Signal Processing: Image Communication 78 (2019) 465–470

Contents lists available at ScienceDirect

Signal Processing: Image Communication

journal homepage: www.elsevier.com/locate/image

Advanced deep learning techniques for image style transfer: A survey ☆

Long Liu *, Zhixuan Xi, RuiRui Ji, Weigang Ma

The Faculty of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048, China

ARTICLE INFO

Keywords:
Image style transfer
Convolutional neural networks
Slow transfer and fast transfer

ABSTRACT

Image style transfer is an emerging technique based on deep learning, which takes advantage of the impressive feature extraction of convolutional neural networks (CNN). The extraction of high-level features of images makes the separation of style information and image content possible. Image style conversion technique aims to learn the style characteristics of various paintings, and then apply the learned style to another image. The combination of artificial intelligence and art makes this technique highly concerned in the relevant technical fields and art fields, and has been applied in many different fields of society. In this paper, we conduct a comprehensive study on image style transfer techniques. Firstly, we analyze and classify the existing algorithms of the current style transfer algorithms, and then elaborate on their applications in different fields. In addition, we also summarize the future development and prospect of the image transfer technique.

1. Introduction

Image style transfer is an interesting research in computer vision. The technique aims to transfer the style of one image to another. In recent years, algorithms based on deep neural networks have shown impressive performance in various intelligent applications, such as face recognition, pedestrian re-identification, vehicle tracking [1–9]. Deep neural network-based image processing algorithms can extract high-level features, which outperform than conventional low-level visual features-based algorithms [10,11]. For example, the shallow layers of VGGNet [12] (such as conv1_1 and conv1_2) extract simple features, such as edge, luminance. While the deep layers (such as conv5_1, conv5_2) extract complicated features. VGGNet aims to extract deep features of the input image. However, different from VGGNet, image style transfer algorithms aim to generate the image based on the input features. Specifically, image style transfer technique is to specify an input image as the base image, also known as the content image. At the same time, another image or more images are specified as the desired image style. As we can see in Fig. 1, there are a lot of different styles that can be regarded as style images. The image style transfer algorithm transforms the image style while ensuring the structure of the content image, so that the final output composite image presents a perfect combination of the input image content and desired style.

Analyzing the image of a certain style and establishing a mathematical or statistical model for this style is significant for image style transfer [13]. Then change the image to be transferred to better conform to the established model as seen in Fig. 2. The results show that the effect is good, but there is still a big disadvantage: a program can only basically achieve a certain style or a certain scene. Therefore, the practical application of traditional style transfer research is very limited. As early as the beginning of 2000, many scholars began to study the problem of image style transfer, but they focused on texture synthesis and transfer at that time [14]. The mathematical methods used were mainly statistical methods of various image transformations, such as wavelet transform [15] with limited effects. Only in recent years, inspired by the excellent performance of deep neural network in large-scale image classification, and with its powerful multi-level image feature extraction and representation ability, has this problem been better solved and its performance improved significantly.

2. Image style transfer without neural networks

In this section, we introduce its derivation to better understand the development of image style transfer. The first important problem of art style automatic conversion is how to model and extract style from images. Before neural networks are applied to style transfer, the following types of style transfer are commonly used.

2.1. Texture synthesis image analogy for style transfer

As far as I know, most papers on image texture before 2015 are manually modeled, and the most important idea used is that texture can be described by the statistical model of image local features. Semmo

☆ No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.image.2019.08.006.

* Corresponding author.
E-mail address: liulong.edu@gmail.com (L. Liu).

https://doi.org/10.1016/j.image.2019.08.006
Received 17 February 2019; Received in revised form 24 July 2019; Accepted 11 August 2019
Available online 13 August 2019

Fig. 1. Some samples of different art style images.

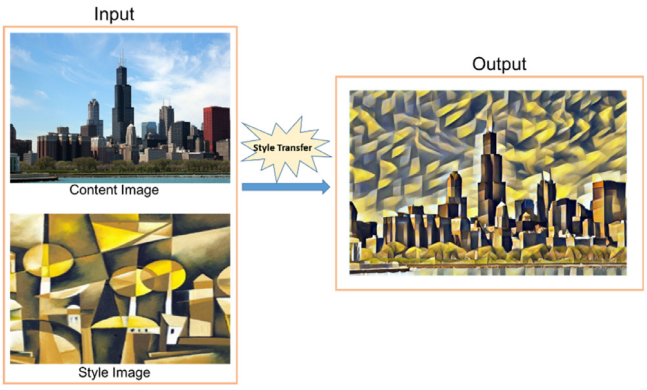

Fig. 2. An example of image style transfer.

et al. [16] proposed a quantitative method of feature perception recolor using the main color of the input image. They used 7 different steps to describe and transfer the features of the oil painting, and finally converted the image into the appearance of the oil painting. Shih et al. [17] proposed a head style transfer algorithm based on local multi-scale style transformation. Due to weak spatial constraints, direct application of general style transfer may deform the head of a character, so it is not suitable for head style transfer. So they typically edit the captured spatial variation images by portrait editing and multi-scale processing of facial texture at different scales. However, the quality of experimental results is easily limited by mask quality and may lead to input noise amplification.

2.2. Image filtering for style transfer

Considering that image style transfer is actually a process of image simplification and abstraction, the image filtering method adopts some combination image filters (such as bilateral and Gaussian filters) to render a given image. Portilla et al. [18] concluded and generated some textures based on the complex mathematical models and formulas established by the basis functions of adjacent spatial positions, directions and scales. Therefore, they proposed a statistical model of texture image based on wavelet transform for synthesizing random images subject to these constraints. However, this method has a large amount of computation and low efficiency.

2.3. Image analogy for style transfer

Image analogy aims at learning the mapping between a pair of source images and a target image and locating stylized images by means of supervised learning [19]. Image analogy training sets include pairs of uncorrected source images and corresponding programmed images with specific styles. The simulation method is effective, but the difficulty lies in the fact that it is difficult to obtain paired training data.

3. Image style transfer with neural networks

After 2014, the development of neural networks have been widely concerned by people. People found that deep learning can be used to train object recognition models. Previous object recognition models used to compare geometric shapes with different parts of an object, some were modeled by color, some by 3d, and some by local features. Completely different from this, the neural network will automatically extract the most useful features after training the image, so the features are no longer generated simply by cutting the original object into small pieces, but by the neural network to select the optimal way to extract. In this section, we classify the current style Transfer methods: Slow Transfer based on image optimization and Fast Transfer based on model optimization. The first type realizes style transfer and image reconstruction by gradually optimizing the image. The second type optimizes the generation of offline models and the generation of stylized images using a single forward transfer, which actually takes advantage of the idea of fast image reconstruction technology.

3.1. Slow transfer based on online image optimization

The basic idea of online image optimization is to extract the content and style features from the content and style images respectively, and recombine the two features into the target image. Then, the online iterative reconstruction of the target image is based on the difference between the generated image and the content and style images. The neuron algorithm of artistic style is defined as follows: (1) if the high-level features extracted from a pre-trained classification network are relatively close (Euclidean distance), the contents of the two graphs are similar; (2) if the low-level features extracted from a pre-trained classification network have common statistical characteristics (Gram matrix), the style of the two graphs are similar.

On this basis, Gatys et al. [15,20] constructed texture modeling with deep learning as seen in Fig. 2, the use of VGG-net for hierarchical image feature extracting for characterization of image effectively thereby, using CNN will image semantic content and fusion of different style,

the style of this new type of migration method has brought good in academia, and led to the subsequent lot about using deep learning style transformation of research results. For a content image I_C and style image I_S, the goal of online optimization is to minimize the following loss function (see Fig. 3):

$$I^* = \arg\min_I L_{total}(I_C, I_S, I)$$
$$= \arg\min_I \alpha L_C(I_C, I) + \beta L_S(I_S, I) \quad (1)$$

where Lc represents the content loss between the generated image and the content image, Ls represents the style loss between the generated image and the style image. Each with a hyper-parameter adjusts the balance between content and style. And the content loss Lc is shown in formula 2:

$$L_C(I_C, I) = \frac{1}{2}\sum \left(F^l\left(I_C\right) - F^l(I)\right)^2 \quad (2)$$

where $F^l\left(I_C\right)$ and $F^l(I)$ respectively represent their features in the layer l, and then define the squared error loss between the two feature representations. Similarly, A^l and G^l respectively represent the styles of the original image and the generated image in the layer l, so the contribution of layer l to total loss is

$$E_l = \frac{1}{4N_l^2 M_l^2}\sum \left(G^l - A^l\right)^2 \quad (3)$$

where And the total style loss is:

$$L_S(I_S, I) = \sum_{l=0}^{L} w_l E_l \quad (4)$$

where w_l is the weight that each layer contributes to the total style loss. These methods use convolutional neural network to perform texture synthesis and style conversion, and can get better quality results, but there are some limitations in texture quality, stability and necessary parameter adjustment. On this basis, multi-scale image synthesis [21,22] based on convolutional neural networks are proposed and solve these problems. Histogram losses are used to synthesize textures that are statistically more compatible with the sample to improve these instabilities. Li et al. [23] proposed a two-dimensional image synthesis method combining Markov random field (MRF) model with well-trained deep convolutional neural network (dCNNs). MRF regularizer reduced the incredible feature mixing commonly seen in previous dCNNs inverse approaches. It can be concluded from the above that for a neural network, shallow network extracts low-dimensional features such as color, while deep network extracts high-dimensional semantic content information. Therefore, the loss of style is often compared with the characteristics of shallow network, while the loss of content is compared with the characteristics of deep network.

3.2. Fast transfer based on online model optimization

Fast neural network method based on online model optimization is to use the fast image reconstruction based on off-line model optimization to rebuild the stylized results and solve the problem of speed and computational cost. That is, for one or more of the style image, it optimizes a feedforward network on a large set of images, which generates the resulting image directly through the network. According to how many a network can generate styles, fast neural style transfer methods can be divided into PerStyle-Per-Model, MultipleStyle-Per-Model and Arbitrary-Style-Per-Model

PerStyle-Per-Model The PerStyle-Per-Model approaches can produce stylized images twice as fast as the previous slow NST approach as seen in Fig. 4, but it is not flexible enough to train a separate generation network for each specific style image. In fact, many art paintings have similar brush strokes and differ only in their palette, so it is unnecessary to train a separate network for each style [24–27]. Ulyanov et al. [24] trains a compact feedforward convolutional networks to generate multiple samples of the same texture of any size and to transfer the artistic style from a given image to any other image. The resulting network is so light that it can produce textures of a quality comparable to Gatys and others, but hundreds of times faster. Analogously, in [25], Johnson et al. extracted advanced features from the pre-trained network, defined and optimized the perceptive loss function, and used the perceptive loss function to train the image transformation feedforward network. This method can achieve better results in the experiment of single image super-resolution. However, there are still some limitations on the efficiency of the deep neural network method. The method proposed by Li et al. [26] solved this problem by calculating a feedforward, striped convolution network in advance. It captures the feature statistics of Markov patches and can directly generate the output of any dimension, rather than the previous numerical deconvolution. Since no optimization is needed in the generation process, the speed is greatly improved.

Multiple-Style-Per-Models A variety of style training networks should be a lot of different kinds of calculations that can be shared, like some impressionist paintings that have similar brushstrokes but different colors [28–31]. It would be wasteful to train a series of N-style individual models in the same way as before as seen in Fig. 5. Therefore, Dumoulin et al. [28] proposes to train a conditional style transfer network to support multiple styles. The main difference here is that this network is a conditional network, and its input is that in addition to the previous content image, it needs to have an applied style id, but it also supports n-style style. The idea is to model a style by scaling and shifting specific styles after normalization. All convolution weights in a style transformation network can be shared in multiple styles. Furthermore, in [31], Li et al. constructed a forward network based on self-coding. Firstly, input images were converted into feature spaces by encoding sub-networks. StyleBank is used to classify input styles, and each filter bank represents one style. StyleBank can generate different stylized results for the corresponding content by convolving with the content features generated by self-coding. This method can not only train a number of Shared self-coding styles, but also learn a new style in increments without changing the self-coding. Zhang et al. [30] proposed a new forward network with residual network. The main points match feature statistics on multiple scales (4 scales) to match different styles of images. In this network, they come up with an inspiration layer for matching the Gram matrix of stylistic images and keeping the content.

Arbitrary-Style-Per-Model Arbitrary-Style-Per-Model is designed to extract a generic style conversion network, inputs are arbitrary content images and arbitrary style images in the pre-training network [32,33], and the network will generate target style based on the activation values of these two features. In order to implement image transfer of any style, Chen et al. [33] added a style swap layer in the core part of the network, which is to replace each one of the most accessible style features of the content of content on this layer, However, the implementation effect of this paper is not very good. The result of processing is very similar to the fusion effect of two pictures. The style information is not retained much and the style features are not obvious. Based on this, Huang et al. [32] have added adaptive instance normalization(AdaIN) layer into the model, which mainly aims at making the generated image features as similar as possible in mean and variance with the features drawn. This approach supports the use of a forward network for any style switching while also ensuring that the flexibility of style and the calculation efficiency are achieved, achieving a real-time effect (see Fig. 6).

4. Application and future challenges

With the emergence of neural style transfer, this technology has been applied in many fields: in social communication, style transfer applications are popular on social networking sites, where users can share

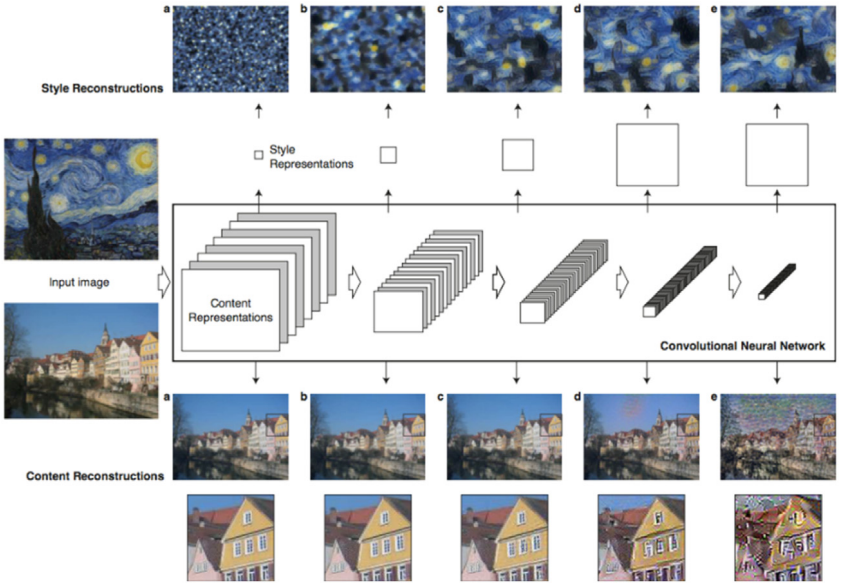

Fig. 3. A system framework of image style transfer using CNN proposed by Gatys et al. The input image is filtered by CNN network, and the network's response at a specific layer reconstructs the input image, so as to visualize the information in different processing stages of CNN.

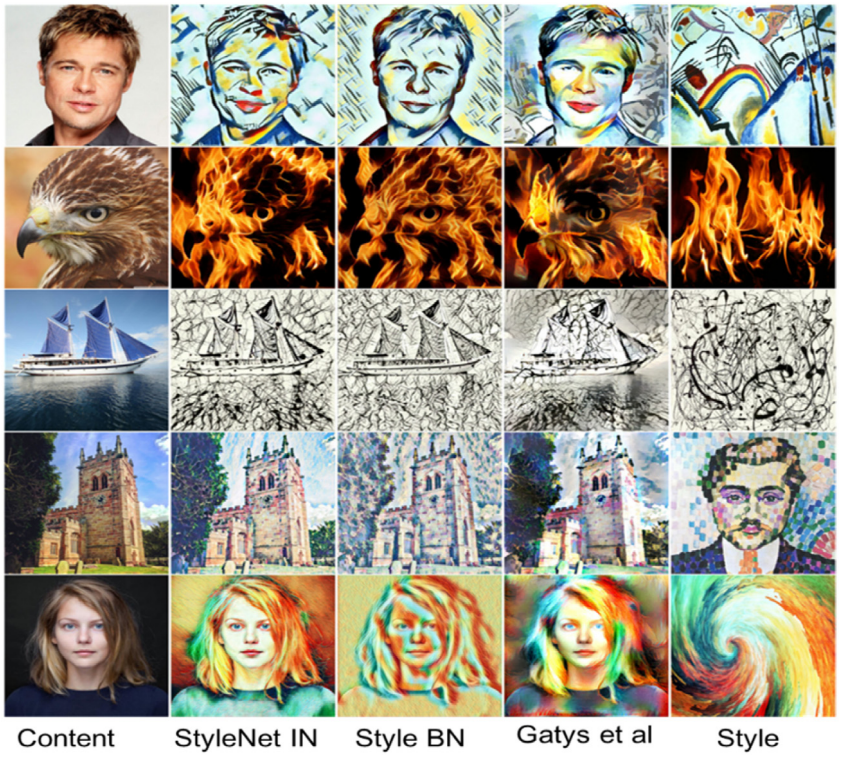

Fig. 4. Stylization consequences obtained by applying different textures to different content images.

their fantastic images, such as the popular Prisma APP. Another use of the NST is as a user-assisted creation tool, especially to help artists more easily create works of art of a particular style. Also in the creation of computer vision, fashion design and other aspects to bring people great convenience. In addition, neural transfer technology is also applied in the field of entertainment, such as film, animation and game creation, which can reduce the creation cost and save the production time. However, this technology still exists some research difficulties. Among many methods, it is difficult to balance speed, flexibility and conversion quality. Due to the black box characteristics of neural network, the process is not controllable and it is difficult to achieve more detailed gaps. In addition, the anti-interference performance is weak, if we add some interference to the image, the result of the network may become unacceptable. Adding an image quality evaluation module might be a good choice, i.e., the quality index is utilized as one component of the loss function. However, how to obtain the reference image is still a big challenge.

5. Conclusion

With the rapid development of science and technology, the style transformation of images processed by neural network has achieved

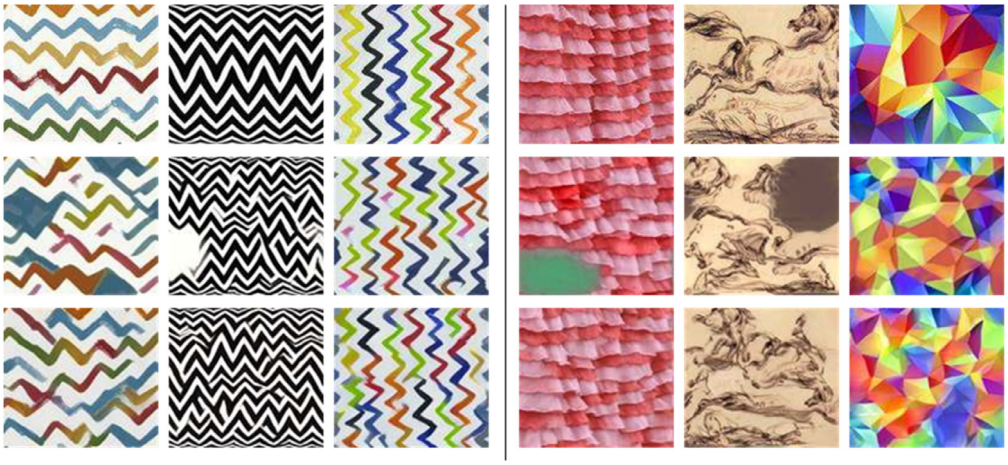

Fig. 5. Comparisons between random training and incremental training.

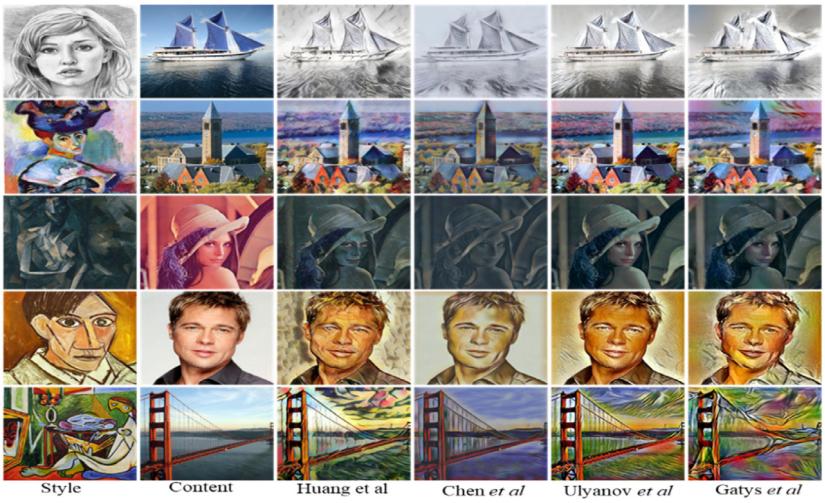

Fig. 6. Example style transfer results. All the tested content and style images are never observed by our network during training.

great achievements. Different style transfer methods have sprang up and developed into a vibrant field, which promoted the development of other different fields. In general, this paper has carried out a wide range of research on the existing neural style transfer, elaborated the origin and development process of image style technology, classified the current methods, and summarized the existing expansion and possible challenges. In the future, image style transfer will have greater development, which may combine deep learning with other methods and extend to different applications to radiate its powerful effect.

Acknowledgments

This work was supported by the Natural Science Foundation of China (NSFC), No. 61673318, No. 61702410, No. U1734210

References

[1] X. Yao, J. Han, D. Zhang, F. Nie, Revisiting co-saliency detection: A novel approach based on two-stage multi-view spectral rotation co-clustering, IEEE Trans. Image Process. 26 (7) (2017) 3196–3209.

[2] G. Niu, Q. Chen, Learning an video frame-based face detection system for security fields, J. Vis. Commun. Image Represent. 55 (2018) 457–463.

[3] L. Zhang, Y. Gao, R. Ji, Y. Xia, Actively learning human gaze shifting paths for semantics-aware photo cropping, IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc. 23 (5) (2014) 2235.

[4] Mingliang Xu, Mingyuan Li, Weiwei Xu, Zhigang Deng, Yin Yang, Kun Zhou, Interactive mechanism modeling from multi-view images, ACM Trans. Graph. 35 (6) (2016) 236.

[5] Luming Zhang, Richang Hong, Yue Gao, Rongrong Ji, Qionghai Dai, Xuelong Li, Image categorization by learning a propagated graphlet path, IEEE T-NNLS 27 (3) (2016) 674–685.

[6] J. Han, D. Zhang, G. Cheng, N. Liu, D. Xu, Advanced deep-learning techniques for salient and category-specific object detection: A survey, IEEE Signal Process. Mag. 35 (1) (2018) 84–100.

[7] Q. Chen, L. Sang, Face-mask recognition for fraud prevention using Gaussian mixture model, J. Vis. Commun. Image (2018).

[8] J. Han, H. Chen, N. Liu, C. Yan, X. Li, CNNs-based RGB-D saliency detection via cross-view transfer and multiview fusion, IEEE Trans. Cybern. 48 (11) (2018) 3171–3183.

[9] Mingliang Xu, Chunxu Li, Pei Lv, Lin Nie, Rui Hou, Bing Zhou, An efficient method of crowd aggregation computation in public areas, IEEE Trans. Circuits Syst. Video Technol. 28 (10) (2018) 2814–2825.

[10] J. Han, G. Cheng, Z. Li, D. Zhang, A unified metric learning-based framework for co-saliency detection, IEEE Trans. Circuits Syst. Video Technol. 28 (10) (2018) 2473–2483.

[11] J. Han, R. Quan, D. Zhang, F. Nie, Robust object co-segmentation using background prior, IEEE Trans. Image Process. 27 (4) (2018) 1639–1651.

[12] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, Comput. Sci. (2014).

[13] L. Zhang, M. Song, N. Li, J. Bu, C. Chen, Feature selection for fast speech emotion recognition, in: International Conference on Multimedia 2009, DBLP, 2009, pp. 753–756.

[14] L. Zhang, Y. Han, Y. Yang, M. Song, S. Yan, Q. Tian, Discovering discriminative graphlets for aerial image categories recognition, IEEE Trans. Image Process. 22 (12) (2013) 5071–5084.

[15] L.A. Gatys, A.S. Ecker, M. Bethge, A neural algorithm of artistic style. arXiv preprint arXiv:1508.06576, 2015.

[16] A. Semmo, D. Limberger, J.E. Kyprianidis, Jürgen Döllner, Image stylization by oil paint filtering using color palettes, in: Workshop on Computational Aesthetics, Eurographics Association, 2015.
[17] Y.C. Shih, S. Paris, C. Barnes, W.T. Freeman, Frédo Durand, Style transfer for headshot portraits, ACM Trans. Graph. 33 (4) (2014) 1–14.
[18] J. Portilla, E.P. Simoncelli, A parametric texture model based on joint statistics of complex wavelet coefficients, Int. J. Comput. Vis. 40 (1) (2000) 49–70.
[19] J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, 2016.
[20] L.A. Gatys, A.S. Ecker, M. Bethge, Image style transfer using convolutional neural networks, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, 2016.
[21] Y. Li, N. Wang, J. Liu, X. Hou, Demystifying neural style transfer. arXiv preprint arXiv:1701.01036, 2017.
[22] E. Risser, P. Wilmot, C. Barnes, Stable and controllable neural texture synthesis and style transfer using histogram losses. arXiv preprint arXiv:1701.08893, 2017.
[23] C. Li, M. Wand, Combining markov random field s and convolutional neural networks for image synthesis, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2479–2486.
[24] D. Ulyanov, V. Lebedev, A. Vedaldi, V.S. Lempitsky, Texture networks: Feed-forward synthesis of textures and stylized images, in: ICML, 2017, pp. 1349–1357.
[25] J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, in: European Conference on Computer Vision, Springer, Cham, 2016, pp. 694–711.
[26] C. Li, M. Wand, Precomputed real-time texture synthesis with markovian generative adversarial networks, in: European Conference on Computer Vision, Springer, Cham, 2016, pp. 702–716.
[27] D. Ulyanov, A. Vedaldi, V.S. Lempitsky, Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis, in: CVPR, Vol. 1, 2017, p. 3, No. 2.
[28] V. Dumoulin, J. Shlens, M. Kudlur, A learned representation for artistic style, in: Proc. of ICLR, 2017.
[29] Y. Li, C. Fang, J. Yang, Z. Wang, X. Lu, M.H. Yang, Diversified texture synthesis with feed-forward networks, in: Proc. CVPR, 2017.
[30] H. Zhang, K. Dana, Multi-style generative network for real-time transfer. arXiv preprint arXiv:1703.06953, 2017.
[31] D. Chen, L. Yuan, J. Liao, N. Yu, G. Hua, Stylebank: An explicit representation for neural image style transfer, in: Proc. CVPR, Vol. 1, 2017, p. 4, No. 3.
[32] X. Huang, S.J. Belongie, Arbitrary style transfer in real-time with adaptive instance normalization, in: ICCV, 2017, pp. 1510–1519.
[33] T.Q. Chen, M. Schmidt, Fast patch-based style transfer of arbitrary style. arXiv preprint arXiv:1612.04337, 2016.