Domain-Aware Universal Style Transfer

Kibeom Hong [1] Seogkyu Jeon[1] Huan Yang[3] Jianlong Fu[3] Hyeran Byun[1,2*]
[1]Department of Computer Science, Yonsei University
[2]Graduate school of AI, Yonsei University
[3]Microsoft Research
{cha2068,jone9312,hrbyun}@yonsei.ac.kr {huayan,jianf}@microsoft.com

1. Additional qualitative results

In this section, we provide additional qualitative results for better comparison with previous state-of-the-arts. Figure 1 shows the results of single-domain methods for both artistic [3, 5, 7, 8, 9, 10] and photo-realistic [6, 11] style transfer. In addition, Figure 2 shows the comparison with previous multi-domain style transfer methods [1, 4]. Our DSTNs outperform previous works on both domains.

Moreover, we conduct style transfer on high-resolution images (4751 $\times$ 3168) which are shown in Figure 3, and Figure 4. It is easy to recognize that DSTNs effectively stylize global and local patches while preserving the original semantic information. In addition, we provide additional qualitative results of ablation study on the effectiveness of domainness α in Figure 5.

2. Ablation study of the number of domain-aware skip connections

In Figure 6, we conduct the ablation study to analyze the effect of each skip connection. As discussed in [2], the structural information is lost in deeper layers of the network. By fully exploiting three skip-connections, DSTNs produce the satisfactory photo-realistic stylized results. With artistic references, our domain indicator produces the higher domainness value thus the results are almost consistent.

3. Adversarial Training via Multi-scale Discriminator

In this section, we describe the details of the multi-scale discriminator and the adversarial training. For adversarial training, we label the original images from both datasets as *real*, while stylized images are marked as *fake*. We adversarially train our decoder and the discriminator so that DSTNs are capable of conducting the stylization in more realistic way.

To further enhance the performance, we adopt the multi-scale discriminator which exploits not only the global texture but also local patch-wise contexts. The multi-scale discriminator consists of four Conv blocks and one convolutional layer as illustrated in Figure 7. We establish two skip-connections on the first and third Conv blocks in order to collect patch-wise features. Finally, the decoder of DSTN and the discriminator are trained with the adversarial loss as follows:

$$\begin{aligned}\mathcal{L}_{adv}^{Dis} &= \mathbb{E}_{I_s \sim I_s^*}[\log D(I_s)] + \mathbb{E}_{\tilde{I} \sim I_{final}}[\log(1 - D(\tilde{I})]\\ \mathcal{L}_{adv}^{Dec} &= \mathbb{E}_{\tilde{I} \sim I_{final}}[\log D(\tilde{I})],\end{aligned} \tag{1}$$

where I_s and $\tilde{I}$ denote the real images and stylized results, respectively.

4. Photo-realistic style transfer with segmentation maps

Following DPST [6] and WCT^2 [11], DSTNs are also capable of utilizing segmentation maps to maintain semantic correspondence between content and reference images. As shown in Figure 8, DSTN successfully produce photo-realistic results with the help of segmentation maps.

Figure 1. Qualitative comparisons with state-of-the-art models. The blue box indicates photo-realistic reference images and the red box indicates artistic ones. We depict the stylized results from existing artistic style transfer models (b-g) and photo-realistic ones (h-i). Previous methods produce unsatisfactory results when they receive images from the opposite domain. The results of (a) demonstrate that DSTNs produce both photo-realistic and artistic well regardless of the domain of reference images.

Figure 2. Comparison with previous state-of-the-arts for multi-domain style transfer.

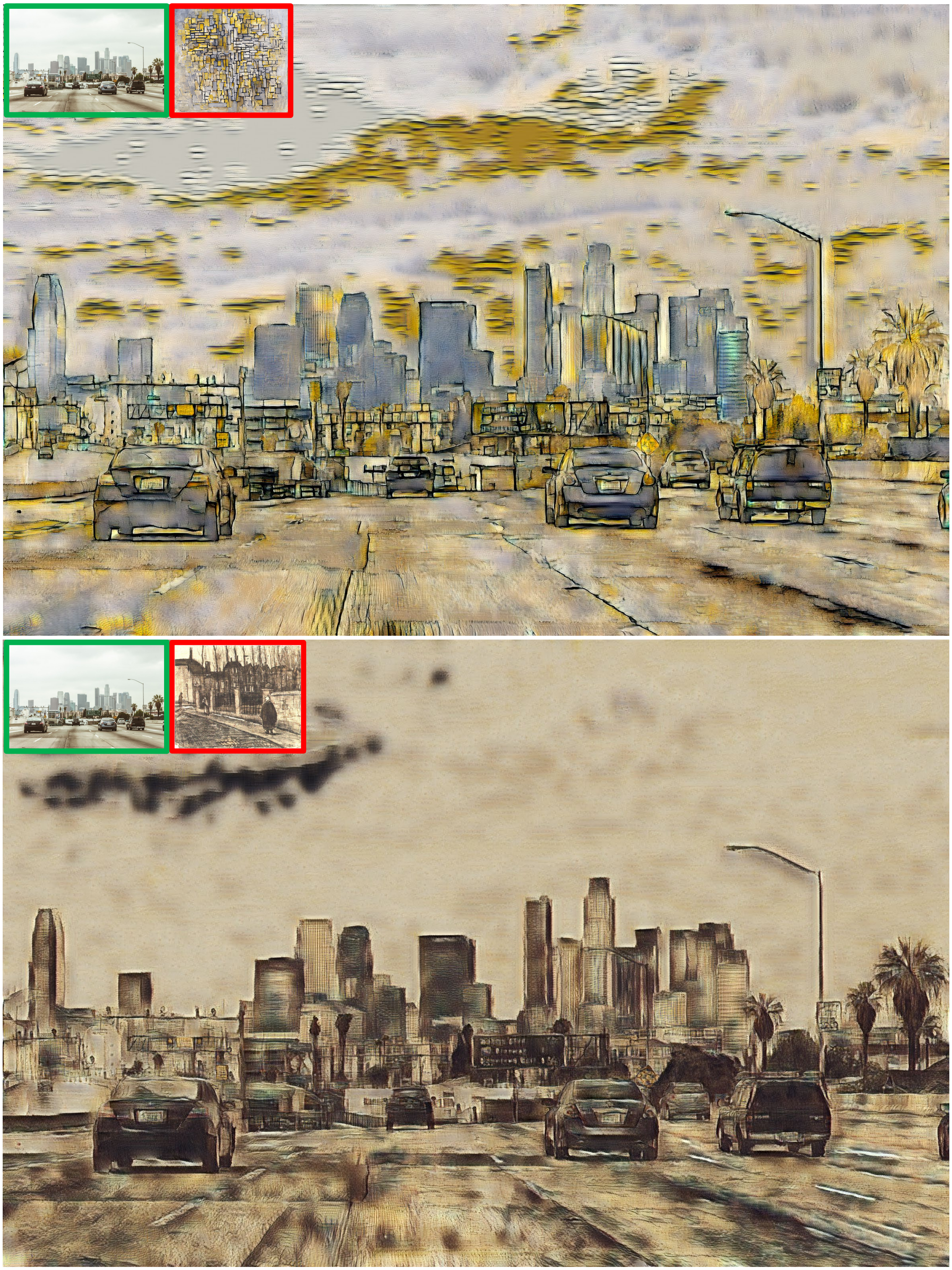


Figure 3. Qualitative result of artistic style transfer with high resolution (4752×3168). The red box indicates artistic reference images, and the green box indicates the content image.

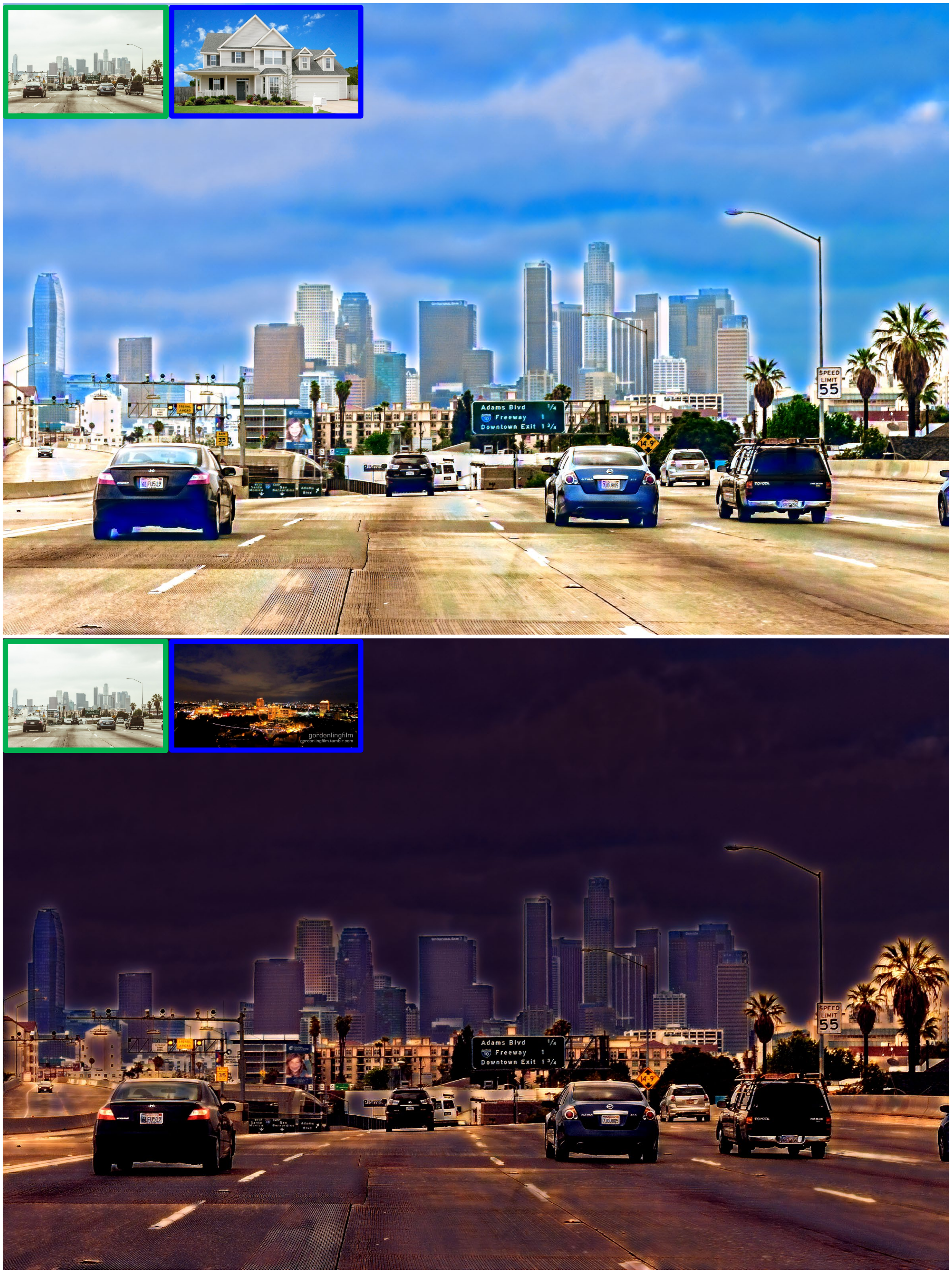
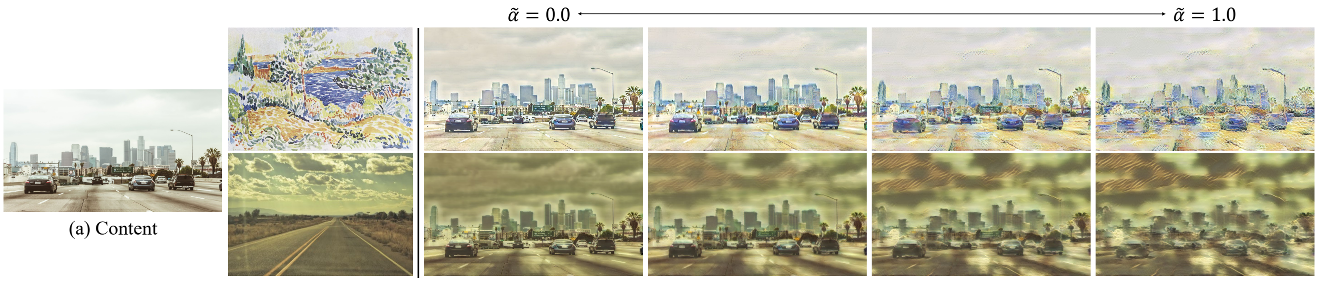
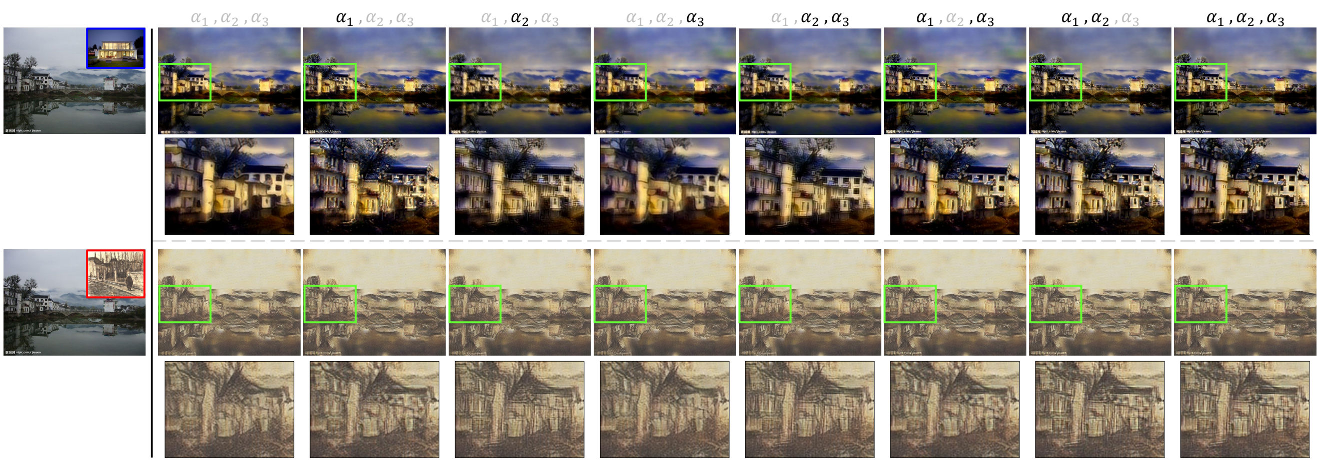


Figure 4. Qualitative result of photo-realistic style transfer with high resolution (4752×3168). The blue box indicates photo-realistic reference images, and the green box indicates the content image.



(a) Content (b) Reference (c) Interpolated results

Figure 5. Ablation study on the effect of domainness α.



Inputs Stylized results

Figure 6. Ablation study on the domain-aware skip connections. α_l denotes the domainness value from each domain-aware skip connection on the level of l. The gray text (α_l) indicates the removal of skip connection of corresponding level. We also display the zoomed patch from the green box.

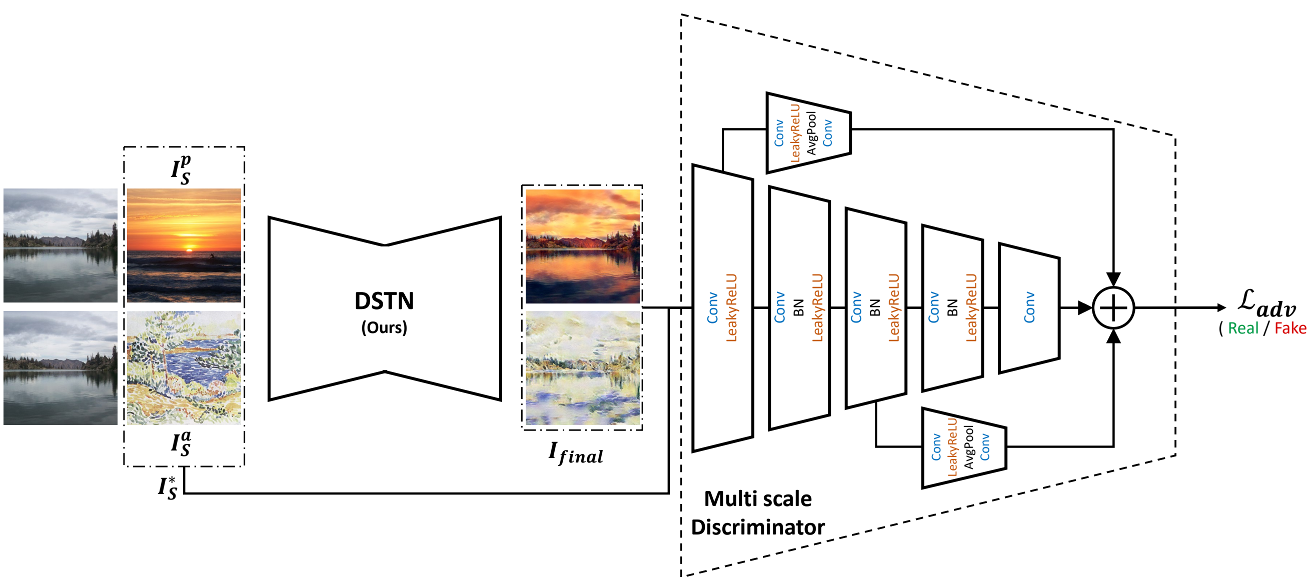


Figure 7. Overview of the multi scale discriminator.

Figure 8. Photo-realistic stylization results with segmentation maps.

References

[1] Tai-Yin Chiu and Danna Gurari. Iterative feature transformation for fast and versatile universal style transfer. In *ECCV*, pages 169–184, 2020.

[2] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *CVPR*, pages 2414–2423, 2016.

[3] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *ICCV*, pages 1501–1510, 2017.

[4] Xueting Li, Sifei Liu, Jan Kautz, and Ming-Hsuan Yang. Learning linear transformations for fast image and video style transfer. In *CVPR*, pages 3809–3817, 2019.

[5] Yijun Li, Chen Fang, Jimei Yang, Zhaowen Wang, Xin Lu, and Ming-Hsuan Yang. Universal style transfer via feature transforms. In *NeurIPS*, pages 386–396, 2017.

[6] Yijun Li, Ming-Yu Liu, Xueting Li, Ming-Hsuan Yang, and Jan Kautz. A closed-form solution to photorealistic image stylization. In *ECCV*, pages 453–468, 2018.

[7] Dae Young Park and Kwang Hee Lee. Arbitrary style transfer with style-attentional networks. In *CVPR*, pages 5880–5888, 2019.

[8] Lu Sheng, Ziyi Lin, Jing Shao, and Xiaogang Wang. Avatar-net: Multi-scale zero-shot style transfer by feature decoration. In *CVPR*, pages 8242–8250, 2018.

[9] Huan Wang, Yijun Li, Yuehai Wang, Haoji Hu, and Ming-Hsuan Yang. Collaborative distillation for ultra-resolution universal style transfer. In *CVPR*, pages 1860–1869, 2020.

[10] Yuan Yao, Jianqiang Ren, Xuansong Xie, Weidong Liu, Yong-Jin Liu, and Jun Wang. Attention-aware multi-stroke style transfer. In *CVPR*, pages 1467–1475, 2019.

[11] Jaejun Yoo, Youngjung Uh, Sanghyuk Chun, Byeongkyu Kang, and Jung-Woo Ha. Photorealistic style transfer via wavelet transforms. In *ICCV*, pages 9036–9045, 2019.