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

Supplementary Material

A. Experimental setup

A.1. Reward functions

We provide a brief overview of the reward functions involved in this study as follows:

HPS v2 [10]. Leveraging the prompts from both DiffusionDB [9] and COCO Captions [2] in conjunction with the images synthesized by diverse generative models, human annotators offer subjective preferences for images generated in response to identical prompts. HPS v2 is established upon OpenCLIP-H and undergoes fine-tuning using the collected preference data. Noteworthy is that HPS v2 not only evaluates image quality but also scrutinizes the alignment between generated images and the text prompts.

PickScore [4]. Similar to HPS v2, PickScore also employs OpenCLIP-H fine-tuned on preference data. The difference is that the training data for the latter is collected through a web application, where users can generate pairs of images produced by different diffusion models with different hyperparameters and then rate them.

Symmetry [11]. Symmetry, as a low-level reward function defined in pixel space, is formulated as follows: $\mathcal{R}(\mathbf{x}) = \frac{\|\mathbf{x}-\text{flip}(\mathbf{x})\|_1}{\text{std}(\mathbf{x})}$, where $\text{flip}(\cdot)$ denotes the mirror flip operation, $\text{std}(\cdot)$ calculates the standard deviation of pixel values in the input image. In contrast to the usage employed in [11], where CLIPScore [6] is integrated as a regularization component, our study combines HPS v2 [10] and PickScore [4] in a 1:10 ratio to serve as the regularization term. This approach has exhibited enhanced efficacy in fostering text-image alignment and enhancing the overall quality of images.

Aesthetic [8]. The LAION aesthetic predictor is built on top of a CLIP [6] image encoder, combined with a multilayer perceptron (MLP). It is trained on a manually annotated dataset and can rate a given image on a scale of 1 to 10. Aesthetic only measures image quality, without considering the provided text prompts.

Compressibility [1]. Compressibility evaluates the reconstruction error between images before and after `JPEG` compression, thereby encouraging the diffusion model to synthesize simple images.

A.2. AI preference study

This study introduces AI preference evaluation. Specifically, we utilize the advanced multimodal model GPT-4V as an automatic evaluator, which demonstrates the ability to provide logical evaluations and reasons that align with human preferences. Fig. 1 illustrates the instruction and evaluation process.

B. Additional experimental results

B.1. Additional qualitative comparison

Limited by the fact that DRaFT-LV and DRTune have not been open-sourced to the community. Furthermore, for a more comprehensive comparison, we also provide an unaligned qualitative comparison with the results reported in [11], *i.e.*, using the same text prompts. As shown in Fig. 2, our method achieves high-quality results.

B.2. Scaling and mixing LoRA weights

In alignment with prior works [3, 5], SHORTFT also supports controlling the fine-tuning strength by scaling LoRA weights. This is achieved by the multiplication of the LoRA parameters by a scalar $\alpha < 1$, thereby enhancing the proximity of the fine-tuned parameters to those of the original pre-trained model. Fig. 3 visually depicts the seamless interpolation process between the pre-trained model and the SHORTFT fine-tuned model.

Furthermore, we demonstrate the ability of SHORTFT to interpolate between different reward functions during the inference stage. This is achieved by linearly combining the LoRA parameters trained with different reward functions, as shown in Fig. 4.

Method	SHORTFT	SHORTFT + DRaFT-LV
HPS v2$^{\uparrow}$	35.97	36.27

Table 1. **Objective results** of integrating ShortFT and DRaFT-LV.

B.3. Integration with existing methods

As outlined in Sec. 4 of the main paper, our method retains a subset of timesteps without introducing LoRA, providing an opportunity to integrate with existing methods, *e.g.*, DRaFT-LV [3], leading to improved performance. As discussed in Fig. 2 of the main paper, DRaFT-LV overlooks direct supervision in the early stages of the denoising chain, resulting in suboptimal alignment with text prompts. However, it excels in refining appearances, making it complementary to our approach. Therefore, we present the objective results of integrating SHORTFT with DRaFT-LV on the HPDv2 using the HPS v2 reward function, with SHORTFT and DRaFT-LV being executed sequentially. Table 1 indicates that DRaFT-LV further enhances the performance.

B.4. Additional visual results

Fig. 5 and 6 show more visual results of our approach achieved on HPS v2 and PickScore, respectively.

C. Limitation and discussion.

This study has made significant strides in the field of diffusion model alignment, attributed to the utilization of a trajectory-preserving few-step diffusion model. Nevertheless, the success of this approach is contingent upon advancements in diffusion distillation, as illustrated in Fig. 3 of the main paper. While the current technique, Hyper-SD [7], has exhibited dependable performance, there is scope for enhancement in alignment with teacher models. It is posited that the refinement of diffusion distillation methods will supplement our approach.

Additionally, while distillation versions are common for existing mainstream diffusion models within the community, additional resources are required to obtain the distilled version for other models.

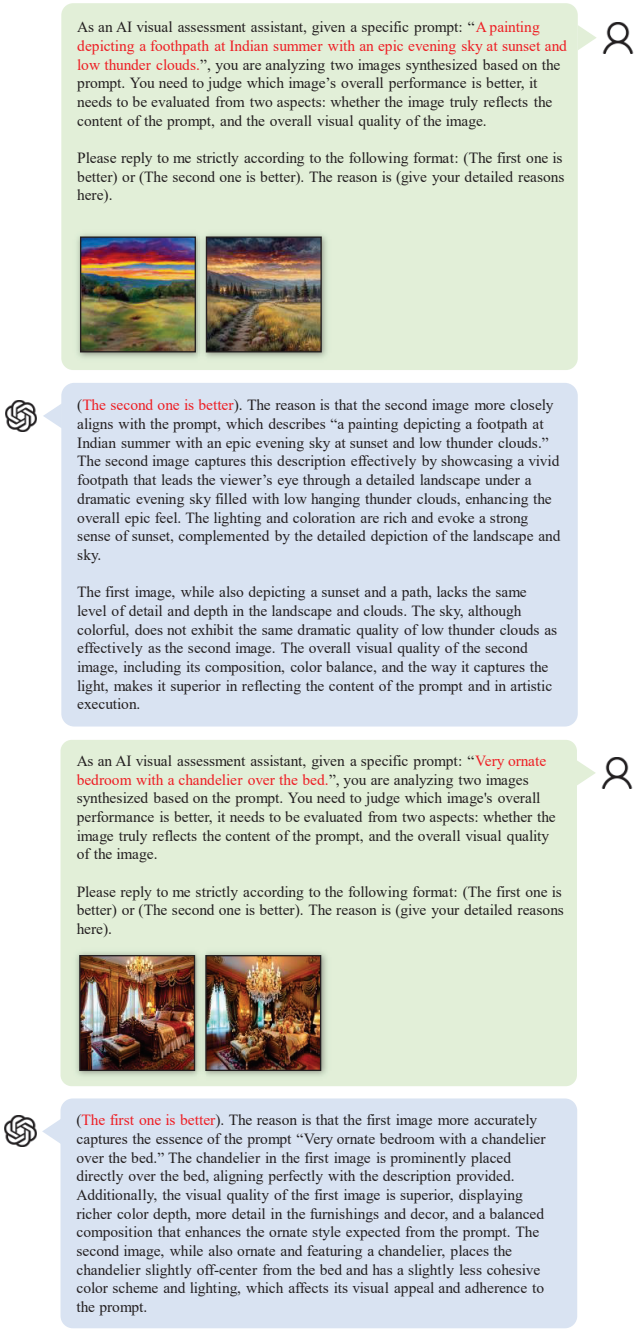

Figure 1. **Illustration of the AI preference study.** We supplement subjective evaluation with the MLLM-assisted assessment.

References

[1] Kevin Black, Michael Janner, Yilun Du, Ilya Kostrikov, and Sergey Levine. Training diffusion models with reinforcement learning. *arXiv preprint arXiv:2305.13301*, 2023. 1

[2] Xinlei Chen, Hao Fang, Tsung-Yi Lin, Ramakrishna Vedantam, Saurabh Gupta, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco captions: Data collection and evaluation server. *arXiv preprint arXiv:1504.00325*, 2015. 1

[3] Kevin Clark, Paul Vicol, Kevin Swersky, and David J Fleet. Directly fine-tuning diffusion models on differentiable rewards. In *ICLR*, 2024. 1

[4] Yuval Kirstain, Adam Polyak, Uriel Singer, Shahbuland Matiana, Joe Penna, and Omer Levy. Pick-a-pic: An open dataset of user preferences for text-to-image generation. In *NeurIPS*, 2023. 1

[5] Mihir Prabhudesai, Anirudh Goyal, Deepak Pathak, and Katerina Fragkiadaki. Aligning text-to-image diffusion models with reward backpropagation. *arXiv preprint arXiv:2310.03739*, 2023. 1

[6] 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. 1

[7] 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

[8] Christoph Schuhmann and Romain Beaumont. Laoin aesthetic predictor. https://laion.ai/blog/laion-aesthetics/, 2022. 1

[9] Zijie J Wang, Evan Montoya, David Munechika, Haoyang Yang, Benjamin Hoover, and Duen Horng Chau. Diffusiondb: A large-scale prompt gallery dataset for text-to-image generative models. In *CVPR*, 2023. 1

[10] 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. 1

[11] 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. 1, 3

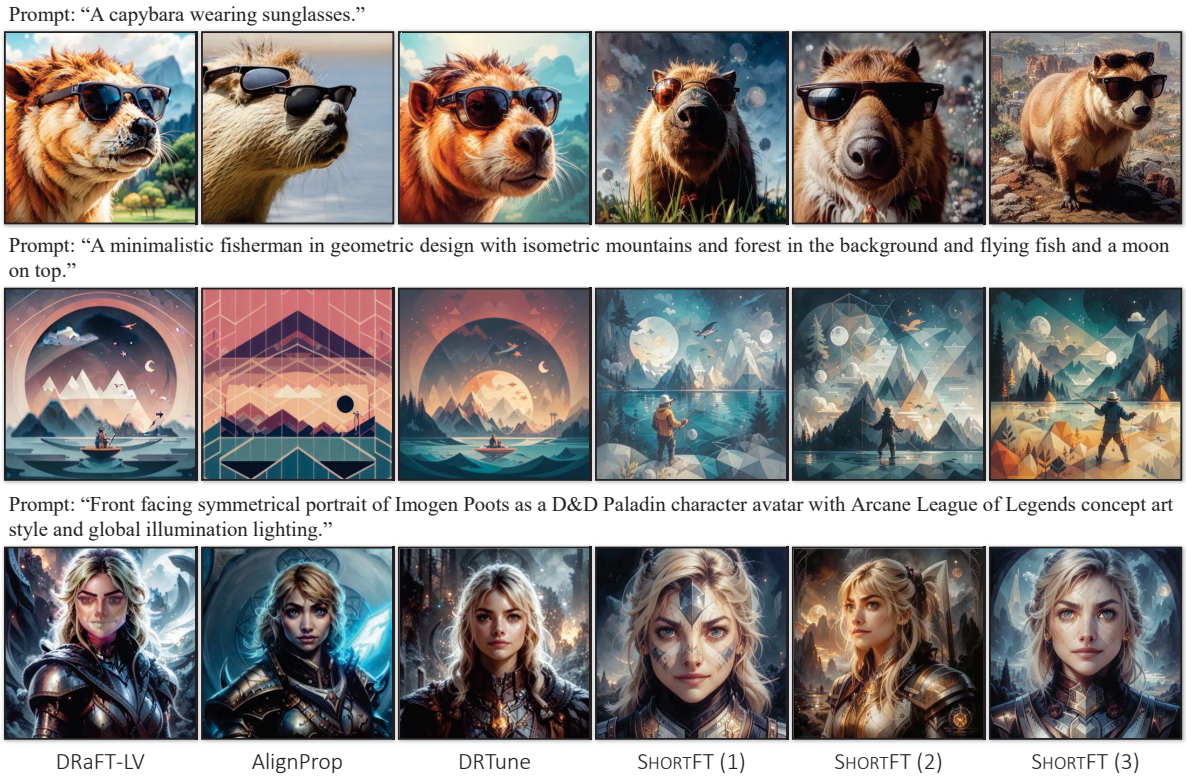

Figure 2. **Qualitative comparison** with results reported in [11]. Our method achieves high-quality results.

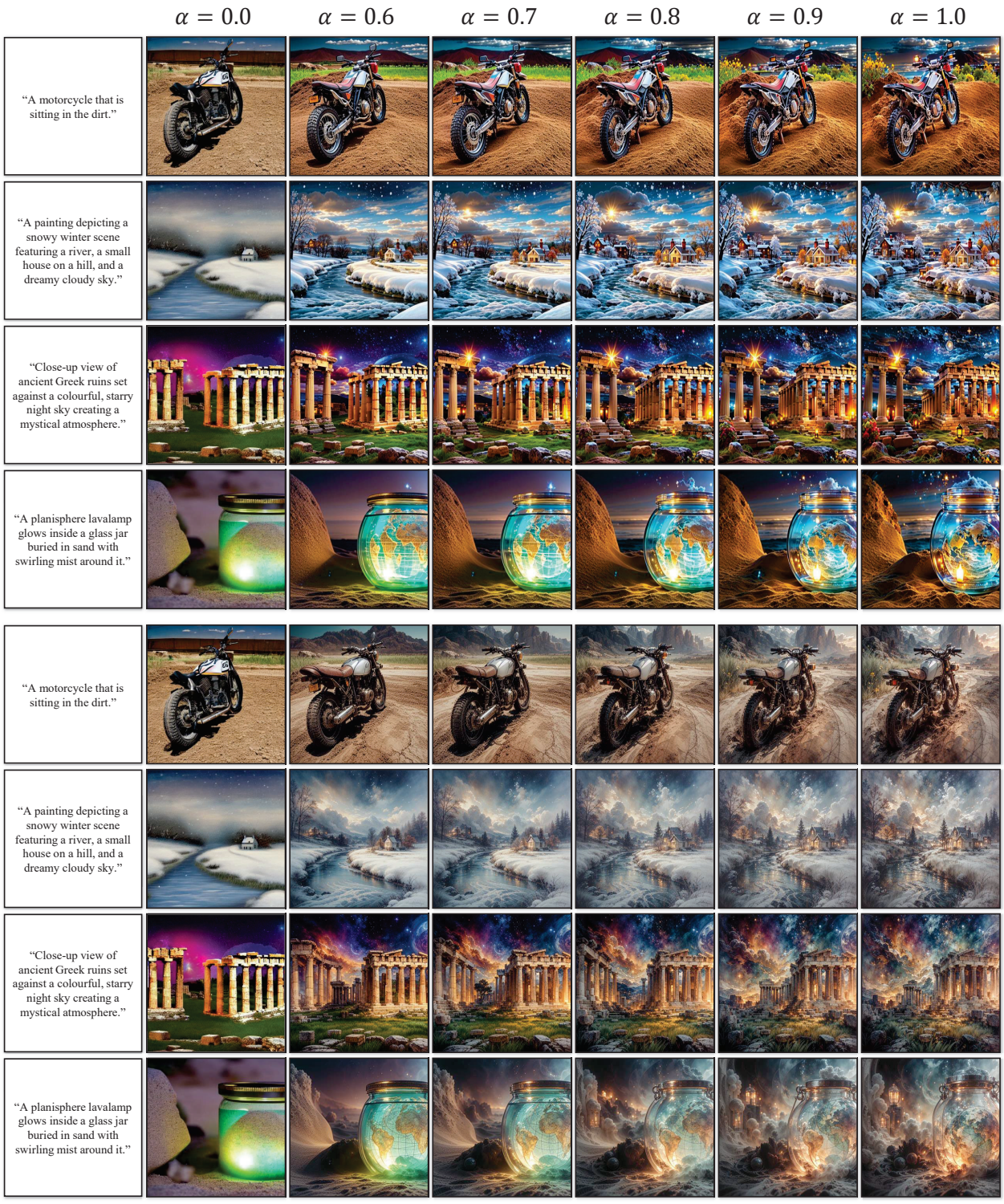

Figure 3. **Visual results** for scaling LoRA parameters. Top: HPS v2; bottom: PickScore.

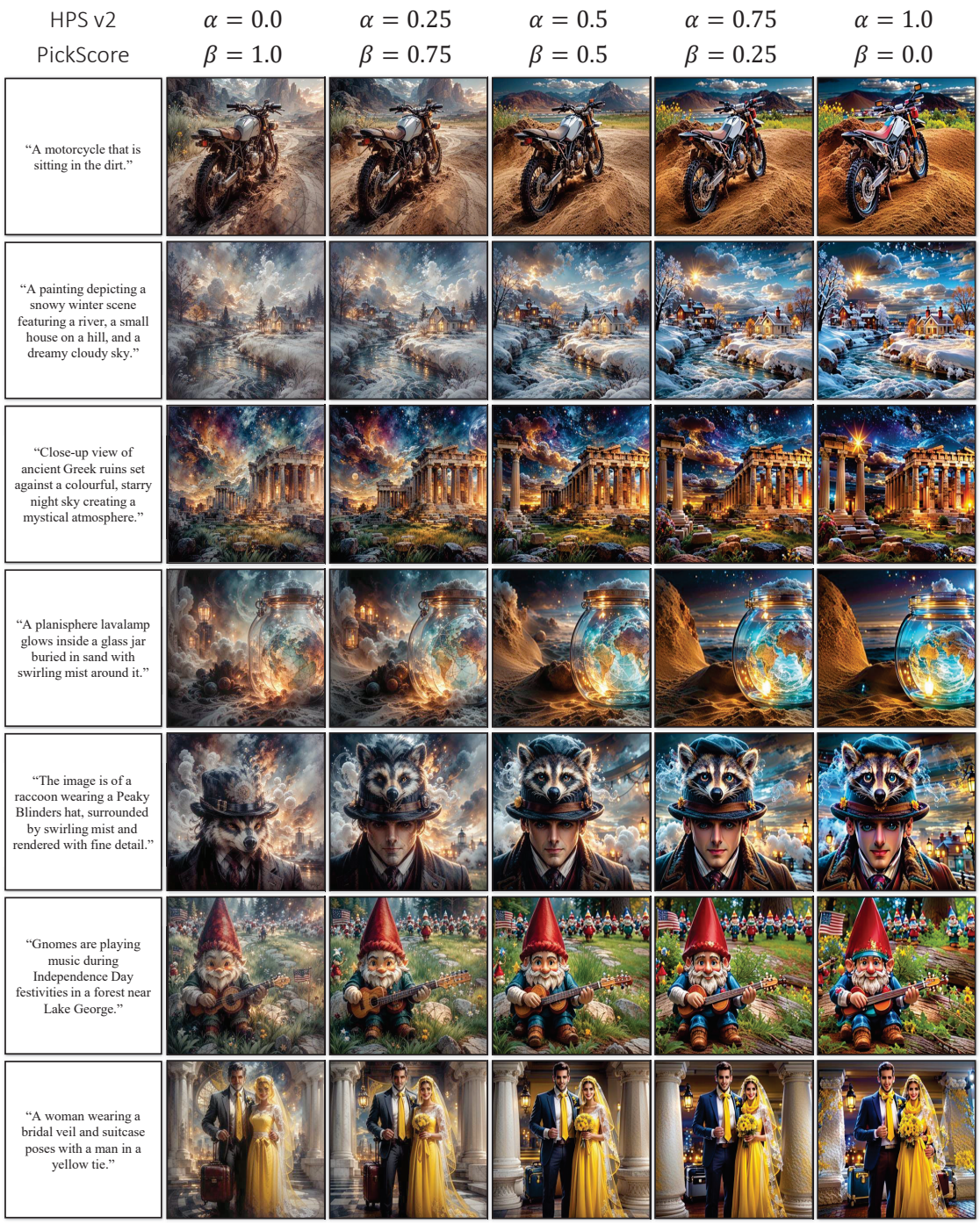

Figure 4. **Visual results** using linear combinations of LoRA parameters adapted for different rewards. Interpolating between LoRA weights allows for smooth transitions between different styles.

Figure 5. **More results synthesized by our proposed SHORTFT using HPS v2.** After training, LoRA weights are scaled down by a factor of 0.75 to reduce reward overfitting.

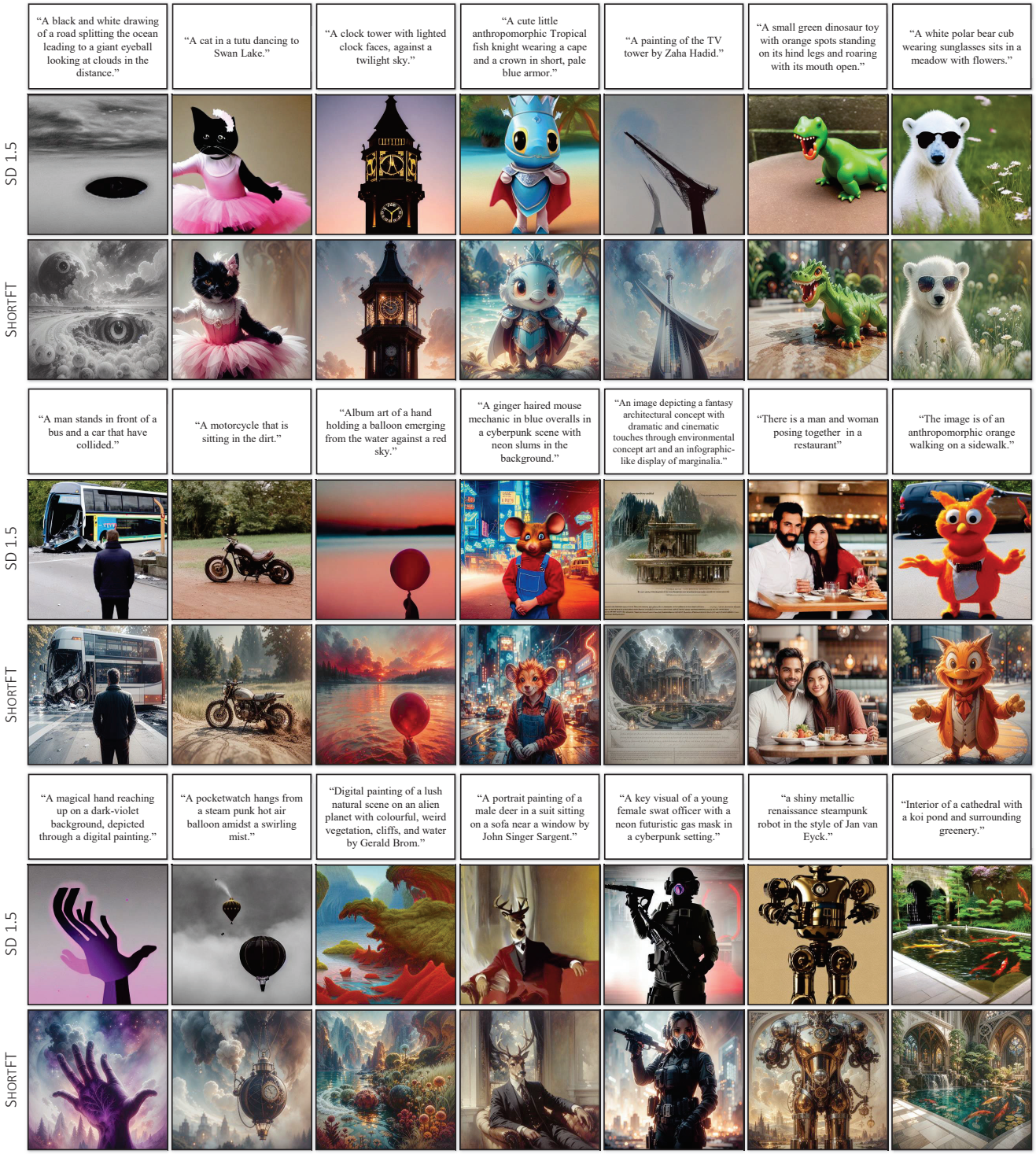

Figure 6. **More results synthesized by our proposed SHORTFT using PickScore.** After training, LoRA weights are scaled down by a factor of 0.75 to reduce reward overfitting.