Puff-Net: Efficient Style Transfer with Pure Content and Style Feature Fusion Network

Supplementary Material

1. Feature Extractor

Figure 1 is an introduction to the architecture of two feature extractors.

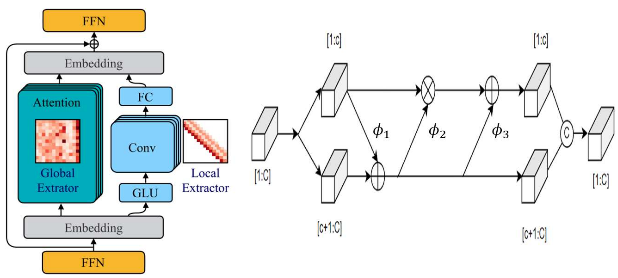

Figure 1. On the left is the workflow of the style extractor, and on the right is the workflow of the content extractor.

Here, [1:c] represents the 1st to the cth channels. $\phi_i(i = 1, 2, 3)$ are the arbitrary mapping functions. To balance the feature extraction ability and computational consumption, we employ bottleneck residual block (BRB) block in MobileNetV2 [2].

2. Different Initialization Results

Figure 2. The first row shows the result of using style image for initialization, the second row shows the result of using zero initialization, and the third row shows the result of using random initialization. In each triplet, it is in the order of the style image, the content image, and the stylization result.

Figure 2 shows the result images generated by initializing ε_o with style images, zero values, and random values. It is easy to see that when using different methods to initialize ε_o, our model will produce different qualitative results. Using style images for initialization will make result images closer to style images, and using zero or random values for initialization will make it difficult for us to obtain visually plausible results. Therefore, we believe ε_o is the basis for style transfer in our proposed model, and the calculation results of the attention mechanism determine the stylization way of each patch.

3. Limitation

Though visually better transfer results have been yeilded, our model still has the drawback of content leak [1] like most existing algorithms. After multiple rounds of style transfer for a set of images, some details of the content image will still be lost. We believe that stylization will disrupt the content features we obtain through the content extractor, such as lines, resulting in fewer and fewer extracted content features. We demonstrate this phenomenon in Figure 3.

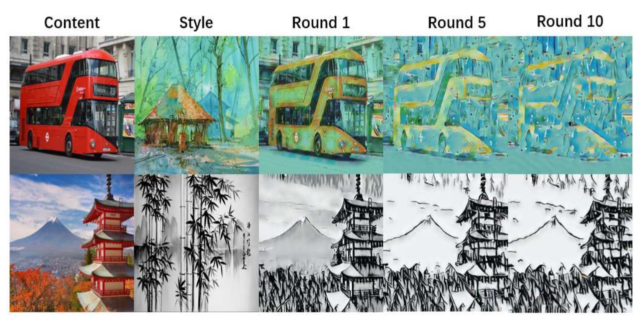

Figure 3. We present the results of stylization for 1, 5, and 10 rounds, respectively.

References

[1] Jie An, Siyu Huang, Yibing Song, Dejing Dou, Wei Liu, and Jiebo Luo. Artflow: Unbiased image style transfer via reversible neural flows. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 862–871, 2021. 1

[2] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018. 1