DRM: Diffusion-based Reward Model With Step-wise Guidance

Supplementary Material

Figure S1. **Qualitative comparison of generation results before and after optimization with our method.** Our approach enhances the model's ability to adhere to complex prompt specifications. (Top) Given a prompt requiring four knives, the baseline model fails on object counting, whereas our optimized model correctly generates the specified number. (Bottom) For a prompt requiring specific attribute binding ("a green bear"), the baseline struggles with color assignment, while our model successfully renders the object with the correct attribute.

1. Result of GenEval

To quantitatively assess text-image alignment, we evaluate our method on the GenEval benchmark. This benchmark contains 553 prompts designed to test compositional understanding, including object counting, spatial relationships, and attribute binding. We apply our Step-GRPO to the SD3.5-M model and compare it against several strong flow matching baselines. Notably, while our reward model is trained exclusively on human preference data, it leads to substantial gains on this objective benchmark. As detailed in Table S1, our method boosts the overall score of SD3.5-M from 0.63 to 0.78, outperforming even the larger SD3.5-L model. The improvements are particularly pronounced in challenging compositional categories. For instance, we observe a massive 60% relative improvement in counting accuracy (from 0.50 to 0.80) and a 35% relative improvement in attribute binding (from 0.52 to 0.70). This highlights our method's ability to generalize from subjective preferences to objective prompt fidelity, instilling a more robust understanding of prompt semantics. These quantitative improvements are further illustrated by the qualitative examples in Figure S1, where our method successfully handles complex prompts that cause the baseline to fail.

2. More Visualization Result

The prompts in Figure S2 are as follows:

1. 16-year-old teenager wearing a white bear-ear hat with a smirk on their face.
2. photo of well done salmon dinner, 8K, Global Illumination, Ray Tracing Reflections
3. A lemon with a McDonald's hat.
4. cat, cute, hat
5. The image is a mixed media collage with broken glass and torn paper elements, featuring intricate oil details and a canvas texture, in a contemporary art style.
6. Kiwi fruit, mint leaves, ice cubes, background yellow, splashing water, soft box, back light, creative food photography, Art by Alberto Seveso,
7. little tiny cub beautiful light color White fox soft fur kawaii chibi Walt Disney style, beautiful smiley face and beautiful eyes sweet and smiling features, snuggled in its soft and soft pastel pink cover, magical light background, style Thomas kinkade Nadja Baxter Anne Stokes Nancy Noel realistic
8. 185764, ink art, Calligraphy, bamboo plant :: orange, teal, white, black –ar 2:3 –uplight
9. A 3D Rendering of a cockatoo wearing sunglasses. The sunglasses have a deep black frame with bright pink lenses. Fashion photography, volumetric lighting, CG rendering.
10. A rock formation in the shape of a horse, insanely detailed
11. a desert in a snowglobe, 4k, octane render :: cinematic –ar 2048:858
12. watercolour beaver with tale, white background

Model	Overall ↑	Single Obj. ↑	Two Obj. ↑	Counting ↑	Colors ↑	Position ↑	Attr. Binding ↑
Flow Matching Models							
FLUX.1 Dev	0.66	0.98	0.81	0.74	0.79	0.22	0.45
SD3.5-L	0.71	0.98	0.89	0.73	0.83	0.34	0.47
SD3.5-M	0.63	0.98	0.78	0.50	0.81	0.24	0.52
GRPO based Methods							
SD3.5-M+Step-GRPO	**0.78**	**0.99**	**0.93**	**0.80**	**0.86**	**0.37**	**0.70**

Table S1. **GenEval Result.** Results for models are from Flow-GRPO. Obj.: Object; Attr.: Attribution.

Figure S2. **Additional qualitative results from our Step-GRPO optimized SD3.5-M model.** These examples showcase the model's ability to generate high-quality, diverse images that faithfully adhere to complex prompts, demonstrating the broad applicability and effectiveness of our method.