Sensors and Materials, Vol. 33, No. 9 (2021) 3325–3332
MYU Tokyo

S & M 2692

Optimized and Improved Methods of Image Style Transfer for Local Reinforcement

Yong Li,[1] Yan Wang,[2,3] Hsien-Wei Tseng,[1*] Hongkun Huang,[1] and Chun-Chi Chen[4**]

[1]College of Mathematics and Information Engineering, Longyan University, Fujian 364012, China
[2]Institute of Scientific and Technological Information of Fujian, Fujian 350003, China
[3]Fujian Provincial Key Laboratory of Information Network, Fujian 350003, China
[4]School of Life Sciences, Longyan University, Longyan, Fujian 364012, China

(Received March 26, 2021; accepted September 2, 2021)

Keywords: image style transfer, deep learning, image segmentation, DeepLab2

Image style transfer, which commonly refers to adding a designated image style to a target content image, is now widely used in the movie industry, animation design, and game rendering, providing strong visual effects and cultural influences. However, there is no common criterion for evaluating the performance of image style transfer. In addition, people are more interested in local regions of images. This paper provides some revised methods to meet customer demand, focusing on an optimized image segmentation method based on DeepLab2, a semantic segmentation method, and fully connected conditional random fields (FCCRFs) for local image style transfer, with experiments demonstrating their usefulness and efficiency.

1. Introduction

Image style transfer, which commonly refers to adding a designated image style to a target content image, is now widely used in the movie industry, animation design, game rendering, and advertisement design, providing strong visual effects and various cultural influences that cannot ordinarily be achieved by other methods.(1,2) Some other fields such as medical image processing and industrial design require non-photorealistic images that can reduce the workload in producing model simulations and virtual reality, making image style transfer a valuable technique in various industries.(3) An example of image style transfer is shown in Fig. 1, where the left image is the target content image, the middle image is the image whose style we wish to emulate, and the right image is the finally generated image after style transfer. In this paper, deep learning (DL) technology with the DeepLab2 image semantic segmentation method has been applied in local image style rendering, with the improvements in the visual effect demonstrated by performing a comparison. In addition, the YIQ color space model has been adopted to preserve color details of original images.

With the rapid development of AI technologies, especially neural networks such as those used in DL, many fields are benefiting from applications of DL, including image style transfer.(4) Although image style transfer based on DL has attracted much attention, many problems remain,

*Corresponding author: e-mail: hsienwei.tseng@gmail.com
**Corresponding author: e-mail: kennath1980@gmail.com
https://doi.org/10.18494/SAM.2021.3402

Fig. 1. (Color online) Example of image style transfer.

such as how to define or judge the effect of different image style transfer methods. This has led to subjective judgment rather than the objective judgment preferred by industry. Moreover, sometimes users want to control the effect of image style transfer by themselves or decide the exact artistic style, so revised methods are needed.

Before DL, image analysis often required programmers to have expert knowledge such as familiarity with feature extraction, edge detection, and critical point detection, making algorithms such as the nonlinear filter, local greedy, region segmentation, morphological, and global optimization algorithms popular. For stroke rendering, Hertzmann provided the snake relaxation algorithm based on stroke-based rendering (SBR), which uses the Sobel operator to confirm image edges and a decision function with a local slide window method.[5] DeCarlo and Santella first adopted regional segmentation in image style transfer, using a down-sampling method to express pixels in a layered manner and the Canny operator to transfer delicate image edges, filling colors, and textures into corresponding areas.[6] Gatys *et al.*, who first used DL in image style transfer, reconstructed the abstract feature representation of the mid-layer in VGG networks with a Gram matrix to capture arbitrary image style transfer information, which required minimization of the maximum mean difference.[7] Johnson *et al.* improved Gatys' work by using feedforward networks and perceptual loss, markedly enhancing the speed of image style transfer and making real-time processing possible.[8] Chen also increased the transfer speed by using image patches to realize the transfer of arbitrary styles.[9] Li and Wand proposed another method of image style transfer that used a Markov random field (MRF) with a deep convolutional neural network. Their core idea was to replace the Gram matrix with the MRF; this method is an example of non-parametric image style transfer.[10,11] Traditional image segmentation methods for the foreground and background include the Otsu, watershed, and GrabCut methods,[12,13] which are unsupervised learning methods and extract low-level features in images. Long *et al.* proposed the fully convolutional networks (FCNs) algorithm, which is a DL method.[14]

2. Materials and Methods

The popularity of commercial products based on image style transfer is very sensitive to the user experience, and their popularity is increased by allowing users to decide the exact style transfer effect they prefer, an example of which is given in Fig. 2.

Fig. 2. (Color online) Images with different ratios of content and style.

Fig. 3. (Color online) Generated image with colors of content image lost.

Fig. 4. (Color online) Unsatisfactory image fusion.

A styled image generated by a traditional image style algorithm is commonly one with the colors replaced by that of the style image, thus losing the original colors of the content image. However, sometimes this is not what users want, especially when the content image has bright colors while the style image is dark, which may lead to an unsatisfactory image, as shown in Fig. 3. That means we want a revised image style transfer method that can inherit the essence of the style image while maintaining the original colors of the content image.

Compared with style transfer on traditional scenic images, image style transfer involving human faces often requires more delicate operations such as target area segmentation, which means we need to combine a good image segmentation method and a local image style transfer method to avoid the generation of images such as that in Fig. 4.

3. Results and Discussion

3.1 Principle of image style transfer

The convolutional neural network of Gatys *et al.* is based on VGG-19, which is an upgrade of the earlier neural networks LeNet and AlexNet. Three images are simultaneously input into the convolutional neural network: the content image, the style image, and a white noise image, and the goal is to establish the content loss function between the content image and the white noise image, and the style loss function between the style image and the white noise image. Both functions should be combined to obtain the sum loss function, with iteration performed to minimize it, from which we can obtain the generated style transfer image.

The content loss function can be expressed as Eq. (1), in which *content_image* refers to the content image, *white_noise_image* refers to the subsequently generated white noise image, *generated_image* refers to the generated style transfer image, and l denotes the designated convolution layer.

$$Loss_{content}\left(\overrightarrow{content\_image},\overrightarrow{white\_noise\_image},l\right)=\frac{1}{2}\sum_{i,j}\left(generated\_image_{i,j}^{l}-content\_image_{i,j}^{l}\right)^{2} \quad (1)$$

The style loss function, which is the product of each convolutional iteration, is calculated through the Gram matrix among different convolutional layers to estimate correlations among features. As the indicator of feature correlation, the Gram matrix is calculated using Eq. (2), in which *white_noise_Gram* is the Gram matrix of the white noise image.

$$white\_noise\_Gram_{i,j}^{l}=\sum\nolimits_{k}\left(generated\_image_{i,k}^{l}-generated\_image_{j,k}^{l}\right) \quad (2)$$

The final style loss function is used to minimize the difference between the Gram matrix of the white noise image and that of the style image, measured by their distance functions. After sufficient iterations to obtain the minimum difference, the style of the white noise image should be close to that of the style image, as expressed below, in which *style_image_Gram* refers to the Gram matrix of the style image.

$$Loss_{style}\left(\vec{a},\overrightarrow{white\_noise\_image}\right)=\frac{1}{4N_{l}^{2}M_{l}^{2}}\sum_{i,j}\left(white\_noise\_Gram_{i,j}^{l}-style\_image\_Gram_{i,j}^{l}\right)^{2} \quad (3)$$

After obtaining the required sum loss function expressed as Eq. (4), back-propagation is used in iteration to minimize the loss function.

$$\begin{aligned}Loss\left(\overrightarrow{content\_image},\vec{a},\overrightarrow{white\_noise\_image}\right)=&\alpha Loss_{content}\left(\overrightarrow{content\_image},\overrightarrow{white\_noise\_image}\right)\\&+\beta Loss_{style}\left(\vec{a},\overrightarrow{white\_noise\_image}\right)\end{aligned} \quad (4)$$

3.2 Adjusted parameters

As mentioned in Sect. 3.1, in Eq. (4), coefficients α and β are determined for each weight value of the generated style transfer image, with much more content image detail remaining for a larger α. Obviously, the effect of the style is a subjective experience, so we provide both parameters for users themselves to decide. According to prior experiments, a ratio of α to β that can be varied within two orders of magnitude will be satisfactory. Here, we recommend the value of α to be between 1 and 5 and the value of β to be between 100 and 500.

3.3 Preservation of original color

The generated style transfer image has the same style and colors as the style image, while losing the colors of the content image. To enrich our art reservoirs, we try to map information in the images in RGB color space into YIQ color space, which has the advantage of extracting luminance information; human eyes are much more sensitive to luminance than to color detail, and both have a linear mapping relationship that is easy to calculate. We repeat the same image style transfer method, transferring luminance information in YIQ format while keeping the original information of the other two color channels, and the result can be seen in Fig. 5.

3.4 Local image style transfer

There has been some research on image style transfer for purposes such as taking a selfie, as mentioned before. For example, the ratio of style parameters has been increased to avoid the transfer style side effect of small texture, and the whole content image has been filtered before performing image style transfer to delete noisy high-frequency dots and obtain a better performance. However, these approaches all have disadvantages such as "face buffing" effects and the generation of image edge blur. In contrast, here, we divide the foreground and background to perform face segmentation using the DeepLab2 segmentation method, which involves ResNet-101, dilated convolution, and fully connected conditional random fields (FCCRFs).[14,15] The segmentation procedures are described as follows:

(1) Input the content image into the pretrained deep neural network, which is ResNet-101 here.
(2) Perform dilated convolution in the neural network to obtain an initial rough grading map, reducing the effect of signal down-sampling.

(a) (b) (c) (d)

Fig. 5. (Color online) Comparison between images with color of content image preserved or not preserved: (a) content image, (b) style image, (c) generated image in RGB, and (d) generated image in YIQ.

(3) Use bilinear interpolation to increase the resolution of the image after applying segmentation procedure (2) above to the original content image.
(4) Enrich pixels by using FCCRFs to confirm their categories to obtain the final image edges.

The resolution of the original content image will be significantly reduced by a normal FCN, whereas most of the resolution can be maintained by performing dilated convolution without adding parameters or increasing the calculation workload. During the process, different scales of segmented images are input into the neural network. Here, atrous spatial pyramid pooling (ASPP) is used for image segmentation under multiple scales, which adopts various dilated convolutions with different sampling sizes to obtain more delicate features.[16]

Traditionally, conditional random fields (CRFs) have been used for pixel classifications by coupling neighboring pixels to label them with the same label, which means they can be recognized as being in the same category. In contrast, FCCRFs can overcome the disadvantages that limit the use of CRFs in FCNs. Figure 6 shows an example of semantic segmentation by DeepLab2.

3.5 Comparison

To verify the effectiveness of our proposed method, we carry out an experimental comparison with two other mainstream image style transfer methods (those of Gatys[7] and Johnson[8]), which is performed using the Ubuntu 16.04 operating system, a 1080-8G graphics card, and the Tensorflow 2.0 AI framework.[17,18] Some of the results are shown in Fig. 7, where the photos taken are of the authors' friends, with different artistic paintings used as style images. As we can see, the edges of the image obtained with the Gatys method are blurred owing to the lack of an efficient segmentation method. For the Johnson method, where the Gram matrix is replaced with the mean value and variance as the style loss parameters, the visual effect is better than that of the Gatys method, but the result is much worse than expected in terms of texture and color. Moreover, the Johnson method requires pretrained measures, which means that models must be generated in advance for some images. In contrast, the images generated by our method can grasp the essence of the style image while maintaining the contents of the original image, and our method is suitable for all types of images owing to its end-to-end semantic segmentation.

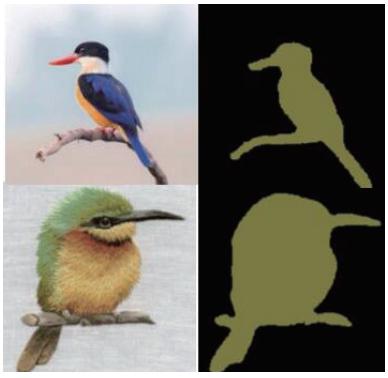

Fig. 6. (Color online) End-to-end semantic segmentation.

(a) (b) (c) (d) (e)

Fig. 7. (Color online) Comparison of results of each image style transfer method: (a) CONTENT, (b) STYLE, (c) GATYS, (d) JOHNSON, and (e) US.

4. Conclusions

As non-photorealistic images are increasingly required with the rapid development of the animation and movie industries, image style transfer is also becoming increasingly important. However, there is no common approach to judging its performance. We proposed an optimized image segmentation method that is based on DeepLab2 semantic segmentation and FCCRFs for local image style transfer, and we experimentally demonstrated its usefulness and efficiency. However, the transfer speed and accuracy should be further improved. In future works, we will attempt to optimize our segmentation method and increase the speed of style-generating models for mobile applications.

Acknowledgments

This work was supported by the Longyan University's Project (LQ2017005), Longyan University's Qi Mai Science and Technology Innovation Fund Project of Liancheng [2018]132 and Shanghang 2019SHQM05 County, Longyan and Longyan University's Research and Development Team Fund (2018)8, and the Great Project of Production, Teaching, and Research of Fujian Provincial Science and Technology Department (2019H6023).

References

1 Y. C. Shih, S. Pari, C. Barnes, W. T. Freeman, and F. Durand: ACM Trans. Graphics **33** (2014) 1. https://doi.org/10.1145/2601097.2601137
2 F. Luan, S. Paris, E. Shechtman, and K. Bala: 2017 IEEE Conf. Computer Vision and Pattern Recognition (CVPR, USA, 2017) 6997–7005. https://doi.org/10.1109/CVPR.2017.740
3 W. H. Qian, D. Xu, Z. Guan, K. Yue, and Y. Y. Pu: Int. J. Pattern Recognit. Artif. Intell. **31** (2017) 620. https://doi.org/10.1142/S0218001417590261
4 S. Strassmann: ACM SIGGRAPH Comput. Graphics **20** (1986) 225. https://dl.acm.org/doi/10.1145/15886.15911
5 A. Hertzmann: Proc. Comput. Graphics Int. 2001 (IEEE, 2001) 47–54. https://doi.org/10.1109/CGI.2001.934657
6 D. DeCarlo and A. Santella: ACM Trans. Graphics **21** (2002) 769. https://doi.org/10.1145/566654.566650
7 L. A. Gatys, A. S. Ecker, and M. Bethge: Comput. Sci. **111** (2015) 98. https://doi.org/10.1167/16.12.326
8 J. Johnson, A. Alahi, and F. F. Li: Perceptual Losses for Real-Time Style Transfer and Super-resolution: Proc. European Conf. Comput. Vision, Amsterdam (Springer, Heidelberg, 2016) pp. 694–711. https://doi.org/10.1007/978-3-319-46475-6_43
9 T. Q. Chen and M. Schmidt: Comput. Vision Pattern Recognit. (2016). https://arxiv.org/pdf/1612.04337
10 C. Li, M. Wand: Combining Markov Random Fields and Convolutional Neural Networks for Image Synthesis (IEEE, 2016) 2479–2486. https://doi.org/10.1109/CVPR.2016.272
11 E. Risser, P. Wilmot, and C. Barnes: Stable and Controllable Neural Texture Synthesis and Style Transfer Using Histogram Losses (2017). https://arxiv.org/abs/1701.08893
12 N. Otsu: IEEE Trans. Syst. Man Cybern. **9** (1979) 62. https://doi.org/10.1109/TSMC.1979.4310076
13 C. Rother, V. Kolmogorov, and A. Lake: ACM Trans. Graphics **23** (2004) 309. https://doi.org/10.1145/1186562.1015720
14 J. Long, E. Shelhamer, and T. Darrel: 2015 IEEE Conf. Comput. Vision and Pattern Recognit. (IEEE, 2015) 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965
15 Liang-Chieh Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille: DeepLab: IEEE Trans. Pattern Anal. Mach. Intell. **40** (2018) 834. https://doi.org/10.1109/TPAMI.2017.2699184
16 V. Badrinarayanan, A. Kendall, and R. Cipolla: IEEE Trans. Pattern Anal. Mach. Intell. **39** (2017) 2481. https://doi.org/10.1109/10.1109/TPAMI.2016.2644615
17 T. Q. Chen and M. Schmidt: Comput. Vision Pattern Recognit. (2016). https://arxiv.org/abs/1612.04337
18 V. Dumoulin, J. Shlens, and M. Kudlur: Proc. Int. Conf. Learn. Representations (Toulon, France, 2017) 24–26.