Hybrid Image Synthesis Using Deep Learning

V.P.V BHARATHI 1,V.SRAVANI 2,M.SOWJANYA 3, K.NAVANEETH SAI KUMAR 4,P.SURYA PRAKASH5, M.SAI KRISHNA TILAK6

#1Assistant Professor in Department of CSE, in Raghu Institute of technology ,Vishakapatnam.

#2#3#4#5 B.Tech in Computer Science and Technology with specialisation in Data Science ,Raghu Institute of Technology & , Vishakapatnam.

ABSTRACT: Neural Style Transfer (NST) is a class of software algorithms that allows us to transform scenes, change/edit the environment of a media with the help of a Neural Network. NST finds use in image and video editing software allowing image stylization based on a general model, unlike traditional methods. This made NST a trending topic in the entertainment industry as professional editors/media producers create media faster and offer the general public recreational use. In this paper, the current progress in Neural Style Transfer with all related aspects such as still images and videos is presented critically. The authors looked at the different architectures used and compared their advantages and limitations. Multiple literature reviews focus on either the Neural Style Transfer (of images) or cover Generative Adversarial Networks (GANs) that generate video. As per the authors' knowledge, this is the only research article that looks at image and video style transfer, particularly mobile devices with high potential usage. This article also reviewed the challenges faced in applying video neural style transfer in real-time on mobile devices and presents research gaps with future research directions. NST, a fascinating deep learning application, has considerable research and application potential in the coming years.

I.INTRODUCTION: The motivation of deep learning is to build a multi-layer neural network to analyse data, with the aim of interpreting data such as images [1], [2], sounds, internet of things [3] and texts by simulating the mechanism of human brain [4]–[6]. Since 2016, deep learning has been applied to a new field, imitating artists' painting style [7], achieving so-called Neural Style Transfer (NST). By inputting a style image and a content image into a trained neural network model, a new image is synthesized. The newly generated image not only has the structure and content features of the original content image, but also has the style or textural features of the original style image. Such NST has become an active research topic in the field of artificial intelligence. Its basic principle is to transfer style from the ''style image'' to the ''content image'' by using a neural network model with these two known images [8]. The purpose is to generate new images with different styles from The associate editor coordinating the review of this manuscript and approving it for publication was Huimin Lu. the same content image according to the guidance of different style images. Nowadays, Neural Style Transfer is widely used to solve many problems, such as video stylization [9], texture synthesis [10], head style transfer [11]–[13], super resolution [14], [15] and font style transfer [16]. In this paper, we specifically consider the problem of image style transfer which is guided by different style loss function strategies. To achieve this, we propose a new loss function which combines a global measure and a local patch based

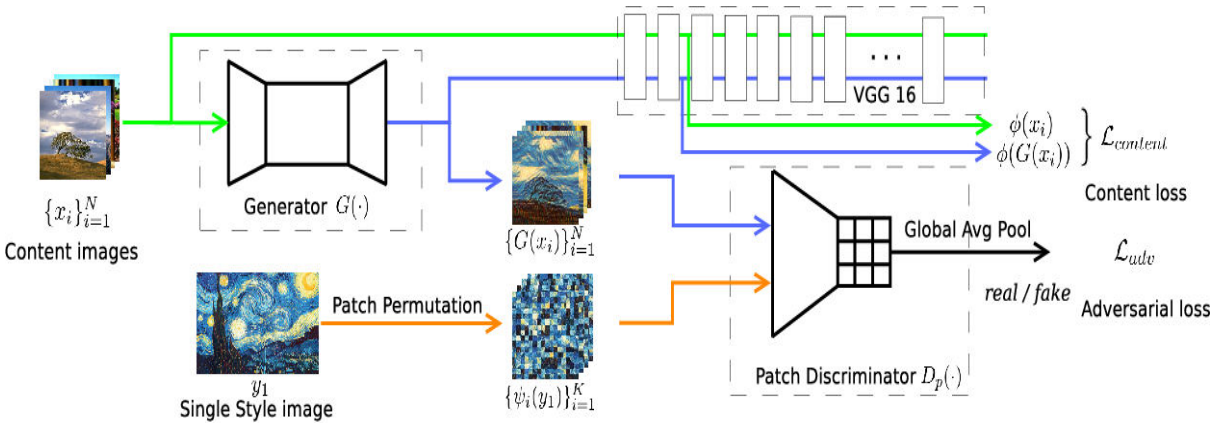
approach. The global style loss helps avoid patch transfer errors, such as mouth, eyes and moustache being transferred into wrong places. An example is shown in figure 1. The local style loss helps better retain detailed styles. By combining both, our method better transfers styles and reduces artifacts.

II. RELATED WORK:

Because neural style transfer can produce impressive results, and can generalize better across different styles than traditional non-photorealistic rendering methods [17], it has become one of the active research topics in academia and industry in recent years. Many research institutes and H.-H. Zhao et al.: Image Neural Style Transfer With Global and Local Optimization Fusion FIGURE 1. Comparison of style transfer results with global style loss and local loss. laboratories have conducted extensive and in-depth research on style transfer. Among them, Stanford University's research group led by Li Fei fei achieved real-time style transfer of images by pre-training the network model of style images, which greatly improves the speed of style transfer and the resolution of image generation [14]. Scholars from the University of Science and Technology of China and Microsoft Research proposed a style library called Style Bank [18] which can be used for image style transfer. Style Bank consists of several convolutional filter banks, each of which explicitly expresses a style. The Durham University team used image style transfer to propose a new method of real-time monocular depth estimation for adaptive synthetic data [19]. The team from Princeton University, Adobe and UC Berkeley proposed a style transfer algorithm called Paired Cycle GAN, which can automatically enhance and remove makeup [20]. Scholars from University of Science and Technology of China, Peking University and Microsoft Research reprocessed the depth features of stylistic images (i.e. arranging the spatial positions of feature maps) to achieve the style transfer of arbitrary images [21]. Researchers from Shanghai Jiaotong University and Microsoft proposed a generalized style transmission network model consisting of a style coder, a content coder, a mixer and a decoder to generate images with target style and content [22]. In order to meet the needs of 3D movies and AR/VR, scholars from the University of Science and Technology of China and Microsoft Research have studied the method of stereo neural style transfer and achieved satisfactory results [23].

Researchers at Tsinghua University and Cardiff University have proposed a Cartoon GAN style transfer algorithm [24]. It can generate arbitrary cartoon images using real scenes as source images where the style is learned from unpaired images from cartoon movies. In

addition, many world-renowned commercial enterprises have joined the style transfer research and its application. For example, Tencent AI along with scholars of Tsinghua University proposed a method of video style transfer using a feedforward network, and adopted a new two-frame cooperative training mechanism to achieve video style transfer [25]. Researchers at Adobe and Cornell University have proposed a method to generate realistic style transfer in various scenarios, which can achieve style transfer for images including daytime, weather, season and art [26]. Adobe, in conjunction with researchers from University of California, proposed a multimodal convolution neural network, which uses separate representations of the color and brightness channels to study hierarchical style transfer with multiple scaling losses [27]. 360 AI Lab, in conjunction with researchers from Peking University and National University of Singapore, proposed a new meta-network model for image style transfer [28]. The meta model is only 449 KB in size and can run the image style transfer program in real time on mobile devices. Sense Time and researchers from the Chinese University of Hong Kong proposed a multi-scale zero-shot style transfer method based on feature decoration [29]. A style decorator was designed to make use of semantic alignment style features from arbitrary style images to form content features. This not only matches their feature distribution on the whole, but also retains the detailed style patterns in decorative features. The partition algorithm is a significant digital image processing technique for image denoising and image reconstruction. [30] proposed a new image denoising method based on Hard partition Weighted Sum filters. In order to solve the multi person pose estimation problem, [31] proposed a novel Pose Partition Network (PPN) which has a good performance in low complexity and high accuracy of joint detection and partition. [32] proposed an Adaptive Triangular Partition Algorithm named IATP for digital images. The method considers the gray-scale distribution of the image and removes the shared edges between the adjacent triangles in the partitioned mesh. Two main types of methods for representing elements of an image are used in deep learning based style transfer: global approaches based on the Gram matrix [7] or other global measures (e.g. histograms [33]) and local approaches based on patch matching [13], [34]. Compared to the global methods, methods based on patch matching are more flexible and better cope with cases in which the visual styles or elements vary across the image. However, they could also produce visible artifacts when there are local matching errors. Compared to 85574 VOLUME 7, 2019 H.-H. Zhao et al.: Image Neural Style Transfer With Global and Local Optimization Fusion FIGURE 2. Comparison of style transfer results with different methods. FIGURE 3. Our neural style transfer framework. the methods based on local approaches, the structure and color of the content image can be preserved better with global approaches, although detailed styles may not be fully captured. An example is shown in figure 2, in which artifacts produced by existing methods are clearly evident.



2.1 Deep image representations :

The results presented below were generated on the basis of the VGG network [28], which was trained to perform object recognition and localisation [26] and is described extensively in the original work [28]. We used the feature space provided by a normalised version of the 16 convolutional and 5 pooling layers of the 19-layer VGG network. We normalized the network by scaling the weights such that the mean activation of each convolutional filter over images and positions is equal to one. Such re-scaling can be done for the VGG network without changing its output, because it contains only rectifying linear activation functions and no normalization or pooling over feature maps. We do not use any of the fully connected layers. The model is publicly available and can be explored in the caffe-framework [14]. For image synthesis we found that replacing the maximum pooling operation by average pooling yields slightly more appealing results, which is why the images shown were generated with average pooling.

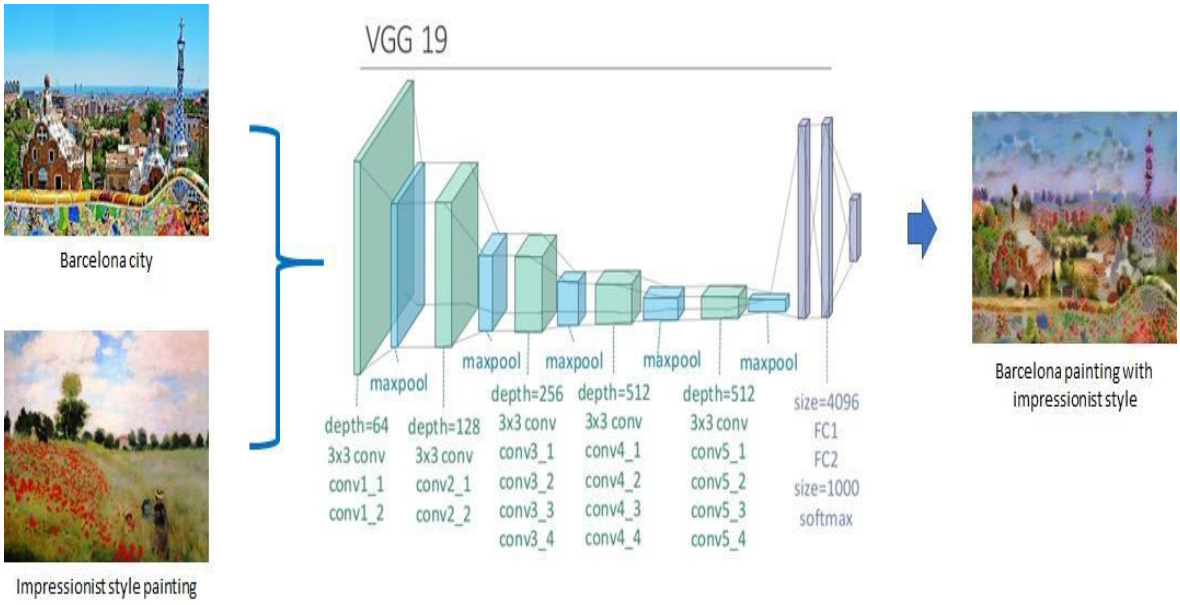
2.2. Content representation

Generally each layer in the network defines a non-linear filter bank whose complexity increases with the position of the layer in the network. Hence a given input image ~x is encoded in each layer of the Convolutional Neural Network by the filter responses to that image. A layer with Nl distinct filters has Nl feature maps each of size Ml , where Ml is the height times the width of the feature map. So the responses in a layer l can be stored in a matrix F l ∈ RNl×Ml where F l ij is the activation of the i th filter at position j in layer l. To visualise the image information that is encoded at different layers of the hierarchy one can perform gradient descent on a white noise image to find another image that matches the feature responses of the original image (Fig 1, content reconstructions) [24]. Let ~p and ~x be the original image and the image that is generated, and P l and F l their respective feature representation in layer l. We then define the squared-error loss between the two feature representations

$$\mathcal{L}_{\text{content}}(\vec{p}, \vec{x}, l) = \frac{1}{2} \sum_{i,j} \left(F_{ij}^l - P_{ij}^l\right)^2 . \qquad (1)$$

The derivative of this loss with respect to the activations in layer l equals

$$\frac{\partial \mathcal{L}_{\text{content}}}{\partial F_{ij}^{l}} = \begin{cases} \left(F^{l} - P^{l}\right)_{ij} & \text{if } F_{ij}^{l} > 0 \\ 0 & \text{if } F_{ij}^{l} < 0\,, \end{cases} \qquad (2)$$



from which the gradient with respect to the image ~x can be computed using standard error back-propagation (Fig 2, right). Thus we can change the initially random image ~x until it generates the same response in a certain layer of the Convolutional Neural Network as the original image ~p. When Convolutional Neural Networks are trained on object recognition, they develop a representation of the image that makes object information increasingly explicit along the processing hierarchy [10]. Therefore, along the processing hierarchy of the network, the input image is transformed into representations that are increasingly sensitive to the actual content of the image, but become relatively invariant to its precise appearance. Thus, higher layers in the network capture the high-level content in terms of objects and their arrangement in the input image but do not constrain the exact pixel values of the reconstruction very much (Fig 1, content reconstructions d, e). In contrast, reconstructions from the lower layers simply reproduce the exact pixel values of the original image (Fig 1, content reconstructions a–c). We therefore refer to the feature responses in higher layers of the network as the content representation.

2.3. Style representation

To obtain a representation of the style of an input image, we use a feature space designed to capture texture information [10]. This feature space can be built on top of the filter responses in any layer of the network. It consists of the correlations between the different filter responses, where the expectation is taken over the spatial extent of the feature maps. These feature correlations are given by the Gram matrix Gl ∈ RNl×Nl , where Gl ij is the inner product between the vectorised feature maps i and j in layer

$$G_{ij}^{l} = \sum_{k} F_{ik}^{l} F_{jk}^{l}. \qquad (3)$$

By including the feature correlations of multiple layers, we obtain a stationary, multi-scale representation of the input image, which captures its texture information but not the global arrangement. Again, we can visualise the information captured by these style feature spaces built on different layers of the network by constructing an image that matches the style representation of a given input image (Fig 1, style reconstructions). This is done by using gradient descent from a white noise image to minimise the mean-squared distance between the entries of the Gram matrices from the original image and the Gram matrices of the image to be generated [10, 25]. Let ~a and ~x be the original image and the image that is generated, and Al and Gl their respective style representation in layer l. The contribution of layer l to the total loss is then

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

and the total style loss i

$$\mathcal{L}_{\text{style}}(\vec{a}, \vec{x}) = \sum_{l=0}^{L} w_l E_l, \qquad (5)$$

where wl are weighting factors of the contribution of each layer to the total loss (see below for specific values of wl I our results). The derivative of El with respect to the activations in layer l can be computed analytically

$$\frac{\partial E_l}{\partial F_{ij}^{l}} = \begin{cases} \frac{1}{N_l^2 M_l^2} \left((F^l)^{\mathrm{T}} \left(G^l - A^l\right)\right)_{ji} & \text{if } F_{ij}^{l} > 0 \\ 0 & \text{if } F_{ij}^{l} < 0\,. \end{cases} \qquad (6)$$

2.4. Style transfer

To transfer the style of an artwork ~a onto a photograph ~p we synthesise a new image that simultaneously matches the content representation of ~p and the style representation of ~a (Fig 2). Thus we jointly minimise the distance of the feature representations of a white noise image from the content representation of the photograph in one layer and the style representation of the painting defined on a number of layers of the Convolutional Neural Network. The loss function we minimise

$$\mathcal{L}_{\text{total}}(\vec{p}, \vec{a}, \vec{x}) = \alpha \mathcal{L}_{\text{content}}(\vec{p}, \vec{x}) + \beta \mathcal{L}_{\text{style}}(\vec{a}, \vec{x}) \qquad (7)$$

where α and β are the weighting factors for content and style reconstruction, respectively. The gradient with respect to the pixel values ∂Ltotal ∂~x can be used as input for some

numerical optimisation strategy. Here we use L-BFGS [32], which we found to work best for image synthesis. To extract image information on comparable scales, we always resized the style image to the same size as the content image before computing its feature representations. Finally, note that in difference to [24] we do not regularise our synthesis results with image priors. It could be argued, though, that the texture features from lower layers in the network act as a specific image prior for the style image. Additionally some differences in the image synthesis are expected due to the different network architecture and optimisation algorithm we use.

2.5 A METHOD BASED ON GAN NETWORK

The GAN network consists of generators and identifiers. The former is dedicated to generating false data, the latter is committed to identifying false data, the two in the confrontation of learning and progress together. Li and Wand et al.[10] get realistic images by training MRFs-based feed-forward networks through adversarial training. The results show that their algorithm is superior to Johnson et al.'s feed-forward generation model algorithm, but the effect is poor in texture look-up because semantic correlation is not taken into account. Mirza et al. proposed CGAN for image generation, and the model added additional information to both the generator and the identifier to guide the model generation direction on the basis of the original GAN model, but the supervised learning algorithm needed to be trained on pre-processed piles of data sets. Zhu et al. proposed CycleGAN without training on paired data sets, a network of two generators and two evaluators for converting between two domains and two evaluators for distinguishing between pictures in two domains. The model not only requires that the image can be converted from the source domain to the target domain, but also that the target image can be converted back to the source domain. Then, Choi et al. proposed SartGAN, a model that trains on multiple cross-domain data sets for multidomain transformation. The use of GAN provides a new way of thinking for the field of NST, not only to improve the speed of NST, but also to ensure the quality of generated images. At the same time, GAN, which meets different needs, promotes the application of style migration technology in the real business field.

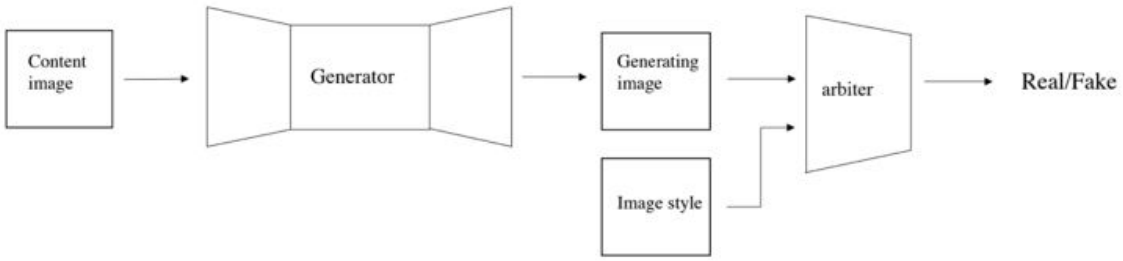


Fig.2. Basic framework of image style transfer based on GAN

3. EVALUATION METHODOLOGY

There are two main methods of NST algorithm, namely qualitative evaluation and quantitative evaluation. Qualitative evaluation depends on the observer's aesthetic judgment, and the evaluation results are related to many factors, such as the age, occupation, educational background, observation conditions and so on. Quantitative evaluation, on the

other being, focuses on precise indicators, such as model generation speed, loss changes, and so on. The combination of qualitative and quantitative assessments can make evaluation more comprehensive and objective.

3.1 APPLICATIONS

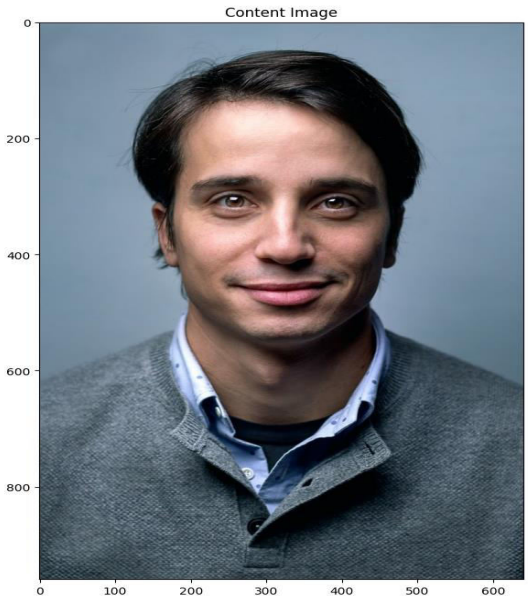
In recent years, NST technology has been widely used in commercial applications as it has been improving in image quality and speed.

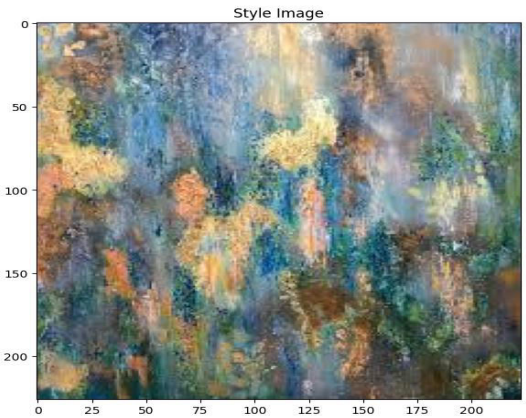
1、 Digital simulation. Style migration can be used for the reconstruction of ink painting, oil painting, traditional Chinese painting and other artistic paintings. With the help of technology, people can easily draw the image style they want without the need for professional skills.

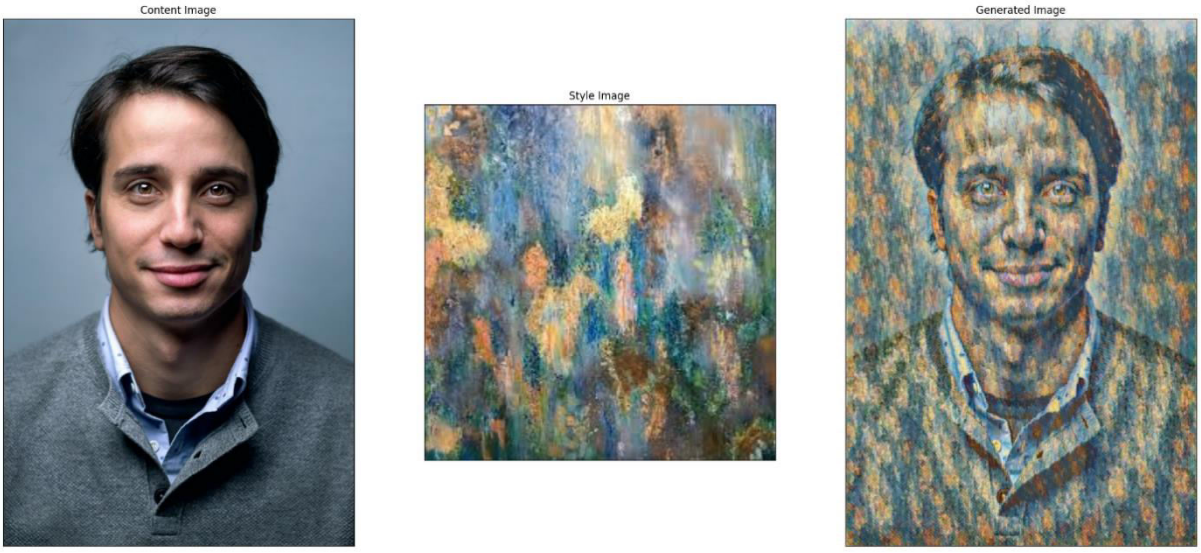
2、 Film and television production. In the film and television industry, if you can use NST instead of manual hand painting, you can greatly improve efficiency and reduce production costs.

3、 Image processing software. The NST algorithm based on deep neural network makes image hard software no longer limited to simply adding a filter to the picture, but uses neural network to learn the style, texture and other details of its works of art.

4.RESULTS AND DISCUSION







The key finding of this paper is that the representations of content and style in the Generative Adversial Networks are well separable. That is, we can manipulate both representations independently to produce new, perceptually meaningful images. To demonstrate this finding, we generate images that mix the content and style representation from two different source images. In particular, we match the content representation of a photograph depicting the riverfront of the Neckar river in Tubingen, Germany and the style ¨ representations of several well-known artworks taken from different periods of art .

6. CONCLUSIONS

Over the past few years, NST has grown rapidly in the arts, academic research, and commercial applications. This paper first summarizes the style migration methods based on image iteration and build model iteration, then introduces the evaluation method and application scenario of the algorithm, and finally summarizes the existing problems and challenges in NST field. In short, image style migration based on deep learning not only promotes the development of computer field, but also receives wide attention in other fields, so the development of NST has important research significance and broad application scenarios.

FUTURE CHALLENGE

Progress in NST is evident, and some algorithms have been used in industry. Although the current results have yielded good performance, there are still some challenges and shortcomings.

1、 The standard for generating image quality assessments is inadequate. At this stage, the evaluation standard in the field of NST is greatly influenced by subjectivity, and a standard

evaluation system is not formed, which is not scientific and standardized enough. You can specify an algorithm as a specification for comparison, and then, when selecting a quality evaluator, cover the population as generally as possible and develop a fixed and comprehensive evaluation form。

2、 The trade-off between speed, flexibility and quality. Image-optimized NST delivers superior performance in quality, but at a higher computational cost. Although PSPM can be stylized in real time, it requires a separate network of aesthetic Chinese types. MSPM increases flexibility by combining multiple styles into one model, but still needs to pre-train the network for a set of target styles. Although ASPM algorithms successfully transmit any style, they are not satisfactory in terms of perceived quality and speed.

3、 The limit of shape change. At this stage, most style migration is only for the texture of the image, color changes, but ignore the impact of its shape. But in a particular scene, people want to generate an image shape as close as the target image shape. For example, when transforming a real face into a comic face, not only does it require a change in style, but it also needs to be exaggerated like an anime character in terms of shape contours. Therefore, the combination of image geometry transformation and style migration is an important way for NST to develop further.

REFERENCES

[1]. Leon Gatys, et al."A Neural Algorithm of Artistic Style." Journal of Vision 16.12(2016):

doi:10.1167/16.12.326.

[2]. Mordvintsev, Alexander, Christopher Olah, and Mike Tyka. "Inceptionism: Going deeper into

neural networks." (2015).

[3]. Berger, Guillaume, and Roland Memisevic. "Incorporating long-range consistency in cnn-based

texture generation." arXiv preprint arXiv:1606.01286 (2016).

[4]. Risser, Eric, Pierre Wilmot, and Connelly Barnes. "Stable and controllable neural texture synthesis

and style transfer using histogram losses." arXiv preprint arXiv:1701.08893 (2017).

[5]. Castillo, Carlos, et al. "Son of zorn's lemma: Targeted style transfer using instance-aware semantic

segmentation." 2017 IEEE International Conference on Acoustics, Speech and Signal

Processing (ICASSP). IEEE, 2017.

[6]. Li, Yanghao, et al. "Demystifying neural style transfer." arXiv preprint arXiv:1701.01036 (2017).

[7]. Li, Chuan, and Michael Wand. "Combining markov random fields and convolutional neural

networks for image synthesis." Proceedings of the IEEE conference on computer vision and

pattern recognition. 2016.

[8]. Johnson, Justin, Alexandre Alahi, and Li Fei-Fei. "Perceptual losses for real-time style transfer and

super-resolution." European conference on computer vision. Springer, Cham, 2016

[9]. Ulyanov, Dmitry, et al. "Texture networks: Feed-forward synthesis of textures and stylized

images." ICML. Vol. 1. No. 2. 2016.

[10].Li, Chuan, and Michael Wand. "Precomputed real-time texture synthesis with markovian

generative adversarial networks." European conference on computer vision. Springer, Cham,

2016.