Rectified-CFG++

Geometry-Aware Guidance for Rectified Flow Models

Shreshth Saini

Laboratory for Image and Video Engineering (LIVE), The University of Texas at Austin
Advised by Prof. Alan C. Bovik

NeurIPS 2025 · November 2025

Saini, Gupta, Bovik. Rectified CFG++ for Flow Based Models. NeurIPS 2025 (arXiv:2510.07631)

The question

Why does guidance break on the best image generators?

Flux, SD3 and SD3.5 sample a rectified flow with a deterministic ODE. Classifier-free guidance (CFG) pushes their trajectories off the manifold, and nothing brings them back.

Why we need guidance at all

- Without guidance, a flow model gives generic images that only loosely match the prompt.
- Guidance should amplify the prompt: sharper detail, fuller color, better alignment.
- CFG does that for diffusion models. On flow models it oversaturates.

Flux-dev, same prompt, same seed



No guidance | CFG, ω = 3.5 | Ours, λ = 3.5

SD3.5, same prompt, same seed



No guidance | CFG, ω = 3.5 | Ours, λ = 3

No guidance is dark and muddy. CFG oversaturates. Ours stays natural and faithful to the prompt.

What CFG does to flow models

Standard CFG on random prompts. These are typical results.

Flux-dev with CFG, ω = 3 (paper Figure 6)



Plastic, cartoon look



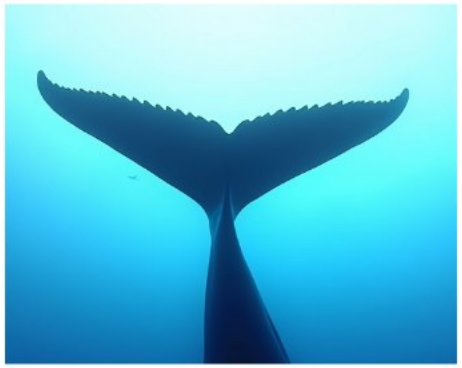
Oversaturated



Flat, clipart look



Blown highlights



Washed out

SD3 with CFG, ω = 3 to 3.5



Oversaturated



Text garbled



Misspelled



Color blow-out

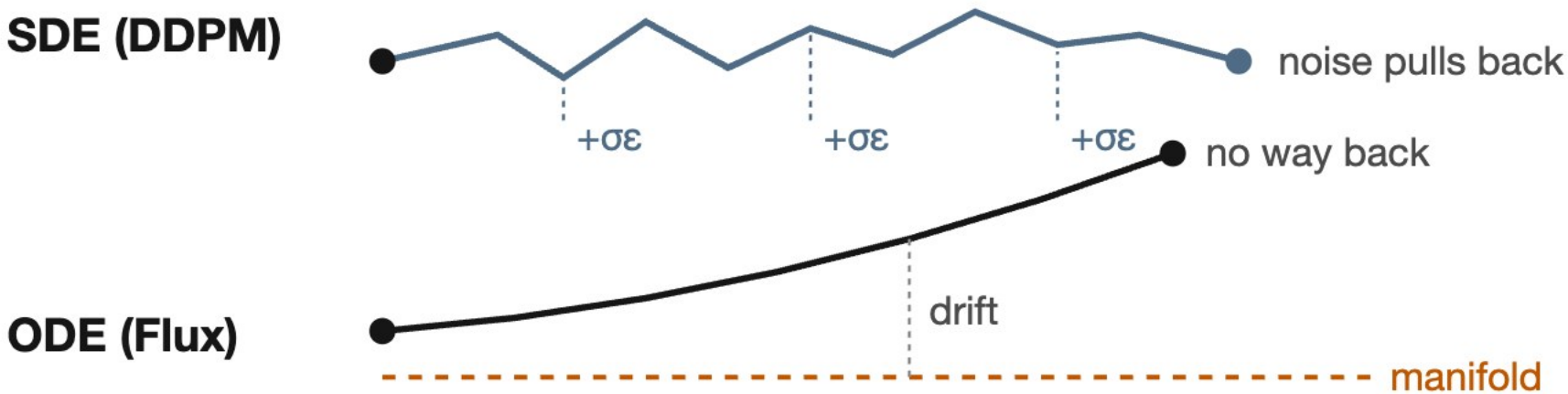


Color distortion

The same four failures every time: oversaturation, a plastic look, garbled text, flat texture.

Why CFG works on diffusion but breaks on flows

Same extrapolation, different samplers



Sampler	Update	After CFG pushes the sample off the manifold
SDE (diffusion)	$\mathrm{d}x = [f(x,t) - g(t)^2\ \nabla \log p_t(x)]\ \mathrm{d}t + g(t)\ \mathrm{d}\bar{w}$	fresh noise at every step partly pulls it back
ODE (rectified flow)	$\mathrm{d}x = v_\theta(x, t)\ \mathrm{d}t$	nothing pulls it back, and the error compounds

Discrete DDPM step: $x_{t-1} = \mu_\theta(x_t, t) + \sigma_t z$ with $z \sim \mathrm{N}(0, \mathrm{I})$. The $\sigma_t z$ term is the fresh noise.

So we need guidance built for deterministic ODE samplers.

What good guidance for a flow model should do

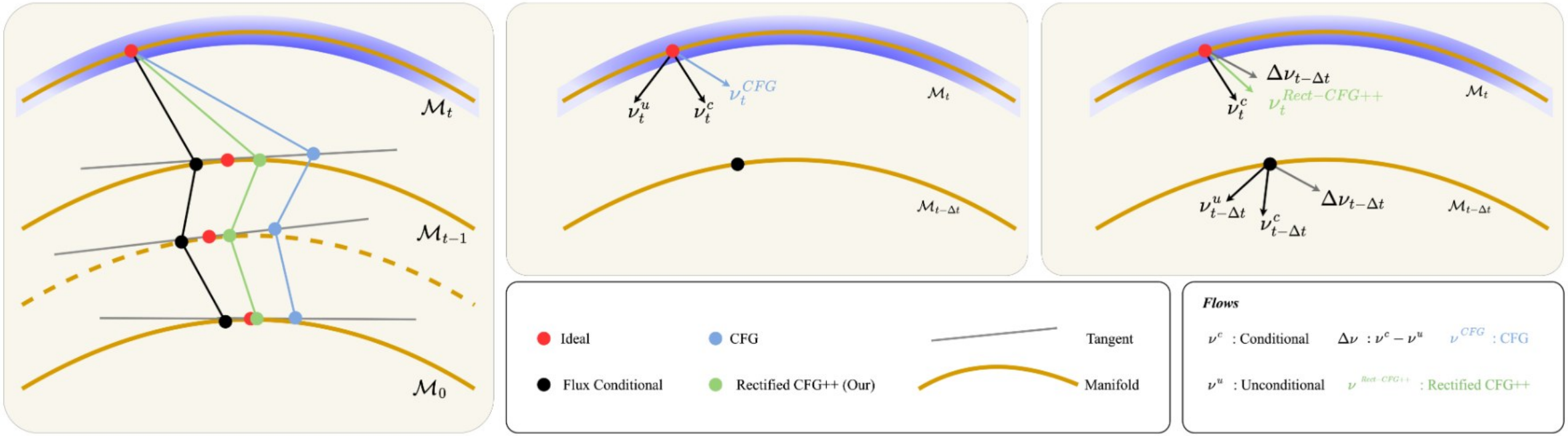
- **D1.** Stay on the manifold, so errors cannot pile up.
- **D2.** Guide strongly early (layout, color), gently late (text, texture).
- **D3.** Come with provable bounds.
- **D4.** Drop into any pretrained model, no retraining.

Method	D1	D2	D3	D4
CFG	no	no	no	yes
CFG-Zero*	no	no	no	yes
APG	partly	no	no	yes
Rectified-CFG++ (ours)	yes	yes	yes	yes

No existing method meets all four. APG is adaptive projected guidance.

We predict a midpoint on the manifold, evaluate guidance there, and add it as a bounded correction.

Why CFG fails and how we fix it, in one picture



Ideal flow

The sample moves along the manifold.

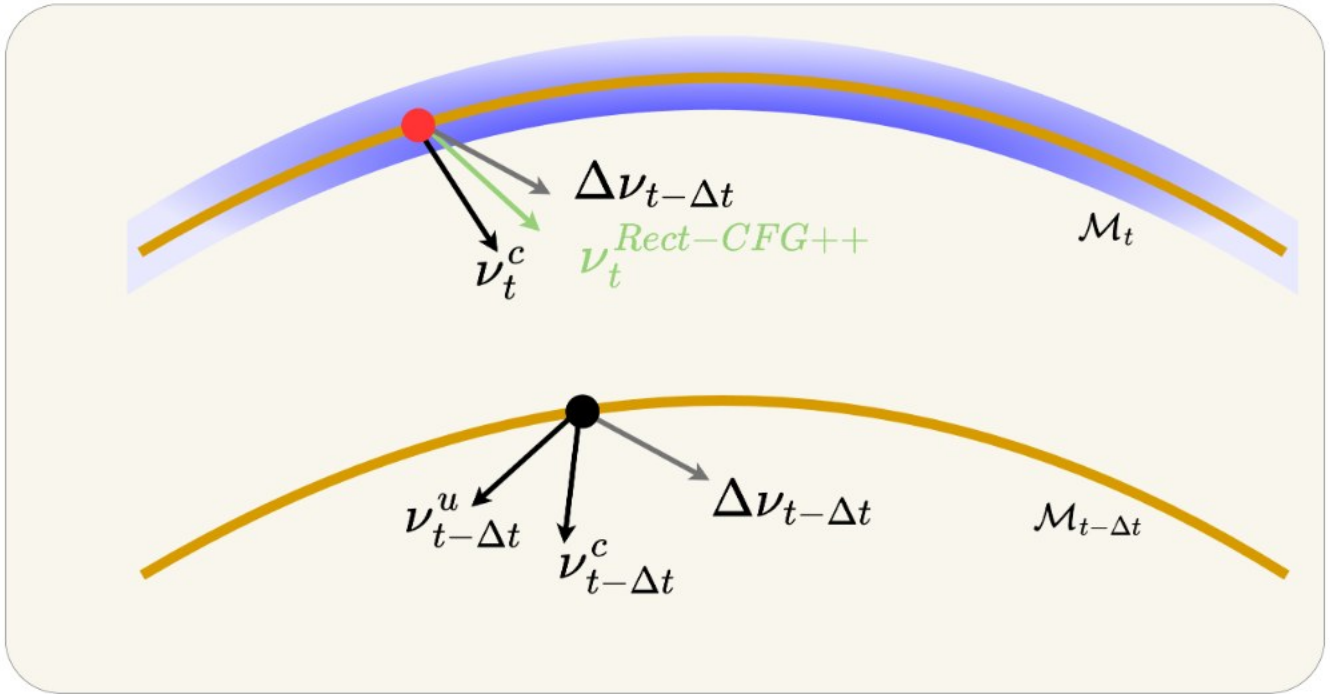
CFG

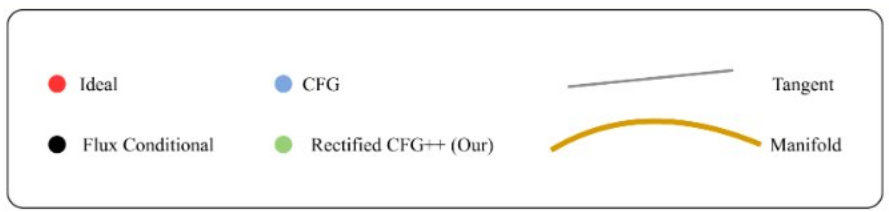
Extrapolation steps off it, and the error stays.

Rectified-CFG++

Predict, then correct. The step lands back on the manifold.

Our predictor-corrector stays on the manifold





1. Predict a half-step with v^c alone	$\tilde{x}_{mid} = x_t + (\Delta t/2)\, v^c{}_t$
2. Evaluate guidance at the midpoint	$\Delta v_{mid} = v^c{}_{mid} - v^u{}_{mid}$
3. Correct a bounded addition to v^c	$\hat{v} = v^c{}_t + \alpha(t)(v^c{}_{mid} - v^u{}_{mid})$

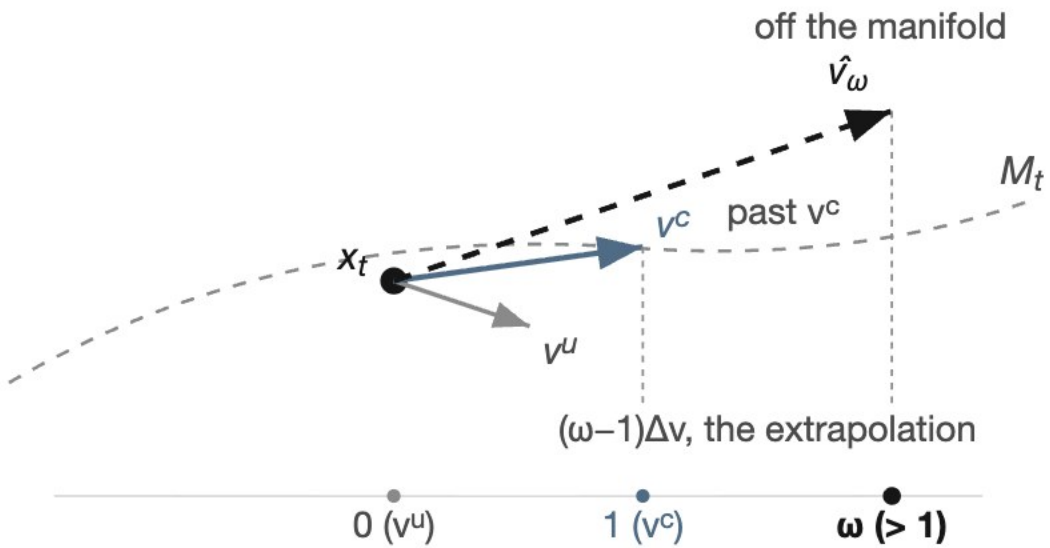
$\alpha(t) = \lambda_{max}(1-t)^\gamma$ fades to 0 near the data, so fine detail survives.

The trajectory stays in a bounded tubular neighbourhood of M_t.

Extrapolation versus interpolation

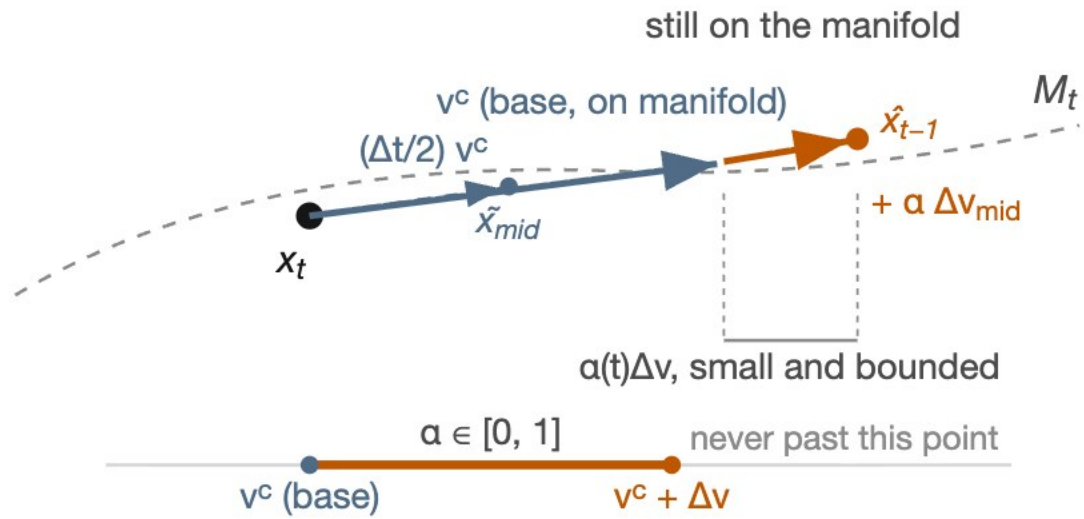
Where each rule puts the guided velocity

CFG extrapolates (ω > 1)



$(1-\omega)v^u + \omega v^c$ with $\omega \in [3, 9]$. $\omega = 3$ means three times the distance from v^u to v^c, where the model never trained.

Ours interpolates (0 ≤ α ≤ 1)



$v^c + \alpha(t)\,\Delta v_{mid}$ with $\alpha \in [0, 1]$. The base is always v^c; the correction is bounded and decays to 0 near the data.

The two algorithms side by side

Highlighted lines are what changes

Standard CFG

```
for t = T down to 1:
  v^c = v_θ(x_t, t, y)
  v^u = v_θ(x_t, t, ∅)
  v̂ = v^u + ω(v^c − v^u)          extrapolation
  x_{t−1} = x_t + Δt v̂
```

2 network evaluations (NFE) per step

Rectified-CFG++

```
for t = T down to 1:
  v^c = v_θ(x_t, t, y)
  x̃_mid = x_t + (Δt/2) v^c                predictor
  v^c_mid = v_θ(x̃_mid, t − Δt/2, y)
  v^u_mid = v_θ(x̃_mid, t − Δt/2, ∅)
  v̂ = v^c + α(t)(v^c_mid − v^u_mid)       interpolation
  x_{t−1} = x_t + Δt v̂
```

3 NFE per step

1. The base is always v^c.
2. Guidance is evaluated at the predicted midpoint.
3. $\alpha(t) = \lambda(1-t)^\gamma$ fades to 0 near the data.

What we can prove

Four mild assumptions

A1	$v_\theta(x,t,y)$ and $v_\theta(x,t,\varnothing)$ are L-Lipschitz in x	**A2**	Bounded guidance direction, $\|\Delta v_t^\theta(x)\| \le B$
A3	$\alpha(t)$ is bounded and integrable	**A4**	Bounded conditional velocity, $\|v^c{}_t\| \le V_{max}$

Standard regularity conditions for a trained network.

Three bounds

Lemma 1 guidance stability	$\|\Delta v^\theta{}_{t-\Delta t/2} - \Delta v^\theta{}_t(x_t)\| \le L\, V_{max}\, \Delta t$	The midpoint moves the guidance direction by only $O(\Delta t)$
Proposition 1 single-step perturbation	$\|\hat{x}_{t-1} - \tilde{x}_{t-1}\| \le \alpha(t)\, B\, \Delta t$	Deviation from the conditional path vanishes near the data
Proposition 2 distributional deviation	$D_{\mathrm{KL}}(\hat{p}_0 \,\|\, p_0) \le C_s\, (B + L\, V_{max}\, \Delta t) \int \alpha(\tau)\, d\tau$	Total KL is set by the integrated guidance

No earlier guidance method for flow models has bounds like these.

CFG versus ours, property by property

Lemma 2, manifold-faithful corrector

$$\mathrm{dist}(\hat{x}, M_{t-\Delta t}) \le C\,\varepsilon\,\Delta t$$

Only the training error ε moves us off the manifold.

Integrated guidance

CFG: ∫(ω−1)dt = ω−1 ≈ 2 to 8



Ours: ∫α(τ)dτ = λ/(γ+1) ≈ 0.5



0.5

4 to 16× lower KL divergence

Property	CFG	Ours
Per-step perturbation	(ω−1) $B\,\Delta t$ constant, never vanishes	**α(t) $B\,\Delta t$** goes to 0 near the data
Manifold distance	Unbounded errors accumulate	**O(ε Δt)** training error only
Integrated guidance	ω−1 ≈ 2 to 8	**λ/(γ+1) ≈ 0.5**
KL($\hat{p}_0 \,\|\, p_0$)	O(ω−1)	**O(0.5)**
Late-step behavior	Same ω everywhere, destroys fine detail	α(t) goes to 0, preserves fine detail
Guidance evaluation	At the current point (possibly off the manifold)	At the predicted midpoint (on the manifold)

Watching the artifacts grow, or not

Standard CFG, ω = 2.5. Each strip shows 7 denoising steps, noise on the left, final image on the right.



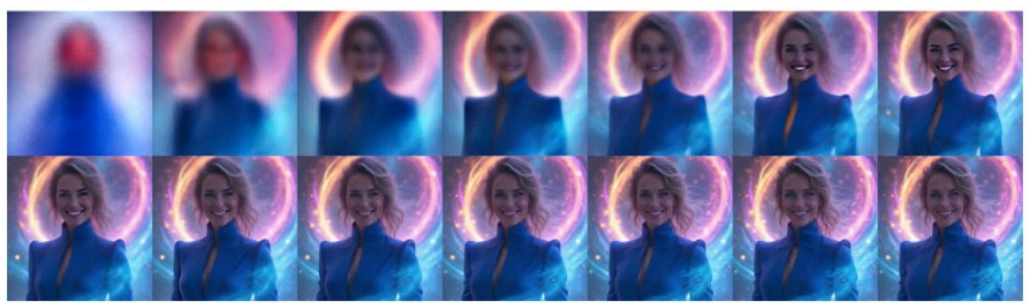
Rectified-CFG++, λ = 0.5, same prompt



CFG, another prompt. Drift starts at step 2 or 3.

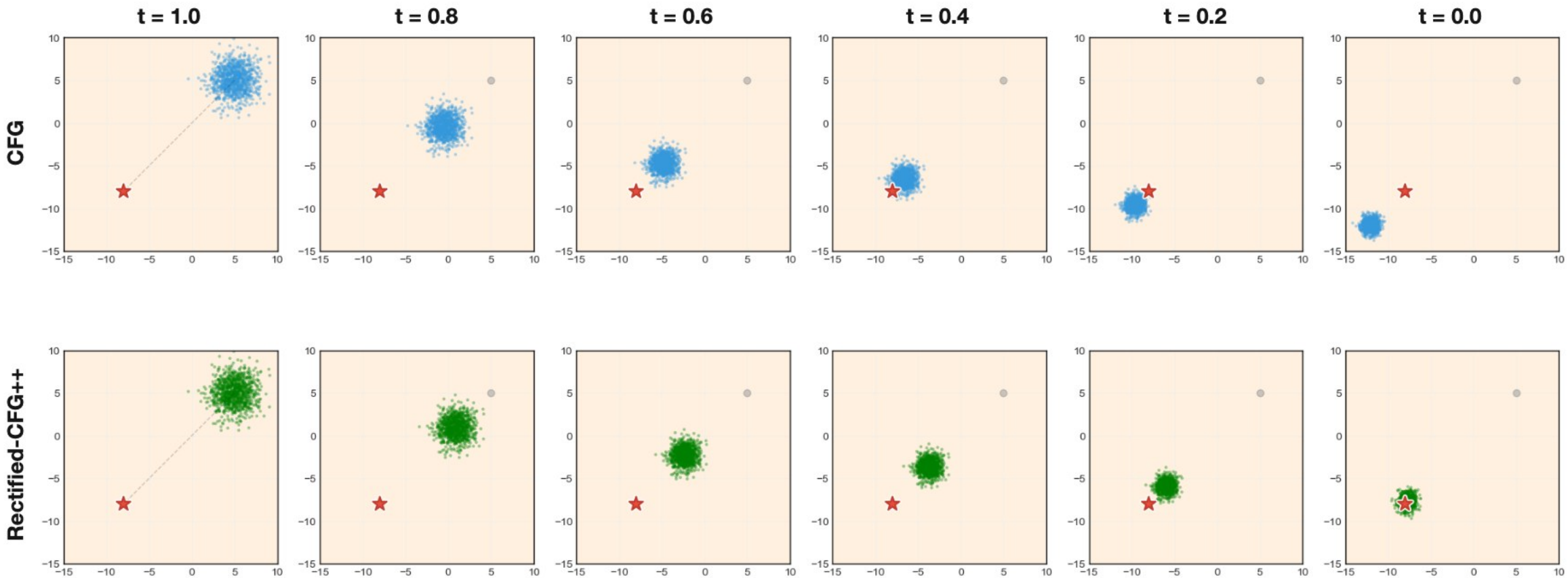


Ours, same prompt. Every intermediate looks natural.



A 2D toy problem shows the same drift

CFG overshoots and needs sharp late corrections. Ours converges smoothly.



200 trajectories on a 2D mixed Gaussian. Top: CFG (blue) drifts off the support. Bottom: ours (green) stays on the manifold.

Baseline, CFG and Rectified-CFG++ on three Flux prompts



Rows from the top: Flux with no guidance, CFG, Rectified-CFG++.

- **Cactus.** CFG draws a cartoon with an emoji face.
- **Dog and moon gate.** CFG loses the stone arch.
- **Parrot.** CFG turns the bird psychedelic.

Ours keeps the realism and the structure in all three.

The same on SD3 and Lumina-Next



SD3, CFG



SD3, ours



SD3, CFG



SD3, ours



Lumina, CFG



Lumina, ours



Lumina, CFG



Lumina, ours

A drop-in replacement across rectified flow models. No retraining, no architecture changes.

Text survives because guidance fades at the end

CFG



Rectified-CFG++



Prompt: "A stop sign with 'ALL WAY' written below it."

$\alpha(t)$ is 0 in the last steps, so the conditional flow alone draws the strokes.

Results on MS-COCO 10K

Model	Method	FID ↓	CLIP ↑	ImgRwd ↑	Pick ↑	HPS ↑
Lumina	CFG	26.93	0.261	0.547	21.03	0.253
Lumina	**Ours**	**22.49**	**0.268**	**0.621**	**21.19**	**0.259**
SD3	CFG	26.33	0.273	0.684	21.48	0.264
SD3	**Ours**	**24.68**	**0.278**	**0.753**	**21.62**	**0.269**
Flux	CFG	37.86	0.285	0.892	22.04	0.279
Flux	**Ours**	**32.23**	**0.289**	**0.961**	**22.18**	**0.283**

CFG against Rectified-CFG++ on three flow models. Bold is best.

Guidance (SD3.5)	FID ↓	ImgRwd ↑	CLIP ↑	HPSv2 ↑
CFG	67.71	1.053	**0.352**	0.294
CFG-Zero*	68.39	0.995	0.346	0.288
APG	67.23	1.075	0.351	0.294
Rectified-CFG++	**67.15**	**1.085**	0.351	**0.296**

Against the other guidance methods on SD3.5.

Best or tied on every metric, for every model.

–5.6
FID on Flux

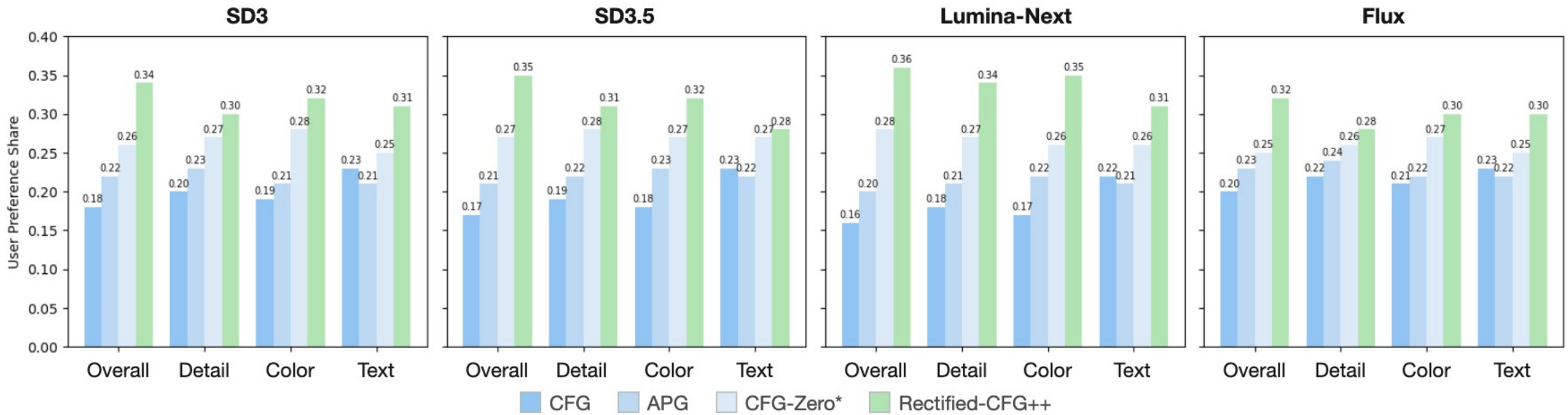
–4.4
FID on Lumina

4
models tested

0
extra training

Experts prefer Rectified-CFG++ on every model

30 experts × 32 prompts × 4 models = 15,360 responses



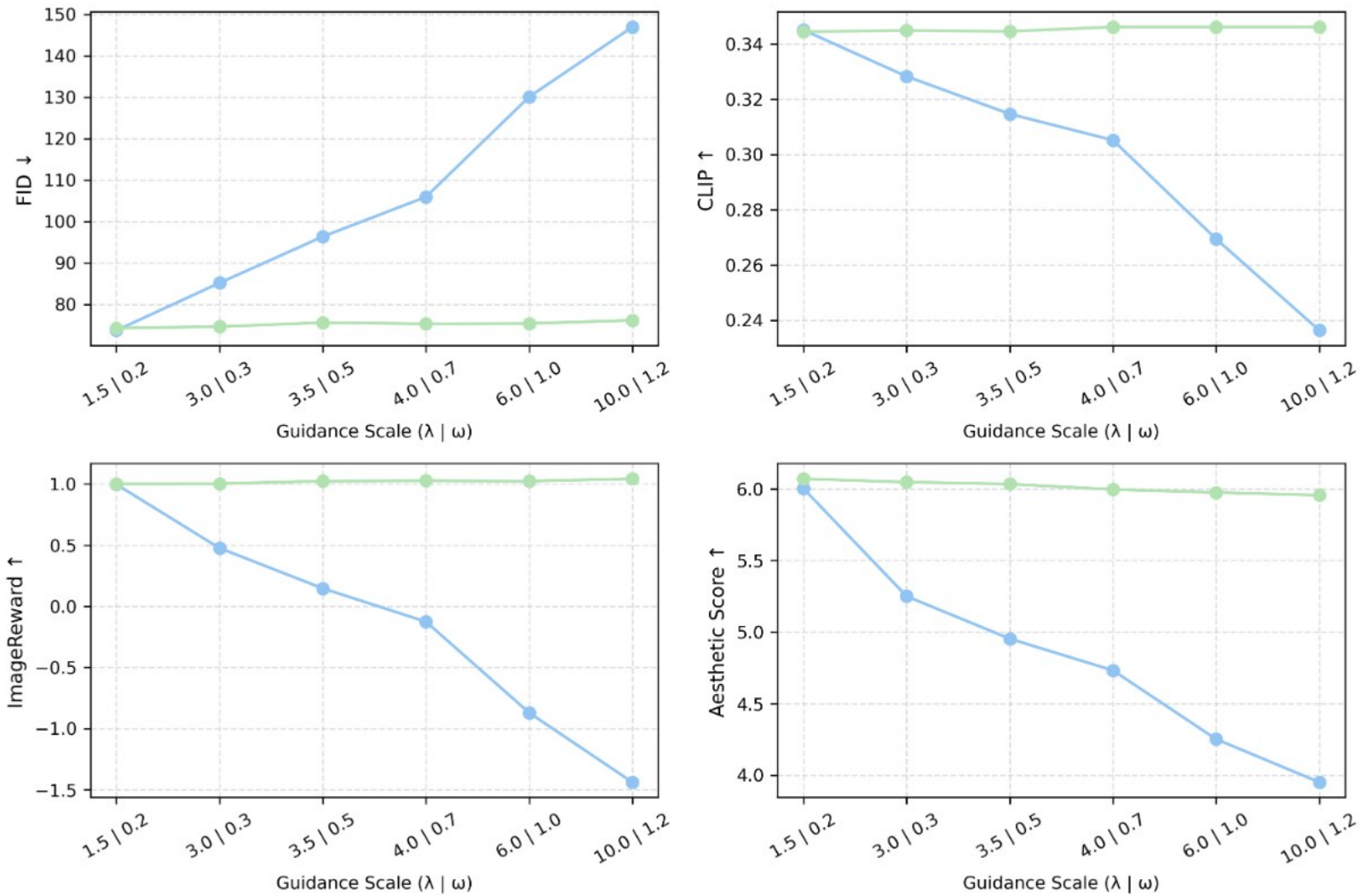
Four-way forced choice on four criteria. CFG dark blue, APG medium blue, CFG-Zero* light blue, Rectified-CFG++ green.

34 to 36%
overall share, against 25% chance ($p < 0.001$ against APG)

28 to 32%
text legibility, against 22 to 25% for the others

$\kappa = 0.61$
Fleiss' kappa, substantial agreement

CFG collapses at high guidance, ours stays flat



x-axis: guidance scale (λ for ours, ω for CFG). y-axis: FID ↓, CLIP ↑, ImageReward ↑, Aesthetic ↑. Blue is CFG, green is ours.

CFG

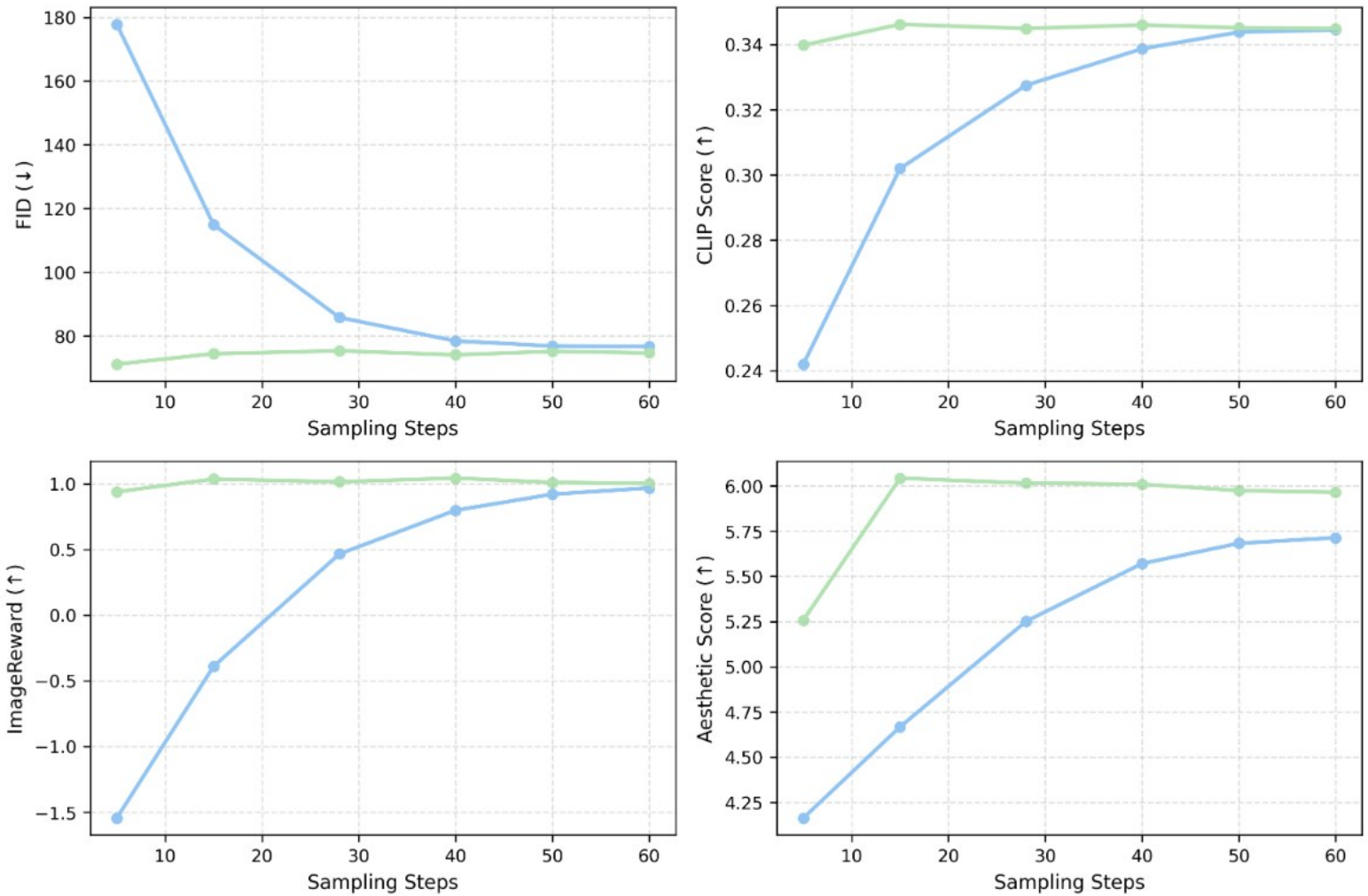
FID goes from 77 to 148 at ω = 10. ImageReward falls from 1.0 to −1.5, CLIP by 30%, Aesthetic by 35%.

Rectified-CFG++

FID stays at 75. ImageReward goes from 1.0 to 0.95.

One setting holds across a 50× range of λ.

The same quality in about half the steps



x-axis: sampling steps. Blue is CFG, green is ours. Ours converges faster on every metric.

- At 5 steps, FID 71 against 178 (2.5× better), ImageReward 1.0 against −1.5.
- 15 of our steps match 28 of CFG's, about 2× faster.

Configuration	FID ↓	CLIP ↑	HPSv2 ↑
Unconditional only	91.12	0.144	0.187
Without predictor	73.70	0.341	0.297
Without corrector	74.65	0.341	0.298
Full	**72.97**	**0.345**	**0.300**

Component ablation on SD3.5. Both the predictor and the corrector are needed.

Takeaways

1. On-manifold guidance for rectified flows, a predictor-corrector with decaying strength.
2. Three propositions, and bounds 4 to 16× tighter than CFG ($\lambda/(\gamma+1)$ against $\omega-1$).
3. Drop-in for Flux, SD3, SD3.5 and Lumina. 15 of our steps match 28 of CFG's.
4. 43.5% preference across 15K expert responses, $p < 0.001$.

Related work of mine

This talk: Saini, Gupta, Bovik. Rectified CFG++ for Flow Based Models. NeurIPS 2025. arXiv:2510.07631

Saini, Liao, Ye, Bovik. LGDM: Latent Guidance in Diffusion Models for Perceptual Evaluations. ICML 2025. arXiv:2506.00327

Saini, Gedik, Birkbeck, Wang, Adsumilli, Bovik. LumaFlux: Structure-Aware SDR-to-HDR Conversion with Monotone Luminance Flows. Preprint, 2026. arXiv:2604.02787

Chen, Saini, Adsumilli, Bovik. LumaGuide: Distribution Shaping for Training-Free HDR Generation in Diffusion Models. Preprint, 2026. arXiv:2607.26237

Saini, Birkbeck, Wang, Adsumilli, Bovik. CachedSearch: Training-Free Cached Exploration for Test-Time Search in Video Diffusion. Preprint, 2026. arXiv:2607.23159

Thank you.

TEXAS The University of Texas at Austin

The University of Texas at Austin LIVE Laboratory for Image & Video Engineering

YouTube