Calibri: Enhancing Diffusion Transformers via Parameter-Efficient Calibration

Supplementary Material

Supplementary Material Structure

This supplementary document is organized as follows:

1. Section A elaborates on the limitations of the proposed methodology, providing a detailed analysis.
2. Section B analyzes the diversity of images generated by the method, both before and after incorporating the Calibri technique.
3. Section C explains the rationale behind the chosen reward model, highlighting its impact on the system's performance.
4. Section D discusses the motivation for using the CMA-ES approach as the parameter search method, justifying its effectiveness.

A. Limitations

Our calibration coefficients selection method, *Calibri*, leverages a reward model [6] as its objective function. Reward models are trained to approximate user preferences for generated images, which enables *Calibri* to optimize the selection process effectively.

However, despite substantial advancements in reward modeling in recent years, current reward models still exhibit notable limitations. Specifically, they often demonstrate insufficient sensitivity to generation artifacts such as anatomical inconsistencies—examples include extra limbs, distorted fingers, and other visually unrealistic features, as illustrated in Figure 1.

These shortcomings in reward models can impact the performance of *Calibri*, resulting in the selection of a suboptimal set of coefficients. Addressing these limitations is crucial for further improving the robustness and overall effectiveness of the calibration process.

We anticipate that ongoing advancements in reward modeling techniques will mitigate these issues and significantly enhance their sensitivity to such artifacts, ultimately improving the performance of *Calibri* in future iterations.

B. Generation diversity

Optimizing diffusion models using reward models often leads to a reduction in generation diversity, as highlighted by recent findings [7]. Since *Calibri* employs a reward model as its optimization objective, it is important to evaluate how this approach affects generation diversity. In Table 1, we present a comparison of generation diversity between the original model (SD-3.5M) [1] and models optimized by *Calibri* and Flow-GRPO [5].

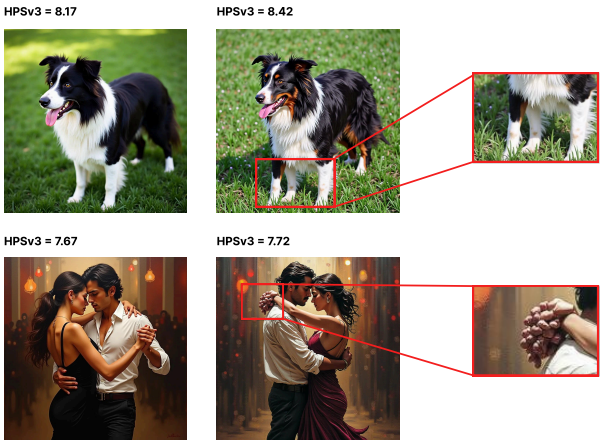


Figure 1. Limitations of modern reward models.

The results demonstrate that the generation diversity of SD-3.5M with 40 inference timesteps remains comparable to that of the model optimized by *Calibri*, which achieves comparable diversity while requiring only 15 inference timesteps. Importantly, despite its reduced inference time, the model optimized by *Calibri* exhibits significantly higher generation quality compared to the original model.

In contrast, Flow-GRPO reduces generation diversity from 0.20 to 0.15 and fails to accelerate inference time. Furthermore, when *Calibri* is applied to the model already optimized by Flow-GRPO, it does not introduce any further changes in generation diversity. This result underscores the efficiency and robustness of *Calibri* in preserving diversity while enhancing generation quality and reducing inference time.

Table 1. Comparison of Generation Diversity for SD-3.5M [1] Optimized by *Calibri* and Flow-GRPO [5].

Flow-GRPO	Calibri	HPSv3	PickScore	Q-Align	Dino Diversity	NFE
✗	✗	11.15	22.40	4.74	0.20 ± 0.06	80
	✗	9.5	22.04	4.51	0.25 ± 0.08	30
	PickScore	**12.47**	**23.13**	**4.91**	0.20 ± 0.06	30
PickScore	✗	12.67	23.78	**4.92**	0.15 ± 0.06	80
	✗	12.514	23.76	4.91	0.15 ± 0.05	30
	PickScore	**12.96**	**23.93**	4.85	0.15 ± 0.05	30

C. Different Rewards

To evaluate the performance of *Calibri* across different objectives, we conducted experiments using various reward

Figure 2. Illustration of *Calibri* with different rewards as objective.

Table 2. Quantitative comparison of *Calibri* layer scale on Flux [4] across different reward models.

Calibri	HPSv3	IR	Q-Align	PickScore	NFE
✗	11.41	1.15	4.85	22.88	30
HPSv3	**13.41**	**1.24**	**4.90**	23.07	15
ImageReward	11.06	1.17	4.70	22.47	15
Q-Align	11.65	1.0	4.89	22.36	15
PickScore	13.34	1.2	4.89	**23.24**	15

models as optimization objectives. The results are summarized in Table 2 and visually presented in Figure 2.

Calibration using the HPSv3 [6] reward model achieved the most significant quality improvement across all metrics, while the Pickscore [3] reward exhibited similarly strong performance. Notably, we observed that calibrating with the most effective reward model not only enhances the target metric but also yields substantial improvements across other metrics. This indicates that *Calibri* is not designed as a reward hacking method tailored to specific objectives, but rather as a general-purpose technique for improving overall generation quality.

D. CMA-ES vs gradient-based parameter search

Gradient-based algorithms are commonly employed for alignment of diffusion models via reward maximization. However, their application to diffusion models presents challenges due to the incompatibility of reward models with noisy latent spaces and generated images, necessitating repeated inference steps during the generation process for accurate reward computation. A recent advancement, Flow-GRPO [5], addresses this issue by reframing the diffusion process as a Markov Decision Process (MDP), enabling reward-driven training in such models. In this section, we evaluate the performance of CMA-ES and compare it to Flow-GRPO, while also providing additional analysis of CMA-ES training dynamics.

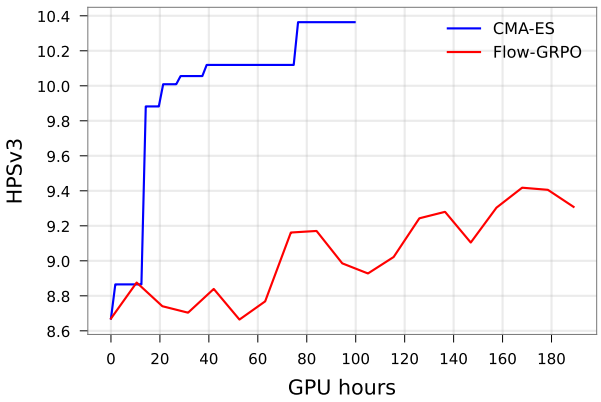


Figure 3. Comparison of CMA-ES and Flow-GRPO performance in optimizing *Calibri* coefficients.

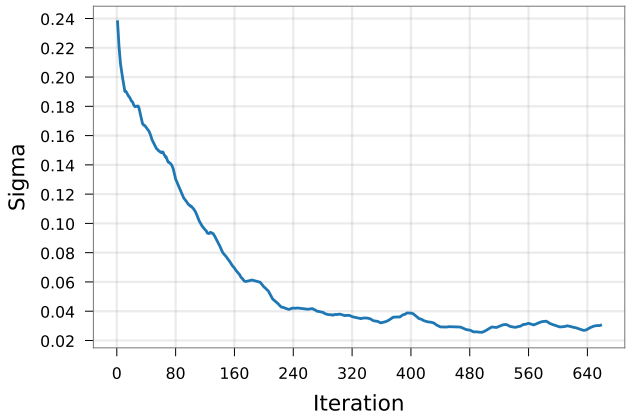
To compare CMA-ES with Flow-GRPO, we trained *Calibri* with layer scale on FLUX using these two optimizers, following the main experimental setup with evaluation during training on the T2I-Compbench++ [2] test prompts. For Flow-GRPO, we adopted the default hyperparameter configuration provided for Flux [4]. The evaluation curves obtained during training are shown in Figure 3 and demonstrate that CMA-ES is substantially more efficient than Flow-GRPO.

Additionally, we present detailed insights into CMA-ES training dynamics in Figure 4. Our analysis indicates that the *Calibri* coefficients converge effectively during training, with optimization reaching a plateau as evidenced by the stabilization of sigma and the stagnation of improvements in the training curve. These observations suggest that training with CMA-ES can be terminated once this convergence behavior is observed, optimizing computational resources without compromising performance.

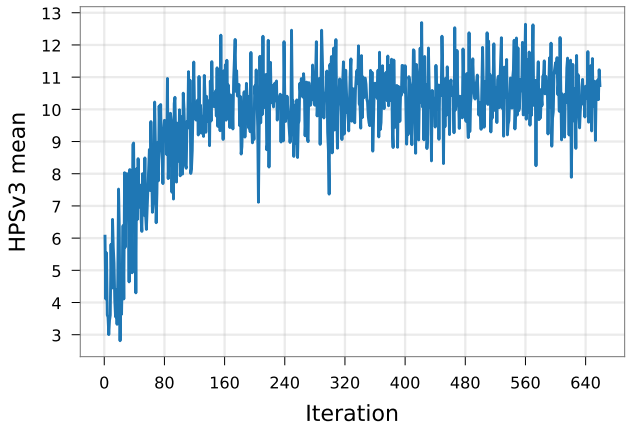
References

[1] Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, et al. Scaling rectified flow transformers for high-resolution image synthesis. In *Forty-first international conference on machine learning*, 2024. 1

[2] Kaiyi Huang, Chengqi Duan, Kaiyue Sun, Enze Xie, Zhenguo



(a) Sigma decrease during training and indicates when the training can be stopped.



(b) Training curve with the mean reward on buckets.

Figure 4. CMA-ES algorithm optimizes *Calibri* coefficients for layer scale FLUX.

Li, and Xihui Liu. T2i-compbench++: An enhanced and comprehensive benchmark for compositional text-to-image generation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47:3563–3579, 2024. 2

[3] 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. *Advances in neural information processing systems*, 36:36652–36663, 2023. 2

[4] Black Forest Labs. Flux. https://github.com/black-forest-labs/flux, 2024. 2

[5] Jie Liu, Gongye Liu, Jiajun Liang, Yangguang Li, Jiaheng Liu, Xintao Wang, Pengfei Wan, Di Zhang, and Wanli Ouyang. Flow-grpo: Training flow matching models via online rl. *arXiv preprint arXiv:2505.05470*, 2025. 1, 2

[6] Yuhang Ma, Xiaoshi Wu, Keqiang Sun, and Hongsheng Li. Hpsv3: Towards wide-spectrum human preference score. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15086–15095, 2025. 1, 2

[7] Dmitrii Sorokin, Maksim Nakhodnov, Andrey Kuznetsov, and Aibek Alanov. Imagerefl: Balancing quality and diversity in human-aligned diffusion models. *arXiv preprint arXiv:2505.22569*, 2025. 1