Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Improved Image Style Transfer Based on VGG-16 Convolutional Neural Network Model

To cite this article: Chenhao Shu and Xichuan Hu 2023 *J. Phys.: Conf. Ser.* **2424** 012021

Improved Image Style Transfer Based on VGG-16 Convolutional Neural Network Model

Chenhao Shu* **and Xichuan Hu**

School of Information Engineering, Shanghai Maritime University, Pudong new area, Shanghai, China

*Email:1360037851@qq.com

Abstract. In the field of computer vision, image style transfer is an important research direction. With the promotion of artificial intelligence, this technology is becoming more and more popular. Compared with the current image style transfer technology based on artificial intelligence, the traditional technology appears to be complex, time-consuming and inefficient. Migration technology is mainly based on images of the current mainstream style VGG - 16 convolution neural network model, adopts TensorFlow framework, in this article to a picture before the smart migration style photo style, texture features such as the methods of improvement, by modifying the style loss function, makes the image content images can transfer a variety of style, Finally, it is verified by relevant experiments.

Keywords. Image style transfer; VGG-16 convolutional neural network model; TensorFlow; Multi-style image.

1. Introduction

Image style transfer technology can be divided into two categories: early traditional style transfer technology and current style transfer technology based on convolutional neural network model. Early traditional image style transfer techniques are mostly based on texture synthesis and non-photo realistic rendering. Julesz[1] proposed that the method of feature extraction is based on texture modeling. Lu[2] et al. proposed that the method based on non-photo realistic rendering is realized through modeling of brush strokes. However, the reason why it is called traditional method is that it has defects and limitations, so it has been eliminated. For the previous methods, image style transfer requires the establishment of a unique mathematical model. This modeling process is very complex and time-consuming, with great limitations, so a new idea of image style transfer emerged later, that is, a new method based on neural network model. Gatys[3] et al. implemented image style transfer by using a pre-trained convolutional neural network model on ImageNet image data set.

In view of the significant style image migration effect of the former, but only one style image can be transferred from one image to another, so in this paper, I will use the VGG-16 model of convolutional neural network to conduct experiments in combination with the TensorFlow algorithm framework developed by Google. The innovation is to change the loss function. A style that allows one content image to migrate multiple style images simultaneously.

2. VGG-16 Neural Network Model

The basic structure of convolutional neural network (CNN) [4] includes five parts: input layer, convolution layer, activation layer, pooling layer and full connection layer. Among them, the Input Layer inputs data, while processing data involves SVD dimension reduction, mean removal, normalization, etc. The second and most important Layer is Conv Layer, whose main work is to carry out local association, treat neurons as a filter, and carry out convolution calculation of local data through sliding window mechanism and filter. The specific operation process is shown in figure 1.

The formula of the operation process is generalized:

$$A(i,j) = (X * F)(i,j) + b \tag{1}$$

Where, $A(i,j)$ represents the position value of filter matrix F corresponding to output matrix (i,j); * denotes convolution operator; X is the input matrix; b is the deviation.

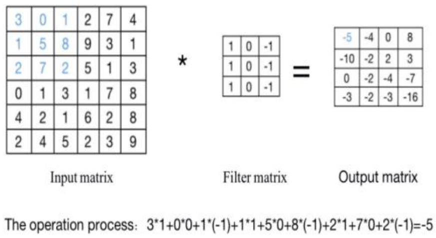

Figure 1. Schematic diagram of convolution

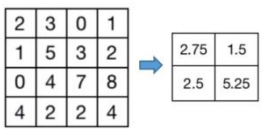

Figure 2. The max pooling

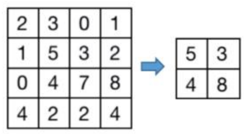

Figure 3. The average pooling operation

If it is the input matrix of order n, the filter matrix of order f, and the padding and stride of border padding are added, then the order x of the final output matrix is:

$$x = \frac{n+2p-f}{s} + 1 \tag{2}$$

At the ReLU Layer, the activation function is used to perform a nonlinear mapping operation on the characteristic matrix output of the convolutional Layer, so as to increase the nonlinearity of the model. ReLU(Rectified Linear Unit) is usually used, which is also one of the most significant unsaturated activation functions.

Pooling Layer, the pooling layer can also be called under sampling, which is mainly used to reduce the dimension and the number of parameters. Its principle is to use specific Windows to divide the feature graph into many non-overlapping regions, and then carry out maximum pooling or average pooling for these regions. The maximum pooling process is shown in figure 2 and the average pooling is shown in figure 3.

At the Fully Connected Layer, after the convolution operation of the convolution Layer and the pooling operation of the pooling Layer, the data features are obtained. Then, the full connection Layer classifies the original image based on these data features, and then adds the bias value, and finally completes the classification output. The expression of the relationship between the outputs of linear neurons and the full connection layer is:

$$f(x) = \sum (w_{ij} x_i) + b_i \tag{3}$$

Where, as the output of the full connection layer, b is the offset quantity.

VGG(Visual Geometry Group),Because it was proposed by a team in Oxford, the VGG model has been updated and iterated with many versions since it was developed [5], and the VGG-16 model was adopted in this paper. The structure of VGG-16 model is shown in figure 4. It can be seen from the figure that the network structure includes 13 Conv layers and 3 Fully Connected layers. VGG model all uses 3*3 filter matrix, and 5 2*2 maximum pooling layers are used for segmentation in the network to reduce the number of parameter.

3. Deep learning framework

Currently, some commonly used deep learning frameworks include TensorFLow, Caffe, Torch, Keras, PyTorch, etc. Compared with other deep learning frameworks, the speed of TensorFLow's training samples is faster. According to online search data, TensorFlow is superior in all aspects to commonly used frameworks such as Caffe and Torch [6].

The deep learning framework adopted in this paper is TensorFlow[7], which is a new generation of learning system in the room launched by Google in 2015. It is also a core open source library that can help us develop and train the machine learning model we need. TensorFlow has a multi-layer structure and is widely used in the realization of various machine learning algorithm programming due to its advantages of deploying various servers and supporting TPU and GPU high-performance computing.

In this article, we also use a third-party high-level API that supports TensorFlow. Based on TensorFlow's Python API, this API provides the building blocks for neural networks, especially deep neural networks. It also encapsulates the development, training, and testing of neural networks, thus enhancing scalability and reducing difficulty.

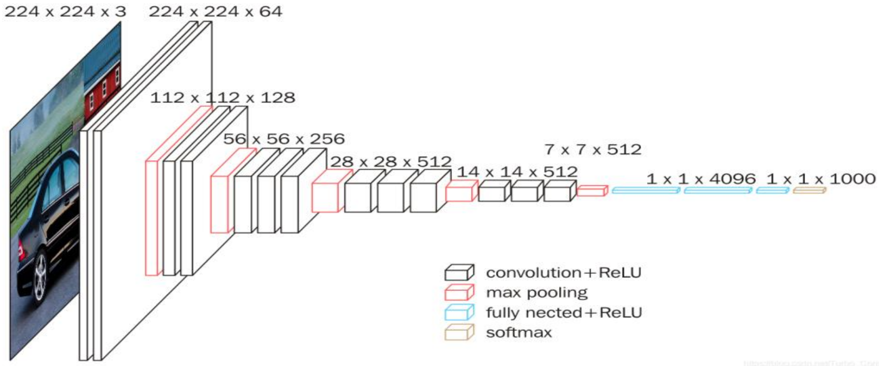

Figure 4. VGG-16 network structure diagram

4. Image Style Transfer

Image style transfer is an important research content in the field of computer vision. As shown in figure 5, this is a simple example diagram of image style transfer. Figure 5 (I) is the picture of Dilieba as the content image; Figure 5 (II) is van Gogh's masterpiece Starry Night, which is mainly blue and yellow; Figure 5 (III) is the output image after the style transfer. According to the comparison of the three pictures, it can be clearly observed that figure 5 (III) has both the character content of Dilieba and the style of the starry sky.

To realize the process of image is probably so, actually this is taken by the VGG - 16[8] neural network model, the first is to obtain the image of the contents and style characteristics, and then use loss loss loss function to calculate the content and style, the key core is to reduce the total error, and constantly adjust the model parameters and the number of iterations, to achieve the best effect of migration image style.

5. Style Migration Loss Function Builds

The loss function of image style transfer is generally used to measure whether the difference between the effect of the image generated after style transfer and the expected effect can be accepted. The value of the loss function is inversely proportional to the effect of the target image. Therefore, in the process of training the model, the value of the loss function is gradually reduced through constant adjustment, so as to finally output a better style transfer image.

5.1. Content Loss Function

The content loss function is:

$$L_{content}(\vec{c},\vec{x},l) = \frac{1}{2}\sum_{i,j}(X_{ij}^{l} - C_{ij}^{l}) \tag{4}$$

Where, $\vec{c}$ is the input content image, and $\vec{x}$ is the output style transfer image, l indicating that the current location is the convolution layer. X_{ij}^{l} denotes the feature vector of the output style image $\vec{x}$ at the l layer, C_{ij}^{l} denotes the feature vector of the input content image $\vec{c}$ at the l layer, X_{ij}^{l} and C_{ij}^{l} denotes the i at the i channel of the convolution layer, and j denotes at the j position of the convolution layer.

5.2. Style Loss Function

For style, because using different filter matrix, so the extracted image features is also different, we are in the feature extraction, naturally one layer, so the extracted features are more complicated, but how to link these complex characteristics, find out their correlation, studies have shown that The correlation between multi-bit features in different dimensions can be calculated by Gram matrix, and then the difference can be quantified, and then the relationship between each feature vector can be obtained. Gram matrix [9] can calculate the characteristic information as follows:

$$G_{ij}^{l} = \sum_{k} F_{ik}^{l} F_{jk}^{l} \tag{5}$$

Where, F_{ij}^{l} represents the feature vector of the image located at j on the i filter matrix of the l layer.

The style loss function is:

$$L_{style}(\vec{s},\vec{x},l) = \frac{1}{4H_l^2W_l^2}\sum_{i,j}(X_{ij}^{l} - S_{ij}^{l})^2 \tag{6}$$

The width and height of the convolution layer are defined as W_l、H_l, the Gram matrix S_{ij}^{l} corresponding to the style image is $\vec{s}$; the Gram matrix X_{ij}^{l} corresponding to the image generated by the style transfer is $\vec{x}$.

The improvement point of the VGG-16 style transfer neural network model in this paper is to improve the style function, so that in the past, only one image can migrate one style image, but now it can migrate k (k=2,3......) after improvement multiple style features on a single content image. So the new style loss function is:

$$L_{style}(\vec{s},\vec{x}_1,\vec{x}_2,..\vec{x}_k,l) = \frac{1}{4H_l^2W_l^2}\sum_{k}\sum_{i,j}(X_{ijk}^{l} - S_{ij}^{l})^2 \tag{7}$$

In VGG-16, multiple convolutional layers are used to calculate the loss value, and a weight coefficient a_l can be added for each convolutional layer, so the final loss function is:

$$L_{style}(\vec{s},\vec{x}_1,\vec{x}_2,..\vec{x}_k) = \sum_{l} a_l\, L_{style}(\vec{s},\vec{x}_1,\vec{x}_2,..\vec{x}_k,l) \tag{8}$$

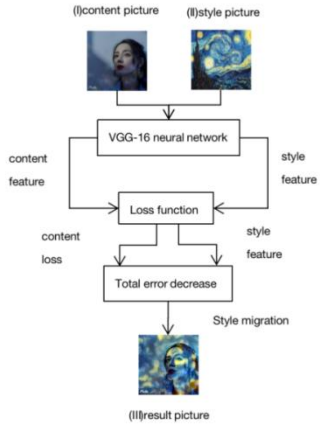

Figure 5. Image style transfer instance

Figure 6. Generated graphs with different content weights and style weights

5.3. Total Loss Function

In the convolutional neural network model based on VGG-16, the total style loss function is summed by the content loss function and the style loss function in a certain proportion, $\alpha\beta$ are respectively used to adjust the corresponding proportion value of the content loss function and the style loss function, that is, the proportion in the total loss function. The purpose of $\alpha\beta$ is to adjust the output style transfer image to style or content [10]. The formula of the total loss function is:

$$L_{total}(\vec{c},\vec{s},\vec{x}) = \alpha L_{content}(\vec{c},\vec{x}) + \beta L_{style}(\vec{s},\vec{x}) \tag{9}$$

Where $\vec{c}$ represents the original content image, $\vec{s}$ represents the original style image, and $\vec{x}$ represents the generated style transfer image.

6. Experiments

6.1. Experimental Environment

During the experiment, the PyCharm compiler was used to run Python codes, and TensorFlow was used as the framework. The computer system for the experiment was IOS12.2.1, the processor was quad-core Intel Core i5, and the memory was 8GB.

The content images used in the experiment are dilireba's character content images, while the style images are clear in color and style characteristics. The image style transfer experiment is carried out under different conditions respectively. Finally, reasonable values of content images and style images are selected to conduct the multi-style image transfer experiment.

6.2. The Experimental Process

Because the experiment content in an image is the focus of the migration on a variety of style characteristics, so the philosophy of this paper is, first, the weight of the contents and style selection of weights, produce the best effect of weights are selected, and then in the screening of different number of iterations, choose to highlight the best effect of the number of iterations, finally to multiple styles of migration, the advantage is, It can make the experimental effect reach the expected effect more quickly, saving the time wasted in the later need to re-adjust parameters.

Table 1. Set content weight and style weight parameters in VGG-16 model

The number	Content image weight α	Style image weight β
1	0.01	4.5
2	0.02	4.5
3	0.05	4.5
4	0.05	3.5
5	0.05	2.5
6	0.05	1.5

Table 2. Setting of different iterations in VGG-16 model

The number	α β	The number of iterations
1	0.02、4.5	10
2	0.02、4.5	20
3	0.02、4.5	30
4	0.02、4.5	40

Table 3. Experiment allocation table

The number	Content image	Style image	output
1	a	b,e	Figure 9
2	a	c,f	Figure 10
3	a	d,g	Figure 11

In figure 6 below, the number of iterations is set to be fixed, followed by the effect diagram after style transfer under six content weight and style weight values.

According to the result of the six groups style after migration figure you can see, in the specified number of iterations of the, if control style image weight remains the same, the smaller the weight of the content image, then the result the content of the image display is not obvious, if the content image of the weight, the greater the style after the migration of the character of the image is, the more obvious. If the weight of the control content image remains unchanged, the smaller the weight of the style image is, the less obvious the style features will be in the image after the style migration; on the contrary, the larger the weight proportion of the style image is, the more obvious the style features will be. By comparing the effect pictures obtained from six groups of different parameters, it can be seen that when setting α=0.02 and β=4.5, that is, the style transfer effect of picture no. 2 is the best, and the character features and style features are clearly displayed.

Then set different iterations in table 2 below, and compare the style transfer effect diagrams under various iterations to obtain the best iteration times.

In figure 7, the content weight and style weight parameter values numbered 2 in table 1 are adopted.

By setting different iterations, the influence of factors with different iterations on the effect of style migration is compared, and the iteration number with the best migration effect is finally selected to facilitate multi-style migration. Reduce the time spent on tuning and error reduction processes for the final multiple style migrations。

Figure7. Generated graphs of different iterations

Figure 8. One content image and six style image

According to the results of the experiment, it can be seen that no. 4, the number of iterations is 40, α=0.02, β=4.5, shows the best style transfer effect, and the character features in the content image and the style features in the style image are clearly expressed.

Finally, based on the conclusions obtained from the above basic experiments, we used these conclusions to conduct the core experiment of this paper. Each group of experiments used a content picture to migrate with two styles of pictures, and the three groups of experiments were compared.

The three groups of experiments were carried out according to table 3. The number of iterations was 40, and the weight of content image was 0.02, The style image weight is 4.5.

In the first group of experiments, the content image is a and the style image is b and e. After the style transfer, the output image is shown in figure 9.

Figure 9. The first set of experimental results

Figure 10. The second set of experimental results

Figure 11. The third group of experimental results

In the second group of experiments, the content image is a and the style image is c and f. After the style transfer, the output image is shown in figure 10.

For the third group of experiments, the content image is A and the style image is D and G. After the style transfer, the output image is shown in figure 11.

7. Conclusion

According to the previous image of the content migration in a style texture on the basis of experiment, we may safely draw the relevant conclusion, when the number of iterations constant control, when increase the content of the image of weight value, so in style after migration results in the image, the character of the content of the image will be increasingly obvious, and style of the image style features in the image will slowly disappear. At the same time, if the weight value of the style image is increased, the style features will be gradually obvious in the result graph of style transfer, while the character features of the content image will gradually disappear.

The purpose of the basic experiment is to better complete the subsequent multi-style transfe r experiment, effectively reduce the time spent in multiple parameter adjustment in the process of multi-style transfer experiment, and make the experiment more effectively achieve the desi red effect.

In a variety of style experiments, we take the three groups of experiments, the contents of the experiment in each group have a fixed image and style of two images, after migration, the image of three groups of output, is a blend of two kinds of style, but there is a common problems, when a certain style style characteristics of the image or color is relatively obvious or bright, it in the style of the migration, It will cover another style to a large extent, which is also the area to be improved in the follow-up research.

8. Future Outlook

At present, a wide variety of styles migration also has some shortcomings, some of the image color is more bright-colored when a wide variety of styles, or the texture characteristics compared to other texture feature is more obvious in figure, after migration, so in a wide variety of styles that is not too bright color or not obvious texture will be overwritten, so in the subsequent migration of the style of study, It is hoped that researchers can continue to update and iterate the technology in this direction.

References

[1] Julesz B. Visual Pattern Discrimination, [J]. IRE Transactions on Information Theora, 1962, 8(2) : 84-92.

[2] Heeaer D J,Bergen J R. Pyramid-based Texture Analysis/Synthesis[C-//Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Technigues,1995: 229-238.

[3] Gatys L A,Ecker A S, Bethge M. Image Style Transfer Using Convolutional Neural Networks, C-//Proceedings of the lEEE Conference on Computer Vision and Pattern Recovnition,2016: 2414-2423.

[4] Ji Changqing, GAO Zhiyong, Qin Jing, Wang Zumin. Summary of image classification Algorithm based on convolutional Neural Network [J/OL].Computer Application:1-6[2022-03-30].

[5] SIMONYAN K , ZISSERMAN A.Very deep convolutional networks for large-scale image recognition [J] .ArXiv preprint arXiv, 2014(9):1-14.

[6] Miao Xu-juan, YU Hao, WANG Lu, GUO Rui-jia, Yang Tian-hui, NIU Ben-jie. General Situation and Research Status of Image Style Transfer Technology [J]. Modern Computer,,2021(02):67-72.

[7] ABADI M, AGARWAL A, BARHAM P, et al.TensorFlow:large-scale machine learning on heterogeneous distribute systems [J] .ArXiv preprint arXiv, 2016(3):1-19.

[8] Wu Ziyang, He Dan, Li Yingqin. Image Style Transfer Based on VGG-19 Neural Network Model [J]. Science and Technology and Innovation,2021(13):171-173.DOI:10.15913/j.cnki.kjycx.2021.13.073.

[9] Yu Zhi-fan, LI Hao, LI Deng-shi, HU Xi. Style Transfer algorithm based on Gram Matrix and Convolutional Neural Network [J]. Journal of Jianghan University (Natural Science Edition),2020,48(03):62-68.DOI:10.16389/j.cnki.cn42-1737/n.2020.03.010.

[10] Wang Ting, LI Hang, Hu Zhi. Design and Implementation of VGGNet Image Style Transfer Algorithm [J]. Computer Applications and Software, 2019, 36(11) : 224 - 228.