DEEP LEARNING JP
[DL Papers]

Training-Free Diffusion Models Alignment with Sampling Demons

Yuta Oshima, Matsuo Iwasawa Lab

Training-Free Diffusion Models Alignment with Sampling Demons

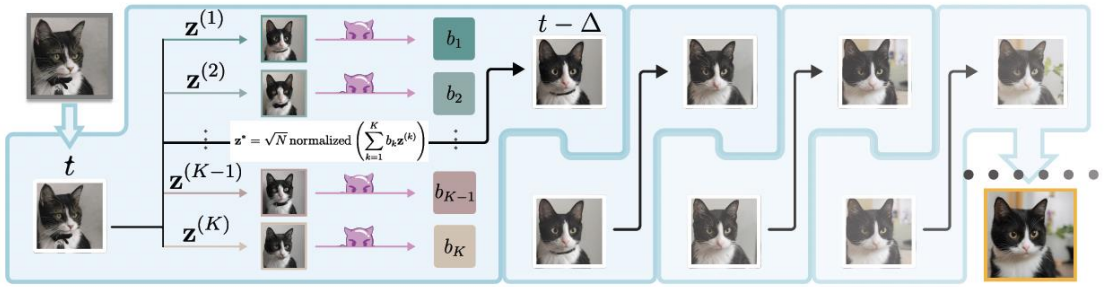
書誌情報

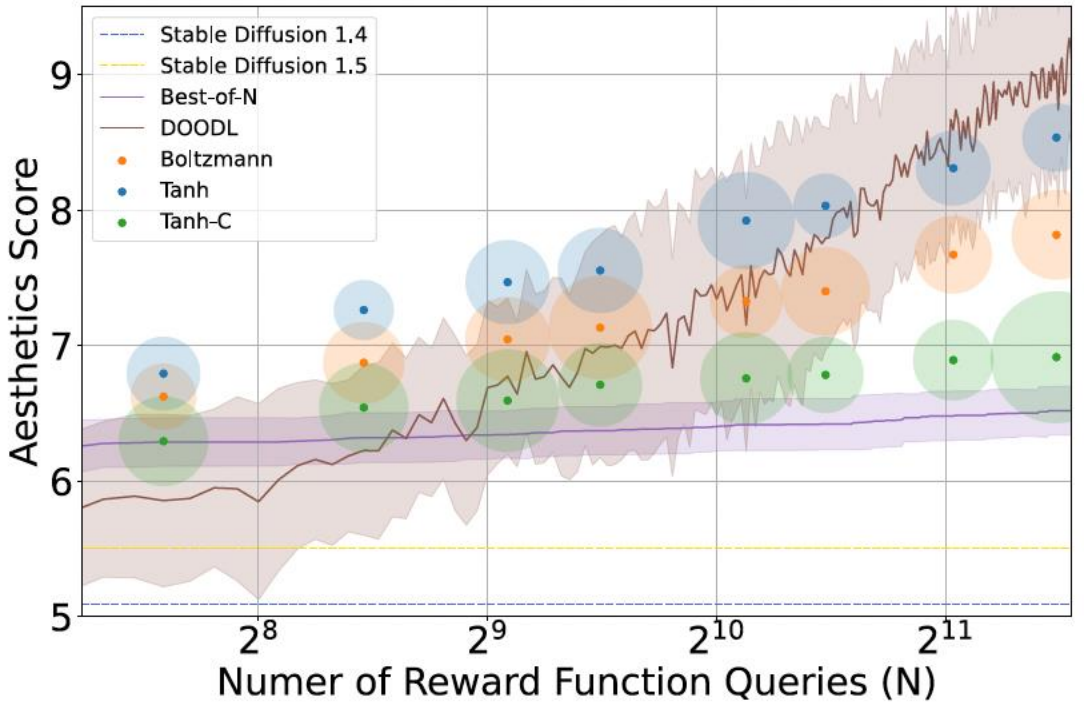
著者

- Po-Hung Yeh1 , Kuang-Huei Lee2 , Jun-Cheng Chen1
- 1.Academia Sinica, 2.Google DeepMind

概要

- Test-time に，ノイズ除去の候補を複数用い，ユーザーが好む生成へと改善する手法
- 再訓練や，評価器の微分可能性が必要ない
- Best-of-N のようなナイーブな手法より効率的に性能向上





拡散モデルのアラインメント

- 拡散モデルは，高精細な動画像生成が大きく注目を集める[Ho et al. 2018]
- ユーザーが好む生成をするよう，拡散モデルをアラインメントする研究も盛ん [Wallace et al. 2023]

Figure 1. We develop Diffusion-DPO, a method based on Direct Preference Optimization (DPO) [33] for aligning diffusion models to human preferences by directly optimizing the model on user feedback data. After fine-tuning on the state-of-the-art SDXL-1.0 model, our method produces images with exceptionally high visual appeal and text alignment, samples above.

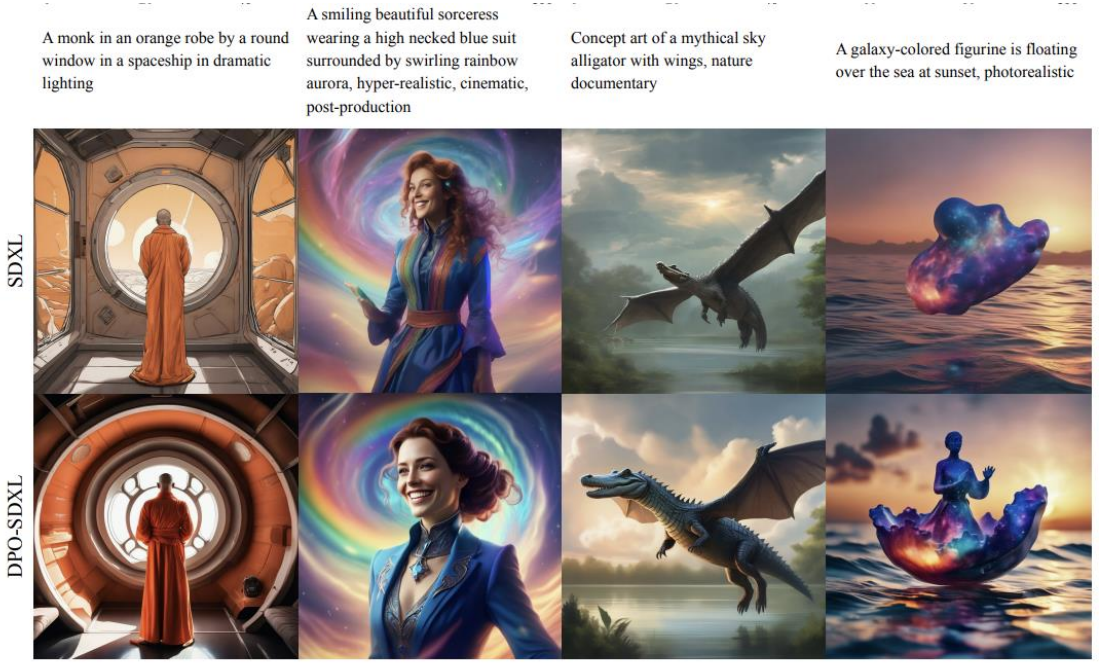


Figure 3. (Top) DPO-SDXL significantly outperforms SDXL in human evaluation. (L) PartiPrompts and (R) HPSv2 benchmark results across three evaluation questions, majority vote of 5 labelers. (Bottom) Qualitative comparisons between SDXL and DPO-SDXL. DPO-SDXL demonstrates superior prompt following and realism. DPO-SDXL outputs are better aligned with human aesthetic preferences, favoring high contrast, vivid colors, fine detail, and focused composition. They also capture fine-grained textual details more faithfully.

拡散モデルのアラインメント

- Fine-tuningによりモデルをアラインメントする手法
  - RL: DPOK [Fan et al. 2023], DDPO [Black et al. 2024]
  - SFT: DPO [Wallace et al. 2023], DRaFT [Clark et al. 2024]

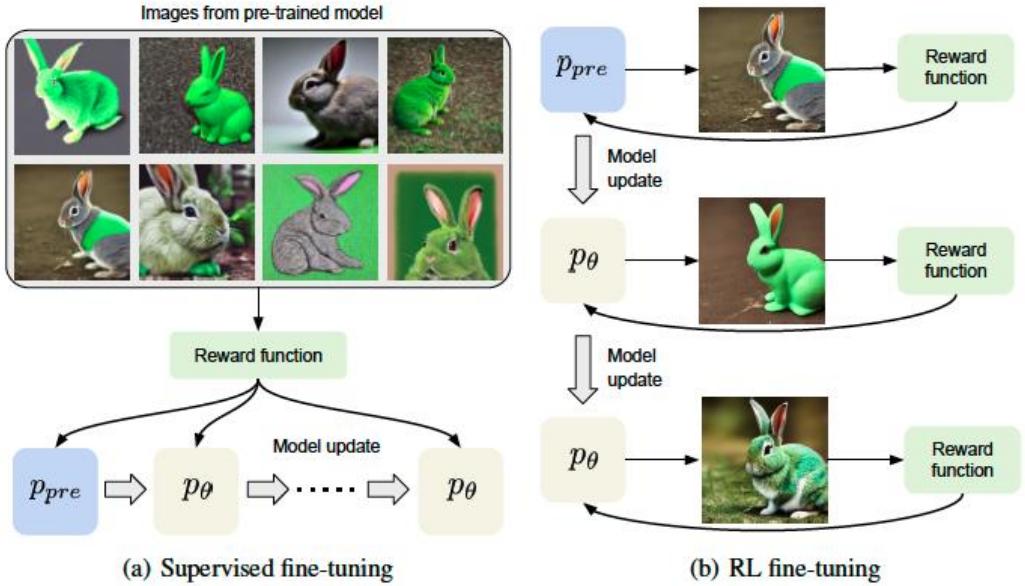


Figure 1: Illustration of (a) reward-weighted supervised fine-tuning and (b) RL fine-tuning. Both start with the same pre-trained model (the blue rectangle). In supervised fine-tuning, the model is updated on a fixed dataset generated by the pre-trained model. In contrast, the model is updated using new samples from the previously trained model during online RL fine-tuning.

拡散モデルのアラインメント

- Test-timeに生成を改善し，所望の生成を得る手法も存在
  - Classifier guidance [Dhariwal et al. 2021]
  - DOODL [Wallace et al. 2023]
  - これらの手法は，勾配を用いて生成を改善するため，評価器が微分可能でないといけない

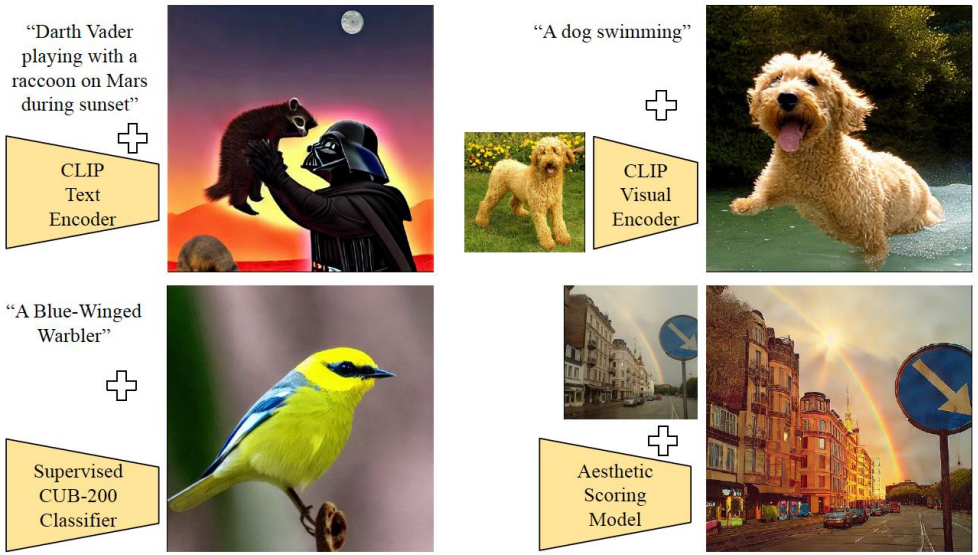


Figure 1: We propose DOODL - a process that directly optimizes diffusion latents w.r.t. a model-based loss on the final generation. Our method improves on vanilla classifier guidance in all tested settings and we demonstrate capabilities novel to this class of methods such as vocabulary expansion, entity personalization, and perceived aesthetic value improvement.

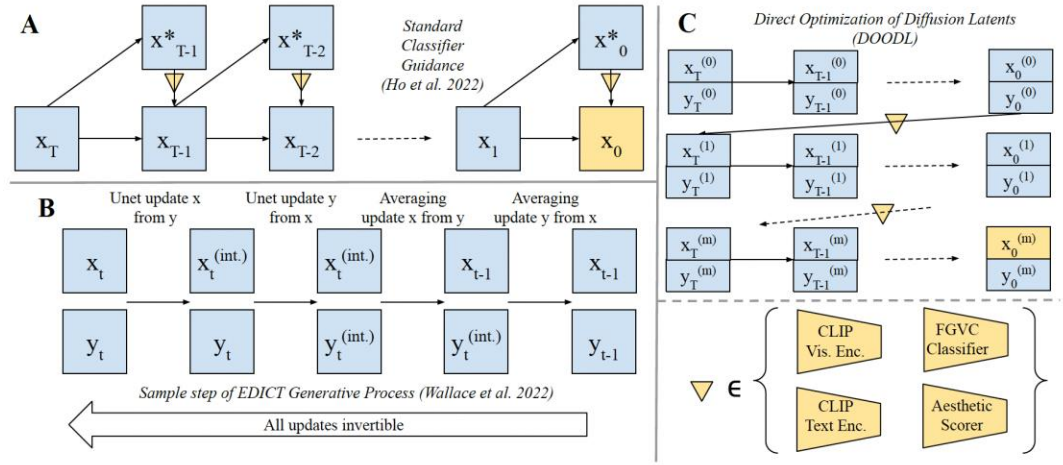


Figure 3: Method diagrams. **A** Standard classifier guidance: at each timestep, t, a one-step denoising approximation of x_0 is computed and the loss is calculated w.r.t the pixels of this generation. The gradient of this loss is incorporated into the subsequent diffusion step. **B** EDICT [45], an invertible variant of the diffusion process which admits backpropagation through the entire chain with no additional memory cost. **C** DOODL, our proposed method. We leverage EDICT and demonstrate that the gradients of model losses computed w.r.t. the final generation can be used to optimize the fully noised x_T directly. ▼ indicates a gradient calculation from a differentiable model-based loss with networks employed in this work displayed.

拡散モデルのアラインメント

- 本研究では，ノイズ除去の候補を複数用い，ユーザーが好む生成へと改善する手法である，Demonを提案
  - 再訓練や，評価器の微分可能性が必要ない
  - Demonとは，熱力学的プロセスを操作する架空の存在であるマクスウェルのデーモンに由来

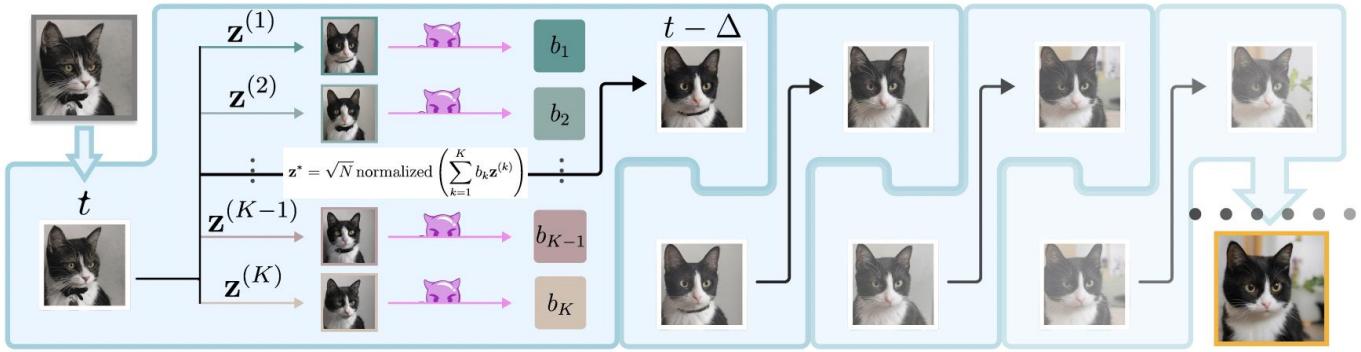


Figure 1: **Illustration of Demon.** Given a reverse-time SDE for denoising and an interval $[t_{\max}, t_{\min}]$, we first discretize it into T steps, $t_{\max} > \cdots > t > t - \Delta > \cdots > t_{\min}$. At every reverse-time denoising step, from t to $t - \Delta$, we synthesize an "optimal" noise $\mathbf{z}^*$ from K i.i.d. noises w.r.t a given reward source and use $\mathbf{z}^*$ to seed the step. This enables guiding the denoising process towards generating images that are more aligned with the reward source and the preference that the reward source represents. More details are presented in Section 4.

提案手法

- **どのようにして，生成過程を評価するのか？**
- 画像の評価器は，ノイズの乗った画像ではなく，クリーンな画像に対してのみ適用できるものが多いので，x_tからx_0へのマッピングを考える
- まず，拡散モデルの逆過程を確率微分方程式で表すと，以下のようになる

$$\mathrm{d}\mathbf{x}_t = \underbrace{\left[-t\nabla_{\mathbf{x}_t}\log p\left(\mathbf{x}_t, t\right) - \beta t^2 \nabla_{\mathbf{x}_t}\log p\left(\mathbf{x}_t, t\right)\right]}_{\mathbf{f}_\beta(\mathbf{x}_t, t)}\,\mathrm{d}t + \underbrace{\sqrt{2\beta t}}_{g_\beta(t)}\,\mathrm{d}\omega_t,$$

- x_0を，拡散ステップtの画像x_tに対応するクリーン画像とすると，

$$\mathbf{x}_0 = \mathbf{x}_t + \int_t^0 \mathbf{f}_\beta(\mathbf{x}_u, u)\,\mathrm{d}u + g_\beta(u)\,\mathrm{d}\omega_u\,,$$

提案手法

- **どのようにして，生成過程を評価するのか？**
- 実際には，SDEによりx_tからx_0へマッピングし，評価の平均を取るのではなく，ODEによるx_tからx_0へのマッピング$\boldsymbol{c}$を用い，それを評価する

$$\mathbf{c}(\mathbf{x}'_t, t) := \mathbf{x}'_0 = \mathbf{x}'_t + \int_t^0 \mathrm{d}\mathbf{x}'_u, \quad \text{where} \quad \mathrm{d}\mathbf{x}'_u = -u\nabla_{\mathbf{x}'_u} \log p\left(\mathbf{x}'_u, u\right) \mathrm{d}u.$$

- サンプルの個数を複数用意する必要がない上に，報酬関数rのラプラシアンが0に近い場合，良い近似になる(補題1参照)
- $\boldsymbol{c}$にはODEやCMを使用できる

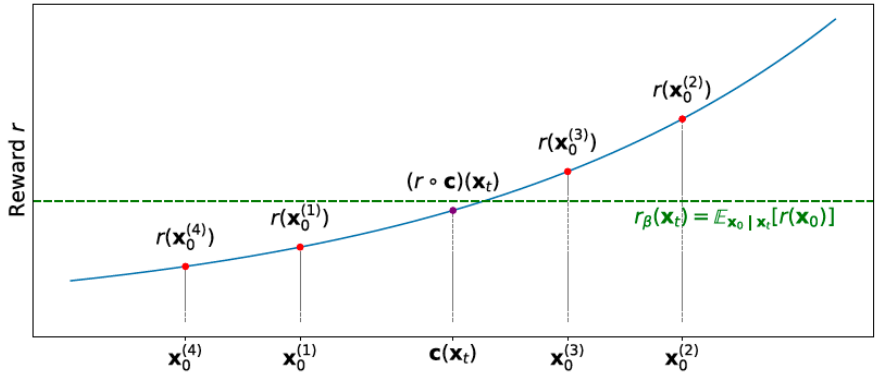


Figure 2: The illustration of the proximity between the r_β and $r \circ \mathbf{c}$. In this figure, the β is nonzero and r is near harmonic (i.e., $\nabla^2 r \approx 0$.). The red points indicate i.i.d. SDE samples and the purple ODE approximation of $\mathbf{x}_t$. The green line indicates the expectation of the rewards of the SDE samples (e.g., an approximate estimation, $\frac{1}{4}\sum_{i=1}^{4} r(\mathbf{x}_0^{(i)})$).

提案手法

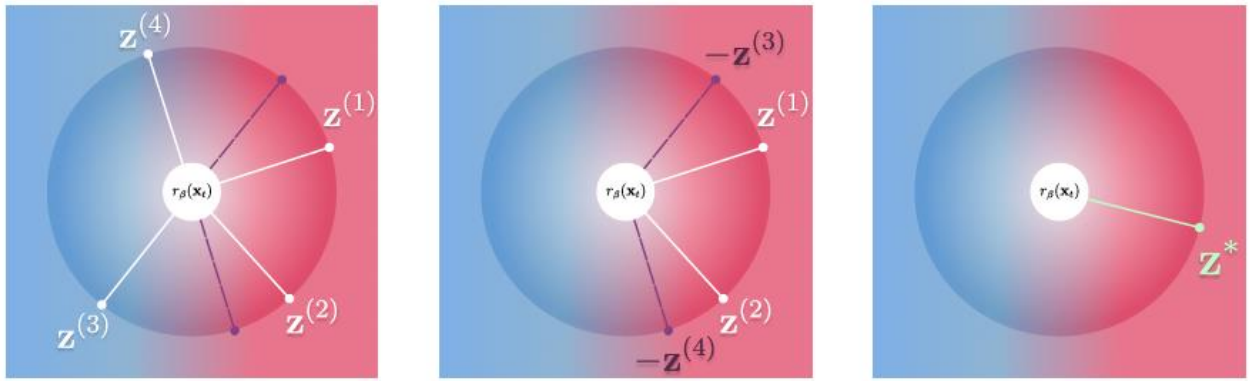
- **評価された候補をどのように使用するか**
- 高次元のガウスノイズ$\boldsymbol{z}$は，半径$\sqrt{N}$の超球面上のほど近くに位置する
  - $\boldsymbol{z}$がN次元標準正規分布に従う場合，高確率で，$||\boldsymbol{z}|| = \sqrt{N} + \mathcal{O}(1)$ (補題5参照，中心極限定理などを利用する)
- よって，以下のように，重み付き和と正規化により，新たなノイズ$\boldsymbol{z}^*$を作成可能

$$\mathbf{z}^* = \sqrt{N}\,\text{normalized}\left(\sum_{k=1}^{K} b_k \mathbf{z}^{(k)}\right),$$

提案手法

$$\mathbf{z}^* = \sqrt{N}\,\text{normalized}\left(\sum_{k=1}^{K} b_k \mathbf{z}^{(k)}\right),$$

- **評価された候補をどのように使用するか**
- 評価の低いノイズには負の重みをつけ，評価の高いノイズには正の重みをつける
  - このときの重み付けの手法をDemonと呼んでおり，この論文ではTanh DemonとBoltzmann Demonの二つを提案している



(a) Before (b) After (c) Final

Figure 3: An illustration of the Tanh Demon sampling method where $K = 4$. (a) A SDE step generates several samples, each determined by sampled noise $\mathbf{z}_k$. We use Tanh Demon to classify each noise sample as "low-reward" or "high-reward" w.r.t $r_\beta(x_t)$ based on their respective reward estimates. (b) We flip the sign of the low-reward noise with tanh, thereby transforming it into high-reward noise. (c) It shows how the post-processed noises are averaged and projected onto the high-dimensional sphere, resulting in a feasible noise representation $\mathbf{z}^*$ with high-reward estimate.

$$b_k^{\text{tanh}} \leftarrow \tanh\left(\frac{(r \circ \mathbf{c})(\hat{\mathbf{x}}_{t-\Delta}^{(k)}) - \hat{\mu}}{\tau}\right)$$

$$b_k^{\text{boltz}} \leftarrow \frac{\exp\left(r \circ \mathbf{c}(\hat{\mathbf{x}}_{t-\Delta}^{(k)})/\tau\right)}{\sum_{k=1}^{K} \exp\left(r \circ \mathbf{c}(\hat{\mathbf{x}}_{t-\Delta}^{(k)})/\tau\right)}.$$

実験

- Tステップの拡散ステップに対して，K個の候補を出す
  - これらを変えることで，評価回数や計算時間を様々に変更して比較
- 使用するモデルは，SD v1.4 とSDXL

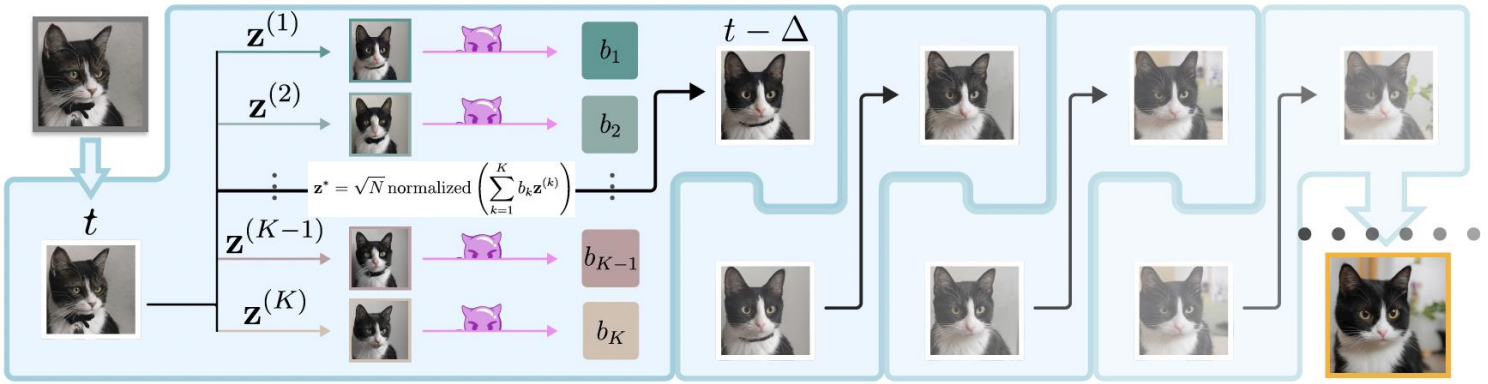
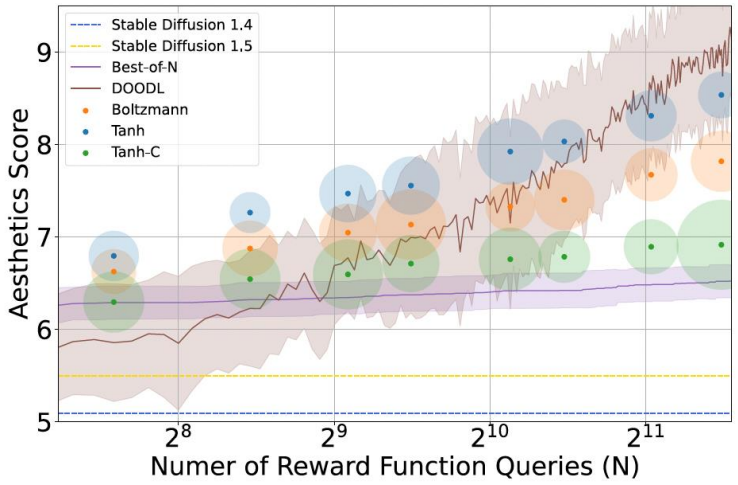


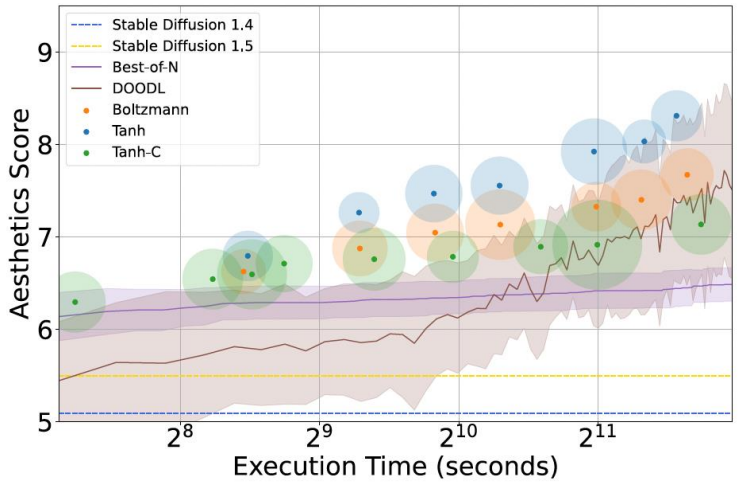
Figure 1: **Illustration of Demon.** Given a reverse-time SDE for denoising and an interval $[t_{\text{max}}, t_{\text{min}}]$, we first discretize it into T steps, $t_{\text{max}} > \cdots > t > t - \Delta > \cdots > t_{\text{min}}$. At every reverse-time denoising step, from t to $t - \Delta$, we synthesize an "optimal" noise $\mathbf{z}^*$ from K i.i.d. noises w.r.t a given reward source and use $\mathbf{z}^*$ to seed the step. This enables guiding the denoising process towards generating images that are more aligned with the reward source and the preference that the reward source represents. More details are presented in Section 4.

実験

- 横軸をコスト，縦軸に評価関数の評価とした場合の結果
  - 横軸は，左図では評価回数，右図では計算時間
- 最終状態でしか評価を行わない手法よりも，効率的に改善可能
  - Best-of-Nよりも，評価関数の評価を効率的に改善できる
  - DOODLよりも，計算時間の面では効率的



(a) Performance w.r.t Reward Query Number



(b) Performance w.r.t Execution Time

実験

- 様々な評価関数に対する結果
  - Demonでは，多少評価関数に対してハックしてしまうものの，一つの評価関数を向上させることで，他の評価関数の評価も向上できている
  - 一方，DOODLでは，他の評価関数の評価が大きく落ちている

Table 3: Results using various reward functions and different generation methods. Each column represents a specific reward objective, with the best performance highlighted in bold.

Generation method	Aes ↑	IR ↑	Pick ↑	HPSv2 ↑
SD v1.4	5.34 ± 0.56	-0.00 ± 0.95	0.202 ± 0.008	0.216 ± 0.036
Tanh + Aes	**7.35** ± 0.40	-0.03 ± 1.24	0.211 ± 0.010	0.257 ± 0.041
Tanh + IR	5.96 ± 0.28	**1.95** ± 0.07	0.216 ± 0.012	0.286 ± 0.033
Tanh + Pick	6.14 ± 0.48	1.39 ± 0.57	**0.245** ± 0.010	0.312 ± 0.033
Tanh + HPSv2	5.98 ± 0.45	1.51 ± 0.63	0.228 ± 0.011	**0.367** ± 0.027
Tanh + Ensemble	6.53 ± 0.50	1.81 ± 0.15	0.236 ± 0.014	0.356 ± 0.030
DOODL + Aes	5.59 ± 0.29	-0.68 ± 1.06	0.197 ± 0.008	0.221 ± 0.028
DOODL + Pick	5.21 ± 0.46	-0.12 ± 0.84	0.204 ± 0.010	0.220 ± 0.035

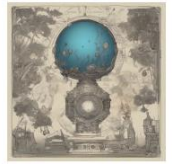
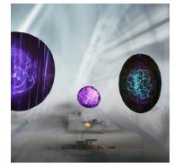
実験

- 微分不可能な評価器へのアラインメント
  - VLM(Google Gemini, GPT4 Turbo)に，各シナリオ(右上)に合うような生成を選ばせ，それを元に評価を作成
    - 選択された候補は+5，それ以外は−5
  - すると，シナリオに沿っている画像を生成できることが，定性的に確認された
  - 一方，DOODLのようなtest-time改善手法は，評価器が微分可能でないと利用できない

The following are the full prompts for the scenarios:

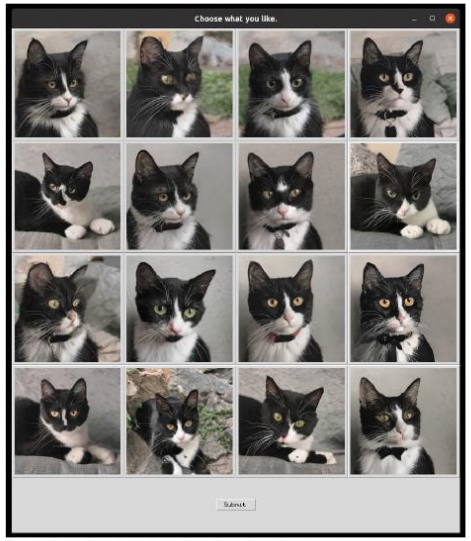
1. **Teacher:** You are a teacher looking to create custom illustrations for your educational materials to make learning more engaging for your students.
2. **Artist:** You are a game or movie concept artist tasked with creating concept art for characters, settings, and scenes to speed up the pre-production process.
3. **Researcher:** You are a researcher needing to visualize complex data, such as molecular structures in chemistry or weather patterns in meteorology, for better understanding or presentation.
4. **Journalist:** You are a journalist who wants to add a visual teaser for your article to grab attention on social media or your news website.

Table 5: Using VLMs to generate images. PF-ODE (baseline) refers to a baseline without using our method for alignment. Columns 3-6 indicate the role that the agent plays in the given prompt.

Model	Baseline	Teacher	Artist	Researcher	Journalist
Gemini-SD v1.4					
Gemini-SDXL					
GPT-SD v1.4					
GPT-SDXL					

実験

- 人間の対話的判断へのアラインメント
  - 参照猫と似た猫と生成する，というタスク
  - ユーザは左図に示される候補の中から，似ている猫を選択
  - 選択に応じて評価
    - 選択された候補は$+1$，それ以外は-1
  - すると，確かに参照猫と似た猫が生成できる



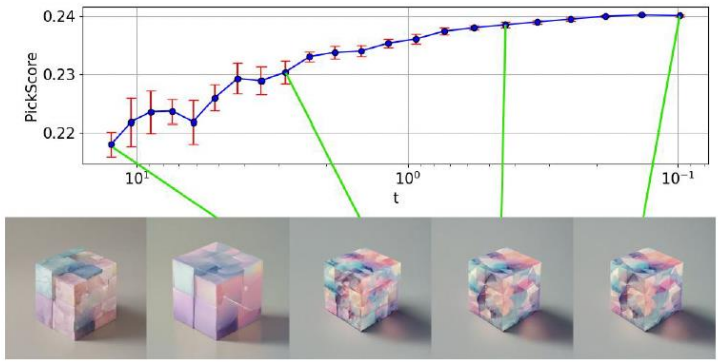
(a) Our user interface for interacting with our algorithm (0.594 cosine similarity).

(b) (Top Left) Image generated by PF-ODE (0.622 cosine similarity). (Bottom Left) Image generated by our method (0.708 cosine similarity). (Right) Reference image.

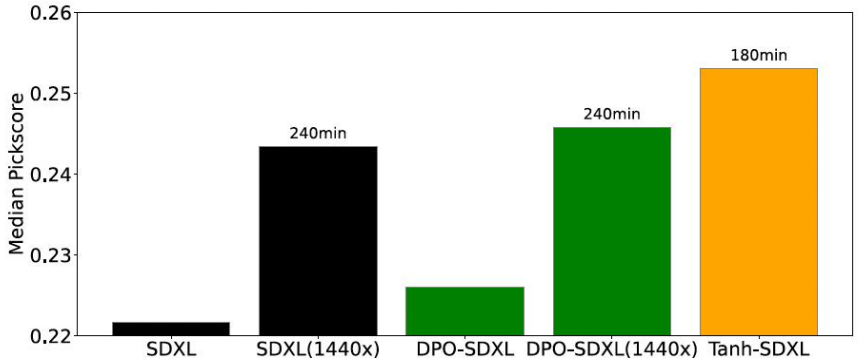
Figure 5: We design an application for manual interaction with our algorithm. Our author selects the images, and the criteria are based on the author's preference (non-preferred images are kept unselected), where the author tries to align the reference image. We evaluate performance by measuring the cosine similarity of DINOv2 features between the targeted and reference images.

実験

- Fine-tuneing手法(DPO)との比較
  - PickScoreを評価関数として用いた場合，DPOよりも精度向上が大きい
  - さらに，DPOでFine-tuneingしたモデルのBest-of-Nを取り，推論にも計算コストをかけたとしても，Demonの方が性能が高い
  - ただ，PickScore以外で同様の結果になることの確認や，指標ハック度合いの比較はない



(a) A Trajectory of Tanh Demon. We plot $(r \circ \mathbf{c})(\mathbf{x}_t)$ for different t.



(b) The performance of each method on PickScore.

Figure 7: Quantitative results for Tanh Demon.

結論

- 拡散モデルにおける，test-timeのアラインメント手法であるDemonを提案
  - 再訓練や，評価器の微分可能性が必要ないため，VLMの出力や人間の対話的判断のような微分不可能な評価器に対してもアラインメント可能
  - ステップごとに評価を行うことができるため，Best-of-Nや既存手法に対して，計算効率的
  - さらに，この論文の検証範囲内では，Fine-tuningを用いたアラインメントよりも，評価向上

感想

興味深かった点

- 拡散モデルにおいても推論に計算量を割くことが，有効であることが示された
  - 特に，ステップごとに評価することが，計算効率的なアラインメントのために有効
  - 限定的な検証ではあるが，fine-tuningを用いる手法よりも良い結果が示されたことも，面白い
- 拡散モデルに対する推論のサーチの方法として加重平均があり得ることが示された
  - ノイズの加重平均をとる，という手法は，連続な出力である拡散モデルだからこそできること

考える余地のある点

- ノイズの加重平均を取るのが，どれだけいい手法なのか分からない
  - 悪いノイズを(-1)倍したら良いノイズになる，という暗黙の仮定があるが，本当？
- ビームサーチとの比較を見たかった
  - Boltzman Demonの場合は，温度パラメタτが小さい方が，アラインメント性能が良い
    - Boltzman Demonで，温度パラメタ$\tau \to 0$の場合は，$N = 1$のビームサーチと等価
  - ただ，さらにそれよりもTanh Demonの場合の方が性能が良い

参考文献

[Ho et al. 2018] Denoising Diffusion Probabilistic Models

[Fan et al. 2023] DPOK: Reinforcement Learning for Fine-tuning Text-to-Image Diffusion Models

[Black et al. 2024] Training Diffusion Mode l s with Reinforcement Learning

[Wallace et al. 2023] Diffusion Model Alignment Using Direct Preference Optimization

[Clark et al. 2024] Directly Fine-Tuning Diffusion Models on Differentiable Rewards

[Dhariwal et al. 2023] Diffusion Models Beat GANs on Image Synthesis

[Wallace et al. 2023] End-to-End Diffusion Latent Optimization Improves Classifier Guidance

[Kirstain et al. 2023] Pick-a-Pic: An Open Dataset of User Preferences for Text-to-Image Generation