Domain Generalizable Portrait Style Transfer

Supplementary Material

Figure 12. **Our method can generate stylized images at 1024×1024 resolution on a GPU of 24GB memory.**

Figure 13. **Results on non-portrait images.**

Metric	PPST	Chen et al.	Deng et al.	StyleID	I.S.+	Ours
Pref.↑	0.056	0.040	0.014	0.116	0.038	**0.736**

Table 4. **User Study results.**

7. Appendix

7.1. User Study

We conduct a user study with 50 participants to further evaluate the methods. We randomly select 100 pairs of portraits from various domains and generate style transfer results using different approaches. Participants are asked to choose the best result in each group, based on both content preservation and stylization strength. The results are shown in table 4, where we can see that our results are more preferred by human subjects.

7.2. High-resolution Generation

Figure 12 shows that our method is capable of performing style transfer at 1024×1024 resolution. Note that this requires less than 24GB of memory, which is fewer than most previous diffusion-based style transfer methods [3, 5, 36].

7.3. Generalization Ability

Generalization on non-portrait images. Figure 13 demonstrates that our model, trained on a portrait dataset, can still produce satisfactory results for non-portrait images, such as animals.

Examples with more abstract or geometrically distorted styles. Our method can handle abstract and geometrically distorted styles like cubist or Picasso artworks, as shown in Figure 14.

Figure 14. **Results on more abstract or geometrically distorted styles.**

Figure 15. **Effect of applying AdaIN-wavelet normalization at different stages.**

Figure 16. **Effect of initializing latents by adding noise to input at different levels.**

7.4. More Ablation Studies on Latent Initialization

Apply the AdaIN-wavelet norm on latent at timestep 0 or in image space. In Figure 15, we show the qualitative results of applying AdaIN-wavelet normalization at different stages. The Gram loss / LPIPS scores for applying it in image space, at timestep 0, and at timestep T are 0.691 / 0.096, 0.804 / 0.114, and **0.657** / **0.083**, respectively. These results demonstrate that applying the proposed normalization at timestep T yields the best performance.

Initializing from latents at different noise levels. We present additional results in Figure 16. Initializing latents by adding noise at different levels (t = 699, 799, 899, 999) preserves content details but weakens stylization as t decreases. In contrast, our AdaIN-Wavelet produces stronger stylization without leading to significant content loss. The Gram loss / LPIPS scores at these timesteps for our method are 4.459 / **0.062**, 3.916 / 0.065, 3.503 / 0.069, 3.132 / 0.076, and **0.657** / 0.083, respectively.

7.5. Effect of ControlNet Conditioning Scale

Our method offers flexibility in controlling the level of detail by adjusting the input to ControlNet and its conditioning scale (s). As shown in Figure 17, with the decrease of s, the preserved content detail is also reduced accordingly, while

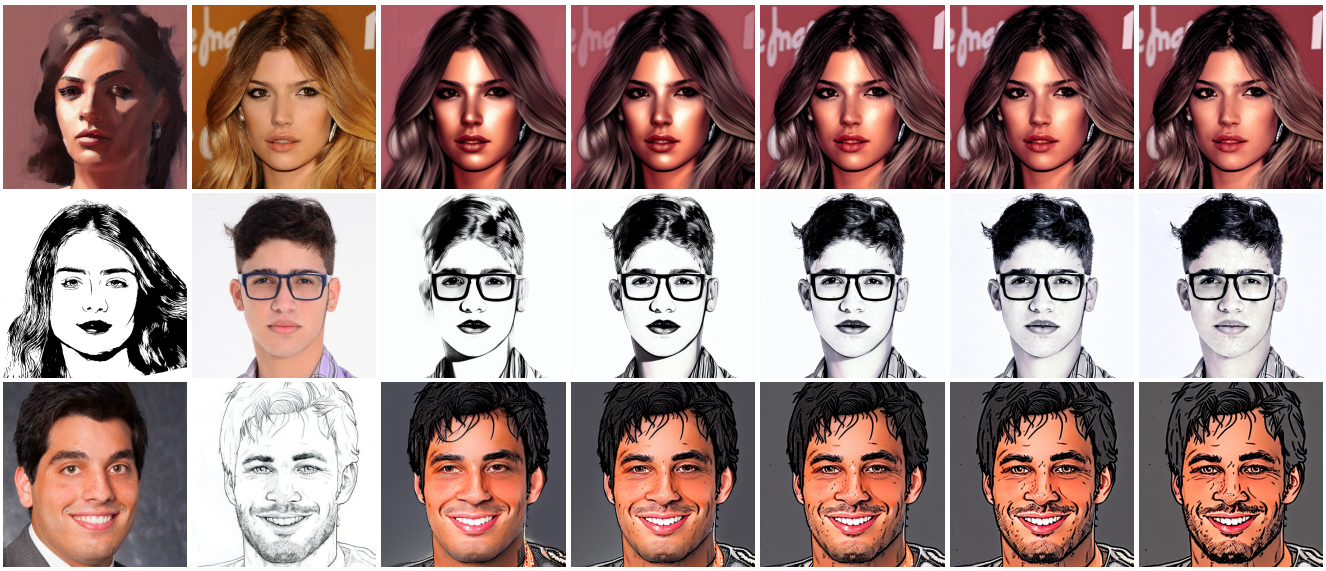

Figure 17. **Effect of the ControlNet conditioning scale.**

Method	Gram loss↓	LPIPS↓	ID↓
s = 0.00	**0.341**	0.338	0.503
s = 0.25	0.558	0.230	0.293
s = 0.50	0.667	0.158	0.167
s = 0.75	0.700	0.126	0.120
Full method (s=0.90)	0.657	**0.083**	**0.087**

Table 5. **Quantitative results of different ControlNet conditioning scale.**

Figure 18. **Effect of input of style adapter.** As can be seen, using original reference may result in inconsistent stylization in hair regions.

the texture will get closer to the reference. The quantitative results are also shown in Table 5.

7.6. Effectiveness of Using Warped Reference

By starting sampling from the latent after the proposed AdaIN-Wavelet transform, the stylization is expected to be semantically aligned. However, as shown in Figure 18, directly feeding original reference into style adapter still may result in stylization between semantic unrelated regions, especially when reducing the ControlNet conditioning scale.

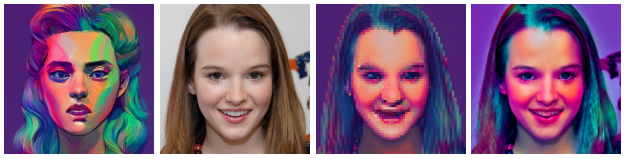

Figure 19. Failure cases.

7.7. Limitations

A failure case is shown in Figure 19. As seen, our method fails to transfer all the color tone of the reference portrait. Because the hair color and the skin of input is homogeneous, the learned correspondence will match these regions to similar target regions. An interesting research direction is to investigate more controllable correspondence to solve the problem.

7.8. More Results

Figure 20 shows more comparison of sketch colorization with previous methods [3, 36, 38]. Figure 21 compares the performance of our method and other method on old photo color restoration. Figure 22, 23, and 24 show more style transfer comparison results with previous methods. As shown, our method consistently produces high-quality results for portraits from various domains. It outperforms in achieving more effective stylization between the semantically corresponding regions and effectively preserving the content details and facial identity, regardless of the domains of input portraits.

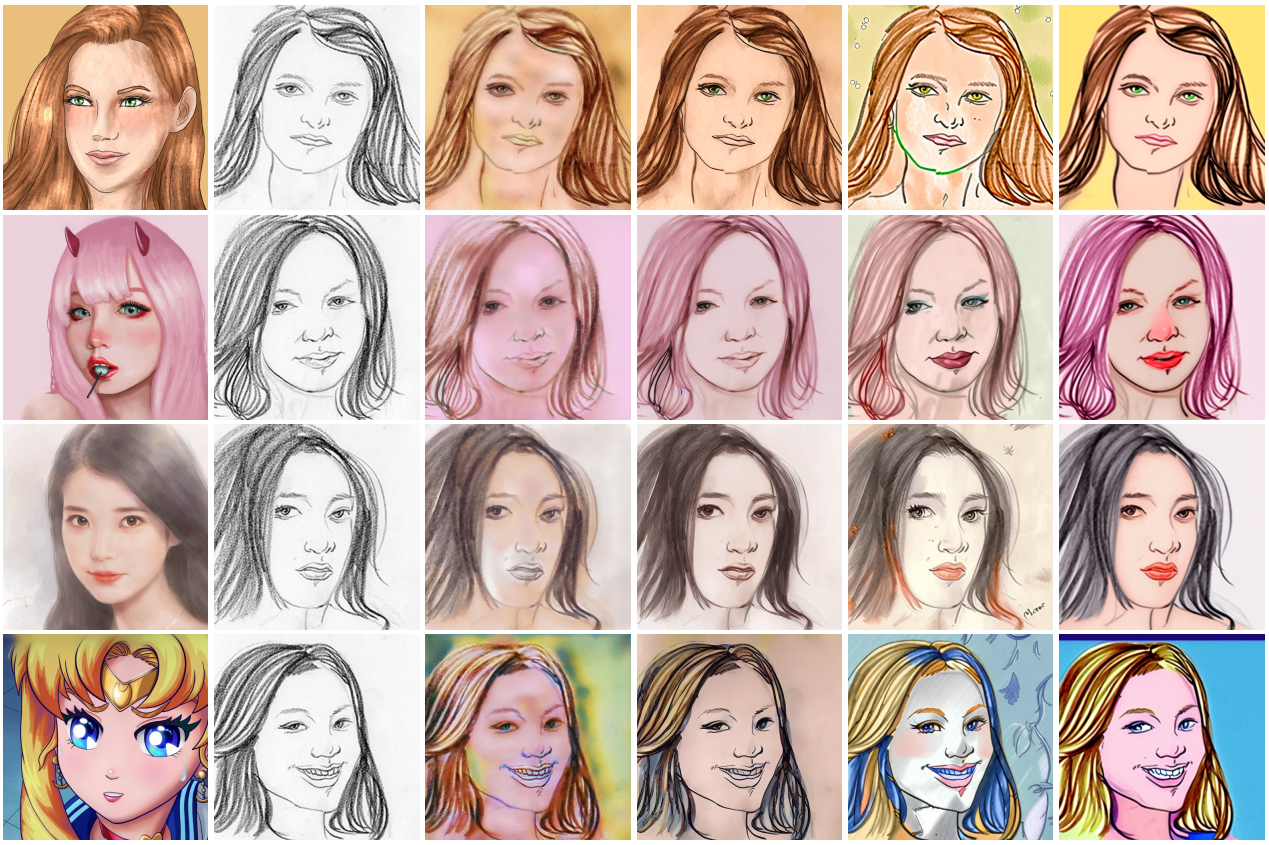

Figure 20. More results on sketch colorization.

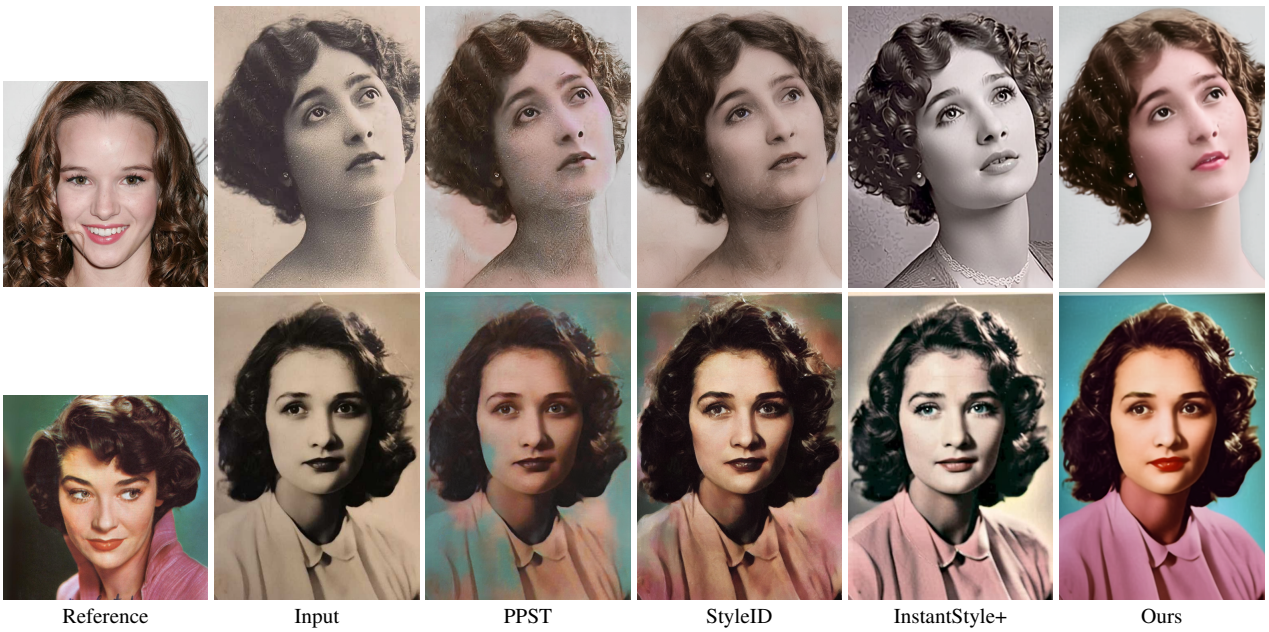

Figure 21. More results on old photo color restoration.

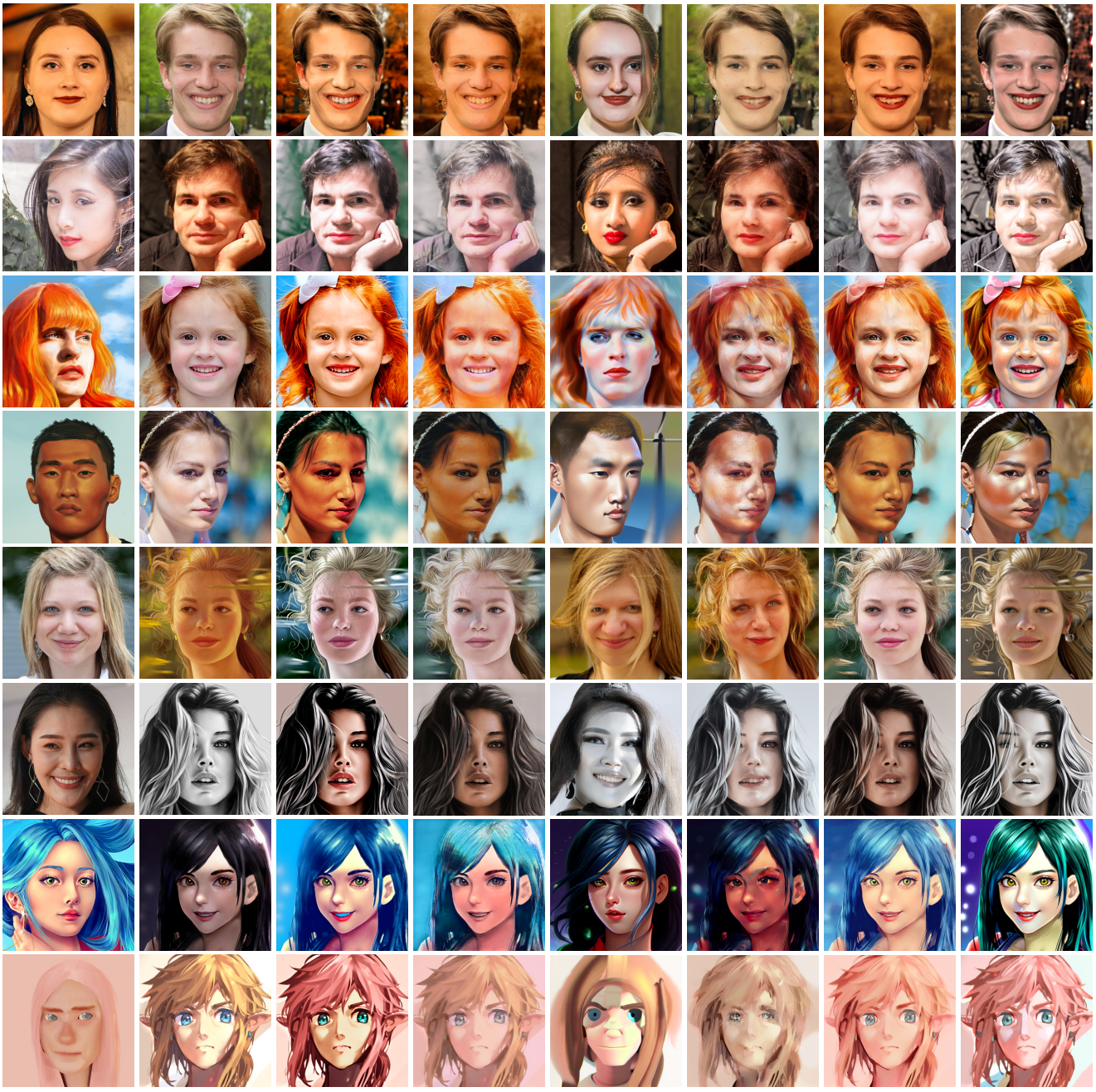

Figure 22. **Visual comparison with previous style transfer methods.**

Figure 23. **Visual comparison with previous style transfer methods.**

Figure 24. **Visual comparison with previous style transfer methods.**