

Research on Image Style Transfer Methods Based on Deep Learning

Jiandong Zhang^a

Department of Mathematics and Computer Science, Nanchang University, Nanchang, China

Keywords: Style Transfer, Convolutional Neural Networks, Generative Adversarial Networks, AdaIN Algorithm.

Abstract: In order to create a new image technology with both properties, the image style transfer technique involves extracting the image's style attributes from the input style pictures and integrating them with the content pictures. As deep learning has advanced over the past few years, style transfer technology problems have seen an increasing application of deep learning networks. This paper summarizes the basic concepts of style transfer technology, introduces the different networks in the deep learning network structure applied in style transfer, as well as the specific models and algorithms to achieve style transfer under different networks, and finally analyzes and compares the migration effects of different networks according to the migration results of different pictures. In addition, this paper also introduces and explains the flow and algorithm structure of AdaIN algorithm, another common technique in style transfer. The purpose of this paper is to summarize and review the transfer technology based on deep learning network used in image style transfer technology, provide theoretical reference for subsequent researchers, and promote the development of this field.


1 INTRODUCTION

Style transfer technology combines the features of style images and content images to create more innovative and visually appealing images. In recent years, style transfer technology has been widely applied in the self-media industry and animation culture industry. Traditional non-deep learning-based style transfer techniques combine style images and content images to generate content images with the target style.

Gatys (Gatys, 2017) et al. trained a convolutional neural network (CNN) model on the Imagenet dataset using transfer learning (Pan, 2009), achieving picture style transfer in the era of deep learning. They also defined a loss function based on CNN style transfer method, using high-level convolutional layer features to provide content loss and integrating feature maps from multiple convolutional layers to provide style loss. This enabled the computer to recognize and learn artistic styles, which could then be applied to regular photos to successfully achieve image style transfer. The deep learning-based style transfer method has much better results than traditional methods. Later, Jin Zhi-gong and others improved the style transfer algorithm by proposing a more suitable

convolutional neural network structure for image style transfer and improving the loss function for style transfer, which can enable a single image to be transferred to multiple different artistic styles at the same time.

One of the most important instruments in the transfer of image styles is the generative adversarial network (GAN). Zero-sum games served as an inspiration for Ian Goodfellow and others in 2014 when they presented the GAN (Goodfellow, 2014). A Cycle-Consistent Generative Adversarial Networks (CycleGAN) was proposed by Zhu et al. (Zhu, 2017). This network enables the original image and target image to be styled in the same way. This breaks the limitation of paired training data in supervised learning and can be used for image style transfer with unpaired training data. This GAN's structure just needs to establish a dynamic balance through an adversarial process between the discriminator and the generator in order to accomplish mutual style transfer between the target image and the original image. It does not require a sophisticated loss function. Many domestic and foreign scholars have improved the CycleGAN algorithm and achieved certain effects. Although the research has achieved good transfer effects, there are still problems such as loss of details

^a <https://orcid.org/0009-0000-7258-9887>

and image authenticity needs to be improved. To solve these problems, Li et al. proposed an improved CycleGAN network model, replacing the original Resnet network with a U-net to better retain image details and structure; integrating self-attention mechanism into the generator and discriminator to further enhance the attention to important details and reconstruction ability, and generate more realistic and delicate transfer effects (Li, 2023).

AdaIN (Huang, 2017) utilizes Encoder, Decoder structures, allowing the transmission of arbitrary styles without training a separate network, but due to the method's failure to retain the content image's depth information, rendering quality is poor. Wu et al. extended and improved the AdaIN method by integrating the depth computation module of the content image into the Encoder, Decoder structure while preserving the structure, resulting in a final output of style-enhanced images that balances efficiency and depth information, thereby improving rendering quality (Wu, 2020).

This paper will introduce and summarize the basic concepts of style transfer, the specific implementation steps of convolutional neural network subnetworks (such as Visual Geometry Group Network(VGG)) in style transfer, and the steps of subnetworks (such as CycleGAN) of generative adversarial networks in style transfer. Finally, the implementation flow of AdaIN algorithm in style transfer is introduced, and the future research directions of style transfer are prospected.

2 IMAGE STYLE TRANSFER BASED ON NEURAL CONVOLUTIONAL NETWORKS

2.1 Introduction to Neural Convolutional Networks

2.1.1 The Basic Mechanism and Principle of Convolutional Neural Networks in Style Transfer

The input layer, pooling layer, fully connected layer, convolutional layer, and output layer are the five levels that make up a CNN.

- (1) Input layer: receives input image information
- (2) Convolution layer: extracting local characteristics of the picture. The convolution layer contains a set of learnable convolution nuclei, each of which

can be used to detect and extract certain features of the input image.

(3) Pooling layer: while keeping sufficient feature information, shrink the feature map's size. Maximum pooling is the most widely utilized of the two basic pooling techniques, the other being average pooling.

(4) Full connection layer: expand the features of the pooled layer to generate a set of one-bit data into the output layer

(5) Output layer: classify images or generate target images

Generally, several convolutional layers are connected to a pooling layer to form a module. The final module will link to at least one complete connection layer after a number of comparable modules have been connected in turn. The final full connection layer will link to the output layer following the extraction of the module's input features by the full connection layer.

2.1.2 The effect of CNN in style transfer

Feature extraction: The convolutional layer of the CNN network facilitates the efficient extraction of both the style and content features from the style and content images, allowing for further mining of the image's contents.

Style learning: By merging the extracted content features with the learned style features, the CNN network is able to transmit the style of the target images while also learning the feature representation of the incoming style images.

2.2 Image Style Migration Based on VGG Network

2.2.1 VGG-19 Network Model

Simonyan created the deep convolutional neural network model known as the VGG (Visual Geometry Group Network) in 2014. The VGG network performs well at extracting content and style elements from images in deep learning-based image style transfer research.

Three fully connected layers, five pooling layers, and sixteen convolutional layers make up the VGG-19 network. The pooling layer is 2×2 , the convolutional step and padding are unified to 1, and the 3×3 convolution kernel is used in all convolutional layers. The maximum pooling method is adopted, and each N convolutional layer and one pooling layer form a block. Each block of the input image passes through, the extracted feature image size gradually decreases and the retained content gradually decreases. Finally, without flattening the

block, a set of one-bit data is generated and passed into the last three layers of full-connection layer. The full-connection layer adopts Relu as the activation function and passes the processed data into softmax classifier for classification. Figure 1 shows the network structure of VGG-19.

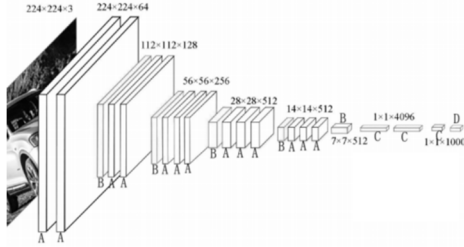


Figure 1: Structure of VGG-19 neural network model (Wu, 2021).

2.2.2 Image Migration Process Based on VGG Network

In order to extract the content features and style features, the target content photos and style pictures are first fed into the VGG network. Next, the loss function computes the style loss and content loss, and the overall picture loss error is examined. By continuously altering the network's parameters and the number of iterations, the overall error is decreased and the image style migration is eventually achieved. Figure 2 displays the style migration flow chart based on the VGG network.

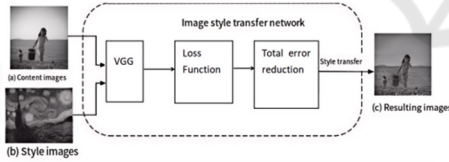


Figure 2: Flow chart of VGG network image style migration (Wu, 2021).

2.2.3 Comparison of Migration Effect Based on Convolutional Neural Network

Figure 3 illustrates how the convolutional neural network affects style migration. Figure 3 shows that whereas Resnet50 and NasnetMobile have low style migration effects, VGG-19 and InceptionV3 have good style migration effects.

Figure (c) achieves the texture transfer of the style picture well, but the image has a certain degree of distortion and detail loss.

Figure (d) not only has no transfer style, but also produces a lot of noise to blur the picture;

Figure (e) preserves the content of the picture well, but the texture is not strong, and finally figure (f) changes little compared to the content picture, only the color of the picture has changed a little.



Figure 3: Comparison of migration effects of different convolutional neural networks (Jin, 2021).

2.3 Research on Style Transfer Based on Improved VGG Network

In recent years, there have been many research efforts on technical improvements for style transfer based on VGG networks. Among them, Jin(Jin, 2021)et al. improved the VGG network by combining it with the Inception V3 network and adjusting the weights of the partial convolutional layers in both networks, which further improved the transfer effect.

2.3.1 Experimental Environment

In Windows10 64-bit system, the Tensorflow framework based on Python is used, and the pre-trained VGG19 and InceptionV3 networks with weights from the ImageNet dataset are used. The machine configuration is an Intel i7 - 9750H CPU, 16G of memory, and an NVIDIA GeForce GTX 1660Ti 6G graphics card.

2.3.2 Experimental Procedure and Result

The convolutional layer weight of the VGG19 part of the style migration network is set to w_{vgg} , and the convolutional layer weight of the InceptionV3 part is set to $w_{inception}$. Let the ratios of $w_{inception} : w_{vgg}$ be $10^0, 10^2, 10^4, 10^6$, and the number of iterations is 500 times. The experiment results show that, The migration network can adjust the effect of style migration by adjusting the ratio of $w_{inception} : w_{vgg}$. Finally, adjust the ratio of $w_{inception} : w_{vgg}$ to the order of 10^3 to achieve the best style transfer effect, and then get a better style transfer method.

2.4 Application of Style Transfer Technique Based on Convolutional Neural Network

The technique of style transfer based on convolutional neural network has many applications in practice. Xu(Xu,2024) et al. used the VGG neural network model and the maximum mean difference (MMD) to extract the features of content images and style images, set different weight ratios with TensorFlow2 as the frame, and used MMD to reduce the deviation between the target image and the training image, and then realized the image transfer technology of traditional Chinese painting style. However, Jiang(Jiang,2020) et al. realized the style extraction of content pictures and style pictures through the VGG network, and combined the content of images with many well-known oil painting styles to realize the style transfer of oil painting styles, and then obtained artworks of high perceived quality.

3 IMAGE STYLE TRANSFER BASED ON GENERATIVE ADVERSARIAL NETWORK

3.1 Structure and Principle of Generative Adversarial Network

The generative adversarial network was proposed in 2014 by Goodfellow et al. A GAN structure consists of a Generator (G) and a Discriminator (G). The generator's job is to produce more real images in order to fool the discriminator, while the discriminator's job is to determine if the sample image is generated or real. At the same time, the discriminator will constantly adjust the parameters to improve the accuracy of the judgment. The generator and the judge are updated iteratively and finally Nash equilibrium is obtained. The specific working process of generator and discriminator is as follows: generator obtains a set of random noise and outputs data G (z). Meanwhile, discriminator accepts data G (z) and real sample y from generator.

The discriminator D will give the probability P (G) and P (y) that the two are true, and the closer the probability value is to 1, the more it is considered to be true data, otherwise it is considered to be generated data. An ideally trained G should be such that P (G (z)) is always 1, and the output of an ideally trained D should satisfy the following formula:

$$D(x) = \begin{cases} 1, & x = y \\ 0, & x = G(z) \end{cases} \quad (1)$$

3.2 Image Style Transfer Process Based on CycleGAN

Based on GAN, Figure 4 depicts the network structure of CycleGAN, which has two generators and two discriminators.

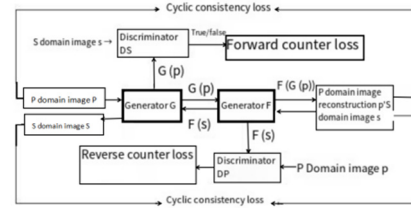


Figure 4: CycleGAN network structure (Li, 2023).

The following is CycleGAN's primary process: The generator G transforms the input picture of domain P into the forged picture of domain S. The generator F transforms the input picture of domain S into a forged picture of domain P. Reducing the disparity between the generated and original images is the generator's main objective. The real domain P image and the produced domain P image (G(p)) are distinguished by the discriminator DP, while the real domain S image and the generated domain S image (F(s)) are distinguished by the discriminator DS. Accurately identifying the input image's source is the discriminator's main objective. In order to maintain the consistency of image conversion, cyclic consistency loss is introduced. Generator G transforms the image of domain P into the forged image of domain S, which is subsequently transformed back into the reconstructed image p' of domain P by generator F. A cyclic consistency loss is computed as the difference between the original domain P picture p and the reconstructed image p'. The difference between the reconstructed image s' and the original image S of domain B is calculated. In a similar manner, generator F converts the image of domain S into the forged image S of domain s, and generator G converts the reconstructed image S' of domain s back to the reconstructed image s'. This completes the transformation of the CycleGAN model.

3.3 Comparison of Migration Effect Based on GAN Network

The result of style transfer by generating adversarial network is shown in the Figure 5. The transfer effects of Wasserstein Generative Adversarial Networks(WGAN) and CylceGAN are compared

from two perspectives of content retention and transfer effects. As shown in the figure below, the first is the content picture of style transfer, the second is the input style picture, where the first and second are the style transfer of sketch style, and the third and fourth are the style transfer of traditional ink painting style. In terms of content retention, the migrated images of the two networks have high content retention of the original images. From the migration effect, it can be seen that the migration effect of CycleGAN is better than that of WGAN. For example, it can be seen in line 3 and 4 that the color of the migrated images of CycleGAN is closer to the target style than that of WGAN.

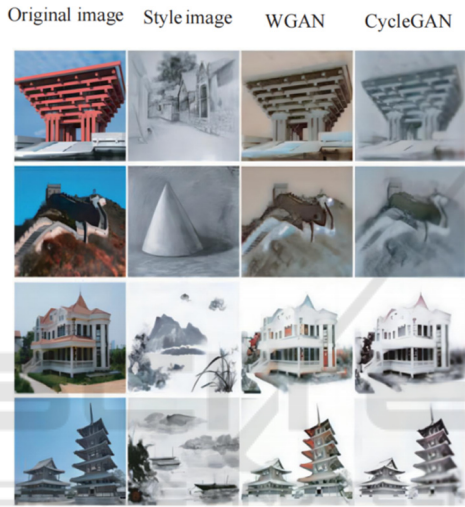


Figure 5: Comparison of migration effects of different generation adversarial networks (Shi, 2020).

4 INTRODUCTION TO STYLE TRANSFER BASED ON ADAIN ALGORITHM

In the AdaIN algorithm, the input image is first encoded by convolutional neural network to obtain the feature representation of different levels. Next, for each feature map, its mean and variance are calculated and standardized. To accomplish style transfer, the target style image's mean and variance are compared with the standardized feature map. Finally, the matched feature map is decoded to the converted image by a decoder. Figure 6 illustrates the AdaIN algorithm's style transfer procedure.

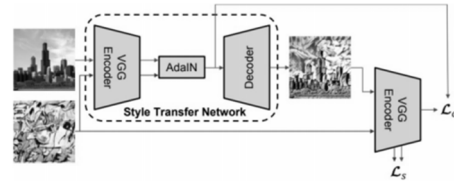


Figure 6: Flow chart of AdaIN algorithm style transfer (Wu, 2020).

5 FUTURE DEVELOPMENT PROSPECTS OF STYLE TRANSFER

The possible future development and research hotspots of style transfer are as follows:

1) Cross-modal style transfer: Cross-modal style transfer, such as music and video style transfer, can be investigated in the future in addition to image style transfer.

2) Fusion of multiple inputs: In addition to a single style image or text description, fusion of multiple input information, such as semantic segmentation, emotion analysis, etc., can provide a richer style transfer effect.

3) Real-time and interactive style transfer: The future development will pay more attention to real-time and interactive, enabling users to carry out instant style transfer, and real-time adjustment and feedback during the iterative process.

4) Style transfer in non-visual fields: The technology of style transfer can also be extended to non-visual fields, such as natural language processing, audio processing, etc., to achieve style transformation in more application scenarios.

5) Introducing the attention mechanism: To improve control over the style transfer, the attention mechanism can be added to the model to make it focus more on key portions of the image.

6) The combination of style transfer technology and Graph neural network(GNN): Graph neural network has attracted much attention for its excellent graph data modelling ability and sensitivity to complex relationships, while style transfer is a very creative area of picture processing. By combining the two, researchers can expect a range of innovations, including better capturing semantic and structural information in images through graph neural networks, which improves the fineness and accuracy of style transformations.

6 CONCLUSIONS

The common techniques in style transfer technology, including the approach based on deep neural network and the AdaIN algorithm-based approach, are compiled in this study. The basic principle of network and its role in style transfer are introduced. In addition, the transfer effects of different neural networks are compared and evaluated. Finally, the future development direction of style transfer technology is discussed, and the application scenarios and transfer methods of style transfer are prospected. From the perspective of application scenarios, style transfer technology can be applied to other forms of input, such as music and text. From the perspective of transfer method, the performance of style transfer technology can be further improved by introducing other techniques (such as attention mechanism) or combining with other neural networks (such as GNN). This paper provides some reference value for the future research of style transfer technology based on deep learning.

REFERENCES

- Gatys, L. A., Ecker, A. S., Bethge, M., 2016. Image Style Transfer Using Convolutional Neural Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2414-2423.
- Pan, S. J., Yang, Q., 2009. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22 (10): 1345-1359.
- Goodfellow, I., Pouget, J., Mirza, M., 2014. Generative adversarial nets. *Proceedings of the 27th International Conference on Neural Information Processing Systems*. 2672-2680
- Zhu, J., Park, T., Isola, P., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. *IEEE International Conference on Computer Vision*. 2242-2251.
- Li, Z. X., Qi, Y. L. , 2023. An improved CycleGAN image style transfer algorithm. *Journal of Beijing Institute of Graphic Communication*.
- Huang, X., Belongie, S., 2017. Arbitrary style transfer in real-time with adaptive instance normalization. *Proceedings of the IEEE International Conference on Computer Vision*. 1501-1510.
- Wu, Y., Song, J. G., 2020. Image Style Transfer Based on Improved AdaIN. *Software guide*. 19(09): 224-227.
- Wu, Z. Y., He, D., Li, Y. Q., 2021. Image style transfer based on VGG-19 neural network model. *Science and Technology & Innovation*.
- Jin, Z. G., Zhou, M. R., 2021. Research on Image Style Transfer Algorithm Based on Convolutional Neural Networks. *Journal of Hefei University (Comprehensive Edition)*. 38(02): 27-33.
- Xu, Z. J., Hu, Y. X., Lu, W. H., 2024. Style migration of Chinese painting based on VGG-19 and MMD convolutional neural network models. *Modern Computer*. 30(03):61-65+70.
- Jiang, M., Fan, Z. C., Sheng, R., Zhu, D., Duan, Y. S., Li, F. F., Sun, D., 2020. A oil Painting Style Migration Algorithm Based on Convolutional Neural Network. *Computer Knowledge and Technology*. 16.34(2020):6-9.
- Shi, Y. C., Zhu, L. J., 2020. Research on image style transfer based on GAN. *Electronic Technology & Software Engineering*. (16): 140-143.