Style-GRPO: Semantic-Aware Preference Optimization for Image Style Transfer Guided by Reward Modeling

Supplementary Material

Contents

1. Preliminary

1.1. Flow Matching

Flow Matching [9] learns a vector field to transport samples from a simple prior distribution X_1 to a complex data distribution X_0. We utilize Rectified Flow [11], which defines a simple linear interpolation between a data sample $x_0 \sim X_0$ and a prior sample $x_1 \sim X_1$:

$$x_t = (1-t)x_0 + tx_1, \quad t \in [0,1]. \tag{1}$$

This path is characterized by a constant target velocity field $v = x_1 - x_0$. A neural network $v_\theta(x_t, t, c)$, conditioned on c , is trained to approximate this target field by minimizing the flow matching objective:

$$\mathcal{L}_{FM}(\theta) = \mathbb{E}_{t,x_0\sim X_0,x_1\sim X_1}[||v - v_\theta(x_t,t,c)||_2^2]. \tag{2}$$

The inference stage is performed by solving a deterministic ordinary differential equation (ODE) for forward process:

$$dx_t = v_\theta(x_t, t, c)dt. \tag{3}$$

A key challenge in applying reinforcement learning (RL) to flow-matching backbone lies in its deterministic sampling process. Flow Matching models rely on an ordinary differential equation sampler, whereas RL requires stochasticity for policy exploration. To bridge this gap, we adopt the ODE-to-SDE conversion strategy from Flow-GRPO [10], reformulating the deterministic ODE into an equivalent Stochastic Differential Equation (SDE) that preserves the marginal distribution $p_t(x)$ at all timesteps t. Using Euler–Maruyama discretization, the final update rule is given by:

$$\begin{aligned} x_{t+\Delta t} = x_t + \left[v_\theta(x_t,t) + \tfrac{\sigma_t^2}{2t}\big(x_t + (1-t)v_\theta(x_t,t)\big)\right]\Delta t \\ + \sigma_t\sqrt{\Delta t}\,\epsilon, \quad \epsilon \sim \mathcal{N}(0,I). \end{aligned} \tag{4}$$

This reformulation defines the policy $\pi_\theta(x_{t-1}|x_t, c)$ as an isotropic Gaussian, enabling the stochastic exploration during the online training process.

1.2. GRPO via Noise-aware Reweighting Strategy

Group Relative Policy Optimization (GRPO) computes a group-normalized advantage, $\hat{A}^i$, from the terminal reward:

$$\hat{A}^i = \frac{R(\boldsymbol{x}_0^i, \boldsymbol{c}) - \text{mean}(\{R(\boldsymbol{x}_0^j, \boldsymbol{c})\}_{j=1}^G)}{\text{std}(\{R(\boldsymbol{x}_0^j, \boldsymbol{c})\}_{j=1}^G) + \epsilon} \tag{5}$$

However, it applies this advantage uniformly across all timesteps, a strategy that ignores the varying influence of different generative stages—from high-impact early steps defining global structure to late-stage refinements. TemporalFlow-GRPO [6] rectifies this by introducing a noise-aware reweighting. The core principle is to leverage the noise magnitude σ_t as a direct proxy for a timestep's importance. This is implemented by weighting each step's log-probability by a factor $\lambda_t \propto \sigma_t$, yielding the objective:

$$\mathcal{J}_{\text{reweighted}}(\theta) = \mathbb{E}\left[\sum_{t=0}^{T-1} \lambda_t \cdot \log p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t) \cdot \hat{A}\right] \tag{6}$$

2. Data Curation Process

2.1. Reference Style Curation

To build a comprehensive and high-fidelity reference style benchmark, we implement a two-phase pipeline designed to move from maximizing stylistic coverage to guaranteeing stringent data quality. The initial phase focuses on expanding and structuring the stylistic asset space, which involves both the meticulous cleaning of existing datasets and

As a strict AI style reference curator, you must evaluate an image's suitability as a style reference and extract its metadata.
Instructions:
1. Analyze & Tag: Generate a descriptive caption, content tags (subjects/concepts), and style tags (aesthetics/techniques).
2. Score Suitability: Rate the image based on these weighted criteria:
- Stylistic Definition & Coherence (40 pts): Style is strong, clear, consistent.
- Technical Quality (30 pts): High-res, no artifacts, good clarity.
- Transferability & Versatility (20 pts): Style is broadly applicable.
- Originality & Uniqueness (10 pts): Style is distinctive.
3. Identify Red Flags: Instantly REJECT for major watermarks, obvious AI errors, extreme low resolution, or conflicting mixed styles.
4. Provide Verdict: Output the full analysis in the specified format.

Required Output Format:
1. Metadata:
Caption: [1-2 sentence descriptive caption of the image, need the global style description.]
Short Caption: [EXACT format: "A {style} style"] (Length: 3–8 words total, No explanations, no extra text, no punctuation)
Content Tags: ["tag one", "tag two", "tag three", "tag four"](describe the main subjects/concepts in the image)
Style Tags: ["tag one", "tag two", "tag three", "tag four"](describe the aesthetics/techniques of the style)
2. Suitability Score:
Stylistic Definition: [Score]/40. Justification: [Brief reason]
Technical Quality: [Score]/30. Justification: [Brief reason]
Transferability: [Score]/20. Justification: [Brief reason]
Originality: [Score]/10. Justification: [Brief reason]
3. Final Verdict:
Red Flags: [List any found, or "None"]
Total Score: [Total]/100
Recommendation: [ACCEPT / CONSIDER / REJECT]
Rationale: [One-sentence final summary]

Example

```
 Filename: "s0005____0919_01_query_1_img_000011_1683443735026_03750905708554526.jpg.jpg":
"1. Metadata:
Caption: Tiny pastel cafe/diorama with a cute storefront, presented in a soft, rounded 3D toy aesthetic.
Short Caption: A cute pastel 3D diorama style.
Content Tags: [ "Pastel cafe", "Miniature architecture", "Whimsical food-cart", \"3D render" ]
Style Tags: [ "Kawaii 3D", "Pastel palette", "Isometric/diorama", "Product-like rendering" ]
2. Suitability Score:
    Stylistic Definition: 35/40. Justification: Distinctive cute-cafe vibe with coherent color language.
    Technical Quality: 28/30. Justification: Clean, well-executed shading and materials.
    Transferability: 17/20. Justification: Works for product concepts, branding, and scenes.
    Originality: 9/10. Justification: Strongly unique cafe/diorama concept within the pastel toy space.
3. Final Verdict:
    Red Flags: Watermark present (visible watermark in corner)
    Total Score: 89/100
    Recommendation: REJECT
    Rationale: A compelling, high-quality pastel diorama with strong originality.",
```

Figure 1. **Reference Style Evaluation Protocol.** The protocol structures style reference assessment through weighted scoring (Style, Quality, Transferability, Originality) and a red flag mechanism to jointly ensure the rigor and high fidelity of our final curated benchmark.

the targeted construction of novel, underrepresented virtual styles. The subsequent phase then applies a rigorous, multi-faceted quality assurance protocol across the entire resulting data pool, employing a multi-stage joint filtering strategy to ensure uncompromising data fidelity and consistency.

Filter Existing Style Dataset. Existing large-scale style datasets, such as WikiArt [13] and Style30k [5] suffer from inherent noise, including corrupted images, pervasive watermarks, and inconsistent stylistic representations. To establish a high-fidelity benchmark for reference style curation, we developed a multi-tiered filtering strategy that leverages complementary AI capabilities and human expertise for efficiency and precision. Our initial pass employs `Qwen2.5-VL-72B` [2] for high-throughput coarse filtering. This stage efficiently identifies and removes egregious artifacts like watermarks, low-quality images, and samples lacking clear stylistic prominence, culling approximately 61.2% of the initial pool. Subsequently, for the remaining more complex cases, we utilize the advanced visual-language reasoning of `GPT-4o` [1] for semi-automated refinement. This stage focuses on more nuanced stylistic attributes, identifying subtle inconsistencies or lack of personalization/diversity in styles, filtering out another 25.9% of the dataset. Then, the most challenging and ambiguous 12.9% of the data undergoes meticulous human annotation for ultimate subjective quality assessment and fine-

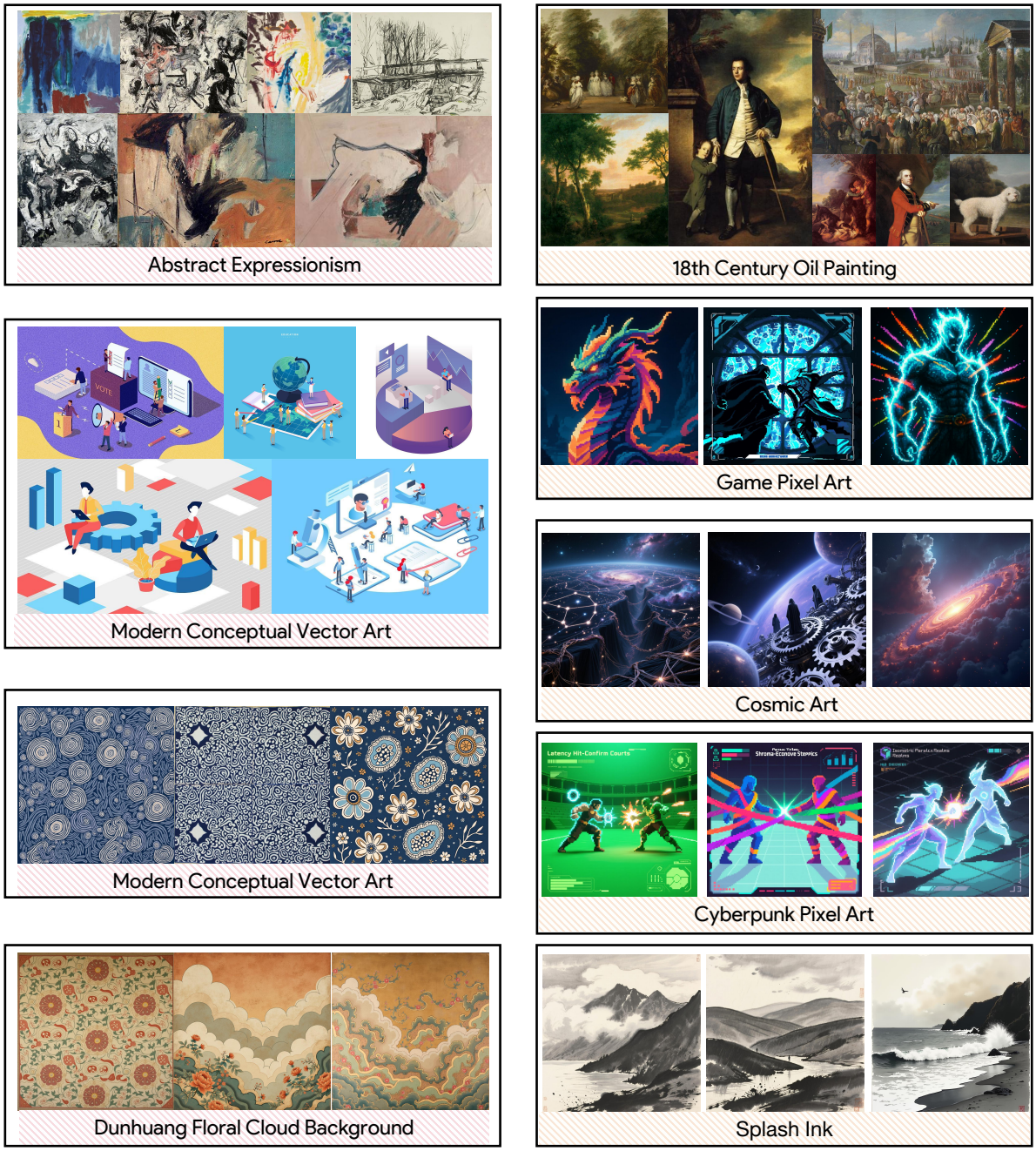

Figure 2. Reference Style Visualization.

grained stylistic consistency, ensuring the final curated dataset meets the highest standards for personalized and diverse style references. All processes use the same prompts and rules to guarantee the consistency of data distribution. Finally, we obtain 12.1k excellent samples with raw images, short/detailed captions, style types and comprehensive scores as shown in Figure 1.

Prompt Design and Data Construction. While existing style benchmarks provide a foundation, they often exhibit a cultural and historical bias, focusing predominantly on global artists while severely lacking coverage of virtual and emerging styles such as creative, anime, and game art aesthetics. To address this gap, we devised a rigorous, taxonomy-guided prompt generation methodology. We first classify these underrepresented styles into 12 distinct categories. We then leverage powerful Large Language Models, `GPT-5` and `Gemini 2.5 Pro`, to perform prompt augmentation for each category, yielding structured outputs including short/detailed captions and explicit style type la-

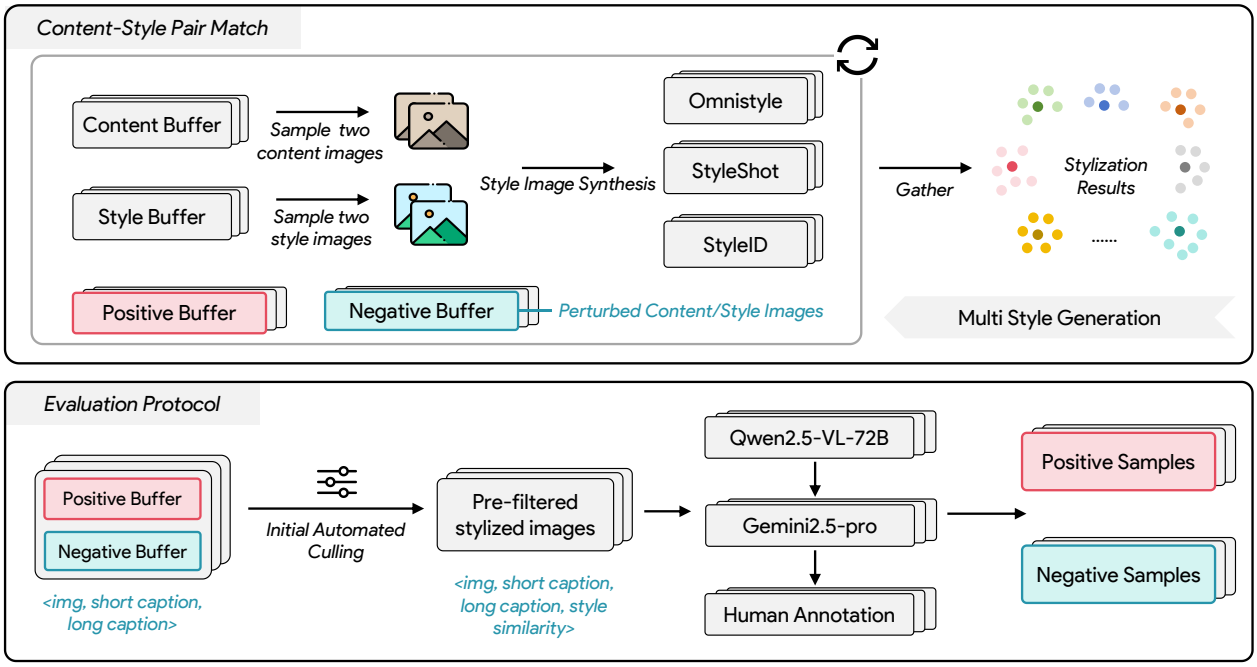

Figure 3. An overview of the style pair design and evaluation protocol.

bels. This augmented data is then used to drive a suite of State-of-the-Art text-to-image (T2I) models Flux [7], Qwen-Image [19], and Hunyuan-Image [3], to generate three candidate images per prompt using three different random seeds to ensure maximal visual diversity. Finally, to guarantee the integrity of this newly constructed dataset, we subject the generated 18k images to a dedicated, multi-tiered quality assurance protocol. This protocol precisely mirrors the three-stage filtering strategy applied to the existing style datasets, employing `Qwen2.5-VL-72B` [2], `GPT-4o` [1], and human annotators sequentially to ensure the final collection is of uncompromising quality, stylistic coherence, and diversity. We obtain 4.3k excellent samples with raw images, short/detailed captions, style types and comprehensive scores. The visualization of the reference styles by class is shown as in the Figure 2.

2.2. Stylized Data Curation

Following the meticulous curation of our style references, the next step is to construct a large-scale, high-quality stylized data curation to capture the transfer dynamics across diverse content. The goal is to generate a comprehensive collection of challenging style-content pairs, which we then subject to a rigorous, multi-level evaluation.

Content-Style Pair Match. To ensure rigorous evaluation and maximize content diversity, we devise a systematic pairing strategy comprising both positive and negative samples. For positive pairs, we align each style reference with a specific content image from GenRef-wds [23] to establish the target stylization task. To construct negative pairs, we implement a bidirectional perturbation strategy: (1) we fix the style reference while randomly sampling distinct content images from other pairs of GenRef-wds [23], and (2) we maintain the content image while substituting the style reference with different examples. This design explicitly separates content structures from stylistic patterns, forcing the model to learn robust disentanglement rather than overfitting to specific image combinations. Consequently, this strategy yields a comprehensive dataset of approximately 600k pairs, ensuring a balanced distribution of challenging style transfer scenarios.

Style Transfer Methods. To prevent bias towards any single style transfer mechanism and to capture the full spectrum of current SOTA capabilities, we intentionally employ three diverse, state-of-the-art style transfer methods to generate the stylized results for every content-style pair. This ensemble approach allows us to ensure that the final dataset represents various levels of transfer fidelity and computational strateg=ies. We use Omnistyle[18], StyleID[8], and StyleShot [5] as the base models of image synthesis.

Evaluation Protocol. The 600k raw pairs are subjected to a comprehensive, multi-layer filtering process to distill the final high-quality dataset of approximately 300k positive and negative pairs. This protocol is critical not merely for culling bad data, but for mining discriminative positive and informative negative samples essential for robust training, as shown in Figure 3. The process is structured as follows:

1. Initial Automated Culling : We first utilize quantitative metrics to perform a high-throughput culling, immedi-

Evaluation Prompt of Style Pairs

Prompt (Style-Transfer Quality Judge). You are an expert judge for image style-transfer quality.
I will provide you with:
- One **ORIGINAL** content image.
- One **STYLE DESCRIPTION** (either an image or text that explains the target style).
- Three **CANDIDATE STYLIZED IMAGES**: **CANDIDATE 1**, **CANDIDATE 2**, **CANDIDATE 3**.

Your task is to:
1. Understand the content of the **ORIGINAL** image.
2. Understand the intended style from the **STYLE DESCRIPTION**.
3. Evaluate each **CANDIDATE** image on how well it transfers the **STYLE** while preserving the **CONTENT**.

Please follow this evaluation protocol and output format **exactly**.

EVALUATION DIMENSIONS
For **each** candidate image, score the following:

1) Stylistic Match (0–30)
- How well does the candidate follow the **STYLE DESCRIPTION**?
- Consider color palette, textures, rendering style, level of abstraction, lighting, and overall mood.
- 0 means "style is completely wrong", 30 means "style matches extremely well".

2) Content Preservation (0–30)
- How faithfully does the candidate keep the main objects, composition, and spatial layout of the **ORIGINAL** image?
- 0 means "content is lost or heavily changed", 30 means "content is preserved almost perfectly".

3) Technical Quality (0–20)
- Rendering quality: sharpness, lighting, shading, geometry, perspective, and absence of obvious artifacts.
- 0 means "serious technical defects", 20 means "clean and professional-quality rendering".

4) Overall Aesthetic & Usability (0–20)
- How visually pleasing and usable is this image as a final stylized result (e.g., for products, branding, or dataset)?
- Consider harmony, clarity, and how well style and content work together.
- 0 means "unattractive or confusing", 20 means "highly attractive and usable".

Total Score = Stylistic Match + Content Preservation + Technical Quality + Overall Aesthetic

---

OUTPUT FORMAT
Return your answer in the following structured format, and **do not** add extra sections:

1. Metadata
- Original Caption: `<1--2 sentences describing the ORIGINAL content image.>`
- Style Summary: `<1--2 sentences summarizing the target style from the STYLE DESCRIPTION.>`
- Content Tags: `["tag1", "tag2", ...]` (for main objects / scene)
- Style Tags: `["tag1", "tag2", ...]` (for style / rendering / mood)

2. Per-Candidate Evaluation
For **each** candidate (1, 2, 3), output:
Candidate X:
- Stylistic Match: `<score>`/30. Justification: `<1--2 sentences>`.
- Content Preservation: `<score>`/30. Justification: `<1--2 sentences>`.
- Technical Quality: `<score>`/20. Justification: `<1--2 sentences>`.
- Overall Aesthetic & Usability: `<score>`/20. Justification: `<1--2 sentences>`.
- Total Score: `<sum>`/100.

3. Ranking & Verdict
Ranking (Best → Worst):
- Candidate `<id>` (Total: `<score>`/100)
- Candidate `<id>` (Total: `<score>`/100)
- Candidate `<id>` (Total: `<score>`/100)

4. Final Recommendation

- If at least one candidate has Total Score ≥ 80 and no severe artifacts:
  Recommendation: **ACCEPT**.
  Comment: `<1--2 sentences summarizing why the best candidate is suitable.>`
- Otherwise:
  Recommendation: **REJECT**.
  Comment:`<1--2 sentences summarizing what is mainly wrong and what should be improved.>`

Important:

- Always give concrete visual reasons comparing to **both** the ORIGINAL image and the STYLE DESCRIPTION.
- Be strict with scores: reserve very high scores (Total ≥ 90) only for truly excellent, near-perfect results.

ately rejecting samples with severe technical defects (extreme distortion or oversmoothing).

2. Positive Sample Mining : Each result is designated as a positive sample only if it achieves a stringent quality consensus across a hierarchical three-tiered system. This protocol requires a successful pass through: (a) Qwen2.5-VL-72B [2] for basic coherence and quality gating; (b) Gemini 2.5 Pro [4] for nuanced semantic style/content alignment; and (c) Human Annotators for final gold-standard validation of aesthetic quality and consistency. Only samples confirming strong style coherence, high technical quality, and faithful content preservation at all stages are retained.
3. Hybrid Negative Sample Mining : We adopt a dual-source strategy to construct a robust negative set. First, we directly utilize the Negative Buffer established via the bidirectional perturbation strategy, where the generated images inherently represent mismatched content-style pairs. Second, we mine hard negatives from the rejected candidates in the positive buffer,specifically targeting instances where state-of-the-art models failed to produce satisfactory transfers (including *style collapse* or *content distortion*). By combining these explicitly mismatched pairs with challenging generation failures, we provide crucial discriminative signals that prevent the model from learning shortcut solutions.

This two-pronged mining strategy ensures the final 300k dataset provides rich supervision, comprising successful examples and discriminative failure cases to maximize the model's learning capacity, as shown in Figure 4.

3. Experimental Details

3.1. Evaluation Metrics

For ImgEdit [21] benchmark, we use the VLM `GPT-4o` [1] and `Gemini 2.5 Pro` [4] as the verifier (Success@4). For AnyEdit [22] benchmark, we adopt four similarity metrics following prior works [15, 17]:

- CLIP image similarity (CLIP_{img}): measuring change between edited and input image with CLIP [14].
- CLIP output similarity (CLIP_{out}): measuring edited image similarity with output caption,
- L1 pixel-distance between input and edit image,
- DINO similarity between the DINO [16] embeddings of input and edited images.

3.2. Training Logs and Generalizability

To verify the stability and effectiveness of StyleScore, we visualize the training dynamics in Figure 6. The training logs record the evolution of the average reward during the post-training phase. As evidenced by the curves, the model exhibits a consistent learning trajectory. In the early stages, the reward scores rise sharply, indicating that the policy is rapidly adapting to the semantic stylistic preferences defined by StyleScore. Subsequently, the curve stabilizes at a high reward level without exhibiting significant oscillations. This stable convergence serves as strong evidence for the effectiveness of StyleScore in guiding preference alignment during the reinforcement learning phase.

4. Additional Experiment Results

4.1. Effect of Reward Model Choice

We next investigate the impact of the reward model choice on the Post-Training stage. We compare the performance of our full pipeline when guided by different reward models. Specifically, we fix the model trained with SFT as the starting point and then apply GRPO using either our StyleScore or other general, publicly available reward models, such as ImageReward [20] and HPSv3 [12]. The results are presented in Table 1. When using the general-purpose ImageReward and Hpsv3 , the performance on ImgEdit only improves marginally from the SFT trained score of 4.50 to 4.55 and 4.52, respectively. In contrast, using our StyleScore significantly boosts the performance to 4.74. This result demonstrates the superiority of our reward model. General models fail to provide a precise gradient signal for the specific task of style-content disentanglement, but Style-

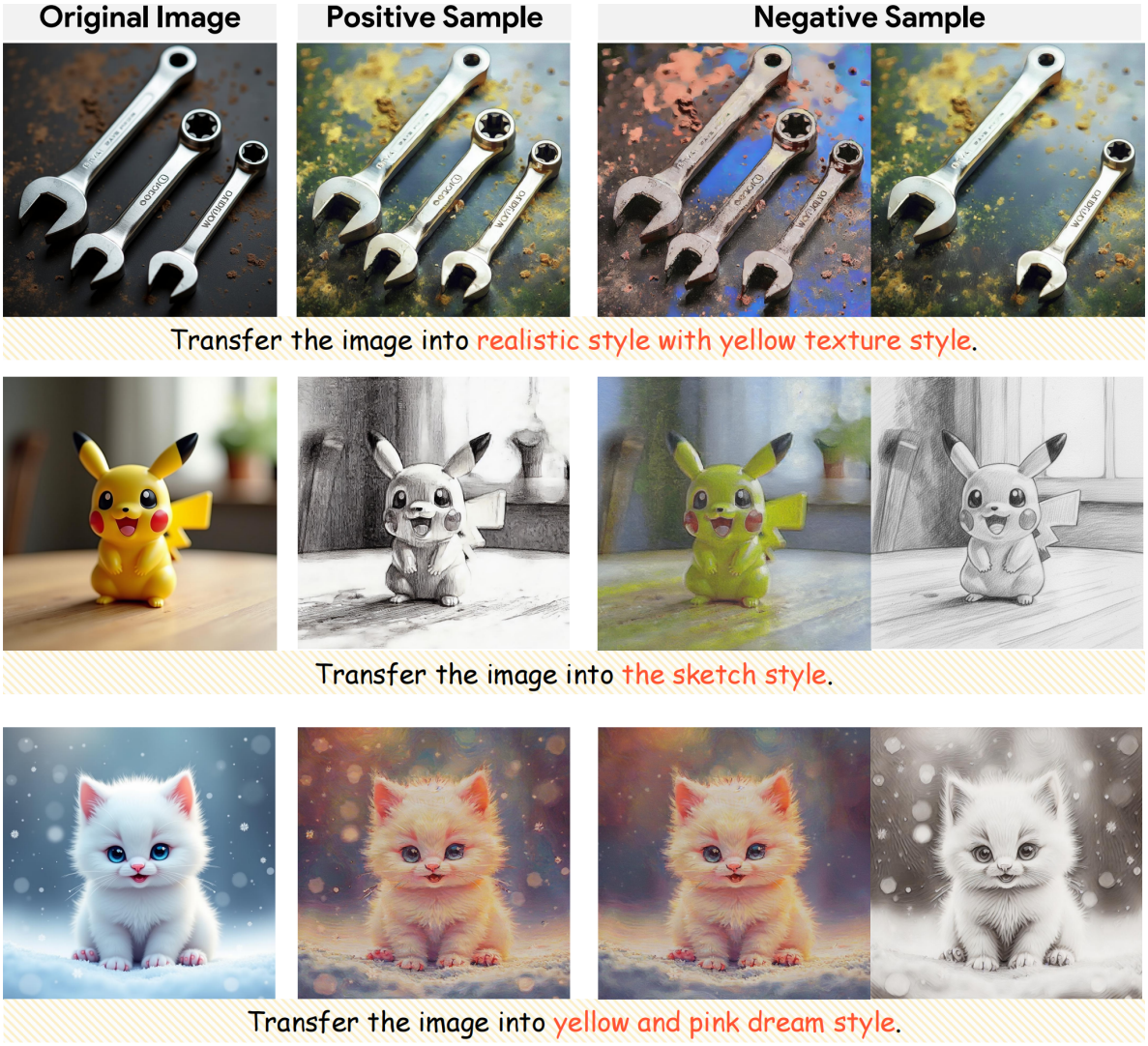

Figure 4. **Examples of positive and negative samples in our dataset.** We show the original content images alongside their corresponding positive stylizations and mined negative samples based on the text instructions.

Table 1. Ablation study on the choice of reward model for the GRPO stage. All models use the same `+SFT` checkpoint.

Method	**ImgEdit** [21]		**StyleScore**
	`GPT-4o`	`Gemini`	
SFT Only	4.50	4.34	3.82
+ GRPO (Hpsv3[12])	4.52	4.12	3.65
+ GRPO (ImageReward[20])	4.55	3.97	3.56
+ GRPO (Ours)	4.74	4.46	3.91
GRPO only (Ours)	4.68	4.30	3.85

Score provides effective and reliable reward.

4.2. Effect of Timestep-Aware Reward Weighting

Standard GRPO optimization applies the reward signal uniformly across all denoising steps. However, Flow Matching-based models inherently exhibit a coarse-to-fine generation hierarchy: the early denoising steps dominate the formation of global semantic structure and stylistic atmosphere, while the later steps focus on refining high-frequency details. For our style transfer task, focusing the reward signal on the early timesteps is more effective than uniform weighting. Consequently, we introduce a timestep-aware reweighting strategy that aligns the reward density with the model's intrinsic generation dynamics as described in Sec 1.2. Specifically, we employ an exponential decay function to prioritize the early timesteps, ensuring the reward signal is most potent during the critical phase of style

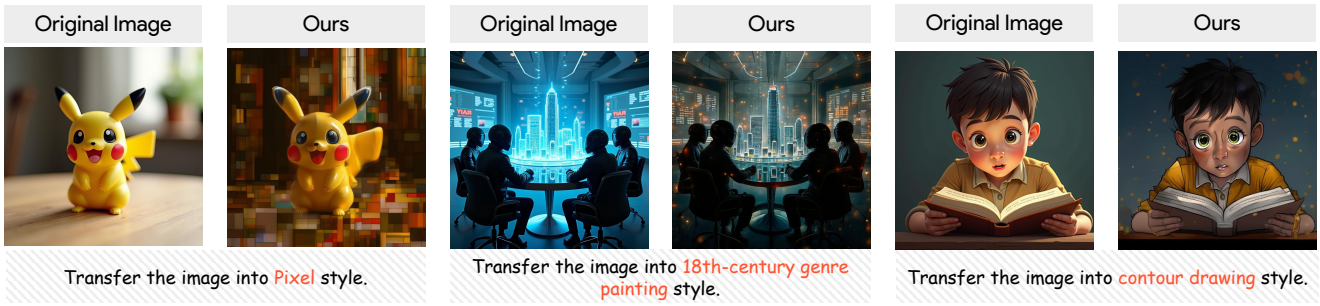

Figure 5. Failure case during the instruction-guided style transfer.

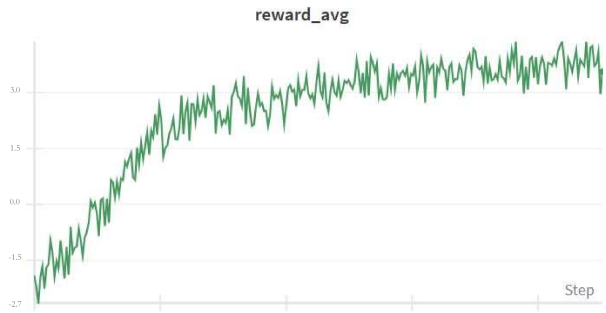

Figure 6. **Training reward curves.** The consistent upward trend and subsequent convergence demonstrate that our proposed method effectively optimizes the objective function, steadily improving the generation quality on both backbone architectures.

injection. As shown in Table 2, comparing our approach against the uniform baseline reveals consistent improvements across both ImgEdit and StyleScore metrics. This validates that concentrating optimization guidance on the structural formation stage significantly enhances the robustness of style-content disentanglement.

Table 2. The timestep-aware reward weighting strategy.

Method	ImgEdit GPT-4o	ImgEdit Gemini	StyleScore
w/o. reweight	4.69	4.37	3.84
w. reweight	**4.74**	**4.46**	**3.91**

4.3. More Qualitative Results

To further demonstrate the diversity and robustness of our Style-GRPO framework, we present additional generation results in Figure 7. Our method generalizes exceptionally well across a broad spectrum of artistic domains defined in our StyleReward-Dataset. Crucially, the model not only captures the textures of these styles, but also maintains high semantic fidelity. This confirms that Style-GRPO effectively decouples style representation from semantic content, ensuring that the structural integrity of original images is preserved while applying globally consistent stylization.

4.4. Failure Case Analysis

While Style-GRPO demonstrates robust stylization capabilities across diverse scenarios, we explicitly analyze its limitations in the field. As illustrated in Figure 5, failure cases predominantly arise in scenarios involving semantic conflicts or abstraction prompts. When the target style requires a alteration of the geometric structure, the model occasionally exhibits under-stylization. This occurs because the model retains a strong bias towards preserving the fine-grained content details of the source image, which inherently contradicts the simplification required by abstract styles. Second, distinct semantic gaps can hinder style injection. In such cases, the strong visual priors of the content may overpower the target stylistic textures, resulting in a hybrid output where the style is not fully dominant. These observations suggest that balancing rigid content preservation with stylistic abstraction remains a significant challenge for current diffusion-based approaches.

5. Broader Impact

5.1. Further Work

Beyond the current focus on text-guided style transfer, our framework opens up several key avenues for future research, leveraging the unique strengths of the StyleScore reward model and the Style-GRPO pipeline. A primary direction is the extension to reference-image-guided style transfer. This transition is non-trivial, requiring adaptation of the StyleScore model to robustly extract style features from an arbitrary reference image and training the Style-GRPO pipeline to manage the complex feature alignment and long-range dependencies inherent in image-to-image style guidance. Success in this area would dramatically enhance the model's versatility and enable finer artistic control.

Furthermore, the StyleScore reward model can be utilized to enhance controllability and interpretability in generative models. By analyzing the gradient of StyleScore with respect to the generated image, the framework could be ex-

tended to identify regions of an image that contribute most to the style score. This could enable novel applications such as localized style editing, where the user can selectively apply, refine, or remove a style based on the StyleScore-derived sensitivity map.

5.2. Limitations

A fundamental constraint is the inherent confinement of the system to static image stylization. The critical barrier to generalization lies in extending this human preference alignment to the more demanding domain of video style transfer. Unlike independent image processing, video stylization imposes stringent requirements for temporal coherence. The successful transfer of style must go beyond aesthetic quality on individual frames to maintain seamless stability and suppress flickering artifacts across the entire sequence. The core human preference learning engine StyleScore operates exclusively within the spatial domain of static images. Consequently, the model lacks the architectural mechanism to perceive or evaluate inter frame consistency, which is pivotal for natural cinematic sequences. Addressing this limitation by scaling the reward model to enforce robust spatiotemporal coherence represents the crucial next step toward the fully generalized and human centric generative systems.

6. Acknowledgment

This work was supported in part by Natural Science Foundation of China (No.62332002, No.62202014), and Shenzhen KQTD (No.20240729102051063)

References

[1] Gpt-4o. https://openai.com/index/hello-gpt-4o/. 2, 4, 6

[2] Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, et al. Qwen2. 5-vl technical report. *arXiv preprint arXiv:2502.13923*, 2025. 2, 4, 6

[3] Siyu Cao, Hangting Chen, Peng Chen, Yiji Cheng, Yutao Cui, Xinchi Deng, Ying Dong, Kipper Gong, Tianpeng Gu, Xiusen Gu, et al. Hunyuanimage 3.0 technical report. *arXiv preprint arXiv:2509.23951*, 2025. 4

[4] Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen Sachdeva, Inderjit Dhillon, Marcel Blistein, Ori Ram, Dan Zhang, Evan Rosen, et al. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint arXiv:2507.06261*, 2025. 6

[5] Junyao Gao, Yanan Sun, Yanchen Liu, Yinhao Tang, Yanhong Zeng, Ding Qi, Kai Chen, and Cairong Zhao. Styleshot: A snapshot on any style. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. 2, 4

[6] Xiaoxuan He, Siming Fu, Yuke Zhao, Wanli Li, Jian Yang, Dacheng Yin, Fengyun Rao, and Bo Zhang. Tempflow-grpo: When timing matters for grpo in flow models. *arXiv preprint arXiv:2508.04324*, 2025. 1

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

[8] Minh-Ha Le and Niklas Carlsson. Styleid: Identity disentanglement for anonymizing faces. *arXiv preprint arXiv:2212.13791*, 2022. 4

[9] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022. 1

[10] 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

[11] Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *arXiv preprint arXiv:2209.03003*, 2022. 1

[12] 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. 6, 7

[13] Saif Mohammad and Svetlana Kiritchenko. Wikiart emotions: An annotated dataset of emotions evoked by art. In *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)*, 2018. 2

[14] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR, 2021. 6

[15] Shelly Sheynin, Adam Polyak, Uriel Singer, Yuval Kirstain, Amit Zohar, Oron Ashual, Devi Parikh, and Yaniv Taigman. Emu edit: Precise image editing via recognition and generation tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8871–8879, 2024. 6

[16] Oriane Siméoni, Huy V Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, et al. Dinov3. *arXiv preprint arXiv:2508.10104*, 2025. 6

[17] Xinghai Sun, Changhu Wang, Avneesh Sud, Chao Xu, and Lei Zhang. Magicbrush: Image search by color sketch. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 475–476, 2013. 6

[18] Ye Wang, Ruiqi Liu, Jiang Lin, Fei Liu, Zili Yi, Yilin Wang, and Rui Ma. Omnistyle: Filtering high quality style transfer data at scale. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 7847–7856, 2025. 4

[19] Chenfei Wu, Jiahao Li, Jingren Zhou, Junyang Lin, Kaiyuan Gao, Kun Yan, Sheng-ming Yin, Shuai Bai, Xiao Xu, Yilei Chen, et al. Qwen-image technical report. *arXiv preprint arXiv:2508.02324*, 2025. 4

[20] Jiazheng Xu, Xiao Liu, Yuchen Wu, Yuxuan Tong, Qinkai Li, Ming Ding, Jie Tang, and Yuxiao Dong. Imagereward: Learning and evaluating human preferences for text-to-image generation. *Advances in Neural Information Processing Systems*, 36:15903–15935, 2023. 6, 7

[21] Yang Ye, Xianyi He, Zongjian Li, Bin Lin, Shenghai Yuan, Zhiyuan Yan, Bohan Hou, and Li Yuan. Imgedit: A unified image editing dataset and benchmark. *arXiv preprint arXiv:2505.20275*, 2025. 6, 7

[22] Qifan Yu, Wei Chow, Zhongqi Yue, Kaihang Pan, Yang Wu, Xiaoyang Wan, Juncheng Li, Siliang Tang, Hanwang Zhang, and Yueting Zhuang. Anyedit: Mastering unified high-quality image editing for any idea. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 26125–26135, 2025. 6

[23] Le Zhuo, Liangbing Zhao, Sayak Paul, Yue Liao, Renrui Zhang, Yi Xin, Peng Gao, Mohamed Elhoseiny, and Hongsheng Li. From reflection to perfection: Scaling inference-time optimization for text-to-image diffusion models via reflection tuning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15329–15339, 2025. 4

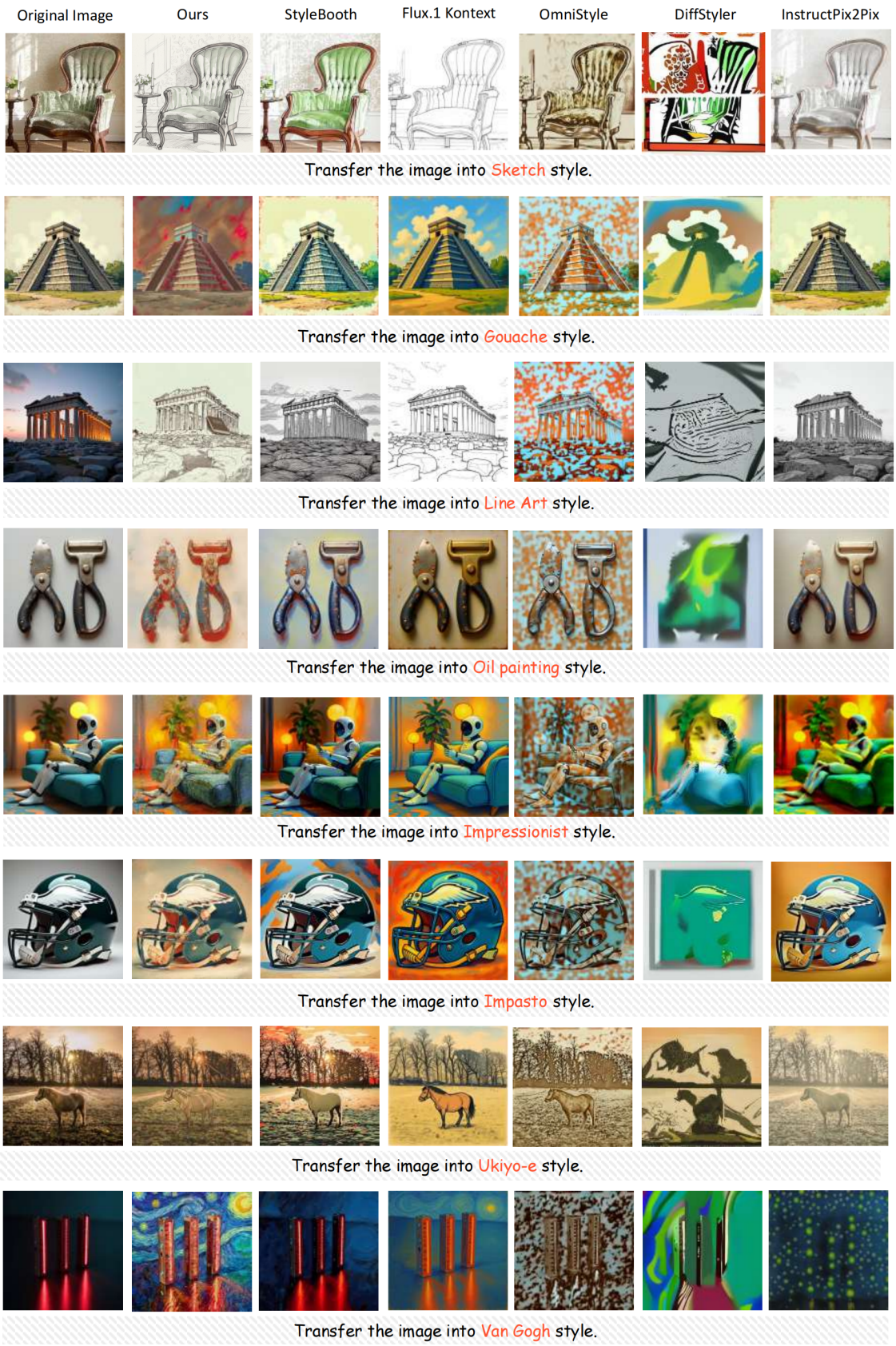

Figure 7. More qualitative results of the instruction-guided style transfer.