DyMO: Training-Free Diffusion Model Alignment with Dynamic Multi-Objective Scheduling

Xin Xie, Dong Gong*

The University of New South Wales (**UNSW**), Australia

https://shelsin.github.io/dymo.github.io/

PROJECT PAGE

Introduction

Diffusion Model Alignment enhances the generated image in two aspects: **semantic alignment** with the textual prompt and **visual alignment** with user preferences.

Figure 1. Sample images generated by DyMO based on SDXL backbones.

Limitations of Existing Works:

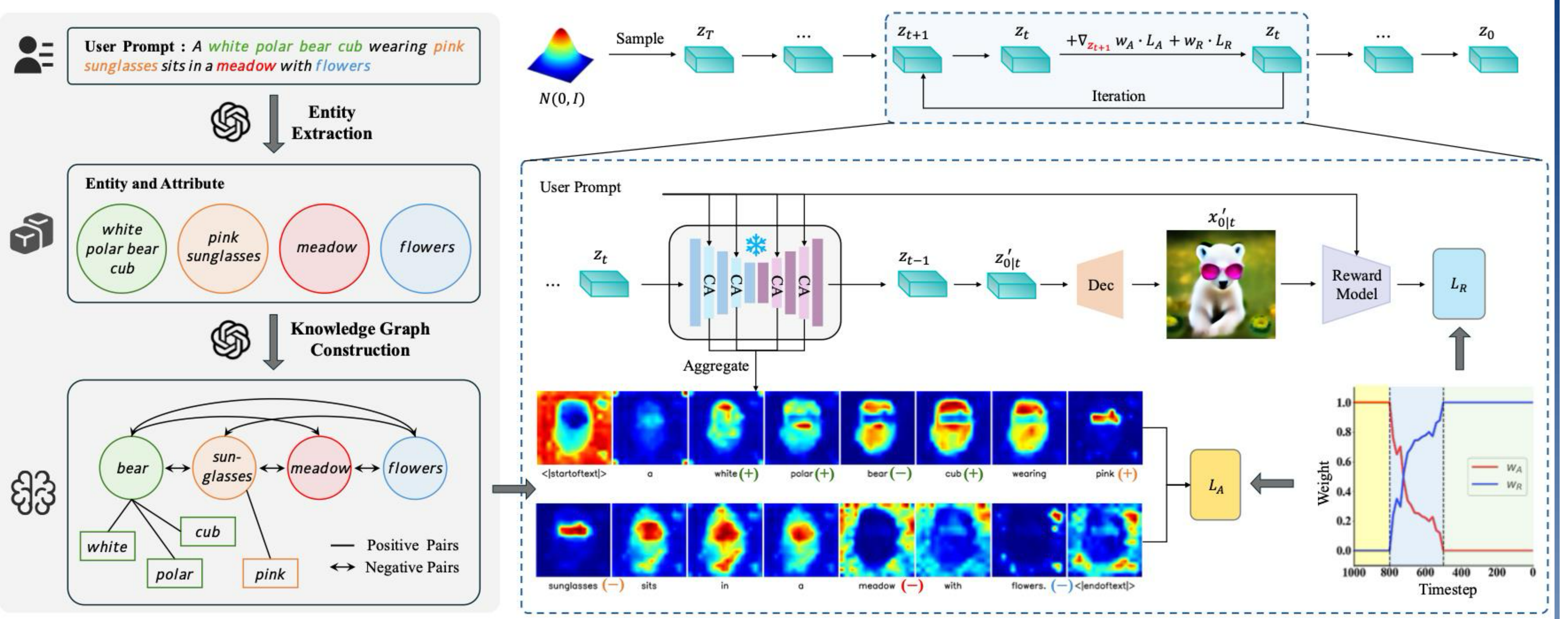
- Training-based alignment methods are resource-intensive and lacks generalization across diverse preferences.
- Training-free alignment methods suffer from inaccurate guidance from the noisy samples or blurred one-step predictions.

Test-time Alignment with Dynamic Multi-Objective Scheduling (DyMO):

- Guiding the denoising process using gradients from the pre-trained text-aware preference scores on one-step predictions.
- One-step prediction is efficient but lacks semantic fidelity, weakening textual alignment in preference-based guidance.
- A semantic alignment objective is introduced to align the visual content (reflected in text-image attention maps) with LLM-extracted text semantics.
- We dynamically schedule two objectives for tailored guidance across timesteps, generating detailed content while keeping the layout.
- A dynamic recurrent strategy is further proposed to adaptively decide iteration count at different stages for improved guidance.

The Proposed DyMO

Overview:



- **Preference Alignment Guidance**: Given a text prompt c, we estimate a text-aware human preference score of $x'_{0|t}$.

$$\mathcal{L}_R(\mathbf{x}'_{0|t}, \mathbf{c}, t) = \exp(\tau \cdot f_V(\mathbf{x}'_{0|t}, t)^T f_T(\mathbf{c}))$$

- **Semantic Alignment Guidance**: We align image content (via cross-attention maps M) with a text semantic graph extracted by LLM.

$$\mathcal{L}_A = -\frac{1}{|\mathbf{S}_{\text{pos}}|}\sum_{(s,l)\in\mathbf{S}_{\text{pos}}} f(M_s, M_l) + \frac{1}{|\mathbf{S}_{\text{neg}}|}\sum_{(s,l)\in\mathbf{S}_{\text{neg}}} f(M_s, M_l)$$

- **Multi-Objective Dynamic Scheduling**: We dynamically balance $\mathcal{L}_A$ and $\mathcal{L}_R$ over denoising steps for tailored guidance.

$$\mathcal{L} = w_A \cdot \mathcal{L}_A(M) + w_R \cdot \mathcal{L}_R(\mathbf{x}'_{0|t}, \mathbf{c}, t)$$

- **Latent Update**: $\mathbf{z}_{t-1} \leftarrow \mathbf{z}_t + \epsilon_\theta(\mathbf{z}_t, \mathbf{c}, t) - \eta_t \nabla_{\mathbf{z}_t}\mathcal{L}$
- **Dynamic Time-Travel Strategy**: Adaptively determine the number of recurrent guidance.

$$r_t = h_t \cdot \|\nabla_{\mathbf{z}_t}\mathcal{L}\|$$

Algorithm 1 Our method + Dynamic Time-Travel Straregy

Input: prompt $\mathbf{c}$, noise predictor $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\cdot, t)$, human preference evaluator $\mathcal{L}_R(\cdot, \mathbf{c})$, semantic alignment loss function $\mathcal{L}_A(\cdot)$, timesteps T, decoder D, guidance strength η_t and pre-defined parameters β_t, $\bar{\alpha}_t$, h_t, k.

1: $\mathbf{z}_T \sim \mathcal{N}(0, \mathbf{I})$
2: **for** $t = T, ..., 1$ **do**
3: $\boldsymbol{\epsilon}_1 \sim \mathcal{N}(0, \mathbf{I})$ if $t > 1$, else $\boldsymbol{\epsilon}_1 = 0$.
4: $\tilde{\boldsymbol{\epsilon}}_t, M = \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\mathbf{z}_t, t)$
5: $\mathbf{z}_{t-1} = (1 + \frac{1}{2}\beta_t)\mathbf{z}_t + \beta_t\tilde{\boldsymbol{\epsilon}}_t + \sqrt{\beta_t}\boldsymbol{\epsilon}_1$
6: $\mathbf{z}'_{0|t} = \frac{1}{\sqrt{\bar{\alpha}_t}}(\mathbf{z}_t + (1 - \bar{\alpha}_t)\tilde{\boldsymbol{\epsilon}}_t)$
7: $\mathbf{x}'_{0|t} = D(\mathbf{z}'_{0|t})$
8: $\begin{cases} w_A = 1, w_R = 0 & \textbf{if } t \geq 800 \\ w = 1 - e^{-k\left(\frac{\|\mathbf{z}'_{0|t} - \mathbf{z}'_{0|t+1}\|}{\|\mathbf{z}'_{0|t+1}\|}\right)} & \textbf{if } 800 > t \geq 500 \\ w_A = w, w_R = 1 - w & \\ w_A = 0, w_R = 1 & \textbf{if } 500 > t \geq 1 \end{cases}$
9: $\mathcal{L} = w_A \cdot \mathcal{L}_A(M) + w_R \cdot \mathcal{L}_R(\mathbf{x}'_{0|t}, \mathbf{c}, t)$
10: $\boldsymbol{g}_t = \nabla_{\mathbf{z}_t}\mathcal{L}$
11: $\mathbf{z}_{t-1} = \mathbf{z}_{t-1} - \eta_t \cdot \frac{\|\tilde{\boldsymbol{\epsilon}}_t\|}{\|\boldsymbol{g}_t\|_2^2} \cdot \boldsymbol{g}_t$
12: $r_t = h_t \cdot \|\boldsymbol{g}_t\|$ ▷ Compute once at each timestep
13: **for** $i = r_t, ..., 1$ **do** ▷ Iterate r_t times
14: $\boldsymbol{\epsilon}_2 \sim \mathcal{N}(0, \mathbf{I})$
15: $\mathbf{z}_t = \sqrt{1 - \beta_t}\mathbf{z}_{t-1} + \sqrt{\beta_t}\boldsymbol{\epsilon}_2$
16: Repeat from step3 to step16
17: **return** x_0

Experiments

Qualitative comparison on both SD V1.5 and SDXL backbones.

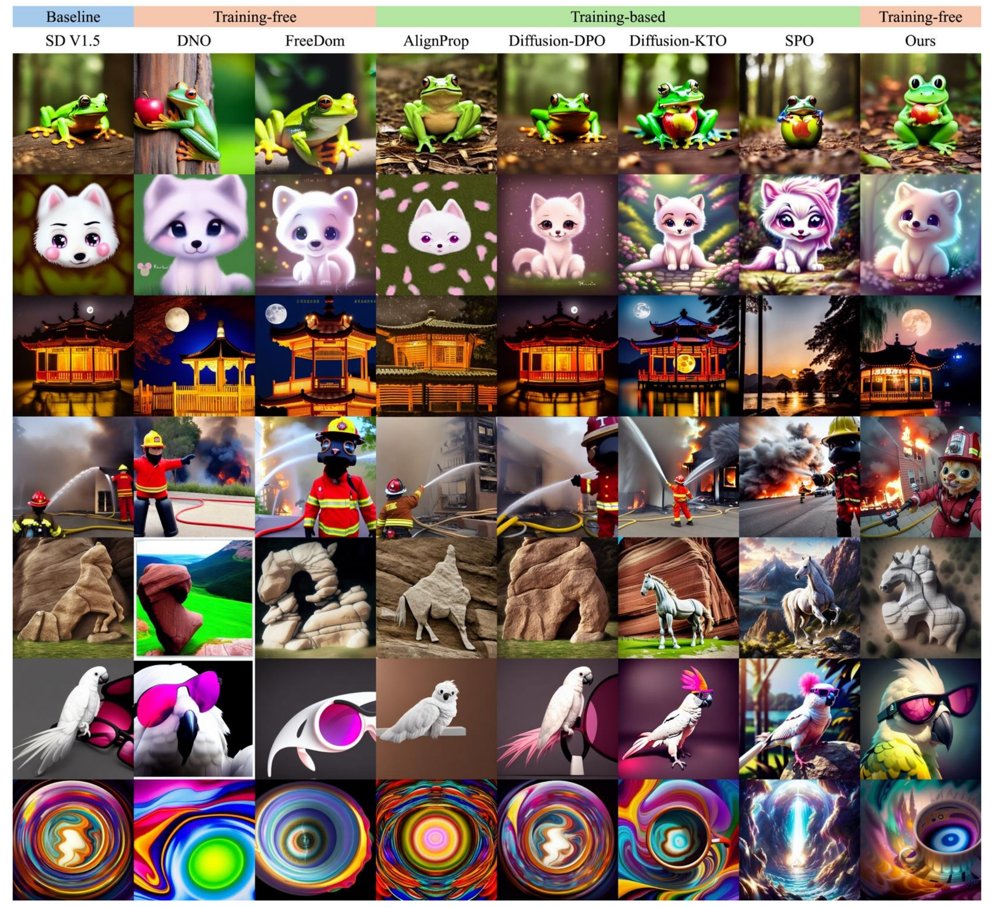


Figure 3. Qualitative comparison based on SD V1.5 backbones.

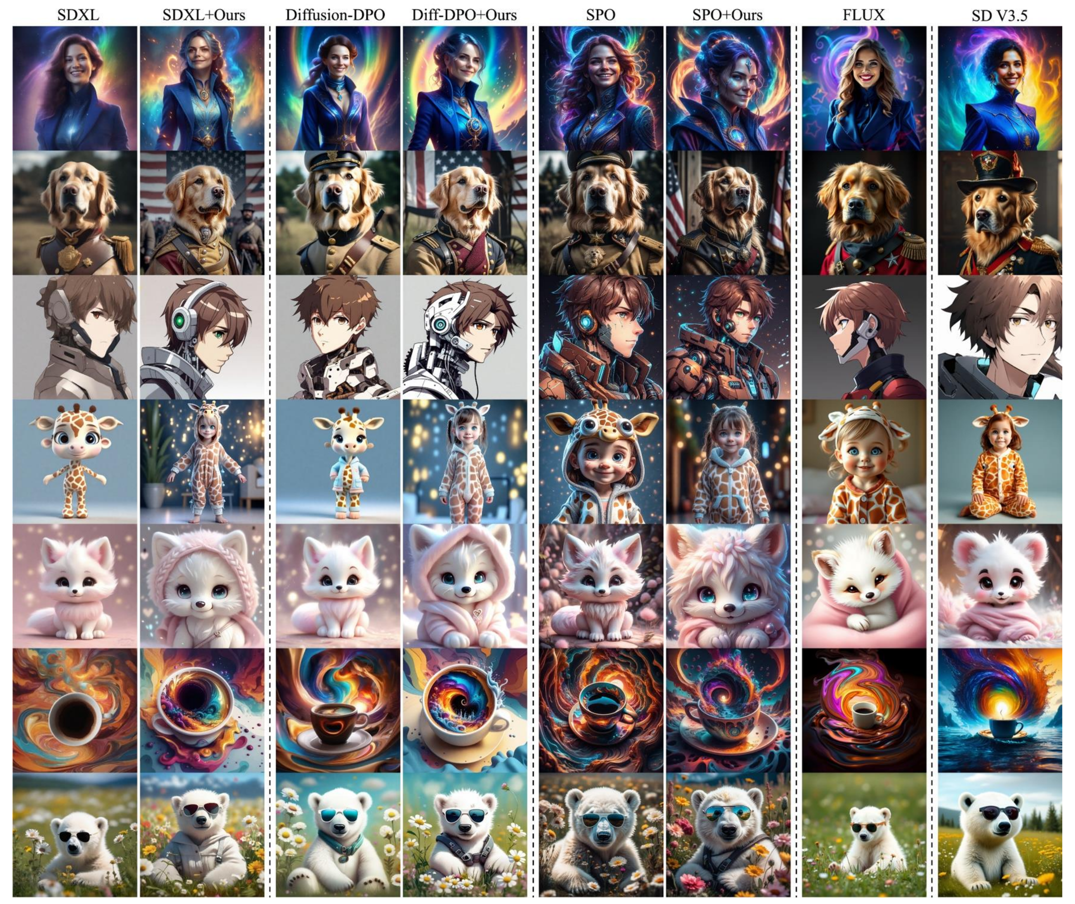


Figure 4. Qualitative comparison based on SDXL backbones.

Comparisons with other alignment methods across four metrics.

Table 1. Comparison of AI feedback on SD V1.5-based methods.

Methods	PickScore	HPSv2	ImageReward	Aesthetics
SD V1.5	20.73	0.2341	0.1697	5.337
DNO	20.05	0.2591	-0.3212	5.597
PromptOpt	20.26	0.2490	-0.3366	5.465
FreeDom	21.96	0.2605	0.3963	5.515
AlignProp	20.56	0.2627	0.1128	5.456
Diffusion-DPO	20.97	0.2656	0.2989	5.594
Diffusion-KTO	21.15	0.2719	0.6156	5.697
SPO	21.46	0.2671	0.2321	5.702
SD V1.5+Ours	**23.07**	**0.2755**	**0.7170**	**5.831**

Table 2. Comparison of AI feedback on SDXL-based methods.

Methods	PickScore	HPSv2	ImageReward	Aesthetics
SDXL	21.91	0.2602	0.7755	5.960
DNO	22.14	0.2725	0.9053	6.042
PromptOpt	21.98	0.2708	0.8671	5.881
FreeDom	22.13	0.2719	0.7722	5.908
SDXL+Ours	**24.90**	**0.2839**	**1.074**	**6.138**
Diffusion-DPO	22.30	0.2741	0.9789	5.891
Diff-DPO+Ours	**24.46**	**0.2836**	**1.049**	**6.116**
SPO	22.81	0.2778	1.082	**6.319**
SPO+Ours	**23.85**	**0.2821**	**1.166**	6.278
SD V3.5	21.93	0.2726	0.9697	5.775
FLUX	22.04	0.2760	1.011	6.077

Performance analysis of core components.

Table 3. Ablation study results.

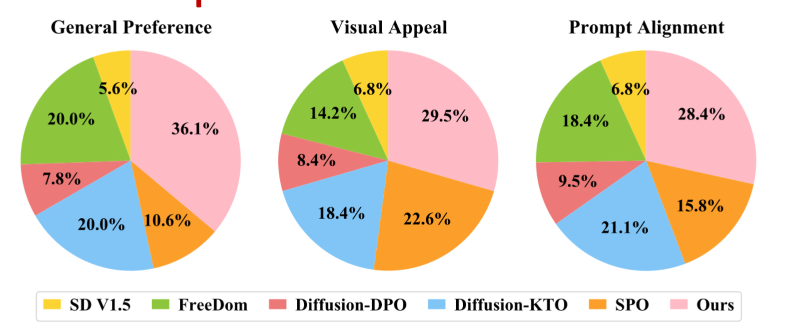
Methods	PickScore	HPSv2	ImageReward	Aesthetics
w/o $\mathcal{L}_A$	22.07	0.2546	0.6413	5.686
w/o $\mathcal{L}_R$	20.37	0.2418	0.1230	5.426
w/o w	22.34	0.2656	0.6748	5.708
w/o Polyak step	20.61	0.2640	0.2838	5.470
w PickSocre	**23.38**	0.2746	0.5463	5.694
Ours (Llama-3.3)	22.58	**0.2829**	0.7048	5.802
Ours (GPT-4)	23.07	0.2755	**0.7170**	**5.831**

Ablation study on effect of recurrent strategy.

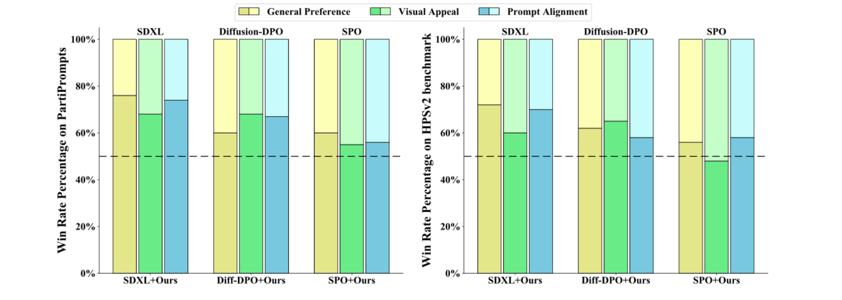
Table 4. Effect of the iteration count in time-travel straregy.

Iteration	PickScore	HPSv2	ImageReward	Aesthetics	Avg Time
1	20.41	0.2423	0.6191	5.652	40s
5	21.27	0.2610	0.6679	5.661	170s
10	22.85	0.2677	0.7055	5.706	295s
Dynamic	**23.07**	**0.2755**	**0.7170**	**5.831**	190s

User preference evaluation.



(a) User preference distribution on HPSv2 benchmark.



(b) Human evaluation w/ & w/o our method on PartiPrompts & HPSv2.

Figure 6. User study results.