PC-Flow: Preference Alignment in Flow Matching via Classifier

Shaomeng Wang[1], He Wang[1], Longquan Dai[1*], Jinhui Tang[2†]

[1]Nanjing University of Science and Technology
[2]Nanjing Forestry University

Abstract

Flow Matching (FM) is an efficient generative modeling framework, but aligning it with human preferences remains underexplored. Although applying Direct Preference Optimization (DPO) to diffusion models has yielded improvements, directly extending DPO-like methods to FM poses three challenges: 1) Incompatibility with ODE-based models, 2) Heavy computational cost from full model fine-tuning, and 3) Reliance on reference model quality. To address these limitations, we propose Preference Classifier for Flow Matching (PC-Flow), a novel reference-free preference alignment framework. Specifically, we reinterpret FM's deterministic ODE as an equivalent SDE to enable DPO-style learning. Then, we introduce a lightweight classifier to model relative preferences exclusively. This approach decouples alignment from the generative model, eliminating the need for costly fine-tuning or a reference model. Theoretically, PC-Flow guarantees consistent preference-guided distribution evolution, achieves a DPO-equivalent objective without a reference model, and progressively steers generation toward preferred outputs. Experiments show that PC-Flow achieves DPO-level alignment with significantly lower training costs.

Introduction

Flow Matching (FM) (Lipman et al. 2023; Albergo et al. 2023; Liu, Gong, and Liu 2022) is a concise and efficient generative modeling framework with successful applications in image generation (Samaddar et al. 2025; Caetano et al. 2025). Despite major advances in quality and efficiency, FM methods still struggle to align with human subjective preferences. In particular, current models fail to capture nuances like aesthetic perception and semantic alignment, often producing content that falls short of human expectations.

To align generative models with human preferences, methods (Wallace et al. 2024; Lu et al. 2025; Liang et al. 2025; Lee et al. 2025; Zhu, Xiao, and Honavar 2025; Wang et al. 2025b) like Direct Preference Optimization (DPO) have been adopted from language modeling and shown great success in diffusion models. However, applying DPO-like approaches directly to FM for image generation poses three challenges: **1) Incompatibility with ODE-based models**, as DPO is designed for SDEs; **2) Heavy computational cost**, due to full model fine-tuning; and **3) Reliance on reference model quality**, which leads to instability and increased training complexity.

To address above limitations, we propose **Preference Classifier for Flow Matching (PC-Flow)**, a novel human preference alignment framework without a reference model. Specifically, we first convert the deterministic ODE in FM into a Stochastic Differential Equation (SDE), enabling the application of DPO-form objectives within FM models. This ODE-to-SDE conversion allows the FM model to incorporate human preferences effectively. We then introduce a lightweight, trainable classifier to model relative preferences. By isolating preference learning within this classifier, PC-Flow completely decouples alignment from the generative model, eliminating the need for costly full-model fine-tuning or reliance on a reference model. We finally provide rigorous theoretical guarantees for the effectiveness of PC-Flow: 1) its preference guidance ensures the consistent evolution of the aligned distribution throughout the entire generation flow; 2) its training objective is provably equivalent to DPO without the need for a reference model; 3) its preference guidance progressively induces the velocity field to evolve toward regions aligned with human preferences.

To summarize, the contributions of PC-Flow are as follows:

- **Reference-Free Preference Alignment:** We propose PC-Flow, a reference-free framework for Flow Matching that utilizes a lightweight, trainable classifier to model preferences. This fully decouples the alignment task from the generative model, eliminating the need for expensive fine-tuning and reliance on reference models, which significantly reduces computational overhead.
- **Rigorous Theoretical Guarantees:** We provide theoretical proof of PC-Flow, establishing that: **1)** its preference-guidance consistently propagate across the flow trajectory; **2)** its training objective is mathematically equivalent to reference-free DPO; and **3)** its preference guidance progressively steers the velocity field toward regions aligned with human preferences.

*Corresponding author
†Corresponding author

- **Superior Performance and Efficiency:** Extensive experiments demonstrate that PC-Flow achieves alignment performance comparable or superior to DPO-like methods, while requiring significantly fewer computational resources and offering greater training stability. These results validate PC-Flow as a practical and scalable solution for preference alignment in FM.

Related Work

This section briefly reviews the literature on diffusion, flow matching and alignment in diffusion models.

Diffusion and Flow Matching

Diffusion models (DMs) (Ho, Jain, and Abbeel 2020; Song, Meng, and Ermon 2020) synthesize data by learning to invert a progressive noising process via a trained denoising network. Leveraging this mechanism, DMs have demonstrated strong performance in modeling complex distributions, particularly in image synthesis (Shen et al. 2025b; Li et al. 2025; Yao, Yang, and Wang 2025; Shen et al. 2025a; Shen and Tang 2024) and video generation (Ho et al. 2022; Wu et al. 2023a; Ni et al. 2023). Despite their success, DMs often require many steps during inference due to the stochastic nature of the sampling process. In contrast to DMs, which are often formulated using Stochastic Differential Equations (SDEs) (Song, Meng, and Ermon 2020), Flow Matching (FM) (Liu, Gong, and Liu 2022; Lipman et al. 2023; Dao et al. 2023) adopts a deterministic approach. Specifically, FM learns a continuous-time deterministic vector field that maps noise distributions to data distributions via ordinary differential equation (ODE) trajectories. By eliminating the need for Markov chains and Gaussian perturbations, FM significantly reduces the number of inference steps required while maintaining generation quality. As a result, flow matching has emerged in recent years as a leading paradigm in image and video generation (Martin et al. 2024; Labs et al. 2025; Wang et al. 2024a) due to its superior inference efficiency.

Preference Alignment for Diffusion Models

Recent efforts (Lu et al. 2025; Liang et al. 2025; Zhu, Xiao, and Honavar 2025) to align generative models with human preferences have increasingly adapted techniques from LLMs (Rafailov et al. 2023; Tang, He, and Qin 2025; Yan et al. 2025; Wang et al. 2025a). Among these, Direct Preference Optimization (DPO) (Wallace et al. 2024) has emerged as a simpler and more effective alternative to Reinforcement Learning from Human Feedback (RLHF) (Wang et al. 2024b) for enhancing image quality. Inspired by its success, we are the first to introduce DPO into the prevailing FM framework to enable efficient text-to-image generation. Despite the success of DPO-like methods in diffusion models, their application to Flow Matching faces three key challenges: **1) Incompatibility with ODE-based models**, as DPO is tied to SDE-based models; **2) Heavy computational cost** from full model fine-tuning; and **3) Reliance on reference model quality**, which can introduce training instabilities and undesirable biases into the final model. To address these issues, we propose PC-Flow, an extension of the previous method (Wang et al. 2025b) and a novel reference-free alignment framework. On one hand, we first converse the ODE process in FM to an equivalent SDE representation to accommodate a DPO-style learning objective. On the other hand, it introduces a lightweight, trainable classifier to directly model relative preferences, thereby achieving efficient and stable preference alignment.

Preliminaries

In this section, we introduce the necessary background for our work, including Flow Matching (FM) as a deterministic generative modeling framework, and Direct Preference Optimization (DPO) as a widely used preference alignment method in diffusion models. We also discuss the key challenges of directly applying DPO to FM-based models.

Flow Matching

Flow Matching (FM) is a deterministic generative modeling framework that learns a continuous mapping from a source distribution $p_0(x)$ (e.g., Gaussian) to a target data distribution $p_1(x)$. This mapping is defined by an Ordinary Differential Equation (ODE):

$$\frac{d\mathbf{x}_t}{dt} = \mathbf{v}_\phi(\mathbf{x}_t, t), \quad t \in [0, T], \tag{1}$$

where v_ϕ is a learnable time-dependent velocity field.

DPO for Diffusion Models

Although Flow Matching is efficient, aligning it with human preferences remains difficult. To provide context, we briefly review DPO in diffusion models. Following (Wallace et al. 2024), DPO compares transition probabilities at intermediate time steps. Given a preference triplet $\left(x_0^w, x_0^l\right)$, it trains a model p_θ to favor preferred outputs over a reference model p_{ref}, omitting the prompt c for simplicity. The DPO-Diffusion objective is then defined as:

$$\begin{aligned} L_{\text{DPO-Diffusion}}(\theta) \leq -\mathbb{E} \; & \begin{array}{r} \left(\boldsymbol{x}_0^w, \boldsymbol{x}_0^l\right) \sim \mathcal{D}, t \sim \mathcal{U}(0,T) \\ \boldsymbol{x}_{t-1,t}^w \sim p_\theta\left(\boldsymbol{x}_{t-1,t}^w \mid \boldsymbol{x}_0^w\right) \\ \boldsymbol{x}_{t-1,t}^l \sim p_\theta\left(\boldsymbol{x}_{t-1,t}^l \mid \boldsymbol{x}_0^l\right) \end{array} \\ & \log \sigma \left(\beta T \log \frac{p_\theta\left(\boldsymbol{x}_{t-1}^w \mid \boldsymbol{x}_t^w\right)}{p_{\text{ref}}\left(\boldsymbol{x}_{t-1}^w \mid \boldsymbol{x}_t^w\right)} \right. \\ & \left. - \beta T \log \frac{p_\theta\left(\boldsymbol{x}_{t-1}^l \mid \boldsymbol{x}_t^l\right)}{p_{\text{ref}}\left(\boldsymbol{x}_{t-1}^l \mid \boldsymbol{x}_t^l\right)} \right), \end{aligned} \tag{2}$$

Here, $\mathcal{D}$ denotes the dataset of human preference pairs, and T is the total number of diffusion steps.

While DPO is effective in diffusion models, applying it to Flow Matching poses challenges due to: 1) incompatibility with ODE-based models, 2) heavy computational cost, and 3) reliance on reference model quality.

To address these limitations, we propose Preference Classifier Guidance for Flow Matching (PC-Flow), a lightweight and reference-free framework that enables effective preference alignment in ODE-based generative models.

Preference Classifier for Flow Matching

We propose an efficient preference alignment framework tailored for FM models, consisting of four key components: **1) ODE-to-SDE Conversion**, which provides the theoretical foundation for enabling preference alignment within FM. **2) Preference Classifier Definition**, which guides the generative process by adjusting output probabilities without retraining the base mode. **3) Preference Classifier Training Objective**, which effectively incorporates human preferences without relying on an reference model. **4) Preference-Guided Sampling**, which uses the human preference prior embedded in the classifier to generate preference samples.

ODE-to-SDE Conversion

The DPO objective (Eq. 2) requires comparing log-probabilities, a feature inherent to SDE-based diffusion models but absent in the deterministic ODE formulation of FM (Eq. 1). To bridge this incompatibility, we reinterpret the deterministic ODE as an equivalent SDE that preserves the original marginal distributions at all timesteps, following by Song, Meng, and Ermon (2020); Liu et al. (2025); Deveney et al. (2025). This process yields the following stochastic formulation for FM:

$$\mathrm{d}\boldsymbol{x}_t=\left(\boldsymbol{v}_t\left(\boldsymbol{x}_t,t\right)-\frac{\sigma_t^2}{2}\nabla\log p_t\left(\boldsymbol{x}_t,t\right)\right)\mathrm{d}t+\sigma_t\mathrm{d}\boldsymbol{W},\quad(3)$$

where $\mathrm{d}\boldsymbol{W}$ represents Wiener process increments, σ_t controls the level of noise during generation, and the marginal score $\nabla\log p_t\left(\boldsymbol{x}_t,t\right)$ is related to velocity:

$$\nabla\log p_t(\boldsymbol{x_t},\boldsymbol{t})=-\frac{\boldsymbol{x}}{t}-\frac{1-t}{t}\boldsymbol{v}_t(\boldsymbol{x},\boldsymbol{t}).\quad(4)$$

For practical implementation, we discretize the SDE using the Euler-Maruyama method, yielding the following single-step update rule:

$$\begin{aligned}\boldsymbol{x}_{t+\Delta t}=\boldsymbol{x}_t+\Bigg[&\boldsymbol{v}_\phi\left(\boldsymbol{x}_t,t\right)\\&+\frac{\sigma_t^2}{2t}\left(\boldsymbol{x}_t+(1-t)\,\boldsymbol{v}_\phi\left(\boldsymbol{x}_t,t\right)\right)\Bigg]\Delta t\\&+\sigma_t\sqrt{\Delta t}\,\boldsymbol{\epsilon},\end{aligned}\quad(5)$$

where $\epsilon\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})$ injects stochasticity.

Preference Classifier Definition

Our ODE-to-SDE conversion equips FM with a marginal-preserving probabilistic structure, enabling likelihood-based preference modeling through well-defined transition probabilities $p_\phi(x_{t-1}\mid x_t)$, similar to those used in DPO.

In DPO, preference alignment is achieved by fine-tuning the entire model to transform a misaligned distribution $p_\phi\left(x_t\right)$ into a preference-aligned one $\hat{p}_\phi\left(x_t\right)$, and propagating it backward via updated transitions $\hat{p}_\phi\left(x_{t-1}\mid x_t\right)$.

In contrast, we propose Preference Classifier for Flow Matching (PC-Flow), which aligns $p_\phi\left(x_t\right)$ using a lightweight preference classifier $\mathcal{S}_\theta(x)$, yielding a preference-guided distribution $\hat{p}_\phi^\theta\left(x_t\right)$ and transition $\hat{p}_\phi^\theta\left(x_{t-1}\mid x_t\right)$, without modifying the base model. We formally define the components of our framework below.

Definition 1 (Preference Classifier Guidance) *The trainable* ***preference classifier*** $\mathcal{S}_\theta(x):\mathbb{R}^d\to(0,1)$ *assigns a human preference score to sample x (higher scores indicate greater preference). Based on $\mathcal{S}_\theta(x)$, we define the preference-guided distribution $\hat{p}_\phi^\theta(x_t)$ and the corresponding preference-guided transition $\hat{p}_\phi^\theta(x_{t-1}\mid x_t)$ respectively:*

$$\hat{p}_\phi^\theta(x_t)=N_t p_\phi(x_t)\exp\left(\log\mathcal{S}_\theta(x_t)\right),\quad(6)$$

$$\hat{p}_\phi^\theta(x_{t-1}\mid x_t)=p_\phi(x_{t-1}\mid x_t)\exp\Big(\log\mathcal{S}_\theta(x_{t-1})-\log\mathcal{S}_\theta(x_t)\Big),\quad(7)$$

where N_t is a normalization term.

Since DPO propagates $\hat{p}_\phi\left(x_t\right)$ by fine-tuning the entire model, we investigate whether our proposed formulation in Eq. 7 can achieve the same. Theorem 1 guarantees the correctness of the preference-guided distribution transformation.

Theorem 1 *Given a preference classifier $\mathcal{S}_\theta(x)$, suppose that x_t is sampled from $\hat{p}_\phi^\theta(x_t)$, and then apply the transition $\hat{p}_\phi^\theta(x_{t-1}\mid x_t)$, we can generate:*

$$\hat{p}_\phi^\theta(x_{t-1})=N_{t-1}p_\phi(x_{t-1})\exp\left(\log\mathcal{S}_\theta(x_{t-1})\right),\quad(8)$$

where N_{t-1} is normalization term.

Thus, this theorem guarantees that preference-guided distributions can be propagated across timesteps solely by optimizing the classifier $\mathcal{S}_\theta$, without modifying the base model.

Preference Classifier Training Objective

Having shown that the preference classifier $\mathcal{S}_\theta(x)$ defines how the base distribution $p_\phi(x_t)$ is transformed into a preference-aware distribution $\hat{p}_\phi^\theta(x_t)$, we now turn to how $\mathcal{S}_\theta(x)$ can be trained to reflect human preferences. Rather than fine-tuning the full generative model as in DPO, our PC-Flow framework isolates the preference learning into the lightweight classifier $\mathcal{S}_\theta(x)$ alone.

This separation is justified by our derivation from Eq. 7, which reveals that the alignment adjustment at each step is governed solely by the ratio of the classifier's scores:

$$\frac{\hat{p}_\phi^\theta(x_{t-1}\mid x_t)}{p_\phi(x_{t-1}\mid x_t)}=\frac{\mathcal{S}_\theta(x_{t-1})}{\mathcal{S}_\theta(x_t)}.\quad(9)$$

Crucially, this insight implies that we only need to ensure $\mathcal{S}_\theta(x)$ accurately reflects human judgment. This allows us to derive a reference-free training objective called PC-Flow Loss for $\mathcal{S}_\theta(x)$ using standard preference tuple $\left(x^w,x^l\right)\sim\mathcal{D}$, as formalized in Eq. 10:

$$\mathcal{L}_{\text{PC-Flow}}(\mathcal{S}_\theta)=-\mathbb{E}_{\left((x_t^w,x_{t-1}^w),(x_t^l,x_{t-1}^l)\right)\sim\mathcal{D}}\left[\log\sigma\Big(\beta T\log\frac{\mathcal{S}_\theta(x_{t-1}^w)}{\mathcal{S}_\theta(x_t^w)}-\beta T\log\frac{\mathcal{S}_\theta(x_{t-1}^l)}{\mathcal{S}_\theta(x_t^l)}\Big)\right],\quad(10)$$

where $\sigma(\cdot)$ denotes the sigmoid function and $\mathcal{D}$ is the preference dataset. This objective is guaranteed by Theorem 2.

Theorem 2 *Given $\mathcal{S}_\theta(x)$, $\hat{p}_\phi^\theta(x_{t-1} \mid x_t)$, and human preference samples across timesteps, the PC-Flow training objective in Eq. 10 can be derived from the DPO loss (Eq. 2) without requiring a reference model.*

Theorem 2 eliminates the need for a reference model, which is central to traditional DPO methods (Eq. 2), by directly modeling relative preferences through the classifier $\mathcal{S}_\theta$.

The PC-Flow framework offers notable practical benefits: **1) Reference-Free Optimization:** The model training does not depend on a reference model, reducing complexity and avoiding instability or bias introduced by imperfect references. **2) High Efficiency:** Only the lightweight classifier $\mathcal{S}_\theta$ is trained, allowing efficient, plug-and-play alignment with pre-trained diffusion models without requiring costly fine-tuning of the base models.

Preference-Guided Sampling

After training the preference classifier $\mathcal{S}_\theta(x)$, we now incorporate it into the sampling process of flow matching models. The objective is to generate samples that follow the learned data distribution while aligning with human preferences. This section presents a principled method to achieve this goal by modifying the velocity field used in generation.

In flow matching, generation is driven by a base velocity field $\boldsymbol{v}_\phi(x_t, t)$, which lacks preference awareness. To steer generation toward preferred outputs while preserving the original modeling structure, we construct a new velocity field $\boldsymbol{v}_\theta(x_t, t)$ by introducing an additive correction based on the preference classifier $\mathcal{S}_\theta(x)$.

To incorporate human preferences into the generative process, we define a reweighted distribution over x_t that adjusts the original model distribution using a learned preference score. By omitting the normalization term, Eq. 6 can be equivalently expressed as:

$$\hat{p}_\phi^\theta(x_t) \propto p_\phi(x_t)\mathcal{S}_\theta(x_t), \tag{11}$$

where $p_\phi(x_t)$ denotes the base generative distribution, and $\mathcal{S}_\theta(x_t)$ represents the human preference score assigned to sample x_t. Based on Eq. 11, we now outline the core steps leading to the construction of the preference-guided velocity field.

To derive the preference-guided velocity field, we begin by computing the score function of the reweighted distribution:

$$\nabla_{x_t} \log \hat{p}_\phi^\theta(x_t) = \nabla x_t \log p_\phi(x_t) + \nabla_{x_t} \log \mathcal{S}_\theta(x_t). \tag{12}$$

Then, from the SDE formulation of flow matching, as defined in Eq. 4, the marginal score of the original distribution satisfies:

$$\nabla_{x_t} \log p_\phi(x_t) = -\frac{x_t}{t} - \frac{1-t}{t}\boldsymbol{v}_\phi(x_t, t). \tag{13}$$

Next, substituting Eq.13 into Eq.12 and rearranging terms yields the preference-guided velocity field:

$$\boldsymbol{v}_\theta(x_t, t) = \boldsymbol{v}_\phi(x_t, t) + \gamma\frac{t}{1-t}\nabla_{x_t} \log \mathcal{S}_\theta(x_t), \tag{14}$$

where γ is a tunable guidance weight. Eq. 14 shows that preference-aligned velocity $v_\theta(x_t, t)$ can be achieved by augmenting the base velocity $v_\phi(x_t, t)$ with a time-scaled classifier gradient $\nabla_{x_t} \log \mathcal{S}_\theta(x_t)$, without modifying the underlying generative model.

Finally, by substituting Eq.14 into Eq.15, we obtain a preference-aligned sample $x_{t+\Delta t}$ at each step. Theorem 3 guarantees that Eq. 15 serves as a good sampling strategy for drawing samples from the distribution defined in Eq 11.

Theorem 3 *Given the $\hat{p}_\phi^\theta(x_t)$ in Eq. 11, marginal score $\nabla_{x_t} \log p_\phi(x_t)$ in Eq. 13, and preference-guide velocity field $v_\theta(x_t, t)$ in Eq. 14, the $x_{t+\Delta t}$ can be sampled by Eq. 15:*

$$\begin{aligned} \boldsymbol{x}_{t+\Delta t} = \boldsymbol{x}_t + \Bigg[& \boldsymbol{v}_\theta\left(\boldsymbol{x}_t, t\right) \\ & + \frac{\sigma_t^2}{2t}\left(\boldsymbol{x}_t + (1-t)\,\boldsymbol{v}_\theta\left(\boldsymbol{x}_t, t\right)\right)\Bigg]\Delta t \\ & + \sigma_t\sqrt{\Delta t}\,\boldsymbol{\epsilon}. \end{aligned} \tag{15}$$

Experiment

In this section, we evaluate the effectiveness of Preference Classifier for Flow Matching (PC-Flow) in aligning generated samples with human preferences while maintaining high-quality outputs. We begin by describing the experimental setup, followed by evaluations on two alignment tasks: **1) Aesthetic Alignment:** This task guides the generative model toward producing visually appealing outputs that reflect human aesthetic preferences. **2) Text-to-Image Alignment:** This task assesses how accurately and comprehensively the generated image reflects the semantic content and specific details of the input text prompt.

Experimental Setup

Models and Datasets: We use the pre-trained Flow Matching model, which named as Stable Diffusion 3.5 Medium (SD-3.5-M) (Esser et al. 2024) equipped with a Multimodal Diffusion Transformer (MMDiT) as the base generative model for all experiments. For aesthetic and text-to-image alignment tasks, we train our preference classifier $\mathcal{S}_\theta$ on the Pick-a-Pic dataset (Kirstain et al. 2023), which consists of pair-wise preferences for images generated by SDXL-beta and Dreamlike, a fine-tuned version of SD1.5. For final testing, we generate images using the PC-Flow models conditioned on captions from the Partiprompt (Yu et al. 2022) and HPSV2 benchmark (Wu et al. 2023b).

Training Setup: Our preference classifier $\mathcal{S}_\theta$ can be easily adapted to various alignment tasks by adjusting its input modalities and architecture. We implement task-specific variants as follows: **1) In the aesthetic alignment task:** The preference classifier $\mathcal{S}_\theta$ is composed of two components: a frozen U-Net without downsampling layers and a trainable two-layer MLP. We utilize the middle layers of the U-Net to extract and refine the features of a given timestep x_t. These features are then fed into the MLP, which uses ReLU and Softplus as activation functions, to predict the final preference score. Furthermore, to enable $\mathcal{S}_\theta$ to distinguish inputs from different timesteps, we also embed the time embeddings into the U-Net. **2) For the text-to-image task:** We extend the aesthetic alignment preference classifier $\mathcal{S}_\theta$ by incorporating a trainable cross-attention module, allowing the model to better capture semantic information from the text prompt. All classifiers are trained using the PC-Flow loss defined in Eq.10 for 2,000 steps with the AdamW optimizer(Loshchilov and Hutter 2017) and an image resolution of 512. We set $\beta = 200$, use a batch size of 8 with gradient accumulation over 2 steps, a learning rate of 1×10^{-5}, and apply a constant learning rate schedule with warm-up. All experiments are conducted on 2 NVIDIA A6000 GPUs.

Inference Setup: For all inference phases across the models, we employ the following configuration to fair comparison. At each timestep t, the sampling update is augmented with a preference-guided correction score derived from the preference classifier $\mathcal{S}_\theta(x_t, t)$, as defined in Eq. 15 based on Eq. 14. We uniformly generate images at a resolution of 512×512 with 50 inference steps and apply Classifier-Free Guidance (CFG) (Ho and Salimans 2022) with a guidance scale set to 7.0. The PC-Flow guidance weight hyperparameter γ is uniformly set to 0.1.

Aesthetic Alignment

Aesthetic alignment aims to orient generative models toward outputs that better conform to human aesthetic preferences. We assess the efficacy of the PC-Flow in enhancing image aesthetics and aligning with human preferences, using four quantitative scores, including PickScore (Kirstain et al. 2023), HPS (Wu et al. 2023b), LAION Aesthetics (Schuhmann et al. 2022), and CLIP (Radford et al. 2021), and qualitative visual comparisons.

Qualitative Comparison. Figure 2 presents a qualitative comparison between our PC-Flow method and four baselines: SD-3.5-M, Flux1.dev, AuraFlow, and SANA-1.5 1.6B. Here, SD-3.5-M represents the original, unaligned flow matching model. As shown in the figure, images generated by PC-Flow demonstrate a consistent improvement in aesthetic quality. By using a learned preference classifier for gradient corrections, our method incorporates human preferences without requiring the entire model fine-tuning, and generates images with vivid colors, dramatic lighting, and coherent composition. In contrast, baseline methods often exhibit issues such as muted tones, flat illumination, or awkward spatial arrangements, leading to visually less engaging results that lack the richness, contrast, and compositional balance achieved by PC-Flow.

Quantitative Evaluation. We evaluate our method against SD-3.5-M, Flux1.dev, AuraFlow, and SANA-1.5 using both automated metrics and user studies. For the automated evaluation, we use Pick Score, HPS, Aesthetics, and CLIP with prompts from HPS and PartiPrompts. As shown in Table 1, our method significantly outperforms SD-3.5-M and AuraFlow, and shows consistent gains over Flux1.dev and SANA, especially in HPS and CLIP. For the user studies, we ask annotators to compare three aspects: general preference, visual appeal, and prompt alignment and summarize the winning percentage in Fig. 1. Results show PC-Flow consistently wins in visual appeal and prompt alignment, enhancing both aesthetic quality and semantic relevance.

Comparison with State-of-the-Art Methods. We evaluate the aesthetic alignment performance of PC-Flow against recent flow matching models, including SD-3.5-M, Flux1.dev, AuraFlow, and SANA-1.5 1.6B, using their original configurations. All models are assessed on the full PartiPrompts and HPS benchmarks across four preference metrics. As shown in Table2, PC-Flow consistently surpasses the base model SD-3.5-M across all metrics and datasets, validating the effectiveness of our preference-guided alignment. Compared with Flux1.dev and AuraFlow, PC-Flow achieves higher scores on nearly every metric, indicating stronger capabilities in both aesthetic quality and prompt alignment. Although SANA-1.5 attains the highest Aesthetics score (5.96 in PartiPrompts and 6.17 in HPS), PC-Flow achieves comparable results (5.94 and 6.14) while outperforming it in HPS and CLIP,suggesting a more balanced performance that better captures both visual appeal and semantic consistency. These results demonstrate that PC-Flow not only improves subjective aesthetics but also maintains faithful adherence to the given prompts.

Text-to-Image Alignment

Text-to-Image Alignment measures how accurately and comprehensively a generated image visually represents the semantic meaning and specific details of its input text prompt. As visually demonstrated in Figure 3, PC-Flow excels at translating challenging prompts into semantically coherent and faithful images. For example, given the prompt *'A baby red panda wearing cake as a hat'*, PC-Flow precisely renders both the character and the unusual object-attribute combination, whereas other models often miss such intricate relationships. Similarly, for prompts like *'A glass of orange juice next to an empty pitcher'* and *'A margarita next to a napkin'*, our method successfully preserves spatial relationships and object attributes, outperforming baselines that may confuse layout or context. These qualitative results showcase PC-Flow's superior capacity for fine-grained text-to-image understanding, resulting in improved semantic alignment and overall fidelity in generated images.

Ablation Study

Effect of temperature parameter β. Ablation on β is conducted for aesthetic alignment on HPSV2, which scales the log-ratio terms in the PC-Flow loss (Eq.10) and modulates the sharpness of preference contrast between positive and negative samples. As shown in Table. 3, setting $\beta = 200$, which we use in all tasks, achieves the best performance across all metrics, outperforming both smaller ($\beta = 20$) and larger ($\beta = 2000$) values. A small β weakens the contrast between preferred and less-preferred samples, reducing the effectiveness of preference modeling. In contrast, an excessively large β may over-amplify the logit differences, leading to optimization instability or overfitting. Notably, since our preference classifier is relatively lightweight, setting β too high can exceed the classifier's representational capacity, causing gradient explosion or training collapse. This suggests that optimal β values depend on model capacity: larger models can tolerate higher scaling, while smaller ones perform better with moderate β. These results suggest that mod-

	PartiPrompts				HPS benchmark			
	PickScore	HPS	Aesthetic	CLIP	PickScore	HPS	Aesthetic	CLIP
vs. SD-3.5-M	**61.36**	**76.82**	**75.22**	**60.72**	**67.35**	**75.47**	**74.54**	**70.14**
vs. Flux1.dev	**57.85**	**51.02**	48.90	**64.88**	**51.53**	**50.13**	49.02	**65.25**
vs. AuraFlow	**80.62**	**71.33**	**77.22**	**70.92**	**87.08**	**75.90**	**73.50**	**70.57**
vs. SANA-1.5	47.77	**54.72**	49.90	**51.60**	44.74	**56.06**	45.57	**55.21**

Table 1: Quantitative Win-rate Comparison Using Automated Preference Metrics. We evaluate the alignment performance of diffusion models using prompts from HPS and PartiPrompts across various evaluators. SD-3.5-M serves as base models. Win rates above 50%—indicating superior performance over the baseline—are highlighted in bold.

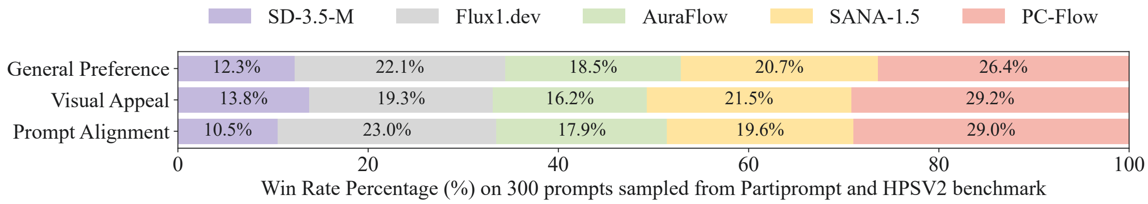

Figure 1: User Study Results Comparing PC-Flow with SD-3.5-M, Flux1.dev, AuraFlow, and SANA-1.5 1.6B. We sampled 100 and 200 prompts for evaluation from Partiprompts and HPSV2 benchmark, respectively. PC-Flow yields clear improvement in visual appeal.

Figure 2: Qualitative Comparison Between PC-Flow and Other Flow Matching Models. Applied to SD-3.5-M (Esser et al. 2024), PC-Flow achieves superior visual and textual alignment compared to baselines including SD-3.5-M, Flux1.dev (Labs et al. 2025), AuraFlow (Fal 2024), SANA-1.5 1.6B (Xie et al. 2025).

erate preference scaling leads to more stable and effective alignment.

Effect of Guidance Weight γ. We further ablate γ in aesthetic alignment using HPSV2, which controls the magnitude of the preference-guidance correction added to the base velocity field (as defined in Eq.14). This parameter directly affects how strongly human preference gradients steer the generation process. As reported in Table. 4, performance improves as γ increases from 0.05 to 0.10, where all metrics reach their peak. However, further increasing γ leads to a slight decline across most metrics, suggesting that excessive guidance may interfere with the underlying generative process and reduce output diversity. These results highlight the importance of calibrating the strength of preference alignment: a moderate value like $\gamma = 0.10$ offers an optimal trade-off between human preferences and maintaining the quality and diversity of outputs. We adopt $\gamma = 0.10$ as the default setting in all subsequent experiments.

	PartiPrompts				HPS benchmark			
	P-S↑	HPS↑	AES↑	CLIP↑	P-S↑	HPS↑	AES↑	CLIP↑
SD-3.5-M	22.53	28.60	5.57	34.37	22.10	27.93	5.90	33.99
Flux1.dev	**22.80**	28.70	5.85	32.86	22.20	28.90	6.01	33.30
AuraFlow	21.25	27.36	5.36	31.31	21.13	28.03	5.61	32.59
SANA-1.5	22.71	28.71	**5.96**	33.80	**22.41**	28.70	**6.17**	34.39
PC-Flow	22.63	**28.76**	5.94	**34.95**	22.18	**28.95**	6.14	**34.45**

Table 2: Alignment Comparison with State-of-the-Art Flow Matching Models Using PartiPrompts and HPS. Evaluation is conducted across four preference metrics with SD-3.5-M as the base model. P-S: PickScore. AES: Aesthetic

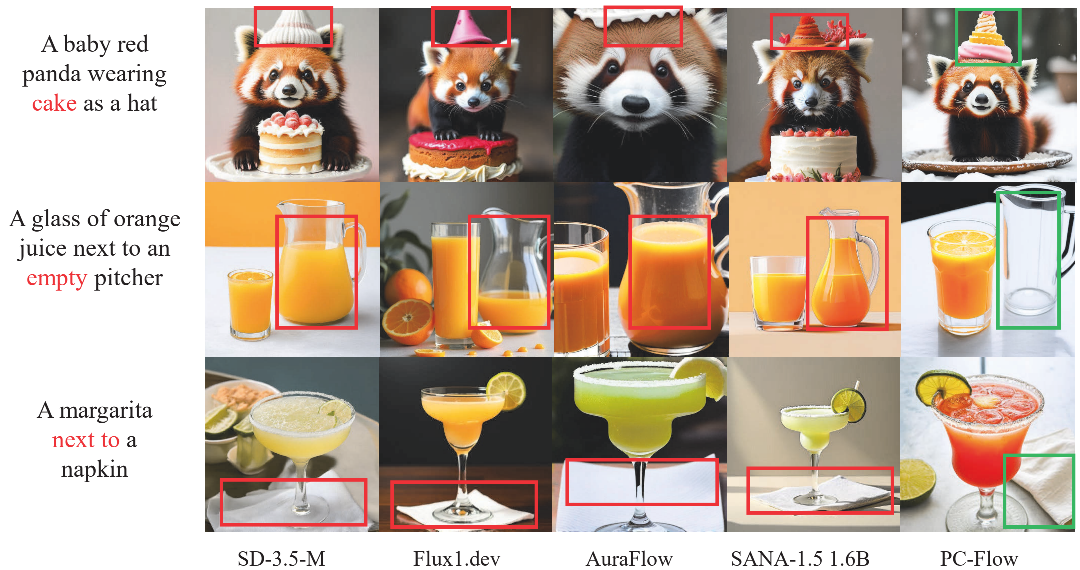

Figure 3: Qualitative Comparison of Text-to-Image Alignment. This figure showcases images generated by PC-Flow and other methods (SD-3.5-M, Flux1.dev, AuraFlow, and SANA-1.5 1.6B) for the text prompts listed on the left. PC-Flow achieves enhanced semantic fidelity, particularly in accurately elements specified in the prompt, such as color and object positioning. Red boxes: Regions mismatched to prompts. Green boxes: Regions matched to prompts.

β	P-S ↑	HPS ↑	AES ↑	CLIP ↑
5	20.34	27.53	5.41	32.53
20	22.18	28.67	5.81	34.16
200	**22.27**	**28.85**	**6.14**	**34.45**
2000	21.92	28.26	5.87	33.93

Table 3: Effect of Temperature Parameter β. We report results using PickScore (P-S) and Aesthetics (AES) as metrics.

γ	P-S ↑	HPS ↑	AES ↑	CLIP ↑
0.05	22.24	28.80	5.84	34.21
0.10	**22.27**	**28.85**	**6.14**	**34.45**
0.15	22.24	28.82	5.87	34.35
0.20	22.20	28.76	5.85	34.23

Table 4: Effect of Guidance Weight γ. We report results using PickScore (P-S) and Aesthetics (AES) as metrics.

Conclusion

In this work, we propose **PC-Flow**, a lightweight and reference-free framework for human preference alignment in Flow Matching (FM) models. Our approach addresses key limitations of directly applying DPO to ODE-based models, namely incompatibility with ODE based models, heavy computational cost, and reliance on reference model quality. By reinterpreting the FM generative process as a marginal-preserving SDE, and introducing a trainable preference classifier, PC-Flow enables efficient and scalable preference alignment without modifying the base generative model. We provide theoretical guarantees that PC-Flow consistently propagates preference-guided distributions, achieves a DPO-equivalent training objective, and effectively steers generation toward human-preferred outputs. Empirical results across various metrics and user studies validate the effectiveness and efficiency of PC-Flow, showing competitive or superior alignment performance compared to DPO-based baselines with significantly lower training cost.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (62372237,62332010) and the Major Science and Technology Projects in Jiangsu Province under Grant BG2024042.

References

Albergo, M. S.; Goldstein, M.; Boffi, N. M.; Ranganath, R.; and Vanden-Eijnden, E. 2023. Stochastic Interpolants with Data-Dependent Couplings. *ArXiv*.

Caetano, F.; Viviers, C.; De With, P. H.; and van der Sommen, F. 2025. Symmetrical Flow Matching: Unified Image Generation, Segmentation, and Classification with Score-Based Generative Models. *ArXiv*.

Dao, Q.; Phung, H.; Nguyen, B.; and Tran, A. 2023. Flow Matching in Latent Space. *ArXiv*.

Deveney, T.; Stanczuk, J.; Kreusser, L.; Budd, C.; and Schönlieb, C.-B. 2025. Closing the ODE-SDE Gap in Score-Based Diffusion Models through the Fokker-Planck Equation. *Philosophical Transactions A*.

Esser, P.; Kulal, S.; Blattmann, A.; Entezari, R.; Müller, J.; Saini, H.; Levi, Y.; Lorenz, D.; Sauer, A.; Boesel, F.; et al. 2024. Scaling Rectified Flow Transformers for High-Resolution Image Synthesis. In *International Conference on Machine Learning*.

Fal. 2024. Introducing AuraFlow v0.1, An Open Exploration of Large Rectified Flow Models. Available at: https://blog.fal.ai/auraflow/.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising Diffusion Probabilistic Models. *Neural Information Processing Systems*.

Ho, J.; and Salimans, T. 2022. Classifier-Free Diffusion Guidance. *ArXiv*.

Ho, J.; Salimans, T.; Gritsenko, A.; Chan, W.; Norouzi, M.; and Fleet, D. J. 2022. Video Diffusion Models. *Neural Information Processing Systems*.

Kirstain, Y.; Polyak, A.; Singer, U.; Matiana, S.; Penna, J.; and Levy, O. 2023. Pick-a-Pic: An Open Dataset of User Preferences for Text-to-Image Generation. *Neural Information Processing Systems*.

Labs, B. F.; Batifol, S.; Blattmann, A.; Boesel, F.; Consul, S.; Diagne, C.; Dockhorn, T.; English, J.; English, Z.; Esser, P.; et al. 2025. FLUX. 1 Kontext: Flow Matching for In-Context Image Generation and Editing in Latent Space. *ArXiv*.

Lee, K.; Li, X.; Wang, Q.; He, J.; Ke, J.; Yang, M.-H.; Essa, I.; Shin, J.; Yang, F.; and Li, Y. 2025. Calibrated Multi-Preference Optimization for Aligning Diffusion Models. In *Computer Vision and Pattern Recognition*.

Li, Z.; Li, H.; Shi, Y.; Farimani, A. B.; Kluger, Y.; Yang, L.; and Wang, P. 2025. Dual Diffusion for Unified Image Generation and Understanding. In *Computer Vision and Pattern Recognition*.

Liang, Z.; Yuan, Y.; Gu, S.; Chen, B.; Hang, T.; Cheng, M.; Li, J.; and Zheng, L. 2025. Aesthetic Post-Training Diffusion Models from Generic Preferences with Step-by-step Preference Optimization. In *Computer Vision and Pattern Recognition*.

Lipman, Y.; Chen, R. T. Q.; Ben-Hamu, H.; Nickel, M.; and Le, M. 2023. Flow Matching for Generative Modeling. *ArXiv*.

Liu, J.; Liu, G.; Liang, J.; Li, Y.; Liu, J.; Wang, X.; Wan, P.; Zhang, D.; and Ouyang, W. 2025. Flow-GRPO: Training Flow Matching Models via Online RL. *ArXiv*.

Liu, X.; Gong, C.; and Liu, Q. 2022. Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow. *ArXiv*.

Loshchilov, I.; and Hutter, F. 2017. Decoupled Weight Decay Regularization. *ArXiv*.

Lu, Y.; Wang, Q.; Cao, H.; Wang, X.; Xu, X.; and Zhang, M. 2025. InPO: Inversion Preference Optimization with Reparametrized DDIM for Efficient Diffusion Model Alignment. In *Computer Vision and Pattern Recognition*.

Martin, S.; Gagneux, A.; Hagemann, P.; and Steidl, G. 2024. PnP-Flow: Plug-and-Play Image Restoration with Flow Matching. *ArXiv*.

Ni, H.; Shi, C.; Li, K.; Huang, S. X.; and Min, M. R. 2023. Conditional Image-to-Video Generation with Latent Flow Diffusion Models. In *Computer Vision and Pattern Recognition*.

Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. 2021. Learning Transferable Visual Models From Natural Language Supervision. In *International Conference on Machine Learning*.

Rafailov, R.; Sharma, A.; Mitchell, E.; Manning, C. D.; Ermon, S.; and Finn, C. 2023. Direct Preference Optimization: Your Language Model is Secretly a Reward Model. *Neural Information Processing Systems*.

Samaddar, A.; Sun, Y.; Nilsson, V.; and Madireddy, S. 2025. Efficient Flow Matching using Latent Variables. *ArXiv*.

Schuhmann, C.; et al. 2022. LAION-Aesthetics v2+ Dataset. https://github.com/christophschuhmann/improved-aesthetic-predictor.

Shen, F.; Jiang, X.; He, X.; Ye, H.; Wang, C.; Du, X.; Li, Z.; and Tang, J. 2025a. Imagdressing-v1: Customizable Virtual Dressing. In *AAAI Conference on Artificial Intelligence*.

Shen, F.; and Tang, J. 2024. Imagpose: A unified Conditional Framework for Pose-guided Person Generation. *Neural Information Processing Systems*.

Shen, F.; Yu, J.; Wang, C.; Jiang, X.; Du, X.; and Tang, J. 2025b. IMAGGarment-1: Fine-Grained Garment Generation for Controllable Fashion Design. *ArXiv*.

Song, J.; Meng, C.; and Ermon, S. 2020. Denoising Diffusion Implicit Models. *ArXiv*.

Tang, H.; He, S.; and Qin, J. 2025. Connecting Giants: Synergistic Knowledge Transfer of Large Multimodal Models for Few-Shot Learning. *ArXiv*.

Wallace, B.; Dang, M.; Rafailov, R.; Zhou, L.; Lou, A.; Purushwalkam, S.; Ermon, S.; Xiong, C.; Joty, S.; and Naik, N. 2024. Diffusion Model Alignment Using Direct Preference Optimization. In *Computer Vision and Pattern Recognition*.

Wang, J.; Zhou, Y.; Zhang, X.; Bao, M.; and Yan, P. 2025a. Self-Evolutionary Large Language Models through Uncertainty-Enhanced Preference Optimization. In *AAAI Conference on Artificial Intelligence*.

Wang, S.; Wang, H.; Wei, X.; Dai, L.; and Tang, J. 2025b. PC-Diffusion: Aligning Diffusion Models with Human Preferences via Preference Classifier. ArXiv:2511.07806.

Wang, Y.; Guo, W.; Huang, R.; Huang, J.; Wang, Z.; You, F.; Li, R.; and Zhao, Z. 2024a. Frieren: Efficient Video-to-Audio Generation Network with Rectified Flow Matching. *Neural Information Processing Systems*.

Wang, Z.; Bi, B.; Pentyala, S. K.; Ramnath, K.; Chaudhuri, S.; Mehrotra, S.; Mao, X.-B.; Asur, S.; et al. 2024b. A Comprehensive Survey of LLM Alignment Techniques: RLHF, RLAIF, PPO, DPO and More. *ArXiv*.

Wu, J. Z.; Ge, Y.; Wang, X.; Lei, S. W.; Gu, Y.; Shi, Y.; Hsu, W.; Shan, Y.; Qie, X.; and Shou, M. Z. 2023a. Tune-A-Video: One-Shot Tuning of Image Diffusion Models for Text-to-Video Generation. In *International Conference on Computer Vision*.

Wu, X.; Hao, Y.; Sun, K.; Chen, Y.; Zhu, F.; Zhao, R.; and Li, H. 2023b. Human Preference Score v2: A Solid Benchmark for Evaluating Human Preferences of Text-to-Image Synthesis. *ArXiv*.

Xie, E.; Chen, J.; Zhao, Y.; Yu, J.; Zhu, L.; Wu, C.; Lin, Y.; Zhang, Z.; Li, M.; Chen, J.; et al. 2025. SANA 1.5: Efficient Scaling of Training-Time and Inference-Time Compute in Linear Diffusion Transformer. *arXiv*.

Yan, Z.; Li, Z.; He, Y.; Wang, C.; Li, K.; Li, X.; Zeng, X.; Wang, Z.; Wang, Y.; Qiao, Y.; et al. 2025. Task Preference Optimization: Improving Multimodal Large Language Models with Vision Task Alignment. In *Computer Vision and Pattern Recognition*.

Yao, J.; Yang, B.; and Wang, X. 2025. Reconstruction vs. Generation: Taming Optimization Dilemma in Latent Diffusion Models. In *Computer Vision and Pattern Recognition*.

Yu, J.; Xu, Y.; Koh, J. Y.; Luong, T.; Baid, G.; Wang, Z.; Vasudevan, V.; Ku, A.; Yang, Y.; Ayan, B. K.; et al. 2022. Scaling Autoregressive Models for Content-Rich Text-to-Image Generation. *ArXiv*.

Zhu, H.; Xiao, T.; and Honavar, V. G. 2025. DSPO: Direct Score Preference Optimization for Diffusion Model Alignment. In *International Conference on Learning Representations*.