An Improved Image Style Transfer Algorithm Based on Deep Learning Network

Yan-ni JI, Yu-de WANG* and Jia CHANG

Qufu Noraml University School of Physics and Engineering, Qufu Shandong 273165, China

*Corresponding author

Keywords: Feature fusion, Algorithm optimization, Image stylization.

Abstract. Aiming at the problem of local style migration distortion in image stylization, an improved image style migration algorithm based on deep learning network is proposed. Firstly, the VGGNet-19 network is used to extract the convolution layer features of images. Then the characteristics of convolution layer are analyzed, and the combination of style and content features is studied. The Block3 layer with the smallest content loss and style loss is selected for feature fusion with conv1_1, conv2_1, conv3_1, conv4_1 and conv5_1 layers. Finally, under the optimal combination of features, Adam algorithm is used to optimize the image style migration. The experimental results show that the proposed algorithm can effectively improve the distortion of local style migration and provide theoretical support for the implementation of style migration technology.

Introduction

With the rapid development of artificial intelligence technology, the combination of neural network and art has attracted great attention in the field of art\cite{1-5}. The technology of image style migration based on neural network can make the image which is not a painter can arbitrarily play the style needed to achieve style migration.

Gats et al. first proposed a new style migration algorithm based on neural network in 2015. They mainly found that the content and style representations in convolutional neural network are separable\cite{6}. That is to say, the two representations can be manipulated independently to generate new and perceptually meaningful images. Yu Chao et al. proposed to decode the style and content of an image by using antagonistic branches, and then reconstruct the style of one image and the content of another image by using an automatic encoder to form a new image\cite{7}. Li Yuni et al. proposed an image style migration algorithm based on multi-dimensional histogram matching, which achieves the balance of overall style similarity while retaining local details\cite{8}. Liu Hongqi et al. proposed a semantics-based strongest gravity method, which measures the similarity of high-dimensional feature points of two images by defining gravity, and then minimizes the most gravitational loss function to achieve style transfer\cite{9}. Guo Meiqin et al. proposed an improved facial image style migration based on DPST algorithm. By maximizing the normalized cross-correlation between the channels of feature maps in convolutional neural network, the corresponding radiation changes of the images were made, thus reducing the duplication of images when there are large spatial differences\cite{10}.

Although some progress has been made, there still exists the problem of local effect distortion in stylized images after style conversion. To solve this problem, an image style transfer algorithm based on VGGNet-19 network is proposed, which integrates features and expresses content texture and style texture clearly. Through the experiment, the Block3 layer with the least content loss and style loss is selected to fuse the features with the conv1_1, conv2_1, conv3_1, conv4_1 and conv5_1 layers, and the Adam optimization algorithm is selected. Compared with the traditional style migration algorithm, the local effect distortion of the stylized image is obviously improved.
VGGNet-19 Network

The network structure of VGGNet-19\(^{[11]}\) consists of INPUT (input layer) - Conv (convolution layer) - RELU (activation function) - Pool (pooling layer) - FC (full connection layer). The convolution layer is the core of the convolution neural network. It carries out feature extraction and convolution layer operation as formula (1).

\[
\chi_j^l = f \left( \sum_{i \in M_j} \chi_{i}^{l-1} \ast W_{ij} + b_j^l \right)
\]

In formula (1), \(\chi_j^l\) represents the characteristic graph of the output of the jth convolution core at level L. \(f(x^0)\) denotes a non-linear function. \(M_j\) represents the set of input feature graphs and \(\ast\) represents the convolution operation. \(W_{ij}^l\) is the convolution core between the jth output graph of layer L and the ith input graph of the upper layer. \(b_j^l\) is a bias value.

Mathematical Principle of Style Transfer

Style migration is to transfer the artistic style of one picture to another and produce a content picture with a certain artistic style to achieve the artistic effect of content rendering. The sketch of style migration is shown in Figure 1.

![Figure 1. Style migration diagram.](image)

Gats and others have proved that the content and style of pictures can be separated. Through the way of neural network, the content of the generated pictures can be as similar as that of the content source pictures, and the style of the pictures can be as similar as that of the style source pictures, so that the style of the pictures can be freely exchanged. The mathematical principles used are as follows.

Content Representation

In convolution neural network, different layers will form feature maps corresponding to the number of convolution kernels. In the dimension of content, it is assumed that each feature map of the generated image is as close as possible to the content source image. Content loss and its derivation (for F derivation) are defined as follows:

\[
L_{content}(\hat{p}, \hat{x}, l) = \frac{1}{2} \sum_{i,j} (F_{ij}^l - P_{ij}^l)^2
\]

\[
\frac{\partial L_{content}}{\partial P_{ij}^l} = \begin{cases} 
(P_{ij}^l - P_{ij}^l)_{ij} & \text{if } F_{ij}^l > 0 \\
0 & \text{if } F_{ij}^l < 0
\end{cases}
\]

In formula (2), \(\hat{p}\) denotes the content source picture and \(\hat{x}\) denotes the generated picture (originally input as the content source picture). \(F_{ij}^l\) denotes the eigenvalues of the position i and j of the content source image on layer L, \(P_{ij}^l\) denotes the eigenvalues of the position i and j of the generated image on layer L.

Style Representation

Gram matrix can be seen as the relationship between different filter features, while ignoring the information on the content. Style loss is represented by Gram matrix. Gram matrix is defined. Style Loss \(E_l\) of the Defined L-Layer and Its Derivative Formula is defined.
\[ G_{ij}^l = \sum_k F_{ik}^l F_{jk}^l \quad (4) \]
\[ E_l = \frac{1}{4N_l^2M_l^2} \sum_{ij} (G_{ij}^l - A_{ij}^l)^2 \quad (5) \]
\[ \frac{\partial E_l}{\partial F_{ij}^l} = \begin{cases} \frac{1}{N_lM_l} (F^l)^T (G^l - A^l)_{ji}, & \text{if } F_{ij}^l > 0 \\ 0, & \text{if } F_{ij}^l < 0 \end{cases} \quad (6) \]

Then the style loss \( L_{\text{style}} \) of all CNN layers is:

\[ L_{\text{style}}(\tilde{a}, \tilde{x}) = \sum_{l=0}^L w_l E_l \quad (7) \]

In formula (5) and formula (6), \( N_l \) is the number of filters and \( M_l \) is the size of the width multiplied by the height of the characteristic graph. \( G_{ij}^l \) and \( A_{ij}^l \) represent the Lagrange Matrix Value of the i and j positions of the generated image and the style source image on layer \( L \). \( \tilde{a} \) and \( \tilde{x} \) represent style pictures and generate pictures, respectively. \( \omega \) is the corresponding weight of each layer.

**Style Transfer**

In order to transfer the style of cartoon art image to the target content image, we use the VGGNet-19 model as feature extractor to extract the feature maps of content image and style image from each layer to minimize the loss of content image and style image, and define the minimum loss function as follows:

\[ L_{\text{total}}(\tilde{p}, a, \tilde{x}) = \alpha L_{\text{content}}(\tilde{p}, \tilde{x}) + \beta L_{\text{style}}(\tilde{a}, \tilde{x}) \quad (8) \]

In formula (8), \( L_{\text{content}} \) denotes the loss function of the target content picture, \( L_{\text{style}} \) denotes the loss function of the target style picture, and alpha and beta denote the weight factor of the target content picture and the target style picture respectively. The style migration degree of the generated image can be controlled by adjusting the size of \( \alpha / \beta \).

**Algorithmic Steps and Flow Char**

The implementation steps of the algorithm are as follows:

1. The sample image is normalized by scale. Then, the input image is convoluted by \( \text{N} \) convolution kernels of \( \text{1x1} \) size. The convolution feature map is output by the activation function, and then the feature map of each layer of the sample image is extracted by downsampling.
2. Adding a new computing layer to the original VGGNet-19 network structure for computing style loss and content loss.
3. Content loss is calculated by mean square error, and style loss is calculated by Gram matrix. Content loss and style loss are applied to style transfer loss function to achieve image stylization.

The flow chart of the algorithm is shown in Figure 2.
Experiments and Results Analysis

Experimental Conditions

The experiment uses computer configuration as Windows 64-bit operating system, processor as Intel (R) Core (TM) i5-4590 CPU @3.3GHz 3.30GHz, machine learning framework tensorflow 1.2.1, programming language as Python 3.6. The 2014 MS-COCO\cite{12} data set is used as the content image training set and the artist's works data set\cite{13} for the target style image training set.

Experiments Result Analysis

During the experiment, 200 pictures were selected from the 2014 MS-COCO database as content images, and 100 pictures were randomly selected from the painter's work data set as style images.

**Experiment 1.** VGGNet-19 convolution neural network was used to extract features. Mean square error and Gram matrix were used to measure content and style differences. The network selection of extracting styles was conv1_1, conv2_1, conv3_1, conv4_1 and conv5_1. The network layer of extracting content was conv4_2. The influence of training times on image style transfer was studied and the optimal iteration times were determined.

The weight of style layer is 0.5, 1.0, 1.5, 3.0 and 4.0, respectively. The weight of style loss is 0.001 and the weight of content loss is 1. The number of iterations is 0, 500, 1000, 1500, 2500, 3000. The results of style transfer are shown in Figure 3 and Figure 4.

![Figure 3. The impact of iteration times on style migration.](image1)

![Figure 4. The impact of iterations on content, style, and total loss.](image2)

As can be seen from Figure 4 and Figure 5, with the increase of iteration times, the loss of content increases gradually, tends to be gentle, the loss of style decreases gradually, and finally tends to be zero. The total loss decreases all the time, and finally tends to be zero. Thus, when the iteration reaches about 500 times, the loss of content is relatively small, the loss of style tends to be 0, and the total loss tends to be 0. The effect is the best.

**Experiment 2.** Adjust the style weight ($\alpha/\beta$) to control the degree of style conversion, and study the influence of style weight ($\alpha/\beta$) on image stylization.

In the experiment, the number of iterations was set to 500, and other parameters were the same as experiment 1. Style adjustment parameters satisfy $\alpha + \beta = 1$. When $\alpha = 1$, $\beta = 0$, the network attempts to reconstruct the content image. When $\alpha = 0$, $\beta = 1$, the network attempts to synthesize stylized images. By changing alpha from 1 to 0, interval 0.25, beta from 0 to 1, interval 0.25, a smooth transition between content similarity and style similarity can be observed. The experimental results are shown in Figure 5.
As can be seen from Figure 5, when $\alpha = 1$, $\beta = 0$, the deep convolution neural network only reconstructs the content image. When $\alpha = 0$, $\beta = 1$, the depth neural network achieves the maximum stylization of the image. Therefore, the whole process of alpha from 0 to 1 and beta from 1 to 0 reflects the gradual stylization of the content image. When the content image changes from 0 to 1, we can see that the depth convolution neural network can only reconstruct the content image. You can set your favorite style weight according to the style weight.

**Experiment 3.** The influence of features extracted from each layer of convolutional neural network on style transfer was studied, and the strategy of feature fusion was determined.

In the experiment, VGGNet-19 convolution neural network is used to extract the features of each layer of a building image\cite{14}. The building image is shown in Figure 6 and the result of feature extraction is shown in Figure 7.

Through the visualization of VGGNet-19 network feature map, we can find some rules. Shallow network is more inclined to detect image edges, textures, contours, shapes and other details. The detected content is comprehensive, and key information will be extracted at the same time. With the deepening of the level, feature map is more abstract and abstract. The deeper the level, the more blank areas, the lower the resolution of the image, which indicates that the convolution core has not extracted the required features.

In view of the above rules, Block1, Block2, Block3, Block4 and Block5 are selected as the layers of extracting content images. The default layers of extracting style are conv1_1, conv2_1, conv3_1, conv4_1 and conv5_1. The weight of style and content is set to 0.5. The selected content images and style images are shown in Figure 8 and Figure 9 shows.
Figure 8. Content and style images.

Figure 9. Style migration contrast diagram.

From Figure 8 and Figure 9, we can see that the feature image extracted by Block3 layer of VGGNet-19 has clear texture and edge, and the feature image is representative, comprehensive detection content and good style migration effect.

**Experiment 4.** Studies the optimal combination of multi-layer feature fusion and optimization algorithm based on convolutional neural network.

In the experiment, the feature extraction layer of content image is set to Block1, and other parameters are the same as experiment 3. The experiment compares the gradient descent algorithms of Adagrad, Adadelta and Adam. The experimental results are shown in Figure 10.

Figure 10. Comparison of different gradient descent algorithms.

As can be seen from Figure 10, the migration effect of Adam algorithm is better than that of Adagrad and Adadelta algorithm.

**Summary**

Aiming at the problem of local style migration distortion in the process of image and video stylization, this paper uses VGGNet-19 convolution neural network to study the feature fusion strategy of image style migration according to the characteristics of image features contained in convolution layer features. The experimental method determines the Adam optimization of style migration algorithm. The method improves the effect of image style migration and provides reference value for future research.

**Acknowledgement**

This research was financially supported by the National Natural Science Foundation and Joint Training Base of Postgraduate Education in Shandong Province. Grant Numbers 11505104 and SDYJD17069.

**References**

[1] Jing Y, Yang Y, Feng Z, et al. Neural Style Transfer: A Review[J]. 2017.
[2] Johnson J, Alahi A, Fei-Fei L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution[J]. 2016.

[3] Gatys L A, Ecker A S, Bethge M. Image Style Transfer Using Convolutional Neural Networks[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.

[4] Meyer S, Cornillère, Victor, Djelouah A, et al. Deep Video Color Propagation[J]. 2018.

[5] Li Y, Liu M Y, Li X, et al. A Closed-form Solution to Photorealistic Image Stylization[J]. 2018.

[6] Gatys L A, Ecker A S, Bethge M. A Neural Algorithm of Artistic Style[J]. Computer Science, 2015.

[7] Yu Chao, Lu Shanfu. Image style migration based on Info-autoencoder [J]. Computer Age, 2019 (03): 45-48+52.

[8] Li Yuni, Wang Peng, Xiao Jianli. An image style migration algorithm using multi-dimensional histogram matching [J]. Computer and Modernization, 2019 (02): 15-18+26.

[9] Liu Honglin, Shuai Renjun, Tao Jing, Zhang Qiuyan. Style migration of spatial non-aligned image data based on the strongest gravity [J]. Minicomputer system, 2019, 40 (03): 651-654.

[10] Guo Meiqin, Jiang Jianmin. An improved algorithm for style transfer of face images [J]. Journal of Shenzhen University (Science and Technology Edition), 2019, 36 (03): 230-236.

[11] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.

[12] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr DollAar, C. Lawrence Zitnick. Microsoft COCO: Common Objects in Context[C]. ECCV, 2014.

[13] Kiri Nichol. Kaggle Dataset: Painter by Numbers, 2016. https://www.kaggle.com/c/painter-by-numbers.

[14] Li Y, Fang C, Yang J, et al. Universal Style Transfer via Feature Transforms[J]. 2017.