

SHORTFT: Diffusion Model Alignment via Shortcut-based Fine-Tuning

Xiefan Guo^{1,2} MiaoMiao Cui Liefeng Bo Di Huang^{1,2*}

¹State Key Laboratory of Complex and Critical Software Environment, Beihang University, Beijing 100191, China

²School of Computer Science and Engineering, Beihang University, Beijing 100191, China

{xfguo, dhuang}@buaa.edu.cn

Project: <https://xiefan-guo.github.io/shortft>

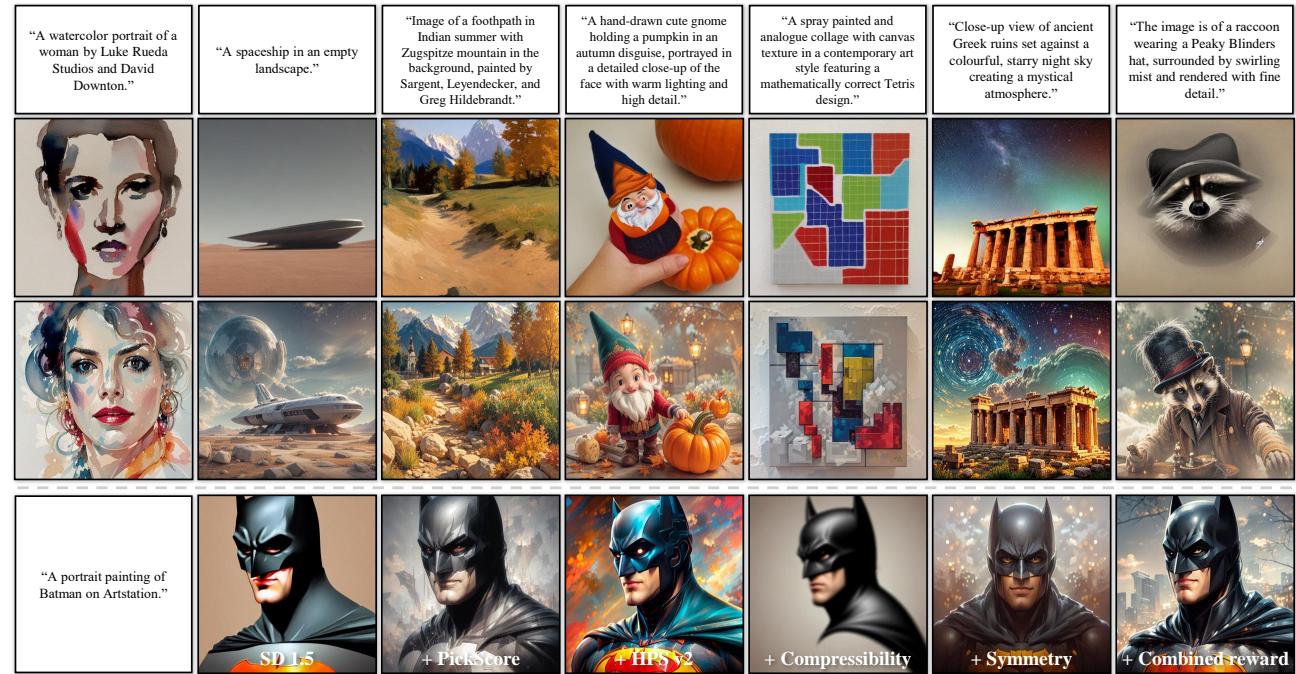


Figure 1. Example results synthesized by SHORTFT. SHORTFT endeavors to achieve the alignment of diffusion models with reward functions by facilitating the end-to-end backpropagation of the targeted reward gradient throughout the denoising chain. Our method has exhibited remarkable efficacy, particularly evident in the realms of text-image alignment and the overall enhancement of image quality (Top). Moreover, its versatility is underscored by the successful application across diverse reward functions, substantially amplifying alignment performance (Bottom). Combined reward is a weighted combination of rewards: PickScore = 10, HPS v2 = 2, Aesthetic = 0.05.

Abstract

Backpropagation-based approaches aim to align diffusion models with reward functions through end-to-end backpropagation of the reward gradient within the denoising chain, offering a promising perspective. However, due to the computational costs and the risk of gradient explosion associated with the lengthy denoising chain, existing approaches struggle to achieve complete gradient backpropagation, leading to suboptimal results. In this paper, we introduce Shortcut-based Fine-Tuning (SHORTFT), an effi-

cient fine-tuning strategy that utilizes the shorter denoising chain. More specifically, we employ the recently researched trajectory-preserving few-step diffusion model, which enables a shortcut over the original denoising chain, and construct a shortcut-based denoising chain of shorter length. The optimization on this chain notably enhances the efficiency and effectiveness of fine-tuning the foundational model. Our method has been rigorously tested and can be effectively applied to various reward functions, significantly improving alignment performance and surpassing state-of-the-art alternatives.

*Corresponding author.

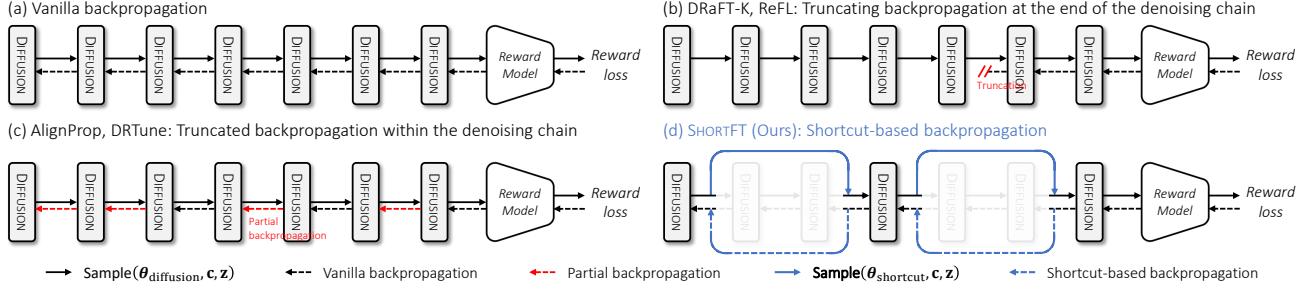


Figure 2. **Comparison of fine-tuning strategies.** (a) The vanilla backpropagation-based fine-tuning strategy, which suffers from lengthy backpropagation chains. (b) DRaFT-K [6] and ReFL [52] truncate the backpropagation chain, focusing on the latter half of the denoising chain, they ignore the direct supervision at the early stage, resulting in suboptimal alignment with text prompts. (c) AlignProp [34] and DRTune [51] truncate part of the backpropagation within the denoising chain by disabling some gradients. Specifically, given the denoising operation: $\mathbf{x}_{t-1} = \alpha_t \mathbf{x}_t + \beta_t \epsilon_\theta(\mathbf{x}_t, t) + c_t \epsilon$, partial backpropagation only utilizes the gradients of the red part ($\alpha_t \mathbf{x}_t$), truncating the green part ($\beta_t \epsilon_\theta(\mathbf{x}_t, t)$). This inevitably introduces gradient errors, leading to unstable optimization. (d) Our method, using the few-step diffusion model to construct denoising shortcuts, facilitates complete gradient backpropagation through the entire denoising chain.

1. Introduction

Diffusion models [1, 9, 17, 18, 22, 32, 38, 44, 45] have established themselves as a pioneering approach in generative modeling, demonstrating exceptional prowess in applications such as photo-realistic text-to-image synthesis. However, the maximum likelihood training objective of these models, which aims to model the training data distribution accurately, can often conflict with downstream goals like aesthetics, fairness, safety, and text-to-image alignment. Therefore, aligning text-to-image diffusion models with human preferences has emerged as a pivotal and practical task.

Directly supervised fine-tuning on small-scale, human-curated datasets, *e.g.*, LAION Aesthetics [42], presents a straightforward solution. However, the prohibitive cost of data collection and the rapid obsolescence of datasets, particularly in terms of resolution compatibility with the latest text-to-image models, make this approach impractical.

Emulating the successful application of Reinforcement Learning from Human Feedback (RLHF) [5, 13, 33, 47, 59] in Large Language Models (LLMs), several studies [3, 10, 12, 25, 50] have experimented with Reinforcement Learning (RL) techniques to align diffusion models with a reward function. Despite promising performance enhancements in specific domains, RL-based methods are notorious for their high-variance gradients, leading to inefficiencies and limited adaptability to diverse prompts.

Recently, backpropagation-based methods [6, 34, 51, 52] have sought to align diffusion models with reward functions using end-to-end backpropagation of the reward gradient through the denoising chain, showing potential. However, these strategies face challenges arising from the lengthy denoising chain, which demands considerable computational resources and is susceptible to gradient explosion. As illustrated in Fig. 2 (b), [6, 52] have made strides by truncating backpropagation to focus on the latter part of the de-

noising chain. However, they overlook direct supervision in the early stages, leading to suboptimal alignment with text prompts. As illustrated in Fig. 2 (c), [34, 51] truncate a portion of backpropagation within the denoising chain by deactivating some gradients and introduce gradient checkpointing, enabling gradient backpropagation to the early stage of the denoising chain. However, they are time-consuming and introduce gradient bias, leading to optimization instability.

This paper revisits the issue of excessively long denoising chain and propose an alternative approach, employing the shorter denoising chain to facilitate full gradient backpropagation throughout the entire denoising chain.

In this paper, we introduce Shortcut-based Fine-Tuning (SHORTFT), which leverages the few-step diffusion model to construct a denoising *shortcut* to **fine-tune the foundational model** (*e.g.*, SD 1.5), inspired by recent trajectory-preserving diffusion distillation methods [23, 37, 39, 46]. This technique enables us to bypass the original lengthy denoising chain and complete the inference, creating a shortcut-based denoising chain of shorter length. In addition, we construct a timestep-aware LoRA as an expert LoRA ensemble, based on the intriguing temporal dynamics exhibited by the text-to-image diffusion model during the denoising process. This approach increases the number of trainable parameters without increasing cost in the inference phase, enabling faster convergence and improved performance. We also devise a custom progressive training strategy to mitigate training inference bias introduced by using denoising shortcuts during the training stage.

Extensive quantitative and qualitative analyses demonstrate that SHORTFT can be effectively applied to various reward functions and architectures, significantly enhancing alignment performance. Furthermore, SHORTFT, benefiting from the short denoising chain and without the need for gradient checkpointing, is particularly efficient, learning faster than DRTune [51], the current most efficient method.



Figure 3. **Denoising shortcut.** The trajectory-preserving few-step diffusion model naturally introduces denoising shortcut, allowing for flexible skipping within the denoising chain while still ensuring high-quality and consistent image synthesis. The 4-step Hyper-SD distilled from SD 1.5 is used in our experiments. $\text{SHORTCUT}(i)$ denotes completing the denoising process from timestep i to 0 using the few-step diffusion model. In addition, we also provide the well-known one-step denoising results $\text{DDIM}(i)$, where $\text{DDIM}(i)$ denotes performing a one-step denoising operation from timestep i to 0. Intuitively, the output of $\text{DDIM}(i)$ is more blurred, lacking accurate structure and texture details, while the output of $\text{SHORTCUT}(i)$ is closer to the original output of SD 1.5. Furthermore, we provide their corresponding HPS v2 scores, with the absolute value of the deviation from the score corresponding to SD 1.5 provided in parentheses, $\text{SHORTCUT}(i)$ exhibits smaller deviations. These observations collectively indicate the reliability and validity of the denoising shortcut.

2. Related Work

2.1. Alignment of diffusion models

Diffusion models [9, 18, 44, 45] have become a dominant force in generative modeling, showing exceptional performance in diverse applications [1, 14–17, 22, 32, 38, 53, 57]. However, certain misalignments with human intentions can arise. Recent research, fueled by the successful alignment of large language models, has sought to align diffusion models with human expectations and preferences.

Fine-tuning via data augmentation. Several studies [7, 10, 20, 25, 50] have explored altering the training data distribution for fine-tuning diffusion models on visually compelling and textually coherent data, which has led to improved results. Other methods [2, 43] involve re-captioning pre-collected web images to enhance textual precision.

Fine-tuning via reward models. Reward models [24, 25, 42, 49, 50, 52] are employed to emulate human preferences given an input prompt and generated images. Several approaches have attempted to integrate these signals to augment text-to-image generation. A significant direction is the utilization of reinforcement learning-based algorithms [3, 4, 8, 11, 58] for fine-tuning text-to-image diffusion models in alignment with these rewards, [29, 30, 36, 48, 54, 55] bypass it entirely with Direct Preference Optimization. However, these methods are costly and have high gradient variance, leading to inefficiency and limited adaptability to diverse prompts. Consequently, backpropagation-based techniques [6, 34, 51, 52] have been explored, which directly fine-tune diffusion models using differentiable rewards [24, 42, 49, 50, 52].

The challenge of backpropagation-based strategies stem from the lengthy denoising chain, which often requires numerous denoising operations (e.g., 50 for DDIM), corresponding to a long backpropagation chain. This process incurs substantial time and memory costs and is prone to gradient explosion. To mitigate this issue, [6, 52] truncate backpropagation by concentrating on the latter part

of the denoising chain. While these approaches yields some improvements, they neglect direct supervision in the early stage of the denoising chain, leading to less precise alignment with text prompts. [34, 51] truncate part of the backpropagation within the denoising chain by deactivating some gradients. By employing gradient checkpointing, they allows for the propagation of gradients to the early stages of the denoising chain. Despite their merits, these techniques can be computationally demanding and induce gradient bias, resulting in optimization instability.

Different from previous methods, this paper revisits the fundamental challenge of excessively long denoising chain and proposes an alternative approach: leveraging the shorter denoising chain to facilitate full gradient backpropagation throughout the denoising chain.

2.2. Diffusion distillation

Existing methods for the distillation of diffusion models can be primarily classified into two categories: trajectory-preserving distillation [23, 37, 39, 46] and trajectory-reformulating distillation [21, 31, 40, 41, 56]. The former aims to preserve the original denoising trajectory dictated by an ordinary differential equation (ODE), while the latter focuses on leveraging the denoising endpoint as the main supervision, disregarding the intermediate trajectory steps. This paper focuses on trajectory-preserving distillation, supporting the establishment of a denoising shortcut within the complete denoising chain.

2.3. Fine-tuning few-step diffusion models

Existing works [26–28] have successfully explored fine-tuning few-step diffusion models. While they share similarities with our research in utilizing few-step diffusion models, a crucial difference is that we focus on fine-tuning the foundational models. In contrast to the foundational models, few-step diffusion models suffer from performance degradation and reduced capacity caused by the distillation process, leading fine-tuning them suboptimal (see Sec. 4.4).

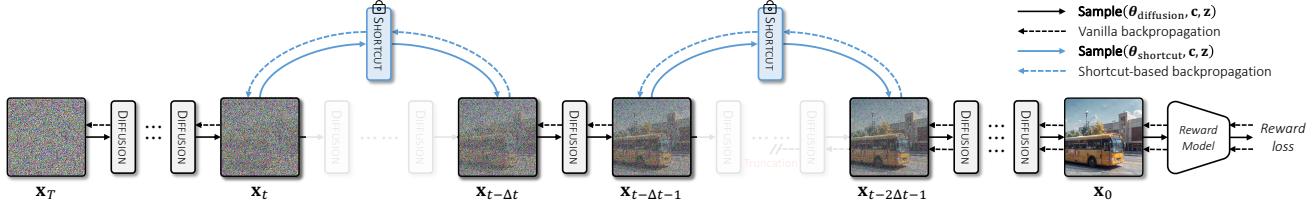


Figure 4. **Illustration of SHORTFT.** The core of the proposed method is the Shortcut-based Fine-Tuning (SHORTFT), which leverages the trajectory-preserving few-step diffusion model as the **shortcut** (identified as blue arrow) to achieve direct end-to-end backpropagation through the diffusion sampling process, fine-tuning the parameters of the pre-trained diffusion model to align it with the reward function.

3. SHORTFT

Our method, Shortcut-based Fine-Tuning (SHORTFT), capitalizes on the trajectory-preserving few-step diffusion model as a **shortcut** to achieve direct end-to-end backpropagation through the diffusion sampling process. This approach fine-tunes the parameters of the pre-trained diffusion model to align it with the reward function.

3.1. Problem formulation

In line with [6, 34, 51], SHORTFT focuses on fine-tuning the parameters θ of pre-trained diffusion models to maximize the differentiable reward function $\mathcal{R}(\cdot)$ for generated images. This can be formally represented as in Eq. 1:

$$J(\theta) = \mathbb{E}_{c, x_T \sim \mathcal{N}(0, 1)} [\mathcal{R}(\text{Sample}(\theta, c, x_T), c)], \quad (1)$$

where $\text{Sample}(\theta, c, x_T)$ represents the denoising process for the timestep $t = T \rightarrow 0$ with prompt condition c .

Consistent with [6, 34, 51], Eq. 1 is resolved by calculating $\nabla \mathcal{R}(\text{Sample}(\theta, c, x_T), c)$ and employing gradient ascent. The computation of this gradient necessitates backpropagation through multiple diffusion models in the denoising chain, akin to backpropagation through time in recurrent neural networks.

3.2. Denoising shortcut

The recent emergence of a series of diffusion-aware distillation algorithms [23, 37, 39, 46] has been instrumental in mitigating the computational burden associated with the multi-step inference process of diffusion models. These algorithms can be roughly classified into two categories: trajectory-preserving distillation [23, 37, 39, 46] and trajectory-reformulating distillation [21, 31, 40, 41, 56].

Among these, trajectory-preserving few-step diffusion models naturally introduce a denoising **shortcut**, which allows for flexible skipping within the denoising chain while still ensuring high-quality and consistent image synthesis. As demonstrated in Fig. 3, the integration of these shortcuts into the denoising chain significantly reduces the total number of denoising steps, thereby shortening the length of the denoising chain. This key insight paves the way for efficient and effective end-to-end backpropagation.

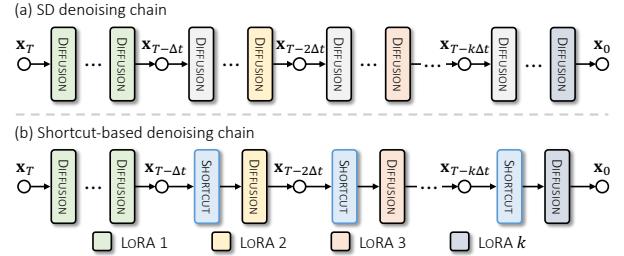


Figure 5. **Illustration of timestep-aware LoRA.** (a) Vanilla SD denoising chain; (b) Shortcut-based denoising chain. In accordance with the interesting time dynamics in the text-to-image diffusion model denoising process revealed by [1], different from existing methods that share the same LoRA parameters at all timesteps, we introduce time-step aware LoRA, which effectively increases the capacity of the diffusion model and accelerates the convergence of training, without increasing the computational cost during the inference stage.

3.3. Shortcut-based fine-tuning

Insight. The naive optimization of Eq. 1 through backpropagation necessitates the construction of the complete denoising chain: $\{x_T, \dots, x_t, \dots, x_{t-\Delta t}, x_{t-\Delta t-1}, \dots, x_{t-2\Delta t-1}, \dots, x_0\}$. This process involves the storage of intermediate activations linked to each neural layer and each denoising timestep within GPU VRAM, which is not feasible due to memory constraints. Furthermore, the typical length of a denoising chain is approximately 50, which results in an overly long backpropagation chain that can lead to issues of gradient explosion.

The key insight of SHORTFT is rooted in the utilization of the trajectory-preserving few-step diffusion model to construct the denoising **shortcut**, significantly reducing the length of the denoising chain. As elaborated in Sec. 3.2, this denoising shortcut enables us to bypass a substantial number of denoising timesteps, leading to create a more streamlined and efficient denoising chain: $\{x_T, \dots, x_t, x_{t-\Delta t}, x_{t-\Delta t-1}, x_{t-2\Delta t-1}, \dots, x_0\}$.

As depicted in Fig. 4, such a design allows for the direct implementation of reward supervision at the early stages of the denoising chain, and facilitates full gradient backpropagation throughout the denoising chain.

Shortcut-based denoising chain. As illustrated in Fig. 5,

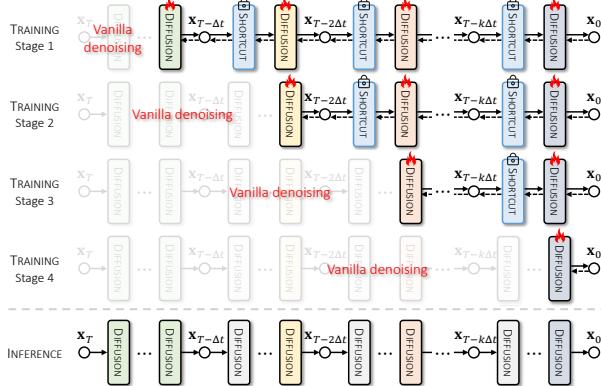


Figure 6. **Illustration of progressive training strategy.** Corresponding to the time-step aware LoRA design, we develop the progressive training strategy, which eliminates the training-inference gap introduced by the shortcut-based fine-tuning.

a vanilla SD denoising chain comprises a sequence of step-by-step denoising operations that transform the input noise \mathbf{z}_T into the output image \mathbf{z}_0 . This process typically necessitates numerous denoising steps. By harnessing the denoising shortcut, we are able to construct a shortcut-based denoising chain and fine-tune the diffusion model through end-to-end backpropagation.

Timestep-aware LoRA. Instead of fine-tuning the full weights of the original diffusion model, Low-Rank Adaptation (LoRA) [19] preserves the weights of the pre-trained model and introduces new low-rank weight matrices alongside the original model weights. The contributions of these matrices are summed to generate the final outputs. Specifically, each linear layer of the UNet of SD is modified from $\mathbf{h} = \mathbf{W}\mathbf{x}$ to $\mathbf{h} = \mathbf{W}\mathbf{x} + \mathbf{B}\mathbf{A}\mathbf{x}$, where $\mathbf{W} \in \mathbb{R}^{d \times d}$, $\mathbf{B} \in \mathbb{R}^{d \times k}$, $\mathbf{A} \in \mathbb{R}^{k \times d}$, and $k \ll d$. LoRA considerably reduces the number of parameters to be optimized, thereby decreasing the memory requirements for fine-tuning.

Moreover, [1] uncovers intriguing temporal dynamics during the denoising process of the text-to-image diffusion model. In the initial sampling stage, the model largely depends on the text prompts to guide the sampling process. As the generation progresses, the model gradually leans on visual features to denoise the image. This indicates that sharing LoRA parameters (standard practice in the existing methods) throughout the entire denoising process may not be optimal and may fail to capture the distinct patterns that emerge during denoising. Therefore, in contrast to [6, 34, 51, 52] which share the same LoRA parameters for all timesteps, we introduce timestep-aware LoRA. This design effectively increases the capacity of the diffusion model and accelerates the convergence of training, without increasing computational cost in the inference phase.

Specifically, as depicted in Fig. 5, we initially divide the entire denoising chain into k segments, $\Delta t = \lfloor \frac{T}{k} \rfloor$. Ex-

cept for the first segment, we assign a corresponding LoRA to the last timestep of each subsequent segment. For the first segment, we adopt the methodology of [6] and share the same LoRA parameters across all timesteps. Notably, a sequence of continuous timesteps lacking LoRA exists in later segments, supporting the denoising shortcuts.

Progressive training strategy. Although the few-step diffusion model is capable of creating the denoising shortcut, bypassing the entire denoising chain, it inherently introduces errors in the output, meaning it still cannot be fully consistent with the output of SD, particularly in fine details, as shown in Fig. 3. This incongruity can lead to a training-inference gap, resulting in suboptimal results. To mitigate this, we design a progressive training strategy.

As illustrated in Fig. 6, in accordance with the design of timestep-aware LoRA, we divide the SHORTFT training process into k stages. For the i -th training stage, we optimize the weights of LoRA i to LoRA k . For the i -th segment and preceding denoising processes, we retain the original denoising chain, while for the denoising processes post the i -th segment, we introduce the denoising shortcut, thereby shortening the depth of the backpropagation chain. Moreover, in line with [6], we also employ the truncated backpropagation technique.

During inference, as displayed in Fig. 6, the denoising shortcut is bypassed, and the original denoising chain is used to generate the final output image.

4. Experiments

4.1. Experimental settings

In our experiments, Stable Diffusion 1.5 serves as the foundational diffusion model. The DDIM schedule [44] is employed to execute 50 steps of denoising, with a classifier-free guidance scale of 7.5.

Shortcut. SHORTFT shortens the denoising chain by bypassing certain steps within the chain. Hence, the selection of few-step diffusion models is focused predominantly on methods that employ trajectory-preserving distillation algorithms. To accommodate the proposed time-aware LoRA, one-step diffusion models are avoided, whose output image quality is also relatively inferior. Specifically, 4-step Hyper-SD [37], distilled from SD 1.5, is utilized to construct the denoising shortcut. The value of k is set to 4, and the timesteps configured for LoRA are $\{761, 501, 261, 1\}$. Consequently, the denoising shortcuts are executed separately between timesteps 741 to 501, timesteps 481 to 261, and timesteps 241 to 1.

Timestep-aware LoRA. As suggested by [6], LoRA is applied to both the feedforward and attention layers in the UNet. The LoRA rank is set to 128. Furthermore, we adopt a stepwise stacking approach, where for LoRA i , we introduce a new LoRA branch on top of LoRA $i - 1$.



Figure 7. **Qualitative comparison on PickScore and HPS v2.** Each image is generated with the same text prompt and random seed for all methods. Our method outperforms existing methods in both text-image alignment and image quality.

Reward functions. The proposed method is evaluated using three reward functions: Human Preference Score v2 (HPS v2) [49], PickScore [24], and Symmetry [51]. HPS v2 and PickScore capture human preference for images based on input prompts, while Symmetry encourages images to have horizontal symmetry features. Different from [51], which utilizes CLIPScore [35] as a regularization term, our experiment amalgamates HPS v2 and PickScore in a ratio of 1:10 to function as a joint regularization term. This approach has demonstrated superior performance in terms of text-image alignment and overall image quality.

Experimental details. All experiments are conducted using 2 A800 GPUs and the AdamW optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and a weight decay of 0.1. SHORTFT is performed with a batch size of 128 and a constant learning rate of 5×10^{-5} . During training, the pre-trained SD parameters are converted to bfloat16 to reduce memory usage, while the LoRA parameters under training remain in float32. Gradient checkpointing is not required.

Datasets. We compare SHORTFT-finetuned diffusion models to those of the state-of-the-art counterparts on the Human Preference Score v2 dataset (HPDv2) [49]. The final reward is computed on the 400 prompts from the test split. Following [51], for fair comparison, we evaluate all methods using the same computational budget. Specifically, all methods are trained for six hours on 2 A800 GPUs.

Method	HPS v2 $^\uparrow$	PickScore $^\uparrow$	Symmetry $^\downarrow$
SD 1.5 [38]	26.91	20.46	0.853
DRaFT-LV [6]	33.13	23.35	0.418
DRTune [51]	32.79	23.22	0.207
SHORTFT	33.88	24.16	0.138

Table 1. **Objective evaluation.** Our method performs over other counterparts, under the same computational cost.

4.2. Qualitative comparison

Fig. 7 and 8 present the quantitative comparison of our results against those of representative methods, including current state-of-the-art techniques, under same text prompts and random seeds. SD 1.5 suffers from low-quality image generation. DRaFT [6], while proficient at managing local image details, struggles to effectively handle global layouts and the optimization of the symmetry reward function. Similarly, DRTune [51] is constrained by gradient inaccuracies, which contribute to instability during training and a deficiency in managing complex semantics. One particular area of weakness is the synthesis of images with specific counts of objects, such as accurately depicting a given number of cows. In contrast, SHORTFT exhibits enhanced capabilities in generating images that are both visually realistic and semantically faithful, outperforming over other counterparts.

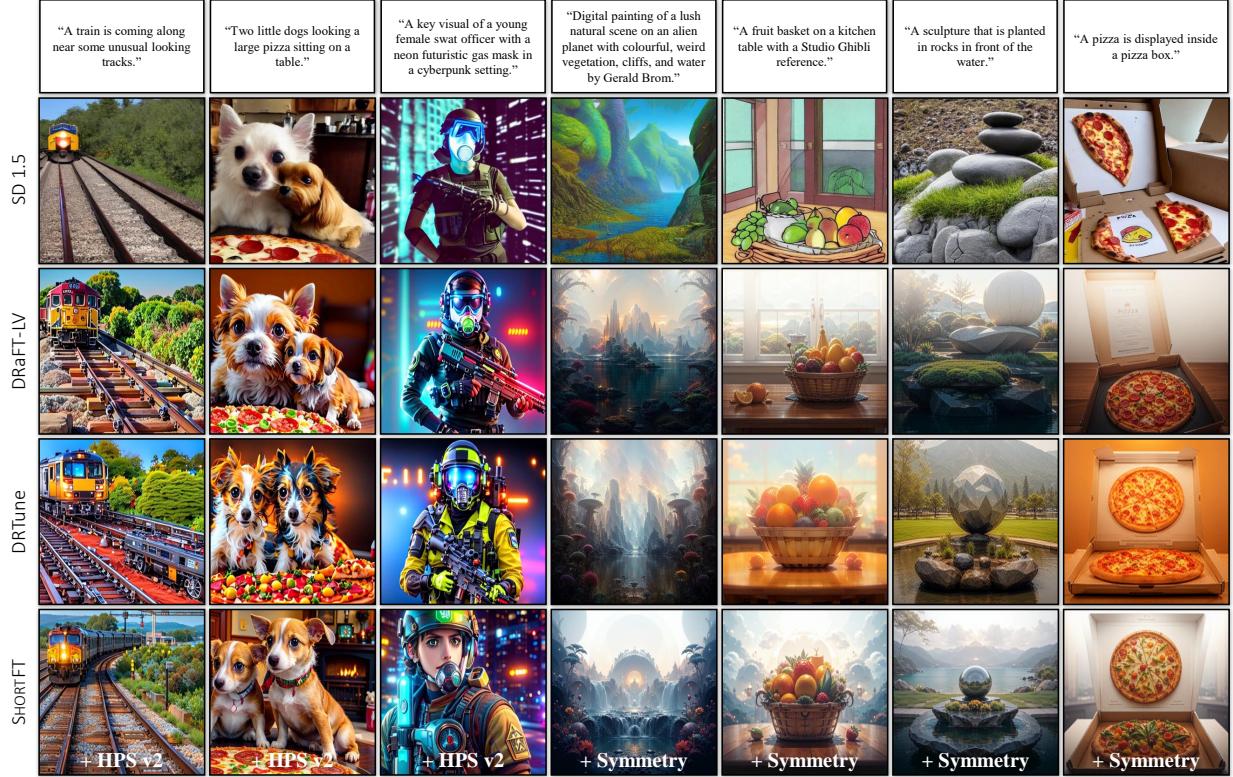


Figure 8. **Qualitative comparison on HPS v2 and Symmetry.** Each image is generated with the same text prompt and random seed for all methods. Our method outperforms existing methods in both text-image alignment and image quality.

	Human	AI (GPT)
SD 1.5 vs. SHORTFT	6.36% 93.64%	5% 95%
DRaFT-LV vs. SHORTFT	35.45% 64.55%	30% 70%
DRTune vs. SHORTFT	20.91% 79.09%	15% 85%

Figure 9. **Human and AI preference evaluation** against current methods. SHORTFT performs over other counterparts.

4.3. Quantitative comparison

Objective evaluation. Table 1 shows the quantitative results achieved on the Human Preference Score v2 benchmark [49], where the proposed method outperforms the other approaches, clearly demonstrating its effectiveness. It is important to highlight that both our method and DR-Tune [51] strategically employ backpropagation of the reward gradient to the initial stages of the denoising chain. This intentional design choice significantly enhances the Symmetry score performance when compared to alternative methods. Despite this, DRTune continue to grapple with the challenge of gradient bias. In contrast, our method significantly mitigates this issue, delivering superior performance.

User study. A subjective user study comprising 11 volunteers is conducted, with five possessing expertise in image processing and the remaining participants having no background in computer vision or graph. Participants are tasked to select the most visually attractive and semantically ac-

curate image among those generated by our method and current state-of-the-art techniques. Each participant has 10 questions for each pair of comparisons. Furthermore, an MLLM-assisted evaluation is employed using GPT-4V. We make 20 queries to GPT-4V for each pair of comparisons. More details are provided in the Appendix. As depicted in Fig. 9, the results exhibit a significant inclination towards SHORTFT in comparison to other techniques.

4.4. More results

SHORTFT, 10k training step. Due to the shorter denoising and backpropagation chains, our method achieves superior performance under the same computational budget. Furthermore, to validate the upper bound of our approach, following the protocol in [6], we conduct the full training process using HPS v2 reward function on HPDv2, comprising 10k training steps, and evaluate it on the corresponding benchmark. The obtained HPS v2 score of 35.97 surpasses the reported score for DRaFT-LV in [6].

Method	Tuning Hyper-SD	Tuning SD 1.5
HPS v2 [†]	32.92	35.97

Table 2. **Objective evaluation** on tuning SD 1.5 and Hyper-SD.

Fine-tuning SD vs. Hyper-SD. As mentioned in Sec. 2.3,

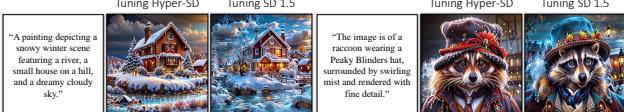


Figure 10. **Qualitative comparison** on tuning SD 1.5 (SHORTFT) and Hyper-SD. Fine-tuning SD 1.5 significantly outperforms fine-tuning Hyper-SD, with the former enjoying more exquisite details.

[26–28] explore fine-tuning few-step diffusion models and have achieved certain successes. However, compared to the foundational model, few-step diffusion models face performance degradation caused by the distillation process. Fine-tuning the few-step diffusion model is actually suboptimal compared to fine-tuning the foundational model. Furthermore, following two strategies, we separately conduct the training processes on the HPDv2 using the HPS v2 reward function. As shown in Table 2 and Fig. 10, fine-tuning SD 1.5 (SHORTFT) significantly outperforms fine-tuning Hyper-SD, with the former enjoying more exquisite details.



Figure 11. **Example results** synthesized by SHORTFT on SD 3.

Fine-tuning SD 3. SHORTFT is an architecture-agnostic fine-tuning strategy, applicable to both UNet-based (SD 1.5) and Transformer-based (SD 3) architectures. As illustrated in Fig. 11, SHORTFT is also capable of mastering SD 3, where SD 3 aligns with HPS v2.



Figure 12. **Generalization to wild text prompts** from Sora. Our method is capable of effectively handling the wild text prompts.

Other reward functions. SHORTFT exhibits remarkable versatility, demonstrating efficacy across a spectrum of reward functions, significantly improving the alignment performance and the quality and fidelity of the output. As shown in Fig. 1 and 12, SHORTFT not only accommodates HPS v2, PickScore, and Symmetry, but also exhibits profi-

Method	HPS v2 \uparrow	PickScore \uparrow	Symmetry \downarrow
w/o T-LoRA	33.46	23.82	0.187
w/o P-Training	33.27	23.97	0.146
SHORTFT	33.88	24.16	0.138

Table 3. **Ablation study** on timestep-aware LoRA.



Figure 13. **Ablation study** on progressive training strategy. The red circle marks the incoherent local details, i.e., unsmooth hair.

ciency in managing Compressibility and Combined reward.

Generalization to wild text prompts. As shown in Fig. 12, we present the qualitative results of prompt generalization. We found that using SHORTFT to fine-tune the model on HPDv2 still enables effective handling of wild text prompts, enhancing the overall quality of the generated images.

4.5. Ablation study

On timestep-aware LoRA. As shown in Table 3, timestep-aware LoRA effectively increases the capacity of the diffusion model and accelerates the convergence of training. Under the same computational cost, the time-aware LoRA achieves better performance.

On progressive training strategy. The progressive training strategy is committed to eliminating the training-inference gap. As shown in Table 3 and Fig. 13, directly integrating the LoRA parameters obtained from training stage 1 into the pre-trained diffusion model results in incoherent local details, resulting in worse results, which can be effectively handled by progressive training strategy.

5. Conclusion

In this paper, we propose a novel Shortcut-based Fine-Tuning (SHORTFT), an advanced technique for aligning diffusion models with reward functions through end-to-end backpropagation in the denoising chain. While existing methods struggle with computational costs and the risk of gradient explosion, SHORTFT leverages shorter denoising chains, markedly improving fine-tuning efficiency and effectiveness. Rigorous evaluations demonstrate that our method can be effectively applied to various reward functions, significantly enhancing alignment performance and surpassing state-of-the-art alternative solutions.

Acknowledgment

This work is supported by the National Key Research and Development Plan (2024YFB3309302).

References

- [1] Yogesh Balaji, Seungjun Nah, Xun Huang, Arash Vahdat, Jiaming Song, Karsten Kreis, Miika Aittala, Timo Aila, Samuli Laine, Bryan Catanzaro, et al. ediffi: Text-to-image diffusion models with an ensemble of expert denoisers. *arXiv preprint arXiv:2211.01324*, 2022. 2, 3, 4, 5
- [2] James Betker, Gabriel Goh, Li Jing, Tim Brooks, Jianfeng Wang, Linjie Li, Long Ouyang, Juntang Zhuang, Joyce Lee, Yufei Guo, et al. Improving image generation with better captions. *Computer Science*. <https://cdn.openai.com/papers/dall-e-3.pdf>, 2(3):8, 2023. 3
- [3] Kevin Black, Michael Janner, Yilun Du, Ilya Kostrikov, and Sergey Levine. Training diffusion models with reinforcement learning. *arXiv preprint arXiv:2305.13301*, 2023. 2, 3
- [4] Chaofeng Chen, Annan Wang, Haoning Wu, Liang Liao, Wenyi Sun, Qiong Yan, and Weisi Lin. Enhancing diffusion models with text-encoder reinforcement learning. In *ECCV*, 2024. 3
- [5] Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. In *NeurIPS*, 2017. 2
- [6] Kevin Clark, Paul Vicol, Kevin Swersky, and David J Fleet. Directly fine-tuning diffusion models on differentiable rewards. In *ICLR*, 2024. 2, 3, 4, 5, 6, 7
- [7] Xiaoliang Dai, Ji Hou, Chih-Yao Ma, Sam Tsai, Jialiang Wang, Rui Wang, Peizhao Zhang, Simon Vandenhende, Xiaofang Wang, Abhimanyu Dubey, et al. Emu: Enhancing image generation models using photogenic needles in a haystack. *arXiv preprint arXiv:2309.15807*, 2023. 3
- [8] Fei Deng, Qifei Wang, Wei Wei, Tingbo Hou, and Matthias Grundmann. Prdp: Proximal reward difference prediction for large-scale reward finetuning of diffusion models. In *CVPR*, 2024. 3
- [9] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. In *NeurIPS*, 2021. 2, 3
- [10] Hanze Dong, Wei Xiong, Deepanshu Goyal, Yihan Zhang, Winnie Chow, Rui Pan, Shizhe Diao, Jipeng Zhang, Kashun Shum, and Tong Zhang. Raft: Reward ranked finetuning for generative foundation model alignment. *arXiv preprint arXiv:2304.06767*, 2023. 2, 3
- [11] Ying Fan, Olivia Watkins, Yuqing Du, Hao Liu, Moonkyung Ryu, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, Kangwook Lee, and Kimin Lee. Reinforcement learning for fine-tuning text-to-image diffusion models. In *NeurIPS*, 2023. 3
- [12] Ying Fan, Olivia Watkins, Yuqing Du, Hao Liu, Moonkyung Ryu, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, Kangwook Lee, and Kimin Lee. Reinforcement learning for fine-tuning text-to-image diffusion models. In *NeurIPS*, 2023. 2
- [13] Shane Griffith, Kaushik Subramanian, Jonathan Scholz, Charles L Isbell, and Andrea L Thomaz. Policy shaping: Integrating human feedback with reinforcement learning. In *NeurIPS*, 2013. 2
- [14] Jiatao Gu, Shuangfei Zhai, Yizhe Zhang, Josh Susskind, and Navdeep Jaitly. Matryoshka diffusion models. *arXiv preprint arXiv:2310.15111*, 2023. 3
- [15] Xiefan Guo, Jinlin Liu, Miaoqiao Cui, Liefeng Bo, and Di Huang. I4vgen: Image as free stepping stone for text-to-video generation. *arXiv preprint arXiv:2406.02230*, 2024.
- [16] Xiefan Guo, Jinlin Liu, Miaoqiao Cui, Jiankai Li, Hongyu Yang, and Di Huang. Initno: Boosting text-to-image diffusion models via initial noise optimization. In *CVPR*, 2024.
- [17] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022. 2, 3
- [18] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *NeurIPS*, 2020. 2, 3
- [19] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. In *ICLR*, 2022. 5
- [20] Kaiyi Huang, Kaiyue Sun, Enze Xie, Zhenguo Li, and Xihui Liu. T2i-compbench: A comprehensive benchmark for open-world compositional text-to-image generation. In *NeurIPS*, 2023. 3
- [21] Minguk Kang, Richard Zhang, Connelly Barnes, Sylvain Paris, Suha Kwak, Jaesik Park, Eli Shechtman, Jun-Yan Zhu, and Taesung Park. Distilling diffusion models into conditional gans. In *ECCV*, 2024. 3, 4
- [22] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *NeurIPS*, 2022. 2, 3
- [23] Dongjun Kim, Chieh-Hsin Lai, Wei-Hsiang Liao, Naoki Murata, Yuhta Takida, Toshimitsu Uesaka, Yutong He, Yuki Mitsufuji, and Stefano Ermon. Consistency trajectory models: Learning probability flow ode trajectory of diffusion. *arXiv preprint arXiv:2310.02279*, 2023. 2, 3, 4
- [24] Yuval Kirstain, Adam Polyak, Uriel Singer, Shahbuland Maitiana, Joe Penna, and Omer Levy. Pick-a-pic: An open dataset of user preferences for text-to-image generation. In *NeurIPS*, 2023. 3, 6
- [25] Kimin Lee, Hao Liu, Moonkyung Ryu, Olivia Watkins, Yuqing Du, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, and Shixiang Shane Gu. Aligning text-to-image models using human feedback. *arXiv preprint arXiv:2302.12192*, 2023. 2, 3
- [26] Jiachen Li, Weixi Feng, Wenhua Chen, and William Yang Wang. Reward guided latent consistency distillation. *arXiv preprint arXiv:2403.11027*, 2024. 3, 8
- [27] Jiachen Li, Weixi Feng, Tsu-Jui Fu, Xinyi Wang, Sugato Basu, Wenhua Chen, and William Yang Wang. T2v-turbo: Breaking the quality bottleneck of video consistency model with mixed reward feedback. *arXiv preprint arXiv:2405.18750*, 2024.
- [28] Jiachen Li, Qian Long, Jian Zheng, Xiaofeng Gao, Robinson Piramuthu, Wenhua Chen, and William Yang Wang. T2v-turbo-v2: Enhancing video generation model post-training through data, reward, and conditional guidance design. *arXiv preprint arXiv:2410.05677*, 2024. 3, 8
- [29] Shufan Li, Konstantinos Kallidromitis, Akash Gokul, Yusuke Kato, and Kazuki Kozuka. Aligning diffusion models by optimizing human utility. In *NeurIPS*, 2024. 3

- [30] Zhanhao Liang, Yuhui Yuan, Shuyang Gu, Bohan Chen, Tiankai Hang, Ji Li, and Liang Zheng. Step-aware preference optimization: Aligning preference with denoising performance at each step. *arXiv preprint arXiv:2406.04314*, 2024. 3
- [31] Shanchuan Lin, Anran Wang, and Xiao Yang. Sdxl-lightning: Progressive adversarial diffusion distillation. *arXiv preprint arXiv:2402.13929*, 2024. 3, 4
- [32] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *ICML*, 2021. 2, 3
- [33] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. In *NeurIPS*, 2022. 2
- [34] Mihir Prabhudesai, Anirudh Goyal, Deepak Pathak, and Katerina Fragkiadaki. Aligning text-to-image diffusion models with reward backpropagation. *arXiv preprint arXiv:2310.03739*, 2023. 2, 3, 4, 5
- [35] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *ICML*, 2021. 6
- [36] Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. In *NeurIPS*, 2023. 3
- [37] Yuxi Ren, Xin Xia, Yanzuo Lu, Jiacheng Zhang, Jie Wu, Pan Xie, Xing Wang, and Xuefeng Xiao. Hyper-sd: Trajectory segmented consistency model for efficient image synthesis. *arXiv preprint arXiv:2404.13686*, 2024. 2, 3, 4, 5
- [38] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, 2022. 2, 3, 6
- [39] Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. In *ICLR*, 2022. 2, 3, 4
- [40] Axel Sauer, Frederic Boesel, Tim Dockhorn, Andreas Blattmann, Patrick Esser, and Robin Rombach. Fast high-resolution image synthesis with latent adversarial diffusion distillation. *arXiv preprint arXiv:2403.12015*, 2024. 3, 4
- [41] Axel Sauer, Dominik Lorenz, Andreas Blattmann, and Robin Rombach. Adversarial diffusion distillation. In *ECCV*, 2024. 3, 4
- [42] Christoph Schuhmann and Romain Beaumont. Laoin aesthetic predictor. <https://laion.ai/blog/laion-aesthetics/>, 2022. 2, 3
- [43] Eyal Segalis, Dani Valevski, Danny Lumen, Yossi Matias, and Yaniv Leviathan. A picture is worth a thousand words: Principled recaptioning improves image generation. *arXiv preprint arXiv:2310.16656*, 2023. 3
- [44] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *ICLR*, 2021. 2, 3, 5
- [45] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *ICLR*, 2021. 2, 3
- [46] Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *ICML*, 2023. 2, 3, 4
- [47] Nisan Stiennon, Long Ouyang, Jeffrey Wu, Daniel Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul F Christiano. Learning to summarize with human feedback. In *NeurIPS*, 2020. 2
- [48] Bram Wallace, Meihua Dang, Rafael Rafailov, Linqi Zhou, Aaron Lou, Senthil Purushwalkam, Stefano Ermon, Caiming Xiong, Shafiq Joty, and Nikhil Naik. Diffusion model alignment using direct preference optimization. In *CVPR*, 2024. 3
- [49] Xiaoshi Wu, Yiming Hao, Keqiang Sun, Yixiong Chen, Feng Zhu, Rui Zhao, and Hongsheng Li. Human preference score v2: A solid benchmark for evaluating human preferences of text-to-image synthesis. *arXiv preprint arXiv:2306.09341*, 2023. 3, 6, 7
- [50] Xiaoshi Wu, Keqiang Sun, Feng Zhu, Rui Zhao, and Hongsheng Li. Human preference score: Better aligning text-to-image models with human preference. In *ICCV*, 2023. 2, 3
- [51] Xiaoshi Wu, Yiming Hao, Manyuan Zhang, Keqiang Sun, Zhaoyang Huang, Guanglu Song, Yu Liu, and Hongsheng Li. Deep reward supervisions for tuning text-to-image diffusion models. In *ECCV*, 2024. 2, 3, 4, 5, 6, 7
- [52] Jiazheng Xu, Xiao Liu, Yuchen Wu, Yuxuan Tong, Qinkai Li, Ming Ding, Jie Tang, and Yuxiao Dong. Imagereward: Learning and evaluating human preferences for text-to-image generation. In *NeurIPS*, 2023. 2, 3, 5
- [53] Zeyue Xue, Guanglu Song, Qiushan Guo, Boxiao Liu, Zhuban Zong, Yu Liu, and Ping Luo. Raphael: Text-to-image generation via large mixture of diffusion paths. In *NeurIPS*, 2023. 3
- [54] Kai Yang, Jian Tao, Jiafei Lyu, Chunjiang Ge, Jiaxin Chen, Weihan Shen, Xiaolong Zhu, and Xiu Li. Using human feedback to fine-tune diffusion models without any reward model. In *CVPR*, 2024. 3
- [55] Shentao Yang, Tianqi Chen, and Mingyuan Zhou. A dense reward view on aligning text-to-image diffusion with preference. *arXiv preprint arXiv:2402.08265*, 2024. 3
- [56] Tianwei Yin, Michaël Gharbi, Richard Zhang, Eli Shechtman, Fredo Durand, William T Freeman, and Taesung Park. One-step diffusion with distribution matching distillation. In *CVPR*, 2024. 3, 4
- [57] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *ICCV*, 2023. 3
- [58] Yinan Zhang, Eric Tzeng, Yilun Du, and Dmitry Kislyuk. Large-scale reinforcement learning for diffusion models. In *ECCV*, 2024. 3
- [59] Daniel M Ziegler, Nisan Stiennon, Jeffrey Wu, Tom B Brown, Alec Radford, Dario Amodei, Paul Christiano, and Geoffrey Irving. Fine-tuning language models from human preferences. *arXiv preprint arXiv:1909.08593*, 2019. 2