# Effective Text-to-Image Alignment with Quality Aware Pair Ranking

**Kunal Singh*** **Mukund Khanna** **Pradeep Moturi**
Fractal AI Research
Mumbai

## Abstract

Fine-tuning techniques such as Reinforcement Learning with Human Feedback (RLHF) and Direct Preference Optimization (DPO) allow us to steer Large Language Models (LLMs) to align better with human preferences. Alignment is equally important in text-to-image generation. Recent adoption of DPO, specifically Diffusion-DPO, for Text-to-Image (T2I) diffusion models has proven to work effectively in improving visual appeal and prompt-image alignment. The mentioned works fine-tune on Pick-a-Pic dataset, consisting of approximately one million image preference pairs, collected via crowdsourcing at scale. However, do all preference pairs contribute equally to alignment fine-tuning? Preferences can be subjective at times and may not always translate into effectively aligning the model. In this work, we investigate the above-mentioned question. We develop a quality metric to rank image preference pairs and achieve effective Diffusion-DPO-based alignment fine-tuning.We show that the SD-1.5 and SDXL models fine-tuned using the top 5.33% of the data perform better both quantitatively and qualitatively than the models fine-tuned on the full dataset. The code is available at this link.

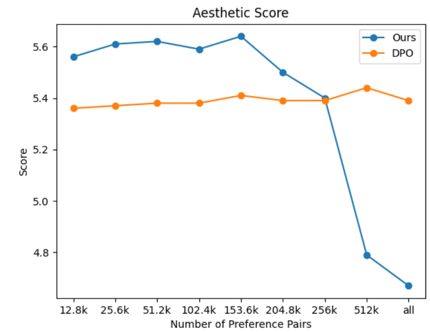
## 1 Introduction

Currently, diffusion-based Text-to-Image (T2I) [3, 4, 11, 14] models are state-of-the-art (SOTA) in image generation. These models are trained in a single stage on a large-scale dataset of images scraped from the internet, enabling them to have huge knowledge. However, their outputs often fail to align with human preferences, as they are not explicitly optimized for this purpose. In contrast, Large Language Models (LLMs) undergo training in two distinct stages: the first stage involves pre-training on large web-scale datasets, while the second stage uses Supervised Fine-tuning (SFT) and Reinforcement Learning based on Human Feedback (RLHF) to align outputs with human preferences. While significant progress has been made in alignment fine-tuning for LLMs, aligning T2I outputs with human preferences remains a difficult challenge.

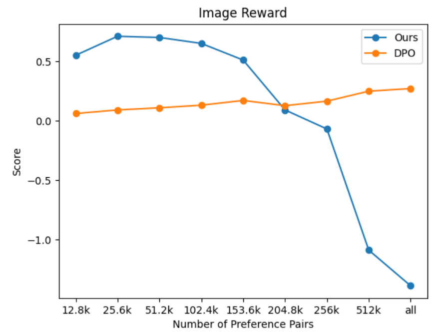
Recent works have begun exploring how to better align T2I models with human preferences. These approaches can be broadly classified into two broad categories – they either use a reward model trained on human preference data to guide the T2I model, or they directly fine-tune the T2I model on pairwise preference data. Reinforcement Learning (RL) based approaches like DRAFT[5], Alignprop[12], ImageReward[19], ReFl[19], do not scale well to large datasets and are highly prone to problems like overfitting and mode collapse. Additionally, training good reward models and using them to fine-tune diffusion models introduces significant operational challenges, as it adds a lot of computational overhead. DPOK[6] and DDPO[2] maximize the scores from the reward model over a set of limited prompts which limits the performance of these methods as the number of prompts increases. DOODL[17] attempts to generate more aesthetically pleasing images by doing iterative

*Corresponding author: Kunal Singh (kunal.singh@fractal.ai)

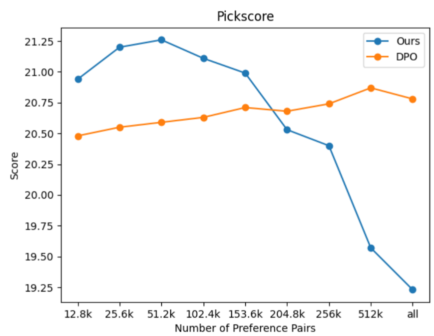
38th Conference on Neural Information Processing Systems (NeurIPS 2024) workshop on Adaptive Foundational Models.



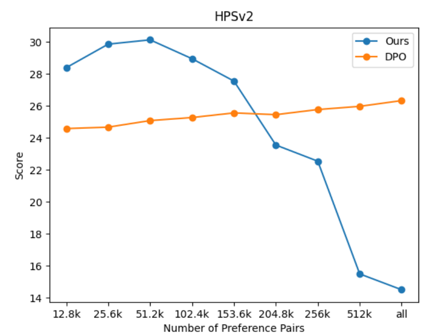

(a) Aesthetic Score vs Number of Preference pairs used for fine-tuning




(b) Image Reward vs Number of Preference pairs used for fine-tuning




(c) Pick Score vs Number of Preference pairs used for fine-tuning




(d) HPS-v2 vs Number of Preference pairs used for fine-tuning

Figure 1: Trend of aesthetic score, Image Reward, PickScore and HPS-v2 while Diffusion-DPO fine-tuning of SD 1.5 on our quality-sorted dataset vs full dataset.

improvements to the generation at run-time. Parrot [9] does multi reward RL finetuning in parallel with a prompt-expansion network.

Taking inspiration from direct preference optimisation (DPO) [13] for aligning LLMs and to overcome challenges with reward modelling based methods, approaches like Diffusion-DPO[16] have emerged, reformulating the loss function to completely remove the reward model and directly fine-tune on pairwise image preference data, which solves the problems of traditional RL-based approaches. In recent works, more preference alignment approaches like Diffusion-KTO[10] and IPO[1] have emerged, building on Diffusion-DPO to further improve diffusion model alignment. Additionally, D3PO[20] suggest creating its own image pairs from a set of prompts and then using a reward model to identify preferred images. Despite all these advances, these approaches still suffer from noisy pairwise preference datasets and over-optimization.

To address these shortcomings, we propose our novel approach — **Effective Text-to-Image Alignment with Quality Aware Pair Ranking**. We introduce a quality metric to assess the quality of a pair of images and the corresponding prompt as a fine-tuning sample. We use this metric to rank all samples from the alignment fine-tuning dataset. We use ranking to prioritise stronger samples over weaker samples during training by fine-tuning on the dataset sorted using this ranking, which we call the Quality-Sorted Dataset (QSD). We observe over 10x improvement in fine-tuning efficiency and demonstrate that over 90% of the samples in Pick-a-Pic v2 [8] dataset send conflicting signals which do more harm than good during RLHF fine-tuning. Finally, we demonstrate through human and AI evaluations that our ranking method improves the performance of SOTA fine-tuning techniques and is preferred by human raters. For brevity, we refer to our approach as DPO-QSD or QSD. Figure 3 (in appendix section) shows the generated image outputs from SDXL base, SDXL-DPO checkpoint fine-tuned on full Pick-a-Pic v2 dataset [8] of approximately 1 million image preference pairs, and SDXL-DPO-QSD fine-tuned on top 50k image preference pairs selected via our method.

## 2 Method: Quality metric for ranking preference pairs

We capture human preferences from online forums, but these judgments are influenced by factors such as malicious intent, varying expertise, and subjectivity. To address this, we focus on pairs that align with overall preferences. Diffusion-KTO selects pairs where the winning image always wins, and the losing image always loses. However, in the Pick-a-Pic dataset [8], less than 5% of images were compared more than five times, and only 25% were compared more than once, making this metric unreliable due to limited comparisons by random users.

We propose a quality metric for pair selection, where a higher score indicates a greater likelihood of the pair being influential. Through experiments, we demonstrate that fine-tuning with higher-quality pairs leads to improved model performance. However, as lower-quality pairs are introduced, performance begins to decline, supporting the importance of ranking image preference pairs.

Consider any paired preference dataset $D = \{(c^1, x_w^1, x_l^1), (c^2, x_w^2, x_l^2), ...., (c^n, x_w^n, x_l^n)\}$, where each sample consists of a caption ($c$), a winning image ($x_w$), and a losing image($x_l$). We use an AI reward model trained to model human preferences to get the probability of the winning image to be winning and the losing image to be losing. We use the HPSv2[18] model that is trained on a expert-reviewed dataset for human preference to output preference for image given the prompt. This preference value will ranges from 0-1, allowing us to interpret them as the probability of the image being preferred. We refer to this model as $\psi$. Now quality $Q$ of each sample pair can be written as

$$Q(c, x_w, x_l) = \psi(x_w/c) * (1 - \psi(x_l/c)) \tag{1}$$

This can be viewed as probability of pair being correct i.e. probability of the winning image being the winning image and the losing image being the losing image. In appendix section A.1, we show the plot of quality scores of all the preference pairs from Pick-a-pic v2 dataset and discuss the segregation of pairs using the plot.

## 3 Experiments

### 3.1 Dataset and Hyper-parameters

We demonstrate the efficacy of our model on the Pick-a-Pic v2 dataset. The dataset contains 1 million rows split into 959.5k rows, 20.5k rows, 20.5k rows of train, validation and test sets respectively. The training set contains approximately 58k distinct captions. We run experiments on SD1.5 and SDXL models. For pairwise preference fine-tuning we use the fine-tuning approach as highlighted in Diffusion-DPO. For both set of experiments we use the ADAMW optimizer. For all SD1.5 and SDXL experiments we use a batch size of 128. All experiments are run on a cluster of 8 NVIDIA 80 GB A100 GPUs. We train at fixed square resolution of 512x512 for SD1.5 and 1024x1024 for SDXL. We train for 1 epoch with a learning rate of $1e^{-4}$ for SD1.5 and $1e^{-5}$ for SDXL. In line with the Diffusion-DPO paper, we use a Beta value of 2000 for SD1.5 and 5000 for SDXL. We do not use any dataset augmentations and keep learning rate constant with no warm-up. For all our experiments we fine-tune using the LoRA approach and use a rank of 64 for both SD1.5 and SDXL.

### 3.2 Evaluation

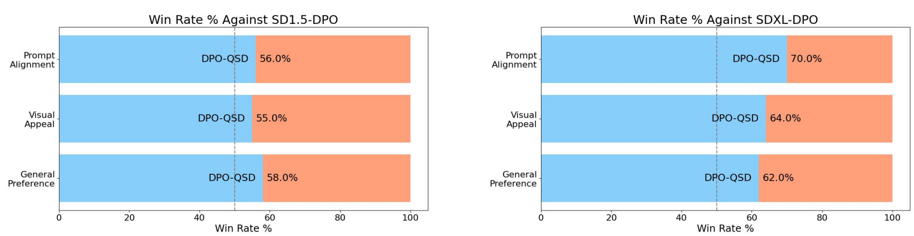
To verify the effectiveness of our approach, we compare against SOTA human preference learning approache, Diffusion-DPO, fine-tuned on the entire training dataset. As we use the LoRA technique, we also fine-tune LoRAs for the SOTA approaches and compare against them. We evaluate all checkpoints on the Pick-a-Pic v2 validation set, which consists of 500 unique prompts. We choose four AI reward models: ImageReward, Pickscore, HPS-v2, and Laion aesthetics classifier. Additionally, we perform a user study to compare our approach to the SOTA Diffusion-DPO. Similar to Diffusion-DPO, we employ reviewers to select the preferred generation under three different criteria: Q1 General Preference (Which image do you prefer given the prompt?), Q2 Visual Appeal (prompt not considered) (Which image is more visually appealing?) Q3 Prompt Alignment (Which image better fits the text description?). 5 responses are collected for each comparison with majority vote (3+) being considered the collective decision. For the user study, we randomly sample 25 prompts from each of the 4 sub-sections of the HPS-v2 test set: photos, anime, paintings and concept-art.

## 4 Results

In Figure 1 for SD1.5 (also, in appendix section, Figure 7 for SDXL), we show that the models fine-tuned using Diffusion-DPO on our QSD significantly outperform the baseline models fine-tuned using Diffusion-DPO on randomly ordered data across 4 key metrics. These results are also presented in Table 1. We also observe a significant improvements in fine-tuning efficiency with our SD1.5 DPO-QSD model and the SDXL DPO-QSD model outperforming the baseline models with just 5.33% of the data. As our fine-tuning data increases, we see a peak in the performance of both models after which the metrics start decreasing or start plateauing. This proves our initial hypothesis that not all fine-tuning pairs are equal and that some fine-tuning data does more harm than good by sending adverse signals. By using only 5.33% of the Pick-a-Pic dataset we achieve our best models, which vastly outperform the baseline models fine-tuned on the full training dataset. This also proves that over 90% of the preference pairs in Pick-a-Pic v2 dataset negatively impact training and can be discarded. Similarly, the user study in Figure 2 shows that our models are preferred by human raters over baseline Diffusion-DPO models. Our SDXL DPO-QSD model is preferred by human annotators 70% of the time in prompt alignment, 64% of the time in visual appeal and 62% of the time in general preference. Similarly, our SD1.5 DPO-QSD model is preferred by human annotators 54% of the time in prompt alignment, 55% of the time in visual appeal and 58% of the time in general preference. In appendix section, we highlight few examples of the high-quality pairs in Figure 5 & low-quality pairs in Figure 6, ranked using our approach.

Table 1: Comparison of our DPO-QSD approach with baseline DPO for SD1.5 and SDXL.

| Method | Aesthetic Score | Image Reward | PickScore | HPSv2 | Samples used |
|---|---|---|---|---|---|
| SD1.5 DPO | 5.39 | 0.27 | 20.78 | 26.34 | 100% |
| **SD1.5 DPO-QSD** | **5.62** | **0.70** | **21.26** | **30.14** | **5.33%** |
| SDXL DPO | 5.97 | 0.85 | 22.20 | 29.40 | 100% |
| **SDXL DPO-QSD** | **6.21** | **1.09** | **22.42** | **31.62** | **5.33%** |




(a) SD1.5-DPO vs SD1.5-DPO

(b) SDXL-DPO-QDS vs SDXL-DPO

Figure 2: SD1.5 & SDXL QSD-DPO outperforms the baseline DPO models in human evaluation.

### 4.1 Effect of Different scoring models

We test the importance of various scoring models by using different reward models to score each pair of images . We run this ablation using LoRA approach for SD1.5 with rank 64, learning rate 1e-4, and a batch size of 128. We keep the quality function constant as $\psi_z(c, x_w) * (1 - \psi_z(c, x_l))$. For this experiment, we try out four different scoring models $\psi_z(c, x_w)$ - HPSv2, Laion aesthetic score predictor, PickScore model, and ImageReward model. To view theses scoring models as probabilities, we standardized the PickScore and Image reward score. We clip the values to +/- 3 and shift scale to 0-1. Aesthetic score is simply divided by 10. As we can observe from Table 2, the model fine-tuned on pairs ranked best using HPS-v2 as the scoring model all other scoring models. ImageReward fails to serve as a good ranking metric for pairs. While the Laion aesthetic predictor shows great improvement in aesthetic score as expected, it fails to show similar improvement across other metrics. HPS-v2 outperforms PickScore and achieves the best results using only 5.33% of the dataset. This ablation reinforces our use of HPS-v2 as a scoring metric.

Table 2: Effect of different scoring models.

| Scoring Method | Aesthetic Score | Image Reward | PickScore | HPSv2 | Samples used |
|---|---|---|---|---|---|
| Baseline Diffusion-DPO | 5.39 | 0.27 | 20.78 | 26.34 | 100% |
| Image Reward | 5.40 | 0.32 | 20.88 | 26.91 | 100% |
| Laion Aesthetics | **5.80** | 0.49 | 21.09 | 27.30 | 16% |
| PickScore | 5.44 | 0.38 | 21.05 | 27.52 | **5.33%** |
| HPS-v2 | 5.62 | **0.70** | **21.26** | **30.14** | **5.33%** |

### 4.2 Effect of LoRA rank

To test the effect of capacity of the LoRA layers and their effect on the model's capability to learn the new information from the dataset, we run experiments with different dimensions of the LoRA layers. Specifically, we want to see how the performance of the model and the fine-tuning efficiency varies with our QSD dataset as we vary the LoRA rank. We run this experiment using SD1.5 as the base model with a learning rate of 1e-4 and a batch size of 128. To this end, we fine-tune with three different LoRA ranks - 32, 64 and 256. For comparison with the baseline, we fine-tune dpo models with the same hyper-parameters and ranks. We present the results in Table 3. As we can observe, we achieve the best results with rank 256 LoRA; however, the improvements over rank 64 are minimal. Therefore, we decide to use rank 64 for our main results. The key observation is that despite the capacity of the LoRA model we get the best fine-tuning efficiency with just 5.33% of the data.

Table 3: Effect of LoRA rank on training efficiency and model performance.

| Model | Rank | Aesthetic Score | Image Reward | PickScore | HPSv2 | Samples used |
|---|---|---|---|---|---|---|
| Diffusion-DPO | 32 | 5.43 | 0.25 | 20.93 | 26.39 | 100% |
| DPO-QSD | 32 | **5.58** | **0.68** | **21.24** | **29.82** | **5.33%** |
| Diffusion-DPO | 64 | 5.39 | 0.27 | 20.78 | 26.34 | 100% |
| DPO-QSD | 64 | **5.62** | **0.70** | **21.26** | **30.14** | **5.33%** |
| Diffusion-DPO | 256 | 5.42 | 0.35 | 20.91 | 26.65 | 100% |
| DPO-QSD | 256 | **5.66** | **0.70** | **21.24** | **30.06** | **5.33%** |

We also make an attempt to study the efficacy of our approach on couple of more fine-tuning methods apart from DPO, using their implementations present in Diffusers [15] in appendix section A.2.

## 5 Conclusion

In this paper, we address the problem of optimal fine-tuning of diffusion models to better align them with human preferences. Unlike previous approaches, we solve this problem by introducing a quality metric that prioritizes high-quality preference pairs and fine-tune in a sorted fashion on this dataset. We demonstrate that our data ranking strategy significantly enhances diffusion model alignment, achieving superior results across multiple AI-based metrics and human evaluators. Our experiments show that models fine-tuned with less than top 10% of the Pick-a-Pick v2 dataset outperform baseline models in both quantitative metrics and human preference evaluations. We run multiple ablations to showcase the effectiveness of our data ranking approach across multiple methods. We validate our initial hypothesis that not all preference pairs contribute equally, and fine-tuning on the entire dataset can be detrimental. By applying our fine-tuning strategy alongside early stopping, one can significantly enhance training efficiency, leading to a more robust & powerful model. **Limitations & Ethics**: We verified our method on pick-a-pic crowd sourced dataset collected from anonymous users whose decisions might be effected by various factors. Any T2I generation poses ethical risks, including the potential for harmful, biased, or explicit content due to web-collected data & humans biases. Efforts to mitigate these risks include diverse labeling & safety filtering during development.

## References


[1] M. G. Azar, M. Rowland, B. Piot, D. Guo, D. Calandriello, M. Valko, and R. Munos. A general theoretical paradigm to understand learning from human preferences, 2023. URL `https://arxiv.org/abs/2310.12036`.

[2] K. Black, M. Janner, Y. Du, I. Kostrikov, and S. Levine. Training diffusion models with reinforcement learning. 2023.

[3] J. Chen, J. Yu, C. Ge, L. Yao, E. Xie, Y. Wu, Z. Wang, J. Kwok, P. Luo, H. Lu, et al. Pixart-$\alpha$: Fast training of diffusion transformer for photorealistic text-to-image synthesis. *arXiv preprint arXiv:2310.00426*, 2023.

[4] J. Chen, C. Ge, E. Xie, Y. Wu, L. Yao, X. Ren, Z. Wang, P. Luo, H. Lu, and Z. Li. Pixart-\sigma: Weak-to-strong training of diffusion transformer for 4k text-to-image generation. *arXiv preprint arXiv:2403.04692*, 2024.

[5] K. Clark, P. Vicol, K. Swersky, and D. J. Fleet. Directly fine-tuning diffusion models on differentiable rewards, 2024. URL `https://arxiv.org/abs/2309.17400`.

[6] Y. Fan, O. Watkins, Y. Du, H. Liu, M. Ryu, C. Boutilier, P. Abbeel, M. Ghavamzadeh, K. Lee, and K. Lee. Dpok: Reinforcement learning for fine-tuning text-to-image diffusion models, 2023. URL `https://arxiv.org/abs/2305.16381`.

[7] J. Hong, N. Lee, and J. Thorne. Orpo: Monolithic preference optimization without reference model, 2024. URL `https://arxiv.org/abs/2403.07691`.

[8] Y. Kirstain, A. Polyak, U. Singer, S. Matiana, J. Penna, and O. Levy. Pick-a-pic: An open dataset of user preferences for text-to-image generation. 2023.

[9] S. H. Lee, Y. Li, J. Ke, I. Yoo, H. Zhang, J. Yu, Q. Wang, F. Deng, G. Entis, J. He, G. Li, S. Kim, I. Essa, and F. Yang. Parrot: Pareto-optimal multi-reward reinforcement learning framework for text-to-image generation, 2024. URL `https://arxiv.org/abs/2401.05675`.

[10] S. Li, K. Kallidromitis, A. Gokul, Y. Kato, and K. Kozuka. Aligning diffusion models by optimizing human utility, 2024.

[11] D. Podell, Z. English, K. Lacey, A. Blattmann, T. Dockhorn, J. Müller, J. Penna, and R. Rombach. Sdxl: Improving latent diffusion models for high-resolution image synthesis, 2023. URL `https://arxiv.org/abs/2307.01952`.

[12] M. Prabhudesai, A. Goyal, D. Pathak, and K. Fragkiadaki. Aligning text-to-image diffusion models with reward backpropagation, 2023.

[13] R. Rafailov, A. Sharma, E. Mitchell, S. Ermon, C. D. Manning, and C. Finn. Direct preference optimization: Your language model is secretly a reward model, 2024. URL `https://arxiv.org/abs/2305.18290`.

[14] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. High-resolution image synthesis with latent diffusion models, 2021.

[15] P. von Platen, S. Patil, A. Lozhkov, P. Cuenca, N. Lambert, K. Rasul, M. Davaadorj, D. Nair, S. Paul, W. Berman, Y. Xu, S. Liu, and T. Wolf. Diffusers: State-of-the-art diffusion models. `https://github.com/huggingface/diffusers`, 2022.

[16] B. Wallace, M. Dang, R. Rafailov, L. Zhou, A. Lou, S. Purushwalkam, S. Ermon, C. Xiong, S. Joty, and N. Naik. Diffusion model alignment using direct preference optimization, 2023.

[17] B. Wallace, A. Gokul, S. Ermon, and N. Naik. End-to-end diffusion latent optimization improves classifier guidance, 2023.

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

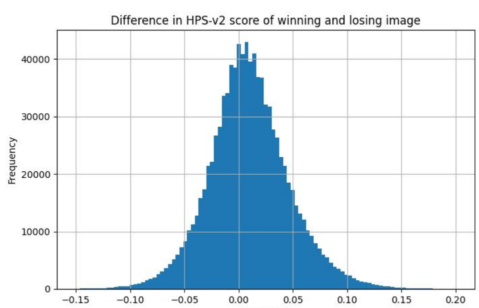
[19] J. Xu, X. Liu, Y. Wu, Y. Tong, Q. Li, M. Ding, J. Tang, and Y. Dong. Imagereward: Learning and evaluating human preferences for text-to-image generation, 2023.

[20] K. Yang, J. Tao, J. Lyu, C. Ge, J. Chen, Q. Li, W. Shen, X. Zhu, and X. Li. Using human feedback to fine-tune diffusion models without any reward model. *arXiv preprint arXiv:2311.13231*, 2023.

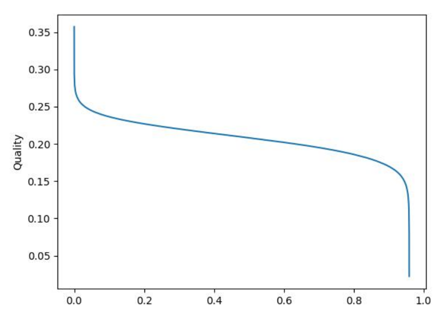
[21] Y. Zhao, R. Joshi, T. Liu, M. Khalman, M. Saleh, and P. J. Liu. Slic-hf: Sequence likelihood calibration with human feedback, 2023. URL `https://arxiv.org/abs/2305.10425`.

# A Appendix / supplemental material

Figure 3: Top to Bottom: *SDXL-DPO-QSD, SDXL-DPO, SDXL* Prompts: *(1) A smiling beautiful sorceress wearing a high necked blue suit surrounded by swirling rainbow aurora, hyper-realistic, cinematic, post-production (2) Concept art of a mythical sky alligator with wings, nature documentary (3) A galaxy-colored figurine is floating over the sea at sunset, photorealistic (4) close up headshot, steampunk middle-aged man, slick hair big grin in front of gigantic clocktower, pencil sketch (5) A swirling, multicolored portal emerges from the depths of an ocean of coffee, with waves of the rich liquid gently rippling outward. The portal engulfs a coffee cup, which serves as a gateway to a fantastical dimension. The surrounding digital art landscape reflects the colors of the portal, creating an alluring scene of endless possibilities.*




(a) Difference in HPSv2 scores (which can be viewed as probability of being preferred) of the winning image and the losing image.




(b) Quality metric plotted in a sorted order for preference pairs in the Pick-a-Pic dataset.

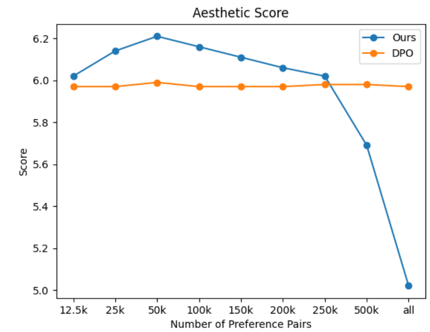
Figure 4: Left - plot of difference in HPSv2 sores for Pick-a-Pic train dataset, Right - plot of quality metric on Y-axis with the sorted dataset index on X-axis.


Figure 5: Examples of good pairs ranked best using our method. *Top row: Winning Image, Bottom row: losing image*. The winning images of good samples have better prompt adherence, aesthetic score and are more preferable to humans. Caption from left to right: *(1) A closeup portrait of a playful maid, undercut hair, apron, amazing body, pronounced feminine features, kitchen, freckles, flirting with camera, (2) A nun holding a sign that says repent, (3) Roman emperor, photo, palace background, (4) A rabbit in a 3 piece suit, sitting in a cafe. Hyper Realistic, ultra realistic, 8k, (5)a painting of a woman with an owl on her shoulder, james gurney and andreas rocha, owl princess with crown, also known as artemis or selene, wlop and sakimichan, detaild, portrait character design, falcon, portrait of modern darna, crowned, golden goddess, white witch, by Johannes Helgeson, goddess of travel*

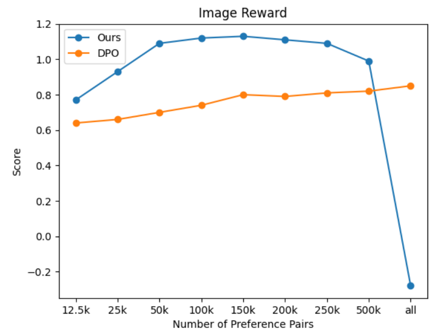
Figure 6: Examples of bad pairs identified by our method. *Top row: Winning Image, Bottom row: losing image*. As can be observed, in these pairs the losing image is better in some quality like aesthetics or prompt adherence over the winning image. Caption from left to right: *(1)a little faery floating in the style of dan hipp, (2) cat wearing a hat, (3) "Hello world" text, space, planets style, (4) face close up woman Jean-Baptiste Monge, watercolour and ink, intricate details, a masterpiece, dynamic backlight, (5) Design a logo for a modern, high-end medical clinic that specializes in personalized, holistic healthcare. The clinic is called "C" and focuses on improving patients' overall well-being through nutrition, exercise, and mental health support. The logo should be simple, sleek, and convey a sense of warmth and approachability while still exuding professionalism and expertise*

### A.1 Quality

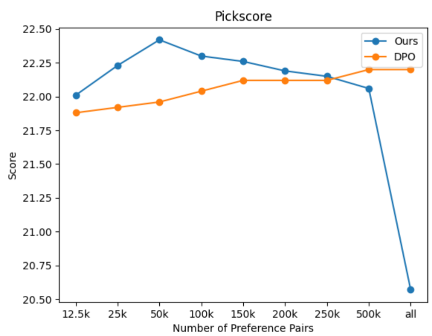
In Figure 4b, we see a sharp decrease in quality score for the initial 100k pairs, followed by a gradual decline for the majority of the dataset, and finally, another sharp drop towards the end, where the samples are of the poorest quality. This plot illustrates that the dataset has good samples where the winning image is clearly better, average samples where the preference is more subjective and bad samples where the reward model does not agree with human labels.



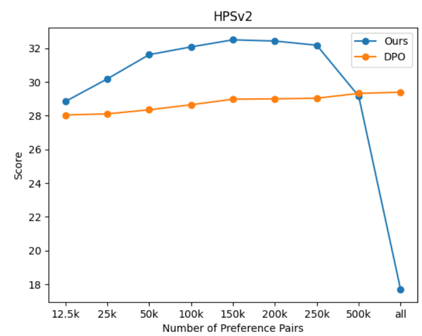

(a) Aesthetic Score vs Number of preference pairs used for fine-tuning




(b) Image Reward vs Number of preference pairs used for fine-tuning




(c) Pick Score vs Number of preference pairs used for fine-tuning




(d) HPS-v2 vs Number of preference pairs used for fine-tuning

Figure 7: Trend of aesthetic score, Image Reward, PickScore and HPS-v2 while Diffusion-DPO fine-tuning of SDXL on our quality-sorted dataset vs full dataset.

### A.2 Efficacy with different fine-tuning methods

We fine-tune the base model using different fine-tuning methods to show that our QSD is effective in improving performance across different fine-tuning approaches. For all experiments, we fine-tune the baseline on the train dataset with random ordering, while our approach uses the quality sorted dataset. We experiment with the loss function of Diffusion ORPO and loss function defined in SLIC-HF[21], we use their implementations from diffusers[15]. We run this ablation using LoRA approach for SD1.5 with rank 64, a batch size of 128, and a learning rate of 1e-4. For Diffusion-ORPO inspired from ORPO[7], we use a learning rate of 1e-3 for baseline model and our model as well. We use the quality metric as described in the methodology section. For these experiments, we select the best-performing model and present the results in table 4. As we can observe, our approach performs considerably better than the baseline across both the methods. Moreover, our approach achieves these results while using only the top 5.33% of the data in case of SLIC-HF and top 10.6% of the data for ORPO, demonstrating over a 10x gain in fine-tuning efficiency. This ablation proves that our pair ranking method improves performance across different fine-tuning paradigms and is not limited to the Diffusion-DPO loss formulation. We believe that the loss in efficiency for Diffusion-ORPO stems from the inclusion of the mean squared error loss of the winning image in the overall loss function, which dominates the other loss terms

Table 4: Efficacy of our ranking method on different fine-tuning paradigms using Pick-a-Pic dataset.

| Method | Aesthetic Score | Image Reward | PickScore | HPSv2 | Samples used |
|---|---|---|---|---|---|
| SLIC-HF baseline | 5.45 | 0.33 | 20.93 | 26.71 | 100% |
| **SLIC-HF-QSD** | **5.69** | **0.72** | **21.24** | **29.65** | **5.33%** |
| ORPO baseline | 5.51 | 0.30 | 20.57 | 26.97 | 100% |
| **ORPO-QSD** | **5.60** | **0.60** | **20.80** | **28.25** | **10.6%** |

## NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Yes we clearly highlight the claims that we intend to prove in our abstract and we run thorough experiments to establish those claims.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Explained in Conclusion

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

Justification: The scope of the paper is purely experimental and we present empirical results to prove our hypothesis. No theoretical proofs are required.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Yes we have clearly explained in the paper how to reproduce all our experiments. Additionally we have open sourced all the relevant code required to reproduce our exact experiments and results.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: Yes we have fully open sourced all the code to run our exact experiments and achieve the same results. The exact command and environment is also shared along with the code base.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: Yes we have specified all the details of data splits, hyperparameters, optimizers and evaluation settings necessary to understand the results.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No]

Justification: No we do not report error bars as the computational cost of running evaluations and doing user study was extremely high. To keep the results as consistent as possible we use the same seed for all generations and use multiple annotators to reduce the variance that comes from human subjectivity.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Yes we do describe how many resources we used to train our models and we also mention for how long we train our models for.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification:Yes we have reviewed the NeurIPS code of Ethics and our paper conforms with respect to everything mentioned in it.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: Explained in Conclusion

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.

- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: We do not release any new dataset or models.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: Yes all the creators and original owners are properly referenced. We have clearly mentioned the versions of the assets we used. All assets used in the paper are fully open source.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.

- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: Yes we release our code base to reproduce the experiments. All assets used have an open source license.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [Yes]

    Justification: We share the full set of instructions in our paper.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: This is not relevant to our paper.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.