SCAdapter: Content-Style Disentanglement for Diffusion Style Transfer Supplementary Material

Luan Thanh Trinh
LY Corporation
ttrinh@lycorp.co.jp

1. Additional applications - Text-driven stylized synthesis.

Text-driven stylized synthesis is a highly sought-after feature that has attracted significant research interest. In this task, given a text prompt and a style reference image, users aim to generate images that embody similar styles. The proposed method, with its ability to extract pure style features from the style reference image, is promising for delivering good results in text-driven stylized synthesis.

To adapt our approach for this task, we removed the content extractor branch from SCAdapter. Fig. 1 showcases example results where our method demonstrates strong alignment with both reference styles and input prompts. For an objective evaluation, we conducted a quantitative comparison with several prominent models in text-driven stylized synthesis, as shown in Tab. 1. We assessed text alignment capability (TA) by measuring cosine similarity within the CLIP text-image embedding space between the textual prompts and their corresponding synthesized images. To evaluate style similarity, we utilized SFS from Eq. 13.

The quantitative results in Tab. 1 indicate that, although not specifically designed for this task, the proposed method performs competitively when compared to other methods. Notably, it achieves highest TA scores, indicating a strong alignment between input prompts and output content. This success is due to the extraction of pure style features from the style reference image, effectively minimizing unnecessary influences from its content features.

Table 1. Quantitative comparison with other text-driven stylized synthesis baselines.

Method	InST	CAST	StyTR2	T2I-Adapter	DEADiff	Ours
TA	0.237	0.282	0.282	0.224	0.284	**0.285**
SFS	0.238	0.244	0.234	**0.271**	0.259	0.266

2. Qualitative comparison with ablation of KVS Injection and Style Consistency Loss

To highlight the effects of KVS Injection (denoted as KVS in Fig. 2) and the style consistency loss (denoted as SCL in Fig. 2), we present qualitative examples of style transfer results obtained by ablating these components. Consistent with the quantitative findings reported in the main paper, both KVS Injection and the style consistency loss contribute to improving the final performance of our proposed model, with KVS Injection playing a particularly significant role. For the photo-realistic transfer task (row 1), snowflake effects are added more abundantly and in a more harmonious manner when KVS Injection is applied. For the artistic transfer task (row 2), the painting style characteristics are expressed more vividly. Meanwhile, although the style consistency loss provides measurable improvements, its impact is less apparent from a qualitative perspective. These observations indicate that the proposed model remains effective when trained on standard image datasets, without the need to rely on a content–style–stylized triplet dataset.

3. More results of artistic transfer

Fig. 3 presents several additional results of our proposed model applied to artistic style transfer. As shown, the model produces visually compelling outputs that effectively capture the artistic characteristics of the reference styles while preserving the structural content of the original images.

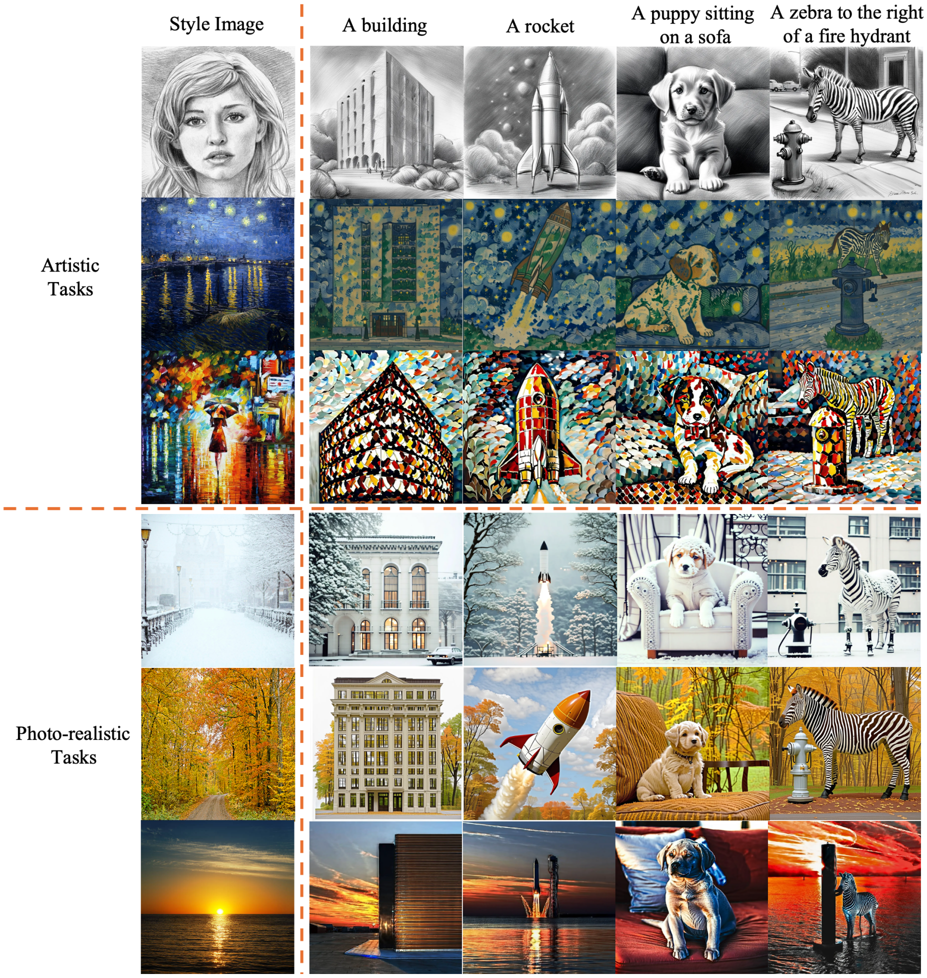

Figure 1. SCAdapter delivers high-quality text-driven stylized synthesis for both artistic and photo-realistic tasks.

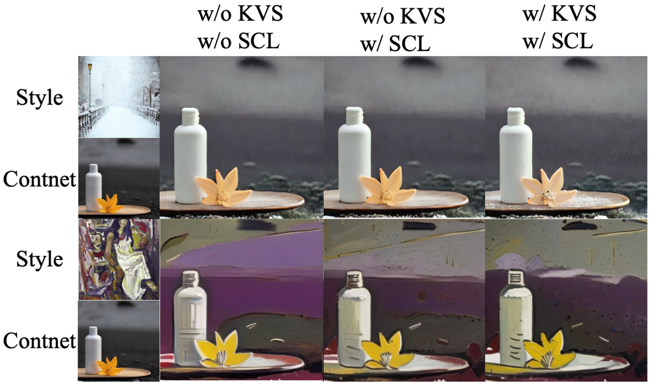

Figure 2. Qualitative comparison while ablating the KVS Injection and Style Consistency Loss.

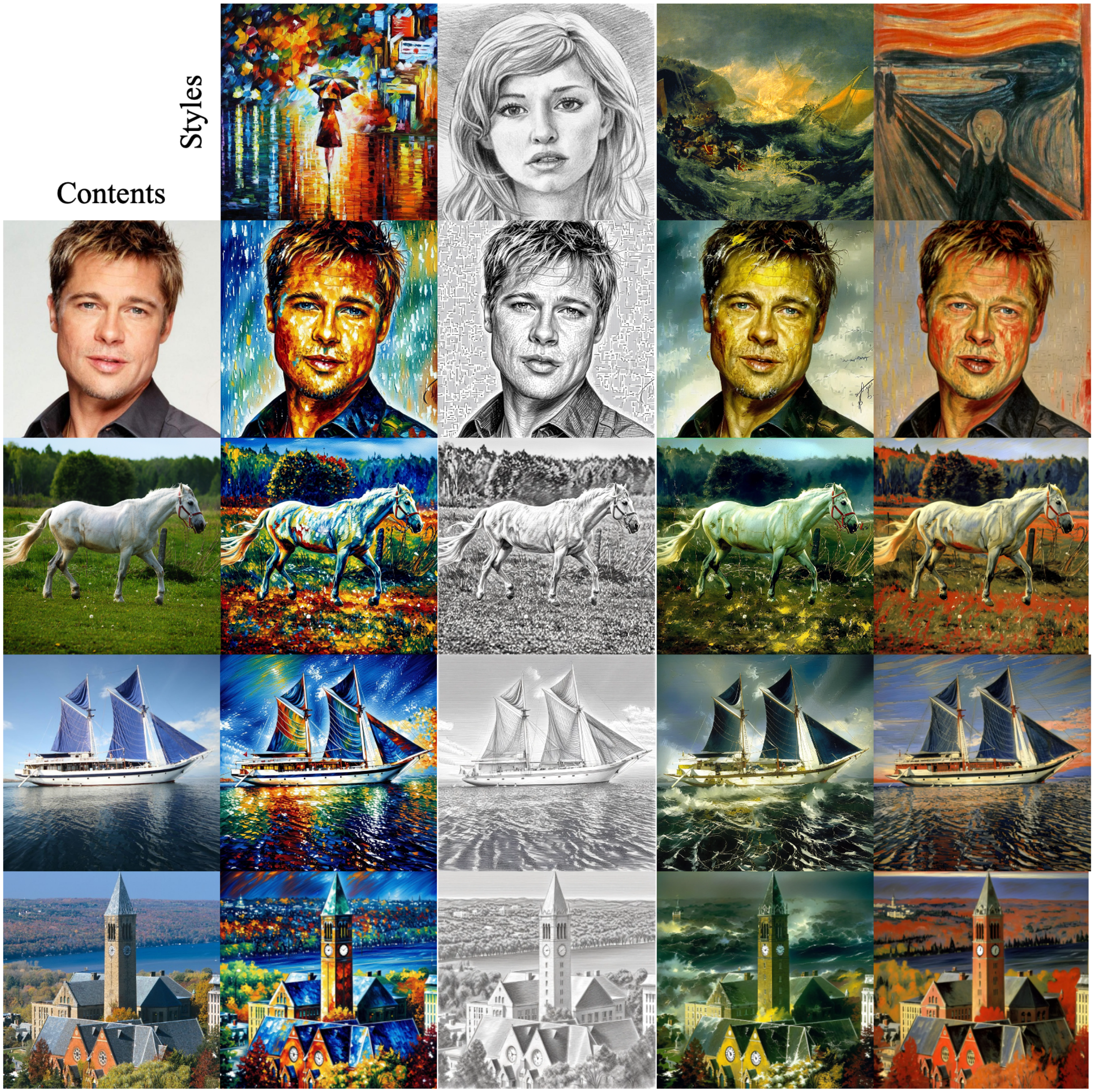

Figure 3. More results of artistic transfer. Zoom in for viewing details.