Self-Improving Diffusion Classifiers with Minority Preference Optimization

Hyunsoo Kim*[1], Jungmyung Wi*[1], Soobin Um[2], Donghyun Kim†[1], and Suhyun Kim†[3]

[1]Korea University, [2]Kook Min University, [3]Kyung Hee University
{climba, wjm333, d_kim}@korea.ac.kr
soobin.um@kookmin.ac.kr
dr.suhyun_kim@gmail.com

Abstract. Prior studies have demonstrated that diffusion classifiers achieve robust zero-shot classification performance. However, their effectiveness is strongly tied to the pretraining data distribution: they perform well in majority, high-density regions of the data manifold, but are significantly less accurate in minority, low-density regions. Although prior works on minority sampling have focused on generating more minority-like images, what minority sampling fundamentally enables beyond generation remains underexplored. In this paper, we reveal a direct relationship between minority sampling in generation and the perception capability of diffusion classifiers. Specifically, we show that enhancing minority sampling broadens the coverage of underrepresented regions on the data manifold, thereby improving diffusion-based recognition. To exploit this connection, we propose *Self-Improving Diffusion Classifiers with Minority Preference Optimization* (MiPO), which fine-tunes a pretrained diffusion model using minority preference rewards. Using only arbitrary caption data, MiPO generates candidate samples, rewards those that better cover minority regions, and optimizes the model with LoRA and Group Relative Policy Optimization, without additional image data, external foundation models, or external reward models. This enables stable, prompt-adaptive minority sampling and translates low-density generative coverage into improved zero-shot diffusion classification. To sum up, we show that diffusion classifier perception is biased toward majority regions, demonstrate that this bias can be alleviated through minority preference optimization, and evaluate MiPO on five standard datasets.

Keywords: Diffusion Classifier · Minority Sampling · Reinforcement Learning for diffusion model

1 Introduction

Beyond the astonishing generative performance of diffusion [23, 39, 43] and flow models [16, 33, 35], a new wave of research is emerging that seeks to leverage

*These authors contributed equally to this work.
†Co-corresponding author.

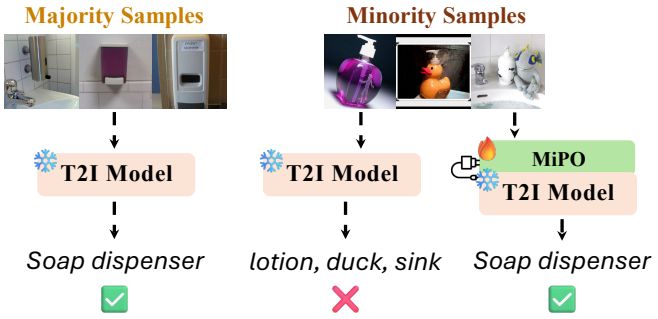

Fig. 1: MiPO expands the perceptual coverage of diffusion classifiers. A standard diffusion classifier reliably recognizes majority (*i.e.*, high-density) samples that are well covered by the pretrained generative distribution (e.g., *soap dispenser*), but often fails on underrepresented (*i.e.*, low-density) concepts such as *lotion*, *duck*, and *sink*. This illustrates that the recognition ability of diffusion classifiers is closely tied to the generative coverage of the underlying diffusion model. By optimizing minority preference rewards, MiPO encourages the model to better cover such low-density regions, thereby improving the classifier's perception of minority visual categories without requiring additional downstream image data.

these generative models to solve various understanding tasks. This trend mirrors the evolution in Natural Language Processing (NLP), where the advent and advancement of large language models (LLMs) resolved numerous downstream tasks such as machine translation [1] or sentiment analysis [37]. Similarly, recent works [30, 55] leverage large-scale vision generative model to solve vision perception tasks. These studies explore methods such as extracting useful features [45, 46], learned during the generative training process, from the intermediate layers of the models to aid in segmentation or classification, or employing the generative models themselves to solve these tasks directly [30, 55].

A prominent example within this direction is the "Diffusion Classifier" [7, 10, 30] which utilizes the generative capabilities of diffusion models to perform zero-shot classification. This approach has garnered significant attention for its distinct advantages, particularly its notable robustness against out-of-distribution (OOD) samples, reduced reliance on spurious shortcut features [31], and understanding capability of compositionality [25]. However, despite these promising properties, enhancing the performance of generative classifier has not been explored yet. To tackle this, one fundamental question is: **"When do the generative models fail at classification?"** Inspired by Richard P. Feynman's assertion, *"What I cannot create, I do not understand,"* our analysis begins by examining the model's generation failures. A widely recognized characteristic of diffusion models is their tendency to excel at generating majority samples from high-density regions, while comparatively struggling with samples from minority (low-density) regions. As illustrated in Fig. 1, the performance of the Diffusion Classifier (DC) [30] on images from minority groups is markedly lower than on

those from majority (high-density) groups, measured across various datasets. This phenomenon arises because the noise-matching loss function used to train diffusion models is inherently biased towards these majority modes. While existing research [50, 51] attempts to address this minority sampling problem by training auxiliary minority classifiers or using prompt optimization to guide the diffusion model, these methods are computationally expensive, requiring additional training or per-prompt optimization. Furthermore, they do not inherently enhance the underlying representation of the diffusion model itself, making it difficult to translate these generation improvements into better diffusion classifier performance.

Therefore, we propose a Self-improving Diffusion Classifier with ***Minority Preference Optimization*** (MiPO), an effective approach designed to enhance the ability of the diffusion model to represent and generate minority concepts. Prior minority image generation methods [49–51] primarily focus on generating diverse samples by applying classifier guidance or per-prompt optimization at inference time, without updating the diffusion model itself. As a result, these methods enhance sample diversity but leave the model's internal representation unchanged, limiting their ability to improve perception tasks such as image classification. In contrast, MiPO enables the diffusion model to self-train its minority representation using only an arbitrary set of prompts, requiring neither images nor external reward models (*e.g.*, CLIPScore [22], ImageReward [58], PickScore [26], HPS [57]) typically used for diffusion alignment. To encourage the generation of minority concepts from these prompts, we leverage online reinforcement learning (*e.g.*, GRPO [42]) to fine-tune with LoRA by computing group-wise minority reward and updating a reliable generating policy across diffusion timesteps. Furthermore, we employ Kullback–Leibler regularization to enable minority sampling without deviating excessively from the original knowledge of the diffusion model (*i.e.*, preserving majority sampling), allowing the model to generate both majority and minority samples effectively.

Leveraging a LoRA, MiPO offers plug-and-play compatibility, allowing users to attach and detach it from various diffusion backbones (*e.g.*, SD and SD2 [39]). Once trained, the adapter enables arbitrary prompt-adaptive minority generation in an efficient manner and allows users to control the strength of minority expression via a LoRA scaling value. Beyond improving minority representation, MiPO also brings practical benefits. It adds only a small number of parameters, introduces negligible inference overhead, and requires no prompt-specific optimization, making it suitable for large-scale or real-time generation scenarios. Moreover, because the adapter updates a compact, isolated parameter set rather than the full model, it avoids catastrophic shifts in the generative prior and preserves compatibility with downstream alignment or fine-tuning pipelines. This efficient and modular design contributes to the strong performance observed in diffusion-based classification tasks. MiPO demonstrates an accuracy gain of over up to 3.8% on standard benchmarks CIFAR10 [28], ImageNet [14])-Tiny as well as on out-of-distribution datasets (CIFAR10-C [21], Caltech [20], SUN09 [8]), achieving these improvements without using any additional image

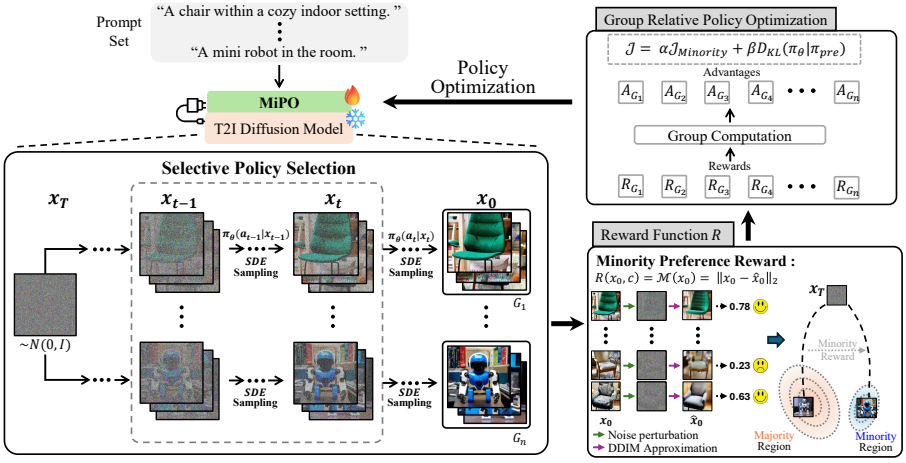

Fig. 2: Training pipeline of Minority Preference Optimization (MiPO). MiPO is integrated into a pretrained text-to-image diffusion model, where multiple SDE trajectories are sampled using the same prompt and initial noise. Only the early denoising steps are updated through *Selective Policy Selection*. For each generated sample, the *Minority Preference Reward* is obtained from DDIM-based minority scores, which reflects how strongly the sample lies in a minority region. Rewards within each prompt group are aggregated to compute group-normalized advantages via GRPO, and the final objective combines the minority reward with KL regularization to maintain consistency with the original model. MiPO guides the model toward better coverage of minority regions while preserving a stable denoising.

data and showing particularly substantial gains for minority samples. We believe these are significant findings that can be extended to other classes of generative models. Our contributions can be summarized as follows:

- We propose MiPO, a self-improving diffusion classifier that enhances minority representation without using additional training images or external reward models.
- MiPO efficiently learns minority preference through GRPO with a compact LoRA adapter, enabling prompt-adaptive minority generation.
- MiPO expands the model's perceptual coverage and achieves significant gains in zero-shot diffusion classification across datasets and backbones.

2 Related Works

Minority Sampling with Diffusion Models. The task of generating minority samples has advanced substantially with the advent of diffusion models, owing to their ability to faithfully capture complex data distributions [41,48–51]. Pioneering works in this direction commonly adopt the idea of classifier guidance [15], where separately trained classifiers are employed to steer the generative process

toward low-density regions [41,49]. Subsequent studies [48,50,51] have addressed practical limitations, such as the dependence on external classifiers [50], and have enabled practical minority sample generation that can be performed solely using a pretrained diffusion model across diverse scenarios (*e.g.*, text-to-image generation [51]). However, these methods often incur substantial computational overhead during inference due to backpropagation-based guidance for minority representations [49–51]. While the approach in [48] has addressed the complexity issue by developing a guidance-free minority generator, it is limited to stochastic samplers (like DDPM [23]) and not applicable to deterministic ODE solvers (*e.g.*, DDIM [44]) that are popularly used in modern diffusion-based frameworks. In contrast, our approach is compatible with ODE solvers and requires only marginal additional computation for LoRA construction. Moreover, unlike existing works, we are the first to demonstrate that promoting minority representations can directly enhance the performance of diffusion classifiers, a previously unexplored benefit.

Reinforcement Learning for Generative Model Alignment. Recent studies have explored aligning diffusion models with downstream objectives through reinforcement learning based fine-tuning, inspired by the success of RLHF in large language models. These methods optimize generated samples according to reward signals while keeping the updated model close to the pretrained distribution. Online RL approaches such as DDPO [3], DPOK [18], PRDP [13], AlignProp [38], and DRaFT [11] enable adaptive reward-driven updates but often suffer from instability and high computational cost, whereas offline approaches such as Diffusion-DPO [52], SPIN-Diffusion [60], and ID-Aligner [6] offer greater stability at the expense of flexibility. To mitigate the instability of reward-based optimization, DNO [47] reformulates alignment as a noise-space optimization problem, directly updating denoising trajectories without relying on policy gradients. More recently, DanceGRPO [59] and FlowGRPO [34] further stabilize RL-based diffusion alignment by introducing GRPO [42] with enabling robust and scalable policy learning across both diffusion and rectified flow models. We extend these advances by fine-tuning diffusion models to generate samples from minority regions, enhancing diffusion classifier accuracy.

Diffusion Model for Zero-shot Image Classification. Diffusion classifier [10, 30] leverages the noise prediction loss of diffusion models to perform zero-shot classification. Given that the noise-matching loss during the diffusion model's training enables the learning of rich image features, DCs not only demonstrate performance comparable to conventional classifiers but are also recognized for distinct advantages, such as robustness to OOD datasets and avoidance of spurious shortcut features [5, 31]. Furthermore, [25] revealed that DC possess an understanding of compositionality. More recently, [53] explored the role of noise in DC, investigating the concept of "good noise." However, these studies do not fundamentally explore methods for improving the intrinsic performance of the diffusion classifier itself. In this paper, we are the first to propose a simple yet effective method for enhancing the performance of the diffusion classifier by strengthening its minority generation capabilities.

3 Method

3.1 Preliminaries

Fine-tuning Diffusion Models with Reinforcement Learning. Following recent works [3, 34, 59], the denoising trajectory of a diffusion model can be formulated as a Markov Decision Process (MDP), where the policy $\pi_\theta(a_t \mid x_t) = p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{c})$ sequentially generates denoised samples conditioned on the state $s_t = (\boldsymbol{c}, t, \boldsymbol{x}_t)$. For each prompt $\boldsymbol{c}$, a group of G trajectories $\{\tau^{(g)} = (\boldsymbol{z}_T^{(g)}, \ldots, \boldsymbol{z}_0^{(g)})\}_{g=1}^G$ is sampled from the same initial noise under stochastic SDE solvers. Each trajectory produces a final output $\boldsymbol{z}_0^{(g)}$ and reward $r^{(g)} = R(\boldsymbol{z}_0^{(g)}, \boldsymbol{c})$. The group-relative advantage is then computed as

$$A^{(g)} = \frac{r^{(g)} - \text{mean}\big(\{r^{(j)}\}_{j=1}^G\big)}{\text{std}\big(\{r^{(j)}\}_{j=1}^G\big)}, \tag{1}$$

where the group corresponds to multiple SDE sampling trajectories sharing the same prompt and initialization noise. This group-level normalization ensures that policy updates remain invariant to reward scale and the stochastic variance of different SDE solvers.

The policy is then updated by maximizing the clipped GRPO objective:

$$\mathcal{J}(\theta) = \mathbb{E}\Bigg[\frac{1}{G}\sum_{i=1}^{G}\frac{1}{T}\sum_{t=1}^{T}\min\Big(\rho_{t,i}(\theta)A_t^{(i)}, \\ \text{clip}\big(\rho_{t,i}(\theta),\, 1-\epsilon,\, 1+\epsilon\big)A_t^{(i)}\Big)\Bigg]. \tag{2}$$

where $\rho_{t,i}(\theta) = \frac{\pi_\theta(a_{t,i}|x_{t,i})}{\pi_{\theta_{\text{old}}}(a_{t,i}|x_{t,i})}$ is the likelihood ratio between current and previous policies. This group-relative normalization makes the optimization invariant to the absolute scale of rewards and heterogeneous prompt distributions, enabling stable and consistent RL-based fine-tuning of diffusion models.

Diffusion Classifier. Diffusion classifier [10, 30] is a framework that applies pre-trained diffusion models to classification tasks. Since calculating the class-conditional likelihood $p(x|y)$ in diffusion models is intractable [30], DC uses the noise prediction error as a surrogate objective. This is based on the intuition that the noise prediction error will be smaller when conditioned on the correct class.The classification process is as follows: A noisy image x_t is generated from an original image x_0 and random noise ϵ. The model then predicts the conditioned noise $\hat{\epsilon}_\theta(x_t, c_i, t)$ for each possible class c_i. The final predicted class $\hat{c}$ is the one that minimizes the L2 distance between the predicted noise and the ground-truth noise ϵ.

$$\hat{c} = \arg\min_{c_i} \|\hat{\epsilon}_\theta(x_t, c_i, t) - \epsilon\|_2^2$$

However, diffusion models are known to favor majority regions because the standard noise-matching objective emphasizes reconstruction of frequently occurring samples [41, 49–51]. As a result, rare or underrepresented concepts, referred to as *minority regions*, are often neglected, leading to biased generation and degraded performance of the diffusion classifiers [10, 30].

3.2 Minority Preference Optimization (MiPO) for Diffusion Classifier

Building on the formulation in Sec. 3.1, we now describe how diffusion models can be fine-tuned to better capture underrepresented, minority regions of the data distribution. The previous works [41, 49–51] introduce inference-time guidance methods that focus solely on generating diverse or minority images without updating the model itself. In contrast, we aim to go beyond merely generating diverse samples, instead enhancing perceptual capability of diffusion model by strengthening its intrinsic understanding of minority regions. To achieve this, we fine-tune the diffusion model with GRPO [12,42], **and, unlike the prior works, we apply GRPO updates through a LoRA [24], enabling efficient and stable policy learning without modifying the full diffusion backbone.** We adopt a minority preference reward design that encourages minority sampling while preserving the stability and fidelity of the pretrained model. The overall training process is illustrated in Fig. 2, which depicts the selective policy updates, group-relative optimization, and KL regularization used in MiPO. MiPO does not require any additional real data; instead, it improves generalization by self-generating training samples and refining itself in a self-improving method with minority preference rewards.

Minority Preference Reward. Inspired by recent inference-time guidance methods for minority generation [49–51], we adapt their intuition into a training-time optimization framework. To encourage the diffusion model to generate samples from minority regions of the data distribution, we introduce a *minority preference reward* that quantifies how underrepresented a generated sample is. Given a generated image $\boldsymbol{x}_0$ conditioned on prompt $\boldsymbol{c}$, the reward is computed as $R(\boldsymbol{x}_0, \boldsymbol{c}) = \mathcal{M}(\boldsymbol{x}_0)$, where $\mathcal{M}(\cdot)$ measures the degree of minority.

We compute $\mathcal{M}(\boldsymbol{x}_0)$ using a DDIM-based approximation of the Tweedie's formula [9, 27]. Specifically, for a generated image $\boldsymbol{x}_0$, we add Gaussian noise to obtain a perturbed latent $\boldsymbol{x}_t = \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Following the common practice in a diffusion study [49], we set the perturbation strength to an intermediate diffusion step (*e.g.*, $t = 0.9T$) to ensure that the denoiser uncertainty is sufficiently exposed while preserving perceptual structure. We then reconstruct $\hat{\boldsymbol{x}}_0$ via the DDIM [44]:

$$\hat{\boldsymbol{x}}_0 = \frac{\boldsymbol{x}_t - \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\epsilon}_{pre}(\boldsymbol{x}_t, c, t)}{\sqrt{\bar{\alpha}_t}}, \tag{3}$$

The minority score is then defined as the reconstruction discrepancy between the generated sample and its DDIM-approximated posterior mean:

$$\mathcal{M}(\boldsymbol{x}_0) = \|\hat{\boldsymbol{x}}_0 - \boldsymbol{x}_0\|_2. \tag{4}$$

Intuitively, samples with higher reconstruction errors indicate regions where the denoiser exhibits greater uncertainty or weaker representation, corresponding to minority areas of the data manifold. In practice, $\mathcal{M}(\boldsymbol{x}_0)$ can also be combined with feature-space density estimation, LPIPS [61] based classifier-based likelihood proxies which further refine minority sensitivity by leveraging complementary perspectives of data rarity. This reward formulation encourages the model to explore and synthesize samples that lie in sparsely populated regions of the training distribution, leading to enhanced coverage of long-tail semantics and improved diffusion-classifier accuracy.

To integrate this objective into GRPO training, the group-normalized advantage in Eq. 1 is computed using the minority reward $R(\boldsymbol{x}_0, \boldsymbol{c}) = \mathcal{M}(\boldsymbol{x}_0)$. Hence, higher-reward (*i.e.*, minority) samples yield positive advantages, while low-reward (*i.e.*, majority) samples produce negative ones, guiding the policy toward minority-aligned generation behavior. This design ensures that the policy optimization remains prompt-adaptive and that minority exploration emerges organically from the reward signal alone.

Robust Fine-tuning with KL Regularization. While the minority reward encourages exploration in minority regions, it can also induce instability or degrade visual fidelity if the policy diverges too far from the pretrained distribution. To mitigate this issue, we incorporate a Kullback–Leibler (KL) regularization term during GRPO updates:

$$\mathcal{J}(\theta) = \alpha \mathcal{J}_{Minority}(\theta) + \beta D_{\mathrm{KL}}\big(\pi_\theta \,\|\, \pi_{\mathrm{pre}}\big), \tag{5}$$

where π_{pre} denotes the pretrained diffusion model and β controls the regularization strength. Here, α determines the contribution of the minority reward, while β regulates how closely the updated policy should remain aligned with the pretrained prior. We set $\alpha = 0.7$ and $\beta = 0.15$ to balance minority enhancement and fidelity preservation. This term penalizes excessive deviations from the reward optimization, acting as the constraint that preserves the generative prior of the base diffusion model. Empirically, KL regularization stabilizes reward-driven optimization and prevents reward hacking while enabling reliable minority sampling. Ablation results on the KL term are presented in Sec. 5.1.

3.3 Selective Policy Optimization

Although reinforcement learning can, in principle, update the policy across all denoising timesteps, recent works suggest that optimizing over a subset of timesteps is sufficient for effective diffusion fine-tuning. DanceGRPO [59] demonstrated that stable policy learning can be achieved even when updates are applied to

only a fraction of the diffusion steps, showing that the policy does not need to be optimized across the entire trajectory to obtain performance gains. Similarly, [4] introduced a *stop guidance* mechanism, highlighting that early-stage timesteps play a dominant role in determining sample quality and that applying guidance at later stages provides diminishing returns.

Motivated by these findings, we update the policy only for the initial 40% of diffusion timesteps. This selective optimization not only reduces computational cost but also focuses learning on the early denoising phase, where coarse semantic and structural information is formed. Empirically, we find that restricting updates to these earlier timesteps preserves visual fidelity while maintaining strong reward optimization performance, demonstrating that fine-tuning need not encompass the entire diffusion trajectory to yield effective alignment. Ablation results for our selective policy optimization strategy are presented in Sec. 5.1.

4 Experiments

We evaluate our approach through three complementary perspectives: (i) zero-shot diffusion classification across multiple datasets, (ii) ablation studies that analyze the effect of KL regularization and timestep selection, and (iii) minority image generation to assess whether the learned policy captures minority visual concepts. This structure provides a comprehensive understanding of how minority policy optimization improves both the perceptual ability of the diffusion model and its capacity to represent minority samples.

4.1 Experimental Setup

Training Datasets. We employed captions from HPSv2 [56] for minority policy optimization without any images. The HPSv2 caption set is derived from DiffusionDB [54] and MSCOCO [32], consisting of over 100k cleaned and de-biased text prompts that were curated for large-scale human preference evaluation. It is noteworthy that our approach requires only arbitrary caption data without needing additional images.

Evaluation Datasets. We utilized the CIFAR10 [28] and Tiny-ImageNet [29] validation sets to measure the accuracy of the diffusion classifier. Furthermore, we evaluated our method on robustness benchmarks, specifically CIFAR10-C [21], and Caltech [20], LabelMe [40], SUN09 [8], and PASCAL VOC [17] in VLCS [19] to assess performance against corruptions and perturbations. Due to the significant computational overhead of diffusion classifier evaluation on CIFAR10-C and Tiny-Imagenet, we perform measurements on a random subset of 2,000 images. For the generation and evaluation of minority images, we randomly sampled 1,000 captions from the MS-COCO [32] validation set. In addition, to intuitively demonstrate how the diffusion model understands majority and minority samples, we further divide the CIFAR10 dataset into two groups of 1,000 images each, with one representing the majority (*i.e.*, majority group) and the other

Table 1: Zero-shot classification accuracy of diffusion classifiers on SD 1.5 and SD 2.0. MiPO improves performance across most datasets for both SD 1.5 and SD 2.0, notably achieving gains of about 2.0% on CIFAR10-C and 1.2% on ImageNet-Tiny. While CIFAR10 under SD 2.0 shows a slight decrease, the overall trend demonstrates that minority-aware fine-tuning generally enhances robustness and extends perceptual coverage of the diffusion models across diverse domains.

Model	CIFAR10	CIFAR10-C	ImageNet-Tiny	Caltech	SUN09
SD1.5	85.38	57.72	46.50	97.81	64.69
MiPO	**87.89**	**59.76**	**47.30**	**99.51**	**68.25**
SD2.0	**88.58**	64.61	51.45	99.65	64.72
MiPO	86.98	**65.33**	**55.25**	**99.93**	**69.26**

Table 2: Effect of the KL regularization. Introducing KL regularization leads to consistently better performance for both minority and majority groups, indicating that constraining the policy to stay close to the pretrained diffusion prior stabilizes MiPO.

	Minority group	**Majority group**
SD1.5	73.07	79.18
w/o KL	78.58	82.08
w/ KL (Ours)	**80.58**	**82.68**

representing the minority (*i.e.*, minority group), and evaluate the classifier performance separately on both groups. A detailed description of the evaluation dataset and additional dataset configurations is provided in the Appendix.

Baselines and Evaluation Metrics. To the best of our knowledge, this work is the first to improve the diffusion classifier in a self-improving method without using any additional image data such as few-shot examples. Hence, we adopt the original diffusion classifier as the main baseline. Ablation studies on the proposed components (*e.g.*, minority reward, KL regularization, and selective policy optimization) are presented in Sec. 5. Further analyses on the reinforcement learning design choices and additional ablations are provided in the Appendix. For minority image generation, we compare our method with two baselines: vanilla Stable Diffusion and Minority Prompt, which represents the current state of the art in minority sampling.

For evaluation metrics, we follow the standard diffusion classifier protocol and report zero-shot classification accuracy for each dataset. For minority image generation, we assess the perceptual and preference alignment of generated images using CLIPScore [22], ImageReward [58], HPSv2 [56], and HPSv3 [36]. To quantify the degree of minority representation in the generated images, we additionally measure Minority Score [49] and JEPA-SCORE [2].

Table 3: Ablation study on time step selection. We compare the time step selection strategies, Full (0–50), Late (30–50), Middle (15–35), and Early (0–20, ours). Updating the early timesteps results in the most effective policy optimization and achieves the best overall performance.

	Minority group	**Majority group**
SD1.5	73.07	79.18
Full (0–50)	<u>80.58</u>	**82.68**
Late (30–50)	73.77	78.18
Middle (15–35)	79.90	78.28
Early (0–20, Ours)	**81.68**	<u>82.48</u>

Table 4: Quantitative comparison of minority-oriented image generation on SD 1.5. *Minority Gen* indicates whether a method can generate minority samples; *DC Eval* denotes compatibility with diffusion-classifier evaluation; *Prompt-adaptive* reflects whether the method works without per-prompt optimization at inference. Across all quantitative metrics, our MiPO approach achieves the second-best performance while running at the same speed as DDIM (1s per image), unlike MinorityPrompt, which requires heavy prompt-specific optimization (63s). These results show that MiPO enables efficient minority generation while maintaining strong perceptual alignment.

Method	Minority Gen.	DC Eval.	Prompt-adaptive	CLIPScore ↑	HPSv3 ↑	Minority Score ↑	JEPA-SCORE ↓	Time(s) ↓
DDIM [44]	✗	✓	✓	**31.7571**	**5.7581**	0.2132	-568.7979	**1**
MinorityPrompt [51]	✓	✗	✗	30.4351	1.5324	**0.2311**	**-647.9724**	63
Ours (DDIM + MiPO)	✓	✓	✓	<u>30.8026</u>	<u>4.4090</u>	<u>0.2289</u>	<u>-646.0296</u>	**1**

4.2 Zero-shot Classification Results

Our proposed method demonstrates consistent performance improvements in diffusion classifier accuracy across all primary evaluation datasets. Notably, these gains are achieved without using any additional image data, relying solely on arbitrary captions (HPSv2) and the minority preference as a reward signal. As summarized in Table 1, MiPO consistently improves zero-shot classification accuracy across both SD 1.5 and SD 2.0. For SD 1.5, all datasets exhibit clear improvements, while for SD 2.0 we observe gains of about 2% on CIFAR10-C and roughly 1.2% on ImageNet-Tiny, with Caltech and SUN09 showing similar upward trends. Although CIFAR10 under SD 2.0 shows a slight decrease, the overall pattern still indicates that MiPO enhances the perceptual coverage of diffusion models and generalizes effectively across diverse domains.

Interestingly, we also observe performance degradation on datasets such as LabelME and VOC2007 (Table 6). These datasets contain noisy, multi-object, or inconsistently annotated images, making them substantially harder and inherently unstable for zero-shot diffusion classification. Since they are not strictly clean benchmarks, we provide a detailed analysis and dataset-specific denoising procedures in the Appendix.

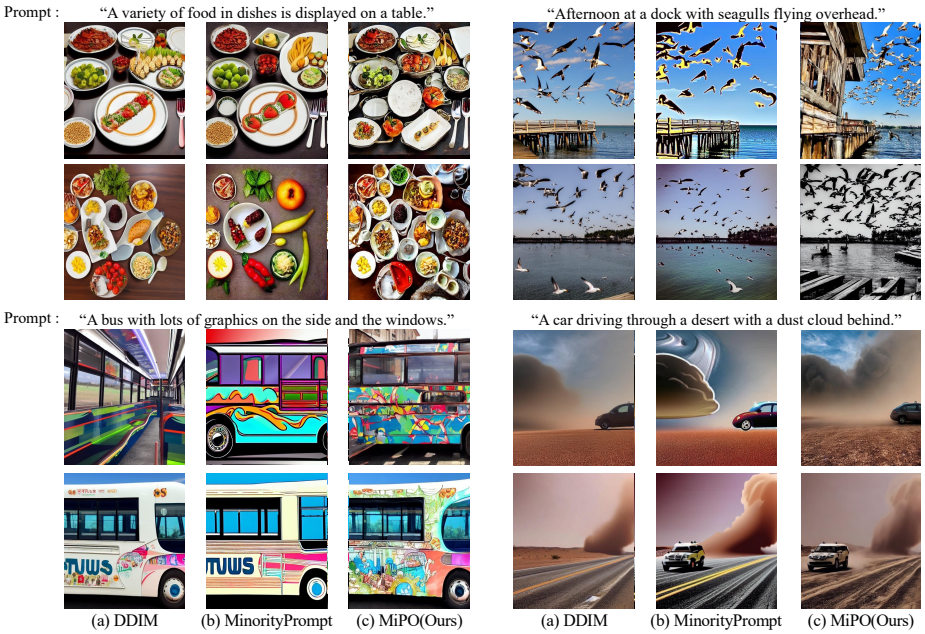

Fig. 3: Qualitative results of minority samples. Across diverse MSCOCO validation prompts, MiPO produces images that exhibit richer minority details and greater concept diversity compared to DDIM, while remaining competitive with the prompt-optimized MinorityPrompt baseline. Unlike MinorityPrompt, MiPO achieves these results in a fully prompt-adaptive method without per-prompt optimization. **Images are not cherry-picked.**

5 Further Analyses

5.1 Ablation Study

To verify whether enhancing minority representation genuinely improves diffusion classifier performance, we conduct ablation studies using the CIFAR-10 dataset by dividing it into minority and majority groups. For a fair comparison, we define these groups using a k-Nearest Neighbors (KNN)–based density estimator rather than the minority score used during training. This setup ensures that improvements are not an artifact of our reward function but reflect real gains in representation quality. Furthermore, prior minority sampling approaches [49,51] have not been shown to improve diffusion classifier performance directly, highlighting the distinct advantage of our self-improving minority preference optimization framework.

Effect of KL Regularization. Table 2 reports the impact of KL regularization on accuracy. Introducing the KL term yields consistent improvements in both minority and majority groups (approximately +2% and +0.6%, respectively).

This demonstrates that KL regularization plays a crucial stabilizing role during minorit policy optimization by preventing the policy from drifting too far from the pretrained diffusion prior. By constraining the update magnitude, the model is able to explore minority regions while preserving majority representations and overall denoising fidelity. Without the KL constraint, the policy tends to overfit high-reward states, leading to instability and degraded consistency.

Effect of Timestep Selection. We further examine how the choice of diffusion timesteps affects fine-tuning performance. As shown in Table 3, among the four configurations – Full (0–50), Late (30–50), Middle (15–35), and Early (0–20) – updating only the early timesteps (0–20) yields the best performance for both minority and majority groups. This finding is consistent with prior work [4, 59], which similarly highlights that early denoising steps play a central role in effective policy optimization. Concentrating policy updates at early stages enables effective learning of meaningful semantic features without disrupting structural coherence. In contrast, applying updates predominantly at later stages leads to weaker gains or even degradation, likely due to overfitting on local pixel statistics rather than enhancing semantic diversity.

5.2 Evaluation on Minority Sample Generation

Although our primary objective is to enhance diffusion classifier performance, a natural question remains: *Does the LoRA truly learn minority concepts?* To verify this, we attach our Minority LoRA to the diffusion model, generate images using MSCOCO captions, and compare both qualitative and quantitative results with the state-of-the-art MinorityPrompt baseline.

Qualitative Results on Minority Sample Generation. As illustrated in Fig. 3, MiPO enables the diffusion model to generate images that express minority characteristics more clearly and consistently than the DDIM baseline. Across diverse MSCOCO prompts, MiPO produces richer details, more diverse local structures, and visually coherent patterns that reflect minority concepts, demonstrating that the model has genuinely internalized minority-oriented generation rather than relying on prompt engineering.

Quantitative Results on Minority Sample Generation. Table 4 summarizes the quantitative comparison on SD 1.5. MiPO improves both perceptual-quality metrics (e.g., CLIPScore, HPSv3) and minority-sensitive metrics (Minority Score, JEPA) over the DDIM baseline, indicating stronger minority representation while maintaining high visual fidelity. Importantly, MiPO is fully *prompt-adaptive* and does not require any per-prompt optimization, unlike MinorityPrompt [51], which optimizes a special token for each input prompt. Despite this difference, MiPO achieves comparable minority-sensitive scores, demonstrating that our minority policy optimization effectively generalizes across prompts without additional optimization.

Table 5: Human preference evaluation on minority samples. We compare the baseline DDIM sampler, MinorityPrompt, and our MiPO using ImageReward and HPSv2. While MinorityPrompt promotes minority-oriented generation, it noticeably lowers human preference scores. In contrast, MiPO achieves the best ImageReward score and remains close to the baseline on HPSv2, suggesting a better balance between minority-oriented generation and perceptual quality.

Method	ImageReward ↑	HPSv2 ↑
Baseline	<u>0.184</u>	**0.254**
MinorityPrompt	-0.026	0.241
Ours (MiPO)	**0.228**	<u>0.252</u>

Table 6: Zero-shot classification accuracy on LabelME and VOC2007. Due to domain gap (*e.g.*, multi objectes) or inconsistent annotations in these datasets, improvements are less stable compared to clean benchmarks. A detailed analysis is provided in the Appendix.

Model	LabelME	VOC2007
SD 1.5	**62.90**	**81.84**
Ours	55.28	76.81
SD 2.0	**63.84**	**82.38**
Ours	58.79	80.78

5.3 Human Preference Evaluation

We further evaluate human preference using ImageReward [58] and HPSv2 [56]. Table 5 compares images generated by the baseline DDIM sampler, MinorityPrompt, and MiPO. While minority-oriented generation can improve coverage of low-density regions, it may also introduce a trade-off: aggressively pushing samples toward minority regions can reduce visual fidelity or human preference alignment. This trend is reflected in the results, where MinorityPrompt produces minority-biased samples but obtains lower ImageReward and HPSv2 scores.

In contrast, MiPO maintains stronger human preference scores while promoting minority-oriented generation. Notably, MiPO achieves the highest ImageReward score, surpassing both the baseline and MinorityPrompt, and remains very close to the baseline on HPSv2. These results suggest that MiPO does not simply increase minority characteristics at the cost of perceptual quality; rather, it provides a better balance between minority-oriented generation and human-preferred visual quality.

5.4 Limitations

While MiPO consistently improves zero-shot diffusion classification across most datasets, it does not yield gains uniformly. Because our method relies solely on caption-based self-improving policy optimization without access to additional

image data, certain datasets exhibit weaker or unstable improvements. As shown in Table 6, datasets such as LabelME and VOC2007 contain a number of multi-object or ambiguous scenes, which introduce uncertainty in zero-shot evaluation. A detailed analysis of these cases is provided in the Appendix.

Another limitation arises in minority-oriented image generation. Although our policy effectively increases minority representation, it introduces an inherent trade-off between minority distinctiveness and overall perceptual fidelity. When the optimization places stronger emphasis on minority reward, the model may generate images that better reflect minority characteristics but drift slightly from the high-level aesthetic or semantic alignment of the pretrained diffusion prior. Conversely, enforcing stronger regularization preserves perceptual quality but limits the strength of minority expression. Balancing this trade-off remains an open problem and a promising direction for future work.

6 Conclusion

In this paper, we propose self-improving diffusion classifiers with Minority Preference Optimization (MiPO), which enhances the performance of diffusion classifiers by self-generating training samples guided by minority-preference rewards. To achieve stable and prompt-adaptive optimization, we fine-tune the diffusion model with LoRA using Group Relative Policy Optimization (GRPO), allowing effective policy updates across diverse prompts. As demonstrated on five standard benchmarks, MiPO successfully improves overall diffusion classifier performance without requiring any external model or image data. In summary, we show that the perception capability of diffusion classifiers is inherently focused on majority groups and demonstrate that enhancing minority group representation can improve the model's overall recognition capability.

Acknowledgements

This research was partly supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grants funded by the Korea government (MSIT) (IITP-2026-RS-2026-25543726, Leading Generative AI Human Resources Development, 23%; No. IITP-2026-RS-2023-00258649, Information Technology Research Center (ITRC), 17%; No. RS-2019-II190079, Artificial Intelligence Graduate School Program (Korea University), 1%), the Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency (KOCCA) grant funded by the Ministry of Culture, Sports and Tourism in 2024 (Project Name: International Collaborative Research and Global Talent Development for the Development of Copyright Management and Protection Technologies for Generative AI, Project Number: RS-2024-00345025, 10%), and the National Research Foundation of Korea (NRF) grants funded by the Korea government (MSIT) (No. RS-2025-00562437, 10%; No. RS-2024-00341514, 20%; No. RS-2025-02263628, 19%).

References

1. Bahdanau, D.: Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014)
2. Balestriero, R., Ballas, N., Rabbat, M., LeCun, Y.: Gaussian embeddings: How jepas secretly learn your data density. arXiv preprint arXiv:2510.05949 (2025)
3. Black, K., Janner, M., Du, Y., Kostrikov, I., Levine, S.: Training diffusion models with reinforcement learning. arXiv preprint arXiv:2305.13301 (2023)
4. Chan-Santiago, J.A., Tirupattur, P., Nayak, G.K., Liu, G., Shah, M.: Mgd3: Mode-guided dataset distillation using diffusion models. arXiv preprint arXiv:2505.18963 (2025)
5. Chen, H., Dong, Y., Wang, Z., Yang, X., Duan, C., Su, H., Zhu, J.: Robust classification via a single diffusion model. arXiv preprint arXiv:2305.15241 (2023)
6. Chen, W., Zhang, J., Wu, J., Wu, H., Xiao, X., Lin, L.: Id-aligner: Enhancing identity-preserving text-to-image generation with reward feedback learning, 2024. URL https://arxiv. org/abs/2404.15449
7. Chen, Y.C., Inouye, D.I., Gao, J.: Your var model is secretly an efficient and explainable generative classifier. arXiv preprint arXiv:2510.12060 (2025)
8. Choi, M.J., Lim, J.J., Torralba, A., Willsky, A.S.: Exploiting hierarchical context on a large database of object categories. In: 2010 IEEE computer society conference on computer vision and pattern recognition. pp. 129–136. IEEE (2010)
9. Chung, H., Kim, J., Mccann, M.T., Klasky, M.L., Ye, J.C.: Diffusion posterior sampling for general noisy inverse problems. arXiv preprint arXiv:2209.14687 (2022)
10. Clark, K., Jaini, P.: Text-to-image diffusion models are zero shot classifiers. Advances in Neural Information Processing Systems **36**, 58921–58937 (2023)
11. Clark, K., Vicol, P., Swersky, K., Fleet, D.J.: Directly fine-tuning diffusion models on differentiable rewards. arXiv preprint arXiv:2309.17400 (2023)
12. DeepSeek-AI: Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning (2025), `https://arxiv.org/abs/2501.12948`
13. Deng, F., Wang, Q., Wei, W., Hou, T., Grundmann, M.: Prdp: Proximal reward difference prediction for large-scale reward finetuning of diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7423–7433 (2024)
14. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
15. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. Advances in neural information processing systems **34**, 8780–8794 (2021)
16. Esser, P., Kulal, S., Blattmann, A., Entezari, R., Müller, J., Saini, H., Levi, Y., Lorenz, D., Sauer, A., Boesel, F., et al.: Scaling rectified flow transformers for high-resolution image synthesis. In: Forty-first international conference on machine learning (2024)
17. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. International journal of computer vision **88**(2), 303–338 (2010)
18. Fan, Y., Watkins, O., Du, Y., Liu, H., Ryu, M., Boutilier, C., Abbeel, P., Ghavamzadeh, M., Lee, K., Lee, K.: Dpok: Reinforcement learning for fine-tuning text-to-image diffusion models. Advances in Neural Information Processing Systems **36**, 79858–79885 (2023)

19. Fang, C., Xu, Y., Rockmore, D.N.: Unbiased metric learning: On the utilization of multiple datasets and web images for softening bias. In: Proceedings of the IEEE international conference on computer vision. pp. 1657–1664 (2013)
20. Fei-Fei, L., Fergus, R., Perona, P.: Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. In: 2004 conference on computer vision and pattern recognition workshop. pp. 178–178. IEEE (2004)
21. Hendrycks, D., Dietterich, T.: Benchmarking neural network robustness to common corruptions and perturbations. Proceedings of the International Conference on Learning Representations (2019)
22. Hessel, J., Holtzman, A., Forbes, M., Le Bras, R., Choi, Y.: Clipscore: A reference-free evaluation metric for image captioning. In: Proceedings of the 2021 conference on empirical methods in natural language processing. pp. 7514–7528 (2021)
23. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
24. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al.: Lora: Low-rank adaptation of large language models. ICLR **1**(2), 3 (2022)
25. Jeong, Y., Uselis, A., Oh, S.J., Rohrbach, A.: Diffusion classifiers understand compositionality, but conditions apply. arXiv preprint arXiv:2505.17955 **2** (2025)
26. Kirstain, Y., Polyak, A., Singer, U., Matiana, S., Penna, J., Levy, O.: Pick-a-pic: An open dataset of user preferences for text-to-image generation. Advances in neural information processing systems **36**, 36652–36663 (2023)
27. Kotz, S., Johnson, N.L.: Breakthroughs in Statistics: Foundations and basic theory. Springer Science & Business Media (2012)
28. Krizhevsky, A., Hinton, G., et al.: Learning multiple layers of features from tiny images (2009)
29. Le, Y., Yang, X.: Tiny imagenet visual recognition challenge. CS 231N **7**(7), 3 (2015)
30. Li, A.C., Prabhudesai, M., Duggal, S., Brown, E., Pathak, D.: Your diffusion model is secretly a zero-shot classifier. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2206–2217 (2023)
31. Li, A.C., Kumar, A., Pathak, D.: Generative classifiers avoid shortcut solutions. In: The Thirteenth International Conference on Learning Representations (2024)
32. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: European conference on computer vision. pp. 740–755. Springer (2014)
33. Lipman, Y., Chen, R.T., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. arXiv preprint arXiv:2210.02747 (2022)
34. Liu, J., Liu, G., Liang, J., Li, Y., Liu, J., Wang, X., Wan, P., Zhang, D., Ouyang, W.: Flow-grpo: Training flow matching models via online rl. arXiv preprint arXiv:2505.05470 (2025)
35. Liu, X., Gong, C., Liu, Q.: Flow straight and fast: Learning to generate and transfer data with rectified flow. arXiv preprint arXiv:2209.03003 (2022)
36. Ma, Y., Wu, X., Sun, K., Li, H.: Hpsv3: Towards wide-spectrum human preference score. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 15086–15095 (2025)
37. Medhat, W., Hassan, A., Korashy, H.: Sentiment analysis algorithms and applications: A survey. Ain Shams engineering journal **5**(4), 1093–1113 (2014)
38. Prabhudesai, M., Goyal, A., Pathak, D., Fragkiadaki, K.: Aligning text-to-image diffusion models with reward backpropagation (2023)

39. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10684–10695 (2022)
40. Russell, B.C., Torralba, A., Murphy, K.P., Freeman, W.T.: Labelme: a database and web-based tool for image annotation. International journal of computer vision **77**(1), 157–173 (2008)
41. Sehwag, V., Hazirbas, C., Gordo, A., Ozgenel, F., Canton, C.: Generating high fidelity data from low-density regions using diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11492–11501 (2022)
42. Shao, Z., Wang, P., Zhu, Q., Xu, R., Song, J., Bi, X., Zhang, H., Zhang, M., Li, Y., Wu, Y., et al.: Deepseekmath: Pushing the limits of mathematical reasoning in open language models. arXiv preprint arXiv:2402.03300 (2024)
43. Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In: International conference on machine learning. pp. 2256–2265. pmlr (2015)
44. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020)
45. Stracke, N., Baumann, S.A., Bauer, K., Fundel, F., Ommer, B.: Cleandift: Diffusion features without noise. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 117–127 (2025)
46. Tang, L., Jia, M., Wang, Q., Phoo, C.P., Hariharan, B.: Emergent correspondence from image diffusion. Advances in Neural Information Processing Systems **36**, 1363–1389 (2023)
47. Tang, Z., Peng, J., Tang, J., Hong, M., Wang, F., Chang, T.H.: Inference-time alignment of diffusion models with direct noise optimization. arXiv preprint arXiv:2405.18881 (2024)
48. Um, S., Kim, B., Ye, J.C.: Boost-and-skip: A simple guidance-free diffusion for minority generation. In: Forty-second International Conference on Machine Learning (2025), https://openreview.net/forum?id=IH8OwjOGzM
49. Um, S., Lee, S., Ye, J.C.: Don't play favorites: Minority guidance for diffusion models. arXiv preprint arXiv:2301.12334 (2023)
50. Um, S., Ye, J.C.: Self-guided generation of minority samples using diffusion models. In: European Conference on Computer Vision. pp. 414–430. Springer (2024)
51. Um, S., Ye, J.C.: Minority-focused text-to-image generation via prompt optimization. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 20926–20936 (2025)
52. Wallace, B., Dang, M., Rafailov, R., Zhou, L., Lou, A., Purushwalkam, S., Ermon, S., Xiong, C., Joty, S., Naik, N.: Diffusion model alignment using direct preference optimization. pp. 8228–8238 (2024)
53. Wang, Y., Chen, L.: Noise matters: Optimizing matching noise for diffusion classifiers. arXiv preprint arXiv:2508.11330 (2025)
54. Wang, Z.J., Montoya, E., Munechika, D., Yang, H., Hoover, B., Chau, D.H.: Diffusiondb: A large-scale prompt gallery dataset for text-to-image generative models. In: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 893–911 (2023)
55. Wiedemer, T., Li, Y., Vicol, P., Gu, S.S., Matarese, N., Swersky, K., Kim, B., Jaini, P., Geirhos, R.: Video models are zero-shot learners and reasoners. arXiv preprint arXiv:2509.20328 (2025)

56. Wu, X., Hao, Y., Sun, K., Chen, Y., Zhu, F., Zhao, R., Li, H.: Human preference score v2: A solid benchmark for evaluating human preferences of text-to-image synthesis. arXiv preprint arXiv:2306.09341 (2023)
57. Wu, X., Sun, K., Zhu, F., Zhao, R., Li, H.: Better aligning text-to-image models with human preference. arXiv preprint arXiv:2303.14420 **1**(3) (2023)
58. Xu, J., Liu, X., Wu, Y., Tong, Y., Li, Q., Ding, M., Tang, J., Dong, Y.: Imagereward: Learning and evaluating human preferences for text-to-image generation. Advances in Neural Information Processing Systems **36**, 15903–15935 (2023)
59. Xue, Z., Wu, J., Gao, Y., Kong, F., Zhu, L., Chen, M., Liu, Z., Liu, W., Guo, Q., Huang, W., et al.: Dancegrpo: Unleashing grpo on visual generation. arXiv preprint arXiv:2505.07818 (2025)
60. Yuan, H., Chen, Z., Ji, K., Gu, Q.: Self-play fine-tuning of diffusion models for text-to-image generation. Advances in Neural Information Processing Systems **37**, 73366–73398 (2024)
61. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 586–595 (2018)