Learning and Generation of Personal Handwriting Style Chinese Font

Conference Paper · March 2019

DOI: 10.1109/ROBIO.2018.8665297

5 authors, including:

Liguang Zhou
The Chinese University of Hong Kong, Shenzhen

Huihuan Qian
The Chinese University of Hong Kong, Shenzhen

Zhenglong Sun
The Chinese University of Hong Kong, Shenzhen

Learning and Generation of Personal Handwriting Style Chinese Font

Yutian Lei, Liguang Zhou, Tianjiao Pan, Huihuan Qian, and Zhenglong Sun*

***Abstract*— Personal handwriting style fonts generation is a diverting but time-consuming task due to the large size of Chinese character set. In addition, unlike standard printed style fonts, hand-writing style fonts are of more complicated stroke and glyph feature. In this paper, an improved network architecture is proposed for learning and generation of personal hand-writing style fonts based on small character set. The network is composed of three sub-networks: 1) a classification network for identifying the general style of the target fonts; 2) a generating network for transferring the identified fonts to the target fonts; 3) a discriminating network for differentiating the generated image from real ones. The experiments revealed the effectiveness of the model for generating personal hand-writing style font with relatively small data size, reduction by a scale of 10 comparing to previous reported works.**

I. INTRODUCTION

The writing of Chinese characters is always intractable for the beginner of no matter brush calligraphy or hard-pen calligraphy due to the delicate structure of the Chinese characters. Considering saving the human costs in calligraphy teaching, a calligraphy robot is investigated by us to mimic the writing habits of humans, or even further provide guidance in improving the user's writing skills. In this paper, the initial step for the robot to learn humans' writing styles and generate personal font is first probed into.

Due to the large size of Chinese character set, designing or generating a new font is never an easy task even for the standard printed fonts, let alone the hand-writing ones. Taking Chinese official character set GB 18030-2000 as an example, it contains 27,533 simplified character. Currently, to design a new font for Chinese character generally requires a large amount of manual work taken by professional designer. It is a time-consuming task to design the font character by character that the general design cycle of a new professional printed font is about 1-2 years. As for the design of a new hand-writing font, the most common approach is just to hand-pick and scan the writings of certain calligrapher to the computer. Obviously, the traditional way for typographic design of Chinese character is far from being effective. And that is why there are some recent works focusing on the style transfer between different Chinese fonts. Some proposed methods tried to first decompose characters into strokes (components), and then assembled the transferred components to obtain a character in target font [1] [2]. But it has two major shortcomings that 1) decomposition does not perform well for the characters with complicated layout, or for the fonts with joined-up writing structure; 2) the transformation of strokes between two different fonts usually varies from character to character and it is difficult to find a general pattern for stroke transferring. Other methods adopted the neural network like Fully Convolutional Network or Generative Adversarial Nets, and demonstrated promising results [3][4][5][6]. However, these network-based methods usually required a large amount of characters as training set (for example, 3,000 characters used in [6] and 6,000 in [4]), which is still a tough work to collect in application. Basically, they have only demonstrated the feasibility of using machine learning for font style learning and transferring, which could not be directly applied to learning and generation of a personal hand-writing style font. In addition, previous works generally only used one specified standard font as training source, as a result sometimes the performance was unsatisfactory, especially when the target fonts were in cursive scripts or semi-cursive script.

To overcome these problems, an improved network structure is proposed, by introducing a pre-trained classification network to first identify the style information of target font before training. It is more intuitive and natural for the network to learn the hand-writing font style from a known font style that is close in appearance, thus less training data is required to achieve desired results. Experiments show that our method can actually help in decreasing the required training data set and improving the learning and generation performance of the handwriting fonts. As for the detailed training process, we adopted an FCN-based generating network which can learn the overall structure of a font and transfer the character as a whole without any stroke extraction or other pre-processing. In addition, a discriminating network is borrowed from GAN architecture to further improve the effectiveness and accuracy of the generating network.

The contributions of the paper are mainly reflected in the following aspects:

1) A pre-trained classification network is first proposed to reduce the required training data set size and meanwhile improve the generation efficiency.
2) Proposed model generated promising result in both standard printed style font and hand-writing fonts. Furthermore, the required data size for training is significantly reduced with the proposed model by a scale of 10 comparing to previous research works.
3) A large font set, which is consist of 43 handwriting style fonts is constructed for further usage.

.All authors are with the Institute of Robotics and Intelligent Manufacturing, the Chinese University of Hong Kong, Shenzhen, China.

*corresponding author e-mail:zhenglong.sun@ieee.org.

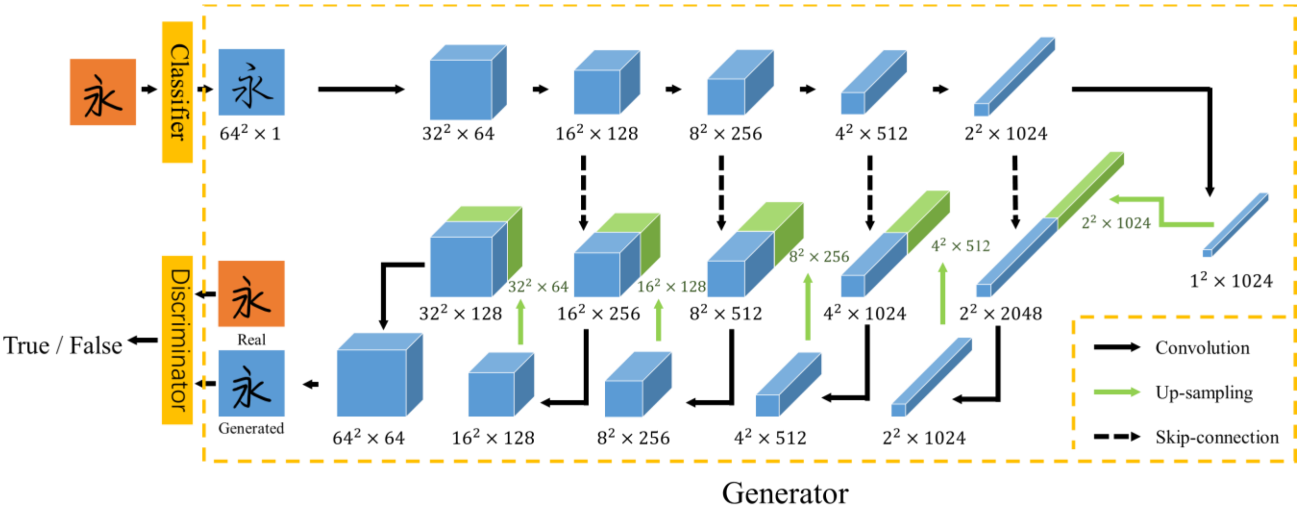

Figure 1. Illustration of the structure and data flow of the proposed model.

II. Method

As shown in Fig 1, the model consists of three sub-networks, *classification network (the classifier)*, *generating network (the generator)*, and *discriminating network (the discriminator)*: the *classifier* is used to first identify the most similar font in our pre-collected font sets, the *generator* implements the learning and transferring from the identified font to our target font using the provided training data, and the *discriminator* tries to differentiate the generated images from the real ones. The latter two networks are optimized jointly. Details of the three sub-nets are discussed below.

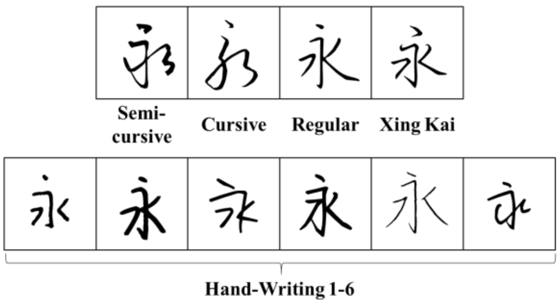

Figure 2. Ten pre-collected fonts used as the source fonts in this study.

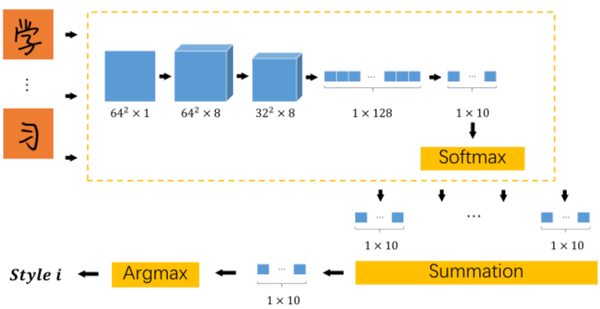

Figure 3. Illustration of the classification network structure.

A. Classification Network

Due to the fact that different Chinese character fonts, especially for hand-writing style, vary a lot in glyph structure, the previous work that using only one specified font as transfer source didn't reaches a good result when the target font and source font are of large difference. Deriving from various calligraphic and historical models, there are numerous styles, or scripts, in which Chinese characters are commonly written, namely, Semi-cursive script (行书), Cursive script (草书), Regular script (楷书), Xing Kai (行楷). For these four scripts, we picked one representative font of each into our source font set. Also, considering the large variation in people's writing style, other six handwriting fonts are collected to capture extra features of people's handwriting. Fig 2 shows how the character "永" can be written in the ten source font styles.

The classification network is pre-trained using the all the characters of the ten source fonts and can finally receive a classification accuracy of over 95%.

Let P_{source_i} be the source font data domain of style i $(i \in \{1,2 \ldots 10\})$, P_{target} be the target font data domain. As shown in Fig 3, the pre-trained classification network $C(y)$ will take the target font data $y \in P_{target}$ as input to output a 1×10 vector that measures the similarities between the target font and each possible source font. We summate the similarities for each character in target font data and find the index with maximum value to represent the style information i we need. The aforementioned progress can be expressed as

$$i = \arg\max\left(\sum\nolimits_{y \in P_{target}} C^*(y)\right) \qquad (1)$$

B. Generating Network

The separation of content and style information is important in most researches about image to image style transfer tasks like in [7][8]. For style transfer between Chinese fonts, however, it is of more significant importance to keep the content information during transferring since any twist or break in the strokes and layout of a character will be apparent and thus fatal. But due to the large size of Chinese character set, it is not an easy work to build a classification network to distinguish every character. Here, an FCN based network is adopted [9], which can keep the character's structure information and meanwhile capture the abstract style information, to take source font data $x_i \in P_{source_i}$ as input and generate a fake image $G(x_i)$ to treat the discriminating network.

The first half part of generating network is composed by a series of convolution blocks which contains a convolution layer followed by Batch-Normalization layer and Leaky Rectified Linear Unit. It works like an encoder, extracting the stroke, structure and style information from a $64 \times 64 \times 1$ character layer by layer until a $1 \times 1 \times 1024$ space is abstracted to present latent feature of the character. Although the 1-D feature actually will destroy the space information compared with a 2-D feature, it is still considered that the deeper encoding layer can extract more latent information, while the skip-connection structure, which will be introduced in the decoder, can help recover the destroyed content information.

Then second half of the network is composed by a series of up-sampling blocks, which contains a deconvolution (transposed convolution) layer, a concatenate layer and a convolution layer. The decoder employs the skip-connection structure [10], which concatenates the layers in decoder and encoder with the same channel before further deconvolution, to recover the glyph and layout information kept in previous layers of encoder. Experiments shows that the skip-connection structure can actually achieve good performance especially for some characters with complicated structure.

C. *Discriminating Network*

Generative Adversarial Network is introduced by [11] to be used in unsupervised machine learning at first, which consists of two neural networks contesting with each other and aims at generating a target image from a random noise. It was future developed to be used in supervised learning tasks that transfer an image from source domain to a target domain. Borrowed from GAN, a discriminating network is introduced in our model that takes fake image $G(x_i)$ as input and output a single scalar $D(G(x_i))$ that represents the probability $G(x_i)$ comes from the P_{target} rather than $P_{generated}$.

D. *Optimization*

The classification network is optimized separately and trained beforehand to offer the style information, while the generating network and discriminating network are optimized jointly after the classification is done.

Like most of multiclass classification problems, cross entropy is typically used in our model to measure the effectiveness of the classification. And the objective of classification network is

$$C^* = \min_C \mathcal{L}_{classification} = \mathbb{E}_{x_i \sim P_{source_i}}[-\log C(x_i)] \quad (2)$$

Borrowed from [11], we define the generating loss as

$$\mathcal{L}_{generator} = -\mathbb{E}_{(x_i,y)\sim\left(P_{source_i},y\right)}[y \cdot \log \sigma(G(x_i)) + (1-t) \cdot \log\left(1 - \sigma\left(G(x_i)\right)\right)] \quad (3)$$

where $i = \arg\max\left(\sum_{y \in P_{target}} C^*(y)\right)$, and σ is sigmoid activation. And similar to GAN, we define the adversarial loss as,

$$\mathcal{L}_{adversarial} = \mathbb{E}_{y\sim P_{target}}[\log D(y)] + \mathbb{E}_{x_i\sim P_{source_i}}\left[\log\left(1 - D\left(G(x_i)\right)\right)\right] \quad (4)$$

Then D and G optimize themselves following a minimax strategy as

$$G^*, D^* = \min_G \max_D \mathcal{L}_{generator} + \mathcal{L}_{adversarial} \quad (5)$$

III. EXPERIMENTS

A. *Data Sets*

As there are no directly available public datasets for different Chinese hand-writing style fonts, we build a data set by downloading forty-two *.ttf* font files from *http://font.chinaz.com/*. Using the methods introduced in 2.1, we choose 10 representative fonts as our source font set, which were processed into the 64×64 numpy array for each character, and 4 of the rest as our testing fonts. According to the *List of Frequently Used Characters in Modern Chinese* [12], 2,500 commonly used and 1,000 sub-commonly used characters are chosen as our character set. It's experimentally demonstrated that the character set can cover over 99% of Chinese corpus. For the classification network, all the 3500 characters are used to train the classification network for improving the prediction accuracy. For the generating and discriminating network, the size of character set used is variant under different configurations.

B. *Performance of Classification Network*

To justify the effectiveness of our classification network, other three handwriting fonts are collected for testing. As shown in Fig 4, our classifier successfully identified the test font style. Although different handwriting fonts will still vary in some details, we believe that the identified font anyhow is more similar to the test font compared with standard printed font like Song or Hei.

C. *Baseline Model*

To further evaluate the transferring part of our proposed model, two baseline models are selected for comparation:

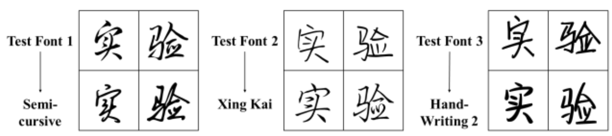

Figure 4. Examples of the outcome of the classification network

Figure 5. Experiments with Different Baseline Models

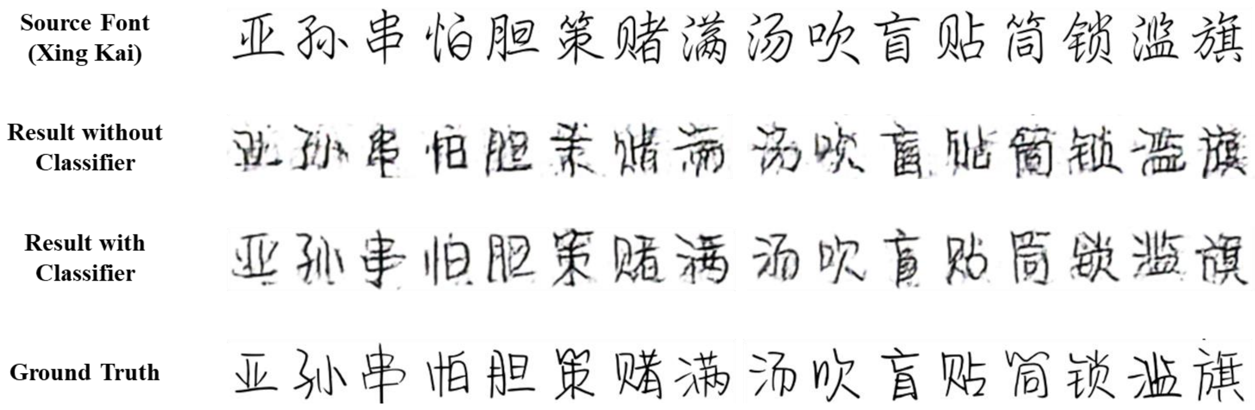

Figure 6. Effect of Classification Network

1) Pix2pix: pix2pix is proposed by [8] using conditional adversarial networks as a general-purpose solution to image-to-image translation problems. These networks not only learn the mapping from input image to output image, but also learn a loss function to train this mapping.

2) Zi2zi: Zi2zi [13] is an application and extension of the pix2pix model, and, as far as we know, is the most effectual recent work for Chinese Character style transfer. The network structure is based off pix2pix with the addition of category embedding and two other losses, category loss and constant loss, from AC-GAN and DTN respectively.

IV. Results and Discussions

1) Experiments with Different Baseline Models

We train the aforementioned baseline models as well as ours using 500, approximate 15% characters from the source font and target font, and the rest of characters is used for testing. The training set and testing set are randomly selected and same for all the models. Some samples of the generated results by our models and baseline models are shown in Fig 5. Specifically, with such a small data size, pix2pix does not achieve a good result: the results generated by pix2pix can only learn the general style of the target font, but for the detail of strokes and layout, most information is lost during transferring. The results generated by zi2zi show a much better result that its generated images are already close to the ground truth. However, noises are still obverse in the strokes of the generated characters, it is believed that the reason is that it cannot capture the complex personal writing style fully with such a small data size. Another evident is that in some complicated characters, the zi2zi model still fails to generate clear and recognizable strokes, for example the characters “矗” and “巍”. In comparison, our model significantly outperforms the two baseline models, especially for those complicated characters and cursive fonts. Some signs of noises are still noticeable but can be easily removed by basic filters. Therefore, the results show that the proposed model can capture and learn the personal hand-writing style through a relatively small size of training data.

2) Effect of Classification Network

The classification network is proposed in our model to solve the dilemma that it is difficult generate a new handwriting font using only one pre-determined standard font like Song. Here specifically, only 300 characters are randomly

Figure 7. Effect of Skip-connection Structure

selected as the training set to test the effectiveness of the classification network in a relatively small data size.

As shown in Fig 6, the experiments are taken and compared between the case without classifier using a randomly selected source font (Song in this case) and the case with classifier using the fonts identified by the classification network as the source font. It is obvious that the classification network improves the performance of learning and generation for the reason that it can provide priori knowledge to subsequent networks, and thus help to save their efforts in learning.

3) Effect of Skip-connection Structure

As mentioned before, one of the feature in the proposed model is to consider the skip-connection structure. It is shown in Fig 7, the result with skip-connection structure outperform significantly than without. It is an evidence that the skip-connection structure is able to recover the details of content and layout information that preserve in the previous layers since the style transfer in Chinese fonts is more sensitive to the stroke or layout distortion compared with other type of image to image translation tasks.

4) Experiments with Different Training Set Sizes

To reduce the workload of generating or designing a new font, it is one of the main objectives for the proposed model to reduce the training data size. In order to investigate how the size of the training data set may affect the style learning and generation performance, three experiments using 300, 500, and 1,000 characters as the training data are taken. The testing character set is randomly selected and kept the same for all the cases. As shown in Fig 8, the experiment with 300 characters learns the global style successfully but fails in some local details. The experiments with 500 and 1,000 characters doesn't vary a lot, expect that the larger size of training set will generate more coherent structure.

Figure 8. Experiments with Different Training Set Sizes

5) Experiments with Different Fonts

To further demonstrate validity of the proposed model, the experiments with 43 other style fonts are also taken. Three of these fonts are shown in Fig 9 as the ground truth 1-3. For each font, 500 characters are randomly selected from the list to be used as the training set and the rest characters are used for testing. According to the training characters, the classification network classified the three test fonts as Semi-cursive, Regular and Hand-Writing Font 2 in our source font set.

The experiment shows satisfactory result for the test font 2, which has highly regular pattern, while the results for the test font 1 and 3 are of a few minor flaws, since for font 1 the joined-up writing style is more difficult to learn using a small data size, and for the font 3, it's of more personal style and thus less regular pattern.

V. Conclusion and Future Works

In this paper, a model consisting of three subnets: classification network, generating network and discriminating network, are proposed for generating personal hand-writing

Figure 9. Experiments with Different Fonts

style fonts. The classification network is pre-trained and can provide the style information identified by it to the generating network and discriminating network. The generating network and discriminating network are trained jointly to transfer the source font to our target one. Experiments results demonstrated that, compared with recent Chinese font generation works, our model can generate satisfactory results with much smaller size of training data size. Using the propose model, it is possible to learn and generate a whole character set of personal handwriting style out of only a small amount of sample characters which are carefully selected.

Our future work will focus on 1) improving the model to further reduce the size of training set; 2) combining with robots to automatically learn and imitate the writing style of human from the information collected from camera or other sensors.

ACKNOWLEDGMENT

This work is supported by the Shenzhen Science and Technology Innovation Commission, fundamental research grant JCYJ20170818104502599.

REFERENCES

[1] S. Xu, F. C. Lau, W. K. Cheung and Y. Pan, "Automatic Generation of Chinese Calligraphic Writings with Style Imitation," *IEEE Intelligent Systems*, vol 24, no. 2, pp. 44 - 53, 2009.

[2] B. Zhou, W. Wang and Z. Chen, " Easy generation of personal Chinese handwritten fonts," *IEEE International Conference on Multimedia and Expo*, pp. 1-6, 2011.

[3] Z. Lian, B. Zhao and J. Xiao, "Automatic Generation of Large-scale Handwriting Fonts via Style Learning," SIGGRAPH ASIA 2016 Technical Briefs, 2016.

[4] P. Lyu, X. Bai, C. Yao, Z. Zhu, T. Huang and W. Liu, "Auto-Encoder Guided GAN for Chinese Calligraphy Synthesis," arXiv:1706.08789, 2017.

[5] Y. Zhang, J. Chang and Y. Gu, "Chinese Typeface Transformation with Hierarchical Adversarial Network," arXiv:1711.06448, 2017.

[6] J. Chang and Y. Gu, "Chinese Typography Transfer," arXiv:1707.04904, 2017.

[7] L. A. Gatys, A. S. Ecker and B. Matthias, "Image Style Transfer Using Convolutional Neural Networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* pp. 2414-2423, 2016.

[8] P. Isola, J. Zhu, T. Zhou and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," arXiv:1611.07004, 2017.

[9] J. Long, E. Shelhamer and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* pp. 3431-3440, 2015.

[10] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in proceeding of International Conference on Medical Image Computing and Computer Assisted Intervention (*MICCAI),* pp. 234-241, 2015.

[11] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, "Generative Adversarial Nets," in proceeding of the Conference on Neural Information Processing Systems (*NIPS),* 2014.

[12] Y. Fu, "Research on Frequently Used Characters in Modern Chinese," *Yu Wen Jian She,* vol. 2, pp. 22-31, 1988.

[13] "zi2zi," [Online]. Available: https://github.com/kaonashi-tyc/zi2zi.