Arbitrary Style Transfer with Parallel Self-Attention

Tiange Zhang*, Ying Gao*, Feng Gao, Lin Qi, Junyu Dong
School of Information Science and Engineering, Ocean University of China
*Authors contribute equally

Introduction

The main contributions of our work can be summarized in threefold:

- We introduce the self-attention mechanism into the arbitrary style transfer task so that the network can synthesize high-quality stylized images with a comparable speed to AdaIN, giving more appealing and reasonable transfer results.
- We present a new learning approach for the improved feed-forward AdaIN network composed of an attention module and a decoder that is optimized by using a trade-off of the content reconstruction loss, style loss, and a Laplacian loss.
- Extensive experiments have been implemented on MS-COCO and WikiArt datasets. The results demonstrate that the proposed method can effectively preserve the content structure with more details and transform the style patterns with fewer artifacts.

Method

Adaptive Instance Normalization

AdaIN receives a content input x and a style input y, and adaptively aligns the channel-wise mean and variance of x to match those of y to perform style transfer:

$$AdaIN(x) \quad = \quad \sigma(y)(\frac{x-\mu(x)}{\sigma(x)}) + \mu(y),$$

where σ and μ are computed across the spatial locations.

Laplacian Loss

The Laplacian filter is defined as:

$$D = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix},$$

Therefore, the Laplacian matrix of an RGB image x can be obtained by convolving the three channels altogether with D. Hence, the MSE loss between the Laplacian matrixs of the stylized image I_{cs} and content image I_c can be measured in three channels separately:

$$\begin{aligned} \mathcal{L}_{lap} = & \sum_{ij}(D(I^R_{cs}) - D(I^R_c))^2_{ij} \\ & + \sum_{ij}(D(I^G_{cs}) - D(I^G_c))^2_{ij} \\ & + \sum_{ij}(D(I^B_{cs}) - D(I^B_c))^2_{ij}, \end{aligned}$$

where i and j denote the pixel of the image.

Training Overflow

The total loss function for training the attention module and the decoder:

$$\mathcal{L} \quad = \quad \lambda_c\mathcal{L}_c + \lambda_s\mathcal{L}_s + \lambda_{lap}\mathcal{L}_{lap},$$

The content loss is the Euclidean distance between the AdaIN output features F_t and the stylized image features $E(I_{cs})$:

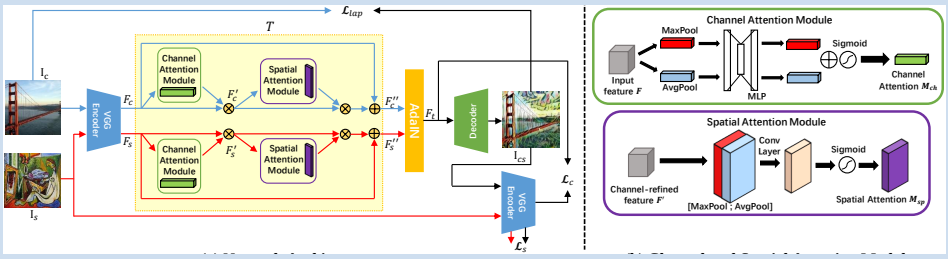
$$\mathcal{L}_c \quad = \quad \|E(I_{cs}) - F_t\|_2.$$

The style loss restricts the features of stylized output image I_{cs} and input style image I_s as:

$$\begin{aligned} \mathcal{L}_s \quad = & \quad \sum_{i=1}^{L}\|\mu(E_i(I_{cs})) - \mu(E_i(I_s))\|_2 \\ + & \quad \sum_{i=1}^{L}\|\sigma(E_i(I_{cs})) - \sigma(E_i(I_s))\|_2, \end{aligned}$$

Overview of Proposed Style Transfer Network

Our proposed style transfer network, which is composed of an encoder–decoder framework, a parallel attention module and an AdaIN layer, can effectively synthesizes high-quality stylized images that appropriately reflect global and local patterns of the image content and style.

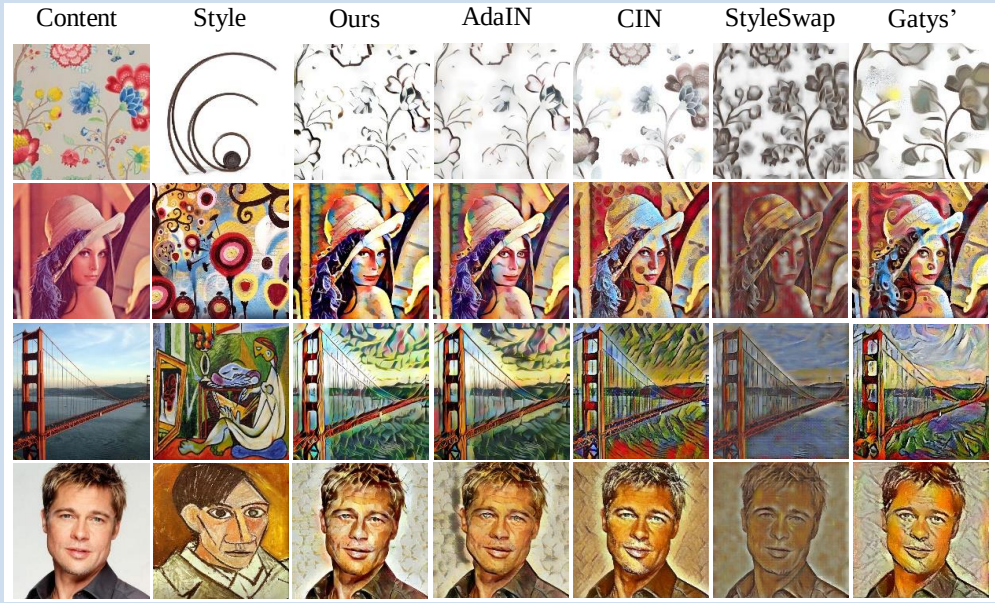


Quantitative Evaluations:

For quantitative evaluations, we conduct an efficiency analysis on **processing speed**, and evaluate the **style loss** defined by the method from Gatys' [1].

Method	Times(s) 256×256	Times(s) 512×512	Styles	$\ln(\mathcal{L}_{style})$	Learning-free
Gatys' [1]	15.024	34.110	∞	11.56	✓
CIN [2]	0.2315	0.9033	Limited	11.81	×
StyleSwap [3]	0.0202	0.1012	∞	13.04	×
AdaIN [4]	0.0181	0.0384	∞	11.89	×
Ours	0.0204	0.0396	∞	11.72	×

Qualitative Comparisons:



Conclusion

We present an improved arbitrary style transfer model based on the self-attention mechanism. By introducing an improved convolutional block attention module into the style transfer network and applying a Laplacian loss, more appealing stylized image can be generated. Experimental results demonstrate that the proposed method can eliminate unexpected artifacts, maintain more content structure details and transfer the most important style patterns properly. Meanwhile, we believe that there is still room for improvement.

References

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

[2] Vincent Dumoulin, Jonathon Shlens, and Manjunath Kudlur. A learned representation for artistic style. In *ICLR*, 2017.

[3] Tian Qi Chen and Mark Schmidt. Fast patch-based style transfer of arbitrary style. *arXiv:1612.04337*, 2016.

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