Supplementary for Painterly Image Harmonization via Adversarial Residual Learning

Xudong Wang, Li Niu*, Junyan Cao, Yan Hong, Liqing Zhang*
Department of Computer Science and Engineering, MoE Key Lab of Artificial Intelligence,
Shanghai Jiao Tong University
{wangxudong1998,ustcnewly,Joy_C1,hy2628982280,zhang-lq}@sjtu.edu.cn

In this supplementary, we will first provide more visualization results of different methods in Section 1. Then, we will analyze the impact of adding residual features to different layers in Section 2. Then, we will show the harmonized results of the same foreground pasted on different backgrounds in Section 3. We will compare different adversarial losses in Section 4 and show the results of multiple foregrounds in Section 5.

1. Visual Comparison with Baselines

We choose the competitive baselines SANet [5], AdaAttN [3], StyTr2 [2], E2STN [6], DPH [4] from two groups of baselines, in which E2STN and DPH are from painterly image harmonization group while the rest are from the artistic style transfer group. In Figure 2, we show the harmonized results generated by baseline methods and our method. Compared with these baselines, our method can successfully preserve the foreground content and transfer style from background image.

For example, our method can preserve fine-grained details (*e.g.*, row 1, 2) and sharp contours (*e.g.*, row 8) while transferring the style, which achieves a better balance between style and content. In contrast, the baseline methods may under-stylize the foreground so that the foreground is not harmonious with background, or severely distort the content structure so that the foreground is hardly recognizable. In some challenging cases, our method can better transfer the style (*e.g.*, color, texture) and obtain more visually appealing results (*e.g.*, row 4, 5, 6, 9), while the baselines fail to make the foreground style compatible with the background. Overall, in our harmonized images, the foreground is properly stylized and harmonious with the background so that the whole image appears to be an intact artistic painting.

*Corresponding author.

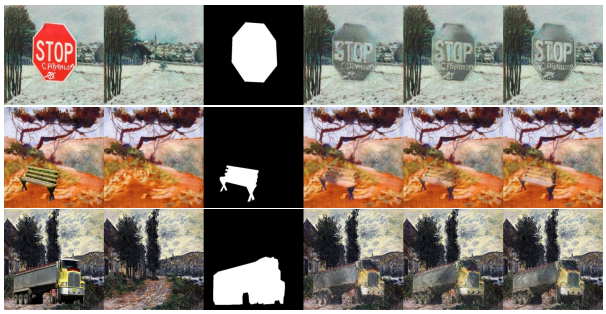

Figure 1. The harmonization results obtained by adding residual features to different layers.

2. Adding Residuals to Different Layers

As described in Section 3.2 in the main paper, we employ four residual blocks to learn four layers of residual features. For the l-th layer of the main encoder, the learned residual feature F_r^l, which is the output from the l-th residual block, is added to the foreground region in the stylized feature map F_a^l, leading to the refined feature map $\tilde{F}_a^l$.

By default, we add residual features to all four encoder layers, that is, $l = 1, \ldots, 4$. In this section, we investigate the impact of adding learned residual features to only two shallow layers ($l = 1, 2$) or only two deep layers ($l = 3, 4$). As shown in Figure 1, we observe that adding residual features only to partial layers may lose some detailed information (*e.g.*, small letters on the stop sign in row 1, the front of the truck in row 3) or generate undesired artifacts (*e.g.*, black spots on the chair in row 2), probably because some layers of feature maps are not well-harmonized. Instead, after adding residual features to all layers, our method can produce harmonized results with sharp details, smooth appearance, and reasonable colors, which demonstrates the effectiveness of modulating all layers of feature maps.

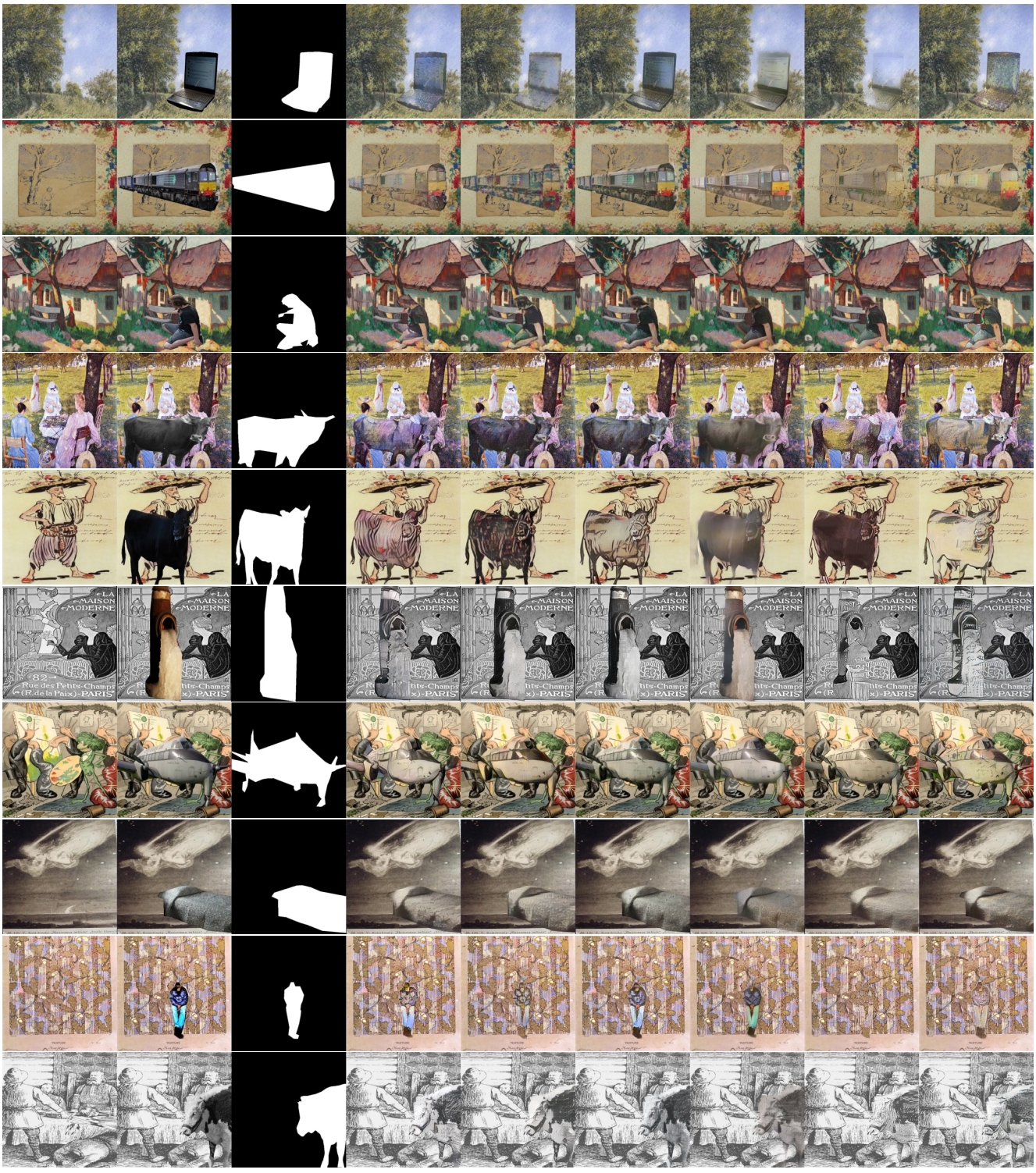

Figure 2. From left to right, we show the background image, composite image, foreground mask, the harmonized results of SANet [5], AdaAttN [3], StyTr2 [2], E2STN [6], DPH [4], and our PHARNet.

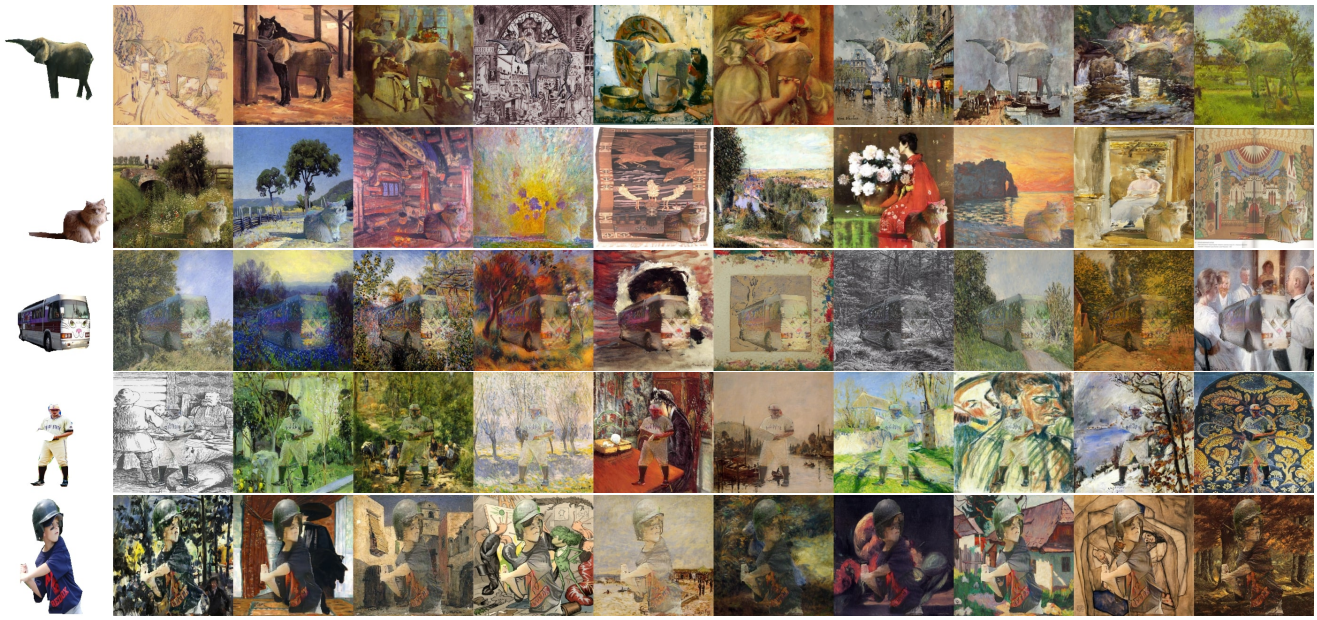

Figure 3. From left to right, we show the foreground object, the harmonized results of the same foreground pasted on different ten background pictures.

3. The Same Foreground on Different Backgrounds

We show the harmonized results when pasting the same foreground on different background images in Figure 3. We observe that with the preserved content structure, the foreground could be sufficiently stylized and harmonious with different backgrounds, which demonstrates the generalization ability of our method to cope with various combinations of foregrounds and backgrounds.

4. Comparing Different Adversarial Losses

We change our pixel-wise adversarial loss used for both encoder feature maps and output images to the vanilla adversarial loss in [6] and the domain verification adversarial loss in [1], while keeping the other network components unchanged. The adversarial losses in [6] and [1] represent image-wise adversarial loss and region-wise adversarial loss respectively. Therefore, we actually compare three types of adversarial losses: image-wise, region-wise, and pixel-wise adversarial losses.

We show the visual comparison below, which demonstrates that pixel-wise adversarial loss performs far better than other types of adversarial losses. Additionally, we invite 50 users to select from three methods for 100 composite images, which shows that 87% users choose our method.

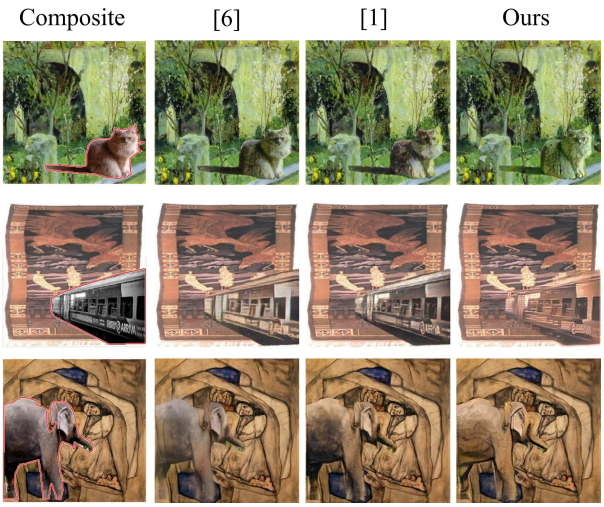

Figure 4. From left to right, we show the composite image, the harmonized results of using adversarial loss of [6], [1] and our method.

5. Multiple Foregrounds on One Background

Our method can be directly applied to the test images with multiple composite foregrounds. We can just feed the composite image and mask with multiple foregrounds, passing through the network once. We show some results in Figure 5, which shows that our method can harmonize multiple foregrounds simultaneously.

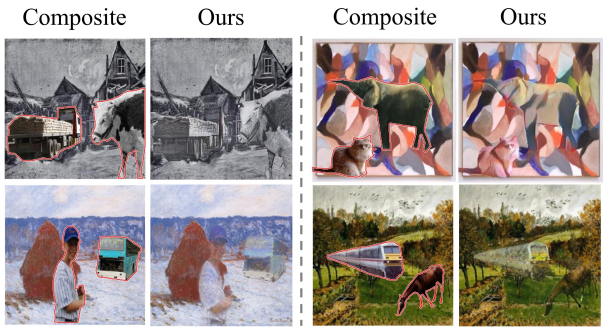

Figure 5. Example of multiple foregrounds on one background.

References

[1] Wenyan Cong, Jianfu Zhang, Li Niu, Liu Liu, Zhixin Ling, Weiyuan Li, and Liqing Zhang. Dovenet: Deep image harmonization via domain verification. In *CVPR*, 2020. 3

[2] Yingying Deng, Fan Tang, Weiming Dong, Chongyang Ma, Xingjia Pan, Lei Wang, and Changsheng Xu. Stytr2: Image style transfer with transformers. In *CVPR*, 2022. 1, 2

[3] Songhua Liu, Tianwei Lin, Dongliang He, Fu Li, Meiling Wang, Xin Li, Zhengxing Sun, Qian Li, and Errui Ding. Adaattn: Revisit attention mechanism in arbitrary neural style transfer. In *ICCV*, 2021. 1, 2

[4] Fujun Luan, Sylvain Paris, Eli Shechtman, and Kavita Bala. Deep painterly harmonization. In *CGF*, 2018. 1, 2

[5] Dae Young Park and Kwang Hee Lee. Arbitrary style transfer with style-attentional networks. In *CVPR*, 2019. 1, 2

[6] Hwai-Jin Peng, Chia-Ming Wang, and Yu-Chiang Frank Wang. Element-embedded style transfer networks for style harmonization. In *BMVC*, 2019. 1, 2, 3