647
DPS [7]	25	-0.130	0.9621	0.9350	0.5867	0.6738	0.9398	0.9473
FreeDoM [70]	25	-0.211	0.9632	0.6900	0.5880	0.6897	0.9586	0.9570
NTRK **(Ours)**	25	3.465	0.9630	0.9500	0.6120	0.6870	0.9591	0.9646

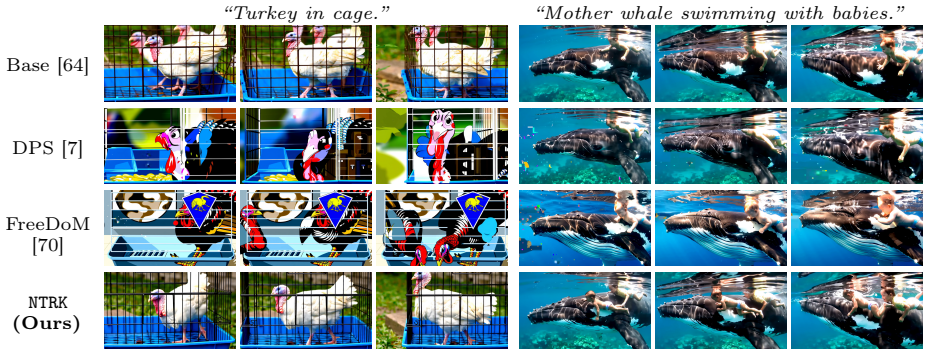

Fig. 6: Qualitative comparison on preference-aligned video generation using VideoReward [39]. NTRK produces videos with better text alignment and visual quality, where we present video results in the **supplementary document**.

Results. Table 3 and Figure 6 summarize the quantitative and qualitative comparisons, respectively. NTRK yields the highest VideoReward score and outperforms all the baselines. Beyond the target reward, NTRK achieves the best results on Dynamic Degree, Aesthetic Quality, and Subject Consistency, and remains marginally runner-up on the other held-out metrics. The qualitative examples further support this observation, with NTRK producing videos that better align with the text prompts. In particular, NTRK generates videos that clearly captures the elements (*e.g.*, turkey and babies) described in the prompt.

5.5 Integration with Fine-Tuned Models

Fine-tuning and inference-time alignment improve reward alignment along orthogonal directions: the former adapts model parameters, while the latter guides the sampling process without modifying them. Because of this orthogonality, NTRK can be applied on top of a fine-tuned model whenever the goal is to maximize a specific reward beyond what fine-tuning alone provides. In this section, we integrate NTRK with MixGRPO [35], which fine-tunes the base FLUX model [33]. The target reward is PickScore [29], and we report the same held-out rewards used in the text-aligned image generation task described in Section 5.3.

Table 4: Quantitative comparison on text-aligned image generation with fine-tuned model. We compare DPS [7] and NTRK integrated with a fine-tuned model, MixGRPO [35]. For single-particle methods we augment sampling with Best-of-N, denoted with †. Dark green cells indicate the best result for each metric across all runs, while light green cells denote the second best.

Method	NFE	Target Reward	Held-Out Reward			
		PickScore ↑	Aesthetic Score ↑	HPSv2 ↑	Image Reward ↑	VQA Score ↑
Base [33]	25	0.2054	5.4664	0.2316	0.1710	0.8011
MixGRPO [35]	25	0.2166	6.5245	0.2679	0.7605	0.8239
⌞ DPS† [7]	500	0.2235	6.6966	0.2840	1.0501	0.8376
⌞ NTRK† **(Ours)**	500	0.2545	6.7281	0.3224	1.2648	0.8498

Results. As shown in Tab. 4, MixGRPO [35] improves both the target reward and all held-out rewards compared to the base model [33]. We further observe that applying inference-time reward alignment remains highly effective on top of the fine-tuned model, yielding additional improvements across all metrics. In particular, NTRK consistently outperforms DPS on both the target reward and all held-out rewards, achieving the highest overall scores. These results confirm that NTRK complements fine-tuning: the two can be composed to push reward alignment beyond what either approach achieves alone.

6 Conclusion

In this work, we identify a practical limitation of widely used inference-time reward alignment techniques that steer the reverse diffusion kernel by directly shifting its mean using the reward gradient. To address this limitation, we propose NTRK, which keeps the reverse mean fixed and injects reward guidance through a whitened, noise-compatible direction in the noise term. Experiments demonstrate that our method consistently achieves stronger reward alignment while maintaining sample quality.

Acknowledgments

We thank Takeo Igarashi for valuable discussions. I-Chao Shen acknowledges support from the UTokyo-Google AI Symbiotic Future Society Program. This work was also supported by the NRF grant (RS-2026-25486000); IITP grants (RS-2024-00399817, RS-2025-25441313, RS-2025-25443318, and RS-2026-25526850), funded by MSIT, Korea; the Industrial Technology Innovation Program grant (RS-2025-02317326), funded by MOTIE, Korea; the InnoCORE program of MSIT (AI Meta-Scientist, N10260110); the National Supercomputing Center with supercomputing resources and technical support (KSC-2025-CRE-0475); the Advanced GPU Utilization Support Program; and the DRB-KAIST SketchTheFuture Research Center.

References

1. Bansal, A., Chu, H.M., Schwarzschild, A., Sengupta, S., Goldblum, M., Geiping, J., Goldstein, T.: Universal guidance for diffusion models. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (2023)
2. Ben-Hamu, H., Puny, O., Gat, I., Karrer, B., Singer, U., Lipman, Y.: D-Flow: Differentiating through flows for controlled generation. In: International Conference on Machine Learning. pp. 3462–3483 (2024)
3. Black, K., Janner, M., Du, Y., Kostrikov, I., Levine, S.: Training diffusion models with reinforcement learning. In: International Conference on Learning Representations (2024)
4. Cai, H., Cao, S., Du, R., Gao, P., Hoi, S., Hou, Z., Huang, S., Jiang, D., Jin, X., Li, L., et al.: Z-image: An efficient image generation foundation model with single-stream diffusion transformer. arXiv preprint arXiv:2511.22699 (2025)
5. Cardoso, G., Idrissi, Y.J.E., Corff, S.L., Moulines, E.: Monte carlo guided diffusion for bayesian linear inverse problems. In: International Conference on Learning Representations (2024)
6. Chefer, H., Alaluf, Y., Vinker, Y., Wolf, L., Cohen-Or, D.: Attend-and-excite: Attention-based semantic guidance for text-to-image diffusion models. ACM Transactions on Graphics **42**(4), 148:1–148:12 (2023). `https://doi.org/10.1145/3592116`
7. Chung, H., Kim, J., Mccann, M.T., Klasky, M.L., Ye, J.C.: Diffusion posterior sampling for general noisy inverse problems. In: International Conference on Learning Representations (2023)
8. Chung, H., Sim, B., Ryu, D., Ye, J.C.: Improving diffusion models for inverse problems using manifold constraints. In: Advances in Neural Information Processing Systems. vol. 35, pp. 25683–25696 (2022)
9. Clark, K., Vicol, P., Swersky, K., J, F.D.: Directly fine-tuning diffusion models on differentiable rewards. In: International Conference on Learning Representations (2024)
10. Daras, G., Nie, W., Kreis, K., Dimakis, A., Mardani, M., Kovachki, N., Vahdat, A.: Warped diffusion: Solving video inverse problems with image diffusion models. In: Advances in Neural Information Processing Systems. vol. 37, pp. 101116–101143 (2024)
11. Dou, Z., Song, Y.: Diffusion posterior sampling for linear inverse problem solving: A filtering perspective. In: International Conference on Learning Representations (2024)
12. Doucet, A., De Freitas, N., Gordon, N.J., et al.: Sequential Monte Carlo methods in practice. Springer (2001)
13. Eyring, L., Karthik, S., Roth, K., Dosovitskiy, A., Akata, Z.: ReNO: Enhancing one-step text-to-image models through reward-based noise optimization. In: Advances in Neural Information Processing Systems. vol. 37, pp. 125487–125519 (2024)
14. Fan, Y., Watkins, O., Du, Y., Liu, H., Ryu, M., Boutilier, C., Abbeel, P., Ghavamzadeh, M., Lee, K., Lee, K.: DPOK: reinforcement learning for fine-tuning text-to-image diffusion models. In: Advances in Neural Information Processing Systems. vol. 36, pp. 79858–79885 (2023)
15. Gao, L., Schulman, J., Hilton, J.: Scaling laws for reward model overoptimization. In: International Conference on Machine Learning. pp. 10835–10866 (2023)
16. He, Y., Murata, N., Lai, C.H., Takida, Y., Uesaka, T., Kim, D., Liao, W.H., Mitsufuji, Y., Kolter, J.Z., Salakhutdinov, R., Ermon, S.: Manifold preserving guided diffusion. In: International Conference on Learning Representations (2024)

17. Hertz, A., Mokady, R., Tenenbaum, J., Aberman, K., Pritch, Y., Cohen-Or, D.: Prompt-to-prompt image editing with cross attention control. In: International Conference on Learning Representations (2023)
18. Hertz, A., Voynov, A., Fruchter, S., Cohen-Or, D.: Style aligned image generation via shared attention. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4775–4785 (2024)
19. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems. vol. 33, pp. 6840–6851 (2020)
20. Huang, K., Duan, C., Sun, K., Xie, E., Li, Z., Liu, X.: T2I-CompBench++: An enhanced and comprehensive benchmark for compositional text-to-image generation. IEEE Transactions on Pattern Analysis and Machine Intelligence **47**(5), 3563–3579 (2025). https://doi.org/10.1109/TPAMI.2025.3531907
21. Huang, Z., He, Y., Yu, J., Zhang, F., Si, C., Jiang, Y., Zhang, Y., Wu, T., Jin, Q., Chanpaisit, N., Wang, Y., Chen, X., Wang, L., Lin, D., Qiao, Y., Liu, Z.: VBench: Comprehensive benchmark suite for video generative models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 21807–21818 (2024)
22. Hwang, J., Kim, J., Sung, M.: Moment- and power-spectrum-based Gaussianity regularization for text-to-image models. In: Advances in Neural Information Processing Systems. vol. 38, pp. 18235–18264 (2025)
23. Hwang, J., Sung, M.: Gradient preconditioning for efficient and reliable reward-guided generation. arXiv preprint arXiv:2602.08646 (2026)
24. Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. John Wiley & Sons (2001)
25. Jang, S., Ki, T., Jo, J., Yoon, J., Kim, S.Y., Lin, Z., Hwang, S.J.: Frame guidance: Training-free guidance for frame-level control in video diffusion models. In: International Conference on Learning Representations (2026)
26. Kessy, A., Lewin, A., Strimmer, K.: Optimal whitening and decorrelation. The American Statistician **72**(4), 309–314 (2018). https://doi.org/10.1080/00031305.2016.1277159
27. Kim, J., Yoon, T., Hwang, J., Sung, M.: Inference-time scaling for flow models via stochastic generation and rollover budget forcing. In: Advances in Neural Information Processing Systems. vol. 38, pp. 30830–30864 (2025)
28. Kim, S., Kim, M., Park, D.: Test-time alignment of diffusion models without reward over-optimization. In: International Conference on Learning Representations (2025)
29. Kirstain, Y., Polyak, A., Singer, U., Matiana, S., Penna, J., Levy, O.: Pick-a-Pic: An open dataset of user preferences for text-to-image generation. In: Advances in Neural Information Processing Systems. vol. 36, pp. 36652–36663 (2023)
30. Korbak, T., Elsahar, H., Kruszewski, G., Dymetmant, M.: On reinforcement learning and distribution matching for fine-tuning language models with no catastrophic forgetting. In: Advances in Neural Information Processing Systems. vol. 35, pp. 16203–16220 (2022)
31. Kumari, N., Zhang, B., Zhang, R., Shechtman, E., Zhu, J.Y.: Multi-concept customization of text-to-image diffusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1931–1941 (2023)
32. Kwon, T., Ye, J.C.: Solving video inverse problems using image diffusion models. In: International Conference on Learning Representations (2025)
33. Labs, B.F.: FLUX. https://github.com/black-forest-labs/flux (2024), accessed 26 June 2026 (AoE)

34. Lee, Y., Kim, K., Kim, H., Sung, M.: Syncdiffusion: Coherent montage via synchronized joint diffusions. In: Advances in Neural Information Processing Systems. vol. 36, pp. 50648–50660 (2023)
35. Li, J., Cui, Y., Huang, T., Ma, Y., Fan, C., Yang, M., Zhong, Z., Bo, L.: Mixgrpo: Unlocking flow-based grpo efficiency with mixed ode-sde. arXiv preprint arXiv:2507.21802 (2025)
36. Li, X., Zhao, Y., Wang, C., Scalia, G., Eraslan, G., Nair, S., Biancalani, T., Regev, A., Levine, S., Uehara, M.: Derivative-free guidance in continuous and discrete diffusion models with soft value-based decoding. In: Advances in Neural Information Processing Systems. vol. 38, pp. 95507–95545 (2025)
37. Lin, Z., Pathak, D., Li, B., Li, J., Xia, X., Neubig, G., Zhang, P., Ramanan, D.: Evaluating text-to-visual generation with image-to-text generation. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 366–384 (2024)
38. Liu, J., Liu, G., Liang, J., Li, Y., Liu, J., Wang, X., Wan, P., Zhang, D., Ouyang, W.: Flow-GRPO: Training flow matching models via online RL. In: Advances in Neural Information Processing Systems. vol. 38, pp. 40783–40818 (2025)
39. Liu, J., Liu, G., Liang, J., Yuan, Z., Liu, X., Zheng, M., Wu, X., Wang, Q., Qin, W., Xia, M., et al.: Improving video generation with human feedback. In: Advances in Neural Information Processing Systems. vol. 38, pp. 82155–82192 (2025)
40. Lu, Y., Liang, Y., Zhu, L., Yang, Y.: Freelong: Training-free long video generation with spectralblend temporal attention. In: Advances in Neural Information Processing Systems. vol. 37, pp. 131434–131455 (2024)
41. Ma, N., Tong, S., Jia, H., Hu, H., Su, Y.C., Zhang, M., Yang, X., Li, Y., Jaakkola, T., Jia, X., Xie, S.: Inference-time scaling for diffusion models beyond scaling denoising steps. arXiv preprint arXiv:2501.09732 (2025)
42. Ma, X., Wang, Y., Chen, X., Wong, T.T., Chen, C.: Training-free stylized text-to-image generation with fast inference. arXiv preprint arXiv:2505.19063 (2025)
43. Prabhudesai, M., Mendonca, R., Qin, Z., Fragkiadaki, K., Pathak, D.: Video diffusion alignment via reward gradients. In: International Conference on Learning Representations (2025)
44. Qiu, H., Chen, Z., Wang, Z., He, Y., Xia, M., Liu, Z.: Freetraj: Tuning-free trajectory control in video diffusion models. arXiv preprint arXiv:2406.16863 (2024)
45. Rafailov, R., Sharma, A., Mitchell, E., Manning, C.D., Ermon, S., Finn, C.: Direct preference optimization: Your language model is secretly a reward model. In: Advances in Neural Information Processing Systems. vol. 36, pp. 53728–53741 (2023)
46. Ramesh, V., Mardani, M.: Test-time scaling of diffusion models via noise trajectory search. arXiv preprint arXiv:2506.03164 (2025)
47. Robbins, H.E.: An Empirical Bayes Approach to Statistics. Springer (1992)
48. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10674–10685 (2022)
49. Rout, L., Chen, Y., Kumar, A., Caramanis, C., Shakkottai, S., Chu, W.S.: Beyond first-order tweedie: Solving inverse problems using latent diffusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 9472–9481 (2024)
50. Rout, L., Raoof, N., Daras, G., Caramanis, C., Dimakis, A., Shakkottai, S.: Solving linear inverse problems provably via posterior sampling with latent diffusion models. In: Advances in Neural Information Processing Systems. vol. 36, pp. 49960–49990 (2023)

51. Rozet, F., Andry, G., Lanusse, F., Louppe, G.: Learning diffusion priors from observations by expectation maximization. In: Advances in Neural Information Processing Systems. vol. 37, pp. 87647–87682 (2024)
52. Samuel, D., Ben-Ari, R., Darshan, N., Maron, H., Chechik, G.: Norm-guided latent space exploration for text-to-image generation. In: Advances in Neural Information Processing Systems. vol. 36, pp. 57863–57875 (2023)
53. Schuhmann, C.: LAION aesthetics. `https://laion.ai/blog/laion-aesthetics` (2022), accessed 26 June 2026 (AoE)
54. Seo, J., Veer, S., Tian, R., Ding, W., Sharma, A., Leung, K., Schmerling, E., Pavone, M., Bajcsy, A.: Stressdream: Steering video world models for robust policy evaluation and improvement. arXiv preprint arXiv:2606.00267 (2026)
55. Singhal, R., Horvitz, Z., Teehan, R., Ren, M., Yu, Z., McKeown, K., Ranganath, R.: A general framework for inference-time scaling and steering of diffusion models. arXiv preprint arXiv:2501.06848 (2025)
56. Song, J., Vahdat, A., Mardani, M., Kautz, J.: Pseudoinverse-guided diffusion models for inverse problems. In: International Conference on Learning Representations (2023)
57. Song, J., Zhang, Q., Yin, H., Mardani, M., Liu, M.Y., Kautz, J., Chen, Y., Vahdat, A.: Loss-guided diffusion models for plug-and-play controllable generation. In: International Conference on Machine Learning. pp. 32483–32498 (2023)
58. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. In: International Conference on Learning Representations (2021)
59. Stiennon, N., Ouyang, L., Wu, J., Ziegler, D.M., Lowe, R., Voss, C., Radford, A., Amodei, D., Christiano, P.: Learning to summarize from human feedback. In: Advances in Neural Information Processing Systems. vol. 33, pp. 3008–3021 (2020)
60. Tang, Z., Peng, J., Tang, J., Hong, M., Wang, F., Chang, T.H.: Inference-time alignment of diffusion models with direct noise optimization. In: International Conference on Machine Learning. pp. 58905–58930 (2025)
61. Uehara, M., Zhao, Y., Black, K., Hajiramezanali, E., Scalia, G., Diamant, N.L., Tseng, A.M., Biancalani, T., Levine, S.: Fine-tuning of continuous-time diffusion models as entropy-regularized control. In: International Conference on Learning Representations (2025)
62. Uehara, M., Zhao, Y., Hajiramezanali, E., Scalia, G., Eraslan, G., Lal, A., Levine, S., Biancalani, T.: Bridging model-based optimization and generative modeling via conservative fine-tuning of diffusion models. In: Advances in Neural Information Processing Systems. vol. 37, pp. 127511–127535 (2024)
63. Wallace, B., Dang, M., Rafailov, R., Zhou, L., Lou, A., Purushwalkam, S., Ermon, S., Xiong, C., Joty, S., Naik, N.: Diffusion model alignment using direct preference optimization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8228–8238 (2024)
64. Wan Team, A.G.: Wan: Open and advanced large-scale video generative models. arXiv preprint arXiv:2503.20314 (2025)
65. Wu, L., Trippe, B., Naesseth, C., Blei, D., Cunningham, J.P.: Practical and asymptotically exact conditional sampling in diffusion models. In: Advances in Neural Information Processing Systems. vol. 36, pp. 31372–31403 (2023)
66. Wu, X., Hao, Y., Sun, K., Chen, Y., Zhu, F., Zhao, R., Li, H.: Human preference score v2: A solid benchmark for evaluating human preferences of text-to-image synthesis. In: International Conference on Learning Representations (2024)

67. Xu, J., Liu, X., Wu, Y., Tong, Y., Li, Q., Ding, M., Tang, J., Dong, Y.: ImageReward: Learning and evaluating human preferences for text-to-image generation. In: Advances in Neural Information Processing Systems. vol. 36, pp. 15903–15935 (2023)
68. Ye, H., Lin, H., Han, J., Xu, M., Liu, S., Liang, Y., Ma, J., Zou, J.Y., Ermon, S.: TFG: Unified training-free guidance for diffusion models. In: Advances in Neural Information Processing Systems. vol. 37, pp. 22370–22417 (2024)
69. Yoon, T., Min, Y., Yeo, K., Sung, M.: Psi-sampler: Initial particle sampling for smc-based inference-time reward alignment in score models. In: Advances in Neural Information Processing Systems. vol. 38, pp. 104745–104781 (2025)
70. Yu, J., Wang, Y., Zhao, C., Ghanem, B., Zhang, J.: FreeDoM: Training-free energy-guided conditional diffusion model. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 23174–23184 (2023)
71. Zhang, Y., Wei, Y., Jiang, D., Zhang, X., Zuo, W., Tian, Q.: Controlvideo: Training-free controllable text-to-video generation. In: International Conference on Learning Representations (2024)