

A Survey on Image Style Transfer Based on Deep Learning

Mingyu Huang *

School of Software, Northwestern Polytechnical University, Xi'an, Shaanxi, China

* Corresponding author Email: huangmingyu@mail.nwpu.edu.cn

Abstract: With the rapid development of deep learning technology, style transfer has achieved significant breakthroughs at both algorithmic and application levels, greatly promoting innovative interactions between content and style. This paper comprehensively reviews the fundamental concepts, classifications, and widespread applications of style transfer within neural networks, with a particular focus on the principles, diverse variants, and synthesis techniques of neural network-based style transfer. Additionally, the article explores image style transfer as a research hotspot in the field of artificial intelligence and computer vision, highlighting that traditional methods primarily rely on physical and texture synthesis techniques, which often result in coarse outcomes and poor robustness. Through comparative analysis, this paper outlines the development trajectory and latest research advances in image style transfer and proposes future research directions.

Keywords: Image Style Transfer; Deep Learning; Neural Networks; Model Algorithms; Computer Vision.

1. Introduction

In today's digital age, advancements in image processing technologies have opened new avenues for artistic creation and information communication. Among these, Image Style Transfer has emerged as a revolutionary technique that, leveraging the powerful capabilities of deep learning, enables the transformation of artistic images between different styles. The core task of image style transfer is to transfer the style, texture, and other features of one image onto another. This is achieved through deep neural networks that extract and reconfigure the features between content images and style images, thereby creating a new image that retains the original content structure while incorporating the desired style characteristics. As shown in Fig.1, this process involves blending the content features of the content image in Fig.1 with the style features of the style image in Fig.2 to generate the new effect image in Fig.3 [1]. The complexity and diversity of images make achieving ideal style transfer effects highly challenging. Researchers from various fields, including mathematics, physics, and computer science, continue to explore and refine the theories and methods of image style transfer, proposing new algorithms and models to enhance the quality of style transfer.

The rise of deep learning technologies, with their exceptional expressiveness and learning capabilities, has provided a new approach to image style transfer and accelerated its development within the field of computer vision. This paper aims to review the trajectory of image style transfer techniques, discuss the working principles and the advantages and disadvantages of different style transfer methods, and analyze current challenges and potential future research directions. By comprehensively reviewing related research in this field, this paper seeks to further advance the development of image style transfer technology, providing theoretical and practical references for achieving more accurate and efficient style transfer results.



Fig.1 Content Image

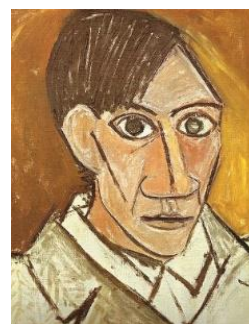


Fig.2 Style Image



Fig.3 Result Image

2. Early Image Style Transfer Methods

The origins of image style transfer can be traced back to the mid-1990s, marked by the development of Non-Photorealistic Rendering (NPR)[2], which represented an initial exploration in this field. NPR, one of the earliest research topics in computer graphics, contrasts with Photorealistic Rendering, which aims for realistic effects. Instead, NPR focuses on mimicking artistic painting styles, creating visual effects similar to hand-drawn illustrations. Common NPR techniques include cartoon rendering, oil painting rendering, pixel art, pencil sketching, crayon drawing, and ink wash painting, among others. These algorithms primarily focus on simulating specific brushstroke textures rather than comprehensive painting style transfers. However, these methods are often limited to single styles and struggle to adapt to other styles or higher-level semantic style transfers.

In the early 2000s, researchers further advanced NPR algorithms and developed texture transfer techniques. This technology moves beyond simply mimicking brushstrokes in paintings and instead focuses on extracting more abstract texture features from images. These texture features, which represent key information of the images, are separated from the original images using methods such as convolution, downsampling, and Harris operators. After acquiring these texture features, researchers employed various techniques to integrate these features into target images. For instance, Efros and Freeman's team developed a mapping table containing texture features of the target image to guide the texture transfer process [3]. Simultaneously, Hertzmann and colleagues implemented this process through image analogies [4].

However, these algorithms mainly employ non-parametric approaches for style transfer. Although this method represents some progress over the initial NPR algorithms, it remains confined to relatively shallow feature transformations. As a result, there is a pressing need in the market for a more general and efficient style transfer algorithm.

3. Neural Style Transfer Algorithms

With the development and widespread application of deep learning technologies, neural network-based image style transfer has become a significant direction in style transfer research. This field encompasses various methods, each with specific applications, advantages, and limitations.

3.1. Convolutional Neural Networks (CNNs)

Around 2015, with the rapid rise of Convolutional Neural Networks (CNNs) and their substantial advancements in fields such as image recognition and natural language processing, image style transfer techniques also began to follow this trend. This period marked the increasing penetration of deep learning technologies into broader application areas, especially within the arts and creative industries, where the application of CNNs opened up new possibilities and fueled a new wave of integration between visual arts and technology. The technological advancements of this era not only improved the performance of algorithms but also significantly expanded their scope of application, making style transfer more practical and creatively rich beyond mere academic exploration.

Convolutional Neural Networks (CNNs) are a type of deep learning model widely used in image style transfer tasks. A CNN consists of multiple layers, including convolutional layers, activation function layers, pooling layers, and fully connected layers, which work together to extract image features progressively from local to global scales, enabling the model to learn high-level representations of image data.

In style transfer, CNNs play a crucial role due to their exceptional feature extraction and learning capabilities. By stacking multiple convolutional layers, CNNs can extract image features at various levels, ranging from low-level to high-level, which are essential for capturing the content and style of images. Additionally, CNNs can learn the feature representations of style images and effectively combine these with the features of content images, thus achieving style transfer. This process allows style transfer algorithms to blend the input content and style images, creating a new composite image with the desired style.

In 2016, Gatys and his team introduced a CNN-based image style transfer algorithm[5], building upon prior research in the field and marking a significant technological breakthrough. They utilized the well-known VGG convolutional neural network model to extract style features and optimized the generated image using gradient descent. Specifically, the Gatys team used the VGG model to extract features from both style and target (content) images at different hierarchical levels, with higher levels representing more abstract semantic information. They defined separate loss functions for the style and content images and calculated the total loss function as follows:

$$L_{total} = \alpha L_{content} + \beta L_{style} \quad (1)$$



Fig.4 Results of Neural Style Transfer 1

In the style transfer process, the coefficients α and β play a crucial role, as they respectively control the extent to which

the resulting image is stylized and the degree to which the original structure is preserved. At the beginning of the training, the resulting image is initialized by setting each pixel value as a parameter. The VGG network is then used to extract the image features, and the loss function is calculated according to the aforementioned method. Using gradient descent and backward algorithms, the network undergoes forward and backward iterations. After multiple iterations, the resulting image gradually converges, ultimately achieving the desired style transfer effect, as shown in Fig.4 [6].

3.2. Cycle-Consistent Generative Adversarial Networks (CycleGAN)

In 2014, Goodfellow [7] proposed the Generative Adversarial Network (GAN). Currently, GANs have been widely researched and applied in the field of image processing in computer vision, such as image classification, image recognition, object tracking, and semantic segmentation, achieving superior results compared to traditional convolutional neural networks. Zhu et al. [8] introduced the Cycle-Consistent Generative Adversarial Network (CycleGAN), specifically designed for style transfer without the need for paired images. Traditional style transfer techniques often rely on paired training data, where each style must be matched with specific content. However, in practice, obtaining such paired data is both challenging and time-consuming. CycleGAN addresses this problem through an innovative network structure, providing an effective solution.

CycleGAN utilizes an adversarial training framework with two pairs of generators and discriminators, achieving style transformation from images in the source domain to the target domain. The CycleGAN architecture includes two generators, F and G, which act as inverse mappings of each other. An inverse mapping means that one generator, G, converts images from source domain X into fake images of target domain Y, while the other generator, F, converts these fake images back into the original source domain X. In other words, the transformation relationships of F and G can be expressed as $F(G(x)) \approx_x$ and $G(F(y)) \approx_y$, where x represents a sample from domain X, and y represents a sample from domain Y. Additionally, the adversarial losses DX and DY are associated with the adversarial discriminators of source domain X and target domain Y, respectively. The objective function maps the original image to the target domain:

$$L_{adv}(F_{X \rightarrow Y}, G_{Y \rightarrow X}, D_X, D_Y) = E_x[\log D_Y(y)] + E_x[\log(1 - D_Y(F_{X \rightarrow Y}(x)))] + E_y[\log D_X(x)] + E_y[\log(1 - D_X(G_{Y \rightarrow X}(y)))] \quad (2)$$

The cycle-consistency loss is designed based on a dual learning mechanism. Although the adversarial loss guides generators F and G to learn the mapping relationships, F and G are essentially random functions that cannot guarantee consistently expected outputs. To address this issue, the cycle-consistency loss is introduced. This loss function ensures that images transformed through the cycle can be restored to their original states, thereby further constraining the possible space of the mapping functions to achieve better transformation results:

$$L_{ccl}(F_{X \rightarrow Y}, G_{Y \rightarrow X}) = E_x(\|G_{Y \rightarrow X}(F_{X \rightarrow Y}(x)) - x\|) + E_y(\|F_{X \rightarrow Y}(G_{Y \rightarrow X}(y)) - y\|) \quad (3)$$

The complete objective function of CycleGAN, L_{total} , is derived by combining Equations (2) and (3), as shown in Equation (4):

$$L_{total}(F_{X \rightarrow Y}, G_{Y \rightarrow X}, D_X, D_Y) = L_{adv}(F_{X \rightarrow Y}, G_{Y \rightarrow X}, D_X, D_Y) + \lambda_{ccl} L_{ccl}(F_{X \rightarrow Y}, G_{Y \rightarrow X}) \quad (4)$$

where λ is the weighting coefficient that controls the contribution of the cycle-consistency loss LCCL in the total objective function.

3.3. Diffusion Models (DDPM)

Diffusion Models, specifically Denoising Diffusion Probabilistic Models (DDPM) [9], are a class of probabilistic deep learning models that generate high-quality images through a process of gradually adding and then removing noise. The core concept of DDPM involves defining a forward diffusion process and a reverse generation process. In the forward process, given an original image x_0 , the model sequentially adds noise to generate a series of noisy images $x_1, x_2, \dots, x_T$, with each step described by:

$$q(x_t | x_{t-1}) = N(x_t; \sqrt{1 - \beta_t} x_{t-1}, \beta_t I) \quad (5)$$

where β_t controls the intensity of noise at each step. The ultimate goal is to learn the reverse diffusion process, which starts from a noisy image x_t and gradually removes the noise to regenerate an image x_0 that matches the original image distribution. The reverse process is defined as:

$$p_\theta(x_t | x_{t-1}) = N(x_t; \mu_\theta(x_t, t); \sum_\theta(x_t, t)) \quad (6)$$

where $\mu_\theta(x_t, t)$ and $\sum_\theta(x_t, t)$ are the mean and variance learned by the model, defining the denoising distribution at each step. By maximizing the likelihood, the model incrementally learns to restore high-noise images to low-noise images, eventually generating samples that align with the training data distribution.

When using diffusion models for style transfer, DDPM's generative mechanism is ingeniously applied to the fusion of style and content [10]. To achieve style transfer, an additional

style embedding vector ϕ_s , representing the target style image's features, is introduced during the diffusion and denoising process. This embedding vector works in conjunction with the content image features ϕ_c during the diffusion process, allowing the target style to be gradually applied to the image while denoising.

Specifically, during the generation process, the model starts with a high-noise image x_T and iteratively applies denoising steps as follows:

$$x_{t-1} = \mu_\theta(x_t, t, \phi_s) + \sum_\theta(x_t, t) \epsilon \quad (7)$$

where $\mu_\theta(x_t, t, \phi_s)$ represents the predicted mean based on the current noisy image x_t , time step t , and style embedding ϕ_s , while $\sum_\theta(x_t, t)$ is the predicted variance. In this setup, the model gradually transmits style features from high levels to lower levels during the denoising process, resulting in a final generated image that retains the content structure while exhibiting the visual effect of the target style.

Moreover, because diffusion models operate on multiple scales, they can handle the fusion of style and content both globally and locally. At each time step t , the model can adjust the balance between style and content, allowing DDPM to achieve complex and natural style transfer. This approach

significantly enhances the quality and consistency of the generated images, especially in maintaining content integrity and accurately conveying style characteristics, outperforming traditional methods.

4. Image Style Transfer Datasets

In the field of image style transfer research, selecting the

right datasets is crucial for training models and evaluating their performance. Different methods and models rely on specific content and style images to achieve style transfer tasks through diverse scenes and artistic styles. The main methods discussed in this paper are CNN, CycleGAN, and Diffusion Models, each having its own focus and application scenarios when using datasets.

Table 1. Datasets and Application Scenarios for Style Transfer Algorithms

Algorithm Type	Representative Works	Dataset	
		Content	Application Scenarios
CNN	References [11]	MS-COCO	Content Images (Complex Scenes)
	References [12]	ImageNet	Content Images (Multiple Categories)
	References [13]	WikiArt	Style Images (Artistic Styles)
	References [14]	Oxford 102 Flowers	Content Images (Natural Styles)
CycleGAN	References [15]	Cityscapes	Urban Style Transfer
	References [16]	Maps	Map and Satellite Image Transfer
	References [17]	Horse2Zebra	Animal Style Transfer
	References [17]	Monet2Photo	Artistic Style Transfer
	References [18]	Yosemite Summer to Winter	Seasonal Style Transfer
Diffusion Model	References[10]	LSUN Church	Architectural Style Transfer
	References[19]	WikiArt	Artistic Style Transfer
	References[20]	CelebA-HQ	Facial Style Transfer
	References[9]	FFHQ (Flickr-Faces-HQ)	High-Quality Facial Generation

For CNN-based style transfer, selecting the appropriate dataset is key to achieving high performance. MS-COCO is one of the important datasets widely used for training content images. It covers a wide variety of everyday scenes and objects, providing complex and realistic visual content that helps CNNs better capture image structures when handling diverse scenarios. On the other hand, ImageNet, as a large-scale image database, contains millions of images spanning a wide range of categories from animals to objects. Due to its extensive categories and high-quality annotations, ImageNet is often used for pre-training CNN models, allowing them to extract multi-level features from images, thereby maintaining the integrity and accuracy of content during style transfer. In contrast, the WikiArt dataset focuses on collecting artworks from various artistic styles and historical periods. In style transfer tasks, WikiArt provides diverse style learning materials, enabling CNNs to learn and transfer distinctive features of artistic styles, generating images with an artistic feel. The Oxford 102 Flowers dataset provides high-quality images of flowers from nature, rich in detail and vibrant in color, making it suitable as a source of content images for natural style transfer.

CycleGAN has achieved remarkable results in unsupervised image style transfer and is widely used in various style transfer tasks. Common datasets include Cityscapes, Maps, Horse2Zebra, Monet2Photo, and Yosemite Summer to Winter. The Cityscapes dataset is mainly used for urban scene style transfer, allowing CycleGAN to effectively migrate visual styles in urban environments. The Maps dataset is used to achieve style transfer between maps and satellite images, helping the model learn to transform two-dimensional maps into realistic satellite images. The Horse2Zebra dataset demonstrates style transfer between animals, where the model can convert horse images into zebra styles and vice versa. The Monet2Photo dataset focuses on transferring artistic styles to real images, allowing CycleGAN to transform Monet-style paintings into realistic photo effects. The Yosemite Summer to Winter dataset is used for seasonal style transfer, enabling CycleGAN to learn and achieve

transformations from summer scenes to winter scenes.

In recent years, Diffusion Models have made significant progress in the fields of image generation and style transfer, with commonly used datasets including LSUN Church, WikiArt, CelebA-HQ, and FFHQ (Flickr-Faces-HQ). The LSUN Church dataset is primarily used for architectural style generation and transfer, allowing models to achieve style conversion in complex architectural structures. WikiArt continues to serve as an important dataset for artistic style transfer, covering various art genres and providing a rich source of artistic style materials for the models. The CelebA-HQ and FFHQ datasets focus on high-quality facial image generation and style transfer. Their detailed image quality enables Diffusion Models to generate realistic and stylistically rich facial images at high resolutions.

In the study of image style transfer, different deep learning methods rely on datasets with unique characteristics, which provide abundant training materials and evaluation benchmarks for the models. CNN methods achieve effective integration of content and style through diverse scenes and artworks; CycleGAN demonstrates strong adaptability in various complex style transformation tasks thanks to its unsupervised learning advantage; and the introduction of Diffusion Models further improves the precision and quality of style transfer, especially in generating high-resolution images. As datasets diversify and model technologies advance, the field of image style transfer will continue to expand its application scope and showcase its innovative potential in broader scenarios.

5. Conclusion

This paper provides a comprehensive review of image style transfer techniques based on deep learning, from early methods to the latest neural style transfer algorithms, along with an analysis of relevant datasets. As a powerful artistic and technical tool, style transfer has demonstrated its wide-ranging applications and innovative potential in the field of image processing.

We first explored early image style transfer methods, which,

although based on traditional techniques, laid the groundwork for the development of style transfer. Next, we detailed the evolution of neural style transfer algorithms, which utilize deep neural networks to effectively separate and merge style and content, significantly enhancing the quality and efficiency of style transfer. CNNs, with their robust feature extraction capabilities, are well-suited for integrating content and style across diverse scenes and artistic works, making them ideal for a wide variety of image datasets. CycleGAN, leveraging its unsupervised learning nature, excels in style transfer tasks without paired data, particularly in urban landscapes, animal transformation, and artistic styles. As a recent emerging method, Diffusion Models have shown immense potential in generating high-resolution images, particularly in complex architectural style generation and high-quality facial style transfer. Additionally, the selection of datasets and the application of evaluation metrics are crucial for advancing research in this field, and this paper provides an in-depth analysis of these aspects.

Despite the significant achievements in style transfer technology, many challenges remain, such as balancing content and style and improving computational efficiency for large-scale data processing. Future research needs to continually explore new methods and technologies, such as the application of graph neural networks and style transfer across modalities and non-visual domains, which are essential directions for advancing the field.

Through this systematic and comprehensive survey, this paper aims to provide researchers with a clear perspective to further the development of style transfer. By deeply analyzing the theoretical foundations, algorithmic principles, and practical applications, we hope to inspire more innovative thinking and exploration of research directions in style transfer. Looking ahead, we anticipate more interdisciplinary collaborations that will create richer, more diverse, and highly personalized experiences for users in the realm of style transfer technology.

References

- [1] Tang Renwei, Liu Qi, and Tan Hao. "A Review of Neural Style Transfer Models." *Journal of Computer Engineering & Applications*, 57.19 (2021).
- [2] Li Benjia. "A Review of the Development of Non-Photorealistic Rendering Techniques." *Computer Knowledge and Technology*, 2018, 14(35): 188-190.
DOI:10.14004/j.cnki.ckt.2018.4166.
- [3] Efros, AA, Freeman, WT. Image quilting for texture synthesis and transfer[C].//COMPUTER GRAPHICS.2001:341-346.
- [4] Hertzmann, Aaron, et al. "Image analogies." *Seminal Graphics Papers: Pushing the Boundaries*, Volume 2. 2023. 557-570.
- [5] Gatys, Leon A.Ecker, Alexander S.Bethge, Matthias. Image Style Transfer Using Convolutional Neural Networks[C]. // IEEE Conference on Computer Vision and Pattern Recognition. 2016: 2414-2423.
- [6] Liu Jiaxiong, Zhou Jun. "A Review of Style Transfer Methods Based on Deep Learning." *Science and Engineering of Artificial Intelligence*, 2024, (01): 1-19.
- [7] Goodfellow, Ian, et al. "Generative adversarial nets." *Advances in neural information processing systems* 27 (2014).
- [8] Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." *Proceedings of the IEEE international conference on computer vision*. 2017.
- [9] Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." *Advances in neural information processing systems* 33 (2020): 6840-6851.
- [10] Dhariwal, Prafulla, and Alexander Nichol. "Diffusion models beat gans on image synthesis." *Advances in neural information processing systems* 34 (2021): 8780-8794.
- [11] Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. *European Conference on Computer Vision (ECCV)*, 740-755.
- [12] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 248-255.
- [13] Saleh, Babak, and Ahmed Elgammal. "Large-scale classification of fine-art paintings: Learning the right metric on the right feature." *arXiv preprint arXiv:1505.00855* (2015).
- [14] Nilsback, Maria-Elena, and Andrew Zisserman. "Automated flower classification over a large number of classes." 2008 Sixth Indian conference on computer vision, graphics & image processing. IEEE, 2008.
- [15] Cordts, Marius, et al. "The cityscapes dataset for semantic urban scene understanding." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- [16] Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
- [17] Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." *Proceedings of the IEEE international conference on computer vision*. 2017.
- [18] Park, Taesung, et al. "Contrastive learning for unpaired image-to-image translation." *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX* 16. Springer International Publishing, 2020.
- [19] Nichol, Alexander Quinn, and Prafulla Dhariwal. "Improved denoising diffusion probabilistic models." *International conference on machine learning*. PMLR, 2021.
- [20] Saharia, Chitwan, et al. "Palette: Image-to-image diffusion models." *ACM SIGGRAPH 2022 conference proceedings*. 2022.