

Artificial Intelligence for Art Creation with Image Style

Jinglun Shi *

College of Computer and Information Technology, Liaoning Normal University, Liaoning, China

* Corresponding author: 631402110122@mails.cqjtu.edu.cn

Abstract. Artificial intelligence (AI) has become a great success in the past decade. Powered by efficient hardware and neural network models, AI has made a wide change to almost all disciplines. While still in an early stage, AI has been used in the art field much more frequently than before, with various methods, e.g., generative adversarial networks. However, whether AI can replace human beings in art creation is still an unsolved question. Therefore, how to use AI technology to achieve artistic creation has become very meaningful, and has also been favored by many research groups. In this paper, we demonstrate the ability of AI for art creation with a case study of image style transfer. The result shows that AI is still not satisfactory and consumes too much computation. However, a new creative idea is proposed, and we will conduct lightweight research on this basis in the future. So as to realize the artistic creation of real artificial intelligence.

Keywords: Artificial Intelligence, Art Creation, Image Style Transfer.

1. Introduction

In 1956, John McCarthy, Marvin Lee Minsky, and other scientists met at Dartmouth College in the United States to discuss "how to use machines to simulate human intelligence" and first proposed "Artificial Intelligence (AI)". This concept marked the birth of artificial intelligence. But however, there has not been much progress since it was proposed. In the last few years, artificial intelligence has made breakthrough advances. The rapid development of the Internet has produced a large amount of data that need to be handled and processed, and the need for data processing technology has spurred the rapid development of artificial intelligence [1]. The second is the improvement of computation abilities based on cloud computing and computer servers. Last, the development of many advanced algorithms and models also helps, e.g., deep neural networks. In summary, data, computation, and algorithms contribute to the progress of artificial intelligence, which has achieved a great success in different domains, including finance [2-4], transportation [5-8], communication [9-12], energy [13, 14] and environmental sciences [15, 16]. Artificial intelligence technologies and products have now come into people's lives. To name some examples. GPS navigation systems often give voice commands to users, such as "at the next corner, turn left". Siri, the smart butler in cell phones, and Google's AlphaGo, which invokes reinforcement learning, have defeated the best human chess players. Self-driving technology benefits from advances in neural network technology, etc. There are also many applications of artificial intelligence in the field of art creation, such as in architectural design [17]. Because of the many factors to be considered, the hardness of the material, the choice of type, etc., it is difficult to meet the diversity of architectural design in traditional building design, but with artificial intelligence technology, to a certain extent, this problem can be solved. Building designers can take the building they want to build, adjust the right data, simulate the terrain and environment, create a 3D model and visualize the result of the building once it is built. In everyday life, the beauty camera is a reflection of the development of artificial intelligence. In simple terms, a Beauty camera is a camera that recognizes the information of a face in a picture and then adjusts the face in terms of shape, color, position, and proportion to be slim, smooth, and whiten the face. Some beauty cameras now can "make up" the face in the picture, enlarging the eyes, putting on eye make-up, blush, and lipstick.

In this paper, we explore the current maturity of artificial intelligence models in the field of art and how they are being used, using image style transfer as an example. Image style transfer is the blending of the style of the original image with that of another image. This blending is done so that the final

result is close to the original image in terms of content and close to the style of the other image in terms of style. The basic steps are as follows. The first step is to use the intermediate features of the convolutional layer of the neural network to restore the original image corresponding to this feature. The second step is to use the pretrained network as the recognizer for identifying features. The third step is to get a system that separates image content and image style and finally combines the content of the input image with the style of another image. The current image style transfer is not mature enough, and there are still some problems. For example, the effect of style migration will be related to the number of training sessions. It is difficult to get the ideal image for different styles. Based on the results of the numerical experiments, we found that AI technology is not powerful for artistic creation. AI appears to be able to simulate human thinking and thus make judgments, but it is still unable to reach the same level of activity and thought as the human mind and is also a certain barrier to use, such as high computer performance requirements.

2. Related Work

We found in the article that there have been some similar studies where they applied AI techniques to various types of image style transfer, including cartoon style transfer and ink and animation movie styles. The basic methods are also based on convolutional neural networks and generative adversarial networks and have achieved good results. Follow-up studies are also exploring newer models to achieve a more realistic image generation effect or to reduce the overhead of the process. We provide a brief overview of the relevant papers in this section; a more detailed discussion can be found in the related review [18].

2.1. Style Image Transfer Based on Pretrained Models

Zhao completed the experiment through two convolutional neural network models VGG-16 and VGG-19 pretrained by ImageNet and used the pretrained model to extract style features from style images and content images [19]. Finally, the final target image is continuously optimized by constructing a loss function. The main process of the experiment revolves around building a loss function, which is a weighted combination of the content loss function and the style loss function. Get the total loss function to achieve the purpose that the style is sufficiently similar to the original image. Both models reflect good results in the final results. However, the loss value of VGG-16 is lower, so the conclusion is that the VGG-16 model is better. Based on the pretrained convolutional neural network in ImageNet, the training time is saved to promote the application more efficiently and promote the practical application and development of computer vision technology in different fields.

2.2. Cartoon Style Transfer

This article summarizes and discusses the existing mainstream methods of animation style transfer. Until now, anime style transfer was mainly performed based on GAN networks. The authors classify the types of animation style transfer into three categories: landscape animation style transfer, portrait animation style transfer, and video frame animation style transfer. The transfer of scenery is the CartonGAN network proposed by Chen et al. [20] in 2018. It mainly uses the general characteristics of animation images, such as animation images with clear and simple textures and smooth color blocks in the area. In 2020, Chen et al. [21] proposed a network parameter quantity, which is a more lightweight AnimeGAN than CartoonGAN, which effectively reduces the problem of local color artifacts in the resulting images of CartoonGAN. The migration of portraits is very different from the migration of landscapes. The training data of real faces and cartoon faces are unpaired, and their appearances are quite different and have no correspondence, so it is difficult to capture faces with the same characteristics as cartoon faces. Therefore, Wu et al. [22] proposed a landmark-assisted CycleGAN, which introduced landmark consistent loss on the basis of the original CycleGAN to solve the problem of structural mismatch between unpaired training data and also designed global

and local discrimination. The tool optimizes the entire portrait and local features, such as eyes, nose, mouth, and ears, and achieves a more refined effect. Video frame migration is intended to provide anime creators with an automatic tool for converting videos to manga.

On the problem of converting real pictures into cartoons, Gao [23] proposed an improved generative adversarial network based on the existing VGG-generated adversarial network. In this article, a generative adversarial network based on the residual network is proposed. It is divided into two parts: the generator model and the discriminator model. The generator model first extracts the features of the picture, then uses it to reduce the dimension, and finally reconstructs the Comic style image and output image. The discriminator model encodes and classifies basic local features and then connects them to output an image. In terms of the loss function, Gao believes that the loss function in the CartoonGAN network proposed by Chen can theoretically generate enough fake pictures, but in actual training, there are many problems. By studying the ideas in the article "From GAN to WGAN", he also made improvements to the loss function. Finally, Gao conducted a comparative experiment between his Re-GAN model and the original model. In the experiment, the picture of Makoto Shinkai's "Me and My Cat" was selected. In terms of the color tone and picture outline, the two improved models were compared, and it was found that the cartoon effect of Re-GAN was better than that of CartoonGAN. Therefore, the Re-GAN model can convert pictures in real life into pictures with a better cartoonist style.

Wang [24] believes that most style transfer models are relatively broad and do not have specific models. To achieve the desired effect in different styles. The author has made improvements in the following aspects. 1. Introduce a multilayer perceptron module (Multi-Layer Perceptron, MLP) into the generator of the GAN network and add noise input. 2. Modify the network structure so that content features can be well transformed into style features. 3. Divide the loss function into two parts and calculate the loss separately for the network and the content. 4. Increase the picture data set, expand the animation picture and real picture set, and improve the training effect.

Mao [25] proposed a new Chinese-style animation generative adversarial network model CCGAN (Chinese Cartoon GAN) for video-oriented style transfer. This model can effectively reduce the constraints of painting skills in Chinese-style animation creation and improve the efficiency of artistic creation. Its main principle is divided into two steps: first, build a lightweight GAN model, then, for the content, style, color, etc., of the image, the grayscale style loss and color reconstruction loss are constructed in the generator and the discriminator. Build a grayscale adversarial loss function and an edge-promoting adversarial loss function. Finally, the Adam algorithm is used to minimize the above loss function, realize the image style transfer, and combine the pictures into a video. The technical novelty and advantage of this model are that it breaks through the existing GAN's limitation of texture and boundary line reconstruction so that the reconstructed generated image maintains the content characteristics of the original image in structure and integrates the style image. In addition, because the unmatched source data are used for training, there is no need to force the content of the original image and the sample image to remain relevant, which expands the scope of application of the model and makes it more widely used

2.3. Ink Wash Style Transfer

Huang [26] proposed a new stylization method for ink animation films based on generative adversarial networks, which realizes the transfer of ink paintings drawn by artists from one picture to the whole video. According to the same training strategy based on the patch proposed by Texler, the author does not pretrain a large number of data sets, but the problem of video flickering occurs. To solve the problem, the video author performs a depth check on the video after the style transfer. In the final comparison experiment, the author selects a certain shot of the animated movie "Big Fish & Begonia" as the experimental data because the Texler method does not perform video deflickering, and the generated results are all noisy. However, the pictures generated by the method in this paper have clear effects, have good stability on moving objects, and can remove flickers well. This also verifies the effect of the method in this paper on the style transfer of ink animation.

2.4. Chinese Character Style Transfer

The method of Chinese character style transfer [27] is realized by combining a generative adversarial network and a recurrent generative adversarial network model. Combined with the experimental results, it is found that there are model collapse problems and generation failure problems during the training process. Because the LBP algorithm can effectively extract the local features of the image (characteristics of Chinese characters), it has the advantages of rotation invariance, illumination invariance and grayscale invariance, so the author combines the LBP algorithm with the recurrent generative adversarial network to enhance Chinese characters. Migration performance. In the experiment, we selected printed font with large differences in writing style and the classic calligraphy italic style of handwriting as the data set. The Classic Song Dynasty font was used as the source font, and the Classic Handwritten Calligraphy Kai font was used as the target font. The result proves that the method proposed in this paper achieves the basic requirements of invariant font and outline in Chinese font style transfer, and the generated font image is realistic.

3. Case Study Of Image Style Transfer

In this section, image style transfer is used as a case study to demonstrate the ability of artificial intelligence for art creation.

3.1. Image Style Transfer Basics

Image style transfer refers to the fusion of the original image and the style image, as shown in Figure 1. This fusion makes the final result image close to the original image in content and the style image in style. Before using deep learning, image style transfer is mainly based on statistical methods, which extract the patterns manually. These traditional methods have many limitations. Each model can only be used to generate a specific style without the generalization ability. The styles that can be modeled by these statistical methods are also very limited, making them less preferable compared with popular image editing tools.



Figure 1. The basic flowchart of image style transfer.

Neural style transfer is proposed as a new idea with the introduction of deep neural networks to the application of image style transfer [28, 29]. The basic idea is as follows. The first step is to use the intermediate features of the convolutional layer of the neural network to restore the original image corresponding to this feature. An original image is selected, and the features of each convolutional layer are obtained through model training. Then, the feature is restored to mimic the original image according to the obtained features. The second step is to use the pretrained network as the recognizer for identifying features, then calculate the correlation of those different local features, turn it into a statistical model, and make a method that can generate textures without manual modeling. The third step is to get a system that separates image content and image style and finally combines the content of the input image with the style of another image. After the introduction of neural style transfer, many follow-up studies have been proposed, with two main objectives. The first objective is to lower the computation requirement of the neural networks. While different styles can be obtained by training different neural networks, the computational burden increases considerably with this process. Many efficient neural network structures have been proposed for similar purposes, e.g., image classification and object detection, and some of them have been successfully applied to neural style

transfer applications. Another idea is network pruning when a sophisticated model is trained first and then a simplified version is created by pruning some unimportant neurons for a faster and more efficient implementation. Another objective is to improve the fidelity of the output image. While some models can be more efficient, there is a tradeoff between the fidelity and the computation cost. Without enough training epochs, the output image can be very easily identified as a fake image. In practice, a cloud-based approach is often used to support relevant applications. For example, a picture is captured by the smartphone and uploaded to the central server, in which the image style transfer process actually happens. Then, the output image is sent back to the smartphone for display.

3.2. Model Description

In this study, a specific model named AdaIN [30] improves the original neural style transfer method in multiple aspects. The basic workflow of AdaIN is shown in Figure 2.

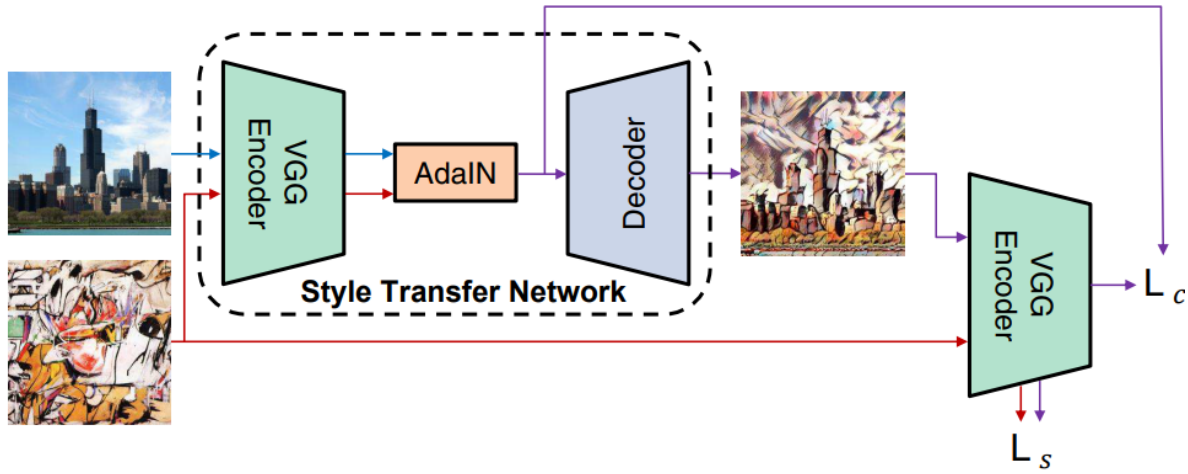


Figure 2. The basic workflow of AdaIN [30].

The main improvement of AdaIN is the use of Adaptive Instance Normalization. In previous relevant studies, batch normalization was widely used for neural style transfer, with the aim of solving the problem of internal covariate shift. As indicated by its name, the batch normalization operation is conducted on a batch of data samples, making the mean zero and the standard deviation one. Afterward, it is found that instance normalization is both more efficient and effective than batch normalization in the task of neural style transfer. The instance normalization operation is conducted for each data sample, which is different from batch normalization. AdaIN further purposes as an improved version of instance normalization, as shown in the following formulation:

$$\text{AdaIN}(x, y) = \sigma(y) \left(\frac{x - \mu(x)}{\sigma(x)} \right) + \mu(y) \quad (1)$$

where x is the input image, y is the style image, $\mu(x)$ and $\mu(y)$ are mean values, and $\sigma(x)$ and $\sigma(y)$ are variances. Based on this function, the whole process of AdaIN is explained as follows, which consists of three parts: the VGG encoder, the AdaIN layer, and the encoder. VGG [31] is a deep convolutional neural network designed for image recognition that introduces the idea of using blocks instead of layers for neural network structure design. With the pretrained VGG model, the features can be extracted from both the input and style images, which can be denoted as $f(c)$ and $f(s)$. Then, the AdaIN layer is used to extract the feature mapping as $t = \text{AdaIN}(f(c), f(s))$. The decoder is used to generate the desired output image as the final output $T(c, s) = g(t)$. The tradeoff between the content loss and the style loss is achieved by a weighted loss function.

3.3. Experiment and Results

The whole experiment is conducted with Python 3 and PyTorch. An NVIDIA GeForce RTX 2070 graphics card is used as the GPU to train the deep learning model. An open-source implementation

of AdaIN is used in this study, which is publicly available: <https://github.com/naoto0804/pytorch-AdaIN>. The input content image is shown in Figure 3, and the input style image is shown in Figure 4. The input content image is a picture captured by a digital camera, while the input style image is a cartoon image, which was originally drawn by a cartoonist. The purpose of this case study is to validate the AI ability for creating cartoon images without the burden of human drawings.



Figure 3. Content image.



Figure 4. Style image.

The generated output images after different epochs are shown in Figure 4. After epoch 1000, the generated output image is already similar to a cartoon image. The longer training epoch may be unnecessary. This case study demonstrates the ability of artificial intelligence for art creation, with a case study of cartoon image style transfer.

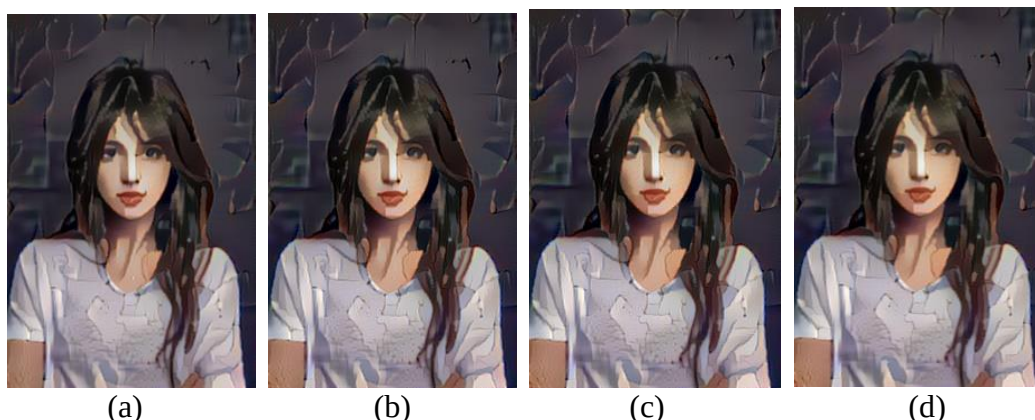


Figure 5. The results in different epochs. (a) 1000; (b) 2000; (c) 3000; (d) 4000

However, many problems still exist. The first problem is that during the experiments, only those style images with an obvious pattern can be well used. If the style image is a normal picture without any obvious pattern, the output image would be similar to the original content image, since no style can be learned and transferred. The second problem is the high computational cost. For training 1000 epochs, at least three hours are consumed even with GPU acceleration. If the AI ability is to be used in the art industry with a huge demand, the computation requirement would be prohibitory if more efficient solutions cannot be found.

4. Conclusion

In this paper, we use image style transfer as a case study to explore the maturity and use of AI techniques in the field of art. While AI manages to achieve a satisfactory output, the computation cost and the unstable performance cannot be ignored. There is still a long way before AI replaces human beings for art creation.

References

- [1] Xu Z, Tang N, Xu C, et al. Data science: connotation, methods, technologies, and development [J]. Data Science and Management, 2021, 1 (1): 32 - 37.
- [2] Jiang W. Applications of deep learning in stock market prediction: recent progress [J]. Expert Systems with Applications, 2021, 184: 115537.
- [3] Li Z, Jiang W. Evaluation of Machine Learning Techniques for Stock Market Movement Prediction [C]//2020 Management Science Informatization and Economic Innovation Development Conference (MSIED). IEEE, 2020: 248 - 253.
- [4] Kumbure M M, Lohrmann C, Luukka P, et al. Machine learning techniques and data for stock market forecasting: a literature review [J]. Expert Systems with Applications, 2022: 116659.
- [5] Jiang W, Luo J. Big data for traffic estimation and prediction: a survey of data and tools [J]. Applied System Innovation, 2022, 5 (1): 23.
- [6] Jiang W, Zhang L. Geospatial data to images: A deep-learning framework for traffic forecasting [J]. Tsinghua Science and Technology, 2018, 24 (1): 52 - 64.
- [7] Jiang W, Luo J. Graph neural network for traffic forecasting: A survey [J]. Expert Systems with Applications, 2022: 117921.
- [8] Jiang W. Bike sharing usage prediction with deep learning: a survey [J]. Neural Computing and Applications, 2022: 1 - 17.
- [9] Tan J, Guan W. Resource allocation of fog radio access network based on deep reinforcement learning [J]. Engineering Reports, 2022, 4 (5): e12497.
- [10] Jiang W. Graph-based deep learning for communication networks: A survey [J]. Computer Communications, 2021.
- [11] Pampapathi B M, Guptha N, Hema M S. Towards an effective deep learning-based intrusion detection system in the internet of things [J]. Telematics and Informatics Reports, 2022, 7: 100009.
- [12] Jiang W. Cellular traffic prediction with machine learning: A survey [J]. Expert Systems with Applications, 2022: 117163.
- [13] Zhao E, Sun S, Wang S. New developments in wind energy forecasting with artificial intelligence and big data: A scient metric insight [J]. Data Science and Management, 2022.
- [14] Jiang W. Deep learning based short-term load forecasting incorporating calendar and weather information [J]. Internet Technology Letters, 2022, 5 (4): e383.
- [15] Fan G, Deng Z, Ye Q, et al. Machine learning-based prediction models for patient's no-show in online outpatient appointments [J]. Data Science and Management, 2021, 2: 45 - 52.
- [16] Jiang W, Luo J. An evaluation of machine learning and deep learning models for drought prediction using weather data [J]. Journal of Intelligent & Fuzzy Systems, 2021 (Preprint): 1 - 16.
- [17] Zhao M, Zhang Y. GAN-based deep neural networks for graph representation learning [J]. Engineering Reports, 2022: e12517.
- [18] Tang R, Liu Q, Tan H. Review of Neural Style Transfer Models. Computer Engineering and Applications. 2021, 57 (19): 32 - 43.
- [19] Zhao W, Shi S, Zhou C. Image Style Transfer Based on ImageNet Pre-Trained Convolutional Neural Network. Journal of Chengdu University (Natural Science Edition). 2021, 40 (4): 367 - 373.
- [20] Chen Y, Lai Y K, Liu Y J. Cartoongan: Generative adversarial networks for photo cartoonization [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 9465 - 9474.
- [21] Chen J, Liu G, Chen X. AnimeGAN: a novel lightweight GAN for photo animation [C]//International Symposium on Intelligence Computation and Applications. Springer, Singapore, 2019: 242 - 256.
- [22] Wu R, Gu X, Tao X, et al. Landmark assisted cyclegan for cartoon face generation [J]. arXiv preprint arXiv: 1907.01424, 2019.
- [23] Gao X, Ding Z, Li G, et al. Image Generation based on Improved Generation Adversarial Network Style Transfer. Journal of Yunnan Minzu University (Natural Sciences Edition), 2022.
- [24] Wang Y, Zhao L, Li Y. Image Animation Stylization based on Generative Adversarial Network. Computer Engineering and Applications, 2022.

- [25] Mao W, Wu G, Wu C, et al. Animation Video Generation Model based on Chinese Impressionistic Style Transfer. *Journal of Computer Applications*, 2022.
- [26] Huang D, Li N, Yu L, et al. A discussion on ink animation stylization method based on generative adversarial network. *Advanced Motion Picture Technology*, 2022, 3: 21 - 28.
- [27] Niu Y. Chinese Character Style Transfer Method Based on Generative Adversarial Network. *China Computer & Communication*, 2022, 5: 69 - 71.
- [28] Gatys L, Ecker A S, Bethge M. Texture synthesis using convolutional neural networks [J]. *Advances in neural information processing systems*, 2015, 28.
- [29] Gatys L A, Ecker A S, Bethge M. Image style transfer using convolutional neural networks [C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016: 2414 - 2423.
- [30] Huang X, Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization [C]//*Proceedings of the IEEE international conference on computer vision*. 2017: 1501 - 1510.
- [31] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. *arXiv preprint arXiv: 1409.1556*, 2014.