TempFlow-GRPO: When Timing Matters for GRPO in Flow Models

Authors: Xiaoxuan He, Siming Fu, Yuke Zhao, Wanli Li, Jian Yang, Dacheng Yin, Fengyun Rao, Bo Zhang

https://github.com/Shredded-Pork/TempFlow-GRPO

ICLR 2026 Presentation

Motivation

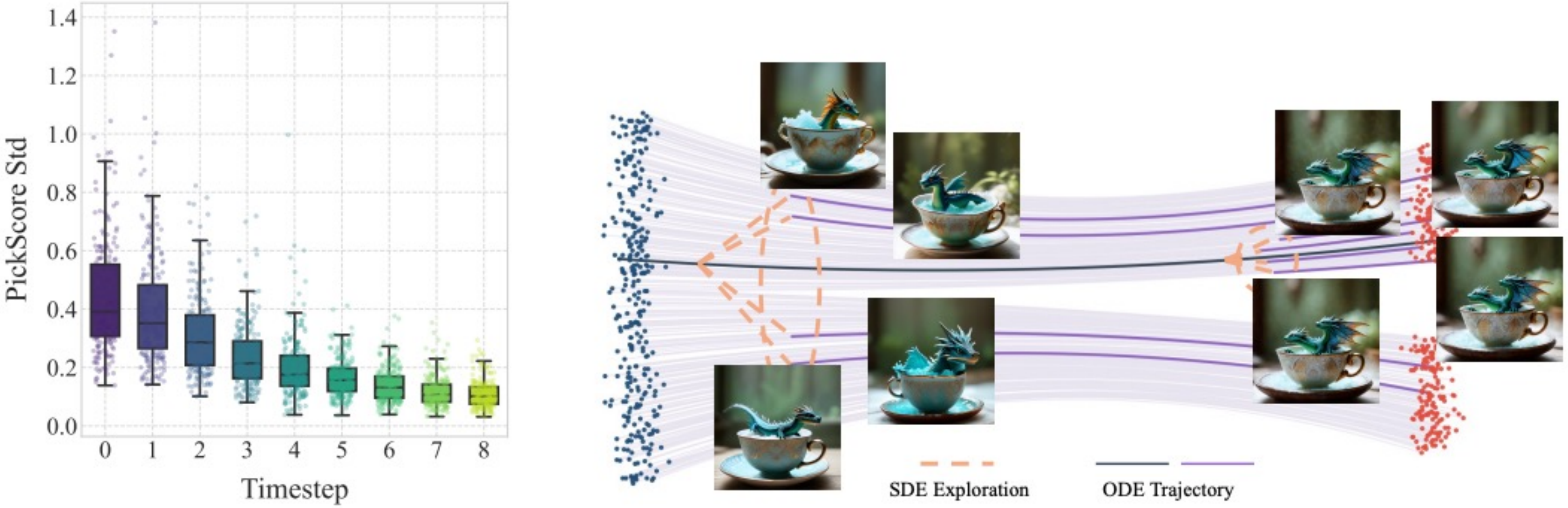


Figure 2: (Left) Reward Variance Analysis: We plot the standard deviation of PickScore at each denoising step for 200 prompts, per prompt group size is 24. The results, obtained via applying SDE at only one step, reveal that reward variance is highest in the initial steps, indicating that early-stage interventions are most impactful for exploration. (Right) Method Illustration: By branching a stochastic (SDE) exploration from a specific, known state on a deterministic (ODE) trajectory, the resulting difference in the final reward can be unambiguously attributed to the exploration action taken at that precise branching point.

Method

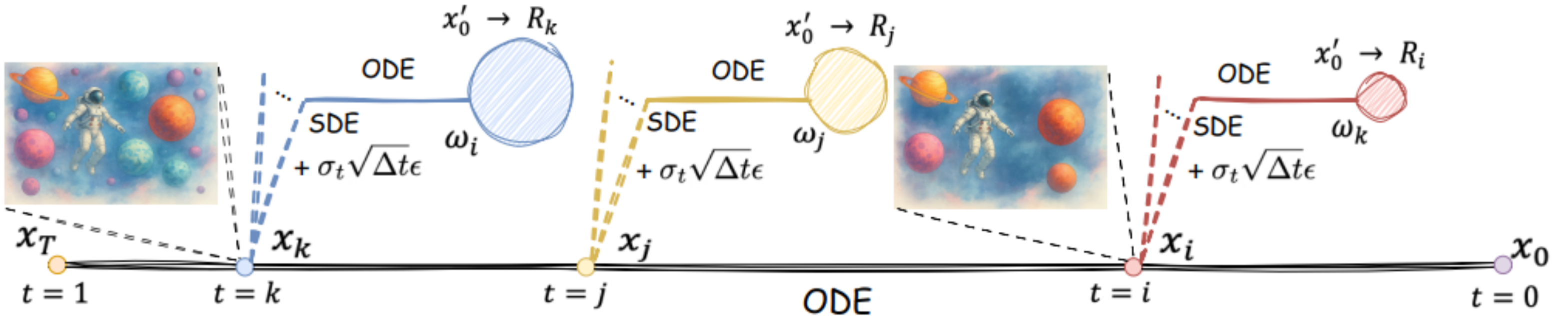


Figure 4: **Overview of TempFlow-GRPO Framework.** Our method performs trajectory branching by switching from ODE to SDE sampling at selected timesteps (t=k, j, i), injecting noise $\sigma_t\sqrt{\Delta t}\epsilon$ to create exploratory branches. Each branch generates a distinct outcome with reward R_i, enabling precise credit assignment. The framework applies noise-aware weighting where $\omega_i > \omega_j > \omega_k$, prioritizing optimization at high-noise early stages (larger circles) over low-noise refinement phases (smaller circles), aligning learning intensity with each timestep's intrinsic exploration capacity. We visualize the model's learning process as an astronaut exploring unknown planets: in early stages , the model explores vast possibility spaces with high uncertainty, while later stages involve focused navigation toward the final destination.

Analysis

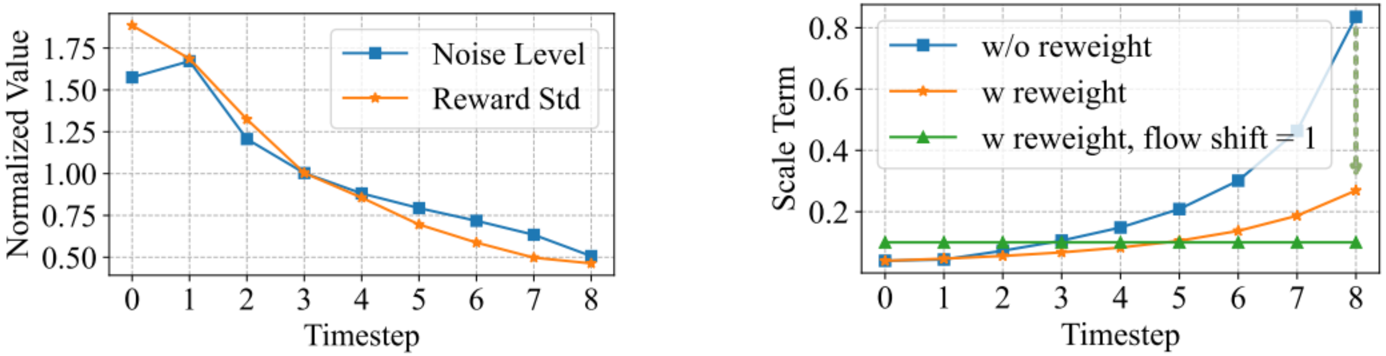


Figure 5: **(Left)** Strong correlation between reward standard deviation and noise level across generative timesteps. **(Right)** Scale term analysis reveals a fundamental mismatch in standard GRPO: scale terms are inversely proportional to noise levels, causing low-noise refinement steps to dominate optimization despite minimal impact on image content.

$$\nabla_\theta \mathcal{J}(\theta) = \sum_{k=0}^{T-1} \mathbb{E}_{\boldsymbol{x}_T \sim \mathcal{N}(0,\boldsymbol{I}), \boldsymbol{\epsilon} \sim \mathcal{N}(0,\boldsymbol{I})} \left[\underbrace{\left(\frac{1}{a} + \frac{a}{2}\right) \Delta k}_{\text{Scale Term}} \cdot \boldsymbol{\epsilon} \cdot \nabla_\theta \boldsymbol{v}_\theta(\boldsymbol{x}_k, k) \hat{A}_k \right]$$

Results

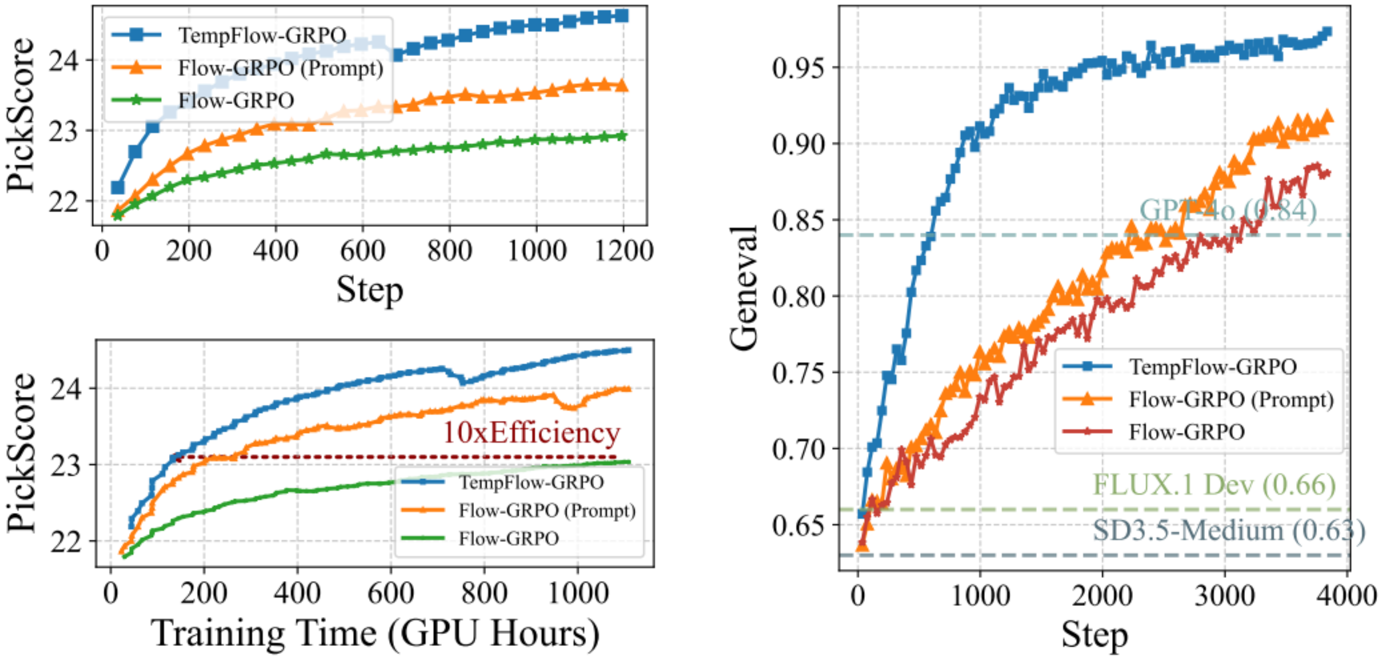


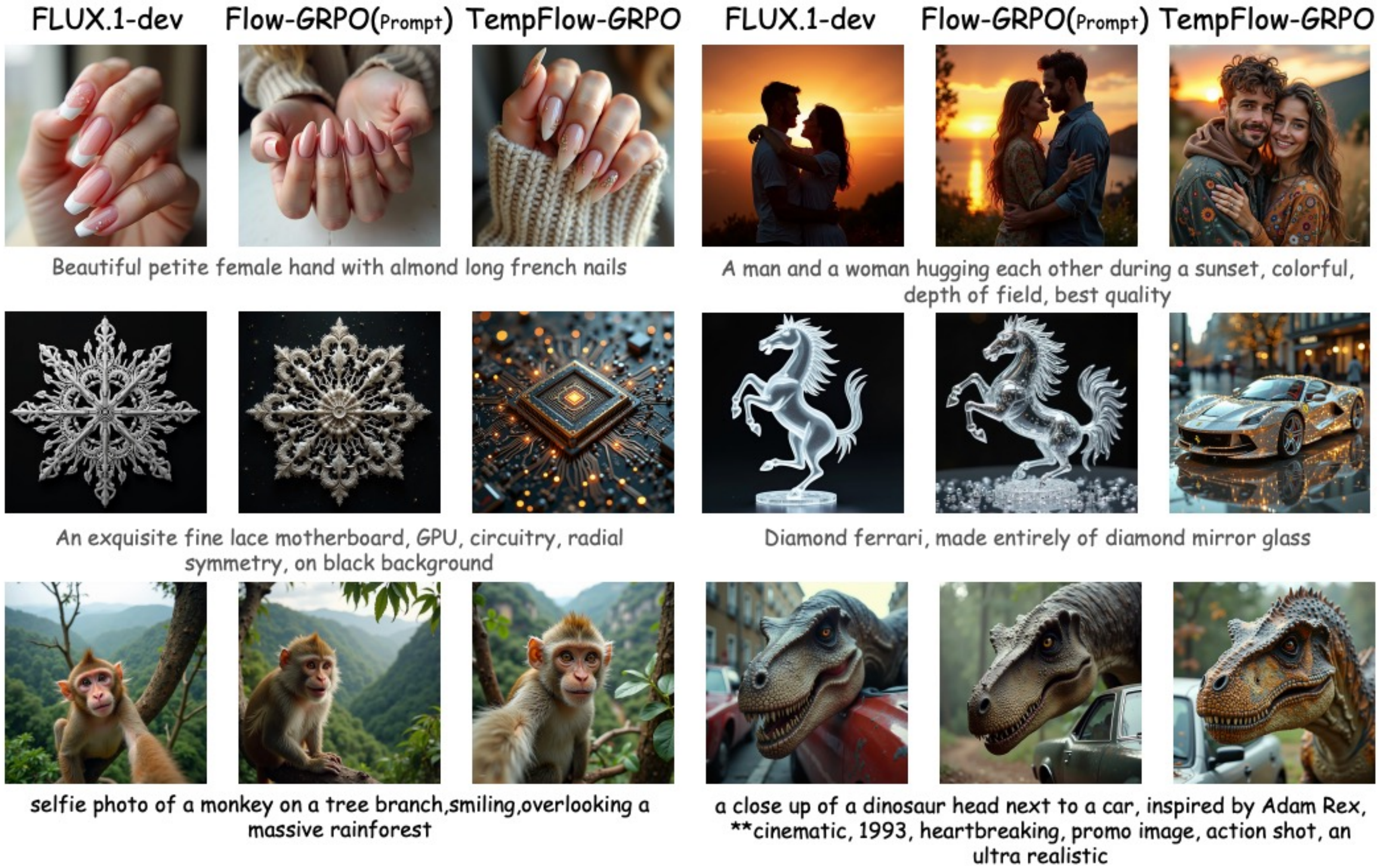
Figure 3: (Left) Performance comparison on the PickScore benchmark across training steps and GPU hours. Flow-GRPO (Prompt) represents **an improved baseline with group-wise standard deviation stabilization**. TempFlow-GRPO consistently outperforms both Flow-GRPO variants in both sample efficiency (steps) and computational efficiency (GPU hours), **demonstrating superior training efficiency while achieving the best performance**. (Right) On the Geneval benchmark, TempFlow-GRPO achieves the highest performance, significantly outperforming Flow-GRPO and surpassing state-of-the-art models including GPT-4o, FLUX.1 Dev, and SD3.5-Medium.

Results

Table 1: **GenEval Result.** Best scores are in blue, second-best in green. Results for models are from Flow-GRPO. Obj.: Object; Attr.: Attribution.

Model	Step	Overall ↑	Single Obj. ↑	Two Obj. ↑	Counting ↑	Colors ↑	Position ↑	Attr. Binding ↑
Diffusion Models								
LDM (Rombach et al., 2022)	-	0.37	0.92	0.29	0.23	0.70	0.02	0.05
SD1.5 (Rombach et al., 2022)	-	0.43	0.97	0.38	0.35	0.76	0.04	0.06
SD2.1 (Rombach et al., 2022)	-	0.50	0.98	0.51	0.44	0.85	0.07	0.17
SD-XL (Podell et al., 2023)	-	0.55	0.98	0.74	0.39	0.85	0.15	0.23
DALLE-2 (Ramesh et al., 2022)	-	0.52	0.94	0.66	0.49	0.77	0.10	0.19
DALLE-3 (Betker et al., 2023)	-	0.67	0.96	0.87	0.47	0.83	0.43	0.45
Autoregressive Models								
Show-o (Xie et al., 2024b)	-	0.53	0.95	0.52	0.49	0.82	0.11	0.28
Emu3-Gen (Wang et al., 2024)	-	0.54	0.98	0.71	0.34	0.81	0.17	0.21
JanusFlow (Ma et al., 2025a)	-	0.63	0.97	0.59	0.45	0.83	0.53	0.42
Janus-Pro-7B (Chen et al., 2025)	-	0.80	0.99	0.89	0.59	0.90	0.79	0.66
GPT-4o (Hurst et al., 2024)	-	0.84	0.99	0.92	0.85	0.92	0.75	0.61
Flow Matching Models								
FLUX.1 Dev (Black et al., 2025)	-	0.66	0.98	0.81	0.74	0.79	0.22	0.45
SD3.5-L (Esser et al., 2024)	-	0.71	0.98	0.89	0.73	0.83	0.34	0.47
SANA-1.5 4.8B (Xie et al., 2025)	-	0.81	0.99	0.93	0.86	0.84	0.59	0.65
SD3.5-M (Esser et al., 2024)	-	0.63	0.98	0.78	0.50	0.81	0.24	0.52
GRPO based Methods								
SD3.5-M+Flow-GRPO (Liu et al., 2025)	5600	0.95	1.00	0.99	0.95	0.92	0.99	0.86
SD3.5-M+Flow-GRPO (Liu et al., 2025)	3800	0.88	0.99	0.96	0.90	0.87	0.83	0.78
SD3.5-M+TempFlow-GRPO	3800	0.97	1.00	1.00	0.96	0.95	0.99	0.91

Results


FLUX.1-dev
Flow-GRPO(Prompt)
TempFlow-GRPO
FLUX.1-dev
Flow-GRPO(Prompt)
TempFlow-GRPO
Beautiful petite female hand with almond long french nails
A man and a woman hugging each other during a sunset, colorful, depth of field, best quality
An exquisite fine lace motherboard, GPU, circuitry, radial symmetry, on black background
Diamond ferrari, made entirely of diamond mirror glass
selfie photo of a monkey on a tree branch,smiling,overlooking a massive rainforest
a close up of a dinosaur head next to a car, inspired by Adam Rex, **cinematic, 1993, heartbreaking, promo image, action shot, an ultra realistic

THANKS