Members: Wenyan Guan

School Name: The High School Affiliated to RENMIN University of China

Province: Beijing

Country/Area: China

Mentor(s): Jun Zhu, Hang Su

Title: Chinese Character Style Transfer Using WGAN

The team declares that the paper submitted are conducted the research work conducted and results achieved under the guidance of the mentors. As far as the team is aware, the paper does not contain research that has been published or written by others other than the ones specifically listed and acknowledged in the text. If there is any inaccuracy, I am willing to assume all relevant responsibilities.

Member:关文妍

Mentors:朱军　苏航

2017 Sep. 15

Chinese Characters Style Transfer Using WGAN

Wenyan Guan

The High School Affiliated to RENMIN University of China

Under the direction of

Prof. Jun Zhu

Dr. Hang Su

Tsinghua National Laboratory of Information Science and Technology

Computer Science and Artificial Intelligence Department

Abstract

This work focuses on the problem of high quality style transfer for Chinese characters, converting Chinese printed with standard typograph into those in calligraphic style. Traditional approaches include employing derivative network structures of CNN, RNN, and GAN, yet generation results are not satisfactory enough.

In this paper, we proposed an integrated system of pix2pix structure, auxiliary classifier, and WGAN. Trials were run to evaluate the performance of traditional CNN and GAN. Cross-comparative sets of experiments among GAN, AC-GAN, and our proposed AC-WGAN have been conducted to test the capabilities of our proposed model. Inference and interpolation tests are conducted as well.

Our proposed model outperformed existing style transfer system in generation's delicacy, similarity, and efficiency. Incorporated with WGAN, the model also demonstrated a strong ability in providing a truthful training indicator with its loss. Study also suggested that AC and pix2pix adaption hold huge significance in accelerating the training process.

Research validates the viability of fulfilling Chinese character transformation with style transfer. With handwriting recognition network, fast, high-quality Chinese character style transfer forms the basis for instantaneous conversion between printed and written work. Optimization on algorithm may be another direction for further exploration.

Keywords

Style transfer, Chinese character, generative adversarial network, pix2pix, auxiliary classifier, Wasserstein GAN, end-to-end generation

1 Introduction

Reading and writing are two of the most fundamental linguistic skills of all human beings, with their counterpart in artificial intelligence as text (or character) recognition and generation. There has been a long history in automatic recognition of characters, and significant achievements have been obtained, especially during the past decades. However, the field of text generation (non-semantical) remains relatively under-explored.

This unbalanced situation is disadvantageous for the development of information processing of all languages, especially Chinese, the most widely used language in the world that even has its characters incorporated into many other Asian languages, such as Japanese and Korean, etc. The main difficulties lying in the way include not only the massive efforts required to gather, classify and label the samples, but also the complexity of the physical appearances of Chinese characters.

In order to bridge the gap between the development of these two mutually complementary skillsets, researchers have successfully introduced the concept of style transfer, which is originally proposed regarding images and pictures [1] [2]. The neural networks designed to perform such tasks are primarily consisted of convolutional network (CNN) and its derivatives.

With the ability to extract high-level features from images, CNNs have shown impressive capability handling most problems in the field of computer vision. However, generative adversarial network (GAN) has outperformed CNN in a recent piece of work regarding style transfer [3].

In this paper, optimization on the conditional GAN solution has been proposed based on the Wasserstein GAN (WGAN) [4]. Other major components in the proposed network structure are the pix2pix architecture [5] and category embedding enabled with auxiliary classifier (AC) [6] [7]. The performance of such modified network system in Chinese character style transfer tasks is investigated.

The rest of the paper is organized as follow. Section 2 introduces background information of previous works related to the field of Chinese character style transfer.

Section 3 describes in detail the network architecture and algorithm involved in our proposed model. Experimental settings and results are presented in section 4, and then further discussed in section 5. Final conclusions are drawn in section 6.

2 Research Background

2.1 Chinese character recognition

In order to generate high-quality output, machines are required to be able to receive inputs. Similarly, the foundation of character generation is text recognition. Current works in Chinese recognition are mainly based on two types of databases: offline and online ones. Similarly, two types of networks are used to handle the offline and online data respectively: the recurrent neural networks (RNNs) and CNNs.

The primary difference between these two sources of data is that the online database records the coordinates of the pen-tip during the writing of characters in a time-sequential manner, while the offline database focuses merely on the physical appearances of the characters. In this way, the online database is able to offer more information of how specifically each character is formed.

Based on the distinctive natures of these two databases, different approaches have been taken to tackle the problems. While the offline database stores information with graphic representation, the online database uses numbers. As a result, CNNs are usually employed dealing with offline data, since Chinese characters resemble images whereas CNN demonstrates strong abilities with image and audio recognition [8] [10]. This analogy proves to be a logical one as CNN generates satisfying results for Chinese character recognition [8] [11] under various circumstances. Derivatives of recurrent neural networks (RNNs), on the other hand, have been employed studying the online Chinese character database [12].

2.2 Chinese character generation

RNNs are tested to be able to generate cursive, hand-print text in both English and Chinese [12] [13], as shown in Figure 1. However, results yielded under such framework are not as ideal as the ones produced with CNNs.

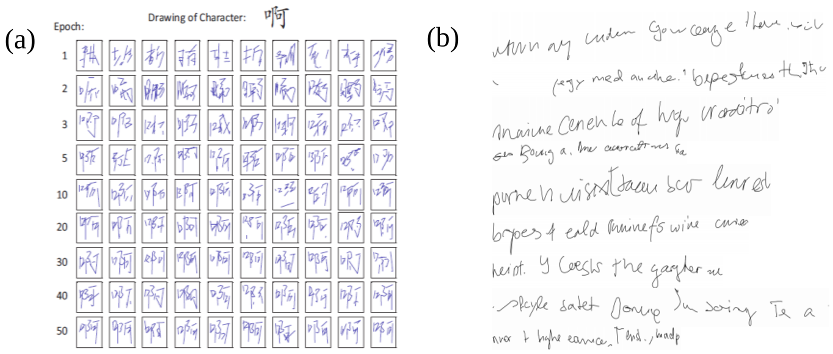


Figure 1 (a): Illustration of one particular character generated by a RNN [12] in different epochs during the training process. (b): Illustration several samples generated by a prediction RNN [13].

The reason behind such an amorphous is because that the sampling of online type of data is conducted in a discrete manner. The physical appearances of RNN generated characters are re-constructed by connecting neighboring sampling points, which cause the characters to loss their thickness and soft contour.

Generation of high quality hand-written style texts is still a missing puzzle. It will be of great significance if realistic characters can be generated automatically as such model will be able to save tremendous amount of human efforts in repetitive tasks, such as font designing and human transcription.

2.3 Previous works with convolutional neural network

CNN has become the choice for complex vision recognition problems for several years. This network architecture was inspired by biological processes [14] in which the connectivity pattern between neurons imitates the organization of animal visual cortex.

Progresses have been made with variants of CNN in the field of style transfer for images [1]. CNN based structure has also been successfully incorporated into Chinese character recognition as a feature extractor [15]. It has also been proved that CNN possesses the ability to learn typographic styles of English characters [16] [17], as shown in Figure 2.

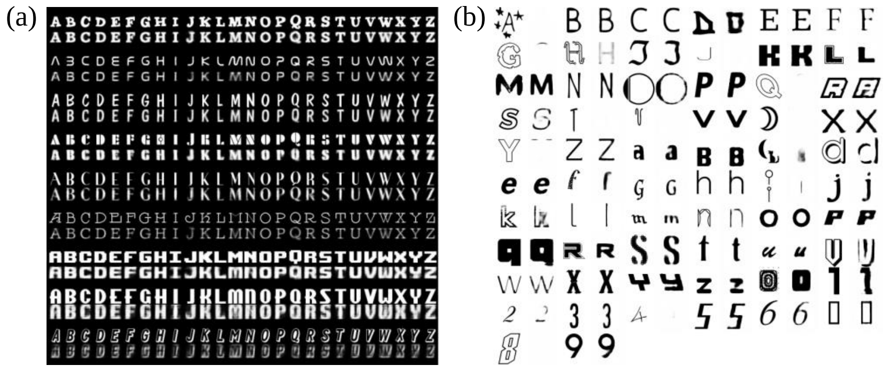


Figure 2 Results generated by CNN in English character style transfer [16] [17] , compared with their ground truths (target fonts).

CNN has also been employed in Chinese character generation [18]. Current structure of CNN in Chinese character style transfer works acceptably for the majority of the fonts. However, all generated characters look blurrier and thicker compared with ground truth, as shown in Figure 3.

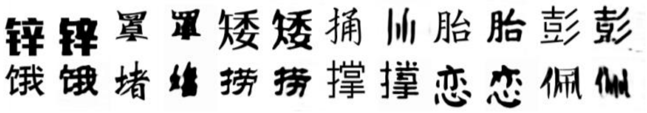


Figure 3 Results generated by CNN in Chinese character style transfer [18], compared with their ground truths

2.4 Generative adversarial network

Meanwhile, improvement in generated results quality has been achieved with GANs [3]. The concept of GAN was first introduced by Goodfellow[19]. A GAN consists of two neural networks is trained in opposition to one another: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G. The training procedure is for D to maximize the probability of assigning the correct label to both training examples and samples from G. Simultaneously, G is trained to minimize term $\log(1 - D(G(z)))$, where z is a prior noise generated by G.

The objective of GAN is shown as follow in Equation (1).

$$\min_{G} \max_{D} V(D, G) = \mathbb{E}_{x\sim p_{data}(x)}[log\, D(x)] + \mathbb{E}_{z\sim p_{data}(z)}[log(1 - D(G(z)))]. \tag{1}$$

The continual competition between D and G urges both networks to evolve, thus forming a more robust system yielding more satisfying results. GANs have been widely employed to deal with image-related problems, including design visualization [20] and text-to-image synthesis [21].

GAN suggest a promising direction to conduct style transfer beyond simple CNN，and this paper aims to provide a few modifications to the basic structure of traditional GAN.

2.5 Wasserstein GAN (WGAN)

WGAN is a freshly proposed piece of work in the area of GAN investigations [4]. Its contributions to the optimization of the current GAN structure are significant.

There are many urging problems that the traditional GAN model is facing[22]: difficulties in training, lack of a truthful indicator of the training progress, and lack of diversity in generated samples, etc. Researchers have been searching for solutions to these deficiencies, yet with no satisfying results. The most well-known improvement of the current GAN structure by far is the deep convolutional GAN (DC-GAN) [7], which relies heavily on laborious enumeration of all possible discriminator and generator set.

Compared with the previous modification on GAN, WGAN manages to:

a) formulate a meaningful loss metric that correlates with the generator’s convergence and sample quality.

b) basically solve the mode collapse of GAN, ensuring the diversity of generated fake samples.

c) improve the stability of the optimization process [4].

It is here provided an approach toward style transfer for Chinese characters via WGAN’s algorithm. With small changes and modifications, it can actually help to speed up the training significantly yielding more exquisite generations.

3 Network Architecture and Algorithm

The overall structure of the proposed model, as shown in Figure 4, can be divided into 3 parts: a base architecture of pix2pix model, an auxiliary classifier, and a WGAN algorithm. Details regarding each one of the three component parts will be discussed in subsections.

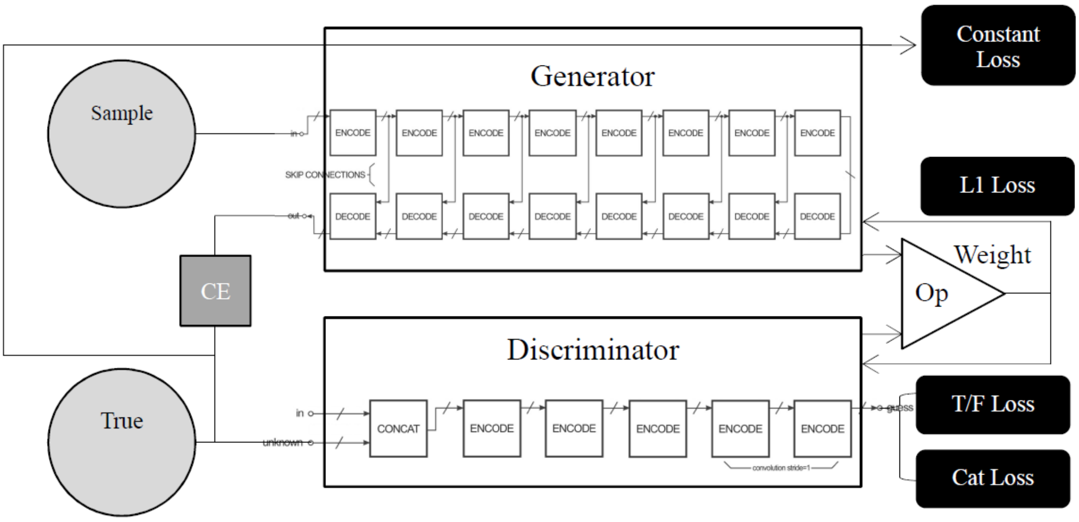


Figure 4 Overall architecture of proposed system (generator and discriminator structure [8]).

3.1 Pix2pix architecture

Proposed in Isola's work [5], the pix2pix uses conditional GAN (cGAN) to learn from a mapping image to an output image. Similarly, this structure, incorporated in proposed system, is able to learn from a mapping Chinese character to an output character. The pix2pix architecture is composed of two main pieces, a *"U-Net"* generator and a *patch GAN* discriminator.

3.1.1 U-Net: generator with skips

In the case of style transfer, though the input and output differ in physical appearances, both are renderings of the same underlying structure. This understanding of the relationship between input and output information is exceptionally critical as it points out a path for algorithm optimization for existing style transfer systems.

Encoder-decoder networks are usually adopted in solutions addressing the problem described above [23] [24] [25]. In such systems, the input passes through a series of layers that progressively down-sample until a bottleneck layers at which point the process is reversed, as show in Figure 5.

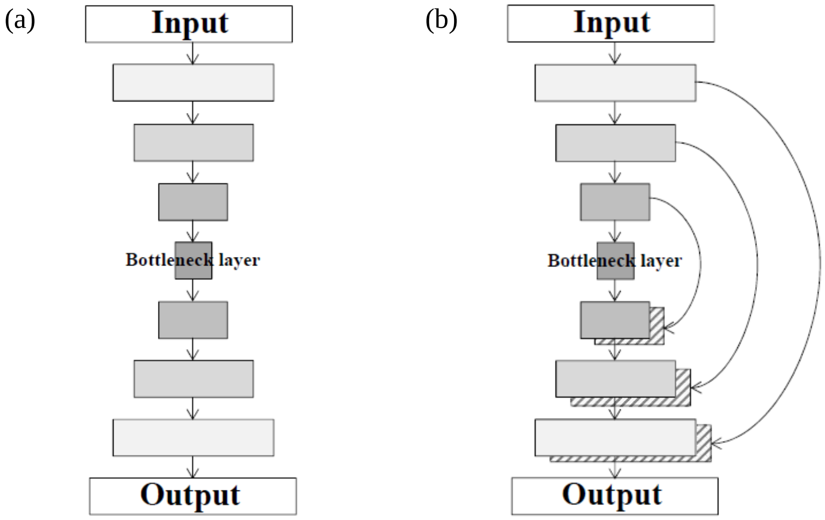


Figure 5 Two types of generator architecture. (a): a typical encode-decoder. (b): the “U-Net” [26], an encoder-decoder with skip connections between mirrored layers in the stack.

Requiring all information flow through every layer compromises the quality of information exchange greatly. The “U-Net” [26] shaped network offers a mean to bypass the information bottleneck with additional connections concatenating all channels in layer No. *i* with those in layer No. *n-i* (*n* stands for the total number of layers in the structure, which is 16 in our proposed structure).

3.1.2 Patch GAN: discriminator

The motivation of adding the “patch” correction mechanism to current discriminator structure is the fact that L1 and L2 losses produce blurry results in image generation problems [27].

The patch GAN only penalizes structure at the scale of patches adopted. Instead of evaluating the whole image, this discriminator tries to classify if each $N \times N$ patch is real or fake. Patch GAN runs convolutionally across the entire image, averaging all responses to provide the ultimate D output.

A 3-layer Patch GAN is involved in our Chinese character style transfer system as a discriminator.

3.2 Category Embedding with AC

Inspired by the previous work [3], an auxiliary classifier (AC) is adapted, which allows the system to realize one-to-multiple training.

The basic GAN framework can be augmented using side information. There are generally two categories of strategies: to supply both the generator and the discriminator with class labels, or to task the discriminator with reconstructing side information [28].

While GANs learn a mapping from random noise vector z to output image y ($G: z \rightarrow y$), cGANs learn a mapping from observed image x and a random noise vector z, to y ($G: \{x, z\} \rightarrow y$). The objective of a cGAN can be expressed as:

$$\mathfrak{L}_{cGAN}(G, D) = \mathbb{E}_{x,y\sim p_{data}(x,y)}[\log D(x, y)] + \mathbb{E}_{x\sim p_{data}(x),z\sim p(z)}[\log(1 - D(x, G(x, z))], \tag{2}$$

where G tries to minimizes this objective against an adversarial D that tries the maximize it.

However, providing Gaussian noise z as an input to the generator is proved to be ineffective as the generator simply learned to ignore the noise [29]. Therefore, a class conditional model is here proposed, with an auxiliary decoder that is tasked with reconstructing class labels [5].

In AC-GAN, every generated sample has a corresponding class label, $c \sim p(c)$ in addition to noise z. The discriminator gives both a probability distribution over sources and class labels. The objective function of can be expressed in two parts: $\mathfrak{L}_S$ and $\mathfrak{L}_C$:

$$\mathfrak{L}_S = \mathbb{E}_{x,y\sim p_{data}(x,y)}[log\, D(x, y)] + \mathbb{E}_{x\sim p_{data}(x),z\sim p(z)}\left[log\left(1 - D\big(x, G(x, z)\big)\right)\right], \tag{3}$$

$$\mathfrak{L}_C = \mathbb{E}_{x,y\sim p_{data}(x,y),c\sim p(c)}[\log P(c\ |D(x, y))] + \mathbb{E}_{x\sim p_{data}(x),z\sim p(z),c\sim p(c)}[\log P(\ c\ |(1 - D(x, G(x, z))))]. \tag{4}$$

D is trained to maximize $\mathfrak{L}_S + \mathfrak{L}_C$, while G is trained to maximize $\mathfrak{L}_S - \mathfrak{L}_C$. AC-GANs learn a representation for z that is independent of class label [22].

The AC mechanism is used to tag and sign different font types in proposed system, which means that aside from conventional true or false loss, the discriminator will also generate a category loss. This loss is expected to urge the discriminator to further involve its discriminative power.

Moreover, the AC mechanism manages to expand the size of the training sets by a great deal without actually introducing more Chinese characters, thus improving the performance of the overall system.

3.3 Wasserstein GAN (WGAN)

As mentioned before in section 2.5, the basic framework of a GAN corresponds to a minimax two-player game with function of $V(G, D)$. Ultimately, the "game" between G and D shall reach an equilibrium, where a unique solution exists, with G recovering the training data distribution and D equals to 0.5 everywhere within the domain.

However, it is observed in practice that as the discriminator improves, the updates to the generator get consistently worse [22]. To solve the problems, the *Earth-Mover* (EM) distance or Wasserstein-1 is included in the proposed architecture to replace the *Jenson-Shannon* (JS) divergence.

The objective of a WGAN can be expressed as:

$$\mathfrak{L}_{V(G,D)} = \max_{w \in W} \mathbb{E}_{x \sim p_{data}(x)} D(X) - \mathbb{E}_{z \sim p(z)} D(G(z)), \tag{5}$$

In general, following modifications are added in our final proposal during WGAN implementation:

a) Cancel the sigmoid applied the original output of D

b) Clipped the weight of parameters in D at the end of every epoch

c) Train D more before training the whole model in a whole package

d) Adapt *root mean square propagation* (RMSProp) instead of momentum based optimizer such as Adam

Last but not the least, the *constant loss* (see in Figure 4) also accelerates the training of propose model [3]. The basic idea of the constant loss is that generator-produced fake samples should not differ dramatically from the input, or the generator is deemed unqualified. This loss forces the generator to continuously improve itself, and thus shortens the training period.

4 Experiments and Results

In this section, experiment procedures and results will be presented in detail. All experiments are conducted on NVIDIA Titan X GPU with 12GB memory.

The model is trained using the RMSProp optimizer in Tensorflow with Python 2.7 interface. The initial *learning rate* (a hyper parameter to RMSProp), set to 0.005 as initial state, decreases by half after a *schedule (a parameter defined in the training process)* number of epochs. *Schedule* is set to 20, with a minimum *learning rate* guaranteed 0.0002. Other parameter settings include *L1_penalty* set to 100, *Lconst_penalty* set to 15, a clipping bottom-line for parameters θ in D, *clip_D*, set to 0.01, and the *source_font* as SIMSUN. Figure 6 listed some typical example of the training set elements.

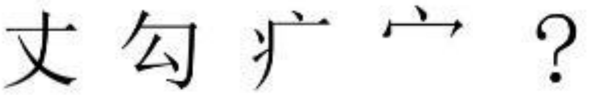


Figure 6 CJK samples included in training set, printed with source font SIMSUN from CJK.

4.1 Dataset and preprocess procedure

This section describes the source of data, and the way in which they are processed before fed into the model.

4.1.1 Characters and fonts

All samples are generated from the CJK (Chinese, Japanese, Korean) family (primarily Chinese characters). The .json file for CJK family is available at source cited in references [3]. Fonts involved in the model are up to users.

Independent single font models and a 32-fonts trained model are constructed in the experiment. The 32 different fonts and the arbitrarily selected 4 independent fonts are printed in Figure 7

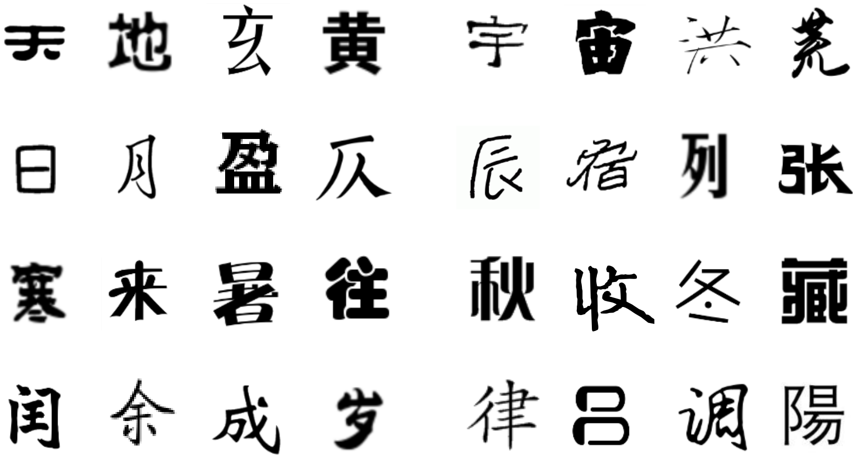


Figure 7 first 8 sentences of *Thousand Character Classic* printed with all 32 target fonts

4.1.2 Preprocess

The preprocess procedure includes two stages. The style information is first converted into imagery files. 1000 characters from the CJK family are randomly generated for each font. The same character is printed twice in both the source font (SIMSUN) and the designated target font on a 512×256 canvas, tagged with a label. Examples of training set elements are shown in Figure 8 below.

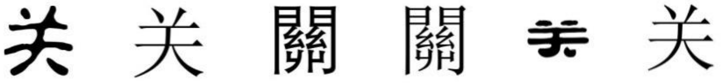


Figure 8 sample No. 730, font label 13, No. 380, font label 1, and No. 579, font 19 from the learning set. In the sample are Chinese character of “guan” printed in in both source font (on the right) and target fonts

Sample imagery files with their corresponding labels are then packaged into binary object files, which are later fed into the system.

4.2 One-to-one style transfer

This section presents the generated results of a CNN [18], a GAN [3], and the WGAN in a one-to-one Chinese character style transfer task, and casts a comparison among them. Shown in Figure 9 are CNN generated results during training. It can be seen that CNN generated results are generally blurry. However, there does exist noticeable resemblance between the target and source fonts.

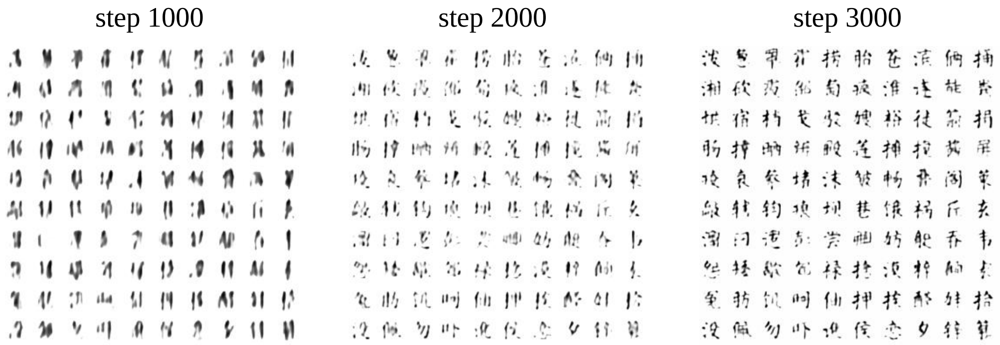
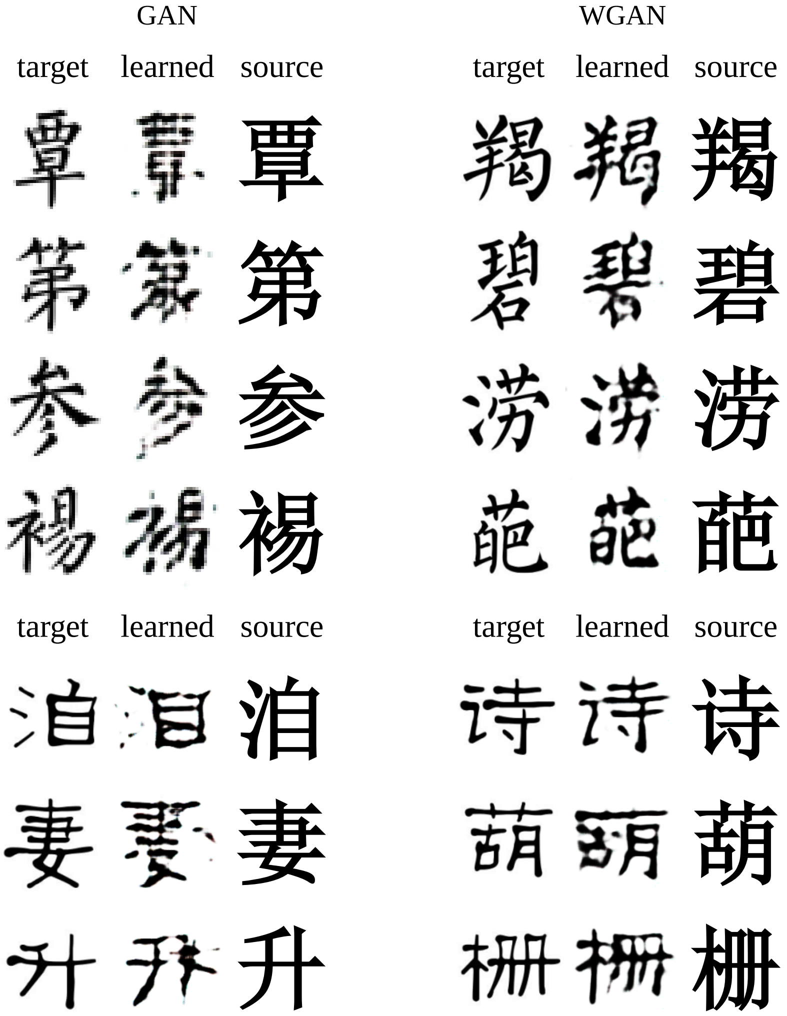


Figure 9 Results generated at step 20, 1000, 2000, and 3000 of a CNN. Target font: 方正苏新诗柳楷简体.

Shown in Figure 10 are results generated by GAN and the proposed model respectively at the end of the training. More results can be found in appendix.



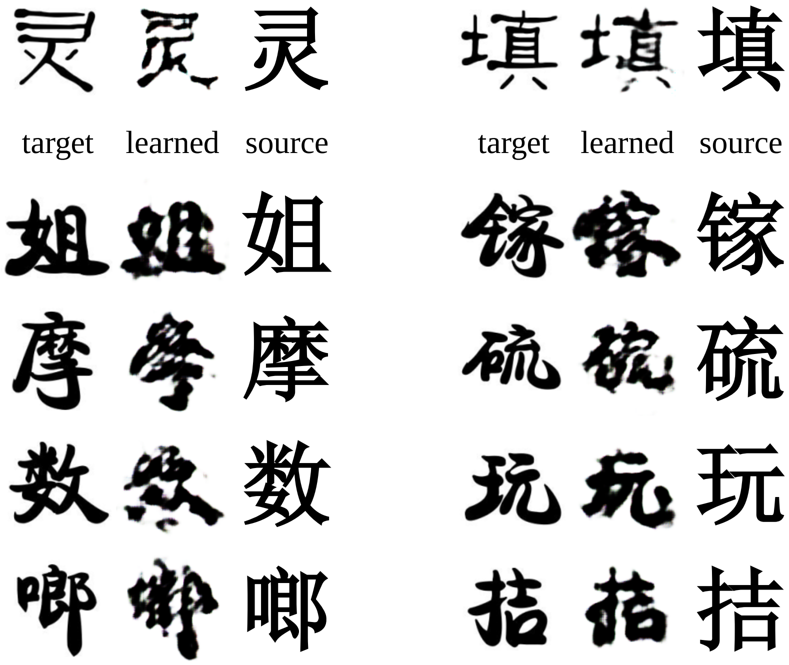


Figure 10 Results generated by a simple GAN and our proposed WGAN system. From left to right printed are ground truth (target fonts), generated results and source fonts. The network has been trained for 60 epochs respectively for every font.

Within the same number of epochs, results generated by the proposed model have clearly contour, more intact structure and higher resemblance to the target font. Although performances of both networks regarding some specific fonts are not so satisfying, results generated with the proposed model are at least readable while its counterpart's generation seems to be rather chaotic.

In order to offer comparison between the presences of AC in the network, an experiment has been conducted using a training set of 32000 characters of the same font. Presented in Figure 11 are the results.

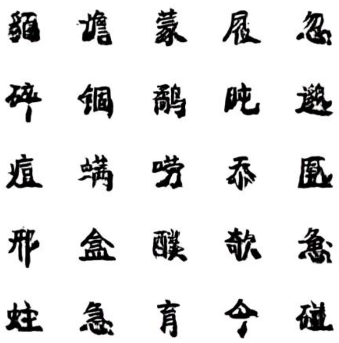


Figure 11 Results generated at the end of the training session by a simple GAN with a training set size of 32000.

4.3 One-to-32 style transfer

This section presents the results generated by a one-to-multiple AC-GAN and our proposed modified structure. As shown in Figure 12

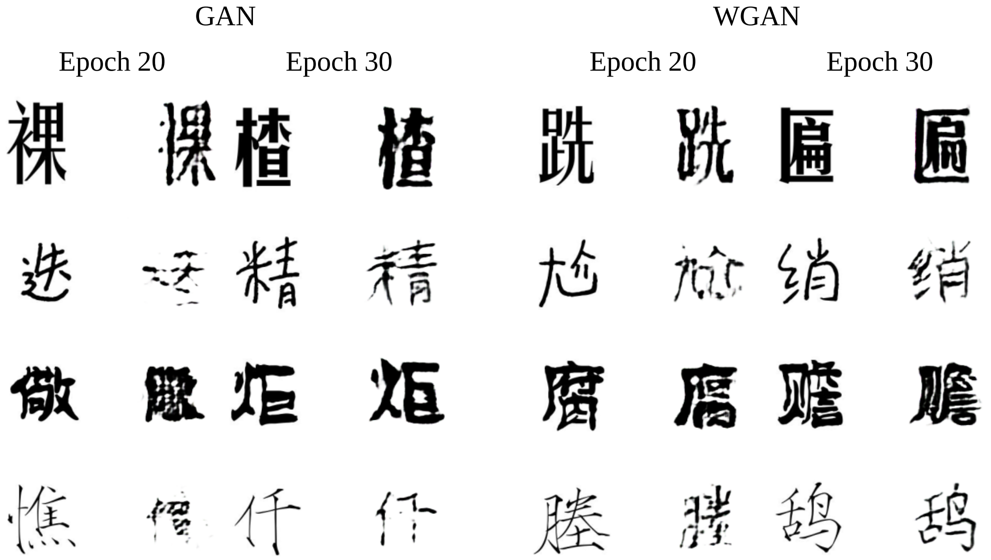


Figure 12 Characters in different fonts generated at epoch 20 and epoch 30 during training by two GANs.

It can be seen that the generation quality of both networks increase along time. However, WGAN's results do have a higher resolution and resemblance to the target font compared with results generated simultaneously by its counterpart. The model is further tested under the inferring mode. Shown in Figure 13 are inference results of characters targeting to learned fonts

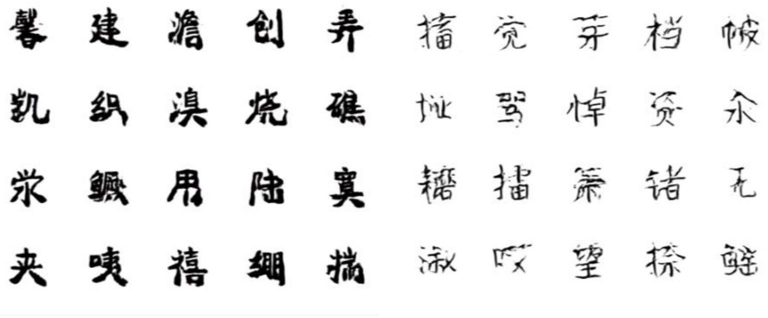


Figure 13 Inference of learned fonts, labeled 7 and 31 respectively.

More inference tests are made on targeting fonts that have not been included in the training set. Part of the results are shown in Figure 14.

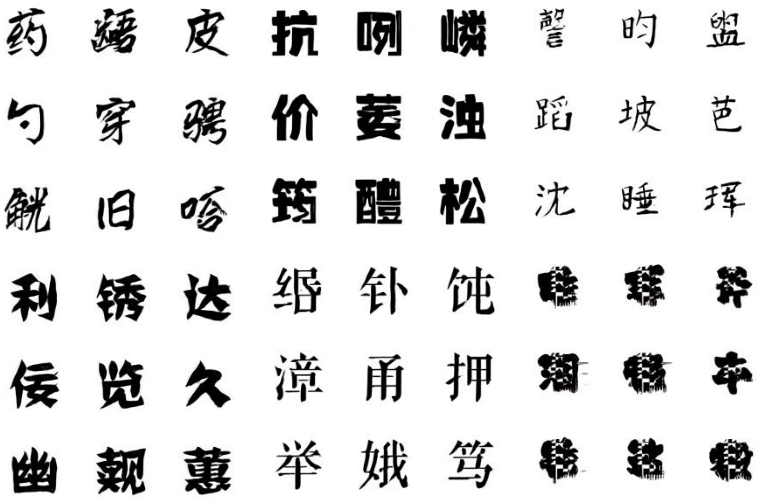


Figure 14 Inference of characters in fonts that have not been included in training sets.

Shown in Figure 15 below is interpolation from one non- acquainted font to another.

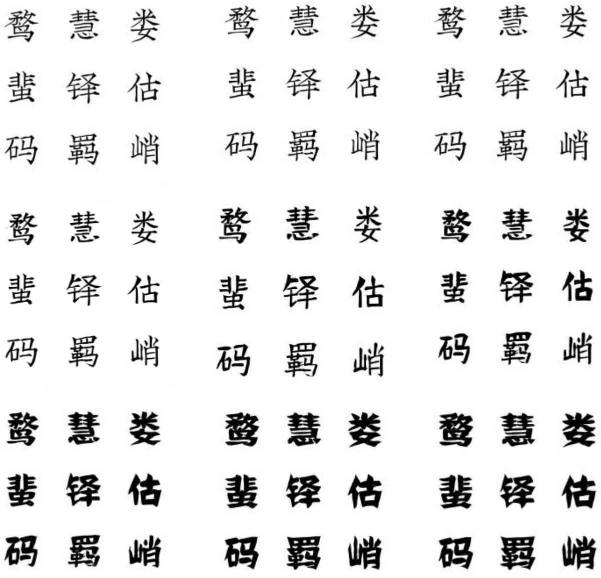


Figure 15 Interpolation from one non- acquainted font to another

Last but not the least, the discriminator and generator losses of the proposed structure during training are investigated here and further compared with those of a non-modified network. The loss vs. time (in sec) figures are presented below in Figure. 16.

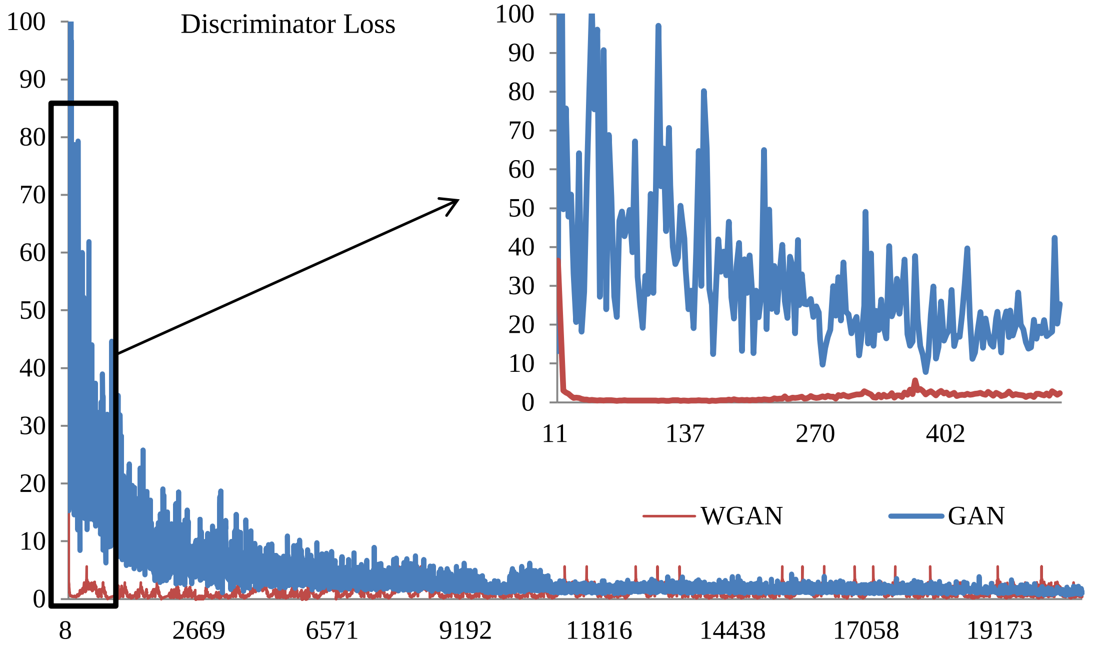


Figure 16 Discriminator loss during the training process base on the 32 fonts sample sets.

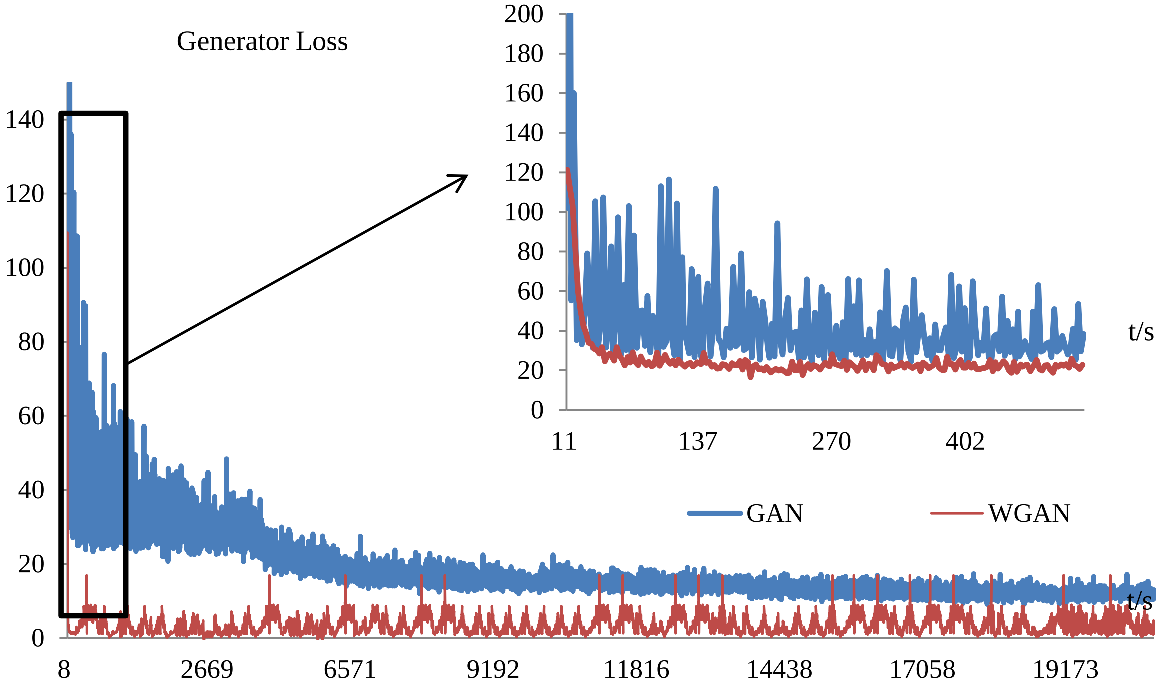


Figure 17 Generator loss during the training process base on the 32 fonts sample sets

While the losses of GAN take about more than 2.5 hours to grow steady, those of WGAN take less than 5 minutes to grow into a relative steady curve. The convergence time is tremendously reduced with WGAN implementation.

5 Discussion

5.1 Auxiliary classifier

From Figure 10, Figure 11, and Figure 13, it could be easily seen that the performance of an AC-GAN greatly surpasses that of a simple GAN, even if the training set of both networks are of the same size. This performance difference is especially significant when it comes to inferring and interpolation involving fonts and styles that have never been studied in the training progress. Learning fonts other than the target ones also seems to be able to improve the performance of the network system. This improvement is probably brought about by the introduction of the constant loss, which helps to speed up the generator's training time.

5.2 WGAN

As described by Arjovsky [4], the WGAN algorithm does have the ability to stabilize the training of GANs. It can be observed from loss graphs that the proposed model, equipped with WGAN, converges significantly faster than a GAN before modification. Although the training time that every epoch takes grows longer (less than 10%) after WGAN implementation (mainly because that the originally cross entropy algorithm is embedded in Tensorflow and can be called upon directly), the proposed system converges at least twice as fast as the non-modified network.

From Figure 12, Figure 16, and Figure 17, it is clear that WGAN seems to provide a reliable indicator of the training progress as the generated sample's quality is actually improving as the losses drop.

It is not hard to imagine that in reachable feature, such proposed WGAN system can be integrated with a style-transfer text recognition network, and thus produce a system that can truly function as an in-situ handwritten style transfer program. Another improvement that can be made to this architecture is to design and add some reliable and universal quantitative indicator of the quality of generated results. In fact, the lack of an objective indicator has long been a problem for generative tasks. It is more than welcome if such authoritative quantity can be created some day.

5.3 Choice of method

As a matter of fact, there is more than one way to approach the problem of Chinese character style transfer. For one, vector operation could be considered for potential solutions. However, different types of methods have been identified with the methods proposed, in consideration of artistic elements and truthfulness in character overall outline. Vector operation does not take into account of the frequent transformations of radicals in characters with different structure. For example, 木 (*Mu*) in 森 (*Sen*) and 林 (*Lin*) would appear differently.

After thorough consideration, it has been evaluated that directly handling characters as incoming images would neither compromise its imagery quality nor cause inaccuracy in outline. Therefore, this work adapts direct image style transfer method to accomplish the goal.

6 Conclusion

In this paper, the performance of the proposed WGAN incorporated structure surpasses those of an ordinary CNN and a non-modified network. Results of simulation are presented to illustrate the effectiveness of WGAN. Not only does the quality of the ultimate generated sample improve, but also the efficiency of the whole system.

The overall approach to problem of Chinese character style transfer is a novel. It has been proved that a AC-WGAN system could carry out both demanding high quality and high efficiency at the same time. Moreover, the significance of the AC mechanism is proven by comparing results generated by a 32-font model and a single-font model (training set sizes both 32000).

With rapid development in the area of GAN studies, different algorithms could be adapted to further promote the system’s performance. Admittedly, the current architecture still depends on a moderate combination of various configurations.

Another potential direction of further investigation would be the applied area of such a network. Together with a handwriting recognition framework and a style extractor, Chinese character style transfer neural net would be able to perform in-situ handwriting-printed style transformation. A style transfer network could also be used in various fields such as cultural relic restorations and designs.

7 Acknowledgments

I would like to thank my mentor, Prof. Jun Zhu, from Tsinghua University, for his guidance and help. I would also like to thank Dr. Hang Su, Mr. Danyang Sun, Mr. Yanze Min and everyone in the computer science and artificial intelligence division for kindly accepting me to learn, work and share resources with them. I want to give special thanks to my colleague Shanxiu He, for offering tremendous help during my working in Tsinghua. I also want to thank my tutor Ms. Yaohan He, writing coach Ms. Yuxing Wang, and counselor Mr. Yifei Chen who have encouraged me during the research period and helped refining language and polish my paper. My thank also goes to my parents for their strong support for me to take part in such rigorous and intense independent research opportunity. Last but not the least, I would like to express my gratitude to Tsinghua University and the Center for Excellence in Education for organizing and sponsoring such educative and beneficial program.

The major part of this paper is the independent work of the contestant Wenyan Guan, completed during the Research Science Initiative program co-organized by the Center for Excellence in Education and Tsinghua University during summer 2017.

Experiments and investigations are conducted at Tsinghua National Laboratory of Information Science and Technology, Computer Science and Artificial Intelligence Department.

8 References

[1] Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image style transfer using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2414-2423).*

[2] Gatys, L. A., Ecker, A. S., & Bethge, M. (2015). A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576.*

[3] kaonanshi-tyc. (2017). Zi2zi: learning Chinese character style with conditional GAN. *Online: https://github.com/kaonashi-tyc/zi2zi, 08/14/2017.*

[4] Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein gan. *arXiv preprint arXiv:1701.07875*

[5] Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2016). Image-to-image translation with conditional adversarial networks. *arXiv preprint arXiv:1611.07004.*

[6] Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., ... & Hughes, M. (2016). Google's multilingual neural machine translation system: enabling zero-shot translation. *arXiv preprint arXiv:1611.04558.*

[7] Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434.*

[8] Hesse, C. (2017, January). Image-to-Image Translation in Tensorflow. *Online: https://affinelayer.com/pix2pix/, 08/14/2017*

[9] Cireşan, D., & Meier, U. (2015, July). Multi-column deep neural networks for offline handwritten Chinese character classification. In Neural Networks (IJCNN), 2015 *International Joint Conference on (pp. 1-6). IEEE.*

[10] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

[11] Chen, L., Wang, S., Fan, W., Sun, J., & Naoi, S. (2015, November). Beyond human recognition: A CNN-based framework for handwritten character recognition. In *Pattern Recognition (ACPR), 2015 3rd IAPR Asian Conference on (pp. 695-699). IEEE.*

[12] Zhang, X. Y., Yin, F., Zhang, Y. M., Liu, C. L., & Bengio, Y. (2017). Drawing and recognizing chinese characters with recurrent neural network. *IEEE Transactions on Pattern Analysis and Machine Intelligence.*

[13] Graves, A. (2013). Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850.*

[14] Matsugu, M., Mori, K., Mitari, Y., & Kaneda, Y. (2003). Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Networks, 16(5), 555-559.*

[15] Du, J., Zhai, J. F., Hu, J. S., Zhu, B., Wei, S., & Dai, L. R. (2015, August). Writer adaptive feature extraction based on convolutional neural networks for online handwritten Chinese character recognition. In *Document Analysis and Recognition (ICDAR), 2015 13th International Conference on (pp. 841-845). IEEE.*

[16] Baluja, S. (2016). Learning Typographic Style. *arXiv preprint arXiv:1603.04000.*

[17] Bernhardsson, E. (2016). Analyzing 50k fonts using deep neural networks. *Online: https://erikbern.com/2016/01/21/analyzing-50k-fonts-using-deep-neural-networks.html, 08/14/2017.*

[18] kaonashi-tyc. (2016). Rewrite: neural style transfer for Chinese characters. *Online:*

https://github.com/kaonashi-tyc/Rewrite, 08/14/2017.

[19] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems (pp. 2672-2680).*

[20] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. (2016). Photo-realistic single image super-resolution using a generative adversarial network. *arXiv preprint arXiv:1609.04802.*

[21] Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., & Lee, H. (2016). Generative adversarial text to image synthesis. *arXiv preprint arXiv:1605.05396.*

[22] Arjovsky, M., & Bottou, L. (2017). Towards principled methods for training generative adversarial networks. *arXiv preprint arXiv:1701.04862.*

[23] Johnson, J., Alahi, A., & Fei-Fei, L. (2016, October). Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision (pp. 694-711). Springer International Publishing.*

[24] Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., & Efros, A. A. (2016). Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2536-2544).*

[25] Wang, X., & Gupta, A. (2016, October). Generative image modeling using style and structure adversarial networks. In *European Conference on Computer Vision (pp. 318-335). Springer International Publishing.*

[26] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 234-241). Springer, Cham.*

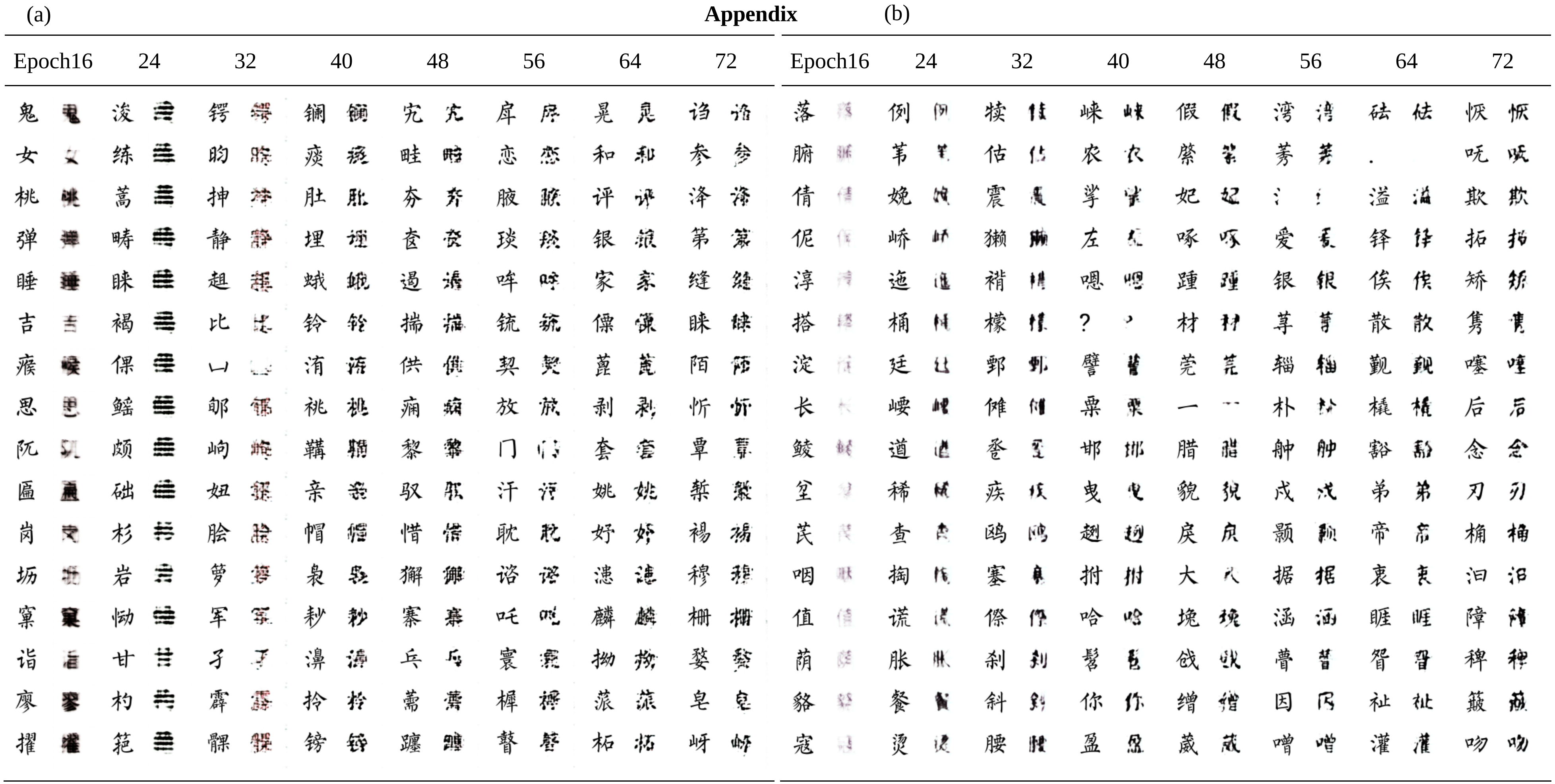
[27] Larsen, A. B. L., Sønderby, S. K., Larochelle, H., & Winther, O. (2015). Autoencoding beyond pixels using a learned similarity metric. *arXiv preprint arXiv:1512.09300.*

[28] Odena, A., Olah, C., & Shlens, J. (2016). Conditional image synthesis with auxiliary classifier gans. *arXiv preprint arXiv:1610.09585.*

[29] Mathieu, M., Couprie, C., & LeCun, Y. (2015). Deep multi-scale video prediction beyond mean square error. *arXiv preprint arXiv:1511.05440.*

[30] Zhang, X. Y., & Liu, C. L. (2013). Writer adaptation with style transfer mapping. *IEEE transactions on pattern analysis and machine intelligence, 35(7), 1773-1787.*

Appendix



Performances of an AC-GAN, as shown in (a), and by the proposed AC-WGAN, as shown in (b) during an 80-epoch training session in a one-to-ton style transfer task. On the left side of each column is the ground truth, on the right side generated sample. Target font: 方正苏新诗柳楷简体, label 15.

参赛队员：关文妍，女，中国人民大学附属中学2018届高中生。自2011年起就读于人大附中第二届早培班。独立项目在 RSI Tsinghua 暑期科研项目中获Top5优秀论文与Top10优秀讲演。课内学习优秀，多年参与数学竞赛，成绩突出；神经生物学课题在2017年人大附中研究性学习评选中获一等奖并全校展示。积极参与校内外活动，任班级宣传委员，校辩论队核心成员。发起或组织过若干学术、公益社团与活动。校外曾作为辩手、队长，两次代表学校参与美国青年物理学家邀请赛，并获得优异成绩。课余时间，爱好绘画，音乐与阅读。

指导老师：朱军，男，清华大学计算机系长聘副教授、卡内基梅隆大学兼职教授、智能技术与系统国家重点实验室副主任。主要从事机器学习基础理论、算法及相关应用研究，在国际重要期刊与会议JMLR, PAMI, ICML, NIPS等发表学术论文100余篇。担任IEEE TPAMI编委、Artificial Intelligence编委、《自动化学报》编委；担任机器学习国际大会 ICML2014 地区联合主席，担任 ICML（2014-2017）、NIPS（2013、2015）、UAI（2014-2017）、IJCAI（2015、2017）、AAAI（2016-2017）等国际会议的领域主席。获中国计算机学会优秀博士论文奖、中国计算机学会青年科学家奖、国家优秀青年基金、中创软件人才奖、北京市优秀青年人才奖；入选国家“万人计划”青年拔尖人才、IEEE Intelligent Systems国际杂志评选的“AI’s 10 to Watch”以及清华大学221基础研究人才计划。

指导老师：苏航，男，博士，IEEE 会员。清华大学智能技术与系统国家重点实验室助理研究员，中国计算机学会计算机视觉专家委员会委员，中国图像与视频大数据专家委员会委员。先后主持国家自然科学基金面上项目、中国博士后基金项目等国家项目，作为核心骨干参与国家973项目、军委前沿科技创新项目和“广东省-国家自然科学基金”联合重大专项等多个国家级项目。在计算机视觉、模式识别和机器学习等领域的重要国际会议和期刊发表论文近50篇（包括CVPR、ECCV、TMI、IJCAI等顶级会议和期刊）。受邀担任大数据领域的重要期刊TBD的专刊编委、人工智能与模式识别顶级期刊 PAMI、TIP 等多个国际期刊的特约审稿人，担任人工智能领域顶级会议ICML、UAI 、NIPS 、CVPR等多个国际会议的技术委员会委员和审稿人。获得国际医学大数据领域顶级国际会议MICCAI的“青年科学家”奖、ACM (中国) 年“卓越博士”提名、和图像视频领域重要会议AVSS的“最佳论文”奖。