WHEN TO TRUST INVERSION: DRIFT-AWARE PREFERENCE OPTIMIZATION FOR DIFFUSION MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Direct Preference Optimization (DPO) has emerged as an effective paradigm for aligning text-to-image diffusion models with human preferences. Recent inversion-based methods replace random forward noising with deterministic inversion to provide more precise, image-specific preference supervision. However, as inversion progresses, the resulting latents increasingly drift away from states typically encountered during Gaussian-initialized denoising. This distributional mismatch is particularly pronounced at high-noise timesteps, undermining the effectiveness of inversion-based preference learning. To address this challenge, we propose Inversion-aware Fallback for Preference Optimization (IFPO), a drift-aware preference optimization framework with a dynamic fallback mechanism. IFPO mitigates the impact of inversion drift by adaptively reverting to standard forward noising at high-noise timesteps, while preserving the benefits of inversion-based supervision where it remains useful. Experiments across diffusion and flow-matching backbones demonstrate that IFPO consistently improves text-to-image alignment and visual fidelity over existing preference optimization methods.

1 INTRODUCTION

Despite the remarkable success of text-to-image (T2I) diffusion models (Ho et al., 2020; Song et al., 2021; Dhariwal & Nichol, 2021) in synthesizing high-fidelity images, aligning their outputs with fine-grained human preferences remains a critical challenge. To address this, Direct Preference Optimization (DPO) and its variants (Wallace et al., 2024; Wang et al., 2025; Zhu et al., 2025; Dai et al., 2025; Yuan et al., 2024; Peng et al., 2025; Na et al., 2025) have emerged as a principled paradigm that bypasses an explicit reward model and directly optimizes preference pairs. However, adapting DPO to diffusion models is inherently bottlenecked by the long Markov chains, leading to training inefficiency and sparse reward signals. To mitigate this, InPO (Lu et al., 2025) conceptualizes the diffusion model as a single-step generator and introduces DDIM inversion (Song et al., 2021; Couairon et al., 2023) to obtain deterministic latents, providing a more precise preference signal and enabling efficient fine-tuning.

Despite its efficiency, InPO's reliance on DDIM inversion surfaces a fundamental challenge: the **Gaussian reachability gap**—the distributional mismatch between inverted latents and states typically encountered during denoising initialized from pure Gaussian noise. As inversion progresses, the latents increasingly drift away from these typical denoising states, with the gap most pronounced in the early (high-noise) stages of denoising. Here, reachability is defined in statistical terms: inverted latents may occupy regions that are rarely visited during ordinary Gaussian-initialized denoising, even when reaching them is theoretically possible. This distinction matters when adapting inversion from image editing to preference optimization. In image editing, DDIM inversion is commonly used to reconstruct an input image, making reconstruction fidelity a primary objective (Mokady et al., 2023; Garibi et al., 2024). In preference optimization, however, the resulting latents serve as supervision for improving ordinary generation. Their ability to reconstruct an image therefore does not by itself establish their suitability for preference learning. The Gaussian reachability gap highlights this discrepancy between reconstruction-oriented inversion and generation-oriented supervision.

To address this reachability gap, we propose Inversion-aware Fallback for Preference Optimization (**IFPO**), a drift-aware preference optimization framework with a *dynamic fallback* mechanism. IFPO mitigates the impact of accumulated inversion drift by adaptively reverting to standard forward

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.



Figure 1: **Qualitative comparison of SDXL and IFPO.** SDXL (top) and IFPO (bottom) are compared across text rendering, objects, landscapes, portraits, and artistic styles. These examples illustrate improvements in text legibility and visual quality. The exact prompts are listed in Appendix D.1.

noising at high-noise timesteps, while preserving the benefits of inversion-based supervision where it remains useful.

Experiments across both diffusion (Podell et al., 2024) and flow-matching (Esser et al., 2024) backbones demonstrate that IFPO improves text-to-image alignment and visual fidelity over existing preference optimization methods. Figure 1 illustrates qualitative improvements over the pretrained SDXL backbone across diverse visual categories. Evaluations across multiple datasets support the effectiveness and generality of drift-aware preference learning.

Our contributions can be summarized as follows:

- We identify and characterize the **Gaussian reachability gap**: inversion-induced drift creates a mismatch between supervisory latents and typical Gaussian-initialized denoising states, compromising preference learning at high-noise timesteps.
- We propose IFPO, a drift-aware framework with dynamic fallback to mitigate the effects of inversion drift while retaining useful inversion-based supervision.
- Experiments on diffusion (SDXL) and flow-matching (SD3.5) backbones show gains in text-to-image alignment and visual fidelity across multiple datasets.

2 Related Work

2.1 Text-to-Image Generative Models

Diffusion models have become a highly effective framework for synthesizing high-quality images, advancing through multiple critical breakthroughs. Drawing inspiration from pioneering research (Ho et al., 2020; Song et al., 2021; Dhariwal & Nichol, 2021; Luo et al., 2023; Choi et al., 2022), initial frameworks (Ramesh et al., 2022; Saharia et al., 2022) showcased the extraordinary capabilities of these techniques in text-driven image generation. To mitigate the heavy computational costs associated with processing in the pixel domain, later progress (Rombach et al., 2022; Podell et al., 2024) proposed latent diffusion models based on VAEs (Kingma & Welling, 2014), which substantially lowered inference overhead without sacrificing visual quality. In recent times, incorporating flow matching (Lipman et al., 2023; Liu et al., 2023) has driven structural advancements, providing novel ways to formulate probability flow dynamics (Esser et al., 2024; Sauer et al., 2024; Batifol et al., 2025). Notwithstanding these strides, existing methods still show shortcomings regarding visual fidelity and alignment with human intent, which drives the exploration of fine-tuning techniques for further improvements.

2.2 Direct Preference Optimization for Diffusion Models

Supervised fine-tuning on carefully filtered datasets continues to be the mainstream strategy for adapting diffusion models. Taking cues from similar progress in the realm of large language models (Rafailov et al., 2023; Hong et al., 2024; Ethayarajh et al., 2024), various preference optimization techniques have been developed, where Diffusion-DPO (Wallace et al., 2024) acts as the primary baseline. Several specific hurdles have driven further improvements in this area. First, the partition function within the DPO formulation introduces computational difficulties, calling for improved approximation techniques to ensure training stability (Zhu et al., 2025). Second, since overall preference annotations fail to accurately reflect the fluctuating quality across intermediate denoising stages, researchers have proposed latent-space reward models to facilitate precise, step-by-step credit distribution (Liang et al., 2025; Zhang et al., 2025; Tee et al., 2025). Third, given the cost of human data, diverse alternatives have emerged to reduce labeling—from positive-only approaches to fully self-supervised frameworks bypassing pairwise comparisons (Peng et al., 2025; Yuan et al., 2024).

3 Preliminaries

3.1 Diffusion Models

Diffusion models (Ho et al., 2020) consist of a forward process $q(\boldsymbol{x}_{0:T})$ that gradually noises data $\boldsymbol{x}_0$ and a learned reverse process $p_\theta(\boldsymbol{x}_{0:T})$ that denoises from pure Gaussian noise. The forward chain is $q(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}) = \mathcal{N}(\boldsymbol{x}_t; \sqrt{\alpha_t}\boldsymbol{x}_{t-1}, (1-\alpha_t)\mathbf{I})$ with schedule α_t, and the reverse is $p_\theta(\boldsymbol{x}_{t-1} \mid \boldsymbol{x}_t) = \mathcal{N}(\boldsymbol{x}_{t-1}; \boldsymbol{\mu}_\theta^t(\boldsymbol{x}_t), \boldsymbol{\Sigma}_t(\boldsymbol{x}_t))$. Training maximizes the variational lower bound, which simplifies to

$$\mathcal{L}_{\mathrm{DM}} = \mathbb{E}_{\boldsymbol{x}_0, \boldsymbol{x}_t, t, \boldsymbol{\epsilon}} \left[\lambda(t) \left\| \boldsymbol{\epsilon}_\theta^t(\boldsymbol{x}_t) - \boldsymbol{\epsilon} \right\|_2^2 \right], \tag{1}$$

where $\boldsymbol{x}_t \sim q(\boldsymbol{x}_t \mid \boldsymbol{x}_0) = \mathcal{N}(\boldsymbol{x}_t; \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0, (1-\bar{\alpha}_t)\mathbf{I})$, $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$, and $\lambda(t)$ is a weighting factor (Ho et al., 2020).

DDIM (Song et al., 2021) offers a deterministic ODE-based variant, with sampling step

$$\boldsymbol{x}_t = \sqrt{\frac{\bar{\alpha}_t}{\bar{\alpha}_{t+1}}}\boldsymbol{x}_{t+1} + \sqrt{\bar{\alpha}_t}\left(\sqrt{\frac{1-\bar{\alpha}_t}{\bar{\alpha}_t}} - \sqrt{\frac{1-\bar{\alpha}_{t+1}}{\bar{\alpha}_{t+1}}} \right) \boldsymbol{\epsilon}_\theta^{t+1}(\boldsymbol{x}_{t+1}). \tag{2}$$

DDIM inversion (Song et al., 2021) approximately reverses sampling, constructing a deterministic trajectory from $\boldsymbol{x}_0$ toward noise using the current-state noise prediction:

$$\boldsymbol{x}_{t+1} = \sqrt{\frac{\bar{\alpha}_{t+1}}{\bar{\alpha}_t}}\boldsymbol{x}_t + \sqrt{\bar{\alpha}_{t+1}}\left(\sqrt{\frac{1-\bar{\alpha}_{t+1}}{\bar{\alpha}_{t+1}}} - \sqrt{\frac{1-\bar{\alpha}_t}{\bar{\alpha}_t}} \right) \boldsymbol{\epsilon}_\theta^t(\boldsymbol{x}_t). \tag{3}$$

3.2 Direct Preference Optimization

DPO leverages human preference data $\mathcal{D} = \{(\boldsymbol{x}_0^w, \boldsymbol{x}_0^l, \boldsymbol{c})\}$, where $\boldsymbol{c}$ is a text prompt and $\boldsymbol{x}_0^w, \boldsymbol{x}_0^l$ are preferred/dispreferred images from a reference model. Under the Bradley–Terry model, $p_{\mathrm{BT}}(\boldsymbol{x}_0^w \succ \boldsymbol{x}_0^l \mid \boldsymbol{c}) = \sigma(r(\boldsymbol{x}_0^w, \boldsymbol{c}) - r(\boldsymbol{x}_0^l, \boldsymbol{c}))$. DPO reparameterizes the implicit reward with the policy p_θ and a frozen reference p_{ref}, giving

$$\mathcal{L}_{\mathrm{DPO}}(\theta) = -\mathbb{E}_{(\boldsymbol{x}_0^w, \boldsymbol{x}_0^l, \boldsymbol{c}) \sim \mathcal{D}} \left[\log \sigma \left(\beta \left(\log \frac{p_\theta(\boldsymbol{x}_0^w \mid \boldsymbol{c})}{p_{\mathrm{ref}}(\boldsymbol{x}_0^w \mid \boldsymbol{c})} - \log \frac{p_\theta(\boldsymbol{x}_0^l \mid \boldsymbol{c})}{p_{\mathrm{ref}}(\boldsymbol{x}_0^l \mid \boldsymbol{c})} \right) \right) \right]. \tag{4}$$

For diffusion models, the exact $p_\theta(\boldsymbol{x}_0 \mid \boldsymbol{c})$ is intractable, so Wallace et al. (2024) optimize a tractable upper bound using latent paths $\boldsymbol{x}_{1:T}$:

$$\begin{aligned} \mathcal{L}_{\text{DPO-Diff}}(\theta) = -\mathbb{E}_{(\boldsymbol{x}_0^w, \boldsymbol{x}_0^l, \boldsymbol{c}) \sim \mathcal{D}} \log \sigma \Bigg(\beta \mathbb{E}_{\substack{\boldsymbol{x}_{1:T}^w \sim p_\theta(\boldsymbol{x}_{1:T}^w \mid \boldsymbol{x}_0^w) \\ \boldsymbol{x}_{1:T}^l \sim p_\theta(\boldsymbol{x}_{1:T}^l \mid \boldsymbol{x}_0^l)}} \Bigg[\\ \log \frac{p_\theta(\boldsymbol{x}_{0:T}^w)}{p_{\mathrm{ref}}(\boldsymbol{x}_{0:T}^w)} - \log \frac{p_\theta(\boldsymbol{x}_{0:T}^l)}{p_{\mathrm{ref}}(\boldsymbol{x}_{0:T}^l)} \Bigg] \Bigg). \end{aligned} \tag{5}$$

where $p_\theta(\cdot)$ abbreviates $p_\theta(\cdot \mid \boldsymbol{c})$.

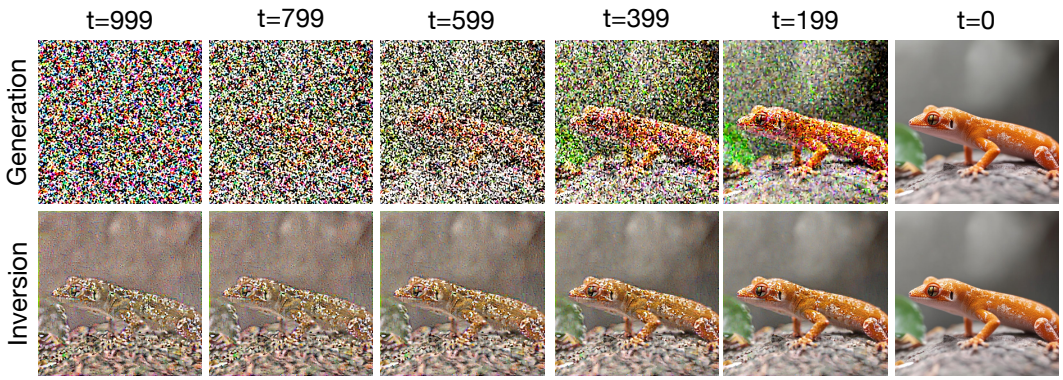

Figure 2: **Inversion drift and persistent image structure.** VAE-decoded latents from ordinary generation (top) and DDIM inversion (bottom) are shown at matched timesteps. At high noise levels, inversion retains recognizable object contours and textures, whereas the corresponding generation states are largely dominated by noise. This difference illustrates the statistical mismatch between inversion states and states visited during Gaussian-initialized generation.

4 METHODOLOGY

Inversion offers image-specific preference supervision, yet its latents can drift substantially from states encountered during ordinary generation. In Section 4.1, we diagnose this mismatch through stage-wise experiments, showing that inversion-based supervision remains valuable at low noise levels despite its limitations at high noise levels. In Section 4.2, we build on these findings and analyze persistent image structure along inversion trajectories to develop a dynamic fallback mechanism that mitigates inversion drift while retaining useful supervision.

4.1 DIAGNOSING THE IMPACT OF INVERSION DRIFT

Figure 2 illustrates the Gaussian reachability gap: at matched high-noise timesteps, inverted latents retain recognizable image structure that is largely absent from generation states. Inversion-based preference optimization therefore supervises states that differ from those visited during Gaussian-initialized denoising. We examine its effects across denoising stages.

We conduct an inference-time checkpoint-switching experiment using the existing DPO- and InPO-trained checkpoints, without additional training. Within each 50-step denoising trajectory, we use the DPO checkpoint for the initial high-noise steps and switch to the InPO checkpoint for the remaining low-noise steps. We vary the number of DPO steps from 0 to 50 and compare the resulting configurations with both pure-model baselines and the reverse assignment. All configurations are evaluated on the same 500 prompts with shared initial noise, using PickScore, HPS, ImageReward, and MPS.

As shown in Figure 3, assigning the first 15 steps to DPO and the remaining 35 steps to InPO achieves the highest mean score for every evaluator, exceeding both pure DPO and pure InPO. Reversing the assignment produces lower scores at every tested split using both models. Together, Figures 2 and 3 motivate reducing reliance on inversion-based supervision at high noise levels, where inversion drift is most pronounced, while preserving its benefits during low-noise denoising.

An exhaustive per-prompt oracle selects switching points independently for each evaluation metric, establishing metric-specific upper bounds within the tested candidate family (Section 5.3). Its gains over the best global split reveal the potential of sample-dependent allocation. This motivates a practical criterion for adapting the boundary between inversion and forward-noising supervision, which we develop next using variance as a proxy.

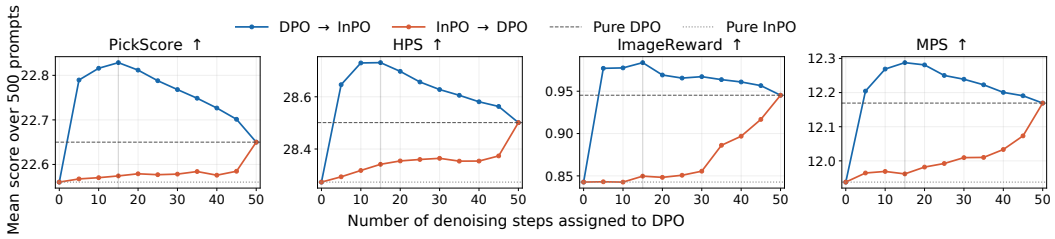

Figure 3: **Stage-wise complementarity of DPO and InPO.** Applying DPO at high noise levels followed by InPO at low noise levels improves all four mean preference scores over either model alone, with the best evaluated split at 15 DPO steps and 35 InPO steps. Reversing the order yields lower scores at every tested split using both models. The asymmetry indicates that inversion-based preference learning remains useful during later denoising, despite its limitations at high noise levels. PickScore and HPS are scaled by 100, as in the tables.

4.2 Drift-Aware Dynamic Fallback

From inversion mismatch to structural retention. Figure 2 reveals more than a trajectory mismatch. The inversion trajectory preserves object contours and textures at timesteps where ordinary generation is largely unstructured. We analyze how this phenomenon can arise from the discrete inversion update. Let $t_0 < \cdots < t_K$ denote the inversion timesteps, and write $\mu_t = \sqrt{\bar{\alpha}_t}$ and $\sigma_t = \sqrt{1-\bar{\alpha}_t}$. A DDIM inversion step is

$$\boldsymbol{x}^{\text{inv}}_{t_i} = a_i \boldsymbol{x}^{\text{inv}}_{t_{i-1}} + b_i \boldsymbol{\epsilon}_\theta(\boldsymbol{x}^{\text{inv}}_{t_{i-1}}, t_{i-1}, \boldsymbol{c}), \quad a_i = \frac{\mu_{t_i}}{\mu_{t_{i-1}}}, \quad b_i = \sigma_{t_i} - a_i \sigma_{t_{i-1}}. \tag{6}$$

For a generation trajectory $\boldsymbol{x}^{\text{gen}}_{t_i}$ that reaches the same clean endpoint, the corresponding rearranged DDIM relation uses the prediction at t_i. Defining $\boldsymbol{e}_i = \boldsymbol{x}^{\text{inv}}_{t_i} - \boldsymbol{x}^{\text{gen}}_{t_i}$ gives

$$\boldsymbol{e}_i = a_i \boldsymbol{e}_{i-1} + b_i \boldsymbol{\delta}_i, \qquad \boldsymbol{\delta}_i = \boldsymbol{\epsilon}_\theta(\boldsymbol{x}^{\text{inv}}_{t_{i-1}}, t_{i-1}, \boldsymbol{c}) - \boldsymbol{\epsilon}_\theta(\boldsymbol{x}^{\text{gen}}_{t_i}, t_i, \boldsymbol{c}). \tag{7}$$

The recurrence in Equation (7) shows that inversion drift accumulates through mismatches between the predictions used by inversion and generation. Diffusion generation typically proceeds from coarse structures to fine details (Qian et al., 2024). Along this process, the transition from a noisier state to the next, less noisy state progressively refines image structure, so the difference between their noise predictions can contain spatially structured components associated with contours and textures. During inversion, these components enter the local prediction mismatch and, when aligned across steps, accumulate rather than cancel. Such accumulation can reinforce residual image structure, leaving recognizable contours and textures at noise levels where ordinary generation is largely unstructured, consistent with Figure 2.

Variance deviation as a drift proxy. To decide when inversion should be retained, we require a lightweight measure of its deviation from ordinary noising. We use spatial variance as a proxy. For a latent $\boldsymbol{x} \in \mathbb{R}^d$, define

$$\mathcal{V}(\boldsymbol{x}) = \frac{1}{d}\sum_{j=1}^{d}(x_j - \bar{x})^2, \qquad \bar{x} = \frac{1}{d}\sum_{j=1}^{d} x_j. \tag{8}$$

For ordinary forward noising, $\boldsymbol{x}^{\text{fwd}}_t = \mu_t \boldsymbol{x}_0 + \sigma_t \boldsymbol{\epsilon}$ with $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, the variance baseline is

$$\mathcal{V}^{\text{fwd}}_t(\boldsymbol{x}_0) = \mu_t^2 \mathcal{V}(\boldsymbol{x}_0) + \sigma_t^2. \tag{9}$$

We measure the variance deviation of an inverted latent by

$$\Delta_t(\boldsymbol{x}^{\text{inv}}_t) = \left| \mathcal{V}(\boldsymbol{x}^{\text{inv}}_t) - \mathcal{V}^{\text{fwd}}_t(\boldsymbol{x}_0) \right|. \tag{10}$$

Spatial variance does not fully characterize the latent distribution or the geometry of inversion drift. It is used here as a computationally inexpensive proxy: the persistent, spatially non-uniform image structure in Figure 2 can induce a substantial change in this statistic.

To calibrate a timestep-dependent reference threshold for the variance deviation Δ_t, we collect $N = 2000$ generation trajectories from the pretrained base model, starting from Gaussian noise and using prompts from the Pick-a-Pic v2 training set. For each trajectory n, we record the intermediate latents $\boldsymbol{x}_t^{\mathrm{gen},n}$ and the corresponding clean endpoint $\boldsymbol{x}_0^{\mathrm{gen},n}$. At each timestep t, we compute their variance deviations and set the reference threshold τ_t to the empirical 95th percentile:

$$\Delta_t^{(n)} = \left|\mathcal{V}(\boldsymbol{x}_t^{\mathrm{gen},n}) - \mathcal{V}_t^{\mathrm{fwd}}(\boldsymbol{x}_0^{\mathrm{gen},n})\right|, \qquad \tau_t = Q_{0.95}\Big(\{\Delta_t^{(n)}\}_{n=1}^{N}\Big), \tag{11}$$

where $Q_{0.95}$ denotes the empirical 95th-percentile operator. This gives a separate reference threshold for each timestep, with the upper 5% of calibration deviations outside the reference range. The thresholds are precomputed once and held fixed throughout preference optimization. Empirically, variance-based selection remains largely stable as training progresses (Section 5.4). Additional calibration details are provided in Appendix B.2.

Tolerance-controlled fallback. Since inversion-based supervision can remain useful despite its deviation from ordinary generation states, we introduce a tolerance parameter $\lambda > 0$ to control the fallback threshold. Specifically, we retain the inverted latent when $\Delta_t \le \lambda\tau_t$ and otherwise use standard forward noising. For the preferred image, IFPO selects the latent

$$\tilde{\boldsymbol{x}}_t = \begin{cases} \boldsymbol{x}_t^{\mathrm{inv}}, & \Delta_t(\boldsymbol{x}_t^{\mathrm{inv}}) \le \lambda\tau_t, \\ \boldsymbol{x}_t^{\mathrm{fwd}}, & \Delta_t(\boldsymbol{x}_t^{\mathrm{inv}}) > \lambda\tau_t. \end{cases} \tag{12}$$

Here, $\boldsymbol{x}_t^{\mathrm{inv}}$ and $\boldsymbol{x}_t^{\mathrm{fwd}}$ are obtained from the same clean latent $\boldsymbol{x}_0$ through inversion and standard forward noising, respectively. A larger λ permits greater variance deviation before fallback.

Preference optimization objective. For a preference triple $(\boldsymbol{c}, \boldsymbol{x}_0^w, \boldsymbol{x}_0^l) \sim \mathcal{D}$, we evaluate the fallback condition in Eq. (12) on the winner trajectory and apply the resulting branch to both trajectories, producing $\tilde{\boldsymbol{x}}_t^w$ and $\tilde{\boldsymbol{x}}_t^l$. IFPO then uses the standard Diffusion-DPO objective on the selected latents:

$$\begin{aligned} \mathcal{L}_{\mathrm{IFPO}} &= -\mathbb{E}_{(\boldsymbol{c},\boldsymbol{x}_0^w,\boldsymbol{x}_0^l)\sim\mathcal{D},\, t} \log\sigma\left(-\beta\omega(t)[\Delta^w - \Delta^l]\right), \\ \Delta^s &= \|\boldsymbol{\epsilon}^s - \boldsymbol{\epsilon}_\theta(\tilde{\boldsymbol{x}}_t^s, t, \boldsymbol{c})\|_2^2 - \|\boldsymbol{\epsilon}^s - \boldsymbol{\epsilon}_{\mathrm{ref}}(\tilde{\boldsymbol{x}}_t^s, t, \boldsymbol{c})\|_2^2, \quad s \in \{w, l\}. \end{aligned} \tag{13}$$

Here, $\boldsymbol{\epsilon}^s$ denotes the target noise associated with the selected trajectory of sample s, and $\omega(t)$ is the standard timestep-dependent weighting. IFPO changes only how the supervisory latent is selected; the preference objective itself remains unchanged. The corresponding flow-matching formulation is given in Section A.3.

5 Experiments

5.1 Experimental Setup

Backbones and training data. We evaluate IFPO on Stable Diffusion XL-base-1.0 (SDXL) (Podell et al., 2024) and Stable Diffusion 3.5 Medium (SD3.5) (Esser et al., 2024; Stability AI, 2024), covering diffusion and flow-matching backbones. For SDXL, we use Pick-a-Pic v2 (Kirstain et al., 2023), obtaining 851,293 preference pairs across 58,960 prompts. For SD3.5, whose output quality is less well represented by Pick-a-Pic v2, we use HPD v3 (Ma et al., 2025) and retain pairs involving at least one SD3-generated image, yielding 144,991 training pairs.

Optimization. As in InPO (Lu et al., 2025), we use 10 inversion steps for both backbones. We train both backbones using Adafactor (Shazeer & Stern, 2018) on eight NVIDIA A100 GPUs. Following Diffusion-DPO (Wallace et al., 2024), we set $\beta = 2000$ and the learning rate to $\frac{2000}{\beta} \times 2.048 \times 10^{-8}$. The default tolerance is $\lambda = 1$ for SDXL and $\lambda = 2.5$ for SD3.5.

Baselines and evaluation. For both SDXL and SD3.5, we compare IFPO with the pretrained backbone, Diffusion-DPO (Wallace et al., 2024), and InPO (Lu et al., 2025); the SDXL comparison additionally includes MaPO (Hong et al., 2025) and IAPO (Sun et al., 2026). We re-implement InPO on both backbones and all fine-tuning baselines on SD3.5, using the same SD3.5 training data and budget as IFPO. Our main benchmarks are the Pick-a-Pic v2 test set (500 prompts) and HPDv2 (3,200 prompts) (Wu et al., 2023). We report mean and median PickScore (Kirstain et al., 2023), HPS v2 (Wu et al., 2023), ImageReward (Xu et al., 2023), and MPS (Zhang et al., 2024).



Figure 4: **Qualitative comparison on SDXL (left) and SD3.5 Medium (right).** Examples examine prompt adherence, text rendering, structural coherence, and visual quality.

5.2 Main Results

Preference alignment across backbones. Table 1 compares IFPO with existing methods on SDXL and SD3.5. On SDXL, IFPO achieves the highest PickScore, HPS, and ImageReward scores on both benchmarks. Mean ImageReward increases from 0.9453 to 1.033 on Pick-a-Pic and from 1.059 to 1.136 on HPDv2. Mean PickScore improves by 0.25 and 0.14 over the best baseline. On SD3.5, IFPO achieves the highest mean and median scores across all four metrics on both benchmarks, improving mean PickScore over the strongest baseline by 0.17 on Pick-a-Pic and 0.15 on HPDv2. Figure 4 provides qualitative comparisons.

User study on SDXL. We conduct a blinded study with 50 participants using 400 image sets generated from randomly sampled Parti-Prompts (Yu et al., 2022). Each participant evaluates 100 sets, selecting the best image among SDXL, DPO, InPO, and IFPO with model identities hidden, yielding 5,000 judgments. As shown in Figure 5, IFPO receives the highest overall preference rate of 60.9%, compared with 20.3% for DPO, 15.6% for InPO, and 3.1% for the pretrained SDXL backbone.

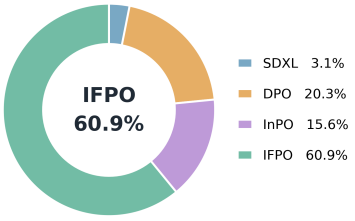

Figure 5: **Human preference on SDXL.** Best-image selection rates.

5.3 Ablation Studies

We first examine the potential of dynamic fallback using an exhaustive per-prompt oracle that evaluates every candidate switching point for each metric. We then compare variance-based selection with reward-selected static splits to assess its effectiveness, and with this oracle to quantify the potential gains from ideal switch-point selection. Finally, we study tolerance sensitivity on SD3.5.

Metric-specific oracle upper bounds. A global split cannot accommodate differences among prompts. We therefore exhaustively select the best split for each prompt and evaluation metric. Let $G_K(c)$ denote the image generated with K DPO steps followed by $50 - K$ InPO steps, holding the models and initial noise fixed. For prompt c_i and evaluator R_m, the oracle chooses

$$K^{\star}_{i,m} = \arg\max_{K \in \mathcal{K}} R_m\big(G_K(c_i), c_i\big), \qquad \mathcal{K} = \{0, 5, \ldots, 50\}. \tag{14}$$

For each metric, we report the mean and median of its per-prompt maxima. The columns therefore represent separate metric-specific upper bounds within the tested candidate family and need not

Table 1: **Preference evaluation on SDXL and SD3.5 Medium.** We report mean and median scores. Best and second-best displayed results are bolded and underlined, including ties.

Benchmark	Method	PickScore (Kirstain et al., 2023)		HPS (Wu et al., 2023)		ImageReward (Xu et al., 2023)		MPS (Zhang et al., 2024)	
		Mean	Median	Mean	Median	Mean	Median	Mean	Median
Pick-a-Pic (Kirstain et al., 2023)	SDXL (Podell et al., 2024)	22.17	22.24	27.97	27.98	0.5679	0.7563	11.54	11.84
	DPO (Wallace et al., 2024)	22.65	22.69	28.50	28.51	0.9453	1.183	**12.17**	**12.39**
	MaPO (Hong et al., 2025)	22.32	22.34	28.54	28.63	0.6998	0.8463	11.66	11.82
	IAPO (Sun et al., 2026)	22.68	22.70	28.60	28.67	0.9108	1.099	12.11	12.22
	InPO (Lu et al., 2025)	22.56	22.67	28.27	28.31	0.8427	1.085	11.94	12.18
	IFPO	**22.93**	**22.92**	**28.83**	**28.92**	**1.033**	**1.283**	**12.17**	12.16
HPDv2 (Wu et al., 2023)	SDXL (Podell et al., 2024)	22.78	22.77	28.63	28.62	0.7801	0.8828	14.27	14.43
	DPO (Wallace et al., 2024)	23.13	23.09	29.00	28.98	1.059	1.246	14.55	14.75
	MaPO (Hong et al., 2025)	22.85	22.82	28.99	29.01	0.8749	0.9947	14.36	14.45
	IAPO (Sun et al., 2026)	23.32	23.28	29.28	29.30	0.9973	1.140	14.74	14.88
	InPO (Lu et al., 2025)	23.28	23.23	29.24	29.26	1.047	1.195	14.66	14.82
	IFPO	**23.46**	**23.44**	**29.43**	**29.42**	**1.136**	**1.302**	**14.88**	**15.12**
Pick-a-Pic (Kirstain et al., 2023)	SD3.5 (Stability AI, 2024)	22.30	22.36	28.57	28.63	1.067	1.311	12.08	12.37
	DPO (Wallace et al., 2024)	22.25	22.22	28.52	28.55	1.045	1.277	12.06	12.31
	InPO (Lu et al., 2025)	22.29	22.33	28.77	28.80	1.115	1.324	12.06	12.18
	IFPO	**22.47**	**22.44**	**28.89**	**28.93**	**1.148**	**1.406**	**12.16**	**12.40**
HPDv2 (Wu et al., 2023)	SD3.5 (Stability AI, 2024)	22.91	22.88	29.36	29.39	1.189	1.366	14.72	14.88
	DPO (Wallace et al., 2024)	22.73	22.70	29.23	29.28	1.173	1.359	14.48	14.67
	InPO (Lu et al., 2025)	22.91	22.85	29.57	29.61	1.229	1.411	14.81	14.97
	IFPO	**23.06**	**23.03**	**29.64**	**29.65**	**1.246**	**1.424**	**14.84**	**15.04**

correspond to the same selected images. Their gains over the reward-selected global split in Table 2 demonstrate the potential of sample-dependent allocation.

Effectiveness of variance-based selection. At inference, Static mix uses the reward-selected 15/35 split for every prompt. Var. mix is designed so that, for most prompts, inverting its generated image selects the same switch point used to generate it (Appendix B.5). Random mix shuffles Var. mix's switch points across prompts. With the same fixed DPO and InPO models, Var. mix improves mean PickScore, HPS, and ImageReward over both controls, supporting variance-based prompt matching. At training time, Static train uses forward-noising supervision at the 15 high-noise positions (30.0%) and inversion supervision at the remaining positions, compared with IFPO's 30.9% fallback rate. IFPO improves the same three mean metrics at a similar rate, supporting sample-dependent fallback placement. The metric-specific oracle bounds leave room for more precise selection.

Tolerance sensitivity on SD3.5. We evaluate $\lambda \in \{2.4, 2.5, 2.6\}$ around the default value of 2.5, with fixed training data, optimization budget, and evaluation protocol. Smaller values trigger fallback more readily, whereas larger values retain more inversion supervision. Table 3 compares these settings with the pretrained SD3.5 Medium baseline, reporting mean preference scores on Pick-a-Pic v2. All three settings outperform the pretrained backbone on every metric; the default $\lambda = 2.5$ achieves the highest ImageReward and MPS, and ties with $\lambda = 2.6$ on PickScore at the reported precision; $\lambda = 2.6$ yields the highest HPS.

5.4 Stability of Variance-Based Selection

Selection stability across checkpoints. Using prompts from the Pick-a-Pic v2 test set, we compare variance-selected switch points from the pretrained SDXL backbone and training checkpoints at steps 300 and 400. Writing $K_c^{(s)}$ for the selected number of DPO steps, we report exact agreement with the pretrained selector, $K_c^{(s)} = K_c^{(0)}$, and agreement within one grid interval (five denoising steps), $|K_c^{(s)} - K_c^{(0)}| \leq 5$, including unchanged selections. The corresponding rates are 76.8% and 96.6% at step 300, and 58.0% and 94.2% at step 400, respectively. Overall, the selected switch points remain largely stable across checkpoints.

Table 2: **Variance-based selection and reference strategies.** SDXL results on the 500 Pick-a-Pic test prompts. Static mix and Static train use the reward-selected 15/35 split. Var. mix selects the inference split using variance. Random mix shuffles Var. mix's switch points across prompts. Oracle (per metric) selects each prompt's split independently.

Variant	PickScore		HPS		ImageReward		MPS	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median
DPO	22.65	22.69	28.50	28.51	0.9453	1.183	12.17	12.39
InPO	22.56	22.67	28.27	28.31	0.8427	1.085	11.94	12.18
Inference-time composition of fixed models								
Static mix	22.83	22.84	28.73	28.75	0.9837	1.236	12.29	12.55
Random mix	22.83	22.86	28.73	28.71	0.9826	1.217	12.29	12.59
Var. mix	22.87	22.93	28.93	29.05	1.048	1.279	12.18	12.41
Oracle (per metric)	**23.10**	**23.12**	**29.18**	**29.18**	**1.213**	**1.435**	**12.92**	**13.14**
Single-model preference training								
Static train	22.78	22.72	28.80	28.81	0.9616	1.230	12.07	**12.31**
IFPO	**22.93**	**22.92**	**28.83**	**28.92**	**1.033**	**1.283**	**12.17**	12.16

Table 3: **Tolerance sensitivity on SD3.5.** Mean preference scores on Pick-a-Pic, with the pretrained backbone as a reference.

Setting	PickScore	HPS	ImageReward	MPS
SD3.5 base	22.30	28.57	1.067	12.08
$\lambda = 2.4$	22.42	28.71	1.143	12.15
$\lambda = 2.5$	22.47	28.89	1.148	12.16
$\lambda = 2.6$	22.47	28.90	1.129	12.13

Table 4: **Reward changes under checkpoint-dependent selection.** Mean scores on Pick-a-Pic v2 test prompts with fixed generation models and initial noise.

Selector	PickScore	HPS	ImageReward	MPS
Pretrained	22.87	28.93	1.048	12.18
Step 300	22.86	28.78	0.9865	12.47
Step 400	22.86	28.79	0.9865	12.46

Reward changes in the stage-wise diagnostic. We reuse the diagnostic in Section 4.1, fixing the DPO and InPO generation models, prompts, and initial noise while changing only the checkpoint-selected switch points. As shown in Table 4, mean PickScore stays between 22.86 and 22.87, and HPS between 28.78 and 28.93 across the three selectors.

6 Conclusion

We identified the **Gaussian reachability gap**: inversion can produce supervisory latents that deviate from states encountered during ordinary Gaussian-initialized generation, particularly at high noise levels. Our stage-wise analysis shows that inversion-based supervision remains useful during low-noise denoising. Building on this finding, **IFPO** uses a variance-deviation proxy and a tolerance-controlled dynamic fallback to retain useful inversion supervision while reverting to standard forward noising when deviations exceed the tolerated level. Experiments on SDXL and SD3.5 demonstrate improved preference alignment across diffusion and flow-matching backbones.

Future work could develop efficient switch-point selectors that approach the metric-specific oracle without exhaustive reward evaluation. Beyond the DPO/InPO pair, we plan to examine whether Diffusion-DPO variants and other diffusion post-training approaches exhibit stage-wise complementarity across prompts and backbones, and to build a general, efficient framework that combines the strengths of two or more models within a single denoising trajectory at inference time.

AI use statement

We used generative AI tools to assist with language editing, reference checking, and LaTeX formatting. The authors reviewed the AI-assisted changes and take responsibility for the final content.

REFERENCES

Stephen Batifol, Andreas Blattmann, Frederic Boesel, Saksham Consul, Cyril Diagne, Tim Dockhorn, Jack English, Zion English, Patrick Esser, Sumith Kulal, Kyle Lacey, Yam Levi, Cheng Li, Dominik Lorenz, Jonas Müller, Dustin Podell, Robin Rombach, Harry Saini, Axel Sauer, and Luke Smith. FLUX.1 Kontext: Flow Matching for In-Context Image Generation and Editing in Latent Space. *arXiv:2506.15742*, 2025.

Jooyoung Choi, Jungbeom Lee, Chaehun Shin, Sungwon Kim, Hyunwoo Kim, and Sungroh Yoon. Perception Prioritized Training of Diffusion Models. In *CVPR*, 2022.

Guillaume Couairon, Jakob Verbeek, Holger Schwenk, and Matthieu Cord. Diffedit: Diffusion-based semantic image editing with mask guidance. In *ICLR*, 2023.

Longquan Dai, Xiaolu Wei, He Wang, Shaomeng Wang, and Jinhui Tang. Aligning Text-to-Image Diffusion Models to Human Preference by Classification. In *NeurIPS*, 2025.

Prafulla Dhariwal and Alex Nichol. Diffusion Models Beat GANs on Image Synthesis. In *NeurIPS*, 2021.

Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, Dustin Podell, Tim Dockhorn, Zion English, Kyle Lacey, Alex Goodwin, Yannik Marek, and Robin Rombach. Scaling Rectified Flow Transformers for High-Resolution Image Synthesis. In *ICML*, 2024.

Kawin Ethayarajh, Winnie Xu, Niklas Muennighoff, Dan Jurafsky, and Douwe Kiela. KTO: Model Alignment as Prospect Theoretic Optimization. In *ICML*, 2024.

Daniel Garibi, Or Patashnik, Andrey Voynov, Hadar Averbuch-Elor, and Daniel Cohen-Or. ReNoise: Real image inversion through iterative noising. In *ECCV*, 2024.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models. In *NeurIPS*, 2020.

Jiwoo Hong, Noah Lee, and James Thorne. ORPO: Monolithic Preference Optimization without Reference Model. In *EMNLP*, 2024.

Jiwoo Hong, Sayak Paul, Noah Lee, Kashif Rasul, James Thorne, and Jongheon Jeong. Margin-aware Preference Optimization for Aligning Diffusion Models without Reference. In *ICLR*, 2025.

Diederik P. Kingma and Max Welling. Auto-Encoding Variational Bayes. In *ICLR*, 2014.

Yuval Kirstain, Adam Polyak, Uriel Singer, Shahbuland Matiana, Joe Penna, and Omer Levy. Pick-a-Pic: An Open Dataset of User Preferences for Text-to-Image Generation. In *NeurIPS*, 2023.

Zhanhao Liang, Yuhui Yuan, Shuyang Gu, Bohan Chen, Tiankai Hang, Mingxi Cheng, Ji Li, and Liang Zheng. Aesthetic Post-Training Diffusion Models from Generic Preferences with Step-by-step Preference Optimization. In *CVPR*, 2025.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow Matching for Generative Modeling. In *ICLR*, 2023.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow. In *ICLR*, 2023.

Yunhong Lu, Qichao Wang, Hengyuan Cao, Xierui Wang, Xiaoyin Xu, and Min Zhang. InPO: Inversion Preference Optimization with Reparametrized DDIM for Efficient Diffusion Model Alignment. In *CVPR*, 2025.

Simian Luo, Yiqin Tan, Longbo Huang, Jian Li, and Hang Zhao. Latent Consistency Models: Synthesizing High-Resolution Images with Few-Step Inference. *arXiv:2310.04378*, 2023.

Yuhang Ma, Xiaoshi Wu, Keqiang Sun, and Hongsheng Li. HPSv3: Towards wide-spectrum human preference score. In *ICCV*, 2025. URL https://openaccess.thecvf.com/content/ICCV2025/papers/Ma_HPSv3_Towards_Wide-Spectrum_Human_Preference_Score_ICCV_2025_paper.pdf.

Ron Mokady, Amir Hertz, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. Null-text inversion for editing real images using guided diffusion models. In *CVPR*, 2023.

Byeonghu Na, Minsang Park, Gyuwon Sim, Donghyeok Shin, HeeSun Bae, Mina Kang, Se Jung Kwon, Wanmo Kang, and Il-Chul Moon. Diffusion Adaptive Text Embedding for Text-to-Image Diffusion Models. In *NeurIPS*, 2025.

Liang Peng, Boxi Wu, Haoran Cheng, Yibo Zhao, and Xiaofei He. SUDO: Enhancing Text-to-Image Diffusion Models with Self-Supervised Direct Preference Optimization. In *NeurIPS*, 2025.

Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis. In *ICLR*, 2024.

Yurui Qian, Qi Cai, Yingwei Pan, Yehao Li, Ting Yao, Qibin Sun, and Tao Mei. Boosting diffusion models with moving average sampling in frequency domain. In *CVPR*, 2024.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. Direct Preference Optimization: Your Language Model is Secretly a Reward Model. In *NeurIPS*, 2023.

Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical Text-Conditional Image Generation with CLIP Latents. *arXiv:2204.06125*, 2022.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-Resolution Image Synthesis with Latent Diffusion Models. In *CVPR*, 2022.

Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily Denton, Seyed Kamyar Seyed Ghasemipour, Burcu Karagol Ayan, S. Sara Mahdavi, Rapha Gontijo Lopes, Tim Salimans, Jonathan Ho, David J. Fleet, and Mohammad Norouzi. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. In *NeurIPS*, 2022.

Axel Sauer, Frederic Boesel, Tim Dockhorn, Andreas Blattmann, Patrick Esser, and Robin Rombach. Fast High-Resolution Image Synthesis with Latent Adversarial Diffusion Distillation. In *SIGGRAPH ASIA*, 2024.

Noam Shazeer and Mitchell Stern. Adafactor: Adaptive Learning Rates with Sublinear Memory Cost. In *ICML*, 2018.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising Diffusion Implicit Models. In *ICLR*, 2021.

Stability AI. Stable diffusion 3.5 medium. Model card, 2024. URL `https://huggingface.co/stabilityai/stable-diffusion-3.5-medium`.

Jiayang Sun, Pin Wang, Hongbo Wang, Xinyue Liu, Huaibo Huang, and Ran He. Towards fine-grained attribution: Instance-aware preference optimization for aligning diffusion models. In *CVPR*, 2026.

Joshua Tian Jin Tee, Hee Suk Yoon, Abu Hanif Muhammad Syarubany, Eunseop Yoon, and Chang D Yoo. A Gradient Guidance Perspective on Stepwise Preference Optimization for Diffusion Models. In *NeurIPS*, 2025.

Bram Wallace, Meihua Dang, Rafael Rafailov, Linqi Zhou, Aaron Lou, Senthil Purushwalkam, Stefano Ermon, Caiming Xiong, Shafiq Joty, and Nikhil Naik. Diffusion Model Alignment Using Direct Preference Optimization. In *CVPR*, 2024.

Fu-Yun Wang, Yunhao Shui, Jingtan Piao, Keqiang Sun, and Hongsheng Li. Diffusion-NPO: Negative Preference Optimization for Better Preference Aligned Generation of Diffusion Models. In *ICLR*, 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. *arXiv:2306.09341*, 2023.

Jiazheng 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 *NeurIPS*, 2023.

Jiahui Yu, Yuanzhong Xu, Jing Yu Koh, Thang Luong, Gunjan Baid, Zirui Wang, Vijay Vasudevan, Alexander Ku, Yinfei Yang, Burcu Karagol Ayan, Ben Hutchinson, Wei Han, Zarana Parekh, Xin Li, Han Zhang, Jason Baldridge, and Yonghui Wu. Scaling Autoregressive Models for Content-Rich Text-to-Image Generation. *TMLR*, 2022.

Huizhuo Yuan, Zixiang Chen, Kaixuan Ji, and Quanquan Gu. Self-Play Fine-Tuning of Diffusion Models for Text-to-Image Generation. In *NeurIPS*, 2024.

Sixian Zhang, Bohan Wang, Junqiang Wu, Yan Li, Tingting Gao, Di Zhang, and Zhongyuan Wang. Learning Multi-dimensional Human Preference for Text-to-Image Generation. In *CVPR*, 2024.

Tao Zhang, Cheng Da, Kun Ding, Huan Yang, Kun Jin, Yan Li, Tingting Gao, Di Zhang, Shiming Xiang, and Chunhong Pan. Diffusion Model as a Noise-Aware Latent Reward Model for Step-Level Preference Optimization. In *NeurIPS*, 2025.

Huaisheng Zhu, Teng Xiao, and Vasant Honavar. DSPO: Direct score preference optimization for diffusion model alignment. In *ICLR*, 2025.

A Inversion Drift and Accumulation of Structural Residuals

We provide the derivation behind Eq. (7), explain how spatially coherent prediction mismatches can persist along inversion, and give the corresponding relation for a flow-matching backbone. Throughout this analysis, the prediction model and conditioning are fixed; we suppress the conditioning argument for readability.

A.1 DDIM: From One-Step Mismatch to Accumulated Drift

Let $t_0 < \cdots < t_K$ be the inversion grid, with $\mu_i = \sqrt{\bar{\alpha}_{t_i}} > 0$ and $\sigma_i = \sqrt{1 - \bar{\alpha}_{t_i}}$. Consider a deterministic DDIM generation trajectory $\{\boldsymbol{x}_{t_i}^{\mathrm{gen}}\}_{i=0}^{K}$, and initialize inversion at its clean endpoint, $\boldsymbol{x}_{t_0}^{\mathrm{inv}} = \boldsymbol{x}_{t_0}^{\mathrm{gen}}$. This construction defines a matched pair of trajectories without assuming that every training image is exactly reachable from a Gaussian starting state.

The generation update from t_i to t_{i-1} is

$$\boldsymbol{x}_{t_{i-1}}^{\mathrm{gen}} = \frac{\mu_{i-1}}{\mu_i}\boldsymbol{x}_{t_i}^{\mathrm{gen}} + \left(\sigma_{i-1} - \frac{\mu_{i-1}}{\mu_i}\sigma_i\right)\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_{t_i}^{\mathrm{gen}}, t_i). \tag{15}$$

With $a_i = \mu_i/\mu_{i-1}$ and $b_i = \sigma_i - a_i\sigma_{i-1}$, rearranging gives

$$\boldsymbol{x}_{t_i}^{\mathrm{gen}} = a_i\boldsymbol{x}_{t_{i-1}}^{\mathrm{gen}} + b_i\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_{t_i}^{\mathrm{gen}}, t_i), \tag{16}$$

$$\boldsymbol{x}_{t_i}^{\mathrm{inv}} = a_i\boldsymbol{x}_{t_{i-1}}^{\mathrm{inv}} + b_i\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_{t_{i-1}}^{\mathrm{inv}}, t_{i-1}). \tag{17}$$

The two updates evaluate the predictor at different states and timesteps. Subtracting them yields the exact discrete identity

$$\boldsymbol{e}_i = a_i\boldsymbol{e}_{i-1} + b_i\boldsymbol{\delta}_i, \tag{18}$$

$$\boldsymbol{\delta}_i = \boldsymbol{\epsilon}_\theta(\boldsymbol{x}_{t_{i-1}}^{\mathrm{inv}}, t_{i-1}) - \boldsymbol{\epsilon}_\theta(\boldsymbol{x}_{t_i}^{\mathrm{gen}}, t_i), \tag{19}$$

where $\boldsymbol{e}_i = \boldsymbol{x}_{t_i}^{\mathrm{inv}} - \boldsymbol{x}_{t_i}^{\mathrm{gen}}$. Iterating this relation gives

$$\boldsymbol{e}_i = \left(\prod_{k=1}^{i} a_k\right)\boldsymbol{e}_0 + \sum_{j=1}^{i}\left(b_j\prod_{k=j+1}^{i} a_k\right)\boldsymbol{\delta}_j. \tag{20}$$

The empty product equals one. For matched endpoints, $\boldsymbol{e}_0 = \mathbf{0}$, and the telescoping product gives

$$\boldsymbol{e}_i = \sum_{j=1}^{i} w_{ij}\boldsymbol{\delta}_j, \qquad w_{ij} = b_j\frac{\mu_i}{\mu_j} = \mu_i\left(\frac{\sigma_j}{\mu_j} - \frac{\sigma_{j-1}}{\mu_{j-1}}\right). \tag{21}$$

Thus, inversion drift aggregates prediction mismatches from the preceding steps. Under the usual increasing noise-to-signal ratio, $w_{ij} \geq 0$, and

$$\|\boldsymbol{e}_i\|_2 \leq \sum_{j=1}^{i} w_{ij}\|\boldsymbol{\delta}_j\|_2. \tag{22}$$

This bound describes accumulation, rather than monotonic growth: vector-valued mismatches can reinforce or cancel one another. The same recurrence applies to SDXL's DDIM latent trajectory, so no separate architecture-specific derivation is required.

A.2 Spatial Coherence and Persistent Structure

To make the role of spatial coherence explicit, let $\boldsymbol{H}$ be a fixed linear spatial-difference operator on the latent grid, applied channel-wise. Equation (21) implies

$$\boldsymbol{H}\boldsymbol{e}_i = \sum_{j=1}^{i} w_{ij}\boldsymbol{H}\boldsymbol{\delta}_j. \tag{23}$$

For any unit direction $\boldsymbol{u}$ in this spatial-difference representation, if the mismatches satisfy $\langle \boldsymbol{u}, \boldsymbol{H}\boldsymbol{\delta}_j \rangle \geq \gamma_j \geq 0$, then

$$\langle \boldsymbol{u}, \boldsymbol{H}\boldsymbol{e}_i \rangle \geq \sum_{j=1}^{i} w_{ij}\gamma_j. \tag{24}$$

This conditional result explains how aligned spatial residuals can survive aggregation across inversion steps. It does not assume that prediction mismatches are independent noise. The retained contours and textures in Fig. 2 provide the corresponding qualitative observation; the alignment condition itself is not guaranteed by DDIM, and latent spatial differences are not identical to edges in VAE-decoded images.

We can also relate the accumulated residual to the variance statistic used by IFPO. Define the centering operator $\boldsymbol{C} = \boldsymbol{I} - \mathbf{1}\mathbf{1}^{\mathsf{T}}/d$, so that $\mathcal{V}(\boldsymbol{x}) = \|\boldsymbol{C}\boldsymbol{x}\|_2^2/d$. Since $\boldsymbol{x}_{t_i}^{\mathrm{inv}} = \boldsymbol{x}_{t_i}^{\mathrm{gen}} + \boldsymbol{e}_i$,

$$\mathcal{V}(\boldsymbol{x}_{t_i}^{\mathrm{inv}}) - \mathcal{V}(\boldsymbol{x}_{t_i}^{\mathrm{gen}}) = \frac{2}{d}\langle \boldsymbol{C}\boldsymbol{x}_{t_i}^{\mathrm{gen}}, \boldsymbol{C}\boldsymbol{e}_i \rangle + \frac{1}{d}\|\boldsymbol{C}\boldsymbol{e}_i\|_2^2. \tag{25}$$

Spatially non-uniform residuals can therefore alter variance through both their energy and their correlation with the generation state. This motivates a calibrated variance proxy; it is not an equivalence between variance deviation and edge strength. Calibration in Appendix B.2 accounts for the variation already present along ordinary generation trajectories.

A.3 Corresponding Analysis for Flow Matching

For the rectified-flow parameterization underlying SD3-family models (Esser et al., 2024), use a noise coordinate $s \in [0, 1]$, with data at $s = 0$ and noise at $s = 1$:

$$\boldsymbol{x}_s^{\mathrm{fwd}} = (1 - s)\boldsymbol{x}_0 + s\boldsymbol{\epsilon}, \qquad \boldsymbol{v}^{\mathrm{target}} = \boldsymbol{\epsilon} - \boldsymbol{x}_0. \tag{26}$$

Let $s_0 < \cdots < s_K$ and $h_i = s_i - s_{i-1}$. For explicit Euler generation and inversion using the same velocity field,

$$\boldsymbol{x}_{s_{i-1}}^{\mathrm{gen}} = \boldsymbol{x}_{s_i}^{\mathrm{gen}} - h_i \boldsymbol{v}_\theta(\boldsymbol{x}_{s_i}^{\mathrm{gen}}, s_i), \tag{27}$$

$$\boldsymbol{x}_{s_i}^{\mathrm{inv}} = \boldsymbol{x}_{s_{i-1}}^{\mathrm{inv}} + h_i \boldsymbol{v}_\theta(\boldsymbol{x}_{s_{i-1}}^{\mathrm{inv}}, s_{i-1}). \tag{28}$$

Rearranging the generation update and subtracting gives

$$\boldsymbol{e}_i^{\mathrm{FM}} = \boldsymbol{e}_{i-1}^{\mathrm{FM}} + h_i \boldsymbol{\delta}_i^{\mathrm{FM}}, \tag{29}$$

$$\boldsymbol{\delta}_i^{\mathrm{FM}} = \boldsymbol{v}_\theta(\boldsymbol{x}_{s_{i-1}}^{\mathrm{inv}}, s_{i-1}) - \boldsymbol{v}_\theta(\boldsymbol{x}_{s_i}^{\mathrm{gen}}, s_i). \tag{30}$$

For a matched clean endpoint,

$$\boldsymbol{e}_i^{\mathrm{FM}} = \sum_{j=1}^{i} h_j \boldsymbol{\delta}_j^{\mathrm{FM}}. \tag{31}$$

This has the same accumulation structure as Eq. (20), with $a_i = 1$ and $b_i = h_i$. Spatially coherent velocity mismatches can likewise persist under the linear projection in Eq. (23). The derivation concerns discrete numerical updates; it does not assert an intrinsic failure of reversibility for an exactly solved continuous flow.

Under this noise coordinate, the variance baseline used in the main paper becomes

$$\mathcal{V}_s^{\mathrm{fwd}}(\boldsymbol{x}_0) = (1 - s)^2 \mathcal{V}(\boldsymbol{x}_0) + s^2. \tag{32}$$

The calibrated selection rule retains its form, $\Delta_s \leq \lambda\tau_s$. For SD3.5, we use the default `FlowMatchEulerDiscreteScheduler` in the Diffusers pipeline for inference, with its pre-trained scheduler configuration. This explicit Euler solver corresponds to the generation update above.

Table 5: Training configuration and generation resolution.

Setting	SDXL	SD3.5 Medium
Training data	Pick-a-Pic v2	HPD v3 (SD3 pairs)
Preference pairs	851,293	144,991
Hardware	8 A100 GPUs	8 A100 GPUs
Optimizer	Adafactor	Adafactor
β	2000	2000
Learning rate	2.048×10^{-8}	2.048×10^{-8}
Tolerance λ	1	2.5
Calibration trajectories	2000	2000
Calibration quantile	0.95	0.95
Training steps	400	175
Warmup steps	200	50
Effective batch (pairs)	1024	1024
Inversion steps	10	10
Training precision	fp16	fp16
Training resolution	1024×1024	512×512
Generation resolution	1024×1024	1024×1024

B Implementation and Evaluation Details

B.1 Backbones, Data, and Optimization

We use SDXL-base-1.0 and SD3.5 Medium. SDXL is trained on 851,293 non-tied Pick-a-Pic v2 preference pairs spanning 58,960 prompts. For SD3.5, we retain HPD v3 pairs involving at least one SD3-generated image, giving 144,991 pairs. Table 5 collects the training configuration.

Following Diffusion-DPO (Wallace et al., 2024), the learning rate is parameterized as $(2000/\beta) \times 2.048 \times 10^{-8}$. The effective batch counts preference pairs across all GPUs and accumulation steps.

B.2 Reference-Scale Calibration

For each pretrained backbone, we collect $N = 2000$ generation trajectories initialized from Gaussian noise with prompts from the Pick-a-Pic v2 training set. Calibration uses a resolution of 1024×1024 for both SDXL and SD3.5, with DDIM sampling for SDXL and the default flow-matching Euler scheduler for SD3.5. We retain each trajectory's intermediate latents and clean endpoint, and compute the deviations in Eq. (11). The empirical 95th percentile at each noise level defines τ_t. The reference thresholds are fixed throughout preference training, with fallback controlled by $\lambda\tau_t$.

B.3 Training Procedure

For each preference pair, IFPO applies the calibrated variance test to the preferred image at the sampled noise level. If it passes, both images use inversion supervision; otherwise, both use standard forward noising. The selected states and their corresponding targets are used for both the trainable predictor and the frozen reference predictor in Eq. (13).

B.4 Baselines and Generation Protocol

For SDXL, we evaluate the official checkpoints released by the authors of DPO, MaPO, and IAPO using our evaluation pipeline. InPO is reproduced on both backbones; all fine-tuning baselines on SD3.5 are reproduced using the same training data and budget as IFPO.

We generate one image per prompt using 50 denoising steps. Other inference settings follow each backbone's default Diffusers pipeline configuration, including guidance scales of 5.0 for SDXL and 7.0 for SD3.5. All HPS evaluations use HPS v2. PickScore and HPS are multiplied by 100 in the tables.

B.5 Stage-Wise and Stability Experiments

The stage-wise diagnostic uses fixed DPO and InPO models and 50 generation steps, varying $K \in \{0, 5, \ldots, 50\}$. Prompts and initial noise are shared across allocations. Static mix uses the reward-selected 15/35 allocation; Static train maps the same division to the training noise schedule.

For Var. mix, we first generate an image with pretrained SDXL for each prompt and initial noise, invert it, and use the variance rule to select K_0 from the same grid. We generate a DPO/InPO mixed image with K_0, invert that image, and select K_1. Holding the prompt and initial noise fixed, we deliberately use K_1 to generate the reported Var. mix image. For most prompts, inverting this final image selects K_1 again: the variance-selected switch point matches the one actually used to generate the image. No reward scores are used to select either switch point.

Random mix randomly reassigns these final switch points across prompts, preserving their distribution and hence the DPO-use rate at every timestep. Each reported generation uses a single switch point. For each evaluation metric, the oracle independently selects the reward-maximizing K for each prompt from the tested candidates, then reports the mean and median of these per-prompt maxima. Its columns are separate metric-specific upper bounds and need not correspond to the same selected images.

For checkpoint stability, the DPO and InPO generation models remain fixed while the checkpoint used to select the switch point changes. We report exact agreement and agreement within one grid interval (five denoising steps), including unchanged selections, together with mean PickScore, HPS, ImageReward, and MPS on the same Pick-a-Pic test prompts.

C Additional Quantitative Results

C.1 Preference Evaluation on Parti-Prompts

We evaluate each method on 1,600 prompts from Parti-Prompts (Yu et al., 2022), covering diverse prompt categories and challenges. Table 6 reports results for both backbones using the same four evaluators as the main paper.

Table 6: Preference evaluation on Parti-Prompts. Each metric reports mean / median. Best and second-best results within each backbone are bolded and underlined, based on unrounded scores.

Backbone	Method	PickScore		HPS		ImageReward		MPS	
		Mean	Median	Mean	Median	Mean	Median	Mean	Median
SDXL	Base	22.59	22.62	28.02	28.03	0.6694	0.9276	11.16	11.45
	DPO	22.90	22.90	28.65	28.62	<u>1.035</u>	<u>1.278</u>	<u>11.67</u>	11.88
	MaPO	22.59	22.54	28.56	28.46	0.7888	1.270	11.28	11.63
	IAPO	<u>23.02</u>	<u>22.99</u>	<u>28.87</u>	<u>28.86</u>	0.9485	1.204	11.58	<u>11.89</u>
	InPO	22.87	22.92	28.38	28.52	0.7792	1.111	11.24	11.70
	IFPO	**23.23**	**23.22**	**29.08**	**29.07**	**1.161**	**1.406**	**11.88**	**12.16**
SD3.5	Base	<u>22.92</u>	<u>22.91</u>	29.26	29.30	1.240	1.453	**11.81**	**12.14**
	DPO	22.83	22.83	29.13	29.14	1.222	1.442	11.64	12.03
	InPO	22.91	22.88	<u>29.38</u>	<u>29.42</u>	<u>1.263</u>	<u>1.481</u>	<u>11.80</u>	12.01
	IFPO	**23.02**	**23.01**	**29.44**	**29.47**	**1.265**	**1.504**	11.77	<u>12.08</u>

IFPO achieves the highest mean and median scores across all four metrics on SDXL. On SD3.5, it achieves the highest mean and median scores for PickScore, HPS v2, and ImageReward.

D Additional Qualitative Results

D.1 Teaser Prompts

The exact prompts in Figure 1, from left to right, are:

1. "A sign that says Deep Learning"

2. “a thumbnail image of an ice cream cone”
3. “New Zealand Beautiful Seaside Scenery”
4. “A photograph of a woman with curly hair reading a book beside a sunny window.”
5. “a painting of a fox in the style of starry night”

D.2 Additional SDXL Examples

Figure 6 shows further SDXL examples. The prompt for each row is printed beneath its images.

E Limitations and Broader Impacts

IFPO uses calibrated variance as an inexpensive proxy for inversion drift. This second-order statistic cannot capture every spatial or semantic mismatch, and its tolerance may depend on the backbone. Preference annotations in Pick-a-Pic v2 and HPD v3 also reflect the coverage and judgments of their contributors.

Preference optimization inherits the capabilities and potential biases of its pretrained text-to-image backbone. Improved visual quality can benefit creative applications, while realistic generated content can also be misused. Deployment should retain the backbone’s applicable content safeguards and clearly communicate when content is generated.

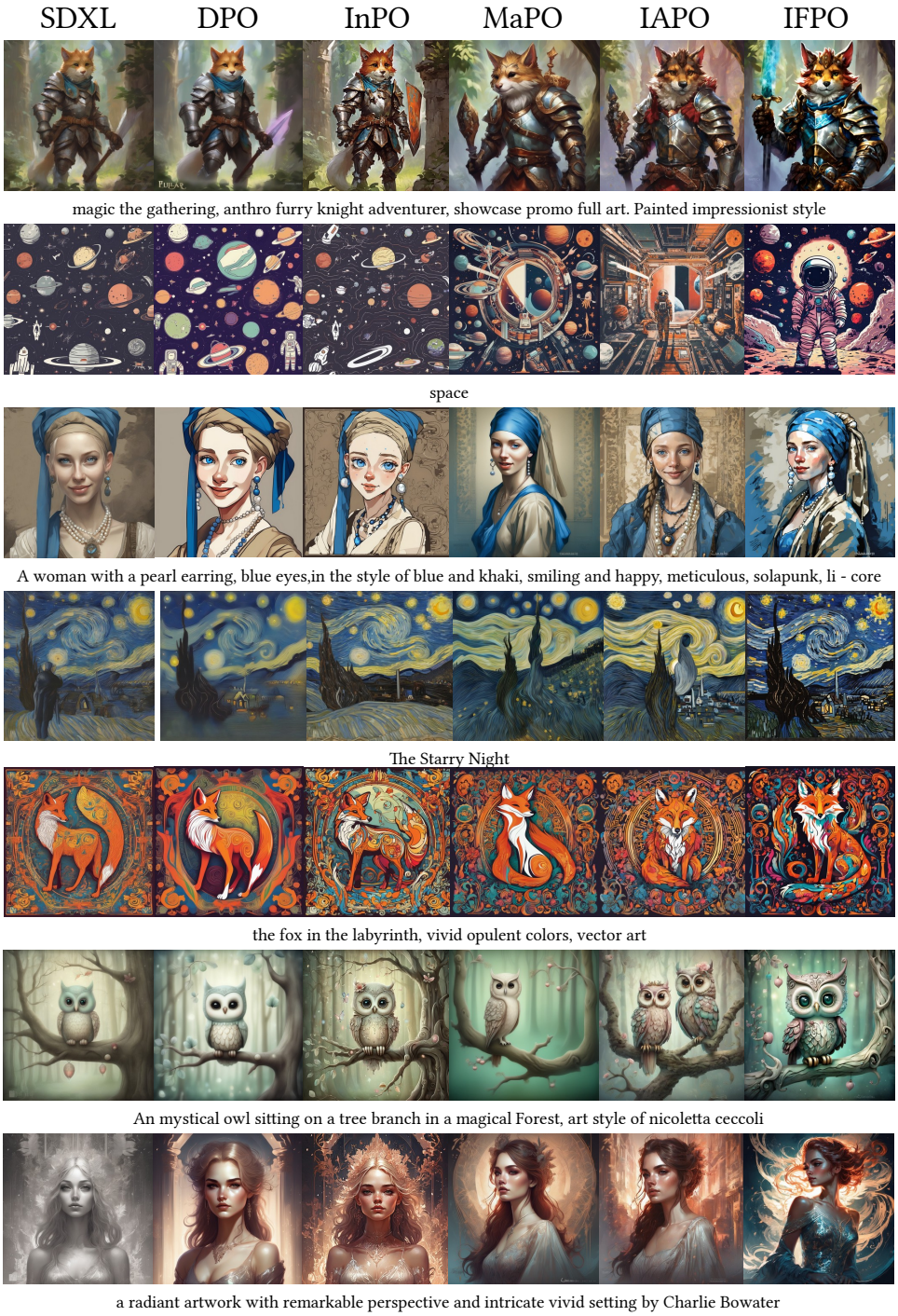

Figure 6: Additional SDXL qualitative examples. Columns show SDXL, DPO, InPO, MaPO, IAPO, and IFPO from left to right. Each row uses the prompt printed immediately below it.