Proceedings of the 38th Chinese Control Conference
July 27-30, 2019, Guangzhou, China

Restore the Incomplete Calligraphy Based on Style Transfer

Mengxi Qin[1,2], Xin Chen[1,2]*
1. School of Automation, China University of Geosciences, Wuhan 430074, P. R. China
2. Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems, Wuhan 430074, P. R. China
E-mail: chenxin@cug.edu.cn

Abstract: Calligraphy is an important humanistic symbol of Chinese civilization. However, most of the calligraphy is incomplete, which has only a small number of Chinese characters circulating in the world. How to use these samples to efficiently restore the remaining calligraphy characters has always been considered as a difficult task. In this work, we propose Densenet-pix2pix model based style transfer method to solve this problem. By training some samples to learn the rules of transferring the printed font images to the calligraphy characters images, Densenet-pix2pix can predict the remaining calligraphy characters. Our method modify the generation network and optimization strategies of style transfer, which improves the generation quality of the calligraphy characters and the stability of the model. In addition, we use pre-trained feature extraction models to extract content information and style information, and scientifically evaluate the quality of our generated calligraphy characters from these two aspects. We compared our method with several other baseline methods. The experimental results show that our method can effectively restore the remaining calligraphy characters, and the generated Chinese characters are more delicate.

Key Words: Calligraphy characters, Style transfer, Restore, Generation

1 Introduction

Unlike other Western languages, Chinese consists of a large number of characters and complex structures. The official GB18030 character set contains 27,533 characters, involving 254 radicals (such as "point", "horizontal", "vertical", "Hooks", "bottom left corner" and "bottom right corner", etc.), which will take a lot of time and labor to restore the remaining calligraphy by some incomplete calligraphy samples. So it is difficult to rely on manual implementation. In addition, calligraphy contains the author's emotions. Only combining the author's original creative environment can we perfectly restore the charm of the calligraphy, which makes the process of restoration more complicated. Therefore, how to use these known calligraphy samples to efficiently generate the remaining calligraphy characters is a long-standing problem in Chinese culture.

As a new research field of deep learning, style transfer [1] can apply the style of one image to another image without needing to establish mathematical models and doing complex image processing. Its simplicity and efficiency have attracted wide attention and interest from researchers. The most common transfers currently include artistic style transfer [2], English font transfer [3], and facial style transfer [4].

In this paper, we use the style transfer technique to restore the remaining calligraphy characters from incomplete calligraphy. As shown in Fig. 1, we turn the calligraphy characters generation into a style transfer problem from learning a mapping from printed font to calligraphy characters. Conditional Generative Adversarial Networks (cGAN) [5] is a commonly used style transfer network. We improve the optimization strategy and network architecture of cGAN to make it more suitable for the transfer of calligraphy style, called Densenet-pix2pix model. Compared with manual restoration which is difficult and time consuming, there is no need to understand the professional background and the calligraphy's creative conditions. Our method can automatically generate calligraphy characters by learning some calligraphy samples. It greatly reduces the design time and cost.

Fig. 1: Learn a mapping from printed font to calligraphy characters.

(a) Some calligraphy characters of Yu Wei

(b) Generated calligraphy characters

Fig. 2: Generate some calligraphy characters by using Densenet-pix2pix mode.

Fig. 2 shows some samples generated by Densenet-pix2pix model, It can be seen that our results are very close to the style of target calligraphy. It proves that our method can effectively restore calligraphy. In summary, the main contributions of this paper are as follows:

- We propose Densenet-pix2pix model based on style transfer to generate calligraphy characters, which can be used to restore the incomplete calligraphy by learning some samples.
- We introduce Wasserstein distance [6] into optimization strategy, which makes the training of style transfer network more stable.
- We use pre-trained VGGNet16 [7] model to extract content and style. The comprehensive evaluation criteria combined with them can objectively evaluate the quality of generated calligraphy characters.

The remaining sections of this paper are arranged as fol-

This work is supported by the National Natural Science Foundation of China under Grants 61873248, the Hubei Provincial Natural Science Foundation of China under Grant 2017CFA030 and Grant 2015CFA010, and the 111 project under Grant B17040.

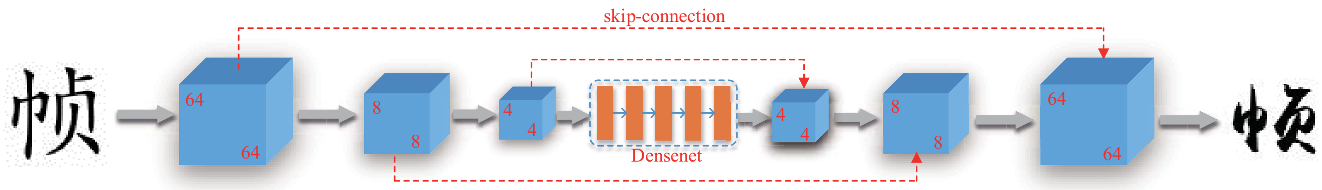


Fig. 3: The architecture of our proposed Densenet-pix2pix model.

lows. Section 2 discusses some related work on the generation of Chinese calligraphy characters. Section 3 describes in detail Densenet-pix2pix model's architecture proposed in this work. Section 4 validates the our method and evaluates the restored calligraphy characters. Section 5 summarizes the contributions of this paper.

2 Related work

In this section, we introduce researches related to the generation of calligraphy characters. Then discuss the advantages and disadvantages between the different methods.

In the past few years, some researchers selected traditional image processing to synthesize calligraphy characters. Xu et al. [8] proposed to generate calligraphy by recombing radical and stroke. Firstly, the calligraphy characters were decomposed into recombinable parts such as word roots and strokes, and then use these parts to synthesize the new calligraphy characters. This algorithm is time consuming and requires calligraphy characters to have a clear stroke structure, so it can't handle calligraphy characters very well. Suveeranont et al. [9] proposed a method for extracting topological structures from font outlines and then generating new fonts by feature retention and weighted blending. This method achieves better results on simple English letters, but the results prone to obvious distortion when the structure of calligraphy characters is complex.

As deep learning has made breakthroughs in image processing, another researchers applied deep learning to study calligraphy characters generation. Deep learning can automatically generate calligraphy characters by learning some samples, which is simpler and more effective than traditional image processing. Zhang et al. [10] proposed using recurrent neural networks to realize the generation of online handwritten Chinese characters. This method relies heavily on online stroke order and time information, so it cannot be used for static calligraphy characters. Tian et al. [11] proposed zi2zi that generates calligraphy characters using pix2pix [12] model based on cGAN. However, the generation network of pix2pix consists of an auto-encoder, and it is impossible to generate an image with a large deformation. For example, the synthetic printed font is perfect, but the generated calligraphy characters is not very well. Chang et al. [13] used the CycleGAN model to turn the printed font into Wang Xizhi's calligraphy characters, and added residual block to the generator module to improve the transfer effect. However, the model uses unpaired Chinese characters for training. There is no standard output for comparison, which loses a lot of details. Sun et al. [14] used variational auto-encoder to generate Chinese characters and embedded stroke information in the middle of the network. This method is able to handle more complex calligraphy characters, and the variational auto-encoder is easier to train than cGAN. However, the disadvantage of this method is that the generated image is more blurred than cGAN.

All of the above methods can't generate calligraphy characters with a large deformation very well, so it is not suitable to restore incomplete calligraphy. To overcome these problems, We propose Densenet-pix2pix model to restore realistic calligraphy characters.

3 Our method

This section will discuss in detail the calligraphy characters's style transfer network proposed in this work, called Densenet-pix2pix model. Firstly, we introduce cGAN and pix2pix which style transfer network we refer to. Then we describe the improved measures for the generator and optimization strategy of the model. Finally, our model design an evaluation criteria including style and content.

3.1 Preliminary

cGAN [5], which is derived from the idea of zero-sum game, is a powerful generation model. It consists of generation network, discriminant network and conditions. The generation network synthesizes the data, the discriminant network distinguishes the synthesized data and the real data. Let the two networks learn the distribution of target sample data through confrontation, the realistic samples will be generated. Besides, cGAN adds a condition in the generation network and the discriminant network. Using this condition, the specified sample with additional condition information is generated.

Pix2pix is an image processing model based on cGAN, which uses images as conditions and input. It applies deep convolutional neural network to generate images and extract features. This model views the problem of traditional image generation as translating one styled image into another styled image, which is often used for image coloring, style transfer, contour extraction, and so on.

In this paper, we use the pix2pix model as style transfer network and propose some improved measures to make our model have a better performance on calligraphy.

3.2 Transfer Network

The generation network of pix2pix only consists of encoder and decoder, which cannot solve the problem of transferring image's style with a large deformation. Therefore, we propose Densenet-pix2pix model to solve this problem. As shown in Fig. 3, the generation network of Densenet-pix2pix is mainly divided into three parts: encoder, transfer module and decoder.

We use Densenet [15] as transfer module to improve the performance of convolutional neural network, which can effectively extract and convert features. Although there are some models commonly used as transfer module to improve the performance of convolutional neural network, these models are discussed from depth and width, respectively. It will cause loss of features during transmission. On the contrary, Densenet analyzes the matter from the perspective of features, directly connecting all the layers, as shown in Fig. 4.

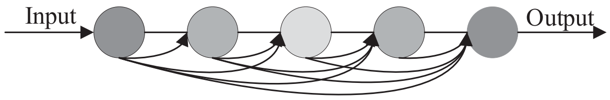


Fig. 4: The architecture of the Densenet-pix2pix model's transfer module.

This dense connection not only makes the network narrower, but also enhances the transfer efficiency of features and uses fewer parameters. It achieves better results on deep convolutional neural networks compared with the above models. Especially, calligraphy characters have a lot of stroke information. Dense connections can effectively transfer feature without causing any information loss. The formula is defined as follows:

$$x_\ell = H_\ell\left([x_0, x_1, ..., x_{\ell-1}]\right), \tag{1}$$

where $[x_0, x_1, ..., x_{\ell-1}]$ refers to the concatenation of the feature maps produced in blocks $0, ..., \ell - 1$, and H_ℓ represents a composite function of operations such as batch normalization [16], leaky rectified linear units (LeakyReLU) and convolution.

The encoder consists of a series of Conv-Norm-LeakyReLU blocks which takes a printed font image as input and produces a $C \times 1 \times 1$ latent feature representation of that image, where C is the dimension of the latent feature. And transfer module converts the latent feature into target style feature by using Densenet, which consists of five dense blocks. Then the decoder is stacked by a series of Deconv-Norm-LeakyReLU blocks, which takes the latent feature representation converted by transfer module as input, and outputs a calligraphy characters image. In addition, we connect layer i and layer $n - i$, where n is the total number of layers of generation network. This structure is able to protect the integrity of information and improve the quality of the generated calligraphy characters. Detailed parameters for generation network are shown in Table 1.

3.3 Optimization Strategy

Pix2pix measures the difference between the two distributions by the Jensen-Shannon divergence. However, Arjovsky et al. [6] shows that the gradient will disappear when there are no overlapping parts or very few overlapping parts. It is difficult to measure whether the discriminant network works via loss function. Consequently, the training of pix2pix is very troublesome, and the parameters of different training sets have to be re-adjusted. Not only the training process of the generation network and the discriminant network needs to be carefully balanced, but also the generated samples lack diversity.

In this paper, we adopt Wasserstein distance [6] to measure the difference between the two distributions, which is defined as follows:

$$W\left(P_r, P_g\right) = \inf_{\gamma \sim \Pi(P_r, P_g)} E_{x,y}\left[\| \; x - y \; \|\right], \tag{2}$$

where P_r and P_g represent two distributions, and $\Pi\left(P_r, P_g\right)$ represents the joint distribution of P_r and P_g. Taking all the x and y from the joint distribution and calculate their means. The minimum mean is the Wasserstein distance. Gulrajani et al. [17] proposed improved Wasserstein distance and built loss function L for its shortcomings, which needs to satisfy the condition of the first-order lipschitz continuity. As shown in Eq. (3) and Eq. (4).

$$\hat{x} = \varepsilon x_r + (1 - \varepsilon)\, x_g, \tag{3}$$

$$\begin{aligned} L = & \underset{x_g \sim P_g}{E}\left[D\left(x_g\right)\right] - \underset{x_r \sim P_r}{E}\left[D\left(x_r\right)\right] \\ & + \lambda \underset{\hat{x} \sim P_r}{E}\left[\left(\| \; \nabla_{\hat{x}} D\left(\hat{x}\right) \|_2 - 1\right)^2\right], \end{aligned} \tag{4}$$

where D represents the discriminant network, x_r subjects to P_r distribution and x_g subjects to P_g distribution.

3.4 Evaluation Method

VGGNet16 [7] increases the network depth by means of repeatedly stacking 3*3 small convolution kernels and 2*2 maximum pooling layer, which effectively improves the ability of extracting feature. According to the research in [18], the deep features contain content information, and the lower layer features contain shape, position, color, texture, etc. In this work, we select the pre-trained VGGNet16 model to obtain the feature matrix on some hidden layers. As shown in Fig. 5, combined with the experimental results of Johnson et al. [19], we take the output characteristics of the $relu3\_3$ layer as content information, called $L_{content}$. The output characteristics of $relu1\_2$, $relu2\_2$, $relu3\_3$, and $relu4\_3$ in the middle layer make up style information, called L_{style}. The final evaluation criteria are combined with the above two kinds of information, which is defined as follows:

$$L_{\text{total}} = \alpha * L_{content} + \beta * L_{style}, \tag{5}$$

Table 1: The architecture and layer specifications of the encoder, transfer, and decoder modules of our proposed generation network. Conv-Norm-LeakyReLU represents a convolution-batch-normalization-LeakyReLU layer, and Deconv-Norm-LeakyReLU represents a deconvolution-batch-normalization-LeakyReLU layer.

Module	Specifications
Encoder	4x4 Conv-Norm-Relu , 64 filter , stride 2
	4x4 Conv-Norm-Relu , 128 filter , stride 2
	4x4 Conv-Norm-Relu , 256 filter , stride 2
	4x4 Conv-Norm-Relu , 512 filter , stride 2
Transfer	dense block
	dense block
	dense block
	dense block
	dense block
Decoder	4x4 Conv-Norm-Relu , 256 filter , stride 0.5
	4x4 Conv-Norm-Relu , 128 filter , stride 0.5
	4x4 Conv-Norm-Relu , 64 filter , stride 0.5
	4x4 Conv-Norm-Relu , 3 filter , stride 1

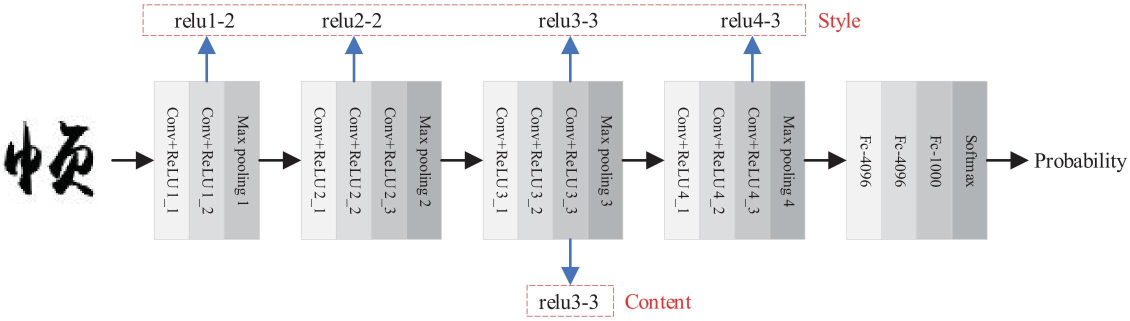


Fig. 5: VGGNet16 extracts content information and style information.

where α and β are two hyperparameters that balance both content information and style information. To better balance the accuracy of content and style, we select α= β=5.

4 Experiment

In this section, we build a dataset using the calligraphy of Yan Zhenqing, Song Huizong, and Yu Wei. Using this dataset, a series of comparative experiments are performed with our method and several related baseline methods. Besides, we apply the evaluation criteria proposed in Eq. (5) to verify the effectiveness of our method.

4.1 Dataset

As we know, there is currently no public dataset for calligraphy's style transfer, so this paper takes the Chinese characters in the Xinhua Dictionary as dataset, and the calligraphy of famous Chinese calligraphers such as Yan Zhenqing, Song Huizong and Yu Wei as the target style. We randomly select 5200 images as training set and the rest images as test set for each calligraphy. We call this dataset CCD-3. Fig. 6 shows some examples in CCD-3.

(a) Yan Zhenqing (b) Song Huizong (c) Yu Wei

Fig. 6: Some examples of CCD-3 dataset.

4.2 Baseline Methods

1) zi2zi: zi2zi is a Chinese font style transfer method based on an image-to-image translation model, which is effective to transfer a printed font to another stylized printed font. It is an end-to-end convolution network with small convolution kernel size and 2x2 stride. Each convolution layer in zi2zi is followed by a Batch Normalization layer and a LeakyReLU layer.

2) CycleGAN: The advantage of this model is that unpaired data can be used to transfer printed words into calligraphy characters, making building dataset more convenient. And the residual block, selected as transfer module, is added to the traditional end-to-end structure to enhance transfer efficiency. In addition, CycleGAN replaces the objective function from the cross entropy loss to the least squares loss in the optimization strategy, which improves the stability of the training process. However, unpaired data can also result in no standard output for comparison, so the details of the generated image are very poor.

4.3 Evaluation and Discussions

In this work, our experiment selects Googles open source software library Tensorflow to construct style transfer network. This deep learning framework can provide a visual interface, which supervises the training process and reduces the debugging cycle. It is very convenient. The computer performed the experiment was equipped with a 3.3GHz Intel Xeon E5-2600 CPU, 32GB RAM and NVIDIA GTX1080Ti GPU.

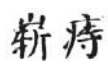
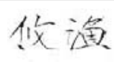
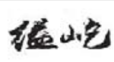
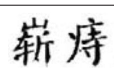
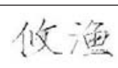
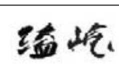
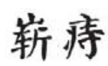
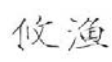
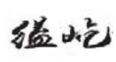
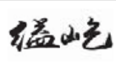
We take the Kaiti as the source font, and use the calligraphy of Yan Zhenqing, Song Huizong and Yu Wei as the target style. Then we evaluate our proposed method and other baseline methods on the CCD-3 dataset. The Kaiti style is simple in structure. Therefore, using Kaiti as the source font can avoid the similarity of input and output, and increase the difficulty of generating calligraphy characters. Besides, it reflects the stability and effectiveness of our network structure from the side.

Next, we analyze our experimental results from quantitative and qualitative, respectively.

1) Qualitative Results

We conduct a series of experiments using our method as well as the above mentioned baseline methods. Table. 2 shows some samples generated by the test set on the three calligraphy subsets individually.

Table 2: The results of the baseline methods as well as our proposed method on the CCD-3 dataset.

Module	Yan Zhenqing	Song Huizong	Yu Wei
zi2zi			
CycleGAN			
Our Method			
Ground Truth			

The results of zi2zi are close to ground truth in the calligraphy of Yan Zhenqing and Song Huizong subsets. But the calligraphy of Yu Wei, which has a large deformation, only correctly restores the Chinese character space structure. The stroke layout and style are almost invisible. This end-to-end structure of the zi2zi cannot solve the problem when large-scale deformation occurred in the style transfer.

Compared with zi2zi, the transfer module has been added to CycleGAN, so the ability of style transfer has been greatly improved. It can be trained to be near optimal with less time. However, the generated stroke details are not satisfactory due to the unpaired data having no standard output for comparison. We can see that the generated calligraphy of Yu Wei is not well and lost a lot of details.

From Table 2, it can be seen that our method obtains the best results. Even evaluated on Yu Wei's calligraphy, our model can still generate images with the similar style of the target image and the local stroke. So it is very suitable to restore incomplete calligraphy.

2) Evaluation Results

Style transfer as a new research direction of deep learning has developed very well, but most researchers evaluated the quality of the images with subjective consciousness. Lacking scientific evaluation is a common problem in this field. We use Eq.(5) to scientifically quantify our results. As shown in Table 3, the result of our method is the closest in content and style to the target Chinese characters.

Table 3: Style and content errors compared to standard calligraphy characters with different methods.

	Yan Zhenqing		Song Huizong		Yu Wei	
	L_{style}	$L_{content}$	L_{style}	$L_{content}$	L_{style}	$L_{content}$
zi2zi	0.193	0.110	0.074	0.026	0.070	0.029
CycleGAN	0.159	0.127	0.306	0.149	0.188	0.107
Our Method	0.146	0.069	0.053	0.026	0.054	0.008

4.4 Stability Analysis

In this section, we compare the inception score (IS) [20] of the generated images with and without using Wasserstein distance, which is a common metric to evaluate generative model. IS measures the quality and the diversity of generated images by calculating the variance of feature outputs from the Inception network [21]. Specifically, a high IS value signifies the fine quality and the variety of a set of images, which means the training of network is more stable.

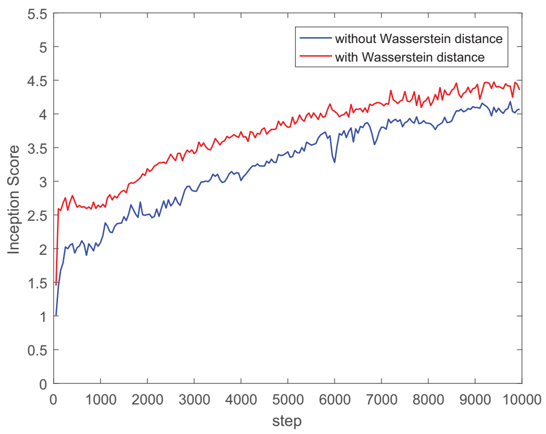


Fig. 7: The inception score of the generated images with and without using Wasserstein distance.

As shown in Fig. 7, since the Wasserstein distance is used in style transfer network, the generated images have a higher inception score and model converge faster. Therefore, even though we don't carefully balance the training process of discriminant network and generation network, high quality and diverse images are generated with the use of Wasserstein distance, and the training process becomes more stable.

5 Conclusion

In this paper, we propose Densenet-pix2pix model based on style transfer to restore some incomplete calligraphy fonts, which can predict the remaining calligraphy characters by learning a mapping from an existing printed font style to calligraphy style. Our method uses Densenet as part of the style transfer network to improve the generation quality. Besides, the training of style transfer network is more stable by using Wasserstein distance. Furthermore, we scientifically evaluate our results from both content and style. The extensive experimental results prove that our method can effectively restore the remaining calligraphy characters from content and style.

References

[1] Jing Y, Yang Y, Feng Z, et al, Neural Style Transfer: A Review, *arXiv preprint arXiv:1705.04058*, 2017.

[2] Gatys L A, Ecker A S, Bethge M, Image Style Transfer Using Convolutional Neural Networks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 2414-2423.

[3] Baluja S, Learning Typographic Style, *arXiv preprint arXiv:1603.04000*, 2016.

[4] Taigman Y, Polyak A, Wolf L, Unsupervised Cross-domain Image Generation, *arXiv preprint arXiv:1611.02200*, 2016.

[5] Mirza M, Osindero S, Conditional Generative Adversarial Nets, *arXiv preprint arXiv:1411.1784*, 2014.

[6] Arjovsky M, Chintala S, Bottou L, Wasserstein GAN, *arXiv preprint arXiv:1701.07875*, 2017.

[7] Simonyan K, Zisserman A, Very Deep Convolutional Networks for Large-scale Image Recognition, *arXiv preprint arXiv:1409.1556*, 2014.

[8] Xu S, Jin T, Jiang H, et al, Automatic Generation of Personal Chinese Handwriting by Capturing the Characteristics of Personal Handwriting, *Twenty-first Innovative Applications of Artificial Intelligence Conference*, 2009.

[9] Suveeranont R, Igarashi T, Example-based Automatic Font Generation, *International Symposium on Smart Graphics. Springer, Berlin, Heidelberg*, 2010: 127-138.

[10] Zhang X Y, Yin F, Zhang Y M, et al, Drawing and Recognizing Chinese Characters with Recurrent Neural Network, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(4): 849-862.

[11] Tian Y, zi2zi: Master Chinese Calligraphy with Conditional Adversarial Network, *https://github.com/kaonashi-tyc/zi2zi*, 2017.

[12] Isola P, Zhu J Y, Zhou T, et al, Image-to-image Translation with Conditional Adversarial Networks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 1125-1134.

[13] Chang B, Zhang Q, Pan S, et al, Generating Handwritten Chinese Characters using CycleGAN, *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2018: 199-207.

[14] Sun D, Ren T, Li C, et al, Learning to Write Stylized Chinese

Characters by Reading a Handful of Examples, *arXiv preprint arXiv:1712.06424*, 2017.

[15] Huang G, Liu Z, Van Der Maaten L, et al, Densely Connected Convolutional Networks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 4700-4708.

[16] Ioffe S, Szegedy C, Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, *arXiv preprint arXiv:1502.03167*, 2015.

[17] Gulrajani I, Ahmed F, Arjovsky M, et al, Improved Training of Wasserstein GANs, *arXiv preprint arXiv:1704.00028*, 2017.

[18] Gatys L A, Ecker A S, Bethge M, A Neural Algorithm of Artistic Style, *arXiv preprint arXiv:1508.06576*, 2015.

[19] Huang X, Belongie S, Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization, *Proceedings of the IEEE International Conference on Computer Vision*, 2017: 1501-1510.

[20] Barratt S, Sharma R, A Note on the Inception Score, *arXiv preprint arXiv:1801.01973*, 2018.

[21] Szegedy C, Liu W, Jia Y, et al, Going Deeper with Convolutions, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 1-9.