Supplementary Material: Progressive Attentional Manifold Alignment for Arbitrary Style Transfer

Xuan Luo[1], Zhen Han *[2], and Linkang Yang[1]

[1] School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an, Shaanxi, China
[2] School of Computer Science, Wuhan University, Wuhan, Hubei, China
https://github.com/luoxuan-cs/PAMA

1 Discussion

The proposed progressive attention manifold alignment (PAMA) consists of linear transformations that redistribute the style feature vectors in a common subspace. The redistributed style feature vectors are then linearly interpolated to their most similar content feature vectors. By recurrently interpolating between the content and style feature vectors, the content manifold is aligned to the style manifold along a geodesic between them. However, why this manifold alignment process can solve the style degradation problem?

Firstly, the manifold alignment process can help the attention module parse the similarity information and establish complex relations. Since the style feature vectors are linearly fused into the content feature vectors, the attention module in the next stage can easily parse the similarity information. As the content feature interpolates with more linear components from the style feature, the content manifold is aligned to the style manifold, enabling the attention module to parse complex structural similarities.

Secondly, all of the transformations applied to the style feature are linear transformations in a common space, which can be considered as rearranging the patches of the style image. The linear property helps to preserve the semantic information of the style feature and avoid information loss.

2 The Effectiveness of PAMA

2.1 The Channel Response of the Channel Alignment Module

To verify the effectiveness of the channel alignment module, we calculate the mean values of the content and style features before and after the channel alignment module. We calculate the mean values of features from the third stage of alignment. Since the features have 512 channels, the mean values are 512-dimensional vectors, which are shown in Fig.1. Although there is a considerable

* Corresponding author, hanzhen_2003@hotmail.com.

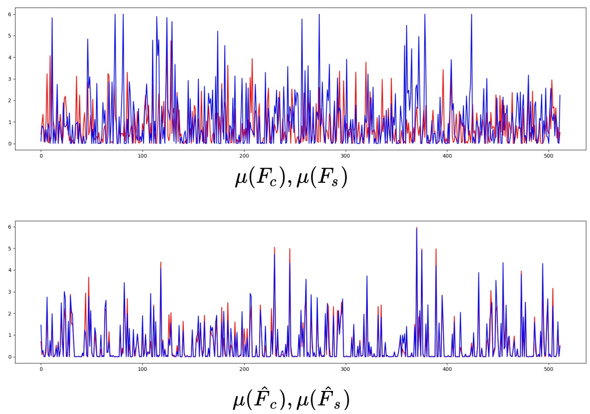

Fig. 1: The channel response of features. The first row shows the mean values of the content and style features before channel alignment, and the second row shows the mean values of the aligned features. The blue and red lines denote the content and style feature, respectively. The content and style images are the same as the example in the left of Fig. 2.

discrepancy between the mean values of the content and style features (the first row), the channel alignment module can re-weight the channels to align their distribution (the second row). This phenomenon proves that the proposed channel alignment module can emphasize the related feature channels to align the two distributions.

2.2 The Attention Map of the Attention Module

To demonstrate that the proposed PAMA can align the content and style manifolds to help the attention module parse similarities, we draw the attention maps of all the manifold alignment stages (Fig.2). The sample on the left side contains paired content and style images with a cat and a tiger. It is evident for humans that the eyes of the cat and the tiger should be matched. However, since the eyes of the cat and tiger are in different colors and shapes, it is challenging for an unsupervised learning algorithm to build this correspondence automatically. The proposed PAMA can gradually align the content manifold to the style manifold to better build complex relations without supervision. Fig.2 left shows that the attention is relatively scattered in the early stages but converges in the late stages. For unpaired content and style images like Fig.2 right, although attention is more dispersed early on, it will eventually converge. The proposed PAMA can align the manifolds and establish stable semantic correspondence without supervision.

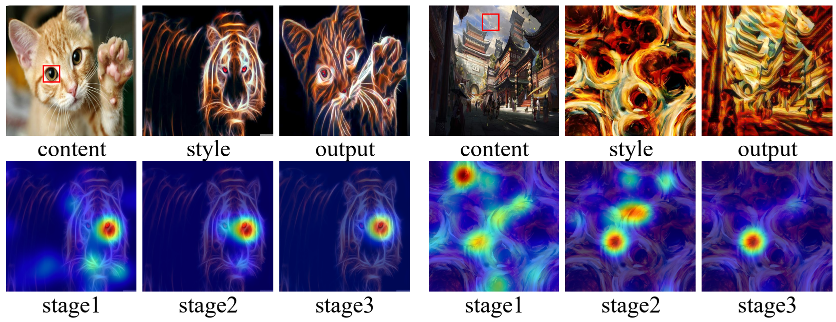

Fig. 2: The attention map. The example on the left side shows the attention map of the eye area of the content image. The example on the right side shows the attention map of the cloud area. The chosen areas are framed by red boxes.

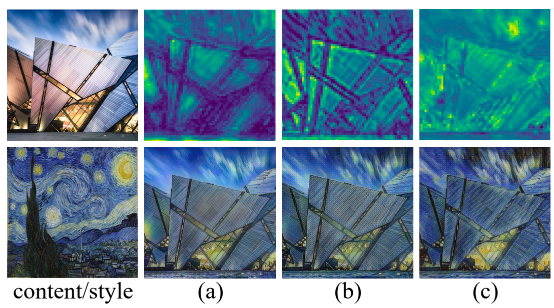

Fig. 3: Space-aware Interpolation results. The first row is the visualization of the adaptive weight W, where the yellow pixels denote higher stylization, and the blue pixels denote higher content preservation. The second row is the interpolation results of the first row respectively. (a) Interpolation of the first stage; (b) Interpolation of the second stage; (c) Interpolation of the third stage.

2.3 The Spatial Interpolation Weights

This part verifies the effectiveness of the space-aware interpolation of all three manifold alignment stages. Fig. 3 demonstrates that the space-aware interpolations are sensitive to edge information and tend to preserve content structures around the edges. This can help the network to remove the local inconsistency (or distortions) in salient areas. Also, the interpolation module has learned to detect the uniform regions and render them with higher strength. In this way, we can align to the style manifold without hurting the content manifold structure significantly. For the reason that we decrease the self-similarity content loss gradually during manifold alignment, the interpolation module fuses more style information in the latter stages (more yellow pixels in Fig. 3), producing results with vivid style patterns.

Fig. 4: Comparison between STROTSS [1] and the proposed PAMA. The first row: content/style images; the second row: STROTSS; the third row: PAMA.

Fig. 5: The updated version of the Fig.1 in our paper.

3 Additional Comparison

To further evaluate the effectiveness of the proposed PAMA, we want to compare it with other style transfer methods [1, 3, 4] adopting the relaxed earth mover distance (REMD). However, the STROTSS is an online optimization based method, which takes around a minute to stylize a single 512px image using a Tesla V100 GPU (PAMA only takes 10ms). The [3, 4] are single style transfer methods that require pre-training for every style. It is an unfair comparison that the proposed PAMA is an arbitrary style transfer method.

Fig.4 shows the results of STROTSS and the proposed PAMA. The single style transfer methods [3, 4] are omitted because they cannot be applied to arbitrary styles. Even if the STROTSS is an online optimization based method, the proposed PAMA can generate results with comparable style quality. But still, the style quality of the STROTSS is better.

Meanwhile, the arbitrary style transfer method StyTr2 [5] also uses the attention mechanism. This is a very recent paper published on CVPR 2022, and it took us some time to redistribute the questionnaires for the user study. We decide to update Fig.1 (the Fig.5 here), Fig.4 (the Fig.6 here), and the user study of our original paper for further comparison. The new user study follows the same method introduced in our paper.

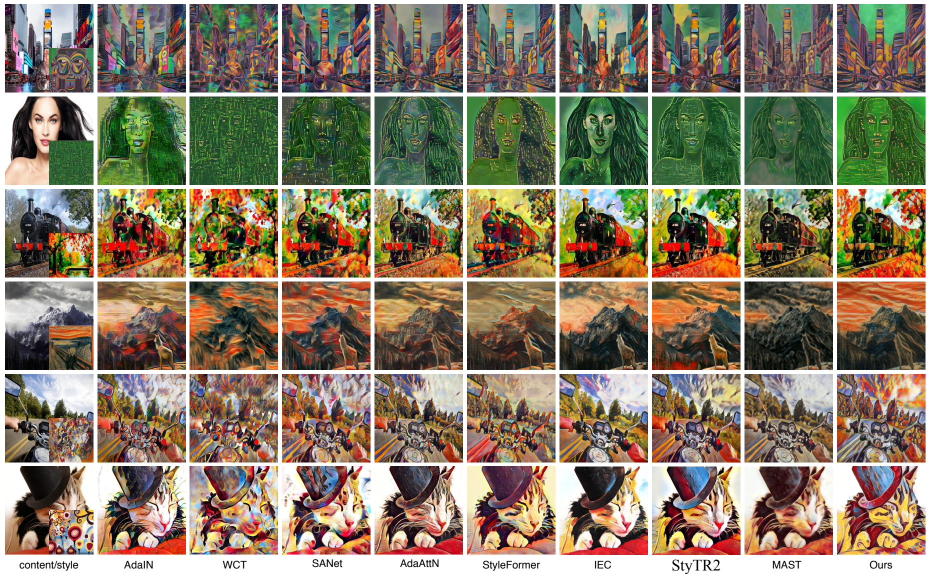

Fig. 6: The updated version of the Fig.4 in our paper.

4 Limitation

In style transfer tasks, the generated images often suffer from checkerboard artifacts (or lattice-like artifacts). The checkerboard artifacts is a common problem of style transfer methods, especially for patch based methods like StyleSwap [13], AdaAttN [9], StyleFormer [10], and AAMS [14]. Since the proposed PAMA is a patch based arbitrary style transfer algorithm, it also has this problem. A common solution is adopting the total variance loss for pixel-level smoothing, which is adopted by the pioneering style transfer methods proposed by Gatys *et al.*[15, 16], Johnson *et al.*, and Ulyanov *et al.*[17, 18]. A simpler solution is using the photo-realistic smoothing technique proposed in PhotoWCT [2], or we can apply gaussian blurring. We applied photo-realistic smoothing to the proposed PAMA, which can smooth the stylization results while preserving their content structure. The smoothed results are demonstrated in Fig.7.

5 More High-resolution Results

Due to the limitation of file size, we used a compressed version of stylization results for the conference paper. Here we provide high-resolution results (Fig.8). More results with 0.5x, 2x, and 4x style losses are also shown in Fig.9, Fig.10, Fig.11, respectively. Please check the following pages.

Table 1: Updated User Study.

method	content quality	style quality	overall quality	total
AdaIN [6]	235	153	204	592
WCT [7]	197	207	189	593
SANet [8]	322	314	336	972
AdaAttN [9]	1164	478	717	2359
StyleFormer [10]	563	325	423	1311
IEC [11]	862	479	744	2085
StyTr$\hat{2}$ [5]	726	615	890	2231
MAST [12]	238	182	219	639
Ours	693	2247	1278	4218

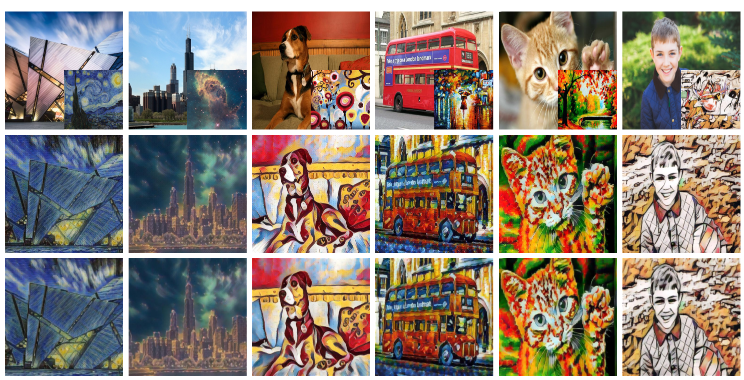

Fig. 7: The photorealistic smoothing [2]. The first row: content/style images; the second row: PAMA; the third row: PAMA with photorealistic smoothing.

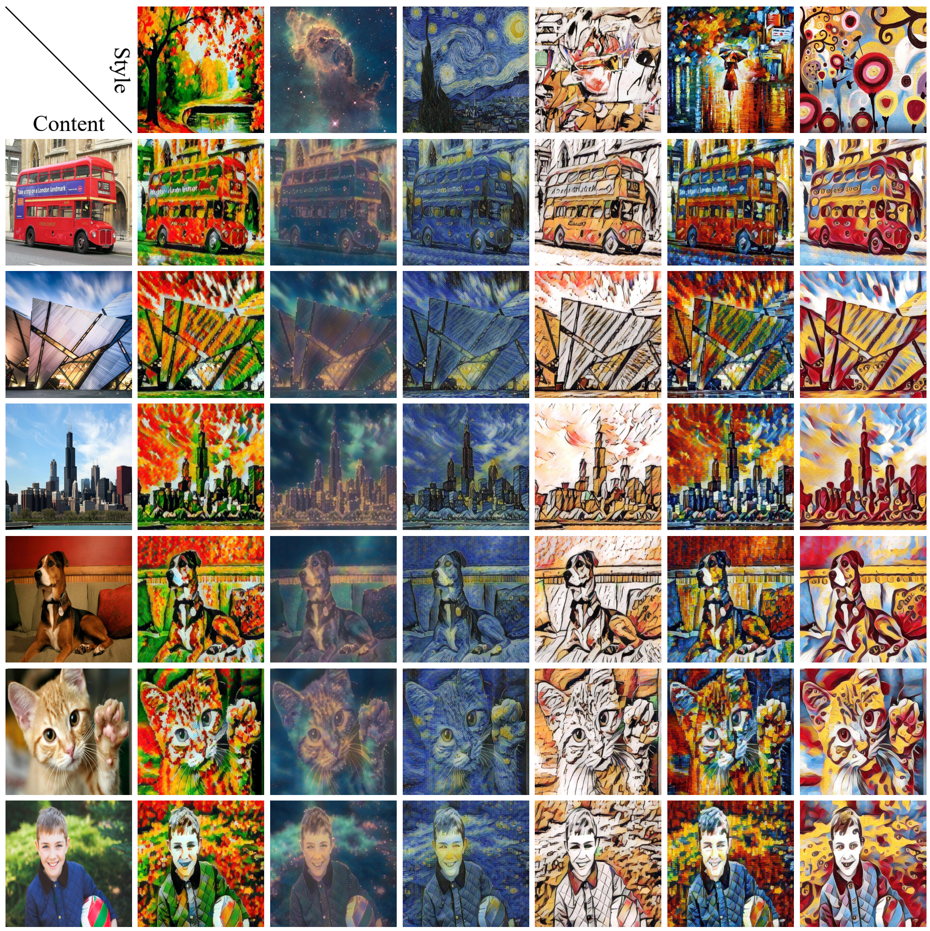

Fig. 8: Stylization results of the original PAMA.

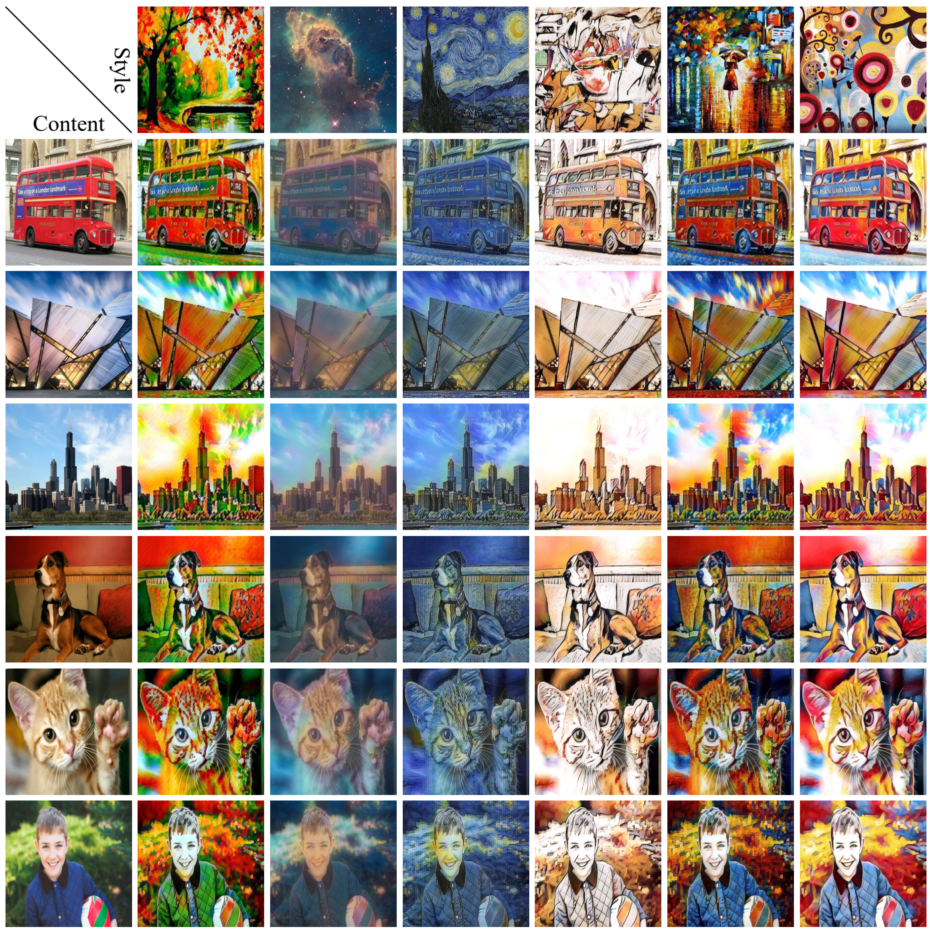

Fig. 9: Stylization results with 0.5x style losses.

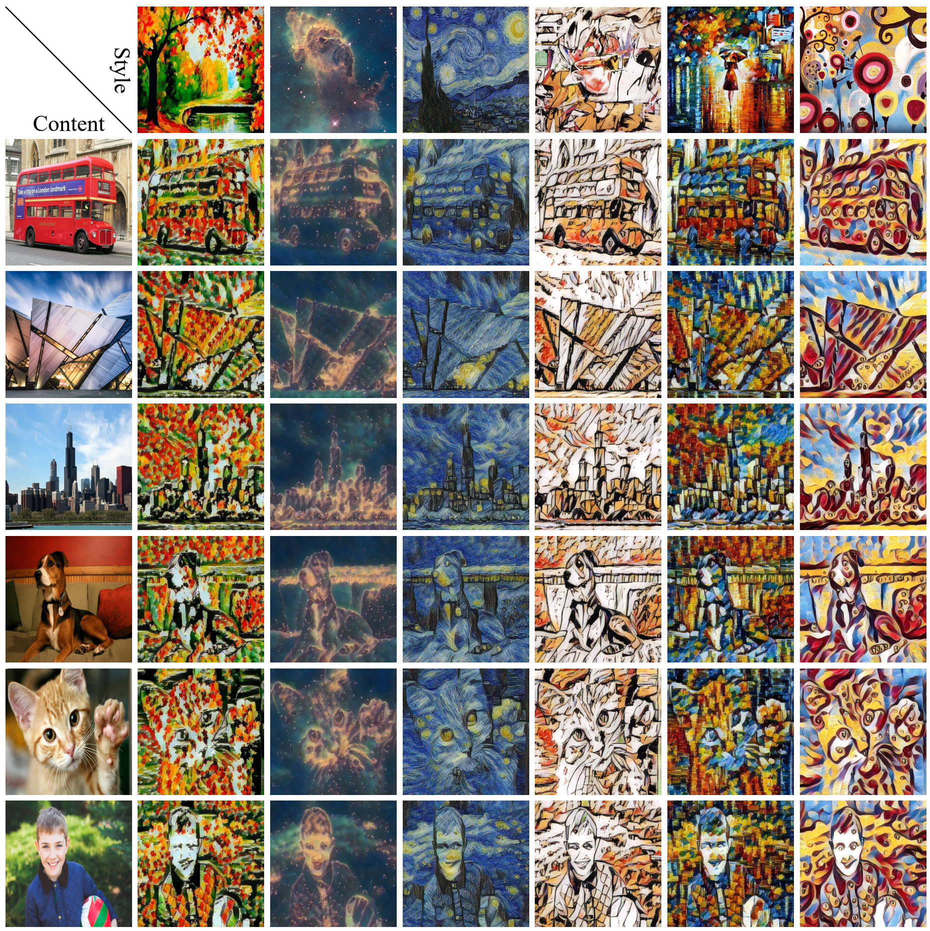

Fig. 10: Stylization results with 2x style losses.

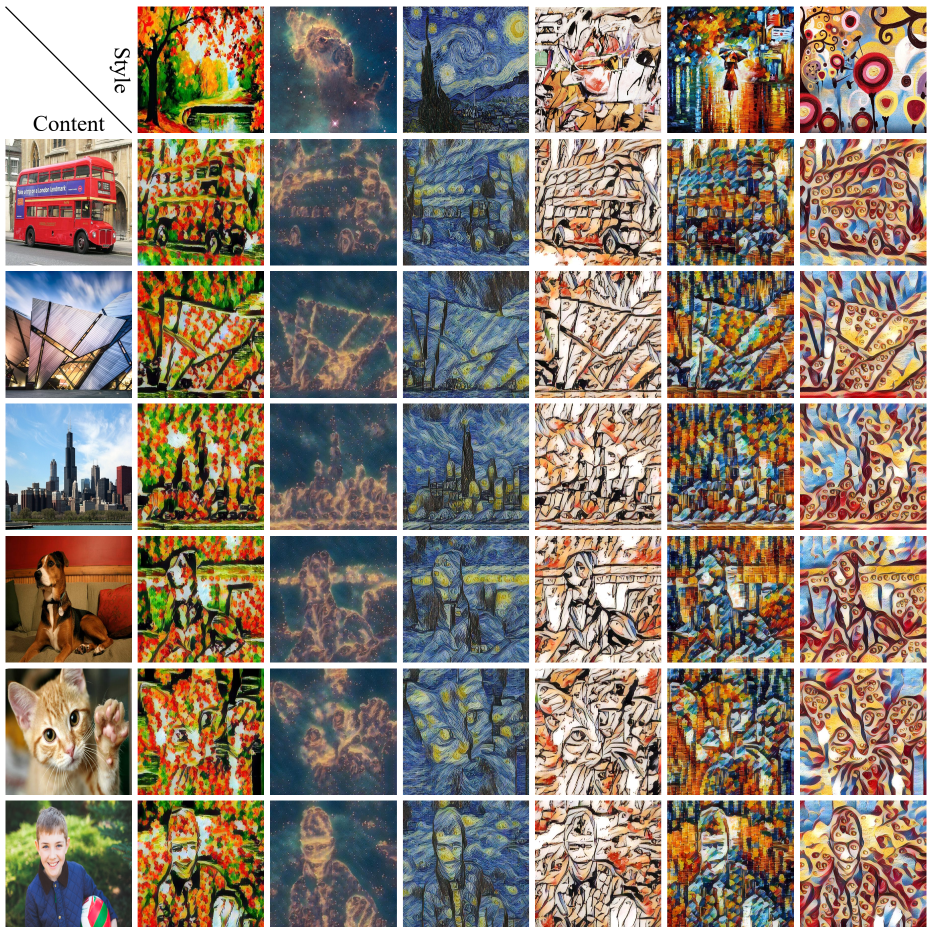

Fig. 11: Stylization results with 4x style losses.

References

1. Kolkin, N.I., Salavon, J., Shakhnarovich, G.: Style transfer by relaxed optimal transport and self-similarity. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019. (2019) 10051–10060
2. Li, Y., Liu, M., Li, X., Yang, M., Kautz, J.: A closed-form solution to photorealistic image stylization. In: Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part III. (2018) 468–483
3. Qiu, T., Ni, B., Liu, Z., Chen, X.: Fast optimal transport artistic style transfer. In: International Conference on Multimedia Modeling. (2021) 37–49
4. Lin, T., Ma, Z., Li, F., He, D., Li, X., Ding, E., Wang, N., Li, J., Gao, X.: Drafting and revision: Laplacian pyramid network for fast high-quality artistic style transfer. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021. (2021) 5141–5150
5. Wang, X., Girshick, R.B., Gupta, A., He, K.: Non-local neural networks. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018. (2018) 7794–7803
6. Huang, X., Belongie, S.J.: Arbitrary style transfer in real-time with adaptive instance normalization. In: IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017. (2017) 1510–1519
7. Li, Y., Fang, C., Yang, J., Wang, Z., Lu, X., Yang, M.: Universal style transfer via feature transforms. In: Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA. (2017) 386–396
8. Park, D.Y., Lee, K.H.: Arbitrary style transfer with style-attentional networks. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019. (2019) 5880–5888
9. Liu, S., Lin, T., He, D., Li, F., Wang, M., Li, X., Sun, Z., Li, Q., Ding, E.: Adaattn: Revisit attention mechanism in arbitrary neural style transfer. In: 2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10-17, 2021. (2021) 6629–6638
10. Wu, X., Hu, Z., Sheng, L., Xu, D.: Styleformer: Real-time arbitrary style transfer via parametric style composition. In: 2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10-17, 2021. (2021) 14598–14607
11. Chen, H., Zhao, L., Wang, Z., Zhang, H., Zuo, Z., Li, A., Xing, W., Lu, D.: Artistic style transfer with internal-external learning and contrastive learning. In: Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual. (2021) 26561–26573
12. Huo, J., Jin, S., Li, W., Wu, J., Lai, Y.K., Shi, Y., Gao, Y.: Manifold alignment for semantically aligned style transfer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. (2021) 14861–14869
13. Chen, T.Q., Schmidt, M.: Fast patch-based style transfer of arbitrary style. CoRR **abs/1612.04337** (2016)
14. Yao, Y., Ren, J., Xie, X., Liu, W., Liu, Y.J., Wang, J.: Attention-aware multi-stroke style transfer. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019) 1467–1475

15. Gatys, L.A., Ecker, A.S., Bethge, M.: Texture synthesis using convolutional neural networks. In: Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada. (2015) 262–270
16. Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016. (2016) 2414–2423
17. Ulyanov, D., Lebedev, V., Vedaldi, A., Lempitsky, V.S.: Texture networks: Feed-forward synthesis of textures and stylized images. In: Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016. (2016) 1349–1357
18. Ulyanov, D., Vedaldi, A., Lempitsky, V.S.: Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017. (2017) 4105–4113