

Supplementary Materials: TreeReward: Improve Diffusion Model via Tree-Structured Feedback Learning

Anonymous Authors

1 ANALYSIS ON FEEDBACK DATA

The quality of feedback data plays a crucial role in successful feedback learning, as it should align well with human preferences. To assess the overall quality of our curated feedback data, we conduct a manual evaluation of the data collected through our automatic pipeline. We engage 50 human annotators to rate the feedback data pairs based on specific fine-grained dimensions. We calculate the preference consistency between our annotations and human judgments as a measure of data quality. For each feedback dimension, we randomly sample 5000 data pairs and have them rated by human annotators. We consider a data pair to be consistent if the preferred sample in our annotated data pair is also rated as preferred by the human annotators. The consistency ratio is then calculated for each feedback dimension. Tab.1 illustrates the consistency ratios of our curated feedback data. Notably, the curated data exhibit significant consistency with human judgments across all feedback dimensions. In particular, the style and content alignment dimensions achieve consistency ratios of 92% and 88% respectively. The feedback data related to aesthetic aspects demonstrate slightly lower consistency ratios, with 74%, 77%, 80%, and 73% for Tint, Texture, Layout, and Vibe respectively. This variation is reasonable given that the judgment of aesthetic aspects can be subjective and ambiguous, actually, the agreement between different annotators on aesthetic preferences only reaches about 63% as stated in ImageReward[1]. Nonetheless, this feedback data remains highly valuable in guiding the diffusion model towards human preferences. We also show some examples of the collected fine-grained feedback data in Fig.1.

Table 1: The agreement degree between the collected feedback data and human preferences. A high consistency rate represents the high quality of the feedback data, which is in line with human preferences.

Feedback Dimension	Style	Content	Color	Detail	Layout	Lighting
Preference Consistency Rate	92.0	88.0	74.0	77.0	80.0	73.0

2 ANALYSIS ON REWARD MODEL

To demonstrate the effectiveness of TreeReward, we evaluate the reward accuracy of TreeReward. Specifically, we carefully collected and human-annotated 1,000 instances of preference data for each feedback dimension with 10 proficient annotators, including content, and style alignment, color, lighting, detail, and layout aesthetic preference data. In this process, given a prompt, the annotator is requested to select the preferred image from two generated candidates along a particular dimension. Then, we validate our TreeReward by

utilizing TreeReward to predict human preference and take the consistency rate as the proxy of the performance of reward modeling. Note that we utilized the reward predictions derived from the corresponding leaf reward nodes for the preference dataset of different dimensions for the fine-grained reward evaluation. The results are shown in Tab.2. It clearly shows that TreeReward demonstrates a higher reward accuracy across all the fine-grained dimensions. Notably, TreeReward outperforms ImageReward in the text-to-image content alignment dimension by 5.2%. Considering that ImageReward is specifically optimized for text-to-image content alignment, this result highlights the superiority of our approach. Additionally, our method achieves higher reward accuracy in almost all other aesthetic dimensions. Note that aesthetic concepts are subjective and abstract and different people may have opposite preferences, this also poses a challenge for the reward model to learn human preference accurately, which leads to a relatively lower preference prediction accuracy(56% on average) than the style and content dimension.

Table 2: The predicted preference accuracy of fine-grained dimensions of different reward models.

Reward Model	Style	Content	Color	Detail	Layout	Lighting
CLIP Score	72.4	58.9	49.1	52.8	50.6	48.7
BLIP Score	74.8	63.4	49.8	51.2	48.8	50.2
Aesthetic Score	50.2	41.0	54.9	50.3	54.8	53.6
ImageReward	81.2	67.3	51.4	52.4	52.0	52.3
Ours	91.5	72.5	58.3	57.9	56.6	55.4

3 DETAILS ON THE HUMAN EVALUATION

To facilitate the evaluation of the generative model, we developed two human evaluation systems with user interface (UI) to gather human assessment results for the generated outputs. These evaluation systems were designed to provide valuable insights into the performance of the models. The first evaluation system is a global evaluation, where users are presented with two generation results and asked to determine which one is superior, as depicted in Fig.2. This method offers a simple and efficient approach to obtain a reliable assessment of the performance comparison between two models. However, it can only provide a coarse evaluation. To obtain a more detailed performance evaluation, we also designed a fine-grained human evaluation system, as illustrated in Fig.3. In this system, users are not only asked to choose the better result but also to provide a more detailed score indicating the degree of improvement of one result over another based on specific aspects. These scores enable a deeper understanding of the generative performance of the model in diverse aspects. These human evaluation systems, with their respective UI interfaces, offer valuable tools for

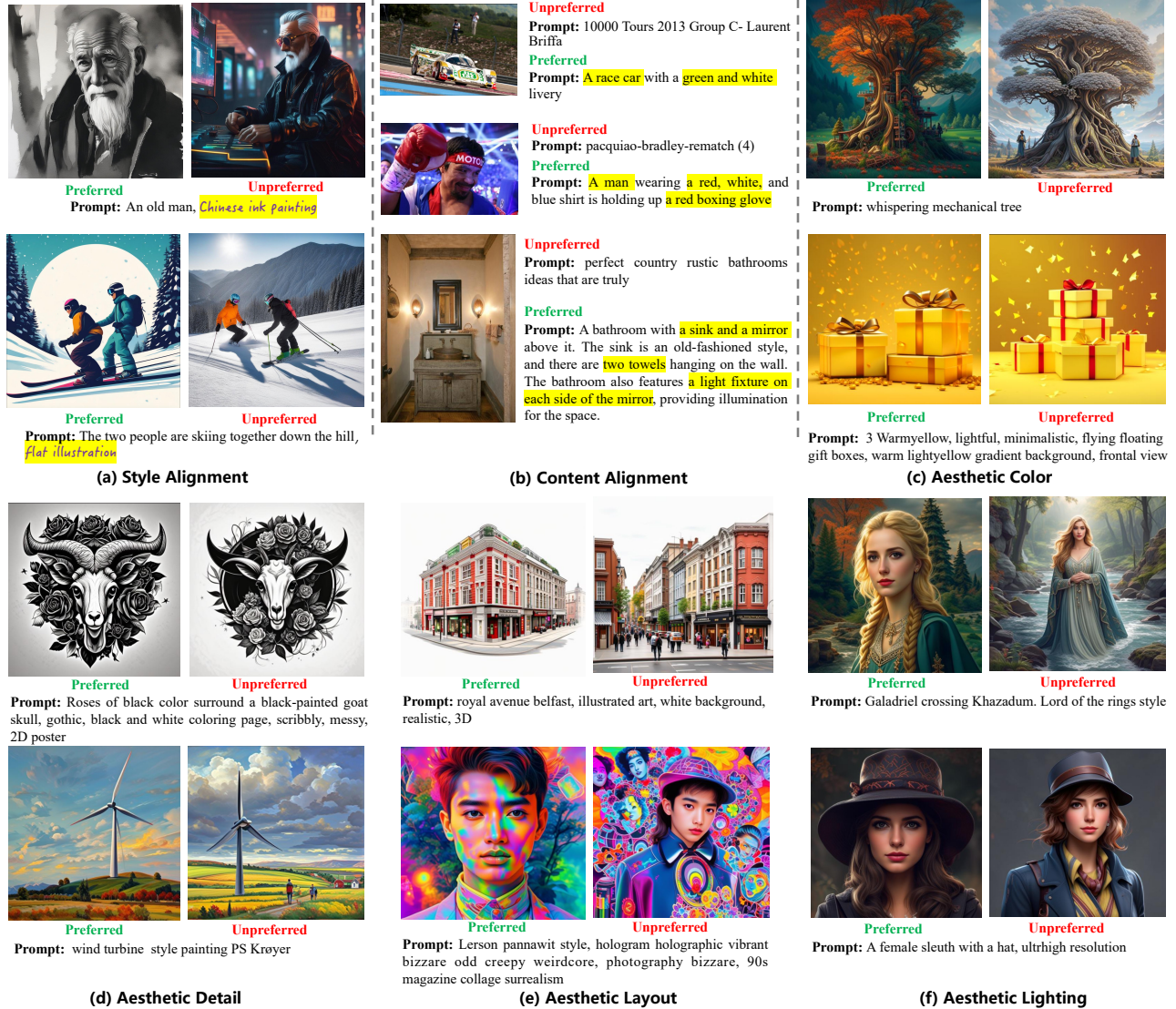


Figure 1: Some examples of the curated fine-grained feedback data, including style alignment, content alignment, aesthetic color, detail, layout, and lighting.

evaluating and analyzing the performance of the generative model in a comprehensive manner.

4 LIMITATION

TreeReward has demonstrated promising outcomes in generating high-quality images. Nevertheless, there exists considerable room for further enhancement:

Tree-Structure Expansion: There is potential for hierarchical tree-structural expansion to refine prompt sensitivity and explore the upper bounds of the generation space.

Benchmark Construction: We are engaged in constructing a unified evaluation benchmark that encompasses more single or fine-grained dimensions for assessment.

Model Expansion: Our proposed methodology holds theoretical applicability to a wide range of models and we plan to conduct experiments on prominent open-source models such as SSD-1B, Kandinsky.

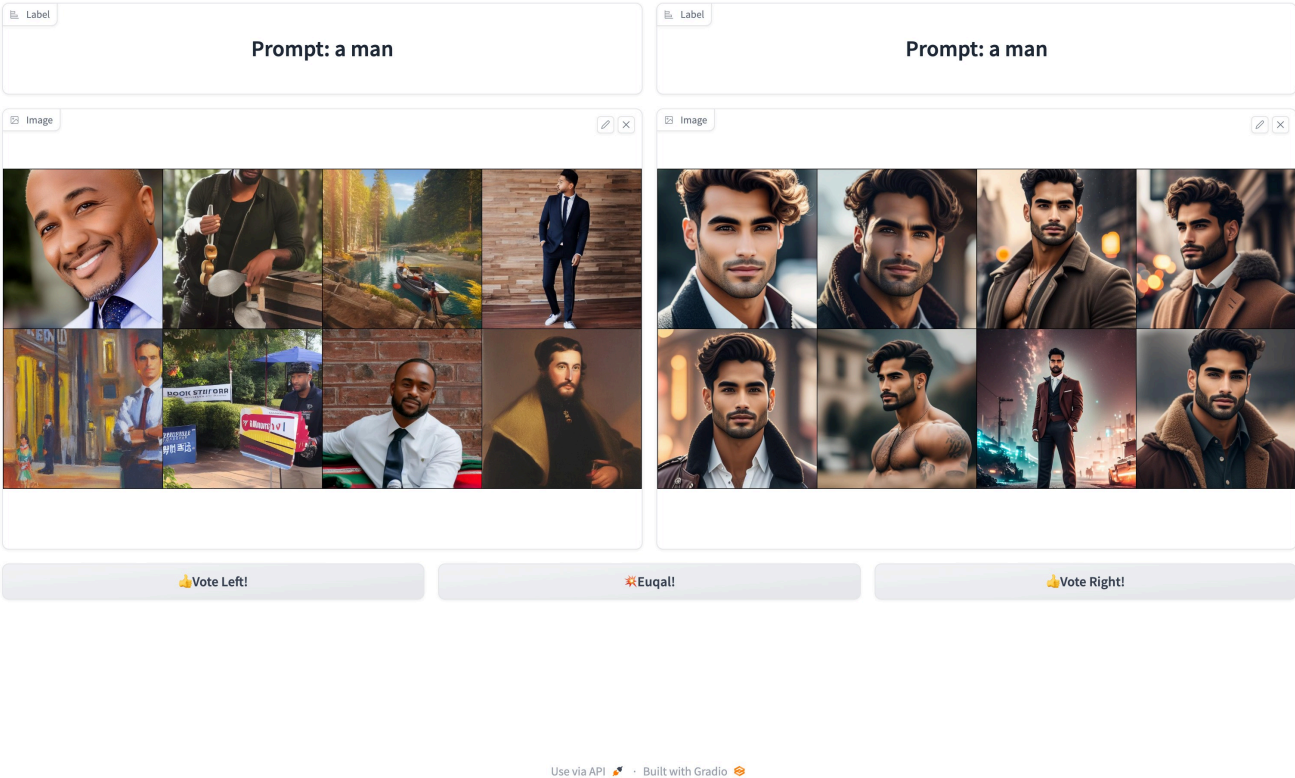


Figure 2: The user interface(UI) of our global human evaluation system. Given the generated results from two models for the same prompt, the users assess whether the two generated results are of the same quality, or otherwise choose the better one.

5 BROADER IMPACT

The impact of our research extends across multiple dimensions. Academically, our method serves as a foundational framework for integrating human feedback with the existing diffusion models. This integration contributes significantly to the advancement of generative models. Practically, our technique holds immense transformative potential across a wide spectrum of industries, including entertainment, portraiture, advertising, and beyond. By providing a means to generate high-quality images with fidelity given the prompt, our approach offers unprecedented opportunities for creativity and innovation. Nevertheless, it is essential to recognize

and address the ethical considerations inherent in the widespread adoption of such technology. The capacity to produce fake images raises legitimate concerns regarding privacy, potential misuse, and the dissemination of false information. Thus, we underscore the critical importance of developing and adhering to stringent ethical guidelines to ensure the responsible and ethical utilization of this groundbreaking technology.

REFERENCES

- [1] Jiazheng Xu, Xiao Liu, Yuchen Wu, Yuxuan Tong, Qinkai Li, Ming Ding, Jie Tang, and Yuxiao Dong. 2023. ImageReward: Learning and Evaluating Human Preferences for Text-to-Image Generation. arXiv:2304.05977 [cs.CV]

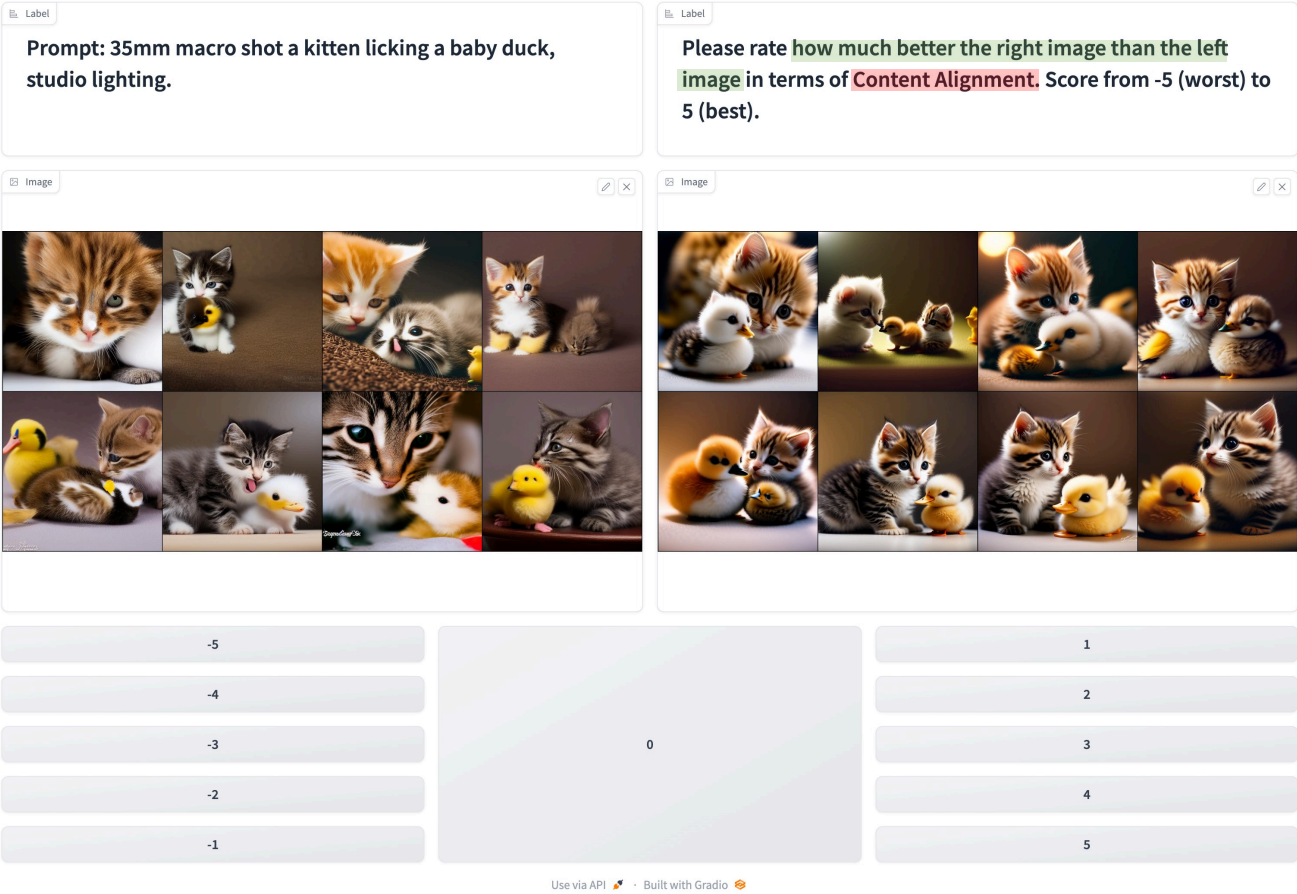


Figure 3: The user interface(UI) of our fine-grained human evaluation system. Given a prompt, and the generated results of the baseline model on the left, the user is asked to evaluate how much better the generated results on the right compare to the results on the left in a particular aspect and give a score from -5 to 5.