

TRUST-REGION NOISE SEARCH FOR BLACK-BOX ALIGNMENT OF DIFFUSION AND FLOW MODELS

Niklas Schweiger¹ Daniel Cremers^{1,†} Karnik Ram^{1,†}

¹TU Munich
niklas.schweiger@tum.de

ABSTRACT

Optimizing the noise samples of diffusion and flow models is an increasingly popular approach to align these models to target rewards at inference time. However, we observe that these approaches are usually restricted to differentiable or cheap reward models, the formulation of the underlying pre-trained generative model, or are memory/compute inefficient. We instead propose a simple trust-region based search algorithm (TRS) which treats the pre-trained generative and reward models as a black-box and only optimizes the source noise. Our approach achieves a good balance between global exploration and local exploitation, and is versatile and easily adaptable to various generative settings and reward models with minimal hyperparameter tuning. We evaluate TRS across text-to-image, molecule and protein design tasks, and obtain significantly improved output samples over the base generative models and other inference-time alignment approaches which optimize the source noise sample, or even the entire reverse-time sampling noise trajectories in the case of diffusion models. Our source code is publicly available*.

1 INTRODUCTION

Generative models such as diffusion and flow-based models have revolutionized diverse domains, including high-fidelity image and video synthesis, molecule generation, and language modeling, by learning complex data distributions from large-scale pretraining data (Ho et al., 2020; Lipman et al., 2023). While scaling model capacity and training compute has significantly improved sample quality, pretrained models often fall short of specific, fine-grained requirements. This includes generating molecules with precise binding affinities (Guan et al., 2023) or images that strictly align with complex, multi-attribute prompts.

This gap has motivated the emergence of inference-time alignment for diffusion and flow-based models, an exciting new paradigm where the quality of generated samples is optimized post-training using feedback from target reward models (Bansal et al., 2024; Eyring et al., 2024). Unlike fine-tuning, this approach requires no additional training data and instead invests additional compute at inference time to steer samples toward desired properties.

Several approaches for inference-time alignment have been proposed. *Gradient-based* methods back-propagate through the entire (often ODE-based) iterative process of the model to adjust the initial noise sample (Ben-Hamu et al., 2024; Wang et al., 2025). But these are prone to high GPU-memory costs and risk drifting off the training-data manifold. *Sequence-based* methods, typically used for SDE samplers, range from filtering approaches that iteratively reweigh and resample (Singhal et al., 2025; Kim et al., 2025) to tree-search based methods (Jain et al., 2025; Li et al., 2025a). But these often require a high number of expensive reward calls or rely on accurate value estimates for terminal rewards, which are not always available or reliable. In contrast, *black-box search* methods which are relative underexplored, treat the

[†]Equal supervision.

*Project page: <https://niklasschweiger.github.io/trust-region-noise-search/>

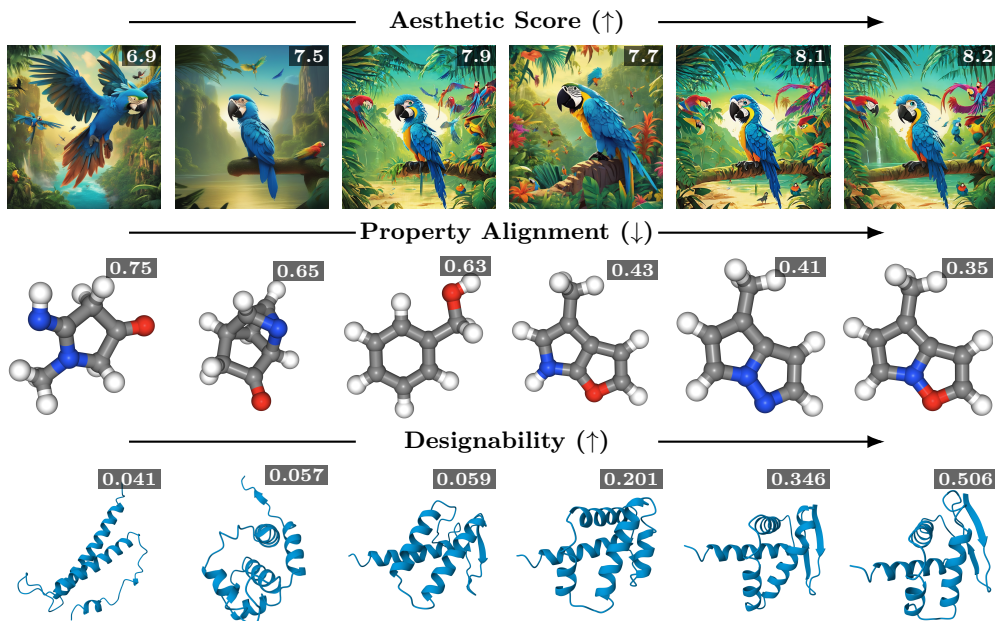


Figure 1: Progression of output samples produced by our algorithm (TRS) across iterations $i \in \{0, \dots, 15\}$ for text-to-image, molecules, and proteins. *Top*: Aesthetic reward alignment for the prompt “Animated movie poster...”. *Middle*: Molecule property alignment. *Bottom*: Protein designability. The horizontal arrow indicates the direction of optimization, while the arrows $\uparrow / \downarrow$ indicate whether the objective is maximized or minimized.

generator and reward models as a black box and apply search heuristics for the source noise sample (Ma et al., 2025; Tan et al., 2025). While versatile and applicable to any generative architecture and reward model, we notice that existing approaches often struggle to find a good balance between global exploration and local exploration, tending toward one extreme or the other.

Inspired by Bayesian optimization algorithms (Eriksson et al., 2019), which also optimize expensive black-box functions, we propose a simple and effective *trust region search* (TRS) approach that balances between global exploration and local exploitation in a structured manner. By controlling only the source noise, TRS is readily applicable to a wide range of generative models and reward functions without internal modifications. TRS begins by exploring multiple seed noise samples, which are then pruned and iteratively refined with local perturbations. Crucially, these perturbations are adaptively controlled in both magnitude and direction based on observed reward values, ensuring the search remains within the data manifold to produce stable and high-quality samples.

The contributions of our work are as follows:

1. We introduce *trust region search* (TRS), a simple approach for inference-time reward alignment of black-box diffusion and flow models via adaptive source noise control.
2. We provide an extensive evaluation on text-to-image generation, demonstrating that TRS yields significantly more aligned and higher-quality samples compared to existing search heuristics and even full-noise sequence search baselines under identical compute budgets.
3. We demonstrate the versatility of our approach through an extended evaluation on small molecule and protein design tasks, where TRS proves effective even with expensive reward functions and requires minimal hyperparameter tuning.

2 BACKGROUND AND RELATED WORK

2.1 DIFFUSION AND FLOW-BASED MODELS

Diffusion and flow-based generative models aim to transport a simple noise distribution $p_{\text{noise}} = p_0$ to a complex data distribution $p_{\text{data}} = p_1$. A sample $\mathbf{x}_0 \sim p_0$ is gradually transformed into a data sample $\mathbf{x}_1 \sim p_1$ through a continuous-time process $\{\mathbf{x}_t\}_{t \in [0,1]}$, typically discretized into T steps with $t_k = k/T$. In continuous data regimes, p_0 is usually white Gaussian noise $\mathcal{N}(\mathbf{0}, \mathbf{I})$.

Diffusion models are characterized by a forward noising process that gradually transforms data into noise. This process is governed by a stochastic differential equation (SDE):

$$d\mathbf{x}_t = \mathbf{f}(t)\mathbf{x}_t dt + g(t)d\mathbf{w}_t,$$

where $\mathbf{f}(t)$ and $g(t)$ are the drift and diffusion coefficients that determine the noising schedule (Song et al., 2021). To generate new samples, this process must be reversed. The generative (reverse) process relies on the score function $\boldsymbol{\epsilon}(\mathbf{x}_t, t) \triangleq \nabla_{\mathbf{x}} \log p_t(\mathbf{x}_t)$, which represents the gradient of the log-density of the noisy sample. Once the score is estimated via a neural network, sampling can be performed using either the deterministic or stochastic formulations shown in Table 1.

Flow matching models instead learn a continuous vector field $\mathbf{v}_\theta(\mathbf{x}_t, t) = d\mathbf{x}_t/dt$ (Lipman et al., 2023) that defines the velocity of the samples. Unlike diffusion, the paths are not necessarily tied to a corruption process; a common choice is the optimal transport (OT) path (Liu et al., 2023), defined by linear interpolation

$$\mathbf{x}_t = (1 - t)\mathbf{x}_0 + t\mathbf{x}_1,$$

which yields a constant target velocity $\mathbf{v}_t = \mathbf{x}_1 - \mathbf{x}_0$. Standard flow matching integrates the learned ODE for sampling, while stochastic variants (Bose et al., 2024) can introduce Brownian motion and optional score corrections as detailed in Table 1.

Unified view. Both paradigms can be expressed in a general form $d\mathbf{x}_t = \mathbf{f}(\mathbf{x}_t, t)dt + g(t)d\mathbf{w}_t$. As summarized in Table 1, the two frameworks differ primarily in how the drift $\mathbf{f}(\mathbf{x}_t, t)$ is parameterized, either through the score function $\boldsymbol{\epsilon}(\mathbf{x}_t, t)$ to reverse a corruption process, or through a velocity field $\mathbf{v}_\theta(\mathbf{x}_t, t)$ to match a probability flow.

Table 1: Generative sampling formulations ($t = 0$ as noise, $t = 1$ as data). Note that for diffusion, $\boldsymbol{\epsilon}(\mathbf{x}_t, t)$ denotes the score function $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$.

Formulation	ODE	SDE
Diffusion	$d\mathbf{x}_t = [\mathbf{f}(t)\mathbf{x}_t - \frac{1}{2}g^2(t)\boldsymbol{\epsilon}(\mathbf{x}_t, t)]dt$	$d\mathbf{x}_t = [\mathbf{f}(t)\mathbf{x}_t - g^2(t)\boldsymbol{\epsilon}(\mathbf{x}_t, t)]dt + g(t)d\mathbf{w}_t$
Flow Matching	$d\mathbf{x}_t = \mathbf{v}_\theta(\mathbf{x}_t, t)dt$	$d\mathbf{x}_t = \mathbf{v}_\theta(\mathbf{x}_t, t)dt + g(t)d\mathbf{w}_t$

2.2 REWARD ALIGNMENT OF DIFFUSION AND FLOW MODELS.

While pre-trained diffusion and flow models generate high-quality samples, they often fail to meet specific downstream criteria, such as aesthetics Wu et al. (2023) in the case of image generation, or specific properties such as stability or binding affinity in the case of molecule (Guan et al., 2023; Xu et al., 2022) and protein design (Ingraham et al., 2023; Watson et al., 2023). This alignment can be achieved by fine-tuning the model’s parameters (Clark et al., 2024; Black et al., 2024) but this requires collecting new samples for every new target property, and is incompatible with new properties that are presented only at inference-time. A promising alternative approach is *reward alignment* where the noise samples during generation are aligned to meet a target reward objective. Conceptually, these aligned or optimized noise

Table 2: **Feature comparison across noise optimization methods.** We compare methods by their ability to support gradient-free objectives, black-box models, parallelized batch efficiency, and global or local optimization.

Method	Grad-free	Black-box	Batch Eff.	Global	Local
OC-Flow (Tang et al., 2025)	✗	✗	✗	✗	✓
Fast Direct (Tan et al., 2025)	✓	✓	✓	✓	✗
DTS* (Jain et al., 2025)	✓	✗	✗	✓	✗
Random Search (Ma et al., 2025)	✓	✓	✓	✓	✗
Zero-Order (Ma et al., 2025)	✓	✓	✓	✗	✓
TRS (Ours)	✓	✓	✓	✓	✓

samples create a new target distribution that balances two competing objectives: staying true to the structural distribution learned during pretraining while shifting the probability mass towards target samples that maximize the reward. We categorize these noise optimization or search methods as follows.

Gradient-based guidance. One approach to noise optimization involves back-propagating the gradients from a reward function through the entire iterative sampling process to update the source (or intermediate) noise samples (Ben-Hamu et al., 2024; Wang et al., 2025; Guo et al., 2024; Tang et al., 2025). While effective, these methods require differentiable reward functions and incur substantial GPU memory and computational overhead, as they necessitate storing or recomputing the full diffusion trajectory during inference. This limitation becomes particularly severe in high-dimensional settings, such as image or 3D generation, or when many solver steps are required. Furthermore, pure gradient-based refinement often shifts the generation off the natural data manifold, necessitating additional regularization or alignment terms to preserve sample quality (Wang et al., 2025). An exception is direct noise optimization (DNO) (Tang et al., 2025), which additionally supports gradient approximations for non-differentiable rewards and an SDE-based mode that can refine intermediate noises along the trajectory. Despite these extensions, the same disadvantages of high runtime and GPU memory consumption remain.

Noise sequence search. Another popular approach is to guide the generation process throughout the whole or parts of the sampling trajectory $\{\mathbf{x}_0, \dots, \mathbf{x}_1\}$. This includes sequential Monte Carlo (SMC) approaches using resampling (Singhal et al., 2025; Kim et al., 2025), tree search methods such as DSearch and DTS (Li et al., 2025a; Jain et al., 2025), and Fast Direct (Tan et al., 2025) which uses Gaussian process-based black-box optimization over the full noise sequence (Tan et al., 2025). While these methods can be effective for stochastic sampling, they typically rely on intermediate reward approximations (Li et al., 2025b) or are difficult to utilize with batched evaluations.

Black-box search. Most versatile are source noise optimization algorithms which treat the generative model and the reward model as a black-box. They can be applied to any generative model, that maps a source noise distribution to some data distribution, and any reward model which can be differentiable or non-differentiable. They are conceptually much simpler and easy to implement, while also leading to surprisingly good results. This category of approaches is relatively underexplored but includes some recent work by Ma et al. (2025) where they apply random search and zero-order search for the source noise. Some additional recent (concurrent) work is by Jajal and Eliopoulos et al. (2025) where they apply evolutionary and genetic search algorithms for the source noise. Our work belongs to this category and we apply trust-region search, inspired by Bayesian optimization algorithms Eriksson et al. (2019), and show strong results across data modalities and reward models.

3 METHODOLOGY

3.1 PROBLEM STATEMENT

Let $\mathcal{F} : \mathbb{R}^M \rightarrow \mathbb{R}^D$ be a pre-trained generative diffusion or flow-based model which maps source noise $\mathbf{x}_0$ to a data sample $\mathbf{x}_1$ and $R : \mathbb{R}^D \rightarrow \mathbb{R}$ a reward model processing $\mathbf{x}_1$ and returning a scalar reward r . Note that generative models typically operate in a

compressed latent space Rombach et al. (2022), leading to $M < D$. We do not intervene with the model in any other way, treating it as a black-box, as illustrated in Figure 2. The resulting objective is:

$$\mathbf{x}_0^* = \arg \max_{\mathbf{x}_0 \in \mathbb{R}^M} R(\mathcal{F}(\mathbf{x}_0)). \quad (1)$$

We treat the entire mapping $R(\mathcal{F}(\mathbf{x}_0))$ as a computationally expensive black-box, as illustrated in Figure 2, either due to the generative model, the reward model, or the use of large batch sizes. Consequently, we focus on methods that can effectively steer generation under a strict evaluation budget.

3.2 TRUST-REGION SEARCH (TRS)

TRS is a trust-region search algorithm with structured noise sampling designed for inference-time alignment with expensive black-box evaluations. Our approach is inspired by TuRBO (Eriksson et al., 2019) for Bayesian optimization, but introduces several important modifications for inference-time steering of large generative models. In particular, unlike surrogate-based Bayesian optimization, TRS relies purely on structured sampling, since we find that surrogates do not contribute meaningfully to performance, due to the highly non-linear noise space, which is difficult to predict under

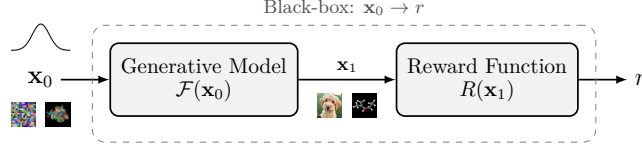


Figure 2: Illustration of the black-box, defined by the generative model $\mathcal{F}$ and the reward function $R(\mathbf{x}_1)$, which connect noise samples $\mathbf{x}_0$ with scalar rewards r .

Algorithm 1 Trust-Region Search (TRS)

Require: Total/Warmup Budget $N_{\text{total}}/N_{\text{warm}}$, Batch size B , Regions k , Model $\mathcal{F}$, Reward function R

- 1: **Warm-up:** Sample $\{\mathbf{x}_{0,i}\}_{i=1}^{N_{\text{warm}}} \sim p_0$; evaluate $r_i = R(\mathcal{F}(\mathbf{x}_{0,i}))$ in batches of size B .
- 2: **Init:** Set centers $\{\mathbf{x}_{0,j}^c\}_{j=1}^k$ to top k ; set lengths $\ell_j \leftarrow \ell_{\text{init}}$; set remaining budget $N \leftarrow N_{\text{total}} - N_{\text{warm}}$
- 3: **while** budget $N \geq B$ **do**
- 4: $\mathcal{S}_{\text{batch}} \leftarrow \emptyset$
- 5: **for** region $\mathcal{T}_j$ and sample $b \in \{1..B/k\}$ **do**
- 6: $p_{j,b} \sim \text{U}(p_{\min}, p_{\max})$; $\mathbf{m}_{j,b} \sim \text{Ber}(p_{j,b})^M$
- 7: $\tilde{\mathbf{x}}_{0,j,b} \sim \text{Perturb}(\ell_j)$ $\triangleright$ Gauss or Sobol
- 8: $\mathbf{x}_{0,j,b} \leftarrow \mathbf{x}_{0,j}^c + (\tilde{\mathbf{x}}_{0,j,b} \odot \mathbf{m}_{j,b})$
- 9: $\mathcal{S}_{\text{batch}} \leftarrow \mathcal{S}_{\text{batch}} \cup \{\mathbf{x}_{0,j,b}\}$
- 10: **end for**
- 11: **Evaluate:** $\{r_{j,b}\} \leftarrow R(\mathcal{F}(\mathcal{S}_{\text{batch}}))$
- 12: **Adapt:** Update lengths $\{\ell_j\}$
- 13: **Shift:** Re-center $\{\mathbf{x}_{0,j}^c\}$ to the global top k .
- 14: $N \leftarrow N - B$
- 15: **end while**
- 16: **return** best $\{\mathcal{F}(\mathbf{x}_0), r\}$

constrained budgets. Additionally, we use a different center selection scheme, by always choosing the top- k observed noises, which is crucial for the performance. An overview of the method is given in Algorithm 1.

Warm-up. We begin with a short warm-up phase for bootstrapping the search. We sample N_{warm} initial noise samples from the model prior p_0 ,

$$\mathbf{x}_{0,i} \sim p_0, \quad \mathbf{x}_{1,i} = \mathcal{F}(\mathbf{x}_{0,i}), \quad r_i = R(\mathbf{x}_{1,i}), \quad (2)$$

with $i = 1, \dots, N_{\text{warm}}$, and select the top-performing k points as initial trust-region centers $\{\mathbf{x}_{0,j}^c\}_{j=1}^k$. All trust regions are initialized with the same side length $\ell_j = \ell_{\text{init}}$. In practice, this warm-up is implemented using standard iterations with batch size B . We find allocating approximately 20% of the total evaluation budget N_{total} to this phase to be a robust heuristic across tasks and budgets; an ablation is provided in Appendix D.3.

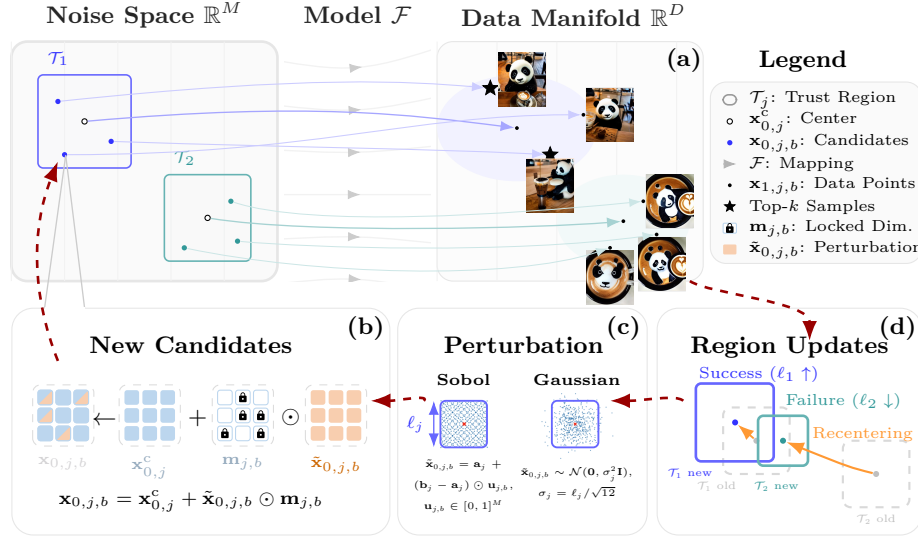


Figure 3: Illustration of the trust region search algorithm with a two-region example for the prompt “A panda making latte art” from DrawBench (Saharia et al., 2022). (a) Samples from the noise space $\mathbb{R}^M$ are mapped to the data manifold in $\mathbb{R}^D$ via the generative model $\mathcal{F}$. Generated samples from the same region ($\mathcal{T}_j$) exhibit visual similarity; here, $\mathcal{T}_1$ shows markedly better prompt alignment than $\mathcal{T}_2$. (b) New candidates $\mathbf{x}_{0,j,b}$ are generated by adding masked ($\mathbf{m}_{j,b}$) relative perturbations ($\tilde{\mathbf{x}}_{0,j,b}$) to the current center $\mathbf{x}_{0,j}^c$. (c) Comparison of Sobol and Gaussian perturbation schemes used to fill the trust-region hypercube. (d) The update and shift logic: $\mathcal{T}_1$ expands upon identifying a top- k sample, while the underperforming $\mathcal{T}_2$ is re-centered (shifted) to a more promising region.

Trust-region iterations. After warm-up, the algorithm maintains k hypercubic trust regions $\mathcal{T}^j \subset \mathbb{R}^M$, each defined by a center $\mathbf{x}_{0,j}^c$ and side length ℓ_j . In each iteration, we perform the following steps, where a global batch set $\mathcal{S}_{\text{batch}}$ of size B across all regions is maintained.

1. **Propose:** For each region j , generate B/k candidate noise vectors $\mathbf{x}_{0,j,b}$ by perturbing the center $\mathbf{x}_{0,j}^c$. Following (Eriksson et al., 2019), perturbations $\tilde{\mathbf{x}}_{0,j,b}$ are generated within an axis-aligned hypercube with side length ℓ_j . Additionally, we combine these perturbations with a stochastic coordinate mask. This mask is built by drawing a perturbation probability $p_{j,b} \sim \text{Uniform}(p_{\min}, p_{\max})$ and applying $\mathbf{m}_{j,b} \sim \text{Bernoulli}(p_{j,b})^M$. The resulting candidates are:

$$\mathbf{x}_{0,j,b} = \mathbf{x}_{0,j}^c + (\tilde{\mathbf{x}}_{0,j,b}) \odot \mathbf{m}_{j,b}. \quad (3)$$

2. **Evaluate:** All B candidates across regions are aggregated into $\mathcal{S}_{\text{batch}}$ and evaluated in parallel to obtain the rewards $\{r_{j,b}\} = R(\mathcal{F}(\mathcal{S}_{\text{batch}}))$.
3. **Update:** After batch evaluation, we adapt trust-region side lengths $\{\ell_j\}$ using success-based rules per region. A key distinction from vanilla TuRBO (Eriksson et al., 2019) is that trust regions are not treated independently. While exploring multiple regions is beneficial early on, we observe that only a subset typically remains promising. Accordingly, after each batch iteration we re-center all trust regions $\{\mathbf{x}_{0,j}^c\}$ at the globally best k points observed so far. This mechanism naturally shifts computation from exploration toward exploitation, reallocating evaluation budget to promising regions over time. An ablation study on different re-centering strategies is provided in Appendix D.1.

Perturbations. For generating the perturbations in the proposal step we use two different schemes. Deterministic low-discrepancy samplers such as Sobol (Sobol’, 1967) are efficient to fill search spaces, but can only be computed only up to 21k dimensions (Joe & Kuo, 2008). For higher dimensional search spaces, like in our experiment with SDXL-Lightning in

Section 4.1 with noise dimension 65,536, we use a Gaussian perturbation scheme, that is designed to follow the trust-region hypercube exploration closely.

For Sobol-based proposals, a point $\mathbf{u}_{j,b} \in [0, 1]^M$ is sampled and mapped affinely into the trust region. Let $\mathbf{a}_j = -\frac{1}{2}\ell_j \mathbf{1}_M$ and $\mathbf{b}_j = \frac{1}{2}\ell_j \mathbf{1}_M$, where $\mathbf{1}_M$ denotes a M -dimensional vector of ones. The resulting proposal is

$$\tilde{\mathbf{x}}_{0,j,b} = \mathbf{a}_j + (\mathbf{b}_j - \mathbf{a}_j) \odot \mathbf{u}_{j,b}. \quad (4)$$

For Gaussian perturbations, we sample

$$\tilde{\mathbf{x}}_{0,j,b} \sim \mathcal{N}(\mathbf{0}, \sigma_j^2 \mathbf{I}), \quad \sigma_j = \ell_j / \sqrt{12}, \quad (5)$$

where $\mathbf{I} \in \mathbb{R}^{M \times M}$ is the identity matrix. The standard deviation $\sigma_j = \ell_j / \sqrt{12}$ is chosen to match the variance of a uniform distribution $\mathcal{U}[-\ell_j/2, \ell_j/2]$, ensuring the Gaussian proposals cover the trust-region hypercube with equivalent spread.

Trust-region adaptation. The update rules for the trust-region side length largely follow the strategy introduced in TuRBO (Eriksson et al., 2019). For each region, we maintain success and failure counters based on whether newly evaluated candidates improve upon the best value observed within that region. If a candidate yields an improvement, the success counter is incremented and the failure counter is reset. Otherwise, the failure counter is incremented and the success counter is reset. Two thresholds, c_{succ} and c_{fail} , govern the adaptation of the trust-region side length ℓ_j . When the success counter reaches c_{succ} , the trust region is expanded according to

$$\ell_j^{\text{new}} = \min(\ell_j \cdot \alpha_\ell, \ell_j^{\text{max}}), \quad (6)$$

where α_ℓ is an expansion factor for the length (we set it to 1.5 by default). Conversely, when the failure counter reaches c_{fail} , the trust region is contracted as

$$\ell_j^{\text{new}} = \max(\ell_j / \alpha_\ell, \ell_j^{\text{min}}). \quad (7)$$

This mechanism enables adaptive control over the exploration scale based on recent optimization progress. We further illustrate the core steps of TRS in Figure 3 and provide additional details in the appendix.

4 EXPERIMENTS

We evaluate our trust-region search (TRS) method across three diverse generative settings: text-to-image diffusion models in Section 4.1, ODE-based molecular flow matching in Section 4.2, and protein backbone design in Section 4.3. These experiments span different modalities, reward models, and sampling procedures, allowing us to assess both the effectiveness and robustness of TRS under varying optimization budgets and reward costs. Across all settings, we compare against representative gradient-based, noise-sequence, and black-box search baselines under matched compute constraints. Overall, the results demonstrate that TRS consistently achieves stronger alignment with target objectives while maintaining sample quality and requiring minimal task-specific tuning.

4.1 TEXT-TO-IMAGE

Setup. We evaluate on the DrawBench (Saharia et al., 2022) benchmark which comprises 200 prompts across diverse categories (e.g., counting, composition). We use two generative models: SD1.5 (Rombach et al., 2022) (50 steps) and SDXL-Lightning (Lin et al., 2024a) (8 steps), with ImageReward (Xu et al., 2023) and HPSv2 (Wu et al., 2023) as the reward models. TRS is compared against gradient-based OC-Flow (Wang et al., 2025), state-of-the-art Diffusion Tree Sampling (DTS)* (Jain et al., 2025), Fast Direct (FD) (Tan et al., 2025), and black-box methods including random and zero-order search (RS, ZO) (Ma et al., 2025). The number of function evaluations (NFE) is fixed across methods. Further implementation details for the baselines are in Appendix F.

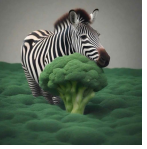
Prompt	TRS (Ours)	RS	ZO	DTS	FD	OC-Flow
Three cats and two dogs sitting on the grass.						
A storefront with 'Google Brain Toronto' written on it.						
A zebra underneath a broccoli.						

Figure 4: Examples of optimized samples from different algorithms for all methods in Section 4.1, with SDXL-Lightning (Lin et al., 2024a). The first row is optimized with ImageReward (Xu et al., 2023) and the lower two rows with HPSv2 (Wu et al., 2023). All prompts are from DrawBench (Saharia et al., 2022). Outputs from TRS adhere to the prompt more closely in terms of specified animal count, text, and relative positions. Further examples (incl. randomized) are given in the appendix.

Metrics. We report the mean of the best per-prompt reward Table 3 and average best rewards across varying NFE budgets in Figure 5. Additionally, we report the number of reward oracle evaluations induced by the fixed NFE budget (Table 3). Direct NFE comparison for backpropagation-based OC-Flow is provided via runtime analysis in Appendix H. Note that the ImageReward outputs are designed to roughly follow a normal distribution, so outputs beyond the range of $[-2, 2]$ are rare. HPSv2 shows differences in a smaller scale where typical high-quality images fall within the range of $[0.25, 0.35]$.

Results. Table 3 demonstrates that TRS is consistently better than all baselines across generative models and reward functions. On both models, TRS exceeds the state-of-the-art DTS* performance with up to $4\times$ reduction in wall-clock time and fewer reward evaluations. OC-Flow and Fast Direct underperform relative to random search, highlighting the difficulty of surrogate modeling and gradient-based optimization in high-dimensional noise spaces, while TRS shows the best results. Figure 4 show some qualitative results and Figure 5 further illustrates the scaling trends from small to high budgets, where TRS generally shows the best trends, where else the other algorithms saturate earlier. Full hyperparameter configurations of all methods are displayed in Appendix B.

4.2 MOLECULE GENERATION

Setup. In this experiment, we move to ODE-based flow matching and consider a different data modality. Our goal is to generate small molecules with specified target values of chemical properties (Ramakrishnan et al., 2014). We use EquiFM (Song et al., 2023) with 50 integration steps, operating on a joint continuous Gaussian coordinate- and encoded feature space (291 total dimensions), with pretrained chemical property prediction models. This setting is a standard benchmark for evaluating classifier-free guidance (Hooeboom et al., 2022) and was also used by OC-Flow (Wang et al., 2025) for noise optimization toward single-property targets. However, we observe that random search based methods already achieve very strong performance in this regime and therefore extend the benchmark to multi-property target matching. Specifically, we define a distance-based loss as the mean absolute deviation between predicted and target values across multiple chemical properties. Since our framework is formulated as reward maximization, we maximize the negative of this

Table 3: Comparison of noise optimization algorithms on DrawBench. We report the mean best rewards (IR, HPSv2) for SD1.5 and SDXL alongside the computational cost in terms of number of reward calls. $\uparrow$ / $\downarrow$ indicate the direction of improvement. Best values are in **bold**, second best are underlined.

Algorithm	SD1.5		SDXL		Reward calls
	IR $\uparrow$	HPS $\uparrow$	IR $\uparrow$	HPS $\uparrow$	
Base	-0.16	0.246	0.50	0.262	1
Gradient-based guidance					
OC-Flow (Wang et al., 2025)	0.42	0.277	0.85	0.287	–
Noise sequence search					
Fast Direct (Tan et al., 2025)	1.35	0.296	1.50	0.314	420
DTS* (Jain et al., 2025)	<u>1.59</u>	0.303	<u>1.62</u>	0.327	503
Black-box search					
Random (Ma et al., 2025)	1.44	0.302	1.54	0.324	400
Zero-order (Ma et al., 2025)	1.50	<u>0.313</u>	1.59	<u>0.332</u>	400
TRS (Ours)	1.62	0.322	1.66	0.340	400

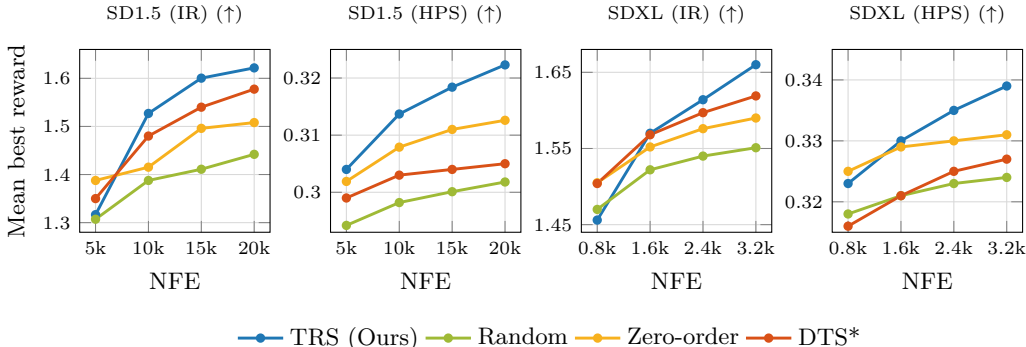


Figure 5: Here we plot the mean best rewards for SD1.5/SDXL optimizing for HPSv2 and ImageReward across different NFE budgets. TRS shows the best scaling performance among all methods.

loss. We consider two reward functions, R_3 and R_6 , involving 3 and 6 chemical properties, respectively and additional property details are provided in Appendix C.2. As baselines, we exclude noise-sequence search methods, as they are incompatible with ODE-based flow matching, and compare against OC-Flow, random search, and zero-order search.

Metrics. As our primary metric, we report the mean best losses $L_3(x) = -R_3(x)$ and $L_6(x) = -R_6(x)$ (mean distance to the target properties) over 200 molecules, with the target property values sampled randomly. In addition, we track other relevant metrics such as the molecule stability percentage (MSP), and valid and unique percentage (VUP).

Results. The quantitative results are summarized in Figure 7. Additionally, we visualize some randomly chosen optimized samples from different solversFigure 6. We observe that TRS consistently achieves the lowest losses, indicating the highest alignment with the specified multi-property targets. Importantly, this improvement does not come at the expense of other quality metrics: molecule stability and novelty remain comparable to those of the base model. In contrast, the gradient-based OC-Flow method exhibits degraded stability and novelty, even with regularization. This suggests that gradient-based optimization tends to drift off the data manifold in this setting. We find that sampling-based approaches are better suited for this task; notably, even random search outperforms OC-Flow. This performance gap likely stems from conflicting gradients between different property classifiers, which makes local optimization difficult. As illustrated in the bottom row of Figure 6, OC-Flow often fails

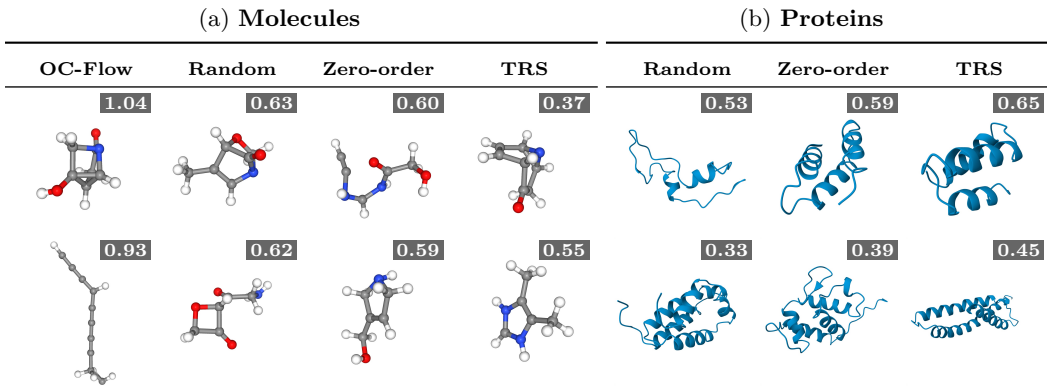


Figure 6: **Optimized molecule and protein samples.** We visualize randomly selected samples produced by different optimization methods in the molecule and protein experiments. (a) In the top row optimization is conducted with three target properties (L_3), while the bottom row with the six (L_6). Lower is better. (b) In the top row $n_{\text{res}}=50$ and in the bottom row $n_{\text{res}}=100$. The proteins are optimized for the designability reward.

to escape unfavorable initial samples, whereas sampling-based methods explore the space more effectively.

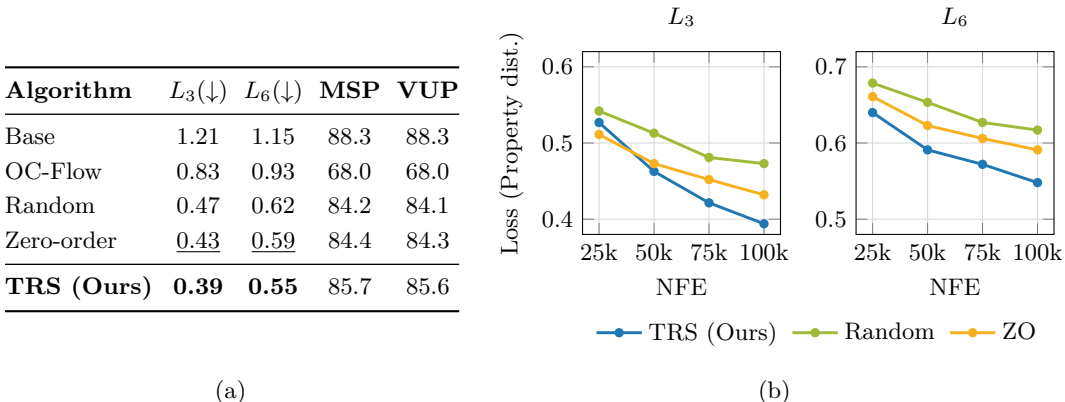


Figure 7: Molecule optimization results. (a) Distance to the combined targets of 3 and 6 chemical property values, together with MSP stability and validity VUP metrics which we do not explicitly optimize. The best alignment values are in **bold**, second best are underlined. (b) Scaling trends over the NFE budget.

4.3 PROTEIN DESIGN

Setup. Protein design is another challenging data modality with expensive reward models and is relatively underexplored by inference-time alignment methods. State of the art 3D protein generation models, like Proteina (Geffner et al., 2025), are flow matching models that are typically trained via ODE integration, but sampling is often performed using an SDE scheduler with noise reduction to improve designability (Bose et al., 2024; Lin et al., 2024b). However, this procedure alters the target distribution and no longer samples from the full flow-matching distribution, which can negatively impact other metrics such as diversity and novelty. In this experiment, we therefore focus on steering ODE-based sampling while preserving the full distribution. For completeness, we provide a comparison to SDE-based sampling in Appendix E. We further note that SDE-based noise-sequence optimization methods are not applicable in this setting: they either rely on value estimation functions that are unavailable for 3D protein design, or are computationally infeasible due to the

high cost of reward evaluation. We design the experiment by fixing the number of residues $n_{\text{res}} = 50$ and $n_{\text{res}} = 100$. For the reward function, we use the computationally expensive designability reward based on large structure extraction and protein folding models (see Appendix C for details). We compare against purely black-box baselines, including random search and zero-order search (Ma et al., 2025).

Metrics. As alignment metric we report the mean best rewards for each run. Additionally we show the cluster diversity and the pairwise TM-score and a novelty metric, which compares the proteins to the entire PDB dataset (Berman et al., 2000). Further details of these are given in Appendix G.2.

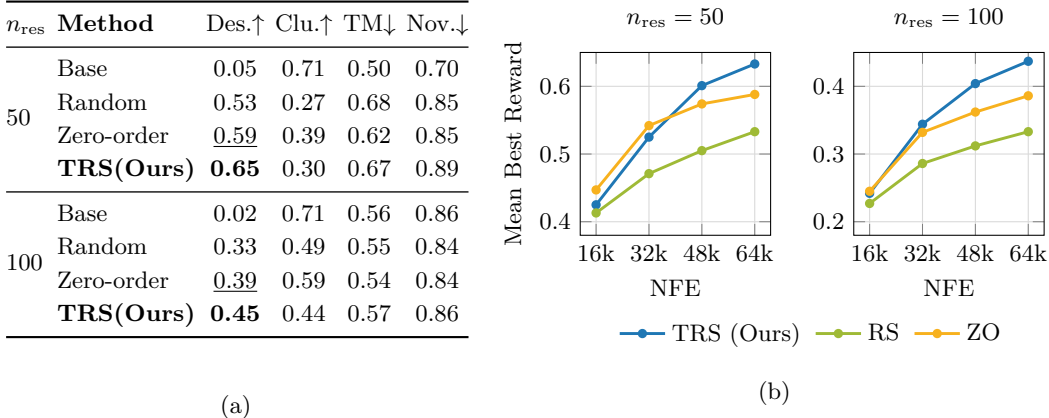


Figure 8: Protein optimization results. (a) Optimized mean designability rewards at 64k NFE, together with diversity and novelty metrics which we do not explicitly optimize. The best values are in **bold**, second best are underlined. (b) Reward improvement across NFE budgets.

Results. In terms of the designability steering, we see in Figure 8, that we significantly improve over other search algorithms in both settings. We see that generally the rewards are higher when we optimize proteins with 50 residues, which is expected, but also their diversity and novelty metrics decrease compared to the base model. We assume that, since the solution space is rather small, it is likely that well designable proteins share similar features. However, this is not comparable to the mode collapse, which we observe for SDE noise reduction (see Appendix E). Naturally, due to a larger solution space, optimizing proteins with 100 residues leads to better diversity and novelty, while still gaining major improvements over the base model. This counts for all optimization methods, but TRS achieves the best reward alignment again.

5 CONCLUSION

In this work we investigate inference-time scaling and preference alignment, where our simple trust-region source noise search achieves state-of-the-art performance across text-to-image, molecule, and protein design tasks. Our approach is model and reward agnostic, making it particularly suited for real-world settings where reward functions are often expensive or unknown. It also offers good balance between exploration and exploitation by searching multiple noise regions early and refining promising ones, and does not drift off the data manifold and remains stable as we observe in Section 4.2. While all methods are limited by the accuracy of the reward models (Appendix I), scaling improvements (Wu et al., 2025) suggest this limitation will diminish and our efficient source noise optimization is particularly well-suited for this development. Future work includes exploring the geometry of the source noise space further and developing improved perturbation schemes that adhere more to this geometry.

REFERENCES

- Arpit Bansal, Hong-Min Chu, Avi Schwarzschild, Roni Sengupta, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Universal Guidance for Diffusion Models. In *The Twelfth International Conference on Learning Representations*, 2024.
- Heli Ben-Hamu, Omri Puny, Itai Gat, Brian Karrer, Uriel Singer, and Yaron Lipman. D-Flow: Differentiating through Flows for Controlled Generation. In *Forty-first International Conference on Machine Learning*, 2024.
- Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, Talapady N Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. The protein data bank. *Nucleic acids research*, 28(1):235–242, 2000.
- Kevin Black, Michael Janner, Yilun Du, Ilya Kostrikov, and Sergey Levine. Training Diffusion Models with Reinforcement Learning. In *The Twelfth International Conference on Learning Representations*, 2024.
- Joey Bose, Tara Akhound-Sadegh, Guillaume Huguet, Kilian FATRAS, Jarrod Rector-Brooks, Cheng-Hao Liu, Andrei Cristian Nica, Maksym Korablyov, Michael M. Bronstein, and Alexander Tong. SE(3)-Stochastic Flow Matching for Protein Backbone Generation. In *The Twelfth International Conference on Learning Representations*, 2024.
- Kevin Clark, Paul Vicol, Kevin Swersky, and David J. Fleet. Directly Fine-Tuning Diffusion Models on Differentiable Rewards. In *The Twelfth International Conference on Learning Representations*, 2024.
- J. Dauparas, I. Anishchenko, N. Bennett, H. Bai, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, A. Courbet, R. J. de Haas, N. Bethel, P. J. Y. Leung, T. F. Huddy, S. Pellock, D. Tischer, F. Chan, B. Koepnick, H. Nguyen, A. Kang, B. Sankaran, A. K. Bera, N. P. King, and D. Baker. Robust deep learning-based protein sequence design using ProteinMPNN. *Science*, 378(6615):49–56, 2022.
- Prafulla Dhariwal and Alexander Quinn Nichol. Diffusion Models Beat GANs on Image Synthesis. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, 2021.
- David Eriksson, Michael Pearce, Jacob Gardner, Ryan D Turner, and Matthias Poloczek. Scalable Global Optimization via Local Bayesian Optimization. In *Advances in Neural Information Processing Systems*, pp. 5496–5507, 2019.
- Luca Eyring, Shyamgopal Karthik, Karsten Roth, Alexey Dosovitskiy, and Zeynep Akata. ReNO: Enhancing One-step Text-to-Image Models through Reward-based Noise Optimization. In *Neural Information Processing Systems (NeurIPS)*, 2024.
- Tomas Geffner, Kieran Didi, Zuobai Zhang, Danny Reidenbach, Zhonglin Cao, Jason Yim, Mario Geiger, Christian Dallago, Emine Kucukbenli, Arash Vahdat, and Karsten Kreis. Proteina: Scaling Flow-based Protein Structure Generative Models. In *The Thirteenth International Conference on Learning Representations*, 2025.
- Jiaqi Guan, Wesley Wei Qian, Xingang Peng, Yufeng Su, Jian Peng, and Jianzhu Ma. 3D Equivariant Diffusion for Target-Aware Molecule Generation and Affinity Prediction. In *The Eleventh International Conference on Learning Representations*, 2023.
- Xiefan Guo, Jinlin Liu, Miaomiao Cui, Jiankai Li, Hongyu Yang, and Di Huang. Initno: Boosting Text-to-Image Diffusion Models via Initial Noise Optimization. In *CVPR*, pp. 9380–9389, 2024.
- Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS ’20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.

- Emiel Hoogeboom, Víctor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant Diffusion for Molecule Generation in 3D. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 8867–8887. PMLR, 17–23 Jul 2022.
- John Ingraham, Max Baranov, Zak Costello, Karl Barber, Wujie Wang, Ahmed Ismail, Vincent Frappier, Dana Lord, Christopher Ng-Thow-Hing, Erik Vlack, Shan Tie, Vincent Xue, Sarah Cowles, Alan Leung, João Rodrigues, Claudio Morales-Perez, Alex Ayoub, Robin Green, Katherine Puentes, and Gevorg Grigoryan. Illuminating protein space with a programmable generative model. *Nature*, 623:1–9, 11 2023.
- Vineet Jain, Kusha Sareen, Mohammad Pedramfar, and Siamak Ravanbakhsh. Diffusion Tree Sampling: Scalable inference-time alignment of diffusion models. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.
- Purvish Jajal, Nick John Eliopoulos, Benjamin Shiue-Hal Chou, George K Thiruvathukal, James C Davis, and Yung-Hsiang Lu. Inference-Time Alignment of Diffusion Models with Evolutionary Algorithms. *arXiv preprint arXiv:2506.00299*, 2025.
- Stephen Joe and Frances Y. Kuo. Constructing Sobol Sequences with Better Two-Dimensional Projections. *SIAM Journal on Scientific Computing*, 30(5):2635–2654, 2008.
- Sunwoo Kim, Minkyu Kim, and Dongmin Park. Test-time Alignment of Diffusion Models without Reward Over-optimization. In *The Thirteenth International Conference on Learning Representations*, 2025.
- Greg Landrum, Paolo Tosco, Brian Kelley, Ricardo Rodriguez, David Cosgrove, Riccardo Vianello, sriniker, Peter Gedeck, Gareth Jones, Eisuke Kawashima, NadineSchneider, Dan Nealschneider, tadhurst cdd, Andrew Dalke, Matt Swain, Brian Cole, Samo Turk, Aleksandr Savelev, Niels Maeder, Alain Vaucher, Maciej Wójcikowski, Hussein Faara, Ichiru Take, Rachel Walker, Vincent F. Scalfani, Daniel Probst, Kazuya Ujihara, Axel Pahl, guillaume godin, and Juuso Lehtivarjo. rdkit/rdkit: 2025_09_5 (Q3 2025) Release, January 2026.
- Xiner Li, Masatoshi Uehara, Xingyu Su, Gabriele Scalia, Tommaso Biancalani, Aviv Regev, Sergey Levine, and Shuiwang Ji. Dynamic Search for Inference-Time Alignment in Diffusion Models. *arXiv preprint arXiv:2503.02039*, 2025a.
- Xiner Li, Yulai Zhao, Chenyu Wang, Gabriele Scalia, Gökçen Eraslan, Surag Nair, Tommaso Biancalani, Shuiwang Ji, Aviv Regev, Sergey Levine, and Masatoshi Uehara. Derivative-Free Guidance in Continuous and Discrete Diffusion Models with Soft Value-based Decoding. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025b.
- Shanchuan Lin, Anran Wang, and Xiao Yang. SDXL-Lightning: Progressive Adversarial Diffusion Distillation, 2024a.
- Yeqing Lin, Minji Lee, Zhao Zhang, and Mohammed AlQuraishi. Out of Many, One: Designing and Scaffolding Proteins at the Scale of the Structural Universe with Genie 2, 2024b.
- Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, Allan dos Santos Costa, Maryam Fazel-Zarandi, Tom Sercu, Salvatore Candido, and Alexander Rives. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.
- Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow Matching for Generative Modeling. In *The Eleventh International Conference on Learning Representations*, 2023.
- Xingchao Liu, Chengyue Gong, and qiang liu. Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow. In *The Eleventh International Conference on Learning Representations*, 2023.

- Nanye Ma, Shangyuan Tong, Haolin Jia, Hexiang Hu, Yu-Chuan Su, Mingda Zhang, Xuan Yang, Yandong Li, Tommi Jaakkola, Xuhui Jia, and Saining Xie. Scaling Inference Time Compute for Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2523–2534, June 2025.
- Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning Transferable Visual Models From Natural Language Supervision. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8748–8763. PMLR, 18–24 Jul 2021.
- Raghunathan Ramakrishnan, Pavlo O. Dral, Matthias Rupp, and O. Anatole von Lilienfeld. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific Data*, 1(1):140022, 2014.
- Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-Resolution Image Synthesis With Latent Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10684–10695, June 2022.
- Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Raphael Gontijo-Lopes, Burcu Karagol Ayan, Tim Salimans, Jonathan Ho, David J. Fleet, and Mohammad Norouzi. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho (eds.), *Advances in Neural Information Processing Systems*, 2022.
- Raghav Singhal, Zachary Horvitz, Ryan Teehan, Mengye Ren, Zhou Yu, Kathleen McKeown, and Rajesh Ranganath. A General Framework for Inference-time Scaling and Steering of Diffusion Models. In *Forty-second International Conference on Machine Learning*, 2025.
- I.M Sobol’. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7(4):86–112, 1967. ISSN 0041-5553.
- Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-Based Generative Modeling through Stochastic Differential Equations. In *International Conference on Learning Representations*, 2021.
- Yuxuan Song, Jingjing Gong, Minkai Xu, Ziyao Cao, Yanyan Lan, Stefano Ermon, Hao Zhou, and Wei-Ying Ma. Equivariant Flow Matching with Hybrid Probability Transport for 3D Molecule Generation. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.
- Kim Yong Tan, Yueming Lyu, Ivor Tsang, and Yew-Soon Ong. Fast Direct: Query-Efficient Online Black-box Guidance for Diffusion-model Target Generation. In *The Thirteenth International Conference on Learning Representations*, 2025.
- Zhiwei Tang, Jiangweizhi Peng, Jiasheng Tang, Mingyi Hong, Fan Wang, and Tsung-Hui Chang. Inference-Time Alignment of Diffusion Models with Direct Noise Optimization, 2025.
- Michel van Kempen, Stephanie S. Kim, Charlotte Tumescheit, Milot Mirdita, Jeongjae Lee, Cameron L. M. Gilchrist, Johannes Söding, and Martin Steinegger. Fast and accurate protein structure search with Foldseek. *Nature Biotechnology*, 42(2):243–246, 2024.
- Luran Wang, Chaoran Cheng, Yizhen Liao, Yanru Qu, and Ge Liu. Training Free Guided Flow-Matching with Optimal Control. In *The Thirteenth International Conference on Learning Representations*, 2025.

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.

Jie Wu, Yu Gao, Zilyu Ye, Ming Li, Liang Li, Hanzhong Guo, Jie Liu, Zeyue Xue, Xiaoxia Hou, Wei Liu, Yan Zeng, and Weilin Huang. RewardDance: Reward Scaling in Visual Generation, 2025.

Xiaoshi Wu, Yiming Hao, Keqiang Sun, Yixiong Chen, Feng Zhu, Rui Zhao, and Hongsheng Li. Human Preference Score v2: A Solid Benchmark for Evaluating Human Preferences of Text-to-Image Synthesis, 2023.

Jiazhen Xu, Xiao Liu, Yuchen Wu, Yuxuan Tong, Qinkai Li, Ming Ding, Jie Tang, and Yuxiao Dong. ImageReward: Learning and Evaluating Human Preferences for Text-to-Image Generation. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. GeoDiff: A Geometric Diffusion Model for Molecular Conformation Generation. In *International Conference on Learning Representations*, 2022.

A APPENDIX

This appendix provides supplementary material to support the results presented in the main text. We begin by detailing the hyperparameters and implementation of TRS in Appendix B, followed by expanded experimental settings in Appendix C. We then present additional ablation studies (Appendix D) and discuss the comparative impact of ODE versus SDE sampling (Appendix E). Furthermore, we provide a comprehensive overview of all baselines (Appendix F), offer deeper insights into the evaluation metrics (Appendix G), and report the runtime for all methods (Appendix H). Finally, we address the limitations of our approach in Appendix I and showcase random samples from the image experiments in Appendix J.

B ALGORITHM DETAILS AND HYPERPARAMETERS

In this section we explain the remaining details about our method of Section 3.2 and show all hyperparameter details from the experiments in Section 4 for TRS and across all baselines.

Restart logic. When the failure tolerance threshold c_{fail} is triggered while the trust-region side length is already at its minimum value $\ell_j^{\min}$, a region restart is triggered. In this case, the side length is reset, and the region is subsequently re-centered at one of the globally best points observed so far. This restart mechanism allows the algorithm to escape from locally saturated regions and to continue allocating evaluations toward more promising areas of the search space.

Length-dependent perturbation constraints. To maintain stable behavior across different trust-region scales, we apply simple constraints that couple the side length ℓ_j with the perturbation probability $p_{j,b}$ used for coordinate masking $\mathbf{m}_{j,b}$, by rejection sampling. These constraints limit the number of perturbed dimensions when the trust region is large, which is particularly important in high-dimensional noise spaces.

- If $\ell_j \geq 2.0 \rightarrow p_j \leq 0.2$.
- If $\ell_j \geq 1.6 \rightarrow p_j \leq 0.5$.
- If $\ell_j \geq 1.2 \rightarrow p_j \leq 0.7$.

These constraints were found to improve robustness without introducing additional tuning complexity. We provide an intuitive visualization for those rules in Figure 10, showing under which combinations of mask and region length the perturbations fail.

Hyperparameters. While TRS contains several hyper-parameters, we find them to be robust across all the experiments we conduct in Section 4. We show this in Table 4, where we can see that the number of regions is the only one we change, which is due to the different batch sizes we use for efficiency. But even this shows good robustness, which we show in an ablation in Figure 11.

Table 4: Hyperparameters for all our experiments in section 4. When entries are separated with a slash / , we refer to the difference between SD1.5 / SDXL. The dagger[†] means that is the setting the entry refers to the highest NFE setting in the respective experiment in section 4. The asterix * excludes OC-Flow and DTS.

Hyperparameter	T2I	Moleculs	Protein
<i>All algorithms</i>			
Noise Dimension D	16384 / 65536	291	$3 \times n_{\text{res}}$
Batchsize B^*	20	100	8
Inference steps	50/8	50	400
OC-Flow (Wang et al., 2025)			
Batch size B	1	1	–
Step size η	0.25	–	–
Weight decay	0.998	0.995	–
Weight constraint	0.4	0.4	–
Number of steps	50	50	–
Optimizer	SGD	L-BFGS	–
Learning rate α	1.0	1.0	–
DTS* (Jain et al., 2025)			
Expansion steps	[50, 40, 30, 20, 10] / [8, 6, 4, 2]	–	–
Exploration Constant λ	0.1 (IR), 0.01 (HPS)	–	–
Progressive Width Constant C	2.0	–	–
Progressive Width α	0.4	–	–
Exploration type	UCB	–	–
Zero-Order Search (Ma et al., 2025)			
Added noise ϵ	0.1	0.1	0.1
Fast Direct (Tan et al., 2025)			
Step size α	80	80	80
Total steps T	6 [†]	6 [†]	6 [†]
Noise sigma (GP) σ_n	0.1	0.1	0.1
TRS (Ours)			
Perturbations	Sobol / Gaussian	Sobol	Sobol
# Trust regions k	15	20	5
Initial Trust Region length l_{init}	0.8	0.8	0.8
Minimal Length l_{min}	0.05	0.05	0.05
Maximal Length l_{max}	2.4	2.4	2.4
Length Update factor α_ℓ	1.5	1.5	1.5
Success counter threshold c_{succ}	3	3	3
Failure counter threshold c_{fail}	3	3	3
Fraction warm-up iterations	20%	20%	20%
Min. prob. for masks p_{min}	0.1	0.1	0.1
Max. prob. for masks p_{max}	0.9	0.9	0.9

C EXPERIMENT DETAILS

In this section, we provide additional details of the experiments in Section 4 with a specific focus on the generative models and the reward functions that are applied. We provide the links to all external resources in Table 5.

Table 5: Summary of generative models and reward functions for the experiments in Section 4.

Domain	Model / Tool	Params	Source / Repository
T2I	SD v1.5	~860M	HF: SD-v1.5
	SDXL-Lightning	~2.6B	HF: SDXL-Lightning
	ImageReward	~446M	GitHub: ImageReward
	HPSv2	~986M	GitHub: HPSv2
	Aesthetic Pred.	~428M	GitHub: Aesthetic
Molecules	EquiFM	~22M	GitHub: EquiFM / OC-Flow
Proteins	Proteina ($\mathcal{M}_{\text{FS}}^{\text{small}}$)	~60M	GitHub: Proteina
	ProteinMPNN	~1.7M	GitHub: ProteinMPNN
	ESMFold-v1	~3B	GitHub: ESMFold

C.1 TEXT-TO-IMAGE

Generative models. We use Stable Diffusion (Rombach et al., 2022) v1.5 (~859.5M parameters), which generates $3 \times 512 \times 512$ images from a noise space of $4 \times 64 \times 64$, and a distilled version of the larger SDXL Lin et al. (2024a) (~2.57B parameters), which generates $3 \times 1024 \times 1024$ images from a noise space of $4 \times 128 \times 128$. Both models encode text prompts using CLIP-based text encoders (Radford et al., 2021). Classifier-free guidance (Dhariwal & Nichol, 2021) is applied during sampling by linearly combining conditional and unconditional model predictions with conditioning signal c . For image sampling, we use the DDIM scheduler. We set $\eta = 0$ (deterministic sampling) for black-box source noise optimization methods and $\eta = 1.0$ (stochastic sampling) for noise sequence optimization methods. The number of inference steps is set to 50 for SD1.5 and 8 for SDXL, and all experiments are conducted in `float16` precision.

Reward functions. We use three reward functions for the text-to-image experiments, each defined by the scalar output of a pretrained model. ImageReward (Xu et al., 2023) and HPSv2 (Wu et al., 2023) evaluate image-prompt alignment and are defined as $R(\mathbf{x}_1, c)$, where $\mathbf{x}_1$ denotes the generated image and c the text prompt. In contrast, the Aesthetic predictor (?) evaluates only the generated image and is defined as $R(\mathbf{x}_1)$.

C.2 MOLECULE GENERATION

Generative model. We use EquiFM (Song et al., 2023) (~22M parameters) to generate small QM9 molecules and base our implementation on the code of OC-Flow (Wang et al., 2025). EquiFM is an ODE-based flow matching model that operates on a combined continuous Gaussian noise representation for both atomic coordinates and atom types, with total dimensionality $M = 291$. The number of atoms per molecule ranges from 3 to 29. To accommodate variable-sized molecules while keeping a fixed input dimension, unused atom entries are masked to zero. In the experiments described in Section 4.2, the number of atoms is sampled according to the empirical distribution in the QM9 training set, which most frequently yields molecules with 15 to 20 atoms.

Reward functions. The reward is defined as the negative sum of absolute deviations between target property values and the corresponding predictions of pretrained regression models. For the R_6 reward, we consider the properties $\{\alpha, \mu, \Delta\epsilon, \epsilon_{\text{HOMO}}, \epsilon_{\text{LUMO}}, c_v\}$, corresponding to isotropic polarizability, dipole moment, HOMO-LUMO gap, HOMO energy, LUMO energy, and heat capacity. For the R_3 reward, we restrict the objective to the subset $\{\alpha, \mu, \epsilon_{\text{LUMO}}\}$.

C.3 PROTEIN DESIGN

Generative model. For protein design, we use the $\mathcal{M}_{\text{FS}}^{\text{small}}$ variant of Proteina (Geffner et al., 2025) with 400 integration steps, which consists of approximately 60M transformer parameters and does not include triangle layers. The source noise is drawn from a continuous Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$ with dimensionality $n_{\text{res}} \times 3$, corresponding to the 3D coordinates of n_{res} residues.

Once the model has learned the ODE dynamics, sampling can be performed either via deterministic ODE integration, $d\mathbf{x}_t = \mathbf{v}_t^\theta(\mathbf{x}_t, \tilde{c}) dt$, or using stochastic SDE sampling, $d\mathbf{x}_t = [\mathbf{v}_t^\theta(\mathbf{x}_t, \tilde{c}) - \gamma g(t) \mathbf{s}_t^\theta(\mathbf{x}_t, \tilde{c})] dt + \sqrt{2\gamma g(t)} d\mathbf{w}_t$. We discuss the implications of these sampling choices in more detail in Appendix E. In practice, setting $\gamma < 1$ improves designability at the expense of diversity and novelty.

Reward functions. Following (Geffner et al., 2025), we optimize the designability of generated 3D protein backbones. Given a generated structure, a pretrained inverse folding model, ProteinMPNN (Dauparas et al., 2022) ($\sim 1.7\text{M}$ parameters), is used to design a compatible amino acid sequence. This sequence is subsequently processed by a folding model, for which we use the transformer-based ESMFold-v1 (Lin et al., 2023) ($\sim 3\text{B}$ parameters), to predict its 3D structure. We compare the predicted structure to the original generated backbone by computing the self-consistency root mean square deviation (scRMSD). The scalar reward is then defined as $R(\mathbf{x}_1) = \exp(-\text{scRMSD}(\mathbf{x}_1))$, which maps the score to the range $(0, 1]$, with values closer to 1 indicating higher designability. Note that usually 8 sequences are extracted and folded and the best of those is chosen, but for computational efficiency we only extract 1 per protein backbone.

D ABLATIONS

Here we perform further ablation studies to justify the design choices in Section 3.2 and hyperparameters in Appendix B. Most importantly, Appendix D.1 emphasises the importance of our top- k center selection instead of keeping them strictly apart. Appendix D.2 provides visual and quantitative insights into the relationship of the trust-region lengths ℓ_j and the mask probabilities p_j , while Appendix D.3 shows that 20% is a good heuristic for the fraction of warm-up budgets across multiple total budgets. Finally, Appendix D.4 shows that even the number of regions k , is a robust hyperparameter, which is the only one that we change across our experiments in Section 4.

D.1 CENTER SELECTION

Setup. We investigate four center selection strategies for text-to-image generation using Stable Diffusion 1.5 (SD1.5) and ImageReward as reward function. *LocalBest* and *LocalLastIter* enforce strict region separation, updating centers using either the historical best or the most recent best sample within each region $\mathcal{T}_j$, respectively. In contrast, *GlobalTopk* and *GlobalLastIter* allow interaction by re-centering regions globally based on the overall top- k samples observed so far or in the latest iteration. All experiments use a batch size $B = 20$ and 20 iterations ($20k$ NFE), consistent with the configuration in Section 4.1.

Metrics. Performance is evaluated on the full DrawBench benchmark. We use the mean best ImageReward as the primary metric to rank the generated noise samples and determine center updates. Random search is included as baseline as a reference to compare the relative improvement of each selection strategy.

Results. As shown in Figure 9, *GlobalTopk* consistently outperforms all other strategies. The results indicate that enforcing strict separation between regions (as in the original TuRBO algorithm (Eriksson et al., 2019)) limits performance in this setting. Conversely, global re-centering enables a more effective allocation of the evaluation budget toward the most promising areas of the search space, significantly improving generation quality.

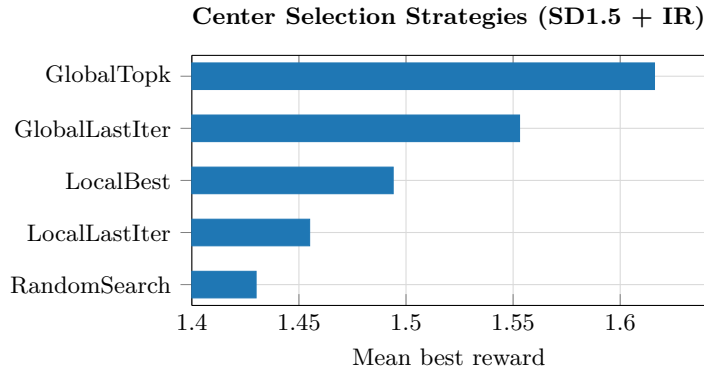


Figure 9: Effect of center selection strategy on DrawBench using SD1.5 and ImageReward. We compare the mean best rewards across the prompts in the benchmark.

D.2 TRUST REGION DYNAMICS: INTERACTION OF LENGTH AND MASKS

Setup. We perform a grid ablation across varying initial trust region lengths ℓ_{init} and masking sparsity ranges defined by p_{min} and p_{max} . First, we do this quantitatively by investigating different combinations of initial trust-region lengths ℓ_{init} and set the mask probabilities in different ranges, with different min and max probabilities $p_{\text{min}}, p_{\text{max}}$. This experiment is conducted using SD1.5 on a representative subset of 55 prompts from DrawBench, covering all categories of the original benchmark to ensure broad coverage of the latent space. Additionally, we visualize the relationship of the trust-region length ℓ_j and sampled mask probability p_j in Figure 10.

Metrics. In Figure 10 we aim to track the point at which the samples begin to exhibit structural degradation and visualize these noising effects, while we show the mean best ImageReward for the quantitative experiment in Table 6.

Results. Figure 10 reveals a clear stability threshold at approximately $\ell_j \approx 1.6$, beyond which images show visible noise. However, we find that low-probability masks (e.g., $p_j = 0.05$) act as a regularizer and they allow for significantly higher exploration lengths, up to $\ell_j = 6.4$, while maintaining valid, coherent structures. Table 6 reveals that the exact masking range is not particularly important, it works well for different ranges. The initial shows also good robustness, even when changing it to very high or low numbers, due to the adaptive nature of the algorithm.

Table 6: Mean best rewards across varying trust-region init lengths (ℓ_{init}) and $[p_{\text{min}}, p_{\text{max}}]$ settings for SD1.5 and ImageReward.

TR init length (ℓ_{init})	Mask probabilities $[p_{\text{min}}, p_{\text{max}}]$			
	[0.05, 0.50]	[0.10, 0.90]	[0.30, 0.70]	[0.50, 1.00]
0.2 (Low)	1.6223	1.6430	1.6353	1.6564
0.8 (Medium)	1.6928	1.6761	1.6820	1.6543
2.0 (Large)	1.6404	1.6197	1.6306	1.6185

D.3 WARM-UP

Setup. We investigate the impact of the warm-up phase on performance to determine the optimal allocation of the total compute budget between initialization and the main optimization process. This experiment uses Stable Diffusion 1.5 on the subset of 55 DrawBench prompts, equally distributed across all categories. We evaluate three total evaluation budgets ($N_{\text{total}} \in \{120, 360, 720\}$) with a fixed batch size of $B = 24$.

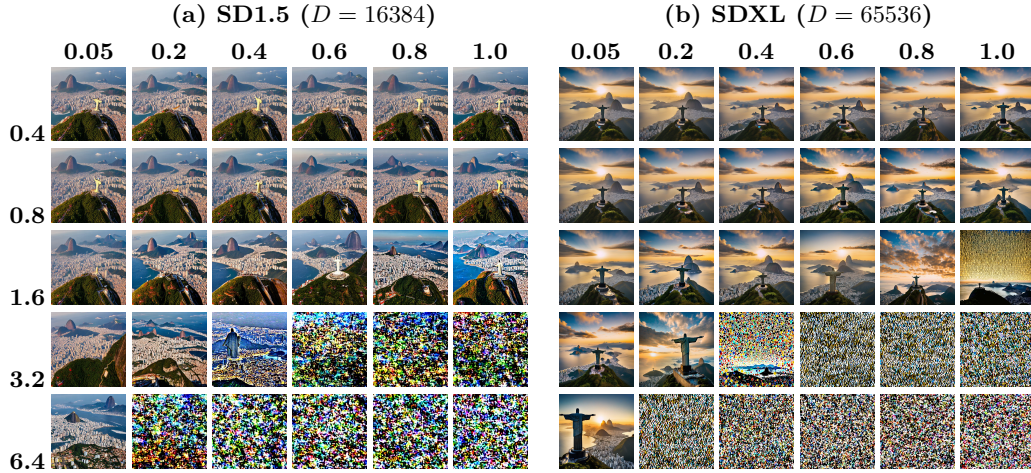


Figure 10: TRS perturbation landscape comparison: Interaction between perturbation length (rows) and masking probability (columns) for SD1.5 and SDXL. For both we used the prompt: "A breathtaking view from behind the Cristo Redentor (Christ the Redeemer) statue in Rio de Janeiro, Brazil, with layered mountains stretching into the distance and the sparkling Atlantic Ocean clearly visible below with blue color; warm golden-hour light, atmospheric haze, ultra-detailed, cinematic, wide-angle landscape, beautiful and serene and the city below."

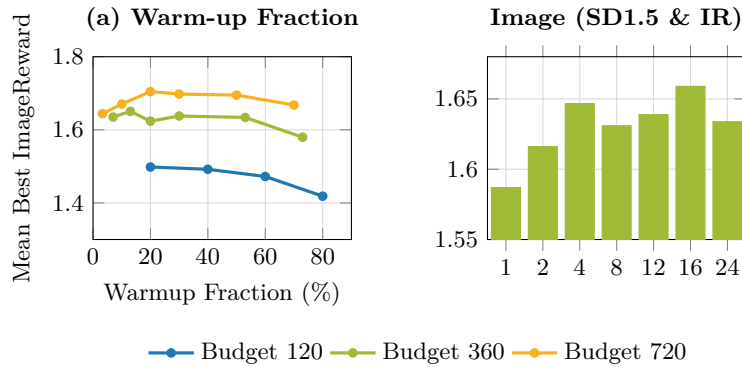


Figure 11: **Ablation Studies.** (a) Impact of the warm-up phase fraction across different compute budgets. (b) Sensitivity of TRS to the number of regions k . Performance remains stable across a wide range of k values.

Metrics. To assess the effectiveness of different budget allocations, we utilize ImageReward as the reward function. We report the final mean best rewards achieved across all 55 prompts, tracking how the optimization efficiency changes as the ratio of warm-up to main optimization varies.

Results. As shown in Figure 11a, TRS achieves peak performance when the warm-up phase accounts for 10% to 20% of the total compute budget across all tested settings. We note that while performance is sensitive to very small warm-up budgets, it degrades only slowly when the warm-up phase is slightly larger than the optimal range. This suggests that the exact choice is not overly critical, provided the initialization period is sufficient; consequently, we set the warm-up fraction to 20% for all main experiments.

D.4 NUMBER OF REGIONS

Setup. We ablate the impact of the number of trust regions $k \in \{1, 2, 4, 8, 16, 24\}$ on the optimization performance for text-to-image generation. These experiments use SD1.5 on a

55-prompt subset of DrawBench, utilizing a fixed batch size of $B = 24$. The optimization process is partitioned into 3 warm-up iterations followed by 12 trust-region iterations.

Metrics. Consistent with our previous ablations, we align the optimization to ImageReward. We evaluate the final mean best rewards across the prompt subset to identify the optimal range for k and to assess how the number of regions interacts with the total evaluation budget and batch size.

Results. As illustrated in Figure 11b, using a very small number of regions leads to sub-optimal performance. However, we find that selecting k in a medium range between 1 and the batch size B consistently yields strong results. These findings suggest that k is not a highly sensitive hyperparameter, as performance remains stable across a wide range of values, indicating that it does not require extensive per-task tuning. Since k is the only hyperparameter that was changed between the main experiments in Section 4, we conclude that TRS is a highly robust algorithm for noise optimization of flow and diffusion models.

E ODE vs. SDE SAMPLING DYNAMICS

TRS is a black-box optimization method that relies on local exploration to find better samples. This approach is more effective with ODE sampling because the deterministic paths provide a clear sense of locality. However, since SDE sampling with noise reduction leads to higher designability in modern generative 3D protein models (Geffner et al., 2025; Bose et al., 2024), we want to see how our algorithm handles the trade-off between the designability gains of SDEs and the diversity preserved by ODEs. We also include experiments for text-to-Image (T2I) generation to see if these sampling dynamics remain consistent across different model types.

Setup. We compare TRS against standard random search (RS) in both deterministic and stochastic settings. For protein generation, we generate 100 proteins of length 50, similar to Section 4.3 optimizing specifically for designability. For T2I generation, we use the same setup as in Section 4.1, utilizing the DDIM scheduler with both a stochastic setting ($\eta = 1.0$) and a deterministic setting ($\eta = 0.0$). A key difference between these setups is the inference budget: protein uses 400 steps, while the T2I model uses 50.

Metrics. We evaluate the protein results based on designability, diversity, and novelty and the Rank-1 percentage and show the results in Table 7. With this last metric, we want to show not only into how many clusters the proteins fall in this analysis, but also what the distribution between those clusters is. Rank-1 is the percentage of proteins that fall into the most dominant cluster. In more detail Figure 12 shows the distribution across the 10 most dominant clusters using rank-size and cumulative mass plots to show how concentrated the generated samples are.

Results and Discussion. As shown in Table 7, SDE sampling does lead to higher designability, but it significantly diminishes diversity and novelty. When we compare TRS with ODE to random search and TRS with SDE ($\gamma = 0.6$), we observe that the designability metrics are on par, while TRS with ODE shows much better diversity and novelty metrics. This becomes especially clear when observing the Rank-1 metric, where we see that SDE-optimized samples, for which 54 or even 68% of the samples belong to the same cluster. Figure 12 further illustrates the mode collapse of SDE-based sampling in protein design, where we can see that in SDE-based optimization, the designable samples are highly concentrated at the few most dominant clusters. ODE-based optimization instead shows a healthier distribution, creating structures distributed at different clusters more equally.

We also find that most current SDE-based scaling methods are not practical for protein, either because they lack necessary value estimations (Li et al., 2025b; Singhal et al., 2025) or because they are too slow without efficient batch utilization (Jain et al., 2025), especially, when many inference steps are necessary and the reward function is expensive.

Interestingly, TRS shows a smaller advantage over random search in proteina when using SDEs compared to T2I. This is likely because the high number of inference steps in proteina allows the stochastic noise to eventually override the local search signal provided by the trust region.

Table 7: Comparison of ODE and SDE variants for random search (RS) and our trust-region search (TRS) in protein design and text-to-image (T2I). ImageReward and designability are the respective alignment rewards and Rank-1, Cluster Divergence, pairwise TM-Score and PDB Novelty are additional metrics to uncover the effects of SDE vs. ODE based alignment with the 400-step proteina model. In T2I we use the 50-step SD1.5 as generative model.

Algorithm	T2I	Protein (SDE with $\gamma = 0.6$)				
	ImageReward $\uparrow$	Design. $\uparrow$	Rank-1 $\downarrow$	Clust. Div. $\uparrow$	TM Div. $\downarrow$	Nov. $\downarrow$
RS + ODE	1.43	0.53	<u>37</u>	<u>0.27</u>	0.68	0.85
RS + SDE	1.45	0.66	68	0.19	0.73	0.94
TRS (Ours) + ODE	1.62	0.65	18	0.30	0.67	0.89
TRS (Ours) + SDE	<u>1.52</u>	0.66	54	<u>0.27</u>	<u>0.70</u>	0.92

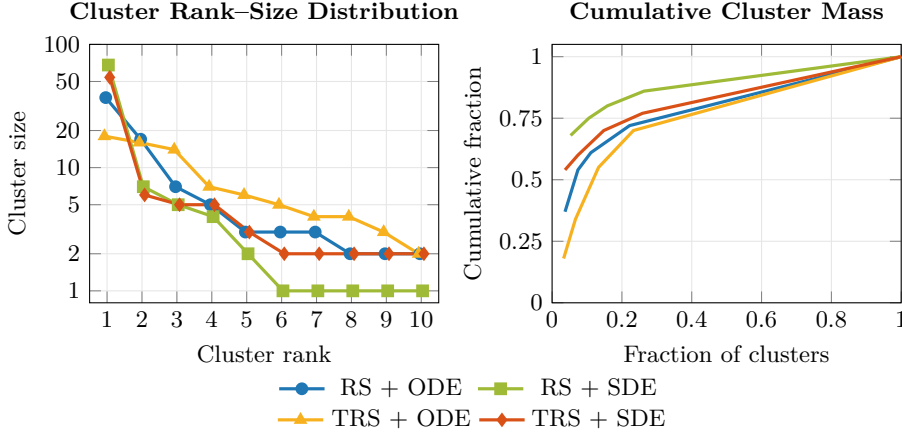


Figure 12: Protein cluster distributions. The rank-size plot (left) shows that ODE methods maintain higher diversity, whereas SDE methods suffer from mode-collapse. The cumulative mass (right) confirms that TRS + ODE provides the most balanced distribution of samples.

F BASELINES

In this section, we provide more detail about the baselines and explain their hyperparameters. The exact numbers used in the experiments are shown in Table 4.

F.1 GRADIENT-BASED GUIDANCE

Optimal Control Flow (OC-Flow). OC-Flow (Wang et al., 2025) provides a theoretically grounded, training-free framework for guided flow matching by framing generation as an optimal control problem. It augments pre-trained flow dynamics with a time-dependent control term $\mathbf{u}_t$: $\dot{\mathbf{x}}_t = \mathbf{v}_t(\mathbf{x}_t) + \mathbf{u}_t$. The framework seeks to minimize a cost functional $J(\mathbf{u}) = R(\mathbf{x}_1) + \int_0^1 \frac{1}{2\lambda} \|\mathbf{u}_t\|^2 dt$, comprising a terminal reward loss $R(\mathbf{x}_1)$ and a quadratic running cost that regulates the trajectory’s deviation from the prior distribution. Leveraging Pontryagin’s Maximum Principle, the control trajectory is optimized iteratively over a fixed number of steps using either SGD or L-BFGS. In each iteration, gradient information is propagated backward via a co-state flow $\boldsymbol{\mu}_t$ to update the control parameters with a step size η and weight decay. To ensure trajectory regularity and stable convergence, a weight

constraint is enforced on the magnitude of the control term throughout the optimization process.

F.2 NOISE SEQUENCE SEARCH

Diffusion Tree Sampling (DTS). DTS (Jain et al., 2025) frames the inference-time alignment of diffusion models as a tree-structured optimization problem over the denoising sequence. The framework employs a recursive value-based search guided by a soft value function $V(\mathbf{x}_t)$, which is estimated via a soft-Bellman backup with an exploration constant λ :

$$V(\mathbf{x}_t) = \frac{1}{\lambda} \log \mathbb{E}_{\mathbf{x}_{t-1} \sim p_{\theta}(\cdot|\mathbf{x}_t)} [\exp(\lambda V(\mathbf{x}_{t-1}))]$$

At each node, the search proceeds for a fixed number of expansion steps, where the selection of trajectories is governed by a specific exploration type (e.g., UCT). To effectively navigate the continuous branching space of the diffusion process, the algorithm utilizes progressive widening to determine the number of children k for a node $\mathbf{x}_t$ based on its visit count $N(\mathbf{x}_t)$:

$$k(\mathbf{x}_t) = \lceil C \cdot N(\mathbf{x}_t)^\alpha \rceil$$

where C is the progressive width constant and α is the progressive width exponent. A stochastic rollout parameter ρ serves as a decision gate: with probability ρ , the algorithm performs a full rollout to the terminal state $t = 0$ to obtain an exact reward, while with probability $1 - \rho$, it continues recursive tree expansion.

Fast Direct. Fast Direct (Tan et al., 2025) is a trajectory-level optimization method designed for black-box guidance and optimizes the complete noise sequence $\{\epsilon_t\}_{t=1}^T$ simultaneously. The method identifies a pseudo-target $\mathbf{x}^*$ on the manifold via Gaussian Process (GP) regression over previous evaluations and computes a universal direction to update the entire sequence:

$$\epsilon_t^{\text{new}} = \text{Norm}(\epsilon_t^{\text{old}} + \alpha(\mathbf{x}^* - \mathbf{x}_0))$$

where α is the step size. This global refinement allows it to converge to high-reward regions in fewer iterations than per-step filtering methods.

F.3 SOURCE NOISE SEARCH

Random Search. Random search is the simplest search-based method one can apply, where N source noises $\mathbf{x}_0$ are randomly sampled and evaluated. Thus, it is often referred to as best-of- N in the literature Jain et al. (2025). This technique can be used both with ODE- and SDE-based samplers.

Zero-Order Search. Zero-order Search (Ma et al., 2025) can be seen as a special case of our algorithm. Like ours, it iteratively refines the initial latent noise to maximize a target reward. The algorithm begins by sampling an initial set of B Gaussian noise vectors $\{\mathbf{x}_0^j\}_{j=1}^B$ and selecting the candidate that yields the highest reward as the initial center $\mathbf{x}_0^c$. This is equivalent to using only one warm-up iteration in TRS and setting the number of regions $k = 1$. In each subsequent search iteration, B new candidates are generated by perturbing the current center:

$$\mathbf{x}_0^{\text{new}} = \mathbf{x}_0^c + \epsilon \cdot \boldsymbol{\delta}, \quad \boldsymbol{\delta} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

where ϵ (scalar) controls the radius of the local neighborhood search. This perturbation scheme is similar to using TRS without probability masks and choosing a fixed region length. Similar to ours, this technique works best for ODE samplers.

G METRICS

In this section, we describe all additional evaluation metrics used in this work that are not employed as reward functions. The reward-based metrics are detailed in Appendix C. We consider evaluation criteria for both molecule generation and protein design.

G.1 MOLECULE GENERATION

For the evaluation of 3D molecule generation, we follow the methodology described by (Song et al., 2023), which assesses the physical and chemical plausibility of the generated Cartesian coordinates and atom types.

Stability. We evaluate the structural integrity of the generated samples using *Atom Stability* and *Molecule Stability*. Following the convention of Hoozeboom et al. (Hoozeboom et al., 2022), chemical bonds are inferred based on the Euclidean distances between atoms using a threshold-based lookup table of covalent radii. An individual atom is considered stable if its inferred bond count matches its expected chemical valency (e.g., 4 for carbon, 1 for hydrogen). *Molecule Stability* is then defined as the percentage of generated molecules for which all constituent atoms are stable. This metric serves as a proxy for the geometric consistency of the generated 3D structures.

Validity, Uniqueness, and Novelty. The chemical validity of a molecule is determined by its ability to be successfully parsed into a molecular graph using RDKit (Landrum et al., 2026). We report the *Valid Fraction*, the proportion of samples satisfying fundamental valency constraints. The *Unique Fraction* indicates the percentage of non-redundant molecules among generated compounds. We also report Novelty, defined as the fraction of generated molecules not present in the training set. To provide a stringent quality assessment, we track the Molecule Stability Percentage (MSP) and the Valid and Unique Percentage (VUP). The former assesses the geometric consistency of the 3D coordinates, while the latter ensures the model explores the chemical space without collapsing into a few valid structures.

G.2 PROTEIN DESIGN

To evaluate the structural quality and variability of the generated protein backbones, we adopt the metric suite introduced by (Geffner et al., 2025). These metrics assess two complementary aspects: the uniqueness of the generated structures relative to known proteins (*Novelty*) and the structural variability within the generated set (*Diversity*). Following (Geffner et al., 2025), we compute these scores only for designable samples, which means that scRMSD $\leq$ 2.0 Å (equivalently designability $\geq$ 0.1353).

Novelty. Novelty measures the extent to which the model generates protein folds that differ from those observed in existing experimental and predicted structure databases. For each generated, designable backbone structure, we compute the maximum TM-score against all entries in a reference set using Foldseek (van Kempen et al., 2024). We then report the average of these maximum TM-scores across the entire sample set. Lower average maximum TM-scores indicate higher structural novelty. We evaluate novelty with respect to the PDB dataset (Berman et al., 2000).

Diversity. We quantify the internal diversity of the generated, designable backbone samples using two complementary metrics:

1. **Average Pairwise TM-score:** We compute the mean pairwise TM-score between all designable backbone samples for each generated protein length. These values are subsequently aggregated to obtain a global average. Since the TM-score measures structural similarity on a scale from 0 to 1, lower values correspond to higher diversity.
2. **Cluster Ratio:** We cluster the generated backbones using Foldseek with a TM-score threshold of 0.5. The diversity score is defined as the ratio of the number of unique clusters to the total number of designable samples. Because a more diverse set of samples results in a higher number of clusters, higher ratios indicate greater structural variety within the generated set.

H RUNTIME COMPARISON

We analyze the runtime of the experiments described in Section 4 for all considered methods on a single NVIDIA A100-SXM4-40GB. The quantitative comparison of computation time and memory usage is visualized in Figure 13. For these benchmarks, we use the settings of ImageReward for T2I, R_6 for molecule generation, and protein designability for sequences of 50 residues.

We observe that DTS* generally exhibits the highest compute time. This arises because the algorithm’s sequential nature is difficult to combine with fixed-batch-size calls of the generative model. In our benchmarks, this results in DTS* being approximately $4\times$ slower than TRS, which is naturally parallelizable.

For OC-Flow, which operates with a batch size of 1, the computation time is high for SD1.5 due to the requirement of back-propagating through 50 integration steps. Conversely, for the distilled SDXL Lightning, OC-Flow is faster but incurs substantially higher memory usage. In the molecule generation task, OC-Flow is the slowest method, while its memory footprint in this domain is lower than that of TRS and other sampling-based methods, which leverage a larger batch size of 100.

Regarding Fast Direct, we employ slightly higher NFE budgets because their iterative step-increasing mechanism makes it difficult to match an exact NFE while maintaining a constant batch size. Beyond this, their computation times and memory usage are only marginally higher than our baselines, as the overhead from their Gaussian Process (GP) is minimal. Finally, Random Search, Zero-Order Search, and TRS are the most efficient overall, as their runtime consists almost entirely of the generative model forward passes and reward function evaluations.

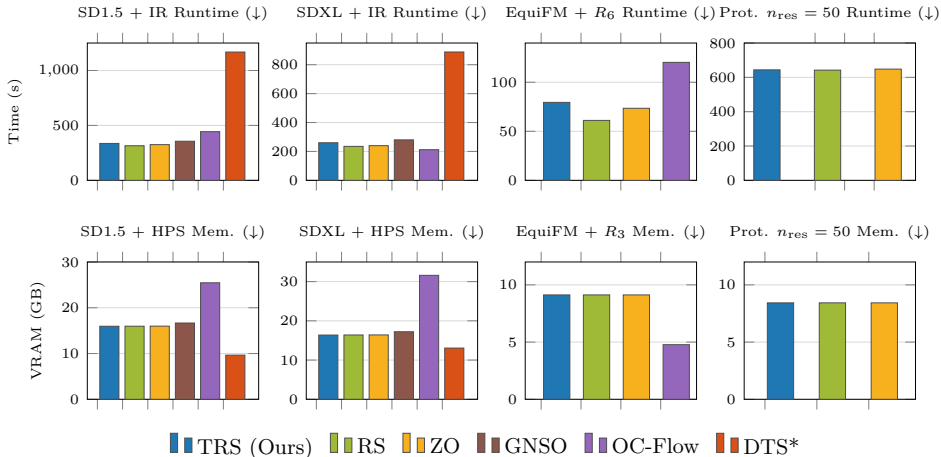


Figure 13: Comparison of runtime and memory consumption across methods on a single NVIDIA A100-SXM4-40GB.

I LIMITATIONS

Reward Functions. A primary limitation of these methods lies in the reliability of reward functions; when derived from neural network predictors, performance is inherently constrained by these models. In our experiments, reliance on pretrained networks for feedback makes optimization susceptible to model bias and failure modes. Figure 14 illustrates this via TRS-optimized SDXL images that receive high rewards despite failing to capture the core prompt intent. This highlights a fundamental challenge: reward models may score samples highly that exploit shortcuts in the learned signal rather than satisfying the true objective. To address this, recent work has focused on larger, more accurate reward models (Wu et al.,

2025). TRS is well-positioned to integrate with these advances, offering a flexible and scalable framework as the field progresses toward more expressive models.

Diversity. While TRS begins with a global exploration phase across multiple promising regions, it ultimately converges to and exploits the most dominant region. Our experimental results demonstrate that this is an effective strategy for identifying a single optimal sample that maximizes the reward objective. However, TRS may be less suited for tasks requiring a diverse ensemble of samples. A direction for future work would be extending the framework to optimize for a diverse set of high-reward samples rather than a single point estimate. For instance, diversity could be explicitly enforced in TRS by incorporating a cosine similarity constraint on the center noise vectors of the trust regions.

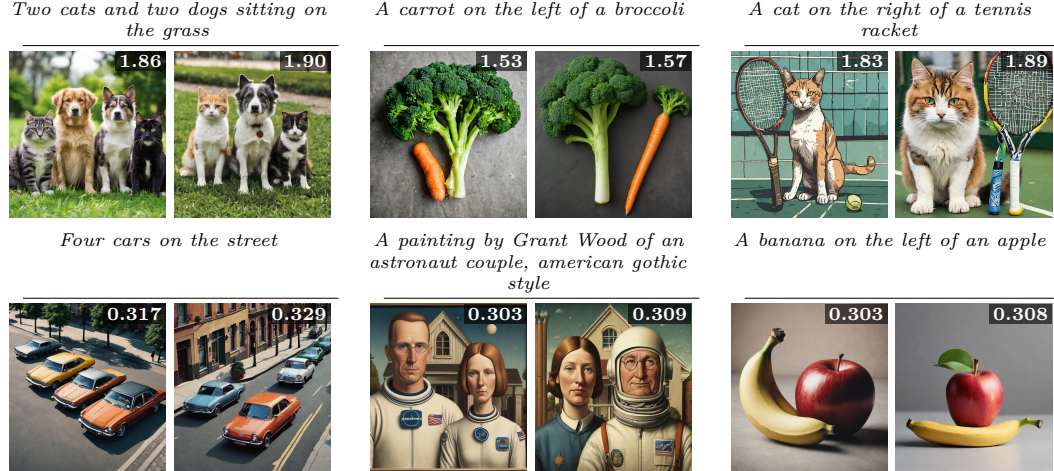


Figure 14: Visualizing reward misalignment: cases where higher reward values are assigned even though the model fails to correctly generate the specific conceptual or spatial requirements of the prompt. All images are generated with SDXL. The upper row shows misalignment to ImageReward and the lower one to HPSv2.

J OPTIMIZED SAMPLES

We randomly sample 10 prompts out of DrawBench and show the qualitative results of the six algorithms we compare in Section 4.1 including our TRS. The images are the best images per prompt from experiment Section 4.1. We provide in total four figures, including Figure 15 for SD1.5 and ImageReward, Figure 16 for SD1.5 and HPSv2, Figure 17 for SDXL-Lightning and ImageReward and Figure 18 for SDXL-lightning optimized for HPSv2.

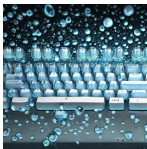
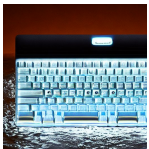
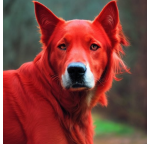
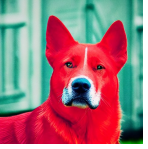
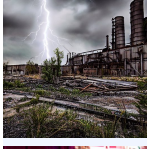
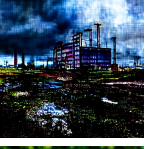
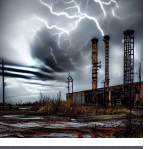
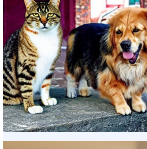
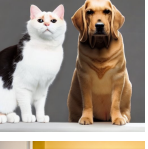
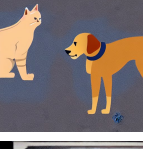
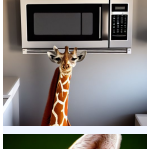
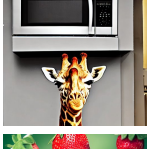
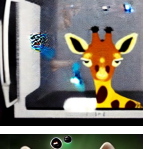
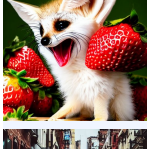
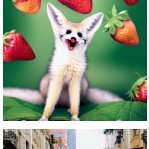
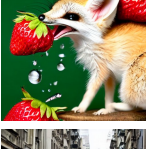
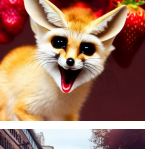
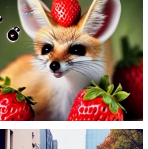
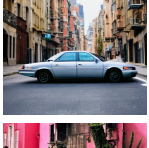
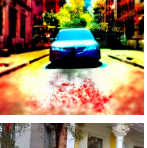
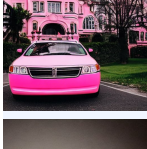
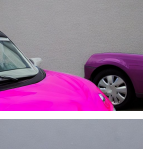
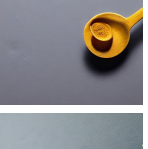
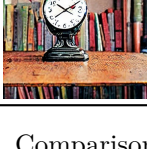
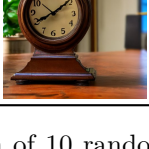
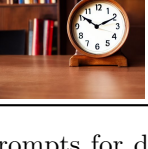
Prompt	TRS (Ours)	RS	ZO	DTS*	FD	OC-Flow
<i>A keyboard made of water, the water is made of light, the light is turned off.</i>						
<i>A red colored dog.</i>						
<i>Hyper-realistic photo of an abandoned industrial site during a storm.</i>						
<i>A cat on the left of a dog.</i>						
<i>A giraffe underneath a microwave.</i>						
<i>A baby fennec sneezing onto a strawberry, detailed, macro, studio light, droplets, backlit ears.</i>						
<i>One car on the street.</i>						
<i>A pink colored car.</i>						
<i>An umbrella on top of a spoon.</i>						
<i>A single clock is sitting on a table.</i>						

Figure 15: Comparison of 10 randomly chosen prompts for different alignment methods using Stable Diffusion v1.5 and ImageReward.

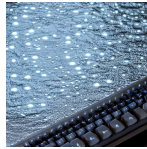
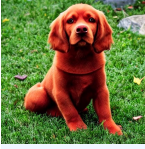
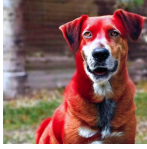
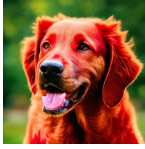
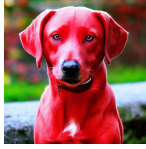
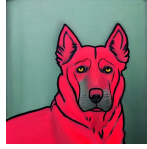
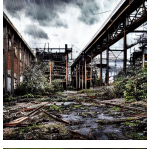
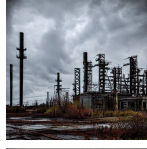
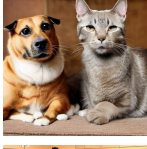
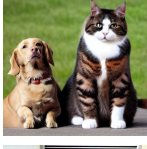
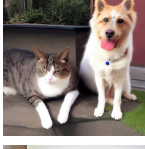
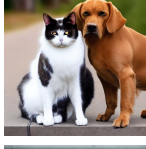
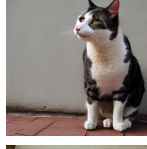
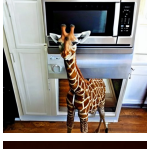
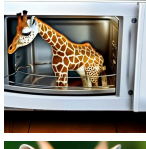
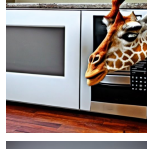
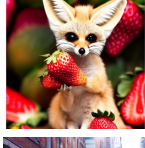
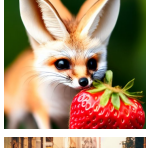
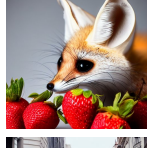
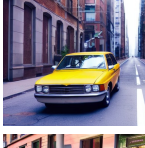
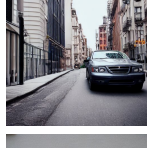
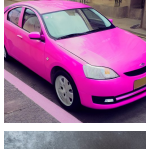
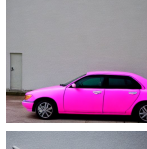
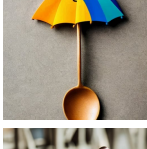
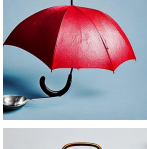
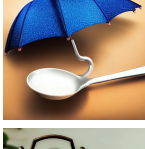
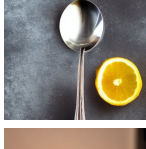
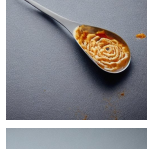
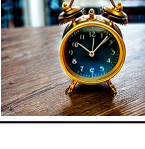
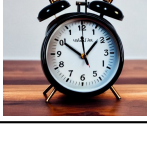
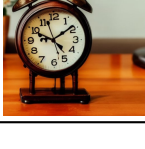
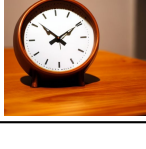
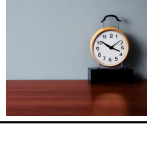
Prompt	TRS (Ours)	RS	ZO	DTS*	FD	OC-Flow
A keyboard made of water, the water is made of light, the light is turned off.						
A red colored dog.						
Hyper-realistic photo of an abandoned industrial site during a storm.						
A cat on the left of a dog.						
A giraffe underneath a microwave.						
A baby fennec sneezing onto a strawberry, detailed, macro, studio light, droplets, backlit ears.						
One car on the street.						
A pink colored car.						
An umbrella on top of a spoon.						
A single clock is sitting on a table.						

Figure 16: Comparison of 10 randomly chosen prompts for different alignment methods using Stable Diffusion v1.5 and HPSv2.

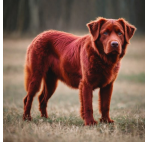
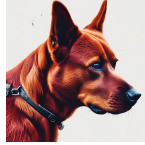
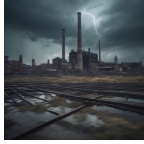
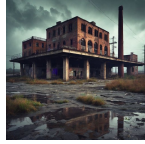
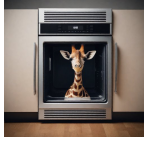
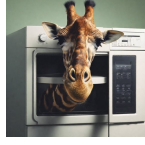
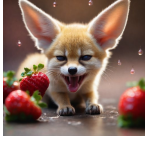
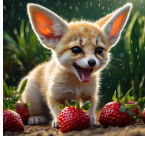
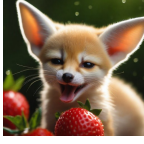
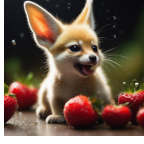
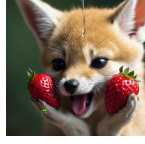
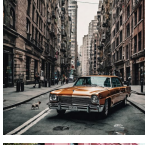
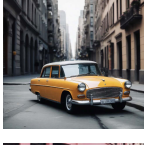
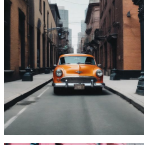
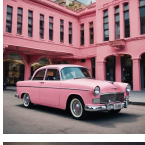
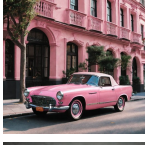
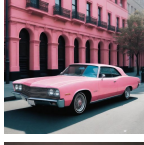
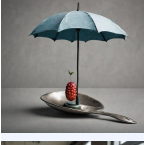
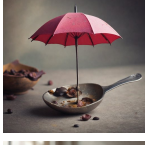
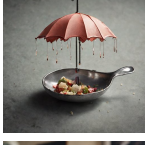
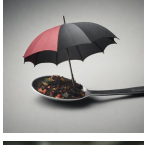
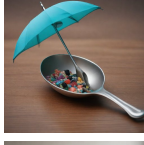
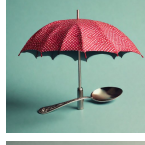
Prompt	TRS (Ours)	RS	ZO	DTS*	FD	OC-Flow
A keyboard made of water, the water is made of light, the light is turned off.						
A red colored dog.						
Hyper-realistic photo of an abandoned industrial site during a storm.						
A cat on the left of a dog.						
A giraffe underneath a microwave.						
A baby fennec sneezing onto a strawberry, detailed, macro, studio light, droplets, backlit ears.						
One car on the street.						
A pink colored car.						
An umbrella on top of a spoon.						
A single clock is sitting on a table.						

Figure 17: Comparison of 10 randomly chosen prompts for different alignment methods using SDXL-Lightning and ImageReward.

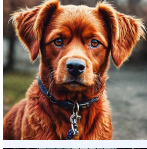
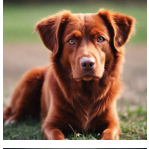
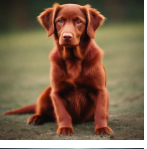
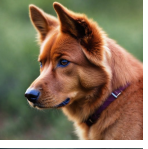
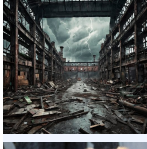
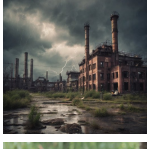
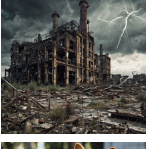
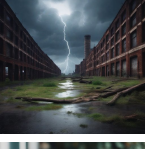
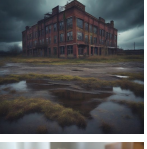
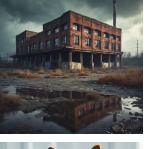
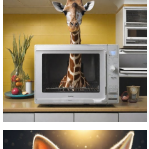
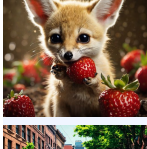
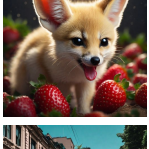
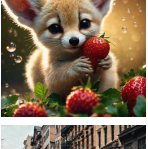
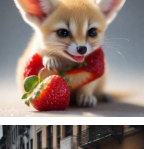
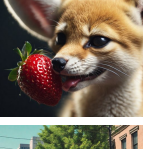
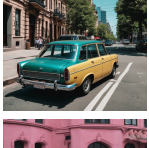
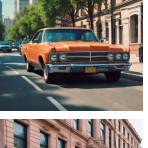
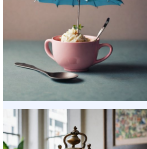
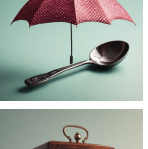
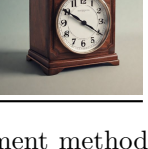
Prompt	TRS (Ours)	RS	ZO	DTS*	FD	OC-Flow
A keyboard made of water, the water is made of light, the light is turned off.						
A red colored dog.						
Hyper-realistic photo of an abandoned industrial site during a storm.						
A cat on the left of a dog.						
A giraffe underneath a microwave.						
A baby fennec sneezing onto a strawberry, detailed, macro, studio light, droplets, backlit ears.						
One car on the street.						
A pink colored car.						
An umbrella on top of a spoon.						
A single clock is sitting on a table.						

Figure 18: Comparison of 10 randomly chosen prompts for different alignment methods using SDXL-Lightning and HPSv2.