Research On Image Stylization Methods Based on Deep Network

Zhuoqun Li†, 1, Zihan Yan†, *, 2

1 NOIC ACADEMY, WuHan, China

2 Shenzhen College of International Education ShenZhen, China

* Corresponding Author Email: s21035.yan@stu.scie.com.cn

†All authors contributed equally

Abstract. For the purpose of promoting the research of deep neural networks, image stylization has received extensive attention. Researchers have proposed a huge number of image stylization methods based on deep neural networks. In this paper, image stylization methods are divided into two categories: image stylization method based on reference and based on domain. A series of stylization methods. Several classicals methods are summarized, and the performance of the two categories of stylization methods on the common data sets of image stylization tasks are compared respectively, the developments of image stylization in video, human face and brushwork are proposed, the current progress of image stylization and the future research direction are mentioned. At the end of the article, the full text is summarized and summarized.

Keywords: Deep neural network, image stylization, reference-based Image stylization, domain-based image stylization.

1. Introduction

Image stylization is an algorithm that maps the style of an artistic image onto other natural images, so that the original natural image has the artistic type of the artistic image while retaining the original semantic message. The concept of image stylization comes from the fact that people are attracted by the works of some art masters and desire to have the same artistic style of images themselves. However, redrawing images of special styles requires a large amount of investment and resource loss of relevant technical personnel, so some researchers begin to study the corresponding algorithm to complete the task of graphics style.

Gatys et al. [1] first proposed the concept of image stylization and the image stylization method Based on Optimization-Based, which was also the main method used in the initial study of image stylization. For the sake of optimize the speed and potency of stylization, Johnson et al. successively put forward deep neural network to replace the slow optimization process, and proposed corresponding content loss function and style loss function for network cultivation, laying a foundation for image stylization research based on forward deep neural network [2]. In addition, the proposal of Generative Adversarial Network (GAN) and its outstanding performance in various image generation and image-to-image conversion tasks have led to the proposal of many GAN-based image stylization methods to optimize the effect and property of actual ways. Greatly promote the research process of image stylization. Therefore, image stylization method be based upon forward deep neural network and image stylization method on th basis of GAN have become the main methods used in image stylization research at present.

The goal of image stylization to change an image range from the original style to the reference image or the style defined by the target domain while maintain the same semantic content through stylized models, that is, the original image with a certain style is transformed into a stylized image of another style through the stylized model. The powerful style conversion ability and diverse and interesting stylization results make the image stylization technology have extensive use of applications in the area of social communication, artistic creation and data production. The development of image stylization research has brought a lot of fun and convenience to the public life, involving life, work and other aspects. Image stylization is a form of artistic expression, because its input and output are pictures, which reproduce the painting techniques of artists and generate ornamental works of art. In this paper, the methods of image stylization are summarized and compared according to image stylization methods based on reference and domain.

2. Image Stylization Method based on Reference

The image on basis of reference stylization method needs to solve the problem of how to influence the original image style into the reference image style while holding the semantic substance unvaried. The key of the problem is how to obtain the semantic content message and style message of the image and to complete the transformation of the target style. According to the feature processing methods in the process of stylization in recent years, the reference-based image stylization methods can be included into two types: image stylization methods on basis of feature decoupling and image stylization methods on basis of feature fusion [2].

2.1. Image Stylization Method Based on Feature Decoupling

The image stylization method based on feature decoupling is to decouple the features of the image into content features containing semantic content information and style features containing style information, and then exchange the style characteristics of the initial image and the reference image to complete the style conversion. This kind of method was widely used in the early stage of image stylization research, and the image stylization process is shown in Figure 1.

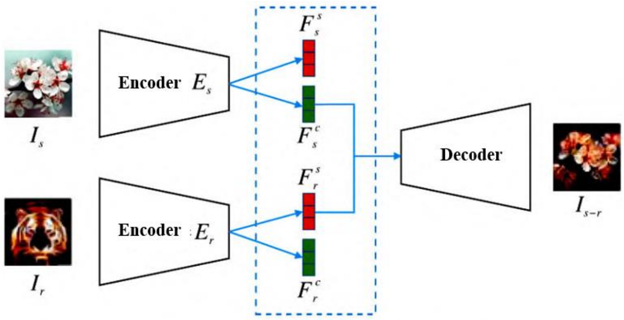

Figure 1. Image stylization process based on feature decoupling [3]

In the process of stylization, the original image Is gets the style feature Fss and content feature Fcs through the encoder Es, and the reference image Ir gets the style feature Fsr and content feature Fcr through the encoder Er. Then the Fcs and Fsr are exchanged, and the stylized image IS-R with Ir style Is obtained through the decoder.

Gatys et al. first proposed the concept of NeuralStyleTransfer (NST) and the corresponding image stylization method based on optimization, which is also the first image stylization method on basis of characteristic decoupling [1]. In this approach, combining the message characteristics of the original image and the style characteristics of the reference image, the generated stylized image is iterated constantly to optimize, and the Gram matrix is used to estimate the stylistic analogy between the stylized image and the reference image. In the field of image stylization, scholars have proposed image stylization methods based on feature decoupling. Li et al. arranged to use Markov Random Field (MRF) regularize replaced of Gram matrix to measure the type similarity between images [4]. MRF considered the semantic correlation between stylized images and reference images when calculating, and used the patch-based way to match and calculate the style similarity. The model optimization on basis of MRF further optimizes the character of the creating stylized images.

Although above methods have obtained high quality stylized images, they are all based on optimization and need to go through several iterations to obtain high quality stylized images, which is very time-consuming in practical application.

The application of forward deep neural network makes the speed of image stylization increase by 2~3 orders of magnitude to achieve real-time image stylization, and the method based on forward deep neural network has gradually become the mainstream method of image stylization.

Since it is difficult to collect image data with both styles, the training data sets used at present are Unpaired, and the network needs to use unsupervised training. Therefore, by introducing the training method of CycleGAN (Cycle-Consis-tentAdversarialNetworks), the generated stylized images Is-r and Ir-s can be obtained from the reconstructed original images Is-r and through the same coding and generation process Ir-s then gets the reconstructed original image through the same coding and generating process, to realize the training of coder and generator. When reasoning, according to the different target style, select the corresponding encoder and generator to complete the target style conversion.

Although the image stylization method based on feature decoupling has obtained high quality stylized images, it is difficult to "cleanly" decouple the style characteristics and content characteristics through the decoupling method of image coding respectively, resulting in the content information contained in the style characteristics, and the content characteristics contained the style information, so that the original image converted from the stylized image will retain more original images Style information or introduce more reference image content message [3]. For one thing, as the stylistic information of the reference image is quite different from that of the initial image, the creating stylized image will be under-stylized due to the retention of more stylistic information of the original image. For another, when the content information of the reference image is quite different from that of the initial image, the creating stylized image will have the problem of excessive stylization due to the content information of more styles of images.

2.2. Image Stylization Method Based on Feature Fusion

The image stylization method based on feature fusion is to collect the characteristics of the original image and the reference image respectively, and then adopt the feature fusion module to get the characteristics containing the semantic substance message of the original image and the style message of the reference image after the fusion of the two characteristics, and complete the style conversion. This kind of method is the mainstream method of reference-based image stylization at present, and the image stylization process is shown in Figure 2.

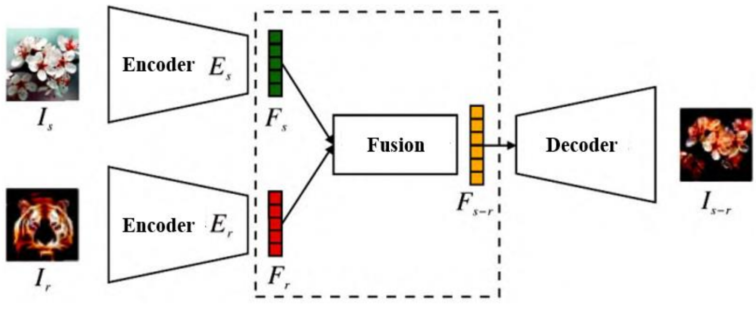

Figure 2. Image stylization process based on feature fusion [3]

In the process of stylization, the original image Is passing the encoder Es to get the image feature Fs, and the reference image Ir passes the encoder Er to get the image feature Fr, then Fs and Fr pass the feature fusion module to get the FS-R after fusion, and then through the decoder to get the stylized image IS-R. Here, the encoder Es and Er can be the same encoder, and the pre-trained VGG network is generally adopted.

Chen et al. [5] first proposed a blocke-based feature fusion method with the early training VGG network to pick up the characteristic picture in the original image and the characteristic picture in the reference image respectively. They then divide each of the overlapping characteristics of equal size and, and swap each original image feature with the most relevant reference one. It then uses the swapped feature block to build a new feature map, giving it a reconstructed feature map. Finally, input feature into the generator to get a stylized image.

Although the method of feature fusion proposed by Chen et al. using block-based feature matching and exchange can obtain high-quality stylized images, this method of feature fusion based on feature matching and exchange needs quantities of computation and is time-consuming [5]. In addition, when there is a large gap among the reference image and the original image, the direct exchange of matched characteristics may introduce more semantic content information of the reference image, which makes the stylized image lose more semantic content information of the original image [6].

In order to solve these problems, such as Li put forward a new way of feature fusion, WCT (Whiting and Coloring Transforms), feature level on the representation of style, the style should be represented by a covariance matrix of image characteristics, and put forward using the original image based on the theory of characteristic figure of covariance [7]. The matrix matches the covariance matrix of the reference image feature map for feature fusion. The proposal of WCT speeds up the processing speed of image stylization based on feature fusion, and obtains higher quality stylized images, which lays a foundation for the subsequent research on image stylization based on feature fusion.

In order to further optimize the quality of the stylized image, Huang et al. proposed the feature fusion method, AdaIN (Adaptive Instance Normalization), which is used in image stylization [6]. They believe that the style of the image is determined by the statistical mean and variance of its feature graph, and propose to modulate the original image feature graph by using the mean and variance of the reference image characteristic map, so that the variance and mean of the modulated feature graph and the reference image feature graph are consistent. AdaIN proposed to greatly accelerate the processing speed of image stylization to achieve the level of true-time stylize image and further optimize the quality of stylized image. In the meantime,AdaIN's proposal also provides a new research direction and ideas for image stylization methods based on feature fusion. Scholars have proposed new methods based on this, and AdaIN has become the main feature fusion method used in image stylization methods based on characteristic amalgamation.

For the sake of further optimize the accuracy and robustness of characteristic fusion in the stylization process, Park et al. proposed SANet (Style Attentional Network) [8]. Firstly, feature maps Fs and Fr of the original image Is and consult image Ir are extracted using the training in advance of VGG network, after that characteristic fusion is implemented using SANet. The proposals of AdaIN [6], WCT [7] and SANet [8] are of great significance for image stylization methods based on feature fusion, and the subsequent proposed methods are basically feature fusion based on these three methods. For the sake of keep the semantic substance information of the intrinsic image as far as possible and optimize the network structure in the stylization process, an et al. [9] adopted the completely reversible encod-decoder structure ArtFlow, that is, the decoder is the reverse process of the encoder. The reversibility of the whole process enables the semantic content information of the original image in the process of stylization to be preserved more and optimize the mass of the stylized image.

Although the image stylization way on basis of feature fusion can solve the problems of under-stylization and over-stylization in the process of stylization to a certain extent, and obtain good quality stylized images, it all the same cannot completely solve the above problems when a big gap between the reference image and the original image. This is also the focus of future research on reference-based image stylization. In conclusion, the characteristics of the representative methods for reference-based image stylization can be summarized as demonstrated in Table 1 below.

Table 1. Characteristics of representative methods in reference-based image stylization

Methods	Characteristics
Optimization-Based	The speed is slow, the iteration times of different stylized tasks need to be manually selected, and the model effect is stable
WCT	The semantic content is easily lost in the process of stylization, and the model effect is unstable and fast, but it cannot reach the real-time level
AdaIN	During the stylization process, there may be excessive stylization. The model effect is relatively stable and the speed reaches the real-time level
SANet	In the process of stylization, more semantic content information is retained, but there is still the possibility of loss, and the speed reaches the real-time level
ArtFlow	In the process of stylization, semantic content information is retained, and the speed reaches the real-time level

2.3. Comparison of Image Stylization Methods Based on Reference

This section mainly compares reference-based image stylization methods in terms of running speed and effect, where running speed refers to the time required to made a stylized image, the result is a generated stylized image quality. There is no fixed data set for the comparison of this kind of method. In this paper, the original image and reference image commonly used in relevant papers are selected as the test data for comparison. This paper uses chose several classical methods to compare the running speed and effect. The running speed comparison result are shown in Table 2, in which the figure in bold represents the optimal value.

As shown in Table 2, Optimization-Based optimization takes the longest time to generate a stylized image because it requires several iterations of optimization. Using the forward deep neural network method WCT, AdaIN, Avatar-Net, SANet and ArtFlow to generate a stylized image takes less time, which optimizes by 2~3 orders of magnitude compared with Optimization-Based method. The comparison shows that the use of forward deep neural network greatly optimizes the processing speed of image stylization. Compared with WCT and Avatar-Net, the matrix operation required by AdaIN, SANet and ArtFlow for feature fusion is simpler, and the time required to generate a stylized image is shorter, enabling real-time image stylization.

Table 2. Running time comparison of different methods

Methods	Image Size 256 × 256	Image Size 512 × 512
Optimization-Based	15.863	50.804
WCT	0.689	0.997
AdaIN	0.011	0.039
Avatar-Net	0.248	0.356
SANet	0.012	0.042
ArtFlow	0.028	0.065

The stylized images generated by SANet and ArtFlow can fully retain texture details such as forehead wrinkles and hair end of the original image. The stylized images generated by CT and AdaIN all have the problem of semantic content information loss to varying degrees, which indicates the effectiveness of SANet and ArtFlow in preserving semantic content information in the process of stylization. Further, it can be found that the stylized images and reference images generated by WCT, AdaIN, Avatar-Net, SANet and ArtFlow have higher style consistency, while the stylized images and reference images generated by Optimization-Based have greater fluctuations in style consistency, indicating Optimizati The quality of the stylized image obtained on-Based will be more dependent on the frequency of iterations During the process of optimization, and the number of iterations required in the conversion to different styles is quite different.

3. Image Stylization Method Based on Domain

Image stylization method based on domain converts two style defined domains in order to transform the styles, realize stylized images. These methods usually use GAN.

Because based on domain - based image stylization method, from the source domain to the target domain, their conversion is unpaired, Zhu and his teammates first came up with CycleGAN, which used for unpaired image converts to image [10]. There are two steps for CycleGAN: style conversion from source domain to target domain to source domain and style conversion from target domain to source domain to target domain. These two processes are similar to each other.

CycleGAN is a new idea for the image stylization method based on domain, it also lays the foundation for the study of domain - based image stylization. Liu and his teammates based on CycleGAN and the hypothesis of characteristics of source domain and target domain are latent space sharing, proposed UNIT (Unsupervised Image-to-Image Translation) [11]. UNIT contains two parallel grid training processes: the reconstruction of images and the style conversion of images. UNIT greatly promotes the study process of image stylization as it provides a new view on stylized images: domain feature latent space sharing.

Kim and his teammates based on CycleGAN, proposed U-GAT-IT (Unsupervised Generative Attentional Networks with Adaptive Layer-Instance Normalization for Image-to-Image Translation), which introduced module of attention and AdaLIN (Adaptive Layer-Instance Normalization) [12]. The module of attention guidance model uses the information taken by auxiliary classifier to make the characteristic pattern pay more attention to the region that makes the most difference between source domain and target domain, as this area is also the most susceptible to deformation, to promote the deforming during the stylization process. AdaLIN means when performing regularization, two modes of regularization, regularization by layer and regularization by instance, are combined together to control the shape and grain of the stylized image more precisely.

U-GAT-IT offers an efficient way for the image stylization tasks that require deformations based on domain - based image stylization method and CycleGAN. This optimizes the quality of the stylized images and solves the problem of low-quality stylized output due to the constraints on cyclic consistency loss in CycleGAN. For the same purpose of solving the low-quality output due to the constraints on cyclic consistency loss, Nie zan and his teammates proposed Council-GAN which realized looser constraint than cyclic consistency loss through coping with multiple loss clustering grid images to gain better quality images [13].

PatchGAN replaces GAN's discriminator with a full convolutional network [12]. It outputs the matrix, so it can take into account the effects of different parts of the image. The feature map it outputs can be traced back to a specific position in the source image, so the effect of this position on the output is obviously shown. Several researches have shown that PatchGAN acts better than ordinary GAN discriminator for tasks which requiring HD details and high resolution. Introducing PatchGAN makes the training model pay more attention to details.

In conclusion, the characteristics of CycleGAN, UNIT, U-GAT-IT and Council-GAN are summarized and shown as Table 3.

Table 3. Characteristics of representative methods in domain - based image stylization method

Method	Characteristic
CycleGAN	The performance is poor, the model is unstable.
UNIT	The performance is unstable, different stylization task effects have difference.
U-GAT-IT	The performance is good, model is stable.
Council-GAN	The performance is good, model training is prone to collapse.

4. Comparison of Image Stylization Method based on Domain

Domain-based image stylization methods can generate a stylized image on the same order of magnitude of time, so the comparison of quantifies of CycleGAN, UNIT, U-GAT-IT and Council-GAN with selfie2anime and CelebA2anime data sets is used.

The adopted quantitative indicators are FID (Frechet Inception Distance) and SSIM (Structural Similarity). FID is the maximum mean square error between characteristics of target domain image and stylized image. A lower FID value indicates a higher stylistic agreement between the target image and the stylized image. SSIM is the similarity of structural between source domain image and the stylized image. The lower SSIM value, the lower the structural consistency between the source domain image and the stylized image.

Selfie2anime data set and CelebA2anime data set both include anime face style and human face style. According to the data from Tu and his teammates [3], 3400 human face images, 3400 anime face images are contained in the training set, 100 anime and human face images are included in the test set, and the result FID values and SSIM values are showed as Table 4.

Table 4. Comparison of FID value and SSIM value of different methods

Method	Selfie2anime		CelebA2anime	
	FID	SSIM	FID	SSIM
CycleGAN	111.1	0.4604	101.1	0.3298
UNIT	120.2	0.4149	103.2	0.2969
U-GAT-IT	105.2	0.5535	95.4	0.4500
Council-GAN	105.7	0.4345	100.6	0.3289

From Table 4, it is showed that the FID values of U-GAT-IT, Council-GAN, CycleGAN and UNIT increase in turn, these increased values indicate U-GAT-IT generated anime face image has the highest style consistency with the target animation face domain. The SSIM values of U-GAT-IT, CycleGAN, Council-GAN and UNIT decrease in turn, which means that the anime face image from using U-GAT-IT has the highest structural consistency with the original face image.

In conclusion, each method has different performance on different type of image stylization tasks, and among these four methods, U-GAT-IT and Council-GAN have better overall performance.

5. Development and Expansion

With the maturity of image stylization technology, some researchers find out that image stylization algorithm has broader research value. Here introduce three expanding directions of image stylization.

5.1. Video Stylization

Video can be seemed as a collection contained continuous processing image after image, so image stylization tasks can also be used in video stylization field. Nie and his teammates [14] mention that the creation of an anime movie based on a comic book not only needs people with drawing skills, but also plenty of time, so an automatic tool that turns videos into comics can help a lot. Film and television special effects technology creation is also facing the same problem as anime movie. The more use of artificial intelligence technology will drastically reduce production costs, so Anderson and his teammates uses luminous flux and deep neural network for film stylization [15]. Ruder and his teammates introduce a time consistency loss function to optimize the coherence between frames after the video is stylized [16].

5.2. Human Face Stylization

Previous algorithms did not consider head features individually, this cause stylized effects on the head are difficult to achieve. Selim and his teammates add gain diagrams to the stylization process to constrain the spatial configuration and preserve the structural features of the face [17]. Zhao and his teammates transfer the brushstroke information extracted from the portraitures of the masters to the content image to achieve the face stylization [18]. Wang and his teammates] use MRF to match the target image to the most appropriate feature information from the training data set [19].

5.3. Brushwork Stylization

Brushwork is an authoring tool of artists and designers; brushwork stylization makes creation of works with particular style easier for artists and designers. Brush-based stylization was first proposed in 1990 by Haeberli [20]. Through continuous development and improvement, brush stylization techniques can produce paintings with different stylistic effects such as oil painting, crayon painting, charcoal painting, chalk painting and so on [21].

6. Conclusion

In this paper, the image stylization method is introduced in detail, and the advantages and disadvantages of the corresponding algorithm and the future research direction are briefly analyzed.

Through the research, it is found that although the traditional method can complete the task of stylization, it is not very ideal either in the aspect of synthesis speed or image effect due to its limitations. With the intervention of deep learning, the traditional image processing methods have been better played, and the feature information of the image has been fully utilized. It can be said that the neural network has made a big step forward in the research of image stylization. The introduction of artificial neural network gives people a new direction in the study of image stylization: convolutional neural network can extract semantic message of the image and better express content and style characteristics. Therefore, in the future research on stylization, artificial neural network can be used for stylization processing, relevant stylization parameters can be extracted, and parameters can be introduced into the brushstroke stylization algorithm, to achieve more efficient and automatic stylization expression. Secondly, existing stylization techniques make more use of the underlying features of images and less consider the semantic information and content information of images. In the future, it is necessary to study the stylization methods that can combine the semantic information and content expression of images.

References

[1] Gatys L A,Eckras, Betchem. Image Style Transfer Using Convolutional Neural Networks, IEEE Conference on Computer Vision and Pattern Recognition, 2016:2414-2423

[2] Johnson J, Alah I am, Li F F. Perceptual Losses for Real. Time Style Transfer and Super.Resolution. European Con.ference on Computer Vision, 2016:694-711.

[3] Tu Pengqi, Gao Changxin, Sang Nong: A Review of Image Stylization Methods Based on Deep Neural Networks, 2022, 35(4):333-347.

[4] Li C, Wang M. Combining Markov Random Fields and Convolutional Neural Networks for Image Synthesis, IEEE Conference on Computer Vision, and Pattern Recognition, 2016: 2479-2486.

[5] Chen T Q, Schmidt M. Fast Patch-Based Style Transfer of Arbitrary Style. https: // arxiv.org / pdf / 1612. 04337.pdf.

[6] Huang X, Belongie S. Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization, IEEE International Conference on Computer Vision, 2017: 1510-1519

[7] Li Y J, Fang C, Yang J M, et al. Universal Style Transfer via Feature Transforms, 31st International Conference on Neural Information Processing Systems, 2017: 385-395.

[8] Park D Y, Lee K H. Arbitrary Style Transfer with Style-Atten- tional Networks, IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 5873-5881.

[9] AN J, Huang S Y, Song Y B, et al. ArtFlow: Unbiased Image Style Transfer via Reversible Neural Flows, IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 862-871.

[10] Zhu J Y, Park T, et al. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks, IEEE International Conference on Computer Vision, 2017: 2242-2251.

[11] Liu M Y, Breule T, Kautz J. Unsupervised Image-to-Image Translation Networks, 31st International Conference on Neural Information Processing Systems, 2017: 700–708.

[12] Kim J, Kim M, Kang H, et al. U-GAT-IT: Unsupervised Generative Attentional Networks with Adaptive Layer-Instance Normalization for Image-to-Image Translation 2022: 1-12. https: arxiv.org/pdf/1907.10830.pdf.

[13] Nie Zan O, Tal A. Breaking the Cycle-Colleagues Are All You Need, IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 7857–7866.

[14] Nie Xiongfeng, Wang Junying, Jiang Shu, Chen Hanhan. Deep Learning Based animation style transfer: A review, Signal Processing: Image Communication, 2019, 78:465-470.

[15] Anderson A G, Berg C Poising D P, et al. DeepMovie: using optical flow and deep neural networks to stylize movies. https: arxiv.org/abs/1605.08153.

[16] Ruder M, Dosovitskiy A, Bronx T. Artistic style transfer for videos, Proc of German Conference on Pattern Recognition, 2016: 26-36.

[17] Selim A, Elgharib M, Doyle L. Painting style transfer for head portraits using convolutional neural networks. ACM Transactions on Graphics, 2016, 35(4): 129-136.

[18] Zhuo M T, Zhu S C. Portrait painting using active templates. ACM SIGG RAPH/Eurographics Symposium on Non-Photorealistic Animation and Rendering, 2011: 117–124.

[19] Wang N N, Tao, et al. Transudative face sketch-photo synthesis. IEEE Transactions on Neural Networks and Learning System, 2013, 24(9): 1364–1376.

[20] Haeberli P. Paint by numbers: abstract image representations Proceedings of the 17th Annual Conference on Computer Graphics and Interactive Techniques, 1990: 207–214.

[21] Hertzmann A. A survey of stroke-based rendering. IEEE Computer Graphics and Applications, 2003, 23(4): 70–81.