Style Transfer Using Whitening and Coloring to Realize Feature Transformation

Li Sijia\textsuperscript{1}\textsuperscript{*} and Chen Wenbai\textsuperscript{2}

\textsuperscript{1}School of information and communication engineering, Beijing University of Post and Telecommunication, Haidian, Beijing, 100876, China
\textsuperscript{2}School of Automation, Beijing Information Science And Technology University, Chaoyang, Beijing, 100101, China
\textsuperscript{*}Corresponding author’s e-mail: lisijia@bupt.edu.cn

Abstract. This paper proposes a general image style transfer method based on VGG network. This method does not need to learn and pre-train any kind of style in advance. The algorithm mainly implements the style transfer task by reconstructing the automatic decoder. In the process of feature extraction, the algorithm matches the content features and style features through whitening and coloring. We conducted a series of experiments with different photographic images. The experimental results show that our method has a simple training process, low computational complexity and high computational speed and excellent generalization ability.

1. Introduction and Background

Image style transfer technology has broad application prospects in the academia and business circles, and can be applied in entertainment media, animation production, movie special effects and so on. Its applications are mainly 3D modeling, texture simulation, medical image remodeling, etc\textsuperscript{[1,2]}. In the traditional method, scientists need to analyze the artistic style of the image manually at first, establishing a statistical or mathematical model based on the analysis results, and then match the established model of style image and content image by changing the target image. The nature of traditional image style transfer methods is the mathematical modeling. However, similar to traditional computer vision and speech recognition, mathematical modeling has shortcomings. For example, manual model extraction requires extensive experience and consumes a lot of time. The cost of it is extremely high and the application scenarios of it is limited. So a new method for style transfer named deep neural networks was proposed\textsuperscript{[3-5]}. Deep neural networks are widely used in style transfer because of their good performance in extracting image features. Therefore, the VGG model is proved to have a good performance in style transfer. Gatys once proposed a scheme using VGG network model to extract image content features and style features separately to realize image style transfer \textsuperscript{[6]}. After the feature matrix is extracted, the Gram matrix will be used to calculate the local features. Then, starting from that of random noise image, the local features of the images are converted into a statistical model. This groundbreaking research method attracted a lot of attention from the academic circles and caused a large number of follow-up studies. Despite the rapid development of neural style transfer theory and research methods, these studies are faced with many problems: the trade-offs of versatility, image quality and computational efficiency. Some optimized methods require high
calculation cost to output high-quality images, while the method using feed-forward neural network can reduce the calculation cost but can not guarantee the quality of the image.

2. Basic principles of deep learning and style transfer
Convolutional Neural Networks (CNN) is a typical deep learning algorithm that can be used to assign weights and deviations to each pixel of the image. Compared with other classification algorithms, the required pre-processing steps in the CNN model are much less. The filters of the CNN model were designed artificially. With sufficient training, the CNN model can learn the characteristics of these filters.

The architecture of CNN is similar to the network of neurons in human brain. It is a neural network composed of neurons. Each neuron contains a mathematical function, which determines the relationship between input value and output value of it. Due to the reduction in the number of parameters and the reusability of the weights, the CNN model can be trained to understand the complexity of the image better.

As shown in Figure 1, CNN can reconstruct the information extracted from each layer. In the first layer, the convolution result is basically the same as the original image. In the output result of third and forth layer, the style of the picture starts to close up that of the style image, and the output result of the fifth layer has learned the style. As the number of layers increases, the information learned by the model also becomes more detailed.

![Figure 1. The Schematic of CNN Model.](image)

In the style transfer scheme proposed by Gatys, we need to input three images, which are the content image, the style image and the image with white noise. The main task of style transfer is to establish the content loss function and and style loss function, and obtain the overall loss function by combining the two loss functions in certain ratio. Then the result image is output. Gatys and other scholars have proposed a method of neural style transfer, which played a pioneering role in the area of style transfer. Their research shows that the features extracted by the trained deep neural network have correlations, which means that the Gram matrix or covariance matrix has a remarkable ability to capture style characteristics.

Inspired by Gatys, many scholars began to study image style transfer. The mainstream methods are image-based iterative methods and model-based iterative methods. Among these, image-based iterative methods are Markov-random-field-based methods, depth-image-analogy-based methods and
maximum-mean-difference-based representative methods. The methods based on model iteration are the method based on image reconstruction decoder and the method based on generative model.

Although these theories and research methods are developing rapidly, current studies are facing some problems. Firstly, traditional style transfer methods cannot perform style transfer to the images without certain style. Secondly, in the process of style transfer, most algorithms are faced with image quality degradation. To solve these problems, this article provides a simple and efficient method to realize style transfer with no need to train and pre-train any style before, and this approach can greatly improve the efficiency of image style transfer.

3. Style transfer based on feature transformation

3.1. Overview
The schematic of the method proposed in this paper is shown in Figure 2. Whitening and coloring are the two most critical steps. We can use these two feature transformation steps to transform the vectorized feature map of the content image to match that of the style image, therefore matching the information and data of style image and content image.

Figure 2 shows the process of image reconstruction, in which the model transforms the content features in the middle layer. We used the VGG-19 network as the feature extractor and the encoder. Also, we needed to restore the images by decoding. This process is called image reconstruction process(Figure 2-a). As shown in Figure 2-b, with VGG and DecoderX, through the given content image C and style image S, our method can transfer the style based on the method of whitening and coloring.
As shown in Figure 2-c, in order to match style information of all layers, we extended the single-layer model to a five-layer composite model. We used the result obtained by the single-layer model as a new content image and decoded it through multiple iterations. At each middle layer, our main goal was to make the extracted content features have the same statistical features as the style features of the same layer.

The output of multi-layer models has been experimentally proven to have higher visual quality than single-layer models. Compared with traditional methods, multi-layer models can produce better visual effects with less computational cost. In addition to optimization at the algorithm level, this model also provides a parameter, which is used by users to adjust the degree of image stylization in the code as needed.

Different from the traditional style transfer algorithm, in our method, VGG only needs to learn the image reconstruction decoder, and does not need to learn the characteristics of the style image. This is the most important and innovative feature that distinguishes this algorithm from other feed-forward networks. With a style image that the network has never seen before, this algorithm can also accurately carry out style transfer only by extracting the feature covariance matrix. Therefore, this algorithm can achieve style transfer without learning and training, while other existing feed-forward networks require a lot of time in pre-training the style images and continuous fine-tuning of new styles.

3.2. Algorithm
We regard style transfer as a process of image reconstruction through feature transformation (whitening and coloring). Feature transformation can match the information and data contained in the style image and the content image.

3.2.1. Reconstruction Decoder
In order to evaluate different features extracted in each layer, we selected the feature map Relu_X(X = 1, 2, 3, 4, 5) in the five layers of VGG-19 and trained the five decoders accordingly.

\[ L = \| I_0 - I_1 \|^2_2 + \lambda \| \Phi(I_0) - \Phi(I_1) \|^2_2 \] (1)

In Equation 1, \( I_0, I_1 \) are the matrixes of the input image and the output image. \( \Phi \) is the encoder who extracts the features of the first layer and \( \lambda \) is the weight used to balance the style loss function and the content loss function.

3.2.2. Reconstruction Decoder
Given a content image \( I_c \) and a style image \( I_s \), we first extracted their vectorized feature maps \( f_c \in R^{C \times H_c \times W_c} \) and \( f_s \in R^{C \times H_s \times W_s} \) at a certain layer, where \( H_c, W_c \) are the height and width of the content image and \( H_s, W_s \) are the height and width of the style image. \( C \) is the number of channels.

If \( f_c \) is input directly, the decoder will reconstruct the original image \( I_c \). Then, we adjust the \( f_c \) with the change of \( f_s \) by whitening and coloring, which directly transforms \( f_c \) to match the covariance matrix of \( f_s \).

1) Whitening. Before this step, we first subtract the mean value vector, and then linearly transform \( f_c \) as shown in Equation 2, thereby obtaining \( \hat{f}_c \), which can ensure that the feature map has correlation.

\[ \hat{f}_c = E_c D_c^{-\frac{1}{2}} E_c^T f_c \] (2)

Where \( D_c \) is the diagonal matrix of the covariance matrix eigenvalue \( f_c f_c^T \in R^{C \times C} \), and \( E_c \) is the corresponding eigenvector orthogonal matrix, which satisfies \( f_c f_c^T = E_c D_c E_c^T \).
(2) Coloring. This process is basically the inverse process of whitening, as shown in Equation 3.

\[ \hat{f}_{cs} = E_s D_s^{\frac{1}{2}} E_s^T \hat{f}_c \]  

Where \( D_s \) is the diagonal matrix with the eigenvalues of the covariance matrix and \( E_s \) is the orthogonal matrix corresponding to the eigenvectors.

In order to prove the effectiveness of WCT, we compared it with histogram matching (HM). HM helps to convey the global color of the style image very well, but it cannot capture a visual pattern precisely. In HM, the style of the image is too fragmented after decomposed and the local structure is also affected, resulting in structural distortion of the image. In contrast, WCT (whitening and coloring) proposed in this paper can better reflect the style features of style images.

After whitening and coloring, we used content feature weights \( \alpha \) to control the intensity of style transfer.

\[ \hat{f}_{cs} = \alpha \hat{f}_{cs} + (1-\alpha) f_c \]  

In Equation 4, \( \alpha \) is the style weight, which can be modified in the code. Users can adjust the degree of stylization by modifying the size of \( \alpha \), whose parameter range is from 0 to 1.

4. Experiment and algorithm efficiency

Figure 3 is an example of image style transfer using our model, where Figure 3(a) is a style image, Figure 3(b) is a content image, and Figure 3(c) is a result image. It can be seen from the figure that our model can capture the style and texture of the style image without the need to learn any style in advance.

In order to evaluate the efficiency of the algorithm, Table 1 compares the efficiency between our method with other methods. The method of Gatys is relative slow due to loop optimization, and usually this method takes more than 500 iterations to produce satisfactory results. Methods [7] and [8] are two similar schemes. Their method is based on a feed-forward network and includes a pre-training process. Method [9] is also based on feed-forward network, but because this method requires thousands of patches to exchange features, the speed of it is relatively slow. In practice, our method will be slightly slower than the TNET method. This is because eigenvalue decomposition in the WCT (whitening and coloring) process requires a certain amount of calculation time. It should be noted that since the size of the covariance matrix will only affect the number of images, so the computational cost of this step will not increase with the increase in image pixels or size. The calculation speed of the remaining methods may be affected by the image size and change.
|                  | Chen et al | TNet | Gatys | Ours |
|------------------|------------|------|-------|------|
| Percent/%        | 13.4       | 32.8 | 24.5  | 29.3 |
| Time/s           | 21.3       | 1.82 | 218   | 8.4  |

5. Conclusion and Prospect

In this article, we proposed a style transfer algorithm. The advantage of this algorithm is that it can realize style transfer without learning any style. This algorithm can create the result image by training an automatic encoder that can be used for image reconstruction. In the feed-forward process, we have implemented two main transformation methods: whitening and coloring, which are used to match the statistical distribution and correlation between content features and style features. We also proposed a multi-level model to obtain better transfer results. In addition, our method has the same excellent performance in algorithm efficiency.

References

[1] Van der Geer, J., Hanraads, J.A.J., Lupton, R.A. (2010) The art of writing a scientific article. J. Sci. Commun., 163: 51–59.
[2] Huang H., Zang Y., Zhang L. (2011) Research on stylized image and video oil painting. Computer Science, 38: 1-6.
[3] S. Dong, Y. Ding, J. Qian (2018) Application of AI-Based Style Transfer Algorithm. Animation Special Effects Design, 01: 104-107.
[4] Luan F., Paris S., Shechtman E., et al (2018) Deep Painterly Harmonization. Computer Graphics Forum, 37: 95-106.
[5] Dumoulin V., Shlens J., Kudlur M. (2017) A Learned Representation For Artistic Style. In: ICLR. arXiv:1610.07629.
[6] Li Y., Fang C., Yang J., et al. (2017) Universal Style Transfer via Feature Transforms. In: NIPS. Long Beach.pp. 386-396.
[7] Gatys L. A., Ecker A. S., Bethge M. (2015) A Neural Algorithm of Artistic Style. Nature Communications, arXiv:1508.06576.
[8] Ulyanov D, Lebedev V, Vedaldi A, et al. (2016) Texture Networks: Feed-forward Synthesis of Textures and Stylized Images. In: Proceedings of the 33rd International Conference on Machine Learning. New York.pp.
[9] Shen T, Lei T, Barzilay R, et al. (2017) Style Transfer from Non-Parallel Text by Cross-Alignment. In: 31st Conference on Neural Information Processing Systems. Long Beach.pp. 6830-6841.
[10] Chen T. Q., Schmidt M. (2016) Fast Patch-based Style Transfer of Arbitrary Style. CoRR, abs/1612.04337.