Generalizing Alignment Paradigm of Text-to-Image Generation with Preferences Through f-Divergence Minimization

Haoyuan Sun, Bo Xia, Yongzhe Chang*, Xueqian Wang*

Tsinghua Shenzhen International Graduate School, Tsinghua University
{sun-hy23, xiab21}@mails.tsinghua.edu.cn; {changyongzhe, wang.xq}@sz.tsinghua.edu.cn

Abstract

Direct Preference Optimization (DPO) has recently expanded its successful application from aligning large language models (LLMs) to aligning text-to-image models with human preferences, which has generated considerable interest within the community. However, we have observed that these approaches rely solely on minimizing the reverse Kullback-Leibler divergence during alignment process between the fine-tuned model and the reference model, neglecting incorporation of other divergence constraints. In this study, we focus on extending reverse Kullback-Leibler divergence in the alignment paradigm of text-to-image models to f-divergence, which aims to garner better alignment performance as well as good generation diversity. We provide the generalized formula of text-to-image alignment paradigm under f-divergence condition and thoroughly analyze the impact of different divergence constraints on alignment process from the perspective of gradient fields. We conduct comprehensive evaluation on text-image alignment performance, human value alignment performance and generation diversity performance under different divergence constraints, and the results indicate that text-to-image alignment based on *Jensen-Shannon divergence* achieves the best trade-off among them. The option of divergence employed for aligning text-to-image models significantly impacts the trade-off between alignment performance (especially human value alignment) and generation diversity, which highlights the necessity of selecting an appropriate divergence for practical applications.

Introduction

Text-to-image generative models have seen significant advances in recent years (Podell et al. 2024; Li et al. 2024b). When presented with appropriate textual prompts, they are capable of generating high-fidelity images that are semantically coherent with the provided descriptions (Zhu et al. 2024; Feng et al. 2024b), which spans a diverse range of topics, piquing significant public interest in their potential applications and societal implications. Existing self-supervised pre-trained generators, although advanced, still exhibit imperfections, with a significant challenge being their alignment with human preferences.

Reinforcement Learning from Human Feedback (RLHF) has established itself as a pivotal research endeavor, demonstrating notable efficacy in aligning text-to-image models with human preferences (Kirstain et al. 2023; Xu et al. 2024; Black et al. 2024). Faced with the intricate challenge of defining an objective that authentically encapsulates human preferences in the realm of Reinforcement Learning from Human Feedback (RLHF), researchers conventionally assemble a dataset to mirror such preferences through comparative assessments of model-generated outputs (Kirstain et al. 2023; Wu et al. 2023). Then, a reward model is trained based on the Bradley-Terry model (Bradley and Terry 1952), inferring human preferences from the collected dataset. And the text-to-image model is fine-tuned with a reinforcement learning (RL) pipeline. It is noteworthy that such process is conducted while ensuring the model remains closely with its original form, which is achieved by employing a *reverse Kullback-Leibler* divergence penalty. Significant complexity has been introduced to the RLHF pipeline due to the requirement to train a separate reward model. Moreover, Reinforcement learning pipelines also present notable challenges in terms of stability and memory demands towards the alignment process of text-to-image models.

Recent research has shown significant success in fine-tuning large language models (LLMs) using methods based on implicit rewards, specially the Direct Preference Optimization (DPO) (Rafailov et al. 2023). Application of similar fine-tuning techniques to text-to-image models has also produced promising results, such as Diffusion-DPO (Wallace et al. 2024), D3PO (Yang et al. 2024). Such results have raised significant interest within the community regarding the alignment of text-to-image models with human value through the methodology of utilizing implicit rewards. Furthermore, researchers have devoted significant efforts to applying such paradigm of aligning human value to text-to-image models, including SPO (Liang et al. 2024), NCPPO (Gambashidze et al. 2024), DNO(Tang et al. 2024), and so on. However, it is the situation that existing research of text-to-image generation alignment predominantly targets solutions subject to the constraint of the *reverse Kullback-Leibler* divergence, with notable underexploitation of strategies that integrate other types of divergences.

It has been pointed out that models would overfit due to repeated fine-tuning on a few images, leading to reduced

*Corresponding Authors

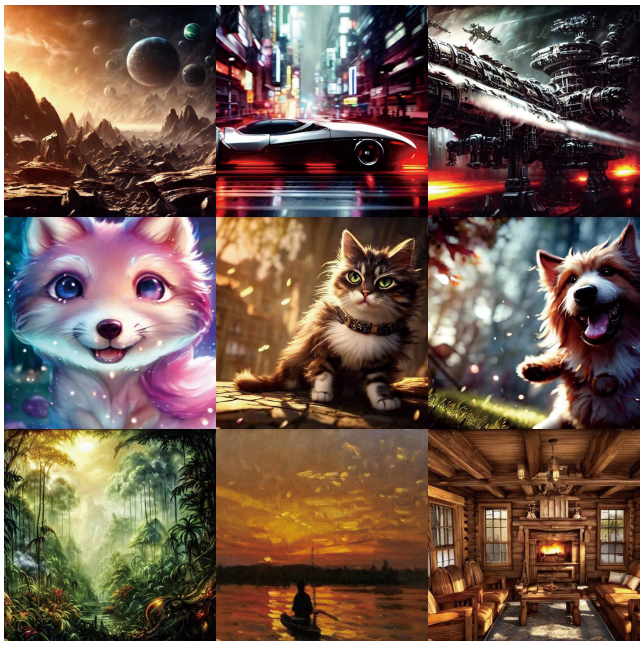

Figure 1: Examples of image generated by the model aligned using the Jensen-Shannon divergence constraint.

output diversity (Ruiz et al. 2023). In the alignment of large language models, similar challenges exist; and some studies (Wiher, Meister, and Cotterell 2022; Perez et al. 2022) have highlighted that the mode-seeking property of reverse KL divergence tends to reduce diversity in generated outputs, which can constrain the model's potential and users' engagement. Studies on aligning large language models (Go et al. 2023; Wang et al. 2024) indicate that the problem of diversity reduction caused by fine-tuning can be alleviated by incorporating diverse divergence constraints. Therefore, in this study, we also focus on exploring effects of employing diverse divergence constraints on the generation diversity.

In this study, we generalize the alignment of text-to-image models based on *reverse Kullback-Leibler* divergence to a framework based on *f-divergence* constraints, which encompasses a wider range of divergences, including Jensen-Shannon divergence, forward Kullback-Leibler divergence, α-divergence, and so on. We comprehensively analyze the impact of diverse divergence constraints on the alignment process from the perspective of gradient fields. Furthermore, we set Step-aware Preference Optimization (SPO) (Liang et al. 2024) as our benchmark method, utilize Stable Diffusion v1-5 (Rombach et al. 2022) as our benchmark model, and evaluate on the test split of HPS-V2 (Wu et al. 2023) with different divergence constraints. Evaluations are carried out to examine the performance of text-image alignment, human value alignment, and generation diversity, which also aim to discern the certain divergence most effectively balances these three aspects. Our results indicate that *Jensen-Shannon divergence* successfully strikes equilibrium among the three aspects, while also achieving the highest standard in human value alignment performance. Moreover, in text-to-image alignment, careful selection of the divergence constraint, tailored to the specific alignment requirements, is paramount. In Figure 1, we show several examples generated by the model that have been aligned under the Jensen-Shannon divergence constraint.

To the best of our knowledge, this is the first work to apply different divergence constraints to text-to-image alignment paradigm. Our contributions are summarized as follows: (1) *Generalized alignment formula*: we propose a generalized formula for text-to-image generation alignment, aiming to provide more choices on divergence constraints in the alignment execution. (2) *Thorough alignment process analysis*: we comprehensively analyze the impact of different divergence constraints on alignment process from the perspective of gradient fields. (3) *Extensive alignment evaluations:* we conducted extensive evaluations on text-to-image generation alignment, meticulously assessing both alignment performance (text-image alignment and human value alignment) and generation diversity.

Related Work

Aligning Text-to-Image Model with Preferences

Recently, inspired by the alignment approaches based on human preferences, notably exemplified by methods such as direct preference optimization (DPO) (Rafailov et al. 2023), eliminating the need for explicit reward models and showing their significant success on Large Language Models (LLMs), and then garnering substantial attention within the community on the development of alignment for text-to-image diffusion models. Diffusion-DPO (Wallace et al. 2024) enables text-to-image diffusion models to directly learn from human feedback in an open-vocabulary setting, and fine-tunes them on the contains Pick-a-Pic (Kirstain et al. 2023) dataset with image preference pairs. Direct Preference for Denoising Diffusion Policy Optimization (D3PO) (Yang et al. 2024) proposes a method on generating pairs of images from the same prompt and identifying the preferred and dispreferred images with the help of human evaluators. Step-aware Preference Optimization (SPO) (Liang et al. 2024) proposes an approach that preferences at each step should be assessed and it utilizes a step-aware preference model and a step-wise resampler to ensure accurate step-aware preference alignment. DenseReward method (Yang, Chen, and Zhou 2024) proposes enhancing the DPO scheme by incorporating a temporal discounting approach, which prioritizes the initial denoising steps. Noise-Conditioned Perceptual Preference Optimization (NCPPO) (Gambashidze et al. 2024) proposes that the optimization process should aligns with human perceptual features, instead of the less informative pixel space. Direct Noise Optimization (DNO) (Tang et al. 2024) optimizes noise during the sampling process of text-to-image diffusion models. PopAlign (Li, Singh, and Grover 2024) is an approach for population-level preference optimization, mitigating the biases of pretrained text-to-image diffusion models. Diffusion-KTO (Li et al. 2024a) generalizes the human utility maximization framework to the alignment of text-to-image diffusion models. While these studies have demon-

strated impressive results in addressing the text-to-image alignment challenge, we also notice that they all rely on *reverse Kullback-Leibler divergence* to minimize the discrepancy between the fine-tuned model and the reference model.

f-Divergence Utilized in Generation Models

In previous studies, researchers have extensively examined the application of f-divergences in generative models. In the classical work done by (Goodfellow et al. 2014), the concept of Generative Adversarial Networks (GANs) and their relationship to the Jensen-Shannon divergence are introduced. f-GAN (Nowozin, Cseke, and Tomioka 2016) proposes that the variational expression of the f-divergence can be regarded as the loss function for Generative Adversarial Networks (GANs). Wasserstein-GAN (Arjovsky, Chintala, and Bottou 2017) offers theoretical insights into the connection between the choice of divergences and the convergence of probability distributions. Moreover, in the work (Theis, van den Oord, and Bethge 2016), it is proposed that utilizing various divergences can result in divergent trade-offs, and distinct evaluations tend to favor specific divergences. The application of f-divergence has also been observed in large language model alignment tasks. f-DPG (Go et al. 2023) shows that Jensen-Shannon divergence strikes a good balance between different competing objectives, and often significantly outperforming the reverse Kullback-Leibler divergence. f-DPO (Wang et al. 2024) generalizes the framework of DPO by incorporating diverse divergence constraints; and it shows that by adjusting the divergence regularization, we can achieve a better balance between alignment performance and generation diversity of large language models (LLMs).

Preliminary

f-Divergence

For any convex function $f(x) : \mathbb{R}^+ \to \mathbb{R}$ with $f(1) = 0$, and p_1, p_2 are two distributions over a discrete set $\mathcal{X}$, the f-divergence between p_1 and p_2 can be defined as (Liese and Vajda 2006):

$$D_f(p_1||p_2) = \mathbb{E}_{x\sim p_2}\left[f\left(\frac{p_1(x)}{p_2(x)}\right) + f'(\infty)p_1(p_2 = 0)\right],$$

where $f'(\infty) = \lim_{t\to 0} tf(\frac{1}{t})$ (Hiriart-Urruty and Lemaréchal 1996), $p_1(p_2 = 0)$ is p_1-mass of the set $\{x \in \mathcal{X} : p_2(x) = 0\}$. Under normal circumstances, we can make the assumption that the support set of p_1 is dominated by the support set of p_2, i.e. $Supp(p_1) \subset Supp(p_2)$, and then we can have $p_1(p_2 = 0) = 0$. Hence, the aforementioned definition can be simplified as:

$$D_f(p_1||p_2) = \mathbb{E}_{x\sim p_2}\left[f\left(\frac{p_1(x)}{p_2(x)}\right)\right] \tag{1}$$

For different functions $f(x)$, the f-divergence class encompasses a wide range of commonly employed divergence measures, such as reverse Kullback-Leibler (KL) divergence, forward Kullback-Leibler (KL) divergence, α-divergence ($\alpha \in (0, 1)$), Jensen-Shannon (JS) divergence, and so on. In previous studies, reverse KL divergence can be regarded as a specific instance of α-divergence with $\alpha = 0$; and forward KL divergence can be considered as a specific instance of α-divergence with $\alpha = 1$. We summarize several commonly used f-divergence, the derivatives and the second derivatives in Table 1.

f-divergence	$f(x)$	$f'(x)$	$f''(x)$
Reverse KL	$x \log x$	$\log x + 1$	$\frac{1}{x}$
Forward KL	$-\log x$	$-\frac{1}{x}$	$\frac{1}{x^2}$
α-divergence	$\frac{x^{1-\alpha}-(1-\alpha)x-\alpha}{\alpha(\alpha-1)}$	$\frac{1-x^{-\alpha}}{\alpha}$	$\frac{1}{x^{\alpha+1}}$
JS divergence	$x \log \frac{2x}{x+1} + \log \frac{2}{x+1}$	$\log \frac{2x}{1+x}$	$\frac{1}{x(1+x)}$

Table 1: Several commonly used f-divergence with their derivatives and second derivatives.

Method

Much like in the alignment tasks of large language models, there are many concepts that are analogous in the alignment tasks of text-to-image models, and we start by elucidating these parallels. Firstly, the question input of LLMs is akin to the text (condition) input of T2I models, i.e. $x \to c$; and the output answer of LLMs is akin to the generated image of T2I models, i.e. $y \to x_0$. Moreover, the policy of LLMs parallels the sampling probability of T2I models (especially for diffusion models), i.e. $\pi(y|x) \to p(x_{0:T}|c)$. Finally, the preference data for output answers of LLMs is analogous to the preference data for generated images of T2I models, i.e. $(x, y_w, y_l) \to (c, x_0^w, x_0^l)$. In the following subsections, we first derive the generalized formula of alignment objective function. Then, we analyze the gradient field of different divergences during the alignment process with respect to the objective function and comprehensively analyze the impact of diverse divergence constraints on alignment performance.

Generalized Formula

In previous studies of Reinforcement Learning from Human Feedback (RLHF), researchers typically aim to maximize the reward function $r(c, x_{0:T})$ while penalizing the reverse KL divergence between the fine-tuned model and the original model to prevent it from collapsing during training. In our study, we generalize such penalty constraint from the reverse KL divergence ($D_{\text{KL}}\left(p_\theta(x_{0:T}|c), p_{\text{ref}}(x_{0:T}|c)\right)$) to the f-divergence ($D_f\left(p_\theta(x_{0:T}|c), p_{\text{ref}}(x_{0:T}|c)\right)$).

We reframe the reinforcement learning objective function as an optimal problem, presenting its formulation as follows:

$$\begin{aligned}\arg\max_{p_\theta} &\mathbb{E}_{c\sim p_c, x_{0:T}\sim p_\theta(x_{0:T}|c)}\left[r(c, x_{0:T})\right] \\ &-\beta D_f\Big(p_\theta(x_{0:T}|c), p_{\text{ref}}(x_{0:T}|c)\Big) \\ \text{s.t.} \quad &\sum_{x_{0:T}} p_\theta(x_{0:T}|c) = 1; \quad \forall x_{0:T} \quad p_\theta(x_{0:T}|c) \geq 0\end{aligned} \tag{2}$$

Such optimization problem can be addressed through the Karush-Kuhn-Tucker (KKT) conditions (Wang et al. 2024).

Firstly, according to the definition of f-divergence, we can construct the following Lagrangian function:

$$\begin{aligned}&\mathcal{L}(p_\theta(x_{0:T}|c),\lambda,\zeta(x_{0:T}))\\&=\mathbb{E}_{c\sim p_c,x_{0:T}\sim p_\theta}\left[r(c,x_{0:T})\right]-\beta\mathbb{E}_{p_{\text{ref}}}f\left(\frac{p_\theta(x_{0:T}|c)}{p_{\text{ref}}(x_{0:T}|c)}\right)\\&-\lambda\left(\sum_{x_{0:T}}p_\theta(x_{0:T})-1\right)+\sum_{x_{0:T}}\zeta(x_{0:T})p_\theta(x_{0:T}|c)\end{aligned}\tag{3}$$

Furthermore, we can derive the Theorem 1 from the *Stationarity Condition* and *Complementary Slackness* of the Karush-Kuhn-Tucker (KKT) conditions, i.e.

$$\begin{gathered}\nabla_{p_\theta(x_{0:T}|c)}\mathcal{L}(p_\theta(x_{0:T}|c),\lambda,\zeta(x_{0:T}))=0;\\\forall x_{0:T},\quad\zeta(x_{0:T})p_\theta(x_{0:T}|c)=0.\end{gathered}$$

Theorem 1. *If $p_{\text{ref}}(x_{0:T}|c)>0$ holds for all condition c, $f'(x)$ is an invertible function and 0 is not in the definition domain of function $f'(x)$, the reward class consistent with Bradley-Terrry model can be reparameterized with the sampling probability $p_\theta(x_{0:T}|c)$ and the reference sampling probability $p_{\text{ref}}(x_{0:T}|c)$ as:*

$$r(c,x_{0:T})=\beta f'\left(\frac{p_\theta(x_{0:T}|c)}{p_{ref}(x_{0:T}|c)}\right)+\text{const}\tag{4}$$

As shown in Theorem 1, the reward function can be represented by a sampling probability $p_\theta(x_{0:T}|c)$, a reference sampling probability $p_{\text{ref}}(x_{0:T}|c)$, and a constant λ that is independent of $x_{0:T}$. Finally, substituting Equation (4) into the Bradley-Terrry model (Bradley and Terry 1952) enables us to derive the generalized formula of text-to-image alignment with preferences in Theorem 2.

Theorem 2. *In the substitution process of Bradley-Terry model, the constant λ is independent of $x_{0:T}$ and thus can be canceled out, resulting in the following form:*

$$\begin{aligned}\mathcal{L}(\theta)=\mathbb{E}_{\substack{(c,x_0^w,x_0^l)\sim\mathcal{D},\\x_{1:T}^w\sim p_\theta(x_{1:T}^w|x_0^w,c),\\x_{1:T}^l\sim p_\theta(x_{1:T}^l|x_0^l,c).}}\\-\log\sigma\left[\beta f'\left(\frac{p_\theta(x_{0:T}^w|c)}{p_{ref}(x_{0:T}^w|c)}\right)-\beta f'\left(\frac{p_\theta(x_{0:T}^l|c)}{p_{ref}(x_{0:T}^l|c)}\right)\right]\end{aligned}\tag{5}$$

where $\sigma(\cdot)$ is the Sigmoid function; $f'(\cdot)$ represents the derivatives of $f(\cdot)$, as listed in Table 1; β is the penalty coefficient.

So far, we have derived the generalized formula for text-to-image generation alignment with preferences. With different divergence constraint choices, we can obtain diverse alignment objectives, thereby offering more options for the alignment process.

Analysis on Gradient Fields of Alignment Process

In this section, we focus on the gradient fields of alignment objective functions derived from various f-divergence constraints, aiming to further elucidate the intricate mechanisms underlying the alignment process.

Let's abstract from the specific details of $f'(\cdot)$, and concentrate instead on a more general formulation of the loss function, so that $\mathcal{L}(\theta)=\mathbb{E}[\mathcal{L}_f(\mathrm{X}_1,\mathrm{X}_2)]$:

$$\mathcal{L}_f(\mathrm{X}_1,\mathrm{X}_2)=-\log\sigma\big(\beta f'(\mathrm{X}_1)-\beta f'(\mathrm{X}_2)\big)\tag{6}$$

where X_1 is the training win ratio, and is equivalent to $\frac{p_\theta(x_{0:T}^w|c)}{p_{\text{ref}}(x_{0:T}^w|c)}$; similarly, X_2 is the training loss ratio, and is identical to $\frac{p_\theta(x_{0:T}^l|c)}{p_{\text{ref}}(x_{0:T}^l|c)}$. We present the gradients of Equation (6) with respect to X_1 and X_2 in the ensuing Theorem 3:

Theorem 3. *The partial derivatives (gradients) of X_1 and X_2 resulting from Equation* (6) *can be expressed as follows:*

$$\begin{aligned}\frac{\partial\mathcal{L}_f(\mathrm{X}_1,\mathrm{X}_2)}{\partial\mathrm{X}_1}&=-\beta\left(1-\sigma\left(\beta f'(\mathrm{X}_1)-\beta f'(\mathrm{X}_2)\right)\right)f''(\mathrm{X}_1)\\\frac{\partial\mathcal{L}_f(\mathrm{X}_1,\mathrm{X}_2)}{\partial\mathrm{X}_2}&=\beta\left(1-\sigma\left(\beta f'(\mathrm{X}_1)-\beta f'(\mathrm{X}_2)\right)\right)f''(\mathrm{X}_2)\end{aligned}$$

*Thus, the **gradient ratio** of $\mathcal{L}_f(\mathrm{X}_1,\mathrm{X}_2)$ between enhancement in probability for human-preferred responses (X_1) and reduction in probability for human-dispreferred responses (X_2) has the expression of:*

$$\left|\frac{\partial\mathcal{L}_f(\mathrm{X}_1,\mathrm{X}_2)}{\partial\mathrm{X}_1}\Big/\frac{\partial\mathcal{L}_f(\mathrm{X}_1,\mathrm{X}_2)}{\partial\mathrm{X}_2}\right|=\frac{f''(\mathrm{X}_1)}{f''(\mathrm{X}_2)}\tag{7}$$

Referencing Table 1, different divergences yield distinct gradient ratios. If selected divergence is reverse Kullback-Leibler divergence, the gradient ratio is $\frac{\mathrm{X}_2}{\mathrm{X}_1}$; if selected divergence is Jensen-Shannon divergence, the gradient ratio is $\frac{\mathrm{X}_2\cdot(\mathrm{X}_2+1)}{\mathrm{X}_1\cdot(\mathrm{X}_1+1)}$; if selected divergence is α-divergence, the gradient ratio is $\frac{\mathrm{X}_2^{1+\alpha}}{\mathrm{X}_1^{1+\alpha}}$; if selected divergence is forward Kullback-Leibler divergence, the gradient ratio is $\frac{\mathrm{X}_2^2}{\mathrm{X}_1^2}$. Previous studies (Feng et al. 2024a; Yan et al. 2024) present detailed theoretical analysis towards the results of original DPO framework, focusing particularly on its application in the context of reverse Kullback-Leibler divergence; while we show the analysis of generalization under diverse divergences.

Furthermore, as the alignment process advances, the value of X_1 tends to increase to more than 1, whereas X_2 tends to decrease to less than 1. Hence, for any pairwise preference data, $\mathrm{X}_2/\mathrm{X}_1<1$ holds during the alignment process. Then, Theorem 4 can be easily derived.

Theorem 4. *As alignment progresses, we have $\mathrm{X}_2/\mathrm{X}_1<1$. Hence,*

$$0<\frac{\mathrm{X}_2^2}{\mathrm{X}_1^2}<\frac{\mathrm{X}_2\cdot(\mathrm{X}_2+1)}{\mathrm{X}_1\cdot(\mathrm{X}_1+1)}<\frac{\mathrm{X}_2}{\mathrm{X}_1}<1\quad\textit{and}$$

$$0<\frac{\mathrm{X}_2^2}{\mathrm{X}_1^2}<\frac{\mathrm{X}_2^{1.8}}{\mathrm{X}_1^{1.8}}<\frac{\mathrm{X}_2^{1.6}}{\mathrm{X}_1^{1.6}}<\frac{\mathrm{X}_2^{1.4}}{\mathrm{X}_1^{1.4}}<\frac{\mathrm{X}_2^{1.2}}{\mathrm{X}_1^{1.2}}<\frac{\mathrm{X}_2}{\mathrm{X}_1}<1$$

Theorem 4 presents the inequality of gradient ratio of different divergences. A lower gradient ratio results in a swifter alteration in the probability of a dispreferred image compared to that of a preferred one, indicating a more pronounced decrease in the probability of dispreferred images.

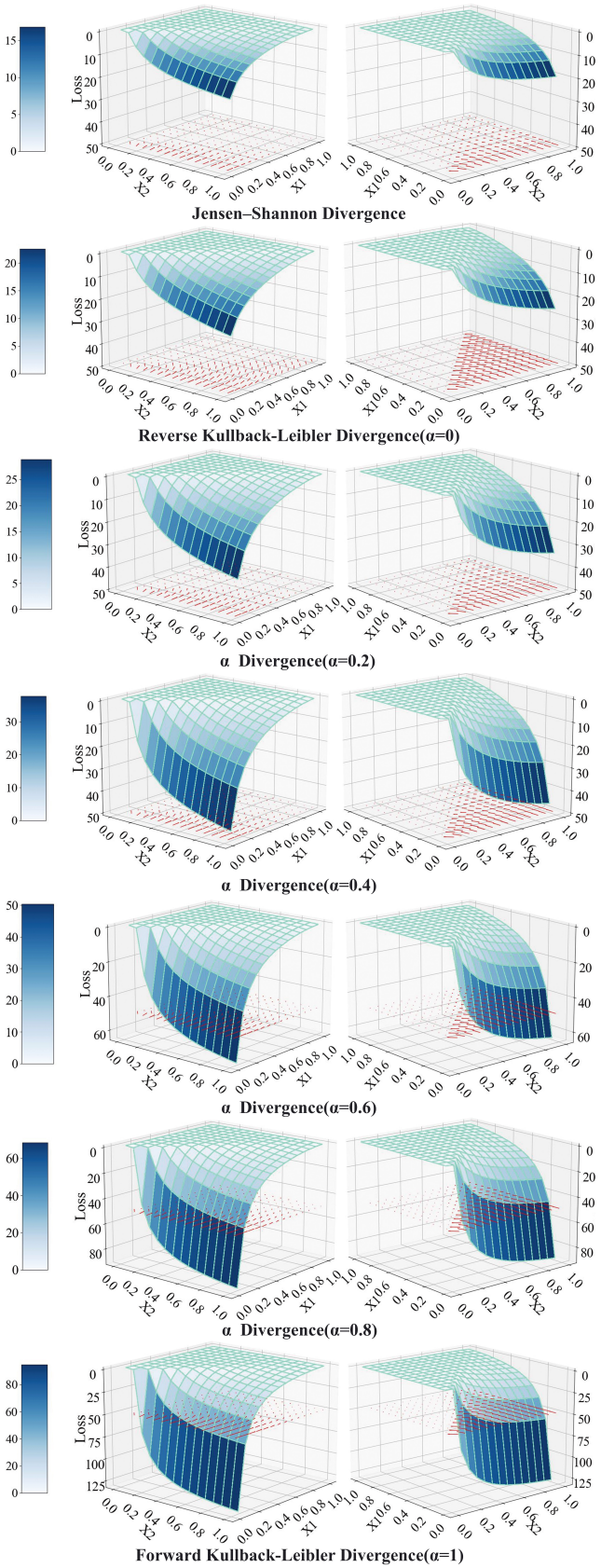

Figure 2: Visualization of landscapes and gradient fields.

Hence, the decline varies in intensity, with *forward KL divergence (α=1)* exhibiting the highest decrease, *reverse KL divergence (α=0)* the lowest, and both *Jensen-Shannon divergence* and *α-divergence ($\alpha \in$(0,1))* falling in between.

In order to obtain a more intuitive understanding of the impact of different divergence choices during the alignment process, we visualize the landscape of alignment objective functions with different divergences in Equation (6) from two viewing angles, as shown in Figure 2 (the penalty coefficient β is selected as 10). Furthermore, to enhance intuition, we plot the gradient field of corresponding loss function on the plane where Z equals 50. When it comes to consider the smoothness within loss function landscape, surface of *Jensen-Shannon divergence* exhibits the best smoothness, suggesting a more stable alignment process. Moreover, this suggests a robust alignment mechanism that effectively prevents the process from merely eliminating dispreferred outputs rather than actively guiding chosen outputs towards optimization. Additionally, it mitigates the issue where the gradient on X_1 gradually diminishes as X_2 rapidly decreases to 0, which consequently leads to a stochastic decline in likelihood of the preferred output (Yan et al. 2024). These improvements are of paramount importance in the scenario of text-to-image alignment, as full alignment has been achieved within 10 epochs (especially for SPO and Diffusion-DPO, D3PO conduct 1000 epochs).

Experiments

In this section, we present extensive experimental evaluations to answer the following questions:
Q1: When choosing different divergences, would it have a significant impact on *text-image* alignment performance?
Q2: When choosing different divergences, would it have a significant impact on the alignment of *human value*? If so, which divergence constraint achieves the best performance?
Q3: When choosing different divergences, would it have an impact on the *generation diversity*? Which divergence can achieve the best trade-off between alignment performance and generation diversity?

Experimental Settings

Benchmark. Step-aware Preference Optimization (SPO) (Liang et al. 2024) employs a step-aware preference model and a step-wise resampler to guarantee precise step-aware preference alignment. Consequently, to support a more tangible experimental assessment, we select SPO as our benchmark approach. To establish a fair basis for comparison with prior methods, we select Stable Diffusion v1-5 model (Rombach et al. 2022) as our benchmark model. In order to conduct a more comprehensive evaluation, we utilize the test set of HPS-V2 (Wu et al. 2023) as our evaluation benchmark dataset, which comprises 400 prompts. We report the mean and standard deviation of metrics of the generated image for these prompts.

Evaluation Metrics. We evaluate the generated images from three aspects (for the aforementioned three questions).

In terms of model's text-image alignment performance (for Q1), we adopt the widely used evaluation metrics in

Model		CLIPScore ↑	VQAScore ↑	Aesthetics Score ↑	ImageReward ↑	PickScore ↑	HPS-V2 ↑
Original Model		0.352±0.049	0.601±0.243	5.648±0.526	0.173±1.011	20.908±1.228	26.933±1.454
Reverse KL Divergence		**0.363±0.049**	**0.654±0.239**	5.812±0.514	0.619±0.921	21.621±1.151	27.801±1.352
α-Divergence	α=0.2	0.360±0.048	0.641±0.234	5.827±0.546	0.561±0.957	21.528±1.177	27.848±1.391
	α=0.4	0.361 ±0.047	0.651±0.240	5.755±0.518	0.622±0.911	21.569±1.204	27.762±1.385
	α=0.6	0.358 ±0.047	0.630±0.241	5.769±0.481	0.491±0.943	21.357±1.180	27.712±1.350
	α=0.8	0.361±0.050	0.633±0.245	5.821±0.511	0.561±0.965	21.483±1.175	27.675±1.379
Forward KL Divergence		0.362±0.050	0.646±0.238	5.844±0.528	0.551±0.946	21.552±1.170	27.822±1.355
Jensen-Shannon Divergence		0.361±0.049	0.647±0.235	**5.884±0.514**	**0.631±0.939**	**21.635±1.149**	**27.850±1.388**

Table 2: Evaluations of the alignment performance, where the CLIPScore and VQAScore evaluates text-image alignment performance, and the remaining four metrics evaluate human value alignment performance.

Text-to-Image models, i.e., Text-Image CLIPScore (Hessel et al. 2021) and VQAScore (Lin et al. 2025). CLIPScore is fundamentally based on the CLIP model, transforming input text and images into distinct text and image vectors, and then followed by calculation of the dot product of these vectors. VQAScore (Lin et al. 2025) is grounded in generative vision-language models (VLMs) that have been specifically trained to undertake visual-question-answering (VQA) tasks (generating an answer from an image and a question). Hence, higher Text-Image CLIPScore and VQAScore indicates a better alignment between the text and the image.

In terms of model's human value alignment performance (for Q2), we adopt four metrics for comprehensive evaluation. Aesthetic score is obtained using the LAION Aesthetics Predictor (Schuhmann et al. 2022), which quantifies the average human appreciation for the visual appeal of generated images. ImageReward (Xu et al. 2024), leveraging a structure that combines ViT-L for image encoding and a 12-layer Transformer for text encoding, effectively modeling the human value and preference. PickScore (Kirstain et al. 2023) is an advanced scoring function built upon a meticulously curated comprehensive dataset named "Pick-a-Pic". Human Preference Score v2 (HPS-v2) (Wu et al. 2023) has been developed through the refinement of the CLIP model on HPD-v2, which enhances the precision of assessing human preferences for generated images. Furthermore, higher Aesthetics Score, ImageReward, Pickscore and HPS-V2 suggest better alignment with human value.

In terms of diversity of generated images from the aligned model (for Q3), we adopt eight metrics for a further comprehensive evaluation. Image-Image CLIPScore (Hessel et al. 2021) has served as a reliable metric for assessing similarity between images. RMSE, PSNR, and SSIM are conventional metrics used to evaluate image similarity, they are also utilized to assess the diversity of generated images. Feature Similarity Index Measure (FSIM) (Zhang et al. 2011) quantifies the similarity between images by assessing the alignment of edges, shapes, visual patterns, and surface attributes. Learned Perceptual Image Patch Similarity (LPIPS) (Zhang et al. 2018) utilizes the feature representations learned by a deep neural network, which is capable of capturing details of human visual perception such as texture, color, and structure; then computing perceptual similarity between two images. Furthermore, it's worth noting that these six metrics all initially describe the similarity between images; and when they are used to describe generation diversity, their properties are the opposite of their properties when describing similarity. Moreover, we opt for Image Entropy, encompassing both Entropy 1D and Entropy 2D, to evaluate the information content diversity within images themselves; they quantifies the average information per pixel, with higher entropy values indicative of a greater diversity and richness in the image's information content.

Text-Image Alignment (For Q1)

For text-to-image models, the alignment performance between text prompts and generated images is of paramount importance. Therefore, we test the Text-Image CLIPScore of all models fine-tuned under different divergence constraints to assess the alignment performance in Table 2. The results indicate that the reverse Kullback-Leibler divergence achieves the best text-image alignment performance; while it is also worth noting that different divergences do not significantly affect the final text-image alignment performance.

Human Value Alignment (For Q2)

We compare human value alignment performances systematically in Table 2. Comparison between results of the fine-tuned models and the original model indicates that alignment process has effectively enhanced the model in terms of its performance in human values. Furthermore, in comparing the influence of diverse divergence constraints on human value alignment, the results reveal that different divergence would significantly affect human value alignment; remarkably, *Jensen-Shannon (JS) divergence* exhibits the best performance across all four human value alignment metrics, suggesting that it serves as a more potent constraint specifically for the scenario of human value alignment. Actually, such observation also aligns with our previous analysis of the gradient fields, where Jenson-Shannon (JS) divergence shows the smoothest loss function surface and suboptimal gradient ratio, resulting in a more stable alignment process.

Model		Image-Image CLIPScore ↓	Entropy 1D ↑	Entropy 2D ↑	LPIPS ↑
Original Model		0.7994 ± 0.0842	7.2547 ± 0.6230	14.0615 ± 1.3478	0.6306 ± 0.0707
Reverse KL Divergence		0.8457 ± 0.0767	7.5382 ± 0.3434	14.4520 ± 0.8549	0.6092 ± 0.0623
α-Divergence	α =0.2	0.8431 ± 0.0813	7.4879 ± 0.4448	14.3615 ± 1.0211	0.6308 ± 0.0633
	α =0.4	0.8390 ± 0.0796	**7.5832 ± 0.3360**	14.4098 ± 0.8855	0.6208 ± 0.0586
	α =0.6	**0.8385 ± 0.0823**	7.4551 ± 0.4636	14.3648 ± 1.0446	**0.6387 ± 0.0656**
	α =0.8	0.8416 ± 0.0766	7.5391 ± 0.3813	14.3011 ± 0.9231	0.6270 ± 0.0639
Forward KL Divergence		0.8435 ± 0.0834	7.5075 ± 0.3663	14.3415 ± 0.8795	0.6202 ± 0.0589
Jensen-Shannon Divergence		0.8474 ± 0.0794	7.5383 ± 0.3505	**14.5798 ± 0.8944**	0.6204 ± 0.0620

Model		RMSE ↑	PSNR ↓	SSIM ↓	FSIM ↓
Original Model		0.0202 ± 0.0043	34.0374 ± 1.8413	0.7344 ± 0.0858	0.2942 ± 0.0307
Reverse KL Divergence		0.0234 ± 0.0042	32.7175 ± 1.5753	0.6660 ± 0.0820	0.3085 ± 0.0247
α-Divergence	α =0.2	0.0249 ± 0.0042	32.1710 ± 1.4683	0.6425 ± 0.0812	**0.3028 ± 0.0249**
	α =0.4	0.0234 ± 0.0039	32.7083 ± 1.4278	0.6726 ± 0.0767	0.3121 ± 0.0235
	α =0.6	**0.0253 ± 0.0040**	**32.0264 ± 1.3737**	**0.6321 ± 0.0767**	0.3059 ± 0.0275
	α =0.8	0.0236 ± 0.0042	32.6651 ± 1.5654	0.6689 ± 0.0819	0.3071 ± 0.0254
Forward KL Divergence		0.0240 ± 0.0040	32.4968 ± 1.4576	0.6573 ± 0.0773	0.3045 ± 0.0236
Jensen-Shannon Divergence		0.0242 ± 0.0040	32.4080 ± 1.4267	0.6508 ± 0.0780	0.3097 ± 0.0258

Table 3: Evaluations of the generation diversity. The metrics originally utilized for evaluating image similarity exhibit an opposite property when evaluating generation diversity. We evaluate the metrics employed for pairwise image comparisons.

Generation Diversity (For Q3)

We evaluate the generation diversity of aligned models with different divergence constraints from multiple perspectives (embedding diversity, pixel-level diversity, structural diversity, perceptual diversity, information complexity, and so on), and the corresponding results are shown in Table 3. From the results, we can observe that different divergence constraints exhibit advantages in different aspects when evaluated with different generation diversity metrics. Firstly, we would like to compare the aligned models under different divergence constraints to the original model: it is demonstrated that the aligned models show a decrease in *embedding diversity*; however, they exhibit improvements in other aspects such as pixel-level diversity, structural diversity, information complexity. Such observation reveals a transformation in the alignment process that the variety of the primary subject diminishes, yet the intricacy and breadth of details and structures of the generated images expand, echoing findings from DreamBooth (Ruiz et al. 2023).

Furthermore, it has also indicated that increased generative diversity is associated with a decline in alignment performance (both text-image alignment and human value alignment). Therefore, careful consideration of the trade-off between alignment performance and generation diversity is essential when choosing divergence constraint and the choice should depend on how much we care about generation diversity compared to alignment performance. Through a deeper comparison and analysis, we can observe that Jensen-Shannon divergence outperforms or matches reverse Kullback-Leibler divergence across most diversity metrics. Combining such observation with the previous evaluation that achieving the best human value alignment, we consider Jensen-Shannon divergence a better trade-off between alignment performance and generation diversity.

Conclusion

In this paper, we extend the alignment framework for text-to-image models, transitioning from the constraint based on the reverse Kullback-Leibler (KL) divergence to a more inclusive framework grounded in f-divergence constraints. Through analysis of gradient fields (gradient ratio and loss function surface) under diverse divergence constraints, we further illustrate advantages of different divergence constraints in the alignment process. Regarding text-image alignment, minimal differentiation is observed among diverse divergence constraints; conversely, for human value alignment, Jensen-Shannon (JS) divergence excels, showcasing its superior performance across all four evaluation metrics. In generative diversity, we observe that diverse divergence constraints demonstrate strengths in various aspects of diversity. Furthermore, it has been observed that increased generation diversity often correlates with a decrease in alignment performance. After thorough comparison, we suggest selecting Jensen-Shannon (JS) divergence as the foremost option in practice, which is a better trade-off between alignment performance and generation diversity.

Acknowledgements

This work is partly supported by the National Natural Science Foundation of China (No.62103225), Natural Science Foundation of Shenzhen (No.JCYJ20230807111604008), Natural Science Foundation of Guangdong Province (No.2024A1515010003) and National Key Research and Development Program (No.2022YFB4701402).

References

Arjovsky, M.; Chintala, S.; and Bottou, L. 2017. Wasserstein generative adversarial networks. In *International conference on machine learning*, 214–223. PMLR.

Black, K.; Janner, M.; Du, Y.; Kostrikov, I.; and Levine, S. 2024. Training Diffusion Models with Reinforcement Learning. In *The Twelfth International Conference on Learning Representations*.

Bradley, R. A.; and Terry, M. E. 1952. Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika*, 39(3/4): 324–345.

Feng, D.; Qin, B.; Huang, C.; Zhang, Z.; and Lei, W. 2024a. Towards analyzing and understanding the limitations of dpo: A theoretical perspective. *arXiv preprint arXiv:2404.04626*.

Feng, K.; Ma, Y.; Wang, B.; Qi, C.; Chen, H.; Chen, Q.; and Wang, Z. 2024b. Dit4edit: Diffusion transformer for image editing. *arXiv preprint arXiv:2411.03286*.

Gambashidze, A.; Kulikov, A.; Sosnin, Y.; and Makarov, I. 2024. Aligning Diffusion Models with Noise-Conditioned Perception. *arXiv preprint arXiv:2406.17636*.

Go, D.; Korbak, T.; Kruszewski, G.; Rozen, J.; Ryu, N.; and Dymetman, M. 2023. Aligning Language Models with Preferences through f-divergence Minimization. In *International Conference on Machine Learning*, 11546–11583. PMLR.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative adversarial nets. *Advances in neural information processing systems*, 27.

Hessel, J.; Holtzman, A.; Forbes, M.; Bras, R. L.; and Choi, Y. 2021. Clipscore: A reference-free evaluation metric for image captioning. *arXiv preprint arXiv:2104.08718*.

Hiriart-Urruty, J.-B.; and Lemaréchal, C. 1996. *Convex analysis and minimization algorithms I: Fundamentals*, volume 305. Springer science & business media.

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. *Advances in Neural Information Processing Systems*, 36: 36652–36663.

Li, S.; Kallidromitis, K.; Gokul, A.; Kato, Y.; and Kozuka, K. 2024a. Aligning diffusion models by optimizing human utility. *arXiv preprint arXiv:2404.04465*.

Li, S.; Singh, H.; and Grover, A. 2024. PopAlign: Population-Level Alignment for Fair Text-to-Image Generation. *arXiv preprint arXiv:2406.19668*.

Li, T.; Tian, Y.; Li, H.; Deng, M.; and He, K. 2024b. Autoregressive Image Generation without Vector Quantization. *arXiv preprint arXiv:2406.11838*.

Liang, Z.; Yuan, Y.; Gu, S.; Chen, B.; Hang, T.; Li, J.; and Zheng, L. 2024. Step-aware Preference Optimization: Aligning Preference with Denoising Performance at Each Step. *arXiv preprint arXiv:2406.04314*.

Liese, F.; and Vajda, I. 2006. On divergences and informations in statistics and information theory. *IEEE Transactions on Information Theory*, 52(10): 4394–4412.

Lin, Z.; Pathak, D.; Li, B.; Li, J.; Xia, X.; Neubig, G.; Zhang, P.; and Ramanan, D. 2025. Evaluating text-to-visual generation with image-to-text generation. In *European Conference on Computer Vision*, 366–384. Springer.

Nowozin, S.; Cseke, B.; and Tomioka, R. 2016. f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization. In Lee, D.; Sugiyama, M.; Luxburg, U.; Guyon, I.; and Garnett, R., eds., *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc.

Perez, E.; Huang, S.; Song, F.; Cai, T.; Ring, R.; Aslanides, J.; Glaese, A.; McAleese, N.; and Irving, G. 2022. Red Teaming Language Models with Language Models. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 3419–3448.

Podell, D.; English, Z.; Lacey, K.; Blattmann, A.; Dockhorn, T.; Müller, J.; Penna, J.; and Rombach, R. 2024. SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis. In *The Twelfth International Conference on Learning Representations*.

Rafailov, R.; Sharma, A.; Mitchell, E.; Ermon, S.; Manning, C. D.; and Finn, C. 2023. Direct preference optimization: your language model is secretly a reward model. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, 53728–53741.

Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; and Ommer, B. 2022. High-Resolution Image Synthesis With Latent Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10684–10695.

Ruiz, N.; Li, Y.; Jampani, V.; Pritch, Y.; Rubinstein, M.; and Aberman, K. 2023. DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 22500–22510.

Schuhmann, C.; Beaumont, R.; Vencu, R.; Gordon, C.; Wightman, R.; Cherti, M.; Coombes, T.; Katta, A.; Mullis, C.; Wortsman, M.; et al. 2022. Laion-5b: An open large-scale dataset for training next generation image-text models. *Advances in Neural Information Processing Systems*, 35: 25278–25294.

Tang, Z.; Peng, J.; Tang, J.; Hong, M.; Wang, F.; and Chang, T.-H. 2024. Tuning-Free Alignment of Diffusion Models with Direct Noise Optimization. *arXiv preprint arXiv:2405.18881*.

Theis, L.; van den Oord, A.; and Bethge, M. 2016. A note on the evaluation of generative models. In *International Conference on Learning Representations (ICLR 2016)*, 1–10.

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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8228–8238.

Wang, C.; Jiang, Y.; Yang, C.; Liu, H.; and Chen, Y. 2024. Beyond Reverse KL: Generalizing Direct Preference Optimization with Diverse Divergence Constraints. In *The Twelfth International Conference on Learning Representations*.

Wiher, G.; Meister, C.; and Cotterell, R. 2022. On decoding strategies for neural text generators. *Transactions of the Association for Computational Linguistics*, 10: 997–1012.

Wu, X.; Hao, Y.; Sun, K.; Chen, Y.; Zhu, F.; Zhao, R.; and Li, H. 2023. Human preference score v2: A solid benchmark for evaluating human preferences of text-to-image synthesis. *arXiv preprint arXiv:2306.09341*.

Xu, J.; Liu, X.; Wu, Y.; Tong, Y.; Li, Q.; Ding, M.; Tang, J.; and Dong, Y. 2024. Imagereward: Learning and evaluating human preferences for text-to-image generation. *Advances in Neural Information Processing Systems*, 36.

Yan, Y.; Miao, Y.; Li, J.; Zhang, Y.; Xie, J.; Deng, Z.; and Yan, D. 2024. 3D-Properties: Identifying Challenges in DPO and Charting a Path Forward. *arXiv preprint arXiv:2406.07327*.

Yang, K.; Tao, J.; Lyu, J.; Ge, C.; Chen, J.; Shen, W.; Zhu, X.; and Li, X. 2024. Using human feedback to fine-tune diffusion models without any reward model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8941–8951.

Yang, S.; Chen, T.; and Zhou, M. 2024. A Dense Reward View on Aligning Text-to-Image Diffusion with Preference. In *Forty-first International Conference on Machine Learning*.

Zhang, L.; Zhang, L.; Mou, X.; and Zhang, D. 2011. FSIM: A feature similarity index for image quality assessment. *IEEE transactions on Image Processing*, 20(8): 2378–2386.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 586–595.

Zhu, C.; Li, K.; Ma, Y.; He, C.; and Xiu, L. 2024. MultiBooth: Towards Generating All Your Concepts in an Image from Text. *arXiv preprint arXiv:2404.14239*.