Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/39/20230589

CNN-based image style transformation--Using VGG19

Yihao Xu

Computer Science and Technology, Donghua university, Shanghai, 201260, China

211380121@mail.dhu.edu.cn

Abstract. Neural Style Transfer is a widely used approach in the field of computer vision, which aims to generate visual effects by integrating the information contained in one image into another. In this paper, this work presents an implementation of neural style transfer using TensorFlow and the VGG19 model. The proposed method involves loading and preprocessing the content and style images, extracting features from both images using the VGG19 model, and computing Gram matrices to capture the style information. A StyleContentModel class is introduced to encapsulate the style and content extraction process. The optimization process is performed using the Adam optimizer, where gradients are applied to iteratively update the generated image. The number of epochs and steps per epoch can be adjusted to control the optimization process and achieve desired results. Experiments show that we are effective in generating an image that is able to integrate the content of one image into the other. The generated images exhibit visually appealing characteristics and showcase the potential of neural style transfer as a creative tool in image synthesis. Future work may involve exploring different variations of the style transfer algorithm, optimizing hyperparameters, and evaluating the performance on a wider range of image datasets. Additionally, the integration of other deep learning architectures and techniques could further enhance the capabilities of neural style transfer.

Keywords: Style Transfer, CNN, VGG19, CV.

1. Introduction

Style transfer is a computer vision technique that generates images with a new style by combining the content of one image with the style of another [1]. It has a wide range of applications in the fields of art creation, image editing, and design. The development of style transfer is inseparable from the application of Convolutional Neural Network (CNN).

A convolutional neural network is a deep learning model with multiple convolutional and pooling layers that learn to extract features from raw images [2]. These features capture different levels of abstraction of images at different levels, from low-level edges and textures to high-level shapes and objects. By extracting features from two input images separately, style transfer can combine the content of the original image with the features of the style image to generate a new image. In the field of image processing, CNN is widely used in many tasks, including image classification, object detection, image segmentation and image generation, etc. Through large-scale training on training data, CNN can learn general image feature representation, so it performs well in many image-related tasks. This makes CNNs ideal for implementing image style transfer, as they are able to extract useful information about content and style from images.

DOI: 10.54254/2755-2721/39/20230589

The concept of style transfer was first proposed by Gatys et al. in 2015. The CNN model was applied to realize the transformation of the picture style. In particular, they are able to get the characteristics of the picture by means of the pre-trained VGG net and express the picture with Gram matrix. High-quality style transmission is realized by minimizing the different styles of the produced and the object styles and keeping the same content of the produced and the original. In particular, they adopt a pre-trained VGG net to extract the character from the image, and then apply the Gram matrix to express the picture's style. The Gram matrix is derived from the relationship among the characteristic maps, which can be used to capture the texture and statistics of the pattern image [3]. Thus, high-quality picture style transmission can be realized by minimizing the different styles of the produced and the object style images and keeping the same content of the produced and the original. Ever since then, style transfer has been extensively researched and used in academic and industrial circles [4,5]. A number of improvements have been put forward in order to enhance the quality and effectiveness of style transmission. Some of these methods include:

Increasing network depth and complexity: By increasing the number of network layers and parameters, the researchers enabled the network to better capture the details and style information of the image.

Multi-scale style transfer: By introducing feature maps of multiple scales in the network, the global and local features of the image can be better captured, thereby improving the effect of style transfer.

Real-time applications of style transfer: To implement style transfer in real-time applications, researchers have proposed some lightweight network structures and optimization algorithms to speed up image generation.

Cross-domain application of style transfer: In addition to the image field, style transfer has also been applied to other fields, such as video style transfer, audio style transfer, and text style transfer [6,7]. Ever since Gatys and others' research, a lot of scholars have made improvements on and expanded the style shift and put forward a number of ways to enhance the performance and effectiveness of style transfer. Some of these improvements include increased network depth and complexity, multi-scale style transfer, real-time application, and cross-domain application. These methods make style transfer more flexible and practical, enabling more diverse style transfers in different domains.

In breif style transfer is a technique that uses convolutional neural networks to extract image features and combine the content and style of different images. It has a wide range of applications in the fields of art creation, image editing and design. In the past few years, researchers have proposed many improved methods to improve the quality and efficiency of style transfer and apply it to other domains. The development of style transfer technology provides us with more possibilities to create and edit images, and also brings new challenges and opportunities to research in the fields of computer vision and artificial intelligence.

2. Method

2.1. Data preprocessing

First convert the pre-image input into tensors. Before loading the image, use the ` function to read the raw data of the image file. Then, use a function to decode the raw data into an image tensor. Next, use the function to convert the image to `tf. float32` type and normalize the pixel values between 0 and 1. Finally, use a function to resize the image to a maximum side length of 512 pixels and add a batch dimension.

2.2. Model-choosing

Use the middle layer of the model to obtain the content and style representation of the image. Starting from the input layer of the network, the excitation response of the first few layers represents low-level features such as edges and textures. As the number of layers deepens, the last few layers represent more advanced features - physical parts such as wheels or eyes. We are using the VGG19 network structure, which is a pre trained image classification network. These intermediate layers are necessary for defining

DOI: 10.54254/2755-2721/39/20230589

content and style representations from images. The custom model is a convolutional neural network composed of a series of convolutional layers and maximum pooling layers. Its structure is similar to the first half of the VGG19 model, but more simplified. The structure of the model is as follows Figure 1:

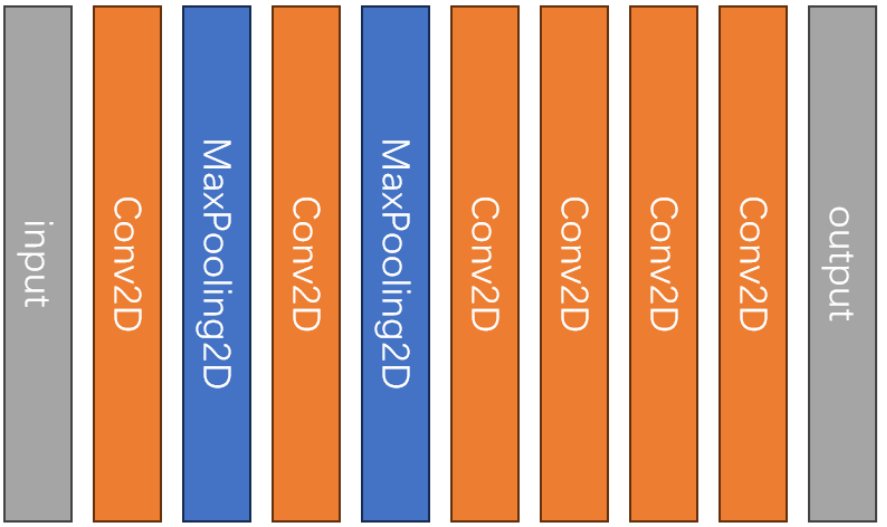

Figure 1. Model structure.

This model includes 5 convolutional layers and 2 maximum pooling layers. Each convolutional layer is followed by a ReLU activation function. The purpose of custom models is to extract feature representations of input images. We have defined a new model and output feature representations at different levels. These feature representations are obtained by extracting outputs on the convolutional layer of the model. The outputs of different levels of the model are considered feature representations of the image. These feature representations are high-dimensional tensors, where each channel corresponds to a specific feature. The feature representations extracted at earlier levels contain more low-level features, while the feature representations extracted at deeper levels contain more high-level semantic features. By adjusting the number and order of convolution layers and maximum pooling layers in the model, feature extraction at different levels can be realized, and feature information at different levels of the image can be captured.

2.3. Computing feature representations

Computing feature representations for style images involves two main steps: loading partial layers of the VGG19 model and computing the Gram matrix. The full VGG19 model is loaded first, and the pre-trained ImageNet weights are used. Then, by iterating through the layers of the model, the output of each layer of interest can be obtained. Here, we focus on the feature representation of style images, so some style layers and content layers are selected.

$$G_{cd}^{1} = \frac{\sum_{ij} F_{ijc}^{1}(X) F_{ijd}^{1}(X)}{IJ} \tag{1}$$

The Gram matrix is a method for representing the statistical information of the feature map, which can capture the style information in the feature map [8,9]. The calculation of the Gram matrix involves the inner product operation of the feature map, which can capture the correlation between different channels in the feature map. Then, by calculating the Gram matrix for every style level of a style image and comparing the Gram matrix of the produced image with that of the Style Losing Function, it is possible to keep the style characteristics of the style image visually. The computation of the Gram matrix involves converting the feature map into a matrix and computing its inner product. In the code, a function called `gram_matrix` is defined to compute the Gram matrix. The `gram_matrix` function uses the `tf.linalg.einsum` function to compute the inner product of feature maps. It takes an input tensor

DOI: 10.54254/2755-2721/39/20230589

`input_tensor` with dimensions `[batch, height, width, channels]`. The function first multiplies the two dimensions `bijc` and `bijd` of the input tensor to obtain a tensor with dimension `[batch, channels, channels]`. Then, the Gram matrix is normalized by dividing by the total number of pixels of the feature map, ie `num_locations`. We can use the `gram_matrix` function to calculate the Gram matrix of each style layer by iterating over the output of the style layer.

2.4. Loss

The definition of the loss function is one of the most important parts of the Style Shift Algorithm, which can be used to determine the difference between the produced picture from the object and the object. The loss function is made up of the content loss and the pattern loss, and I'll explain in more detail how to calculate them.

First, input the target content image and the generated image into the pre-trained convolutional neural network (such as VGG) to obtain their feature representation in one or more layers. Then, by computing the difference between the feature representations of the target content image and the generated image at these layers, the mean squared error (MSE) is used to measure the content loss [10]. The content information of the captured image.

First, this paper inputs the object style and the produced image into a deep convolutional neural network. Calculate the Gram matrix of the object and the generated image for each level. Gram matrix is a kind of internal product matrix for characteristic expression, and it can be used to get the relation and style information among the characteristics. The Average Squared Error is utilized to measure the pattern loss by calculating the difference between the Gram matrix of the object and the Gram matrix of the produced image. Typically, the pattern loss is computed at the bottom level, since the lower level is more precise, which can capture the texture and detail of the picture.

2.5. Training step

First, a trainable variable is defined as a TensorFlow variable whose initial value is the content image.

Next, a function is defined that executes each step of the training. In this function, you can calculate the gradient of a loss function relative to an image by means of TensorFlow's Global Framework Manager. The gradient is calculated by invoking the gradient function, where the loss is the style and content loss calculated by the function.

The image is then updated by applying the gradients using the apply_gradients method of the optimizer (in this case, the Adam optimizer). This will adjust the resulting image towards minimizing the loss.

After applying the gradient, the image's pixel values are clipped to the range [0, 1] by calling the function to ensure that the image maintains valid pixel values.

During training, the training images are iterated through multiple calls. In general, training steps are repeated for one or more epochs.

At each training step, a loss function is calculated, and gradients are applied to the image to update the pixel values of the image. In this way, the generated images are progressively adjusted to better match the style features of style images and the content features of content images.

DOI: 10.54254/2755-2721/39/20230589

3. Result

3.1. Image dataset

Figure 2. Content image.

Figure 3. Style image.

When this paper chooses the target image style image, this paper chooses various types of original images and various styles of style images as the Figure 2, Figure 3, and Figure 4 show.

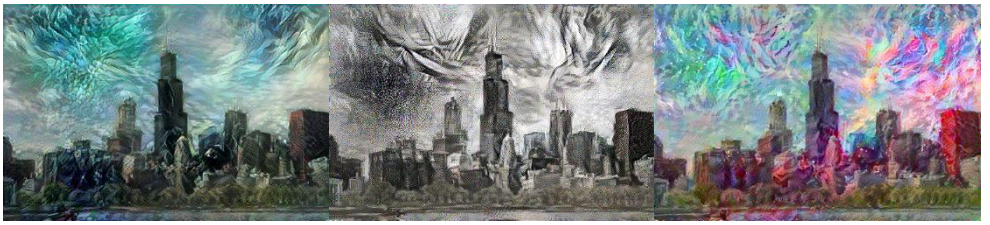

Figure 4. Part output.

3.2. Loss image

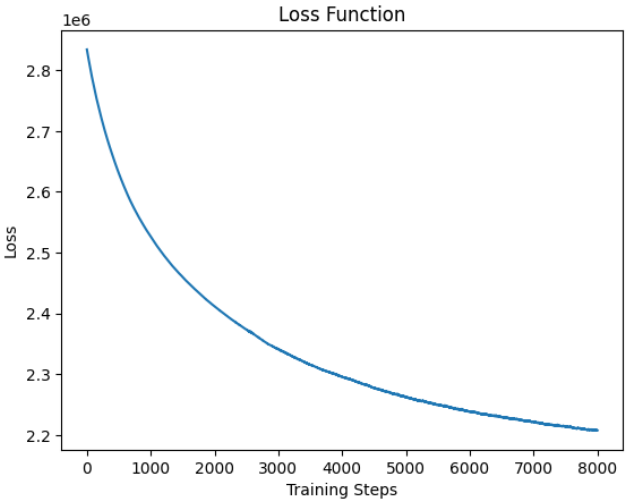

Figure 5. Loss of training.

DOI: 10.54254/2755-2721/39/20230589

According to the image of this loss function, it can be concluded that the overall loss of this model gradually decreases with epoch, so the performance of this model is relatively good, but the problem is that the above output picture is not very good, so there are still some gaps compared with the pictures made by more mature painters. This paper needs to optimize the optimization algorithm and model structure in the future. The loss of training is shown in Figure 5, and the standard picture is shown in Figure 6.

Figure 6. Standard picture.

4. Conclusion

This paper proposes a novel approach to realize the transmission of neural networks based on VGG19 and TensorFlow. Our method is successful in blending one picture's content with the other's, creating an attractive and innovative combination of imagery. This is done by taking VGG19 as a feature extraction tool to extract and manipulate the style and content from an input picture. By computing Gram matrices to capture the style information, this work is able to transfer the desired style onto the content image, producing a unique and visually striking output. The optimization process, performed using the Adam optimizer, iteratively updates the generated image to minimize the style and content losses. By carefully adjusting the number of epochs and steps per epoch, this work has control over the optimization process, allowing us to achieve desired results and enhance the visual quality of the synthesized images. Experiments have shown that we can achieve a good combination of both text and text. The generated images exhibit artistic characteristics and showcase the potential of neural style transfer as a powerful tool in image synthesis and creative expression. While our implementation has achieved promising results, there are still areas for future exploration and improvement. Further research can be conducted to investigate different variations of the style transfer algorithm, optimize hyperparameters, and evaluate the performance on diverse image datasets. Additionally, the integration of other deep learning architectures and techniques could enhance the capabilities of neural style transfer and open up new avenues for creative image synthesis. Overall, the presented neural style transfer framework using TensorFlow and the VGG19 model offers exciting opportunities for artistic expression, visual design, and creative applications in various domains. It provides a powerful tool for generating visually appealing and unique images by combining the content and style of different sources. This work believe that this work contributes to the advancement of image synthesis techniques and serves as a foundation for future research in this field.

References

[1] Gatys, L.A., Ecker, A.S., and Bethge, M., 2016. Image style transfer using Convolutional Neural Networks. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[2] isco, Y., and Carmona, R., 2022. Face recognition using deep learning feature injection: An accurate hybrid network combining neural networks based on feature extraction with Convolutional Neural Network. 2022 XVLIII Latin American Computer Conference (CLEI).

[3] Chu, W.-T., and Wu, Y.-L., 2018a. Image style classification based on learnt deep correlation features. IEEE Transactions on Multimedia, 20(9), pp.2491–2502.

[4] Long Zhang,2022. Research on Image Style Transfer Algorithm Based on Generative Adversarial Networks. Master's thesis, Chongqing University of Posts and Telecommunications, 2023.

[5] Renwei Tang, 2022.Research and application of real image deformable style transfer. Master's thesis, University of Electronic Science and technology.

DOI: 10.54254/2755-2721/39/20230589

[6] zhiwei Wang, 2023.Research on brand visual identity design based on Generative Art. Master's thesis, Jiangnan University
[7] Wang, Y., and Li, L., 2022. The research and implementation of clothing style transfer algorithm based on cyclegan. 2022 4th International Conference on Pattern Recognition and Intelligent Systems.
[8] ZePing Liu, 2022.Research on AGV gesture scheduling based on global vision. Master's thesis, Qingdao University,
[9] Li, Y., Wang, N., Liu, J., and Hou, X., 2017. Demystifying neural style transfer. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence.
[10] yanhua Liang, 2022. Research on the theory and method of image saliency detection model for different modes. Master's thesis, Jilin University