Fusion Image Style Transfer Network

Shuren Lai, Fucheng You, Hechen Gong and Yangze Zhao

Beijing Institute of Graphic Communication. No. 1, Xinghua Street (two section), Daxing District, Beijing
Email: lsr1994104@163.com

Abstract. Generating art automatically and quickly for machines is always a difficult task. The existing image style migration algorithms can realize the rapid generation of art images. However, these algorithms can only generate artistic images by migrating existing content maps, and can not modify the content of images artificially. This paper proposes an image style migration approach which can stylize artificially modified fusion image. Particularly, this strategy can correct the modification of the photographic images and enhance the style shifting details. Aiming at eliminating traces of modification, we adopt an autoencoder network to increase fusion image coordination. This network increases detail representation of fusion image. In order to enhance visual effects of style migration, we improve a neural style migration network to generate fusion image stylized graph. We show that this approach make the fused image more coordinated, resulting in a richer detail of the style image.

1. Introduction

Image style migration has always been a problem, by transforming the style of an art image into another input image. For example, by selecting an appropriate artistic style image, a given input real photographic image is converted into an artistic image with a particular artistic style. Figure 1 shows the results for style transfer on different images. The left column in the figure 1 is the input photographic images that need to be migrated. The middle column in the figure 1 is artistic style images. The right column in the figure 1 is the artistic images after the style migration.

At present, the existing methods can adapt to a large number of artistic styles, but the content of the image scene to be converted is greatly limited. When attempting to perform a style migration on an artificially modified image, the stylized image still has tampering marks. At the same time, the existing style migration methods will lose the original image information when they migrate the artificially modified image, which makes the stylized image generated by the algorithm lack many details and has poor robustness.

In this paper, we propose an image style migration approach which can stylize artificially modified fusion image. Particularly, this strategy can correct the modification of the photographic images and enhance the style shifting details.

This paper is organized as follows. Sections 2 reviews related work of image harmonization and style transfer. The proposed method to transfer fusion image is described in section 3. Sections 4 demonstrate the results of our method and evaluate it.

2. Related Work

2.1. Image Harmonization

The traditional method of blending images is to use linear interpolation to combine foreground and background color values, which is usually done using alpha matting [1]. Pérez et al. Introduced
poisson blending which considers the boundary condition for seamless cloning [2]. Histogram matching method is introduced by Xue et al. [3]. This approach identified key statistical factors that affect the realism of photo compositings such as luminance, color temperature, saturation, and local contrast. Cohen-Or et al. proposed a approach [4] to making image regions compatible is to adjust their colors so that they match a predefined set of templates that are thought to encode color harmony. However, the harmony image does not guarantee authenticity, and the method ignores factors such as brightness and contrast. Aiming at making the objects hard to notice, Camouflage Images [5] adopted an approach to embed objects into certain locations cluttered images. Gijsenij et al. indicated restoring the appearance of input images under neutral illumination is the related problem [6]. However, the problem is highly ill-posed and there is still no effective solution.

The existed image coordination methods are mainly aimed at the visual effect of the fused image itself. Our approach is not aimed at the visibility of fused images, but to enhance the coordination of fusion images after style migration. Therefore, we propose an autoencoder network algorithm to adjust the fused image to make it more conducive to style transfer tasks.

Figure 1. Image style transfer.

2.2. Style Transfer

The early style transfer methods were mostly example-based [7] [8]. The image analogy method [9] finds the relationship between a set of different images and then applies this to other images to achieve style conversion. Analog-based methods [8] often require a large number of images with similar scenes because of the need to find dense correspondences. Therefore, this method is very limited, and it is difficult to implement style migration of arbitrary images.

Recently, deep neural networks have played an increasingly important role in the problem of image generation. Its main idea is closely related to the texture synthesis method. The images generated by these methods have the same style of the semantic information of the target content image and the style image.

Gatys et al. [9] proposed a method to directly modify the random noise so that the overall error is minimized. This approach is based on optimization strategy. For each forward propagation of the method, the original noise is fine-tuned and gradually approaches the final result; there is another way based on a feed-forward network, such as Ulyanov et al. [10], Johnson et al. [11]. These methods train the feed-forward generator network for each specific target image and replace the iterative optimization method with forward propagation, which greatly reduces image generation time. Simultaneously, the resulting networks are remarkably light-weight and can generate textures of quality comparable to Gatys et al.

More recently, a number of methods have been proposed to train a single neural network to transfer multiple styles. Chen et al. [12] propose an optimization objective based on local matching that
combines the content structure and style textures in a single layer of the pre-training network. Liao et al. [13] proposes a two-image direct visual property migration method. This approach is suitable for processing two images with different contents but similar semantic information.

In this paper, we have improved a neural network that enables style migration of the tuned fused image to produce a stylized image that contains more fused image detail features and eliminates fused traces.

3. Fusion Image Style Transfer

3.1. Autoencoder Pre-Training

An automatic encoder is a data compression algorithm in which the compression and decompression functions of the data are data-dependent, lossy, and automatically learned from the sample. Where most of the automatic encoders are mentioned, the functions of compression and decompression are implemented via neural networks.

The autoencoder is data-specific or data-dependent, which means that the autoencoder can only compress data that is similar to the training data. For example, an automatic encoder trained using face training performs poorly when compressing other images, such as trees, because the features it learns are related to the face.

The automatic encoder is lossy, meaning that the decompressed output is degraded compared to the original input, as is the compression algorithm such as MP3 and JPEG. This is different from lossless compression algorithms.

The autoencoder is automatically learned from the data samples, which means that it is easy to train a specific encoder for the input of the specified class without having to do any new work.

Figure 2 shows the basic structure of autoencoder, it consist of encoder and decoder. By converting the original image into the encoder, the original image is converted to compressed representation. Compressed representation and then output the reconstructed image through the decoder. The reconstructed image of the output is the compressed input original image.

![Autoencoder structure diagram.](image)

We pre-train a autoencoder follow the framework proposed by Luo et al.[14]. As is shown in figure 3, it first maps the input vector \( x \) into hidden layer \( y \) through transformation:

\[
y = s(Wx + b)
\]

(1)

Where \( s \) is the activation function, we use hyperbolic tangent. \( W \) and \( b \) are weight matrix and bias vector respectively. The transformation of mapping \( x \) to \( y \) is the encoding process. The hidden representation obtained in this way is the compression representation of input images. After the encoding process, the hidden layer is used as the input layer of the decoder, and the compressed representation is mapped to the reconstructed image. The decoding method is same as encoding:

\[
\hat{x} = s(W'y + b')
\]

(2)
This is the process of decoding, where the output $x$ can be seen as compression of the input image through the hidden layer. The number of nodes in the output layer is equal to the input layer. At the same time, we make the hidden layer's dimension smaller than the input layer. By adding constraints to the hidden layer, it is possible to retain the main features of the original image, and to eliminate the redundant information to feature compression of the input image. Training a self-encoder is to minimize input and output errors. The optimization function is:

$$\min \frac{1}{N} \sum_{i=1}^{N} L(x_i, \hat{x}_i)$$

(3)

Where $L$ is the loss function, this paper uses the squared loss. The training process uses the back propagation algorithm to update the network weights. In this paper, we use CIFAR-100 dataset to train autoencoder, the obtained autoencoder network is recorded as $A$.

![Autoencoder network pre-training](Figure 3)

**Figure 3.** Autoencoder network pre-training.

### 3.2. Fusion Image Modification Trace Elimination Method

We trained a autoencoder $A$ in the previous stage, but this autoencoder still can't fully meet our requirements. We propose a method for eliminating the fused image modification traces, which makes the fused image more suitable for style transfer tasks. The method steps are as follows:

1. The unfused image entered is $I_u$, input $I_u$ into the autoencoder network $A$ for training, use the backward propagation algorithm to update the weight, and stop iterating when the error calculated by the loss function is less than $\alpha$, in this paper $\alpha = 10^{-1}$. The fine-tuned autoencoder network is recorded as $A_f$.

2. Add other images to the unfused image $I_u$ to produce fused images $I_f$.

3. Use the method which proposed in [24] to calculate the brightness value $L_a$ of $I_f$. The formula for calculating the brightness is

$$L_a = \frac{1}{N} \exp \left\{ \sum_{x,y} \log[\delta + L(x, y)] \right\}$$

(4)

Where $L(x, y)$ is the luminance for pixel $(x, y)$, $N$ is the total number of pixels in the image and $\delta$ is a small value to avoid the singularity that occurs if black pixels are present in the image. In this paper, we set $\delta = 0.0001$;
We enhance the pixels of $I_f$ and generate enhanced fused images $I_e$, the formula is

$$I_e = \beta L_u + I_f$$

In this paper we set $\beta = 0.15$;

(5) Input the enhanced mixed image into the autoencoder network $A_e$ to generate image $I_m$.

The images $I_m$ obtained through the above steps have stronger characteristics and image coordination, and can adapt to the style migration task.

3.3. Style Transfer Method

As shown in figure 4, we roughly follow the framework proposed by [11]. This method mainly consists of two parts: an image transformation network $f_w$ and a loss network $\phi$. The image transformation network is a deep residual convolutional neural network parameterized by weights $W$. The input image $x = I_m$ is transformed into an output image $\hat{y}$ via the mapping $\hat{y} = f_w$. Each loss function computes a scalar value $l_i(\hat{y}, y_i)$ measuring the difference between the output image $\hat{y}$ and a target image $y_i$. The image transformation network is trained using stochastic gradient descent to minimize a weighted combination of loss functions:

$$W^* = \arg \min_W E_{x,y} \left[ \sum_{i=1}^{n} \lambda_i l_i(f_w(x), y_i) \right]$$

Network $\phi$ used to calculate the cost function, it is pretrained for image classification as a fixed loss network. Feature reconstruction loss $l_{\text{feat}}$ and style reconstruction loss $l_{\text{style}}$ are defined by the loss network $\phi$ to measure differences in content and style between images:

$$l_{\text{feat}}^j(\hat{y}, y) = \frac{1}{C_j H_j W_j} \| \phi_j(\hat{y}) - \phi_j(y) \|_2^2$$

$$l_{\text{style}}^j(\hat{y}, y) = \| G_j(\hat{y}) - G_j(y) \|_F^2$$

Specific algorithm details are discussed in [11].

![Figure 4. Style migration network.](image)

This paper modifies the network for style transfer of the image $I_m$ generated in the previous step. In lower layers, e.g., $\text{relu}_{1,2}$, $\text{relu}_{2,2}$, the reconstructed style features emphasis on localization that
present as discrete textures and dense fragments. This is in line with our requirements for style transfer. For content target, instead, we need to reduce the representation of low-level modified trace features. For the above reasons, our modifications to the network are shown in the table 1. Enter the image $I_m$ into the modified style migration network to get the fusion image stylized graph. This approach makes the fused image more coordinated, resulting in a richer detail of the style image.

**Table 1.** The adjustment of features reconstruction layers.

| Layers of Style Reconstruction Loss | [11] relu1_2, relu2_2, relu3_3, relu4_3 | ours relu1_2, relu2_1, relu3_1, relu3_3 |
|-----------------------------------|--------------------------------------|--------------------------------------|

| Layers of Content Reconstruction Loss | [11] relu3_3 | ours relu3_3, relu4_3 |
|-------------------------------------|-------------|---------------------|

4. **Experiment**

In this section we show the results of our algorithm processing. As is shown in figure 5. We input three unfused original images $I_{a1}$ (a), $I_{a2}$ (b), $I_{a3}$ (c). We mixed the local contents in $I_{a2}$ and $I_{a3}$ with $I_{a1}$ to generate fused images $I_f$ (d) that is not corrected. For comparison, we input the image directly into the style migration network and output the unmodified fused transfer image $I_t$. At the same time, we use the fusion image modification trace elimination method to correct the image $I_f$. The correction result of output is $I_m$. Finally, the $I_m$ is input to the modified style migration network, and the fusion image stylized graph is $I_T$ (e). The left picture in figure 6 is the unmodified fused transfer image $I_t$, the right picture is the fusion image stylized graph $I_T$ generated by ours method. We can find that our method produces more coordinated image fusion, while retaining more details in the migration process.

![Figure 5. Multi-style migration QR code generation process.](image)
Figure 6. Result comparison.

5. Acknowledgment
This work was partially supported by Joint Funding Project of Beijing Municipal Commission of Education and Beijing Natural Science Fund Committee (KZ201710015010), Project of National Scientific Found (No.61370188), Project of Beijing Municipal College Improvement Plan (PXM2017_014223_000063) and New Project of Green Printing and Publishing Technology by Cooperative Creating Center (PXM_014223_000025) and BIGC Project (Ec201803 Ed201802 Ea201806)

6. References
[1] Porter T, Duff T 1984 Compositing digital images ACM Siggraph Computer Graphics vol 18 pp 253–259
[2] Perez P, Gangnet M 2003 Blake a poisson image editing ACM Trans. Graph vol 22 pp 313-318
[3] Xue S, Agrwala A, Dorsey J, Rushmeier H 2012 Understanding and improving the realism of image composites ACM Trans. Graph. Vol 31 pp 1–10
[4] Cohen-or D, Sorkine O, Gal R, Leyvand T and Xu YQ 2006 Color harmonization ACM Trans. Graph. Vol 25 pp 624–630
[5] Chu H, Hsu W, Mitra N, Cohen-or D, Wong T, Lee T 2010 Camouflage images ACM Trans. Graph. Vol 29 pp 1–8
[6] Gijsenij A, Gevers T, and Weijer V 2011 Computational color constancy: Survey and experiments IEEE Trans. on Image Processing Vol 20 pp 2475–2489.
[7] Hertzmann A, Jacobs C, Oliver N, Curless B and Salesin D 2001 Image analogies Proc. SIGGRAPH vol 1 pp 327–340
[8] Shih Y, Paris S, Durand F and Freeman W 2013 Data-driven hallucination of different times of day from a single outdoor photo Proc. SIGGRAPH vol 32 pp 1–11
[9] Gatys L, Ecker A and Bethge M 2016 Image style transfer using convolutional neural networks Computer Vision and Pattern Recognition pp 2414–2423
[10] Ulyanov D, Lebedev V, Vedaldi A and Lempitsky V 2016 Texture networks: Feed-forward synthesis of textures and stylized images arXiv 1603.03417
[11] Johnson J, Alahi A and Feifei L 2016 Perceptual losses for real-time style transfer and super-resolution Proc. Eur. Conf. Comput. Vis. pp 694–711.
[12] Chen T and Schmidt M 2016 Fast patch-based style transfer of arbitrary style arXiv 1612.04337
[13] Liao J, Yao Y, Yuan L, Hua G and Kang S 2017 Visual attribute transfer through deep image analogy Acm Trans. on Graphics vol 36 p 120
[14] Liu JH, Zheng WQ, Zou YX 2015 A robust acoustic feature extraction approach based on stacked denoising autoencoder IEEE International Conference on Multimedia Bigdata pp 124–127
[15] Reinhard E, Erik M, Stark M, Shirley P and Ferwerda J 2002 Photographic tone reproduction for digital images ACM Transactions on Graphics vol 21 pp 267-276