ISSN: 2414-1895 DOI: 10.6919/ICJE.202211_8(11).0007

Research Progress on Deep Learning based Image Style Migration Methods

Kaige Huang, Ming Lei, Baihong Zhou, Qing Yang

School of College of Electronic Information Engineering, Xi'an Technological University, Shanxi 710032, China

Abstract

Image style migration is a hot research area in the field of computer vision. With the rise of deep learning, the field of image style migration has made a breakthrough. In order to promote the research of image style migration based on nerve network, the main methods and representative works of image style migration based on nerve network are summarised and discussed. The main principles and methods of image wind migration based on neural networks are described in detail, and the application prospects in the field of image wind migration based on neural networks are analysed.

Keywords

Deep Learning; Style Transfer; CNN; GAN.

1. Introduction

Image style migration is an interesting research hotspot in the field of image processing, where the main task is to migrate the style of one image to another, which is also considered as a texture migration problem, i.e. the process of synthesising the source image texture on the target image. Traditional non-parametric texture synthesis methods usually involve resampling on the original texture image to synthesise the new texture, which is simply a process of randomising the original texture without changing its perceptual properties, and can achieve better results in processing images with simple structures, but the results on images with complex colours and textures are hardly satisfactory and cannot meet the needs of practical application scenarios.

With the development of artificial intelligence, Gatys et al [1] have developed an innovative neural network-based image style shifting technique. The core principle of the algorithm is to iteratively optimise the image by pre-training the VGG model [2], with the aim of matching the high-level abstraction feature distribution of the content image and the style image, and then merging the incoming random noise map into the stylised original content image by iterative optimisation. Neural style migration has shown excellent visual results and has attracted a great deal of attention from academia and industry, but the rationale behind neural style migration remains unclear. li et al [3] propose a new interpretation of style migration as a special problem of domain adaptation. Domain adaptation falls under the category of migration learning, which aims to learn models from source data distributions that perform well on different target data distributions, with the key being to measure and minimise the difference between the source and target distributions. They demonstrated mathematically and theoretically that the Gram matrix of a matching feature map is equivalent to the MMD (Maximum Mean Dis-crepancy) [4] statistic of minimising a second-order polynomial kernel, which is commonly used in domain adaptation to measure the difference between two distributions. li et al [3] also experimented with other MMD statistics for different kernel functions, and all obtained This interpretation provides a new perspective on the style migration problem and has inspired much subsequent work on style migration. In Generative Adversarial Networks (GAN) [5], image style migration is an image-to-image translation problem, where images from one domain can be

ISSN: 2414-1895 DOI: 10.6919/ICJE.202211_8(11).0007

transformed to another domain by adversarial training to discriminate the distribution of the image data space.

This paper provides a systematic overview of the origins and development of image style migration, starting with an overview and analysis of the prospects for the application of image style migration, followed by a further discussion of the problems and future directions of image style migration. This provides a solid foundation for further research on style migration and makes some very useful suggestions, concluding with a summary of the future difficulties and directions for style migration.

2. Neural Networks for Image Style Migration

The current algorithms for the wind pattern migration of the nerve network can be divided into two categories: image-based optimization and model-based optimization. The first type transforms the image style by optimising the updated image; the second type optimises the nerve network to generate a model for fast style migration through a forward nerve network, where the optimisation target is the nerve network model.

2.1 CNN-based Image Style Migration

CNN-based image style migration is based on a feature extractor as the core, and approximates the grid loss through feature statistics. The basic framework of CNN-based style migration is shown in Figure 1. Depending on the statistics used to measure style loss, CNN-based methods can be categorised into second-order statistics based on feature distribution and first-order statistics based on feature distribution, which are described below.

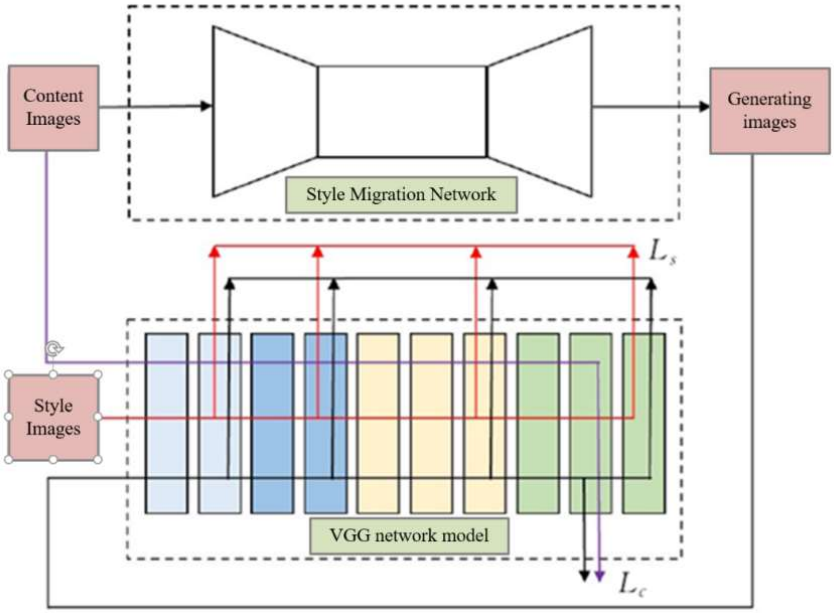

Figure 1. Basic framework diagram for CNN-based image style migration

2.1.1 Second-order Statistics based on the Distribution of Features

Gatys et al [1-2] were the first to discover that convolutional neural networks could extract content feature representations and style feature representations of images separately from arbitrary images. They used pre-trained VGG models as feature extractors and Gram matrices for constructing image features as style representations, and used image iteration to directly optimise the pixels of the initial noisy image to generate a stylised image with the original content and the new style. The The Gram matrix of image features is a second-order statistic of the feature distribution that describes the style of the image through correlations between image features. In the pre-trained VGG model, the shallow features represent the low-level semantic information of the image (such as the edges and colour of the image), while the deeper features of the network capture the high-level semantic information of

ISSN: 2414-1895 DOI: 10.6919/ICJE.202211_8(11).0007

the image. In the selection of feature layers for content loss, the selection of shallow features tends to retain too much content information and affect the stylisation effect, so the selection of deep features for content loss not only achieves the desired stylisation effect, but also preserves the high-level semantic information of the image. For the selection of feature layers for style loss, different depth layers have different granularity of style effect, and choosing multiple feature layers with different depths can obtain better visual effect.

Although the image iteration-based style migration method has good visual effects on synthetic images, this optimisation method is based on changing pixel values by back-propagation at each pixel point of the image, which is a slow and memory-consuming process that greatly limits its extension and application in terms of computational efficiency. To address this problem, the literature [6-7] proposes a fast style migration method, which is three orders of magnitude faster compared to the image iteration-based method, namely the model iteration-based style migration method. The model iteration-based style migration method uses a large number of images to train a style-specific feedforward generative network, shifting the computational burden to the learning phase of the model, and the trained model enables fast style migration in real time, and this method is also the main method used in the application market today. Although fast style migration [6-7] effectively solves the computational efficiency problem, this approach requires a separate model to be trained for each style, making the time cost of extending to other styles too large. To address this problem, Chen et al [8] proposed a method for generating multiple styles in a single model by binding each style to a separate set of convolutional layer parameters, which can be learned jointly to obtain a style library storing different styles. Zhang et al [9] introduced a new mutual matching layer in the generative network, which can learn second-order statistics that directly match style features in the generative network. Li et al [10] designed a texture selection network to generate the style features of the corresponding textures, which were combined with the features of the content images respectively to achieve multi-style migration. They also found that using the covariance matrix of the features instead of the Gram matrix to characterise styles improved the artefacts and blending problems that occur in the generated images. According to the number of styles generated by a single model, the literature [11] classifies three iterative model-based style migration methods, single style generation model, multiple style generation model and arbitrary style generation model. The above-mentioned model iteration-based methods cannot be generalised to unseen styles, although they have high conversion efficiency. The literature [12] proposes a simple and effective generic style migration method that can be generalised to arbitrary unseen styles and does not require training on these styles.

This method is the first learning-free style migration method that implements style migration of images by embedding a pair of linear transforms in the image reconstruction network. Different levels of VGG networks are first selected as encoders and their symmetric decoders are trained to form multiple image reconstruction networks. A generic style transformation is achieved by recursively applying the Whitening and Coloring Transformation (WCT) to each image reconstruction network. The WCT enables direct matching of the covariance matrix of content image features and style image features, in a similar spirit to the method described above for optimally matching the Gram matrix. This linear transformation is identical to the CORAL (CORre-lation ALignment) [13] method in domain adaptation, where the CORAL method first whitens the data in the source domain and then re-associates it to the target domain data; this operation actually aligns the second-order statistical information of the source and target domain data distributions. Xie Bin et al [14] proposed a style migration model based on correlation alignment, where the samples performing correlation alignment are not the image data itself, but the features of the image. Due to the high dimensionality of the depth feature vector, it is computationally expensive for WCT to perform matrix decomposition operations directly. To solve this problem, Li et al [15] used a data-driven approach to learn the output feature transformation matrix instead of direct matrix operations, which is more flexible and efficient.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202211_8(11).0007

2.1.2 First-order Statistics based on the Distribution of Features

Li et al [3] provide an enlightening understanding of image style migration in terms of domain adaptation, finding that the statistics (e.g., mean and variance) in the Batch Normalization (BN) layer [16] contain features from different domains, and that style migration can be achieved by simply adjusting the mean and variance of matched image features in the channel direction. Ulyanov et al [17] found that using an Instance Normalization (IN) layer [18] instead of BN in a fast style migration network not only accelerated the convergence of the network, but also allowed a lower style loss to be achieved during the training process, resulting in a better visual quality. loss, obtaining better visual results. They argue that the superiority of IN lies in its ability to weaken the contrast information between content images in the network, thus making the learning of the network simpler. Another explanation for this proposed in the literature [19] through experiments is that IN itself has the ability to normalise styles to a target style for each style, allowing the rest of the network to focus on the learning of content information. Building on IN, Dumoulin et al [20] found that a multi-style generative network could be trained using different affine coefficients in the normalisation layer of a style migration network. They proposed Conditional Instance Normalization (CIN), where all the convolution parameters in the network are shared among multiple styles, and a normalisation layer with different affine parameters can convert the input content images into different styles.

Each style enables the model to scale to multiple styles by binding to the parameters of the normalization layer in the network, but is limited by the limited number of styles covered, the inability to generalize to untrained styles, and the number of additional parameters growing linearly with the number of styles. To escape this limitation, subsequent work by Ghiasi et al [21] designed a style prediction network by training a large number of images to predict the affine parameters of the CIN in the generative network, a data-driven approach that provides the model with the ability to predict other untrained styles. Also inspired by IN, a simple and effective Adaptive Instance Normalization (AdaIN) layer is proposed in the literature [19], which has no parameters to learn and which adaptively computes the affine parameters of the normalization layer from the input style images, enabling arbitrary style transformation in real time. Given a content input and a style input, the method uses a VGG network as a fixed encoder, after AdaIN adjusts the mean and variance of the content image features in the channel direction to match the mean and variance of the corresponding channel of the style image features, and finally the decoder learns to transfer the matched content features to the image space to complete the style migration of the image. Similar to the AdaBN (Adaptive Batch Normalization) [22] method in domain adaptation, AdaIN can effectively combine the content of the former with the style of the latter by matching the first-order statistics of aligned content and style image features, but it is difficult to synthesize a complex style with rich details and local complexified styles with rich detail and local structure. To enhance the adequacy of style migration in the literature , Park et al [23] were able to effectively balance local and global styles by introducing a novel style attention network that learns semantic correlations between content features and style features.

Shen et al [24] introduced meta-learning to the field of style migration, using the Hyper network [25] approach in meta-learning. The idea of the Hyper network is to use one network to generate the parameters of another network, dynamically generating the parameters of the style transformation network by learning first-order statistics of the feature distribution of the input style image. Their approach provides an effective solution for real-time arbitrary style migration, and the size of the meta-learning-generated model is only a few hundred KB, allowing it to run in real-time on mobile devices. resource-constrained environments. They propose a method for arbitrary style migration based on the lightweight architecture of MobileNet, introducing a Dynamic Instance Normalization (DIN) module to encode styles as learnable convolutional parameters, which is combined with a lightweight content encoder for fast style transformation. The framework is shown in Figure 3, where the DIN module contains IN and Dynamic Convolution, whose parameters change adaptively according to the different styles, allowing more accurate alignment of feature statistics for complex styles and allowing more flexible arbitrary style migration while keeping computational costs low.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202211_8(11).0007

2.2 GAN-based Image Style Migration

2.2.1 Generating Adversarial Networks

The basic framework structure of the GAN is shown in Figure 2, and the generative adversarial network was proposed by Goodfellow et al. in 2014 [1]. Its design is inspired by the idea of gaming in game theory, and this idea is modelled in deep neural networks with generators and discriminators. Figure 2 shows the GAN model, where the generator captures the distribution of data in random data and is used to generate images, while the discriminator uses real images to determine whether the generated images are true or not. The discriminator is essentially a binary classifier, where a true discriminator is denoted by a 1 and a false discriminator is denoted by a 0 for the generated image. When the discriminator is false, it returns the discriminator result to the generator, which continues to learn and optimise the discriminator result, generating a more realistic image to confuse the discriminator, cycling through the above steps until the resulting image appears to the discriminator to be indistinguishable from the real image. The game starts from zero until the discriminator discriminates at 0.5, when the two are balanced.

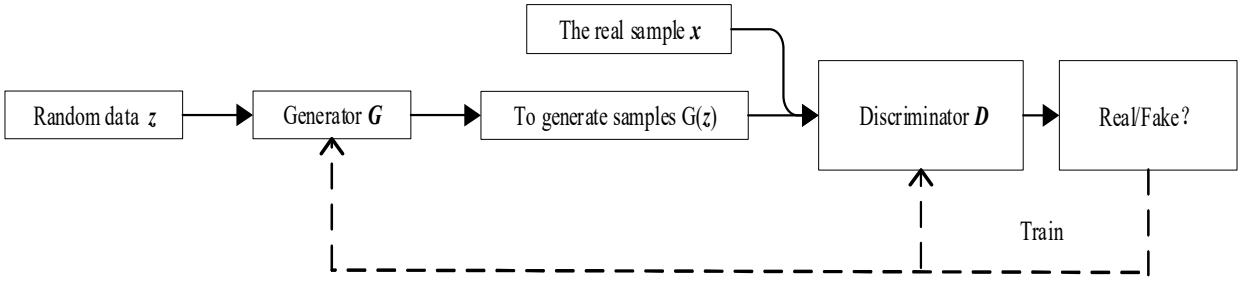

Figure 2. Basic framework diagram for GAN

The training loss of the GAN consists of the adversarial loss and the generative loss, and the objective function $V(D,G)$ is expressed as follows.

$$\min_{G}\max_{D} V(D,G) = E_{x\sim p_{data}(x)}\left[\log D(x)\right] + E_{z\sim p_z(z)}\left[\log\left(1-D\left(G(z)\right)\right)\right]$$

In most cases, discriminator loss and generator loss do not converge in parallel. For example, when the discriminator accuracy is high, its loss may quickly drop to zero, resulting in the inability to provide useful information for the training of the generator, making the training of GAN unstable and also losing the meaning of adversarial. In GAN, image style migration is considered as the process of transforming one class of images to another. pix2pix model proposed by Isola et al [26] as a representative work on image to image translation uses a large number of paired images for supervised training to obtain a one-to-one image translation network, which performs the image style migration task excellently. Although pix2pix can achieve realistic image translation, the training of the model requires a large amount of paired image data, which greatly limits its generalisation and application. To break this limitation, Zhu et al [27] proposed an unsupervised adversarial network, CycleGAN, which contains two pairs of generative adversarial networks designed to achieve bidirectional domain translation, introduces cyclic consistency to remove pairwise constraints between domains, and better preserves the image content structure.

2.2.2 CyceleGAN

The use of cyclic consistency loss and pixel-by-pixel differences in image space as image content loss in CycleGAN allows the content information of the generated images to be over-represented, resulting in the inability to migrate abstract styles, such as artistic styles, well. To better learn artistic

ISSN: 2414-1895 DOI: 10.6919/ICJE.202211_8(11).0007

styles, Sanakoyeu et al [28] introduce a style-aware content loss in GAN that is able to learn the same class of artistic styles rather than being limited to one instance of a style. ma et al [29] observe the feature vectors of content and style images in projection space and find that the features of content and style images are essentially separable in the initial state. Their proposed double consistency loss can learn the relationship between content and style images while maintaining semantic and stylistic consistency. Also focusing on content- and style-aware stylisation of images, Kotovenko et al [30] designed a content transformation model in an adversarial network to learn how to change the details of content during style migration between content and style images with similar content information.

2.3 Comparative Analysis

For image style migration tasks, how to describe and compute styles is a critical issue. Thanks to the feature extraction capability of deep convolutional neural networks, CNN-based image style migration methods can flexibly and efficiently achieve image style migration by extracting abstract feature representations of an image and using the statistics of the feature distribution as a description of the image style. Although this description method can characterize style well, it relies on a feature extraction network with large parameters, which is a pressing problem at present.

Unlike these methods, the GAN-based image style migration method brings a new approach to style description through the mechanism of adversarial learning. In GAN, there is no need for any pre-designed description to compute styles, and the discriminator can compute styles implicitly by fitting image data distributions to achieve style migration of images. Fitting the distribution of image data through adversarial training can result in more realistic style migration, which demonstrates the GAN's ability to understand and perceive image data. Compared to CNN-based style migration methods, GAN is better at generating images, but the style migration process is less controllable and the training of adversarial networks is prone to gradient disappearance and model collapse, which has the disadvantage of being difficult to train.

3. Improvement and Extension Work

The emergence of deep learning-based image style migration methods has greatly The emergence of style migration methods based on deep learning has greatly contributed to the development of the style migration field, and a large number of scholars and researchers have of scholars and researchers have started to focus on image style migration techniques. In recent years, many research results have emerged. Some work has focused on the design of new image style migration methods, while some work is devoted to improving and extending existing methods. This chapter presents some of the improved and extended work on image style migration. This chapter presents some of the improved and extended work on image style migration.

3.1 Texture and Semantic Improvements

Risser et al [31] found that the use of a single Gram style loss in synthetic texture, and also found that the reason for its instability was that feature maps with different same Gram matrix for feature maps with the same mean and variance. Therefore Therefore, they improved the stability of the synthetic texture by additionally matching the histograms of the features to the original network. The histograms of the additional matching features were therefore used to improve the stability of the synthetic texture. In addition, the Gram loss is not satisfactory for rendering images with regular textures and symmetric structures. Berger et al. [32] have shown that Gram loss can be improved by computing Gram loss from a flattened feature map. Wang et al [33] added an orthogonal noise matrix to the depth feature transform to perturb the image. This approach can perturb the style migration process without affecting the quality of the original style migration. significantly improve the diversity of the generated images without affecting the quality of the original style migration. texture diversity without affecting the quality of the original style migration.

In the literature [34], Laplace loss was introduced in the method of Gatys et al [1] to remove artefacts and prevent Liu et al. Liu et al [35] added a depth estimation network to calculate the depth

ISSN: 2414-1895 DOI: 10.6919/ICJE.202211_8(11).0007

information of the generated image, which can This effectively maintains the coherence and overall spatial layout of the original content image. In a subsequent work [36], an edge detection network was added to preserve local detail structure, trade-offs are made to preserve the global and local structure of the generated image.

3.2 Control of Perceptual Factors

The literature [37] improves on existing methods for image style migration, working on on the control of perceptual factors, such as content style interpolation, colour information preservation and stroke size control. They propose two methods that can perform image style migration while preserving the original They propose two ways to perform image style migration while preserving the original colours. Zhang et al. used different sizes of style images to train the network to learn different stroke sizes of strokes according to the size of the input image. Similar to the idea of the literature [37], Wang et al [38] used a sub-network containing multiple model with multiple sub-networks for hierarchical training at different scales, with the stylised The results of each sub-network are up-sampled as the input of the next sub-network, and this coarse-to-fine This coarse-to-fine stylisation process can produce large and fine strokes in high-resolution images. The two methods of scaling the image and training multiple models by resizing the image and training multiple models, can change the size of the stylised strokes, but also entails a loss in quality and cost. The two methods of scaling the image and training multiple models can change the size of the stylised strokes, but they also entail a loss in quality and cost. To address this problem, Jing et al [39] propose an algorithm that can The algorithmic model of merging multiple stroke sizes was introduced by Jing et al. The model introduces an adaptive receptive field module, which learns different stroke sizes through different receptive fields. strokes without sacrificing quality and efficiency. This enables continuous stroke size control of a single model without sacrificing quality and efficiency.

3.3 Realism Style Migration

Luan et al [40] proposed the first realistic style migration algorithm based on deep learning. A deep learning-based realistic style migration algorithm is proposed by Luan et al [40], which is divided into two stages: first, adding semantic segmentation to the style migration process to achieve semantic style migration, and second, introducing realistic regularisation in the post-processing of the generated image to optimise the image. The second stage is to optimise the image by introducing a regularisation term in the post-processing of the generated image. Again using a two-stage optimisation, Li et al.'s [41] realistic style migration method is based on the depth feature transform , where they replace the pooling and upsampling layers in the WCT with pooling and inverse pooling layers with coordinate information, which improves to some extent the loss of spatial information due to multiple pooling. In the post-processing stage, they used a smoothing step based on a streamwise sorting algorithm to generate images by smoothing the image similarity matrix of the content images. In the post-processing stage, they used a smoothing step based on a stream sorting algorithm to generate images by smoothing the image similarity matrix of the content images. Similar to the improvement in [42], Yoo et al [43] replaced the pooling and upsampling layers in the WCT with wavelet pooling and wavelet inverse pooling layers, which, combined with the properties of wavelet information, can preserve the spatial information of the image with almost no loss. information. This method is the first end-to-end realistic style migration method that They use a recursive network instead of the multi-level stylisation strategy in WCT to reduce artefacts. They use a progressive network instead of the multi-stage stylisation strategy in WCT to reduce artefacts, maintain the structure of the image without distortion and achieve The method uses a progressive network instead of the multi-level stylisation strategy used in WCT to reduce artefacts, maintain the structure of the image without distortion and achieve realistic style migration.

4. Conclusion

This paper first provides a comprehensive description of image style migration methods using deep learning, including convolutional neural network-based and generative adversarial network-based

ISSN: 2414-1895 DOI: 10.6919/ICJE.202211_8(11).0007

image style migration methods, and analyses their advantages and disadvantages, then introduces some improved and extended research work, and finally summarises the challenges of current research. This paper categorises the existing research work based on the main principles of image style migration, which can help beginners and researchers in this field to grasp the current research direction and deepen their understanding of the research.

References

[1] GATYS L, ECKER A S, BETHGE M.Texture synthesis using convolutional neural networks[C]// Advances in Neural Information Processing Systems, 2015: 262-270.

[2] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv: 1409. 1556, 2014.

[3] LI Y, WANG N, LIU J, et al.Demystifying neural style transfer[C]//Twenty- Sixth International Joint Conference on Artificial Intelligence(IJCAI), 2017: 2230-2236.

[4] TZENG E, HOFFMAN J, ZHANG N, et al.Deep domain confusion: maximizing for domain invariance[J]. arXiv: 1412.3474, 2014.

[5] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Advances in Neural Information Processing Systems, 2014: 2672-2680.

[6] JOHNSON J, ALAHI A, FEI- FEI L.Perceptual losses for real- time style transfer and super- resolution [C]// European Conference on Computer Vision, 2016: 694-711.

[7] ULYANOV D, LEBEDEV V, VEDALDI A, et al. Texture networks: feed- forward synthesis of textures and stylized images[C]//Proceedings of the International Conference on Machine Learning(ICML), 2016: 1349-1357.

[8] CHEN D, YUAN L, LIAO J, et al.Stylebank: an explicit representation for neural image style transfer[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2770-2779.

[9] ZHANG H, DANA K.Multi- style generative network for real- time transfer[C]//Proceedings of the European Conference on Computer Vision(ECCV), 2018: 349-365.

[10] LI Y, FANG C, YANG J, et al.Diversified texture synthesis with feed- forward networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 266-274.

[11] JING Y, YANG Y, FENG Z, et al.Neural style transfer: a review[J].IEEE Transactions on Visualization and Computer Graphics, 2019, 26(11): 3365-3385.

[12] LI Y, FANG C, YANG J, et al.Universal style transfervia feature transforms[C]//Advances in Neural Information Processing Systems, 2017: 386-396.

[13] SUN B, FENG J, SAENKO K.Return of frustratingly easy domain adaptation[J].arXiv: 1511.05547, 2015.

[14] SIMONYAN K, ZISSERMAN A.Very deep convolutional networks for large-scale image recognition[J]. arXiv: 1409. 1556, 2014.

[15] LI X, LIU S, KAUTZ J, et al.Learning linear transformations for fast image and video style transfer[C] //Proceedings of the IEEE Conference on Computer Visionand Pattern Recognition, 2019: 3804-3812.

[16] IOFFE S, SZEGEDY C.Batch normalization: accelerating deep network training by reducing internal covariate shift[J].arXiv: 1502.03167, 2015.

[17] ULYANOV D, VEDALDI A, LEMPITSKY V.Improved texture networks: maximizing quality and diversity in feed- forward stylization and texture synthesis[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4105-4113.

[18] ULYANOV D, VEDALDI A, LEMPITSKY V. Instance normalization: the missing ingredient for fast stylization[J]. arXiv: 1607.08022, 2016.

[19] HUANG X, BELONGIE S.Arbitrary style transfer in real- time with adaptive instance normalization[C] //Proceedings of the IEEE International Conference on Computer Vision, 2017: 1510-1519.

[20] DUMOULIN V, SHLENS J, KUDLUR M.A learned representation for artistic style[C]//International Conference on Learning Representations(ICLR), 2017.

[21] GHIASI G, LEE H, KUDLUR M, et al.Exploring the structure of a real-time, arbitrary neural artistic stylization network[C]//Proceedings of the British Machine Vision Conference(BMVC), 2017.

[22] LI Y, WANG N, SHI J, et al.Revisiting batch normalization for practical domain adaptation[J].arXiv: 1603. 04779, 2016.

[23] PARK D Y, LEE K H.Arbitrary style transfer with style- attentional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 5873-5881.

[24] SHEN F, YAN S, ZENG G.Neural style transfer viameta networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8061-8069.

[25] HA D, DAI A, LE Q V.Hypernetworks[J].arXiv: 1609.09106, 2016.

[26] JING Y, LIU X, DING Y, et al.Dynamic instance normalization for arbitrary style transfer[C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2020: 4369-4376.

[27] ISOLA P, ZHU J Y, ZHOU T, et al.Image- to- image translation with conditional adversarial networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 5967-5976.

[28] ZHU J Y, PARK T, ISOLA P, et al.Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 2242-2251.

[29] SANAKOYEU A, KOTOVENKO D, LANG S, et al. A style-aware content loss for real-time hd style transfer[C]//Proceedings of the European Conference on Computer Vision(ECCV), 2018: 715-731.

[30] MA Z, LI J, WANG N, et al.Semantic- related image style transfer with dual-consistency loss[J]. Neurocom-puting, 2020, 406: 135-149.

[31] KOTOVENKO D, SANAKOYEU A, MA P, et al.A content transformation block for image style transfer[C]// Proceedings of the IEEE Conference on ComputerVision and Pattern Recognition, 2019: 10024-10033.

[32] CHOI Y, CHOI M, KIM M, et al.StarGAN: unified generative adversarial networks for multi-domain image-to- image translation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8789-8797.

[33] CHEN X, XU C, YANG X, et al.Gated-GAN: adversarial gated networks for multi- collection style transfer[J]. IEEE Transactions on Image Processing, 2018, 28(2): 546-560.

[34] XU Z, WILBER M, FANG C, et al.Adversarial training for fast arbitrary style transfer[J].Computers & Graphics, 2020, 87: 1-11.

[35] CHO W, CHOI S, PARK D K, et al.Image- to- image translation via group- wise deep whitening- and-coloring transformation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 10631-10639.

[36] ZHENG Z, LIU J.P2-GAN: efficient style transfer using single style image[J].arXiv: 2001.07466, 2020.

[37] RISSER E, WILMOT P, BARNES C.Stable and controllable neural texture synthesis and style transfer using histogram losses[J].arXiv: 1701.08893, 2017.

[38] BERGER G, MEMISEVIC R.Incorporating long- range consistency in CNN- based texture generation[J]. arXiv: 1606.01286, 2016.

[39] WANG Z, ZHAO L, CHEN H, et al.Diversified arbitrary style transfer via deep feature perturbation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 7786-7795.

[40] LUAN F, PARIS S, SHECHTMAN E, et al.Deep photo style transfer[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6997-7005.

[41] LI Y, LIU M Y, LI X, et al.A closed-form solution to photorealistic image stylization[C]//Proceedings of the European Conference on Computer Vision(ECCV), 2018: 468-483.

[42] AN J, LI T, HUANG H, et al.Real-time universal styletransfer on high- resolution images via zero- channel pruning[J].arXiv: 2006.09029, 2020.

[43] YOO J, UH Y, CHUN S, et al.Photorealistic style transfer via wavelet transforms[C]//Proceedings of the IEEE International Conference on Computer Vision, 2019: 9035-9044.