RectifiedHR Upgrade Plan

Contributions. We make three contributions. (1) *Energy profiling for diffusion sampling:* we formalize a normalized latent "energy" and introduce stability summaries (total variation, spike magnitude, monotonicity violations, area) that diagnose artifact-prone trajectories. (2) *Energy-guided guidance:* we propose a family of adaptive classifier-free guidance schedules (linear, cosine, step, etc.) with simple, practical rules that stabilize late steps across samplers; we also provide lightweight stabilizers (energy clipping, noise refresh). (3) *Reproducible evidence:* using a unified set of configurations with fixed prompts/seeds, we compare against *best* fixed-CFG baselines per sampler and show consistent gains in CLIPScore/LPIPS and seed-consistency alongside lower energy variation. We include a small SDXL@768 set to demonstrate scalability and perform an overhead analysis.

1. **Datasets, Prompts, Seeds**
   - **Prompts:** 200 public prompts spanning (*i*) object-centric, (*ii*) compositional/relational, (*iii*) artistic/style buckets. (currently 10)
   - **Seeds:** 5 fixed seeds per prompt. (currently 3)
   - **Models & resolutions:** Stable Diffusion 1.5 @ 512×512 (main); add SDXL @ 768 as a lightweight scalability ablation. Optionally add a *tiny* 1024 if space allows.

2. **Suggested Metrics**
   - **Perceptual: CLIPScore** (reference-free), LPIPS, MS-SSIM (you already have these)
   - **Energy-based stability (core):** for step t, $E_t = \|x_t\|_2^2/N$. Report: (i) spike magnitude, (ii) total variation (TV), (iii) # monotonicity violations, (iv) area over trajectory. (mostly done.)
   - **Consistency:** define a metric to measure per-seed agreement across samplers/schedules.
   - **Link metrics:** define a measure to confirm stable trajectories correspond to better quality. For exmaple, *Spearman* ρ between -TV and CLIPScore/LPIPS across configs.

3. **Guidance Schedules & Samplers**
   - **Schedule family for CFG scale s_t:** linear-decrease, cosine, step, exponential, sigmoid. (you have these.)
   - **Samplers:** DDIM, Euler A, DPM++ 2M. (you have these.)
   - **Configurations:** Tables 1 & 2 defines all experiments to run.

4. **Baselines Under Identical Protocol**
   - Reproduce HiResFix and RectifiedFlow results using same configs. (Currently compared conceptually; table cites original-paper numbers not directly comparable)
   - Clearly tag any non-reproduced literature numbers as *reported*. (you have this already)
   - Include best fixed-CFG per sampler = best constant guidance per sampler selected by mean Q over prompts (e.g., EulerA: CFG=8, DPM++2M: CFG=12, DDIM: CFG=10).

5. **Add Some Stress Tests**

6. **Add Some Ablations**

7. **Add a Formal Justification of the Approach**: Currently discussed intuitively; formalize the rationale and notations. (I can help you with this)

Table 1: Experiment factors and levels used to define configurations. Prompts and seeds are fixed and reused across all configurations.

Factor	**Levels**
Model	SD 1.5 (main), SDXL (scalability ablation)
Resolution	512 (main), 768 (SDXL ablation)
Sampler	DDIM, EulerA, DPM++ 2M
Schedule type	linear, cosine, step
Endpoints ($s_0 \to s_1$)	($12 \to 3$), ($10 \to 5$), ($8 \to 3$)
Steps	50 (main)
Prompts	200 public prompts (released)
Seeds	5 fixed seeds per prompt (released)

Table 2: Configurations to run: SD 1.5@512 (main) and a tiny SDXL@768 scalability ablation. Each row is a configuration (cfg_id) run on the same prompts and seeds.

Model	**Res**	**Sampler**	**Schedule**	$\mathbf{s_0}$	$\mathbf{s_1}$	**cfg_id**
SD 1.5 @ 512, steps=50						
SD1.5	512	DDIM	linear	12	3	sd15-512-ddim-linear-s0_12-s1_3-n50
SD1.5	512	DDIM	linear	10	5	sd15-512-ddim-linear-s0_10-s1_5-n50
SD1.5	512	DDIM	linear	8	3	sd15-512-ddim-linear-s0_8-s1_3-n50
SD1.5	512	DDIM	cosine	12	3	sd15-512-ddim-cosine-s0_12-s1_3-n50
SD1.5	512	DDIM	cosine	10	5	sd15-512-ddim-cosine-s0_10-s1_5-n50
SD1.5	512	DDIM	cosine	8	3	sd15-512-ddim-cosine-s0_8-s1_3-n50
SD1.5	512	DDIM	step	12	3	sd15-512-ddim-step-s0_12-s1_3-n50
SD1.5	512	DDIM	step	10	5	sd15-512-ddim-step-s0_10-s1_5-n50
SD1.5	512	DDIM	step	8	3	sd15-512-ddim-step-s0_8-s1_3-n50
SD1.5	512	EulerA	linear	12	3	sd15-512-eulera-linear-s0_12-s1_3-n50
SD1.5	512	EulerA	linear	10	5	sd15-512-eulera-linear-s0_10-s1_5-n50
SD1.5	512	EulerA	linear	8	3	sd15-512-eulera-linear-s0_8-s1_3-n50
SD1.5	512	EulerA	cosine	12	3	sd15-512-eulera-cosine-s0_12-s1_3-n50
SD1.5	512	EulerA	cosine	10	5	sd15-512-eulera-cosine-s0_10-s1_5-n50
SD1.5	512	EulerA	cosine	8	3	sd15-512-eulera-cosine-s0_8-s1_3-n50
SD1.5	512	EulerA	step	12	3	sd15-512-eulera-step-s0_12-s1_3-n50
SD1.5	512	EulerA	step	10	5	sd15-512-eulera-step-s0_10-s1_5-n50
SD1.5	512	EulerA	step	8	3	sd15-512-eulera-step-s0_8-s1_3-n50
SD1.5	512	DPM++2M	linear	12	3	sd15-512-dpmpp2m-linear-s0_12-s1_3-n50
SD1.5	512	DPM++2M	linear	10	5	sd15-512-dpmpp2m-linear-s0_10-s1_5-n50
SD1.5	512	DPM++2M	linear	8	3	sd15-512-dpmpp2m-linear-s0_8-s1_3-n50
SD1.5	512	DPM++2M	cosine	12	3	sd15-512-dpmpp2m-cosine-s0_12-s1_3-n50
SD1.5	512	DPM++2M	cosine	10	5	sd15-512-dpmpp2m-cosine-s0_10-s1_5-n50
SD1.5	512	DPM++2M	cosine	8	3	sd15-512-dpmpp2m-cosine-s0_8-s1_3-n50
SD1.5	512	DPM++2M	step	12	3	sd15-512-dpmpp2m-step-s0_12-s1_3-n50
SD1.5	512	DPM++2M	step	10	5	sd15-512-dpmpp2m-step-s0_10-s1_5-n50
SD1.5	512	DPM++2M	step	8	3	sd15-512-dpmpp2m-step-s0_8-s1_3-n50
SDXL @ 768 (scalability ablation), steps=50						
SDXL	768	DDIM	linear	12	3	sdxl-768-ddim-linear-s0_12-s1_3-n50
SDXL	768	DPM++2M	linear	12	3	sdxl-768-dpmpp2m-linear-s0_12-s1_3-n50

Table 3: Sample main results on SD 1.5@512 (200 prompts, 5 seeds). Mean±95% CI over prompts. Fixed-CFG = best constant guidance per sampler; Ours = best adaptive schedule per sampler. Add all metrics.

Sampler	**Method (CFG)**	**CLIP**	**LPIPS**	**Consistency**	**Energy-TV**
DDIM	Fixed (best const)	???	???	???	???
	Ours (linear 12→3)	???	???	???	???
EulerA	Fixed (best const)	???	???	???	???
	Ours (cosine 12→3)	???	???	???	???
DPM++2M	Fixed (best const)	???	???	???	???
	Ours (linear 12→3)	???	???	???	???

Table 4: Schedule ablation on SD 1.5@512 (steps=50). Mean±95% CI over prompts. Add all metrics.

Sampler	**Schedule ($s_0 \to s_1$)**	**CLIP**	**LPIPS**	**Energy-TV**
DDIM	linear 12→3	???	???	???
	cosine 12→3	???	???	???
	step 12→3	???	???	???
EulerA	linear 12→3	???	???	???
	cosine 12→3	???	???	???
	step 12→3	???	???	???
DPM++2M	linear 12→3	???	???	???
	cosine 12→3	???	???	???
	step 12→3	???	???	???

Table 5: Scalability ablation on SDXL@768 (200 prompts, 3 seeds okay). Add all metrics.

Sampler	**Method**	**CLIP**	**LPIPS**	**Energy-TV**
DDIM	Fixed (best const)	???	???	???
DDIM	**Ours** (linear 12→3)	???	???	???
DPM++2M	Fixed (best const)	???	???	???
DPM++2M	**Ours** (linear 12→3)	???	???	???

Table 6: Overhead vs best Fixed-CFG (SD 1.5@512). Mean runtime per image and peak memory.

Sampler	**Δ Runtime (s/img)**	**Δ Peak Mem (MB)**
DDIM	???	???
EulerA	???	???
DPM++2M	???	???

Table 7: Energy–Perception link (per-prompt Spearman ρ between Q and $-$TV; mean±95% CI across prompts).

Setting	**Link@ρ (CLIP, $-$TV)**
SD 1.5@512 (all configs)	???
SDXL@768 (ablation)	???