A Survey of Deep Learning-Based Image Style Transfer: Postprint

Authors: Chen Shuhuan, Wei Yuke, Xu Le, Dong Xiaohua, Wen Kunzhe

Date: 2018-07-23T00:00:00+00:00

Abstract

Image style transfer is an image processing method that renders the semantic content of images with different styles. With the advent of deep learning, image style transfer has achieved further development and a series of breakthrough research outcomes. Its exceptional style transfer capabilities have attracted extensive attention from both academic and industrial communities, possessing significant research value. To promote technical research on deep learning-based image style transfer, this work systematically summarizes and discusses the current mainstream methods and representative contributions. It begins by reviewing non-parametric image style transfer, then elaborates in detail on the fundamental principles and methodologies of contemporary deep learning-based image style transfer, analyzes the application prospects of image style transfer in related domains, and concludes by summarizing the existing challenges and future research directions for deep learning-based image style transfer.

Full Text

Survey of Image Style Transfer Based on Deep Learning

Chen Shuhuan, Wei Yuke, Xu Le, Dong Xiaohua, Wen Kunzhe
(School of Computer Science, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: Image style transfer is an image processing technique that renders the semantic content of an image in different artistic styles. With the rise of deep learning, image style transfer has achieved significant advancements and breakthrough research results. Its remarkable style transfer capabilities have attracted widespread attention from both academic and industrial communities, making it a topic of important research value. To advance technical research in deep learning-based image style transfer, this paper summarizes and discusses

current mainstream methods and representative works. We first review non-parametric image style transfer approaches, then introduce in detail the fundamental principles and methods of deep learning-based image style transfer, analyze the application prospects of image style transfer technology in related fields, and finally summarize existing problems and future research directions.

Keywords: image style transfer; deep learning; transfer learning; texture synthesis

0 Introduction

Traditional non-parametric image style transfer methods are primarily based on physical models for rendering and texture synthesis. Efros et al. [error! reference not found] proposed a simple texture algorithm that synthesizes new textures by stitching and reorganizing sample textures. Hertzmann et al. [error! reference not found] introduced an analogy-based approach that synthesizes images with new textures through image feature mapping relationships. Zhang Haisong et al. [error! reference not found] utilized multi-layer texture arrays, Chinese painting illumination models, and contour extraction modules to render 3D mountain scenes with real-time Chinese painting effects. Qian Xiaoyan et al. [error! reference not found] proposed a neighborhood consistency metric method that introduces statistical properties into similarity measures to improve the efficiency of searching for image matching points. Although these methods have achieved considerable results, non-parametric image style transfer approaches can only extract low-level image features rather than high-level abstract features, resulting in relatively coarse image synthesis effects when dealing with complex colors and textures, which fails to meet practical requirements.

With the rise of deep learning [error! reference not found][error! reference not found][error! reference not found], Gatys et al. [error! reference not found] pioneered a convolutional neural network-based approach to image style transfer. They discovered that convolutional neural networks could be used to separate the content and style abstract feature representations of images, enabling effective image style transfer by independently processing these high-level abstract features, thereby achieving remarkable artistic effects, as shown in Figure 1 [FIGURE:1]. The core idea of their algorithm involves using a pre-trained VGG model [error! reference not found] to extract high-level abstract feature representations from both content and style images, then starting from a random noise image and generating a synthesized image with the original content and new style through iterative optimization.

The work of Gatys et al. [error! reference not found] has attracted extensive attention from academic and industrial communities. In academia, numerous follow-up studies have been proposed, mainly falling into two categories: image iteration-based and model iteration-based approaches. Among these, image iteration methods can be further categorized based on how style is obtained into Maximum Mean Discrepancy (MMD) [error! reference not found][error!

reference not found], Markov Random Field (MRF) [error! reference not found], and Deep Image Analogy (DIA) [error! reference not found]. Model iteration methods can be summarized as generative model-based [error! reference not found][error! reference not found][error! reference not found] and image reconstruction decoder-based [error! reference not found][error! reference not found] approaches. These methods have been successfully applied in industrial software such as Prisma, Ostagram, and Deep Forger.

The total loss function of Gatys et al.' s [error! reference not found] method is expressed as follows:

$$\mathcal{L}_{total}(x, c, s) = \alpha \mathcal{L}_{content}(x, c) + \beta \mathcal{L}_{style}(x, s)$$

where α represents the weight coefficient of the image content loss function and β represents the weight coefficient of the image style loss function. The image content loss function $\mathcal{L}_{content}(x, c)$ is expressed as:

$$\mathcal{L}_{content}(x, c) = \frac{1}{2} \sum_{l=1}^{L} \sum_{i=1}^{N_l} \sum_{j=1}^{M_l} (F_{ij}^l - P_{ij}^l)^2$$

where F^l represents the content feature representation of the white noise image at layer l in the VGG model, F_{ij}^l denotes the activation value at position j on the i-th filter at layer l, and P^l represents the content feature representation of the content image at layer l in VGG. The total style loss function $\mathcal{L}_{style}(x, s)$ is expressed as:

$$\mathcal{L}_{style}(x, s) = \sum_{l=0}^{L} \omega_l E_l$$

where L represents the total number of convolutional layers in the VGG network used to extract image style feature representations, ω_l denotes the weight factor for the image style loss function corresponding to convolutional layer l, and E_l represents the style loss function at layer l:

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i=1}^{N_l} \sum_{j=1}^{N_l} (G_{ij}^l - A_{ij}^l)^2$$

where G^l and A^l are the Gram matrices of the content image and style image respectively, N_l represents the number of filters at layer l, and M_l represents the size of the feature map at layer l.

Gatys et al. defined the style loss function using Gram matrices. The Gram matrix is expressed as:

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

where F_{ik}^{l} represents the activation value at position k on the i-th filter at layer l.

This paper systematically reviews current deep learning-based image style transfer methods, analyzes their latest research results and application prospects, discusses existing problems in depth, proposes practical suggestions, and lays a foundation for further research. Finally, we summarize future challenges and development trends.

1 Deep Learning-Based Image Style Transfer Methods

This section describes current mainstream deep learning-based image style transfer methods, including two categories: image iteration-based and model iteration-based approaches, as summarized in Table 1 . The first category performs style transfer through optimization iterations directly on white noise images, where the optimization target is the white noise image itself. The second category iteratively optimizes neural network models to achieve fast style transfer in a feed-forward manner, where the optimization target is the neural network model. We discuss these two categories in detail below.

1.1 Image Iteration-Based Methods

The goal of image iteration-based methods is to make a white noise image simultaneously match the content features of a content image and the style features of a style image, ultimately obtaining a stylized synthesized image. We provide detailed discussions of three representative approaches: Maximum Mean Discrepancy-based, Markov Random Field-based, and Deep Image Analogy-based methods.

1.1.1 Maximum Mean Discrepancy-Based Methods Gatys et al. [error! reference not found] first discovered that abstract content representations could be extracted from arbitrary images by reconstructing intermediate layer features of VGG networks, while style feature representations could be extracted by constructing Gram matrices [error! reference not found]. Li et al. [error! reference not found] theoretically proved that matching Gram matrices is equivalent to minimizing a specific Maximum Mean Discrepancy. Therefore, we categorize Gram matrix-based style transfer methods as MMD-based approaches.

Specifically, given a white noise image x, content image c, and style image s, Nikulin et al. [error! reference not found] thoroughly explored the principles of Gatys et al.' s [error! reference not found] method, discussing the impact of different hyperparameters and stylization attributes on image style transfer effects. Novak et al. [error! reference not found] proposed several approaches to

improve Gatys et al.'s [error! reference not found] method. To better explain and refine Gram matrices, Li et al. [error! reference not found] conducted in-depth investigations, proposing the use of different kernel functions to improve the style loss function, such as linear and Gaussian kernel functions. Additionally, Gatys et al.' s [error! reference not found] method suffers from instability during iterative optimization, which can affect texture synthesis. To address this issue, Risser et al. [error! reference not found] introduced a histogram loss function to solve the problem of texture disorder caused by unstable iterative optimization. Yin [error! reference not found] proposed a content-aware method that can effectively control the synthesis of image content and texture, further improving the resolution of synthesized images.

1.1.2 Markov Random Field-Based Methods Markov Random Fields represent a classic framework for non-parametric image synthesis [error! reference not found], describing collections with similar feature information. Li et al. [error! reference not found] first proposed a method combining MRF with deep convolutional neural networks, segmenting image feature maps into numerous patches and matching them to improve the visual plausibility of synthesized images. Specifically, given content image c and style image s, let the synthesized target image be x. The problem can be formulated as:

$$x^* = \arg\min_x \mathcal{E}_s(x, s) + \alpha\mathcal{E}_c(x, c) + \gamma\mathcal{E}_{tv}(x)$$

where $\mathcal{E}_s$ represents the style loss function, $\mathcal{E}_c$ represents the content loss function, $\mathcal{E}_{tv}$ denotes the regularization term for smoothing the synthesized image, Φ represents the set of feature maps at different layers of the neural network model, and α and γ are weight coefficients for the content loss function and regularization term respectively.

The style loss function $\mathcal{E}_s$ is expressed as:

$$\mathcal{E}_s(x, s) = \sum_{i=1}^{m} \psi_i(x) - \psi_{NN(i)}(s)$$

where m is the cardinality of $\psi(x)$, i.e., the number of patches, $\psi_i(x)$ represents a patch in $\psi(x)$, and for each patch $\psi_i(x)$, normalized cross-correlation is used to find its best matching patch $\psi_{NN(i)}(s)$.

The content loss function $\mathcal{E}_c$ is expressed as:

$$\mathcal{E}_c(x, c) = \|\Phi(x) - \Phi(c)\|^2$$

The regularization term $\mathcal{E}_{tv}$ is expressed as:

$$\mathcal{E}_{tv}(x) = \sum_{i,j} \left((x_{i,j+1} - x_{i,j})^2 + (x_{i+1,j} - x_{i,j})^2 \right)$$

Subsequently, Champandard et al. [error! reference not found] built upon Li et al.' s work by adding manually created semantic maps to enhance control over synthesis results, making the structure of synthesized results more reasonable and significantly improving image quality.

1.1.3 Deep Image Analogy-Based Methods The concept of image analogy was originally proposed by Hertzmann et al. [error! reference not found] to handle problems involving deep mining of mapping relationships between images. To better find semantically meaningful dense correspondences between two input images, Liao et al. [error! reference not found] combined the concept of image analogy with deep learning, proposing a deep image analogy method through patch matching and iterative optimization. This approach can apply the concept of image analogy to deep network feature spaces, finding semantically meaningful dense correspondences to improve the effectiveness of image style transfer.

The mapping relationship of deep image analogy can be represented as $A : A^* :: B : B^*$, where A^* and B are unknown variables. This mapping relationship has two constraints: A and A^* or B and B^* share similar image content features; A and B or A^* and B^* share similar image style features. Therefore, the image mapping relationship can be expressed as:

$$\varphi_{a \to b}(p) = \arg\min_{q \in N_b} \sum_{l=1}^{L} \|F_{A^*}^l(p) - F_B^l(q)\|^2 + \sum_{l=1}^{L} \|F_A^l(p) - F_{B^*}^l(q)\|^2$$

Deep image analogy methods first use a pre-trained VGG model to compute abstract feature representations F_A^l and $F_{B^*}^l$ for known images A and B^*, where L represents the number of layers used in the pre-trained VGG model. The mapping relationship for deep image analogy can be obtained through nearest neighbor field search (NNFS) at layer L. Although A^* and B are unknown variables, based on the first constraint of the analogy mapping relationship, we can assume that A and A^* or B and B^* have similar high-level abstract feature representations in the pre-trained VGG model. Therefore, we can have $F_A^L = F_{A^*}^L$ and $F_B^L = F_{B^*}^L$. The mapping relationship function can be expressed as:

$$\varphi_{a \to b}(p) = \arg\min_{q \in N_b} \sum_{l=1}^{L} \|F_A^l(p) - F_B^l(q)\|^2$$

where N_b represents the neighborhood around point p, and similarly for $\varphi_{b \to a}$.

Based on this deep image analogy mapping relationship, deconvolution operations are performed through convolutional mapping functions $\mathcal{C}_l^{-1}$ that inversely map features in the abstract feature space of the pre-trained VGG model, iterating from high layers to low layers to finally obtain synthesized image A^* with A' s content and B^{*}' s style, and synthesized image B with B^{*}' s content and A^{*}' s style. The specific algorithmic process can be described in pseudocode as follows:

Algorithm: Deep Image Analogy Algorithm

Input: Two RGB color space images A and B^*.

Output: Two pixel space mapping relationship functions $\varphi_{a\to b}$ and $\varphi_{b\to a}$; and two RGB color space images A^* and B.

Procedure: 1. Input A and B^* into the pre-trained VGG model to extract abstract features. 2. Randomly initialize mapping relationship functions $\varphi_{a\to b}$ and $\varphi_{b\to a}$. 3. For $L = 5$ to 1 do: - Nearest neighbor search: $\varphi_{a\to b} \leftarrow \text{map}_{a\to b}$, $\varphi_{b\to a} \leftarrow \text{map}_{b\to a}$ - If $L > 1$ then: - Image reconstruction: $\varphi_{a\to b} \leftarrow \text{up-sample}(\varphi_{a\to b})$, $\varphi_{b\to a} \leftarrow \text{up-sample}(\varphi_{b\to a})$

Deep image analogy methods demonstrate excellent performance in texture and color transfer, but suffer from long computation times. He et al. [error! reference not found] built upon Liao et al.' s [error! reference not found] work to achieve one-to-one and one-to-many image color transfer. This method primarily processes image colors, performing iterative optimization through analogy under local and global constraints to finally generate natural-looking new images.

1.2 Model Iteration-Based Methods

Although image iteration-based methods can produce high-quality stylized images, they suffer from low computational efficiency. Model iteration-based image style transfer methods address this issue by using large amounts of images to train generative models that can produce stylized images, significantly improving computational efficiency and enabling combination with image iteration-based methods. Currently, applications in the market primarily use model iteration-based methods. We discuss two representative approaches: generative model-based and image reconstruction decoder-based methods.

1.2.1 Generative Model-Based Methods Johnson et al. [error! reference not found] first proposed a generative model iteration-based image style transfer method, also known as fast style transfer, as shown in Figure 2

. This method builds upon Gatys et al.' s algorithm, using a perceptual loss function to train a generative model for a specific style. Compared with previous loss functions that compared pixels individually, the perceptual loss function computes squared differences on high-level abstract features extracted by a pre-trained VGG model, consistent with Gatys et al.' s algorithm. Specifically, Johnson et al.' s method uses residual networks [error! reference not found]

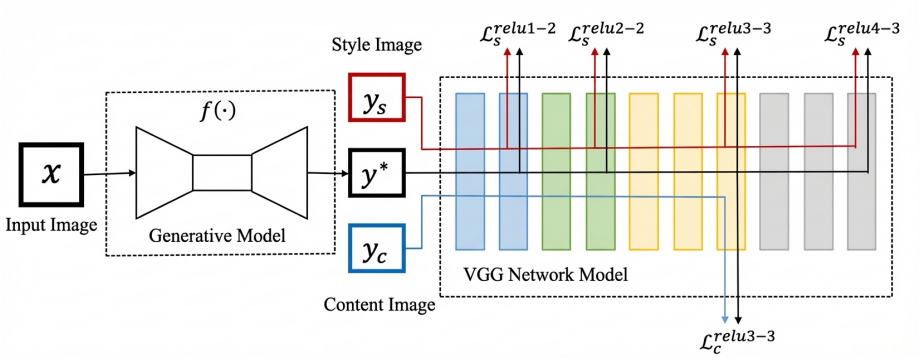

Figure 1: Figure 2

as the basic component of the generative model, with COCO dataset [error! reference not found] as training data. The perceptual loss function can be expressed as:

$$\mathcal{L}_{total}(y, \hat{y}) = \lambda_c \mathcal{L}_{content}(y, \hat{y}) + \lambda_s \mathcal{L}_{style}(y, \hat{y}) + \lambda_{TV} \mathcal{L}_{TV}(\hat{y})$$

where λ_c, λ_s, and λ_{TV} represent the weight coefficients for content loss, style loss, and image smoothness respectively, f denotes the generative model function, $\Phi(y)$ represents content features extracted by the pre-trained VGG model from training images, and $\Phi(\hat{y})$ represents style features extracted from style images.

Johnson et al.' s work [error! reference not found] provides excellent inspiration for improving image style transfer efficiency. Additionally, Ulyanov et al.' s work [error! reference not found] adopted a similar network architecture and demonstrated through experiments that using instance normalization [error! reference not found] instead of batch normalization [error! reference not found] during generative model training can significantly improve generated image quality. Wang et al. [error! reference not found] proposed a multimodal convolutional neural network that considers feature representations of color and brightness channels, performing stylization hierarchically at multiple scales to effectively solve texture scaling issues and produce impressive results on high-resolution images. Zhang et al. [error! reference not found] constructed a generative model that can be trained on multiple styles, enabling fast multi-style transfer. Huang et al. [error! reference not found] proposed an adaptive instance normalization method that eliminates the need for predefining stylization during generative model training.

Furthermore, Generative Adversarial Networks (GANs) [error! reference not found] have demonstrated excellent performance in image style transfer. Li et al. [error! reference not found] combined MRF with GANs, using adversarial training to train generative models, resulting in highly realistic generated im-

ages. Subsequent unsupervised GANs such as CycleGAN [error! reference not found], DiscoGAN [error! reference not found], and DualGAN [error! reference not found], where CycleGAN is based on cycle consistency and DiscoGAN and DualGAN are based on dual learning ideas from machine translation [error! reference not found], have achieved breakthroughs by eliminating the need for paired training data and successfully implementing unsupervised transfer learning with largely consistent network architectures and implementations. We explain the widely applicable CycleGAN model as an example.

Zhu et al. [error! reference not found] proposed the CycleGAN model, which includes two generative models G and F, and two discriminative models D_X and D_Y, using cycle consistency as a constraint on the total loss function. The total loss function in CycleGAN consists of two parts: adversarial loss and cycle consistency loss. The adversarial loss includes forward mapping loss and backward mapping loss. Given datasets X and Y, where $x \in X$ and $y \in Y$, the forward mapping loss function is:

$$\mathcal{L}_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{data}(y)}[\log D_Y(y)] + \mathbb{E}_{x \sim p_{data}(x)}[\log(1 - D_Y(G(x)))]$$

Similarly, the backward mapping loss function is:

$$\mathcal{L}_{GAN}(F, D_X, Y, X) = \mathbb{E}_{x \sim p_{data}(x)}[\log D_X(x)] + \mathbb{E}_{y \sim p_{data}(y)}[\log(1 - D_X(F(y)))]$$

The cycle consistency loss ensures consistency between generative models G and F:

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x \sim p_{data}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{data}(y)}[\|G(F(y)) - y\|_1]$$

Finally, the total loss function for CycleGAN is:

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{GAN}(G, D_Y, X, Y) + \mathcal{L}_{GAN}(F, D_X, Y, X) + \lambda \mathcal{L}_{cyc}(G, F)$$

where λ is a balance parameter for the relative importance of the two mapping objectives. The adversarial training optimization objective for the CycleGAN model is:

$$G^*, F^* = \arg \min_{G,F} \max_{D_X, D_Y} \mathcal{L}(G, F, D_X, D_Y)$$

where G^* and F^* represent the approximately optimal generative models obtained.

However, current GANs are quite unstable during training, and the design of discriminative models makes it difficult to implement directed image style transfer methods. Moreover, GANs perform adversarial training based on iterative optimization of image divergence distributions rather than based on image content, texture, and color, making the style transfer process difficult to control.

1.2.2 Image Reconstruction Decoder-Based Methods While image iteration-based methods suffer from parameter adjustment and low efficiency issues, fast style transfer alleviates the efficiency problem but can only be trained for specific styles and still cannot avoid parameter adjustment. To overcome these problems, Li et al. [error! reference not found] proposed an image style transfer algorithm based on image reconstruction decoders that no longer requires training for specific styles and avoids parameter adjustment.

This algorithm uses a multi-level stylization strategy, as shown in Figure 3 [FIGURE:3]. It first uses a pre-trained VGG model as an encoder, fixing its weights to train a decoder network to invert VGG features back to original images, where the decoder is designed symmetrically to the encoder. The decoder uses pixel reconstruction loss and feature loss as constraints for image reconstruction, with the loss function expressed as:

$$\mathcal{L}_{recon} = \|I_{output} - I_{input}\|_2^2 + \lambda\|\Phi(I_{output}) - \Phi(I_{input})\|_2^2$$

where I_{output} and I_{input} represent the reconstructed output and input images respectively, training data uses the COCO dataset [error! reference not found], Φ represents image feature representations extracted by the pre-trained VGG encoder, and λ is a balance weight coefficient between pixel reconstruction loss and feature loss.

After completing decoder training for corresponding layers, projection functions are set between encoder $\mathcal{E}_n(\cdot)$ and decoder $\mathcal{D}_e(\cdot)$. Through whitening and coloring transform (WCT), stylized image reconstruction is performed. Specifically, given content image c and style image s, their abstract feature representations at specific layers are extracted in the pre-trained VGG model as $\mathcal{E}_n(c)$ and $\mathcal{E}_n(s)$. Through whitening and coloring transforms, the vectorized features for content and style images are obtained as $\mathcal{H}_c$ and $\mathcal{H}_s$, then the stylized encoding result $\mathcal{H}_{cs}$ for the corresponding layer is calculated as:

$$\mathcal{H}_{cs} = \mathcal{E}_s\mathcal{D}_s\mathcal{E}_s^T\mathcal{H}_c$$

where $\mathcal{E}_s$ and $\mathcal{D}_s$ are the orthogonal and diagonal matrices of the covariance matrix $\mathcal{H}_s\mathcal{H}_s^T$ respectively, and $\mathcal{E}_c$ and $\mathcal{D}_c$ are the orthogonal and diagonal matrices of the covariance matrix $\mathcal{H}_c\mathcal{H}_c^T$ respectively. Finally, the trained decoder $\mathcal{D}_e(\cdot)$ decodes the stylized encoding $\mathcal{H}_{cs}$ to obtain the synthesized image Y_{cs} for the corresponding layer:

$$Y_{cs} = \mathcal{D}_e(\mathcal{H}_{cs})$$

To achieve better results, Li et al. [error! reference not found] further improved the image reconstruction encoder structure from [error! reference not found] and added post-processing with local image smoothing [error! reference not found][error! reference not found] to achieve photorealistic fast style transfer, with effects largely consistent with the deep photo style transfer method proposed by Luan et al. [error! reference not found].

2 Application Analysis

With continuous improvements in algorithms and theory, deep learning-based image style transfer has achieved significant enhancements in quality and has broad commercial application prospects. Currently, applications mainly fall into three directions:

a) Image Processing. Most images circulating on social networks today are processed by software, with image beautification being a popular application. Traditional image processing techniques can only perform fixed-pattern processing, while neural network-based image style transfer brings more possibilities for image style design. Chen et al. [error! reference not found] proposed a content-aware style transfer method that can be effectively applied to image inpainting. Zhang et al. [error! reference not found] proposed a method for coloring comic sketches. Prisma was the first mobile application to offer free deep learning-based image style transfer services, capable of transforming users' photos into high-quality artworks within seconds. Subsequently, several paid image style transfer applications emerged, generating certain commercial value. With these applications, people can easily create artworks in their own style without requiring special professional skills.

b) Video Processing. In the film and entertainment industry, such as movies, television, and animation, visual effects technology is ubiquitous. However, creating visual effects requires not only special professional skills but also extensive manual labor. Using more artificial intelligence technology could significantly reduce production costs, and image style transfer is a viable solution. For example, Anderson et al. [error! reference not found] used optical flow and deep neural networks for movie stylization. Ruder et al. [error! reference not found] introduced temporal consistency loss functions to improve coherence between frames after video stylization. Chen et al. [error! reference not found] constructed a temporally correlated network model that can incorporate multiple styles and perform real-time online video stylization. Joshi et al. [error! reference not found] investigated more advanced parameter spaces in image style transfer and identified a set of effective components for impressionist stylization of movie scenes.

c) Style Design Assistance Tools. Image style transfer can serve as a useful

auxiliary tool in areas such as artistic painting creation, architectural design, fashion design, and game scene design. Although there are currently no relevant references or successful application cases, this is likely to become a future research hotspot.

From current research progress, deep learning-based image style transfer is developing rapidly, with substantial research space remaining for improving algorithm efficiency and image quality, and its potential commercial value awaits further exploration.

3 Existing Problems and Research Directions

Deep learning-based image style transfer algorithms have achieved remarkable results, but several problems remain to be solved. This section summarizes the main existing issues and proposes some suggestions.

a) Parameter Adjustment. To obtain satisfactory results, both image iteration-based and model iteration-based methods require manual parameter tuning, particularly model iteration-based methods that need model retraining after each parameter adjustment. Although image reconstruction decoder-based methods alleviate the parameter adjustment problem and eliminate the need for separate model training for different styles, their training process is cumbersome and image generation quality is not ideal. While local smoothing processing can improve image reconstruction decoder-based methods, it causes stylized image textures to disappear, making the final results almost similar to image color transfer [error! reference not found]. Therefore, finding a method that is both simple/controllable and guarantees image quality is an important future research direction. If model storage capacity is not a concern, further improving the image generation quality of image reconstruction encoder-based methods is a worthwhile research direction, as this approach can effectively avoid parameter adjustment issues.

b) Limitations of Pre-trained Models. Gatys et al. [error! reference not found] discovered that using a pre-trained VGG model can extract high-level abstract features from images, enabling image style transfer through iterative optimization. To date, most deep learning-based image style transfer methods use pre-trained VGG models for feature extraction. Although VGG is an excellent CNN model that performs well in feature extraction, it is a heavyweight model with large size and computational requirements, and was not originally designed specifically for image style transfer. Therefore, breaking free from dependence on pre-trained VGG models or designing more compact and effective feature extractors is an important pathway for advancing deep learning-based image style transfer. Generative adversarial networks may solve the limitations of pre-trained models, as their realistic image generation effects help improve quality, their divergence distribution-based optimization is similar to image iteration-based methods, and adversarial training performs well in acquiring new features.

c) Improvement of Transfer Learning Theory. Image style transfer is a typical use case of transfer learning. Currently, deep learning-based transfer learning methods are still in their infancy and require more complete mathematical methods and theoretical guidance. The improvement of transfer learning theory is crucial for the further development of deep learning-based image style transfer. Research on universal models [error! reference not found][error! reference not found] has proposed designing highly generalizable neural network models to improve transfer learning capabilities, providing important guidance for the future development of image style transfer.

d) Preprocessing and Postprocessing Methods. To make final results more practical, preprocessing and postprocessing methods can be employed, such as image semantic segmentation [error! reference not found], image fusion [error! reference not found], image color transfer [error! reference not found][error! reference not found], and image smoothing processing [error! reference not found][error! reference not found]. These preprocessing and postprocessing methods play important roles in improving image style transfer effects. For example, Castillo et al. [error! reference not found] combined image semantic segmentation to perform style transfer on specific objects in images. Li et al. [error! reference not found]' s work combined image fusion technology to provide user-friendly interaction. Gatys et al. [error! reference not found] used image color transfer methods to achieve color control in stylized images. Li et al. [error! reference not found] performed post-processing with local image smoothing on stylized images to achieve photorealistic effects, as shown in Figure 4 [FIGURE:4]. Therefore, combining effective preprocessing and postprocessing methods is an important means to improve style transfer results.

4 Conclusion

This paper provides a detailed introduction to deep learning-based image style transfer, discussing its application prospects, existing problems, and development directions. Although successful application cases already exist, there remains a considerable distance to widespread commercial application, requiring further research and improvement. Overall, deep learning-based image style transfer is a challenging emerging topic that has attracted extensive academic attention and has significant industrial demand, making it a research area with important significance and broad application prospects.

References

[1] Efros A A, Freeman W T. Image quilting for texture synthesis and transfer [C]// Proc of the 28th annual conference on Computer graphics and interactive techniques. New York: ACM Press, 2001: 341-346.

[2] Hertzmann A, Jacobs C E, Oliver N, et al. Image analogies [C]// Proc of the 28th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM Press, 2001: 327-340.

[3] Zhang Haisong, Yin Xiaoqin, Yu Jinhui. Real-time Rendering of 3D Chinese Painting Effects [J]. Journal of Computer-Aided Design & Computer Graphics, 2004, 16 (11): 1485-1489. (in Chinese)

[4] Qian Xiaoyan, Xiao Liang, Wu Huizhong. Fast Style Transfer [J]. Computer Engineering, 2006, 32 (21): 15-17. (in Chinese)

[5] Mao Yonghua, Gui Xiaolin, Li Qian, et al. Study on application technology of deep learning [J]. Application Research of Computers, 2016, 33 (11): 3201-3205. (in Chinese)

[6] Liu Jianwei, Liu Yuan, Luo Xionglin. Research and development on deep learning [J]. Application Research of Computers, 2014, 31 (7): 1921-1930. (in Chinese)

[7] Sun Zhijun, Xue Lei, Xu Yangming, et al. Overview of deep learning [J]. Application Research of Computers, 2012, 29 (8): 2806-2810. (in Chinese)

[8] Gatys L A, Ecker A S, Bethge M. A neural algorithm of artistic style [J]. arXiv preprint arXiv: 1508. 06576, 2015.

[9] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv: 1409. 1556, 2014.

[10] Gatys L A, Ecker A S, Bethge M. Texture synthesis using convolutional neural networks [J]. arXiv preprint arXiv: 1505. 07376, 2015.

[11] Gatys L A, Ecker A S, Bethge M. Image style transfer using convolutional neural networks [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE Computer Society Press, 2016: 241-250.

[12] Gatys L A, Bethge M, Hertzmann A, et al. Preserving color in neural artistic style transfer [J]. arXiv preprint arXiv: 1606. 05897, 2016.

[13] Gatys L A, Ecker A S, Bethge M, et al. Controlling perceptual factors in neural style transfer [J]. arXiv preprint arXiv: 1611. 07865, 2016.

[14] Li Yanghao, Wang Naiyan, Liu Jiaying, et al. Demystifying neural style transfer [J]. arXiv preprint arXiv: 1701. 01036, 2017.

[15] Li Chuan, Wand M. Combining Markov random fields and convolutional neural networks for image synthesis [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition, Piscataway, NJ: IEEE Press, 2016: 2479-2486.

[16] Liao Jing, Yao Yuan, Yuan Lu, et al. Visual attribute transfer through deep image analogy [J]. arXiv preprint arXiv: 1705. 01088, 2017.

[17] Johnson J, Alahi A, Li Feifei. Perceptual losses for real-time style transfer and super-resolution [C]// Proc of European Conference on Computer Vision. [S. I.] : Springer Press, 2016: 694-711.

[18] Huang Haozhi, Wang Hao, Luo Wenhan, et al. Real-time neural style transfer for videos [C]// Proc of IEEE Conference on Computer Vision and Pattern

Recognition. Piscataway, NJ: IEEE Press, 2017: 7044-7052.

[19] Wang Xin, Oxholm G, Zhang Da, et al. Multimodal transfer: a hierarchical deep convolutional neural network for fast artistic style transfer [J]. arXiv preprint arXiv: 1612. 01895, 2016.

[20] Li Yijun, Fang Chen, Yang Jimei, et al. Universal style transfer via feature transforms [J]. arXiv preprint arXiv: 1705. 08086, 2017.

[21] Li Yijun, Liu Mingyu, Li Xueting, et al. A closed-form solution to photorealistic image stylization [J]. arXiv preprint arXiv: 1802. 06474, 2018.

[22] Nikulin Y, Novak R. Exploring the neural algorithm of artistic style [J]. arXiv preprint arXiv: 1602. 07188, 2016.

[23] Novak R, Nikulin Y. Improving the neural algorithm of artistic style [J]. arXiv preprint arXiv: 1605. 04603, 2016.

[24] Risser E, Wilmot P, Barnes C. Stable and controllable neural texture synthesis and style transfer using histogram losses [J]. arXiv preprint arXiv: 1701. 08893, 2017.

[25] Yin Rujie. Content aware neural style transfer [J]. arXiv preprint arXiv: 1601. 04568, 2016.

[26] Efros A A, Leung T K. Texture synthesis by non-parametric sampling [C]// Proc of the 7th IEEE International Conference. Piscataway, NJ: IEEE Press, 1999, 2: 1033-1038.

[27] Champandard A J. Semantic style transfer and turning two-bit doodles into fine artworks [J]. arXiv preprint arXiv: 1603. 01768, 2016.

[28] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition [C]// Proc of IEEE conference on computer vision and pattern recognition. Piscataway, NJ: IEEE Press, 2016: 770-778.

[29] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context [C]// Proc of European Conference on Computer Vision. [S. I.] : Springer Press, 2014: 740-755.

[30] Ulyanov D, Lebedev V, Vedaldi A, et al. Texture networks: feed-forward synthesis of textures and stylized images [J]. arXiv preprint arXiv: 1603. 03417, 2016.

[31] Ulyanov D, Vedaldi A, Lempitsky V. Instance normalization: the missing ingredient for fast stylization [J]. arXiv preprint arXiv: 1607. 08022, 2016.

[32] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift [J]. arXiv preprint arXiv: 1502. 03167, 2015.

[33] Zhang Hang, Dana K. Multi-style generative network for real-time transfer [J]. arXiv preprint arXiv: 1703. 06953, 2017.

[34] Huang Xun, Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization [J]. arXiv preprint arXiv: 1703. 06868, 2017.

[35] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets [C]// Proc of International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2014: 2672-2680.

[36] Li Chuan, Wand M. Precomputed real-time texture synthesis with markovian generative adversarial networks [C]// Proc of European Conference on Computer Vision. [S. I.] : Springer Press, 2016: 702-716.

[37] Zhu Junyan, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks [J]. arXiv preprint arXiv: 1703. 10593, 2017.

[38] Kim T, Cha M, Kim H, et al. Learning to discover cross-domain relations with generative adversarial networks [J]. arXiv preprint arXiv: 1703. 05192, 2017.

[39] Yi Zili, Zhang Hao, Tan Ping, et al. DualGAN: unsupervised dual learning for image-to-image translation [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2017: 2849-2857.

[40] Xia Yingce, He Di, Qin Tao, et al. Dual learning for machine translation [J]. arXiv preprint arXiv: 1611. 00179, 2016.

[41] Zhou Dengyong, Weston J, Gretton A, et al. Ranking on data manifolds [C]// Advances in Neural Information Processing Systems. 2003: 169-176.

[42] Yang Chuan, Zhang Lihe, Lu Huchuan, et al. Saliency Detection via Graph-Based Manifold Ranking [C]// Proc of Computer Vision and Pattern Recognition Conference. Piscataway, NJ: IEEE Press, 2013: 3166-3173.

[43] Luan F, Paris S, Shechtman E, et al. Deep photo style transfer [J]. arXiv preprint arXiv: 1703. 07511, 2017.

[44] Chen Yilei, Hsu C T. Towards deep style transfer: a content-aware perspective [C]// Proc of British Machine Vision Conference. 2016: 8. 1-8. 12.

[45] Anderson A G, Berg C P, Mossing D P, et al. DeepMovie: using optical flow and deep neural networks to stylize movies [J]. arXiv preprint arXiv: 1605. 08153, 2016.

[46] Zhang Lvmin, Ji Yi, Lin Xin. Style transfer for anime sketches with enhanced residual u-net and auxiliary classifier GAN [J]. arXiv preprint arXiv: 1706. 03319, 2017.

[47] Ruder M, Dosovitskiy A, Brox T. Artistic style transfer for videos [J]. arXiv preprint arXiv: 1604. 08610, 2016.

[48] Chen Dongdong, Liao Jing, Yuan Lu, et al. Coherent online video style transfer [J]. arXiv preprint arXiv: 1703. 09211, 2017.

[49] Joshi B, Stewart K, Shapiro D. Bringing impressionism to life with neural style transfer in come swim [J]. arXiv preprint arXiv: 1701. 04928, 2017.

[50] He Mingming, Liao Jing, Yuan Lu, et al. Neural color transfer between images [J]. arXiv preprint arXiv: 1710. 0756, 2017.

[51] Castillo C, De S, Han X, et al. Son of Zorn' s Lemma: targeted style transfer using instance-aware semantic segmentation [J]. arXiv preprint arXiv: 1701. 02357, 2017.

[52] Garcia-Garcia A, Orts-Escolano S, Oprea S, et al. A review on deep learning techniques applied to semantic segmentation [J]. arXiv preprint arXiv: 1704. 06857, 2017.

[53] Levin A, Lischinski D, Weiss Y. A closed-form solution to natural image matting [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2008, 30 (2): 228-242.

[54] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [J]. arXiv preprint arXiv: 1706. 03762, 2017.

[55] Kaiser L, Gomez A N, Shazeer N, et al. One model to learn them all [J]. arXiv preprint arXiv: 1706. 05137v1, 2017.