InPO: Inversion Preference Optimization with Reparameterized DDIM for Efficient Diffusion Model Alignment

Supplementary Material

S1: Background

Conditional Generative Models Inspired by non-equilibrium thermodynamics, diffusion models gradually introduce noise into the data and learn to reverse this process starting from pure noise, ultimately generating target data that aligns with the original data distribution. Broadly, diffusion models can be categorized into two types: denoising diffusion and score-matching [62]. Diffusion models [15, 30, 41] have become the leading approach in generative modeling [4, 64, 73], outperforming earlier methods like Generative Adversarial Networks (GANs) [51] and Variational Autoencoders (VAEs) in both quality and stability. They have shown outstanding results across a wide range of generative tasks, including image [5, 10, 42, 54] and video generation [7, 8, 17, 17, 24, 27, 59]. This paper primarily focuses on diffusion models for conditional image generation [13], encompassing tasks such as text-to-image synthesis, additional control conditions, and image inpainting for restoration. Conditional image generation[25, 29, 37, 38] leverages guidance conditions to synthesize new images from scratch, with the conditions being either single or multiple. Earlier approaches predominantly relied on class-conditional generation, which required training extra classifiers and utilizing classifier-induced gradients for image synthesis. In contrast, Ho et al. introduce classifier-free guidance [26], which eliminates the need for classifier training and allows for more flexible conditioning. This approach also enables control over the degree of guidance, such as text prompts, by adjusting specific coefficients. Beyond text-based prompts [4, 53], more specific conditions [21, 22] can be utilized to achieve finer control over image synthesis. For instance, ControlNet [74] allows the integration of additional input types, such as depth maps, precise edges, poses, and sketches, to guide the generation process more accurately. Moreover, image restoration is a vital task in computer vision that focuses on enhancing the quality of images affected by various degradations. An example is inpainting [58], where the goal is to fill in missing regions of an image to restore its completeness.

Large Language Models Alignment Reinforcement Learning from Human Feedback (RLHF) [11] is a widely used approach to align models with human preferences. It involves first training a reward model on data that explicitly reflects human preferences, followed by using reinforcement learning techniques to optimize the policy/model, aiming to maximize the reward. As is well recognized, the widely popular model ChatGPT leverages RLHF techniques. A pioneering study, the work of [40], is the first to apply RLHF for fine-tuning LLMs, which has since gained substantial recognition. Proximal Policy Optimization (PPO) [56] is a crucial algorithm in RL, but its training process often requires the simultaneous use of a training model, a reference model, a reward model, and a critic, which is particularly demanding in terms of memory consumption, especially when applied to LLMs [40]. Recent research suggests that it is possible to circumvent traditional RL algorithms. For instance, RAFT [16] achieves optimization by fine-tuning on online samples with the highest rewards. Meanwhile, RRHF [72] aligns models using ranking loss, learning from responses sampled from multiple sources to enhance alignment. The work of [34] introduce rejection sampling optimization, where preference data is collected using a reward model to guide the sampling process. DPO bypasses the need for training an explicit reward model by directly optimizing the optimal policy, assuming that pairwise preferences can be approximated using pointwise rewards. To address potential overfitting to preference datasets in DPO, the work of [2] propose Identity Preference Optimization. The work of [28] introduce odds ratio preference optimization (ORPO), which incorporates SFT on preference data. In contrast, the work of [19] avoids reliance on pairwise preference data by combining Kahneman-Tversky optimization (KTO), focusing on directly optimizing utility instead of maximizing the log-likelihood of preferences. Additionally, the work of [60] propose preference ranking optimization (PRO), which leverages higher-order information embedded in list rankings. However, due to the unique characteristics of diffusion models, these methods cannot be directly applied without significant adaptation.

Additional Diffusion Models Alignment Aligning diffusion models with human preferences has recently attracted significant attention. The work of [46] extend this concept to video diffusion models but encountered challenges, such as the linear increase in reward feedback costs due to the added time dimension. To overcome these issues, they optimize the video diffusion model using gradients obtained from publicly available pre-trained reward models. In contrast, InstructVideo [71] fine-tunes text-to-video diffusion models based on human feedback rewards, employs partial DDIM sampling to reduce computational costs, and leverages image reward models to enhance video quality while maintaining the model's generalization

ability. Moreover, reward models [67] and timestep-aware alignment methods [33] for diffusion models warrant further investigation. Human preference alignment techniques developed for LLMs can potentially be adapted for diffusion models. We believe that advances in methods such as IPO, ORPO, and PRO could be extended to diffusion models, potentially enhancing their performance. However, due to the fundamental differences in architecture between LLMs and diffusion models, directly applying LLM techniques to diffusion models may not produce the same level of benefits.

S2: Details of the Primary Derivation

In this section, we present a detailed derivation of the proposed method. From Eq. (4), we can derive the following:

$$
\begin{aligned}
&\max_{p_\theta} \mathbb{E}_{\boldsymbol{x}_0\sim p_\theta(\boldsymbol{x}_0|\boldsymbol{c})}[r(\boldsymbol{x}_0,\boldsymbol{c})]/\beta - \mathbb{D}_{\mathrm{KL}}[p_\theta(\boldsymbol{x}_0|\boldsymbol{c})||p_{\mathrm{ref}}(\boldsymbol{x}_0|\boldsymbol{c})] \\
=&\min_{p_\theta^{\boldsymbol{c}}} -\mathbb{E}_{\boldsymbol{x}_0\sim p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_0)}[r(\boldsymbol{x}_0,\boldsymbol{c})]/\beta + \mathbb{D}_{\mathrm{KL}}[p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_0)||p_{\mathrm{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0)] \\
\leq&\min_{p_\theta^{\boldsymbol{c}}} -\mathbb{E}_{\boldsymbol{x}_0\sim p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_0)}[r(\boldsymbol{x}_0,\boldsymbol{c})]/\beta + \mathbb{D}_{\mathrm{KL}}[p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_0,\boldsymbol{x}_t)||p_{\mathrm{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0,\boldsymbol{x}_t)] \\
=&\min_{p_\theta^{\boldsymbol{c}}} -\mathbb{E}_{p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_0,\boldsymbol{x}_t)}[r_t^{\boldsymbol{c}}(\boldsymbol{x}_0,\boldsymbol{x}_t)]/\beta + \mathbb{D}_{\mathrm{KL}}[p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_0,\boldsymbol{x}_t)||p_{\mathrm{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0,\boldsymbol{x}_t)] \\
=&\min_{p_\theta^{\boldsymbol{c}}} \mathbb{E}_{p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_0,\boldsymbol{x}_t)}\left(\log\frac{p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_0,\boldsymbol{x}_t)}{p_{\mathrm{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0,\boldsymbol{x}_t)\exp(r_t^{\boldsymbol{c}}(\boldsymbol{x}_0,\boldsymbol{x}_t)/\beta)/Z_t(\boldsymbol{c})} - \log Z_t(\boldsymbol{c})\right) \\
=&\min_{p_\theta^{\boldsymbol{c}}} \mathbb{D}_{\mathrm{KL}}(p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_0,\boldsymbol{x}_t)||p_{\mathrm{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0,\boldsymbol{x}_t)\exp(r_t^{\boldsymbol{c}}(\boldsymbol{x}_0,\boldsymbol{x}_t)/\beta)/Z_t(\boldsymbol{c}))
\end{aligned}
\tag{19}
$$

where $Z_t(\boldsymbol{c}) = \sum_{\boldsymbol{x}_0} p_{\mathrm{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0,\boldsymbol{x}_t)\exp(r(\boldsymbol{x}_0,\boldsymbol{c})/\beta)$ is the timestep-aware partition function. From the preceding equation Eq. (19), we can express the closed-form solution for the optimal policy $p_{\theta^*}^{\boldsymbol{c}}(\boldsymbol{x}_0,\boldsymbol{x}_t)$ at timestep t:

$$
p_{\theta^*}^{\boldsymbol{c}}(\boldsymbol{x}_0,\boldsymbol{x}_t) = p_{\mathrm{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0,\boldsymbol{x}_t)\exp(r_t^{\boldsymbol{c}}(\boldsymbol{x}_0,\boldsymbol{x}_t)/\beta)/Z_t(\boldsymbol{c}) \tag{20}
$$

A straightforward transformation of Eq. (20) leads to the solution for the 'joint' reward at timestep t:

$$
r_t^{\boldsymbol{c}}(\boldsymbol{x}_0,\boldsymbol{x}_t) = \beta\log\frac{p_{\theta^*}^{\boldsymbol{c}}(\boldsymbol{x}_0,\boldsymbol{x}_t)}{p_{\mathrm{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0,\boldsymbol{x}_t)} + \beta\log Z_t(\boldsymbol{c}) \tag{21}
$$

Then from Eq. (9), we can derive the expression for the 'initial' reward:

$$
r(\boldsymbol{x}_0,\boldsymbol{c}) = \beta\mathbb{E}_{p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_t|\boldsymbol{x}_0)}\left[\log\frac{p_{\theta^*}^{\boldsymbol{c}}(\boldsymbol{x}_0,\boldsymbol{x}_t)}{p_{\mathrm{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0,\boldsymbol{x}_t)}\right] + \beta\log Z_t(\boldsymbol{c}) \tag{22}
$$

By reparameterizing this reward and substituting it into the maximum likelihood objective of the Bradley-Terry model, as described in Eq. (3), we derive a maximum likelihood objective defined for a timestep-aware single-step diffusion model. At timestep t it is expressed as:

$$
\mathcal{L}_t(\theta) := -\log\sigma\left(\beta\mathbb{E}_{\boldsymbol{x}_t^w\sim p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_t^w|\boldsymbol{x}_0^w),\boldsymbol{x}_t^l\sim p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_t^l|\boldsymbol{x}_0^l)}\left[\log\frac{p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_0^w,\boldsymbol{x}_t^w)}{p_{\mathrm{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0^w,\boldsymbol{x}_t^w)} - \log\frac{p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_0^l,\boldsymbol{x}_t^l)}{p_{\mathrm{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0^l,\boldsymbol{x}_t^l)}\right]\right) \tag{23}
$$

To ensure training stability, we accounted for all time steps, leading to the following result:

$$
\mathcal{L}(\theta) := -\mathbb{E}_{t,(\boldsymbol{x}_0^w,\boldsymbol{x}_0^l,\boldsymbol{c})\sim\mathcal{D}}\log\sigma\left(\beta\mathbb{E}_{\boldsymbol{x}_t^w\sim p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_t^w|\boldsymbol{x}_0^w),\boldsymbol{x}_t^l\sim p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_t^l|\boldsymbol{x}_0^l)}\left[\log\frac{p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_0^w,\boldsymbol{x}_t^w)}{p_{\mathrm{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0^w,\boldsymbol{x}_t^w)} - \log\frac{p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_0^l,\boldsymbol{x}_t^l)}{p_{\mathrm{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0^l,\boldsymbol{x}_t^l)}\right]\right) \tag{24}
$$

where $\boldsymbol{x}_0^w$, $\boldsymbol{x}_0^l$ are from preference dataset.

Preference Optimization via Inversion For the sake of simplicity, we approximate $p_{\mathrm{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0,\boldsymbol{x}_t)$ with $p_{\mathrm{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0|\boldsymbol{x}_t)p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_t)$, Consequently, the above equation can be simplified as:

$$
\begin{aligned}
\mathcal{L}(\theta) :&= -\mathbb{E}_{t,(\boldsymbol{x}_0^w,\boldsymbol{x}_0^l,\boldsymbol{c})\sim\mathcal{D}}\log\sigma\left(\beta\mathbb{E}_{\boldsymbol{x}_t^w\sim p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_t^w|\boldsymbol{x}_0^w),\boldsymbol{x}_t^l\sim p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_t^l|\boldsymbol{x}_0^l)}\left[\log\frac{p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_0^w,\boldsymbol{x}_t^w)}{p_{\mathrm{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0^w,\boldsymbol{x}_t^w)} - \log\frac{p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_0^l,\boldsymbol{x}_t^l)}{p_{\mathrm{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0^l,\boldsymbol{x}_t^l)}\right]\right) \\
&= -\mathbb{E}_{t,(\boldsymbol{x}_0^w,\boldsymbol{x}_0^l,\boldsymbol{c})\sim\mathcal{D}}\log\sigma\left(\beta\mathbb{E}_{\boldsymbol{x}_t^w\sim p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_t^w|\boldsymbol{x}_0^w),\boldsymbol{x}_t^l\sim p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_t^l|\boldsymbol{x}_0^l)}\left[\log\frac{p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_0^w|\boldsymbol{x}_t^w)}{p_{\mathrm{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0^w|\boldsymbol{x}_t^w)} - \log\frac{p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_0^l|\boldsymbol{x}_t^l)}{p_{\mathrm{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0^l|\boldsymbol{x}_t^l)}\right]\right)
\end{aligned}
\tag{25}
$$

Here, we introduce a Gaussian probability density transition function $q^{\boldsymbol{c}}(\cdot|\cdot)$. It is evident that $q^{\boldsymbol{c}}(\boldsymbol{x}_0|\boldsymbol{x}_t, \boldsymbol{x}_0) = 1$, and we have:

$$\begin{aligned}\mathcal{L}(\theta) := -\mathbb{E}_{t,(\boldsymbol{x}_0^w,\boldsymbol{x}_0^l,\boldsymbol{c})\sim\mathcal{D}} \log\sigma(\beta\mathbb{E}_{\boldsymbol{x}_t^w\sim p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_t^w|\boldsymbol{x}_0^w),\boldsymbol{x}_t^l\sim p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_t^l|\boldsymbol{x}_0^l)} \\ \mathbb{E}_{\boldsymbol{x}_0^w\sim q^{\boldsymbol{c}}(\boldsymbol{x}_0^w|\boldsymbol{x}_t^w,\boldsymbol{x}_0^w),\boldsymbol{x}_0^l\sim q^{\boldsymbol{c}}(\boldsymbol{x}_0^l|\boldsymbol{x}_t^l,\boldsymbol{x}_0^l)} \left[\log\frac{p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_0^w|\boldsymbol{x}_t^w)}{p_{\text{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0^w|\boldsymbol{x}_t^w)} - \log\frac{p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_0^l|\boldsymbol{x}_t^l)}{p_{\text{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0^l|\boldsymbol{x}_t^l)}\right])\end{aligned} \tag{26}$$

In this context, $\boldsymbol{x}_0(t)$ serves as a reliable approximation of $\boldsymbol{x}_0$ at timestep t, that is $q^{\boldsymbol{c}}(\boldsymbol{x}_0^w|\boldsymbol{x}_t^w, \boldsymbol{x}_0^w) \approx q^{\boldsymbol{c}}(\boldsymbol{x}_0^w|\boldsymbol{x}_t^w, \boldsymbol{x}_0^w(t))$. Consequently, Eq. (26) can be estimated as:

$$\begin{aligned}\mathcal{L}(\theta) :\approx -\mathbb{E}_{t,(\boldsymbol{x}_0^w,\boldsymbol{x}_0^l,\boldsymbol{c})\sim\mathcal{D}} \log\sigma(\beta\mathbb{E}_{\boldsymbol{x}_t^w\sim p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_t^w|\boldsymbol{x}_0^w),\boldsymbol{x}_t^l\sim p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_t^l|\boldsymbol{x}_0^l)} \\ \mathbb{E}_{\boldsymbol{x}_0^w\sim q^{\boldsymbol{c}}(\boldsymbol{x}_0^w|\boldsymbol{x}_t^w,\boldsymbol{x}_0^w(t)),\boldsymbol{x}_0^l\sim q^{\boldsymbol{c}}(\boldsymbol{x}_0^l|\boldsymbol{x}_t^l,\boldsymbol{x}_0^l(t))} \left[\log\frac{p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_0^w|\boldsymbol{x}_t^w)}{p_{\text{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0^w|\boldsymbol{x}_t^w)} - \log\frac{p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_0^l|\boldsymbol{x}_t^l)}{p_{\text{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0^l|\boldsymbol{x}_t^l)}\right])\end{aligned} \tag{27}$$

Then according to Jensen's inequality, we can obtain:

$$\begin{aligned}\mathcal{L}(\theta) :\leq &- \mathbb{E}_{t,(\boldsymbol{x}_0^w,\boldsymbol{x}_0^l,\boldsymbol{c})\sim\mathcal{D}}\mathbb{E}_{\boldsymbol{x}_t^w\sim p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_t^w|\boldsymbol{x}_0^w),\boldsymbol{x}_t^l\sim p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_t^l|\boldsymbol{x}_0^l)} \log\sigma(\beta \\ &\mathbb{E}_{\boldsymbol{x}_0^w\sim q^{\boldsymbol{c}}(\boldsymbol{x}_0^w|\boldsymbol{x}_t^w,\boldsymbol{x}_0^w(t)),\boldsymbol{x}_0^l\sim q^{\boldsymbol{c}}(\boldsymbol{x}_0^l|\boldsymbol{x}_t^l,\boldsymbol{x}_0^l(t))} \left[\log\frac{p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_0^w|\boldsymbol{x}_t^w)}{p_{\text{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0^w|\boldsymbol{x}_t^w)} - \log\frac{p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_0^l|\boldsymbol{x}_t^l)}{p_{\text{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0^l|\boldsymbol{x}_t^l)}\right]) \\ &= -\mathbb{E}_{(\boldsymbol{x}_0^w,\boldsymbol{x}_0^l,\boldsymbol{c})\sim\mathcal{D},t\sim\mathcal{U}(0,T),\boldsymbol{x}_t^w\sim p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_t^w|\boldsymbol{x}_0^w),\boldsymbol{x}_t^l\sim p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_t^l|\boldsymbol{x}_0^l)} \log\sigma(-\beta(\\ &\quad \mathbb{D}_{\text{KL}}(q^{\boldsymbol{c}}(\boldsymbol{x}_0^w|\boldsymbol{x}_t^w, \boldsymbol{x}_0^w(t))\|p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_0^w|\boldsymbol{x}_t^w)) - \mathbb{D}_{\text{KL}}(q^{\boldsymbol{c}}(\boldsymbol{x}_0^w|\boldsymbol{x}_t^w, \boldsymbol{x}_0^w(t))\|p_{\text{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0^w|\boldsymbol{x}_t^w)) \\ &\quad - \mathbb{D}_{\text{KL}}(q^{\boldsymbol{c}}(\boldsymbol{x}_0^l|\boldsymbol{x}_t^l, \boldsymbol{x}_0^l(t))\|p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_0^l|\boldsymbol{x}_t^l)) + \mathbb{D}_{\text{KL}}(q^{\boldsymbol{c}}(\boldsymbol{x}_0^l|\boldsymbol{x}_t^l, \boldsymbol{x}_0^l(t))\|p_{\text{ref}}^{\boldsymbol{c}}(\boldsymbol{x}_0^l|\boldsymbol{x}_t^l)))\end{aligned} \tag{28}$$

Using Eq. (16) and the definition of the 'initial' variable(Eq. (8)), we can simplify the aforementioned loss function as:

$$\begin{aligned}\mathcal{L}(\theta) = &- \mathbb{E}_{(\boldsymbol{x}_0^w,\boldsymbol{x}_0^l,\boldsymbol{c})\sim\mathcal{D},t\sim\mathcal{U}(0,T),\boldsymbol{x}_t^w\sim p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_t^w|\boldsymbol{x}_0^w),\boldsymbol{x}_t^l\sim p_\theta^{\boldsymbol{c}}(\boldsymbol{x}_t^l|\boldsymbol{x}_0^l)} \log\sigma(-\beta w(t)(\\ &\|\tau_t^w - \epsilon_\theta^t(\boldsymbol{x}_t^w, \boldsymbol{c})\|_2^2 - \|\tau_t^w - \epsilon_{\text{ref}}^t(\boldsymbol{x}_t^w, \boldsymbol{c})\|_2^2 - \|\tau_t^l - \epsilon_\theta^t(\boldsymbol{x}_t^l, \boldsymbol{c})\|_2^2 + \|\tau_t^l - \epsilon_{\text{ref}}^t(\boldsymbol{x}_t^l, \boldsymbol{c})\|_2^2))\end{aligned} \tag{29}$$

where $\tau_t^* = \frac{\boldsymbol{x}_t^* - \sqrt{\alpha_t}\boldsymbol{x}_0^*}{\sqrt{1-\alpha_t}} = \frac{\sqrt{\alpha_t}\boldsymbol{x}_0^*(t) + \sqrt{1-\alpha_t}\delta_t(\boldsymbol{x}_0^*(t)) - \sqrt{\alpha_t}\boldsymbol{x}_0^*}{\sqrt{1-\alpha_t}} = (\boldsymbol{x}_0^*(t) - \boldsymbol{x}_0^*)/\sigma_t + \delta_t(\boldsymbol{x}_0^*(t))$ and $w(t)$ is a weight function defined consistent with [63].

DDIM ODE on image space We now provide a comprehensive derivation of Eq. (14), which represents the reparameterization process outlined in Eq. (2). This derivation follows the approach presented in [36]. To transform the variables from Eq. (14) to Eq. (8), we must differentiate the equation with respect to t and compute $\frac{\mathrm{d}\bar{\boldsymbol{x}}_t}{\mathrm{d}t}$. Performing this differentiation directly yields:

$$\frac{\mathrm{d}\boldsymbol{x}_0(t)}{\mathrm{d}t} = \frac{\mathrm{d}\bar{\boldsymbol{x}}_t}{\mathrm{d}t} - \epsilon_\theta^t\left(\frac{\boldsymbol{x}_t}{\sqrt{\sigma_t^2+1}}, \boldsymbol{c}\right)\frac{\mathrm{d}\sigma_t}{\mathrm{d}t} - \sigma_t\frac{\mathrm{d}}{\mathrm{d}t}\epsilon_\theta^t\left(\frac{\bar{\boldsymbol{x}}_t}{\sqrt{\sigma_t^2+1}}, \boldsymbol{c}\right). \tag{30}$$

By solving for $\frac{\mathrm{d}\bar{\boldsymbol{x}}_t}{\mathrm{d}t}$ and combining the result with Eq. (8), we obtain:

$$\begin{aligned}\epsilon_\theta^t\left(\frac{\boldsymbol{x}_t}{\sqrt{\sigma_t^2+1}}, \boldsymbol{c}\right)\frac{\mathrm{d}\sigma_t}{\mathrm{d}t} &= \frac{\mathrm{d}\boldsymbol{x}_0(t)}{\mathrm{d}t} + \epsilon_\theta^t\left(\frac{\boldsymbol{x}_t}{\sqrt{\sigma_t^2+1}}, \boldsymbol{c}\right)\frac{\mathrm{d}\sigma_t}{\mathrm{d}t} + \sigma_t\frac{\mathrm{d}}{\mathrm{d}t}\epsilon_\theta^t\left(\frac{\bar{\boldsymbol{x}}_t}{\sqrt{\sigma_t^2+1}}, \boldsymbol{c}\right) \\ \frac{\mathrm{d}\boldsymbol{x}_0(t)}{\mathrm{d}t} &= -\sigma_t\frac{\mathrm{d}}{\mathrm{d}t}\epsilon_\theta^t\left(\frac{\bar{\boldsymbol{x}}_t}{\sqrt{\sigma_t^2+1}}, \boldsymbol{c}\right) \\ \frac{\mathrm{d}\boldsymbol{x}_0(t)}{\mathrm{d}t} &= -\sigma_t\frac{\mathrm{d}}{\mathrm{d}t}\epsilon_\theta^t\left(\frac{\boldsymbol{x}_0(t) + \sigma_t\delta_t(\boldsymbol{x}_0(t))}{\sqrt{\sigma_t^2+1}}, \boldsymbol{c}\right) \\ \frac{\mathrm{d}\boldsymbol{x}_0(t)}{\mathrm{d}t} &= -\sigma_t\frac{\mathrm{d}}{\mathrm{d}t}\epsilon_\theta^t\left(\sqrt{\alpha_t}\boldsymbol{x}_0(t) + \sqrt{1-\alpha_t}\delta_t(\boldsymbol{x}_0(t)), \boldsymbol{c}\right)\end{aligned} \tag{31}$$

It represents a velocity field that maps an initial image to a conditional distribution learned by the diffusion model. The discretized form of inversion leads to Eq. (15).

S3: Choice of δ_t

A crucial component of our algorithm is the computation and selection of noise. In theory, $\delta_t(\boldsymbol{x}_0(t))$ should satisfy Eq. (12). Howeverm, solving it directly is computationally impractical. We present several alternative approaches that are less effective compared to our method:

- Fixed-Point Iteration: Given that the model is a known function, we initialize the process with a random noise $\sim \mathcal{N}(0, I)$ as the initial guess and refine the solution iteratively using a predefined iterative formula.
- Gradient Optimization Methods: Approaches such as Newton's method or SGD, initialized with random noise $\sim \mathcal{N}(0, I)$, are employed to refine the solution iteratively.

These methods for estimating the equation are computationally intensive and lack efficiency. Consequently, we opted not to focus on such equation-solving approaches and instead adopted DDIM inversion for its efficiency and practicality.

S4: Discussion of dataset

The Pick-a-Pic v2 train and test sets have several notable drawbacks, including the exploitation of vulnerable groups, misrepresentation or defamation of real individuals, and the portrayal of unrealistic or objectifying body imagery. Additionally, it contains harmful or offensive content, as well as explicit or sexual material. These issues emphasize the need for careful curation to avoid ethical concerns and ensure the dataset's responsible use.

S5: Discussion

Since we conduct experiments on a given dataset Pick-a-Pic v2, the data is not generated by the diffusion model itself. The work of [63] indicates that the quality of the Pick-a-Pic v2 dataset falls between SDXL and SD1.5, as it is sourced from SDXL-Beta and Dreamlike. Our experimental results show that supervised fine-tuning SD1.5 enhances its performance, whereas any level of fine-tuning on SDXL results in a decline in model metrics. Therefore, we recommend maxmizing the likelihood of preferred pairs $(\boldsymbol{x}^w, \boldsymbol{c})$ before aligning the SD1.5 with human preferences, that is, $p^{\boldsymbol{c}}_{\text{ref}} = \arg\max_p \mathbb{E}_{(\boldsymbol{x}^w,\boldsymbol{c})\sim\mathcal{D}}[\log p(\boldsymbol{x}^w|\boldsymbol{c})]$. In contrast, the SDXL model utilizes the parameters of itself.

In T2I diffusion models, human preference feedback is shaped by factors such as image quality, realism, artistic style, and cultural background. These factors are highly subjective, and the presence of noise in datasets makes it challenging for AI to effectively learn from them, underscoring the importance of robust preference learning. Additionally, the diversity and inherent uncertainty of human preferences during the T2I diffusion process introduce significant modeling complexities and may lead to distributional shifts. For instance, most preference data is derived from Stable Diffusion variants, and applying this data to other T2I models (e.g., Midjourney or DALL·E 3) may result in distributional mismatches, causing inconsistencies between the model outputs and human preferences. This issue arises because these models are trained on distinct data distributions, leading to potential training conflicts.

S6: Experiment Details

Pick-a-Pic v2 The Pick-a-Pic dataset is a text-to-image pair dataset that gathers user feedback from the Pick-a-Pic web application. Each image pair (comprising two images) is associated with a text prompt and a label reflecting the user's preference. The dataset includes images generated by various text-to-image models, such as Stable Diffusion 2.1, Dreamlike Photoreal, and variants of Stable Diffusion XL, with a wide range of Classifier-Free Guidance (CFG) values. In this paper, we use its training data. In the supplementary materials, we provided additional quantitative comparative analyses on the test dataset to further validate our approach.

HPDv2 HPDv2 collects human preference data via the "Dreambot" channel on the Stable Foundation Discord server. It contains 25,205 text prompts used to generate a total of 98,807 images. Each text prompt is associated with multiple generated images and paired image labels, where the label denotes the user's preferred choice between two images. The number of generated images per text prompt varies across the dataset. In this paper, we use its test data of text prompt (3200 prompts).

Parti-Prompts Parti-Prompts is a comprehensive dataset consisting of 1,632 text prompts, specifically designed to evaluate and benchmark the capabilities of text-to-image generation models. Covering multiple categories, these prompts offer diverse challenges that facilitate a thorough assessment of model performance across various dimensions.

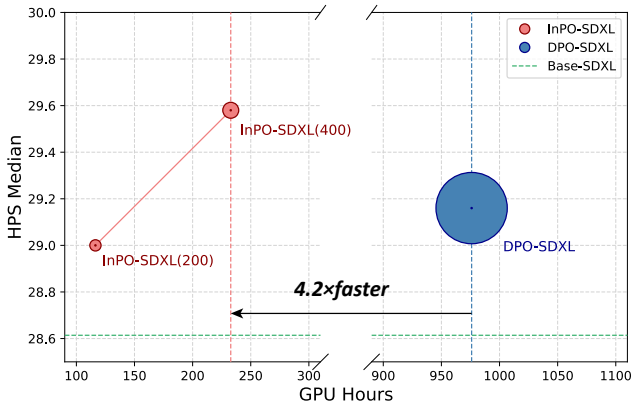


Figure 8. Comparison of the trade-off between the quality of generated images and training efficiency following human preference optimization of SDXL on the HPDv2 test set. Sizes of the circles represent the volume of training data used. Our DDIM-InPO achieves superior performance, with a training speed that is 4.2 times faster than Diffusion-DPO while producing images of higher quality.

Additional Implementation details During the evaluation and comparison phase for text-to-image generation, the inference CFG is set to 7.5 for SD1.5 and 5 for SDXL, which are widely recognized as standard and recommended configurations. In generation tasks conditioned on depth maps and canny edges, we set the ControlNet conditioning scale to 0.5 and the CFG to 5. For inpainting tasks, the strength parameter is set to 0.85, with the CFG also set to 5.

Chart Explanation To facilitate clearer comparisons, we scale PickScore, HPS and the CLIP score by a factor of 100 and retain 5 significant figures for precision, including Fig. 4, Fig. 8, Fig. 10, Tab. 2, Tab. 3, Tab. 4, Tab. 5, Tab. 6, Tab. 7, Tab. 8.

Commitment We are committed to releasing the code and models for the open-source community.

S7: Additional Quantitative Results

In this section, we present additional qualitative results. First, we present additional comparison of the trade-off between the quality of generated images and training efficiency of SDXL on the HPDv2 test set. Subsequently, we also provide automatic preference evaluation comparison of different prompt sets conducted on SDXL and SD1.5.

- Fig. 8 shows the comparison of the trade-off between the quality of generated images and training efficiency following human preference optimization of SDXL on the HPDv2 test set. We present the results of training for 200 and 400 steps using our DDIM-InPO and compare them with those of Diffusion-DPO finetuned for 2000 steps.
- Tab. 3 and Tab. 4 demonstrate evaluation comparison on SDXL and SD1.5 using HPDv2 test set, respectively.
- Tab. 5 and Tab. 6 demonstrate evaluation comparison on SDXL and SD1.5 using Parti-Prompts, respectively.
- Tab. 7 and Tab. 8 demonstrate evaluation comparison on SDXL and SD1.5 using Pick-a-Pic test set., respectively.

Experimental Result Analysis Overall, following fine-tuning with DDIM-InPO, both SD1.5 and SDXL achieve superior performance compared to the baselines across nearly all evaluators and test datasets, thereby validating the effectiveness of our method. Fig. 8 shows that our DDIM-InPO achieves better performance, with a training speed that is 4.2 times faster than Diffusion-DPO while producing images of higher quality. All tables clearly show that while supervised fine-tuning performs well on SD1.5, its application to SDXL results in significant degradation of the base model, making it an ineffective and non-generalizable approach. This limitation stems from the dataset quality, as the Pick-a-Pic dataset is inferior to the outputs generated by the SDXL base model. In comparison, Diffusion-DPO proves to be a more robust alternative, delivering consistent improvements on both SD1.5 and SDXL. However, although Diffusion-KTO achieves notable gains on SD1.5, its high computational demands prevent effective scalability to SDXL models. By contrast, our model emerges as a more effective and efficient solution, achieving state-of-the-art results across nearly all evaluators and test datasets on both SDXL and SD1.5. These findings highlight the suitability of our approach for diffusion models, along with its significant advantage in training speed.

Baselines	Aesthetic		PickScore		HPS		CLIP	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
Base-SDXL	6.1143	6.1346	22.756	22.781	28.614	28.624	38.360	38.155
SFT-SDXL	5.8049	5.8327	21.524	21.380	27.467	27.204	37.217	36.531
DPO-SDXL	6.1124	6.1310	23.133	23.152	29.165	29.174	**38.865**	**38.711**
InPO-SDXL	**6.1676**	**6.1820**	**23.254**	**23.274**	**29.576**	**29.550**	38.627	38.449

Table 3. Automatic preference evaluation comparison to existing alignment baselines on SDXL using prompts from HPDv2 test set. We use median and mean values of four evaluators. To ensure clarity in comparisons, Pickscore, HPS, and CLIP scores are scaled by 100, and all evaluator values retain precision to five significant figures. In the table, the maximum value in each column is bolded, while the second-highest value is underlined.

Baselines	Aesthetic		PickScore		HPS		CLIP	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
Base-SD1.5	5.3491	5.3848	20.719	20.727	26.647	26.633	34.276	33.945
SFT-SD1.5	5.7255	5.7515	21.647	21.648	28.032	27.977	36.292	35.845
DPO-SD1.5	5.5219	5.5841	21.274	21.297	27.428	27.392	35.591	35.197
KTO-SD1.5	5.6922	5.7248	21.566	21.582	28.376	28.306	35.902	35.648
InPO-SD1.5	**5.7734**	**5.8056**	**21.894**	**21.916**	**28.523**	**28.502**	**36.876**	**36.495**

Table 4. Automatic preference evaluation comparison to existing alignment baselines on SD1.5 using prompts from HPDv2 test set. We use median and mean values of four evaluators. To ensure clarity in comparisons, Pickscore, HPS, and CLIP scores are scaled by 100, and all evaluator values retain precision to five significant figures. In the table, the maximum value in each column is bolded, while the second-highest value is underlined.

Baselines	Aesthetic		PickScore		HPS		CLIP	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
Base-SDXL	5.7519	5.7681	22.648	22.628	28.447	28.424	35.550	35.531
SFT-SDXL	5.5373	5.5403	21.666	21.554	27.213	27.085	34.827	34.696
DPO-SDXL	5.8181	5.7942	22.914	22.928	28.885	28.906	**36.401**	**36.457**
InPO-SDXL	**5.8493**	**5.8566**	**23.039**	**23.005**	**29.123**	**29.143**	35.914	35.903

Table 5. Automatic preference evaluation comparison to existing alignment baselines on SDXL using prompts from Parti-Prompts. We use median and mean values of four evaluators. To ensure clarity in comparisons, Pickscore, HPS, and CLIP scores are scaled by 100, and all evaluator values retain precision to five significant figures. In the table, the maximum value in each column is bolded, while the second-highest value is underlined.

Baselines	Aesthetic		PickScore		HPS		CLIP	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
Base-SD1.5	5.3494	5.3132	21.406	21.389	27.291	27.172	33.065	33.128
SFT-SD1.5	5.5798	5.5506	21.803	21.759	28.192	28.129	33.887	33.956
DPO-SD1.5	5.4445	5.3874	21.619	21.631	27.596	27.511	33.551	33.694
KTO-SD1.5	5.5466	5.5110	21.755	21.736	28.240	28.110	34.101	34.013
InPO-SD1.5	**5.6056**	**5.5698**	**21.957**	**21.923**	**28.431**	**28.325**	**34.533**	**34.683**

Table 6. Automatic preference evaluation comparison to existing alignment baselines on SD1.5 using prompts from Parti-Prompts. We use median and mean values of four evaluators. To ensure clarity in comparisons, Pickscore, HPS, and CLIP scores are scaled by 100, and all evaluator values retain precision to five significant figures. In the table, the maximum value in each column is bolded, while the second-highest value is underlined.

Baselines	Aesthetic		PickScore		HPS		CLIP	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
Base-SDXL	5.9775	6.0057	22.219	22.159	27.995	27.978	36.470	36.124
SFT-SDXL	5.6416	5.6452	21.028	21.003	26.790	26.681	35.589	35.427
DPO-SDXL	<u>6.0179</u>	<u>6.0160</u>	<u>22.581</u>	<u>22.627</u>	<u>28.515</u>	<u>28.586</u>	**37.404**	**37.392**
InPO-SDXL	**6.0372**	**6.0558**	**22.606**	**22.692**	**28.824**	**28.817**	<u>37.130</u>	<u>36.842</u>

Table 7. Automatic preference evaluation comparison to existing alignment baselines on SDXL using prompts from Pick-a-Pic v2 test set. We use median and mean values of four evaluators. To ensure clarity in comparisons, Pickscore, HPS, and CLIP scores are scaled by 100, and all evaluator values retain precision to five significant figures. In the table, the maximum value in each column is bolded, while the second-highest value is underlined.

Baselines	Aesthetic		PickScore		HPS		CLIP	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
Base-SD1.5	5.3545	5.3296	20.632	20.661	26.527	26.480	33.023	32.619
SFT-SD1.5	<u>5.6441</u>	<u>5.6285</u>	<u>21.278</u>	<u>21.253</u>	<u>27.707</u>	27.509	<u>34.050</u>	<u>34.144</u>
DPO-SD1.5	5.5258	5.4654	21.020	21.053	27.098	26.913	33.270	33.302
KTO-SD1.5	5.6029	5.5831	21.184	21.190	27.645	<u>27.580</u>	34.003	33.910
InPO-SD1.5	**5.6810**	**5.6585**	**21.456**	**21.490**	**27.866**	**27.765**	**34.782**	**34.728**

Table 8. Automatic preference evaluation comparison to existing alignment baselines on SD1.5 using prompts from Pick-a-Pic v2 test set. We use median and mean values of four evaluators. To ensure clarity in comparisons, Pickscore, HPS, and CLIP scores are scaled by 100, and all evaluator values retain precision to five significant figures. In the table, the maximum value in each column is bolded, while the second-highest value is underlined.

S8: Additional AI Preference

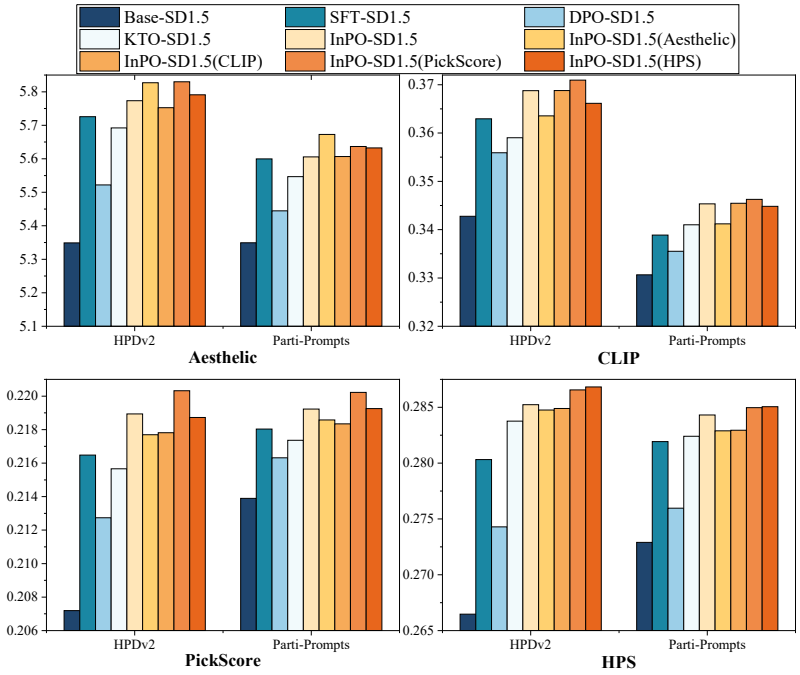


Figure 9. Median of Aesthelc, CLIP score, PickScore and HPS comparisons for all baselines and test datasets on SD1.5.

In this section, we extend the AI preference experiments discussed in the main text. Additionally, we incorporate aesthetic

classifiers and CLIP as evaluators, where higher scores reflect stronger AI preferences. Fig. 9 reveal that training with self-selected images leads to improved scores. Specifically, training with images selected by the aesthetic classifier results in higher aesthetic metrics, and a similar pattern is observed for Pickscore and HPS. Our findings indicate that Pickscore and HPS effectively emulate human preferences, enabling models trained with these metrics to surpass InPO-SD1.5. Conversely, models trained with CLIP-based preference selection exhibit relatively lower scores, suggesting that text alignment plays a less significant role in preference selection.

S9: Additional Ablations

In this section, we introduce additional ablation experiments, primarily exploring whether the number of training timesteps can be reduced. Specifically, we consider the denoiser as timestep-aware, with the total timesteps for DDPM denoising set to 1000. Can we train only on the last 900, 800, or even fewer timesteps to accelerate the training process?

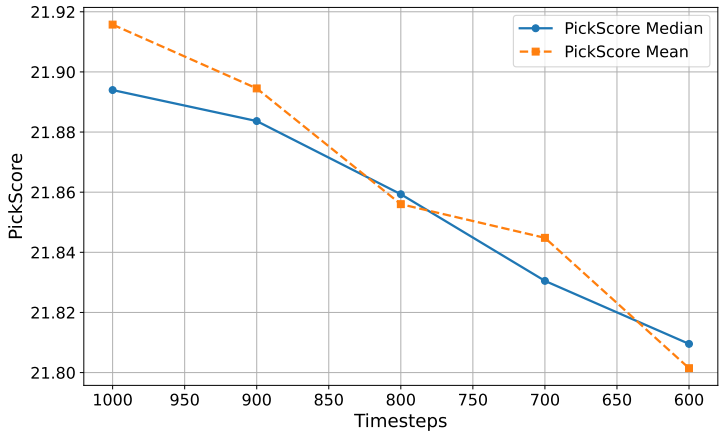


Figure 10. Timestep ablation studies of our DDIM-InPO method for fine-tuning SD1.5, evaluated on the HPDv2 test set using both median and mean PickScore metrics.

As shown in Fig. 10, our method demonstrates strong robustness. While reducing the training timesteps results in a slight performance decline, it still achieves significant improvements over baselines such as KTO-SD1.5, DPO-SD1.5, and SFT-SD1.5 (refer to Tab. 4). This suggests potential for future exploration in optimizing training efficiency by reducing timesteps. Furthermore, investigating the trade-off between training timesteps and inversion steps could provide deeper insights into balancing efficiency and performance, further underscoring the potential of our approach.

S10: Additional Qualitative results

In this section, we present additional qualitative results, including evaluations conducted on SD1.5 and SDXL.

- Fig. 11 showcases the qualitative generation results of InPO-SDXL across diverse prompts. Fig. 12 and Fig. 13 display outputs generated with different random seeds under the same prompt, where the seeds are randomly chosen from the range 0 to 15. These results highlight that our fine-tuned model not only preserves the capabilities of SDXL but also produces high-quality outputs that align with human preferences.
- Fig. 14 and Fig. 15 provide a qualitative comparison between InPO-SD1.5 and the SD1.5 baselines, using prompts sourced from HPDv2. The results reveal that our model exhibits superior text alignment, enhanced visual appeal, and a stronger consistency with human preferences.
- Fig. 16 provides an additional qualitative qualitative evaluation of InPO-SDXL in comparison with Base-SDXL and DPO-SDXL on T2I generation tasks, further highlighting the effectiveness of our approach in demonstrating improved performance.
- Fig. 17 provides an additional qualitative evaluation of InPO-SDXL in comparison with Base-SDXL and DPO-SDXL on conditional generation tasks, including depth map, canny edge, and inpainting.
- Due to safety concerns, failed cases are unsuitable for presentation in the paper. A small subset of images occasionally exhibit an excessively feminized style.

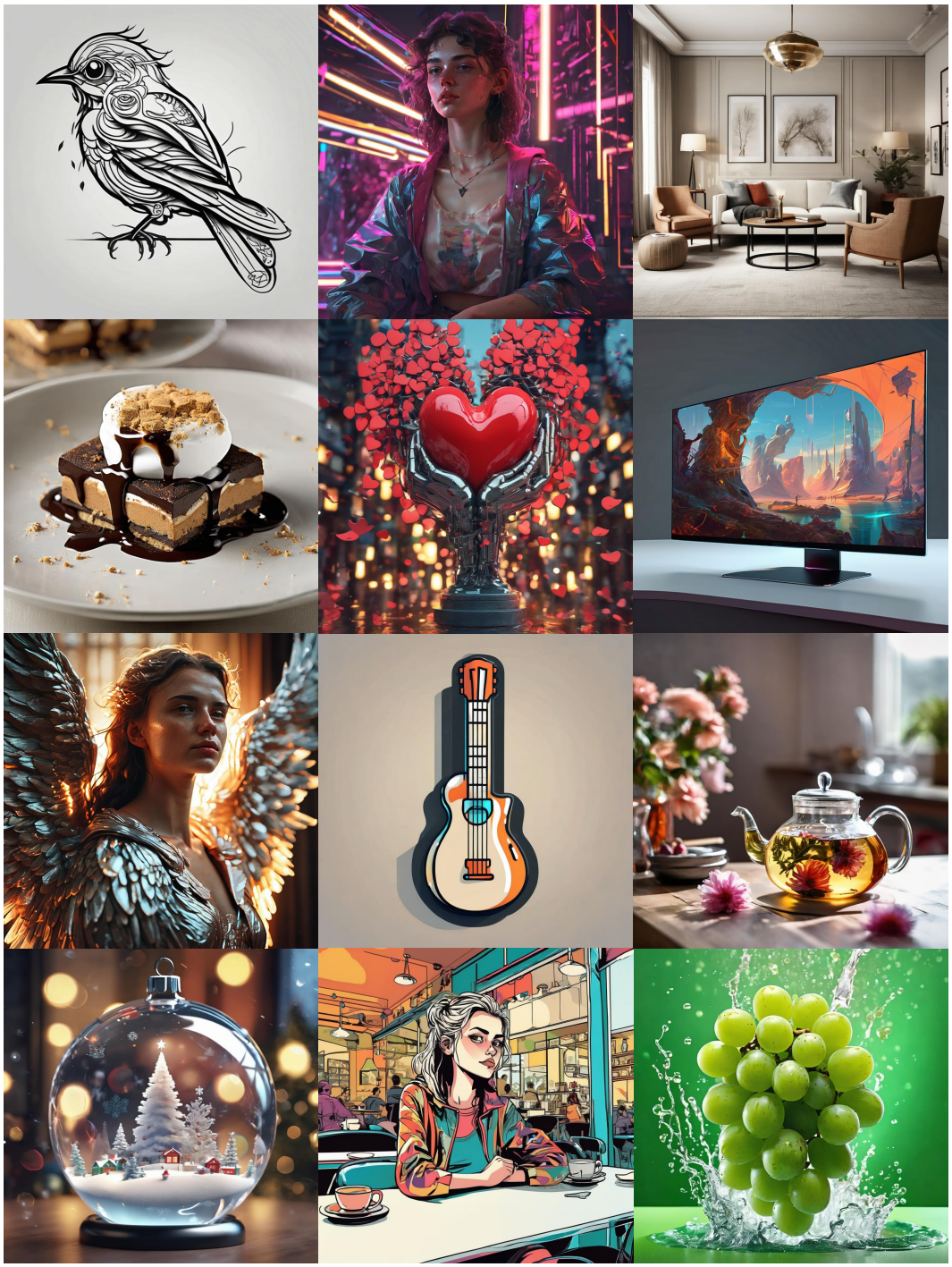
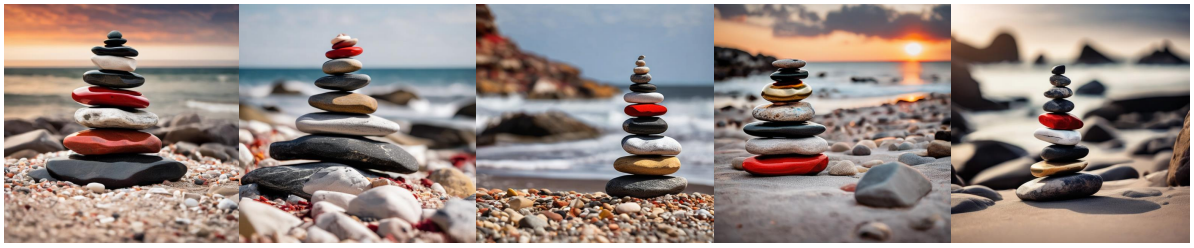


Figure 11. Additional qualitative results of InPO-SDXL for **various prompts**, arranged from left to right and top to bottom. *Prompts: (1) Tattoo design: a tattoo design, a small bird, minimalistic, black and white drawing, detailed, 8k. (2) Young glitchy woman, beautiful girl, 8k, unreal engine, illustration, trending on artstation, masterpiece. (3) A three-seater sofa, capet,end table,west elm chandelier,armchair, modrtn organic style, crrisp lines, neutral colors, backdrop of simplicity, bright environment. (4) An indulgent dessert featuring charred marshmallow, chocolate fondant, and graham cracker crumbs. (5) Hearts, in the style of jamie hewlett killian eng kawase hasui riyoko ikeda, artstation trending, 8 k, octane render, photorealistic, volumetric lighting caustics, surreal. (6) A sleek, ultra-thin, high resolution bezel-less monitor mockup, realistic, modern, digital illustration, trending on Artstation, high-tech, smooth, minimalist workstation background, crisp reflection on screen, soft lighting. (7) Portrait art of female angel, art by alessio albi 8 k ultra realistic, angel wings, lens flare, atmosphere, glow, detailed, intricate, cinematic lighting, trending on artstation, 4k, hyperrealistic, focused, extreme details, unreal engine 5, masterpiece. (8) Icon: a guitar, 2d minimalistic icon, flat vector illustration, digital, smooth shadows, design asset. (9) Drink photography: freshly made hot floral tea in glass kettle on the table, angled shot, midday warm, Nikon D850 105mm, close-up. (10) Looking through a transparent glass Christmas ball , hyper-realistic, minimalist, futuristic background with cute Christmas decorations like Santa Claus and snowflakes, 8k. (11) Comicbook: a girl sitting in the cafe, comic, graphic illustration, comic art, graphic novel art, vibrant, highly detailed, colored, 2d minimalistic. (12) Powerful liquid explosion, green grapes, green background, commercial photography, a bright environment, studio lighting, OC rendering, isolated platform, professional photography.*

B&w photography, model shot, man in subway station, beautiful detailed eyes, professional award winning portrait photography, highly detailed glossy eyes, high detailed skin, skin pores

A mystical white cat with wings standing gracefully in ocean

A beautiful stack of rocks sitting on top of a beach, a picture, red black white golden colors, chakras, packshot, stock photo

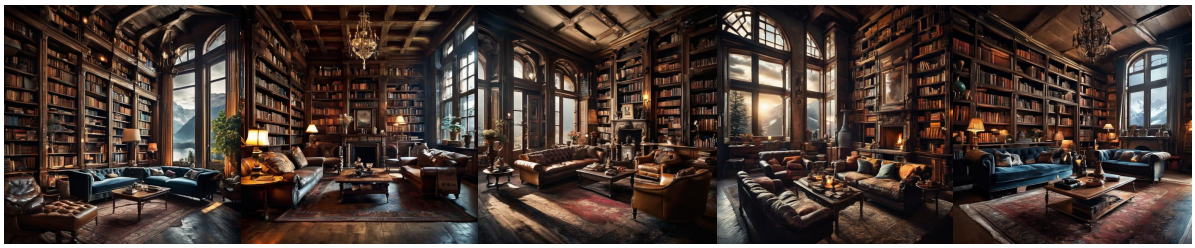


Anime astronaut: a girl astronaut exploring the cosmos, floating among planets and stars, high quality detail, , anime screencap, studio ghibli style, illustration, high contrast, masterpiece, best quality

A beautiful oil painting of a question mark, ?, with thick messy brush strokes, majestic

Figure 12. Additional qualitative results of InPO-SDXL (A). The seeds are chosen from the range of 0 to 15. After fine-tuning with our method, the model not only retains its original generative capabilities but also produces images that align with human preferences.

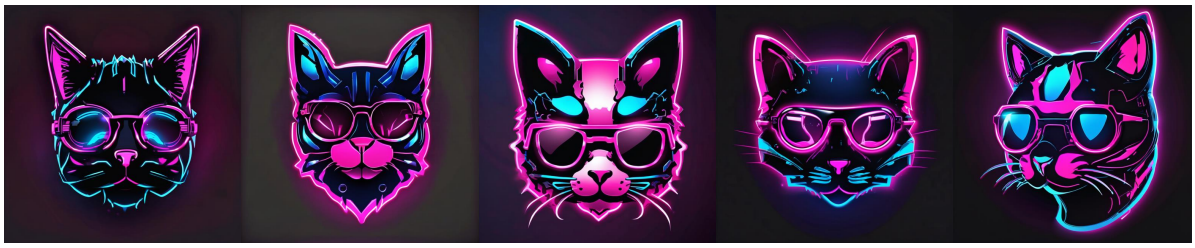
A striking living room interior, sofa furniture, a living room table, bookshelves, shelving, a fireplace, elegant interior design, perfect layout, moody, hazy, cinematic, surreal, high detail, intricate, masterpiece, golden ratio



Papercut-style portrait of a woman wearing a hat

A cottage . flat 2d emoji icon, cute, crisp outlined icon art, colorful, centered.

Neon symbol: symbol of a stylized pink cat head with sunglasses, glowing, neon, logo for a game, cyberpunk, vector, dark background with black and blue abstract shadows, cartoon, simple



Soft pink roses, white Chinese peony, tiny apple blossom flowers, eucalyptus leaves, twigs of cranberries, twigs of copper pepper berries all arrangement into a cute beautiful flowers arrangement on a nickel mug., 8k, Product.

Figure 13. Additional qualitative results of InPO-SDXL (B). The seeds are chosen from the range of 0 to 15. After fine-tuning with our method, the model not only retains its original generative capabilities but also produces images that align with human preferences.

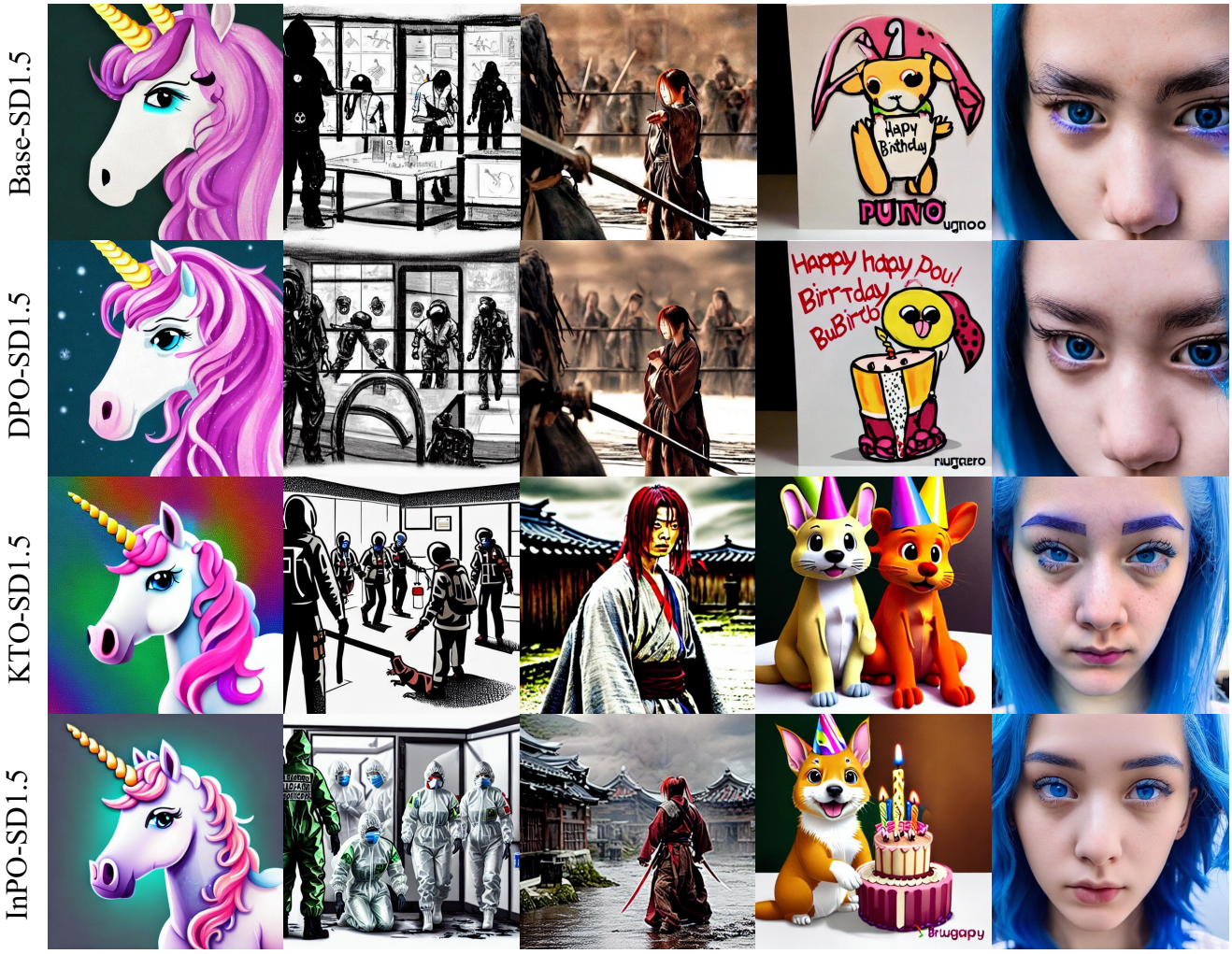


Figure 14. Qualitative comparisons among baselines of SD1.5 (A). InPO-SD1.5 achieves superior prompt alignment and produces images of higher quality. Prompts from left to right: *(1) A cute digital art of a unicorn. (2) A detailed, realistic image of a biohazard lab evacuation with horror influences and multiple art styles incorporated. (3) An apocalyptic scene from Kenshin. (4) A birthday greeting for Pungeroo. (5) A blue-haired girl with soft features stares directly at the camera in an extreme close-up Instagram picture.*

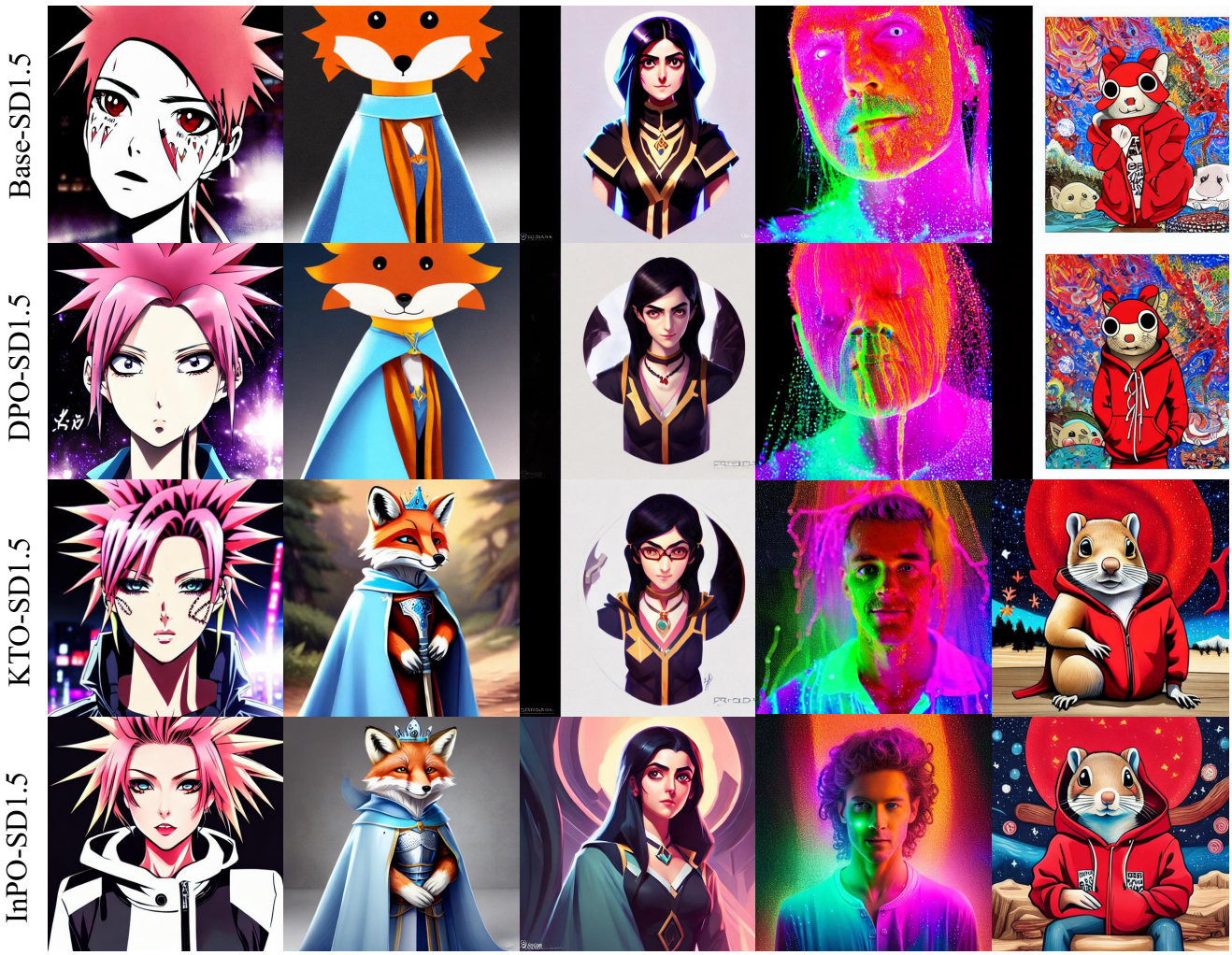


Figure 15. Qualitative comparisons among baselines of SD1.5 (B). InPO-SD1.5 achieves superior prompt alignment and produces images of higher quality. Prompts from left to right: *(1) Anime poster of a woman wearing futuristic streetwear with spiky hair, featuring intricate eyes and a pretty face. (2) A cute anthropomorphic fox knight wearing a cape and crown in pale blue armor. (3) Head-on centered portrait of Maya Ali as a black-haired RPG mage, depicted in stylized concept art for a Blizzard game, by Lois Van Baarle, Ilya Kuvshinov, and RossDraws. (4) A human portrait formed out of neon rain on a galactic background. (5) A water squirrel spirit wearing a red hoodie sits under the stars, surrounded by artwork from various artists.*

Figure 16. Additional qualitative evaluation of InPO-SDXL in comparison with Base-SDXL and DPO-SDXL on T2I generation tasks. Prompts from top to bottom: *(1) A teddy bear on a skateboard in times square. (2) An avocado on a table. (3) A whale breaching near a mountain. (4) A cat drinking a pint of beer. (5) A towel with the word 'stop' printed on it, simple and clear text.*

Figure 17. Additional qualitative evaluation of InPO-SDXL in comparison with Base-SDXL and DPO-SDXL on conditional generation tasks (From left to right: depth map, canny edge, and inpainting). Prompts from left to right: *(1) A full moon rising above a mountain at night. (2) Sunset over misty mountains, cascading waterfalls, and soft god rays breaking through clouds, creating a realistic and serene atmosphere. (3) A big cyberpunk cat with glowing eyes, sitting majestically against a mountain backdrop.*