

Font generation based on least squares conditional generative adversarial nets

Xianming Lin^{1,2} · Jie Li^{1,2} · Hualin Zeng^{1,2} ·
Rongrong Ji^{1,2} 

Received: 10 June 2017 / Revised: 25 October 2017 / Accepted: 26 November 2017 /
Published online: 8 February 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract With the rapid growth of multimedia information, the font library has become a part of people's work life. Compared to the Western alphabet language, it is difficult to create new font due to huge quantity and complex shape. At present, most of the researches on automatic generation of fonts use traditional methods requiring a large number of rules and parameters set by experts, which are not widely adopted. This paper divides Chinese characters into strokes and generates new font strokes by fusing the styles of two existing font strokes and assembling them into new fonts. This approach can effectively improve the efficiency of font generation, reduce the costs of designers, and is able to inherit the style of existing fonts. In the process of learning to generate new fonts, the popular of deep learning areas, Generative Adversarial Nets has been used. Compared with the traditional method, it can generate higher quality fonts without well-designed and complex loss function.

Keywords Least squares conditional GAN · Font generation · Stroke extraction

1 Introduction

On the account of that characters are important carriers of civilization inheritance, Chinese characters, a treasure of Chinese civilization, play a vital role in the process of the inheritance and dissemination of Chinese civilization. As the unique appearance of characters in the digital age, fonts bear a significant intermediary role in the storage, transmission, etc. Whether in the Internet or in print, different fonts offer a completely different visual experience, which is conducive to improving the efficiency of

✉ Rongrong Ji
rji@xmu.edu.cn

¹ School of Information Science and Engineering, Xiamen University, Xiamen 361005, China

² Fujian Key Laboratory of Sensing and Computing for Smart City, Xiamen University, Xiamen 361005, China

information dissemination. Characters written by different persons may reveal emotions or attributions of the writer [33, 36, 37]. But for Chinese, large number of characters and complex strokes require at least 5 to 6 persons work a whole year to produce a complete set of fonts containing more than 20,000 Chinese characters included in the Chinese Internal Code Specification. Therefore, designing a new font is a difficult project needed to spend a lot of manpower and material resources, resulting in little quantity which can't meet the needs of display of the text's diversity. So it has great theoretical value and practical significance to study the automatic generation of Chinese font.

As a result, the study of the problem of automatic generation of characters is a subject of much concern, attracting many scholars to focus on it and proposed a number of models [12, 13, 27, 30–32, 38]. Due to the complexity and the large quantity, automatic generation of Chinese font is much difficulty than English font generation, leaving it an open question. We find from comprehensive analysis of various font generation models that these models can be divided into two categories: model based on the handwriting and image.

- 1) **Model based on the handwriting** employs the movement of the pen to represent characters. For instance, Xuyao Zhang et al. [32] put forward a method training the movement with RNN. This method transforms a character to a series of states which contains whether the pen is on the paper or off and the offset relative to the previous point, and trains the states with RNN. It can output readable character without width information. However the description of the contour and font style can't be represented satisfactorily, and the training process consumes much time.
- 2) **Model based on the image** usually splits Chinese characters into strokes or move levels, re-combines strokes to generate new characters. Zhouhui Lian et al. [13] makes regression prediction on the sampling points from the bitmap; Alfred Zong et al. [38] saves strokes as StrokeBank for reassembling; Songhua Xu et al. [31] adjusted font's details with special shape grammar; Xiaohu Ma et al. [30] use Fourier series to fitting strokes' contour. This method works well for the representation and generation of the stroke level, not the overall structure. Although the local generation has good results, the structure of Chinese characters may be not artistic. Recently, some end-to-end deep network architectures were put forward. Chuang et al. [3] proposed a VAE-based network called variational grid setting network which splits character image into grids. It is able to generate high resolution images while requiring source style and target style share similar structure. Lyu et al. [15] used an end-to-end model combined auto-encoder and GANs to transfer the whole character image to a new character image with target font style. It is not based on strokes extracted from character so it can be applied in a cursive style. However, it not addresses the problem of blur and structure preservation.

At the same time, in the last decade, the research on the deep neural network has attracted much attention. Not only great progress in the research of the discriminative deep neural network model has been made, such as Alexnet [11], VGG [23], GoogLeNet [25], ResNet [8] and other classic models been proposed, but generative model also made a major breakthrough. The Generative Adversarial Networks (GANs) was first proposed by Ian J. Goodfellow et al. [6]; Mehdi Mirza et al. [17] introduced a conditional constraint to learn the distribution of the target under that particular condition by adding inputs to the network G and the network D as conditions to improve the quality

of the generated image. Emily Denton et al. [4], inspired by the DRAW model [7], proposed Laplacian Pyramid of Adversarial Networks (LAPGAN) to generate high-resolution images by increasing the resolution gradually. Xi Chen et al. [2] argues that the input random variable of network G includes some of the characteristics of the target distribution of latent codes in addition to noise which can't be compressed. So they put forward InfoGAN which divided the inputs into noise and latent codes and add network Q in network D to restore the latent codes. Augustus Odena et al. [21] proposed Auxiliary Classifier GANs (AC-GAN) on the basis of InfoGAN, replace latent codes with category tags. [2, 21] proved that addition of more structures to the latent codes, coupled with the corresponding loss function, can effectively produce high quality results. It can be seen that through recent years of research, GANs has made rapid development, and its research results have also been widely used in: image generation [4, 17, 21], image classification [20, 24] and machine translation [28] etc.

In this paper, we proposed a new font generation framework by fusing a variety of Chinese character font styles. In this framework, strokes will be extracted from every Chinese character at first. Then the styles of a variety of fonts will be fused to generate a new font by ensuring the correct structure of every Chinese character. It's common to generate new artistic samples by style transfer or style fusion [5, 10, 34]. The main contributions of this paper are: 1) proposed a stroke extraction algorithm for the Chinese characters based on improved coherent point drift algorithm. (2) proposed the automatic font generation model based on the least squares condition GANs for the style fusion of two kinds of fonts.

The rest of our paper is organized as follows: Section 2 covers the related work. Section 3 presents the LSCGAN and the generation of Chinese font. Section 4 shows the experimental results, and Section 5 presents the conclusions.

2 Related work

Chinese character stroke extraction Stroke extraction plays a very important role in font generation. At present, there are three main methods for extracting strokes: the first kind is extracted directly from the original Chinese character bitmap. For example: Ruini Cao et al. [27] proposed a method requiring the calculation of the point-to-boundary orientation distance and the boundary-to-boundary orientation distance (BBOD) of all points. After that, screen out the BBOD's peaks and project them to 3D space. Look for connect components in 3D space and then cast back to 2D space. Each connect component represents a stroke. This method uses the width and direction etc. of the strokes. However, it needs complex operation and takes a long time to run, and the threshold used for screening out peaks has a great impact on results. The second category is to extract strokes from characters' contours. The method proposed by Li Cheng et al. [12] defines five cases of stroke crossings. After finding the special points (pitches) on the contours, reconnect them according to five different situations to achieve the purpose of separating strokes. This kind of method is good for neat fonts, while may fail on fonts with high degree of artistic effect. The third category is to thinning the bitmap and extract strokes from the skeleton. Skeleton retains the basic features like the length and the direction of the stroke, and contains less redundant information. For example: Xiafen Zhang et al. [29] use crawler method. Let crawler detect path on skeleton according to

writing rules, split the path at the intersection according to some pre-defined rules, and finally split the stroke. As the skeletons are often not enough regular, and situations of bifurcations are often more complex, this method does not cover all the circumstances well. Zhouhui lian et al. [13] proposed a method based on coherent point drift (CPD). In the case of data of marked strokes, sample point from skeleton and map point set to get the information of corresponding point on the skeleton. This method does not require much priori knowledge, and can achieve better results. It is the current mainstream stroke extraction method. Beside, some methods like dynamic time warping [18, 35] being able to measure strokes similarity were used.

Generative adversarial networks The second part of related work is the GANs. As a generative deep neural network model, GANs was first proposed by Ian J. Goodfellow et al. [6]; Mehdi Mirza et al. [7] introduced a conditional constraint to learn the distribution of the target under that particular condition by adding inputs to the network G and the network D as conditions to improve the quality of the generated image. Phillip Isola et al. [9] proposed the pix2pix model, using U-Net network structure to generate image from image instead of from random noise, implementing end-to-end training. In order to overcome the shortcomings of fighting the network instability and slow convergence, researchers continue to improve the network structure, adjust the loss function. For example: Xudong Mao et al. [16] proposed the least squares generative adversarial networks using the least squares loss function which converges steadily and can be trained quickly.

In addition, the work of DCGAN (deep convolutional generative adversarial networks) [22] and Wasserstein GAN [1] has contributed to improving the convergence of the network. At the same time, the GANs are also employed for the generation of domain-related samples (e.g. handwritten digital images of different styles). The Coupled Generative Adversarial Networks (CoGANs) proposed by Mingyu Liu et al. [14] are a pair of GANs sharing weights in the beginning few layers and the last few layers. The pair generated images from different domains independently. Yaniv Taigman gone further and proposed the Domain Transfer Network (DTN) [26] to convert the sample from one domain to another. These research works have played a certain role in the design of the font generation model in this article.

Chinese character font generation The font generation method associated with this article is based on the image. This kind of methods often split Chinese characters into strokes or more levels, re-combine strokes to generate new fonts. For example: Zhouhui Lin et al. [13] proposed that by splitting the stroke, map relative position and stroke shape of reference font to target font stroke with two fully connected artificial neural network including only one hidden layer. They can generate a lot of handwritten character with the same style from a small amount of handwritten characters. Of course, the method in [13] applies only to handwritten fonts without width information, and requires many pre-defined parameters. Alfred Zong et al. [38] proposed the StrokeBank model, which decomposes the standard font and the target font into several components, using the directional element feature and the Fourier spectrum as descriptors to generate the StrokeBank. Decompose corresponding character into components, and then find the best mapping set in StrokeBank to get the corresponding components. These components will be reassembled to get the result. This method uses strokes store in StrokeBank

instead of generating new strokes, and thus do not have the ability to learn stroke style. Songhua Xu et al. [31] proposed the use of a shape grammar to measure the likelihood of a character Y from font X . For the new character to be generated, generate it with existing method along with some parameters that control the visual effects and update the parameters using gradient descent to maximize the likelihood calculated by shape grammar. This method requires the expertise of fonts and it's hard to choose parameters related to visual effects. Xiaohu Ma et al. [30] considered the curve contour as the path on which a particle does the cycle of movement in a constant speed. So the contour of extracted stroke can be expressed as periodic function of time. After transforming the periodic function to the form of Fourier series, character with new style can be got by adding random noise to the Fourier series or by interpolation between two Fourier series got from two different fonts. Since it is not feasible to use an infinite number of series in this method, usually only the first 20 to 25 Fourier series will be chosen as the representation of the font, resulting in poor fitting for angular fonts.

3 LSCGAN for Chinese font generation

3.1 The overview of Chinese font generation

The framework of font generation is composed of two parts: the offline training LSCGAN and the online font generation. The training process of LSCGAN includes extracting strokes, generating the condition vector and training LSCGAN. And the online font generation process includes extracting strokes, generating strokes of a new font and generating the new font. The overviewflow chart of Chinese font generation is shown in Fig. 1.

3.2 Stroke extraction

The stroke is an important component of Chinese character. The main difference between two fonts lies in stroke and structure. Hence, the stroke is vital to a font and the stroke extraction is one of the key step of Chinese font generation as well. In this paper, we proposed a stroke extraction method with the coherent point drift algorithm as the model prototype. Meanwhile, a strategy is used to control the width of the strokes. The coherent

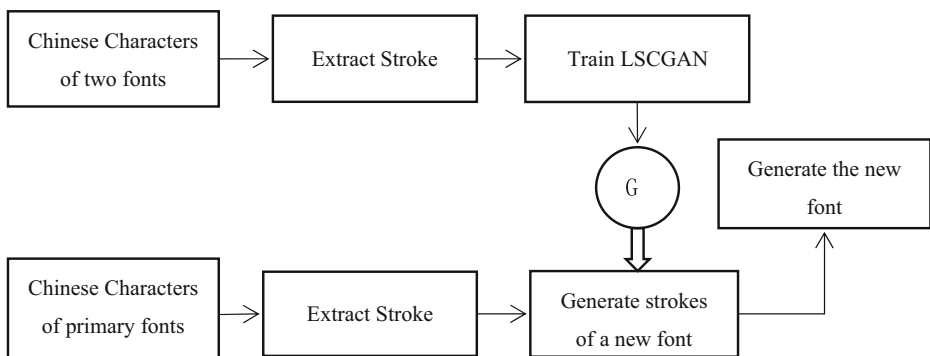


Fig. 1 The framework of new Chinese font generation based on LSCGAN

point drift algorithm [19] was proposed by Andriy Myronenko et al. and used to map a point set to another. The core idea of this algorithm will be briefly introduced below.

Given two point set X and Y, and the point set Y is considered as the center of the Gaussian mixture model. An EM algorithm is derived to compute maximum likelihood estimation for fitting the point set X. The coherent point drift algorithm is widely used in image registration. And the key issue of the stroke extraction in this paper is the matching of two point sets. Hence, the coherent point drift algorithm is introduced in our stroke extraction. Meanwhile, a strategy is adopted for the control of stroke width. For each pixel in the bitmap of a Chinese character, we should calculate the distance between the skeleton and it, and then determine which stroke it belongs. For example, given a pixel p_1 , a stroke set $\{s_1, s_2, \dots, s_m\}$ of a Chinese character skeleton and a threshold value r , if the distance between p_1 and s_i less than r , we can tell that p_1 belongs to s_i with a certain probability. As we can see, a pixel may belong to more than one stroke. The optimal pixel assignment solution will be found by loop iteration.

During the stroke extraction, some points which are sampled randomly from the stroke library compose the point set Y. And the position and stroke type of the every point are stored in Y as well. The skeleton will be extracted from the Chinese character, and a point set X will be generated by sampling from the skeleton randomly. And the coherent point drift algorithm will be used to update point set Y in each iteration. The nearest point will be found from point set Y for every point of point set X. And the stroke category information of the point from set Y will be assigned to the point of set X. The detail of the strokes extraction process can be seen as Fig. 2.

Strokes will be extracted from the common Chinese character by above method. To clear up the position offset, the center of the stroke will be moved to the center of the bitmap. Figure 3 shows the iterative matching process of two point sets.

The basic strokes which make up the Chinese character includes dot, cross, vertical, horizontal, left-falling, right-falling and break. And each basic stroke can be further subdivided. For example, the basic stroke “cross” can be divided into “long cross” and

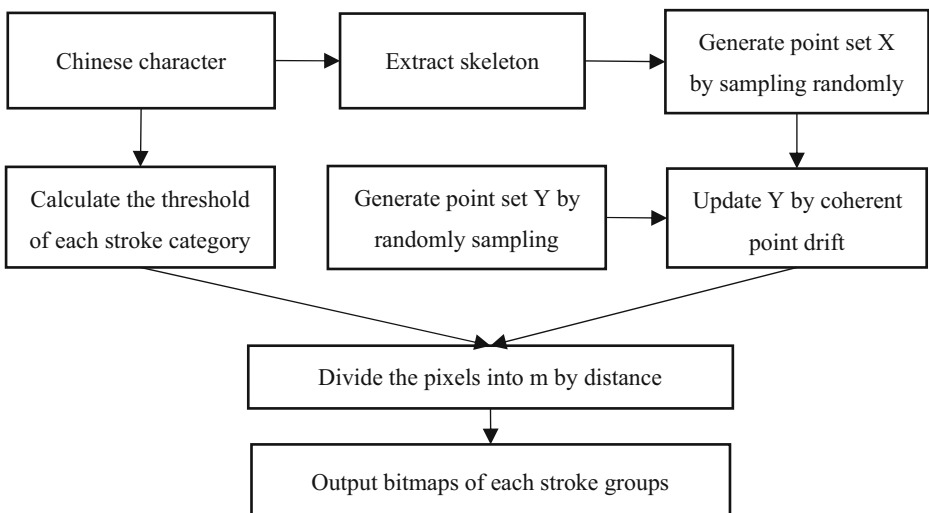


Fig. 2 The Chinese character strokes extraction process based on improved coherent point drift algorithm

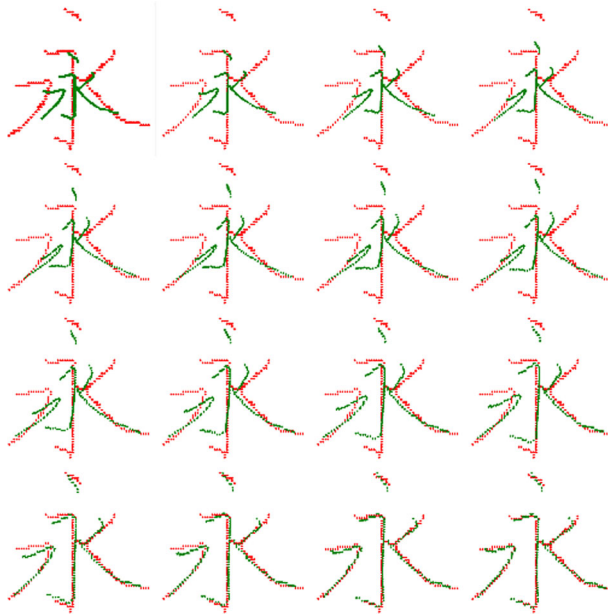


Fig. 3 The matching process during the Chinese character strokes extraction. The red points are sampled from the skeleton of input Chinese character, while the green points are sampled from the standard dataset

“short cross” by the length. It’s hard to categorize the strokes manually. Thus the k-means clustering algorithm is adopted to build the stroke library. And during the clustering process, the distance between two strokes is measured by the formula (1).

$$D = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (S_{i,j} - T_{i,j})^2 \quad (1)$$

We perform the k-means clustering algorithm on the stroke bitmaps and divide the strokes into 200 groups. The number of groups is selected by validation with k from 50 to 300. The details of the stroke extraction algorithm are shown below.

Input: the bitmaps of Chinese character

Output: the bitmaps of strokes extracted from Chinese character

Description:

- Step 1. Extract the skeleton of the Chinese character by erosion;
- Step 2. Generate the point set X by sampling from the skeleton randomly;
- Step 3. Perform the improved coherent point drift algorithm to update the point set Y for fitting the point set X;
- Step 4. Determine the stroke that each point in set X belongs to according to the nearest neighbor point of Set Y;
- Step 5. Determine the stroke that each point of Chinese character belongs to according to the distance between the point and the skeleton;
- Step 6. Place the points which belong to a same stroke into a group to generate the stroke bitmap.

3.3 Least squares conditional generative adversarial net

In original GAN, there is no control on generative modes of the data being generated. Meanwhile, the sigmoid cross entropy loss function is adopted by the original GAN, and it may lead to the vanishing gradients problem when the generative network is updated with the generated samples which lie around the decision boundary. It will result in no convergence. To enhance the controllability and stability of the GAN for generating the new Chinese font, two improvements have been put forward.

The first, we condition the model on the stroke information and the structure information of the font which is possible to direct the font generation process. In [17], Mehdi M. et al. have proved that the controllability of the generation process will be enhanced.

The second, we used the least squares loss function instead of the sigmoid cross entropy loss function for the discriminator. And Xudong M. et al. [16] have verified that the network will perform more stable during the learning process.

The font generation is bound up with the stroke and the font type, so the condition information attached to the GAN is consist of stroke type and font type. The condition information is described as a vector with 202 dims. The generation process of this vector is show in Fig. 4.

The generator includes four layers. The first layer is a FC layer which maps the condition information into a 50 dims vector which is combined with the random noise. The second layer is still a FC layer and extends the input to $8*8*128$ dims. The third

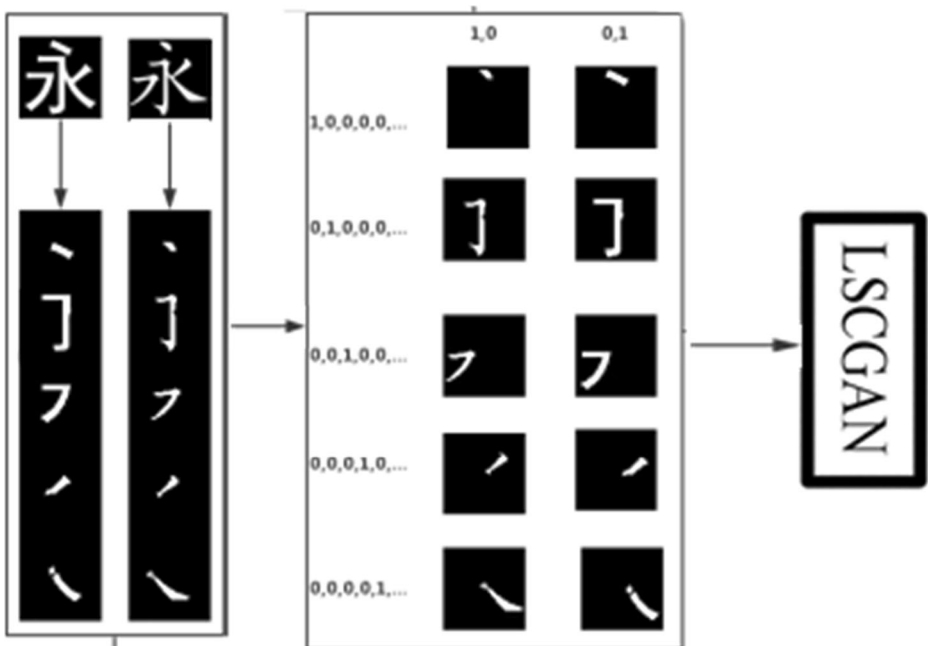


Fig. 4 Preprocessing of training data

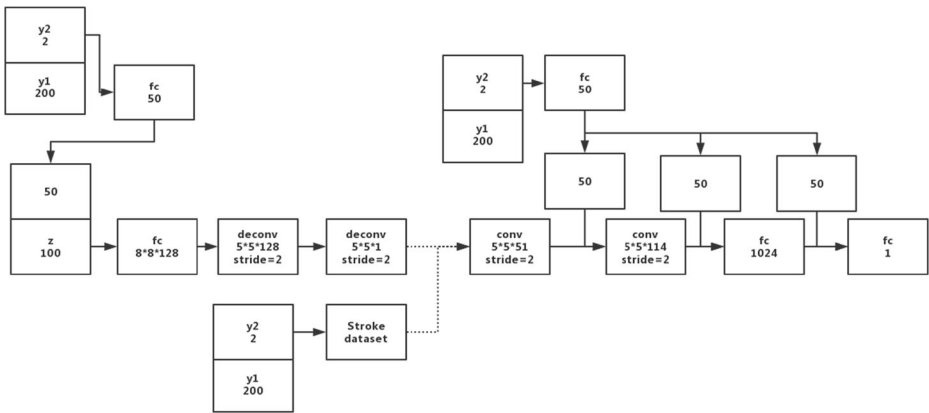


Fig. 5 The architecture of LSCGAN

layer is a deconvolution layer, the kernel size is $5 \times 5 \times 128$ and the stride is 2. The last layer is a deconvolution layer as well. The kernel size is $5 \times 5 \times 1$ and the stride is 2. The final output is $32 \times 32 \times 1$.

The discriminator includes five layers. Like the generator, the first layer of discriminator uses a FC layer to map the condition information into 50 dims. The second layer and third layer are both convolutional layer. The kernel size of second layer is $5 \times 5 \times 51$ and the stride is 2. The kernel size of third layer is $5 \times 5 \times 114$ and the stride is 2. The fourth layer is FC layer which output a vector with 1024 dims. The last layer is FC layer with 1 dim output.

The last layer of the discriminator is the least squares loss function. And during the process of Chinese font generation, the least squares loss function is defined as follows in detail:

$$\min_D L(D) = \frac{1}{2} E_{x \sim P_{data}(x)} \left[(D(x | \Phi(y_1, y_2)) - 1)^2 \right] + \frac{1}{2} E_{x \sim P_z(x)} \left[(D(G(z) | \Phi(y_1, y_2)))^2 \right] \quad (2)$$

$$\min_G L(G) = \frac{1}{2} E_{x \sim P_z(x)} \left[(D(G(z) | \Phi(y_1, y_2)) - 1)^2 \right] \quad (3)$$

Here, y_1, y_2 denote the conditions, $\Phi(y_1, y_2)$ is the conjunction of y_1 and y_2 with a nonlinear mapping. The detail model architecture of LSCGAN can be seen in Fig. 5.

4 Experiments

4.1 Datasets and implementation details

We use Hei and Kai font library as target fonts in this experiment. Hei font library is square, upright, plain, vertical and horizon, with no decoration. The strokes in Hei are

Table 1 Results of stroke extraction

Stroke type	Kai typeface	Hei typeface	Song typeface
cross	0.810	0.861	0.774
vertical	0.957	0.972	0.972
left-falling	0.940	0.880	0.916
right-falling	0.85	0.95	0.95
rising	0.9	0.78	0.76
dot	0.924	0.897	0.853
break	0.891	0.870	0.827

always the same in thickness and edges of strokes are square. Compared with Hei, strokes of Kai are thinner, and different in thickness. The start and finish of strokes are more artistic. We generate bitmaps of 3500 frequently used characters from TrueType files of Hei and Kai font library by simply drawing characters with ttf font files, and extracts 66,334 stroke images using algorithm 1. The standardized point set and corresponding stroke label used by algorithm 1 is extract from the JSON database from hanzi-writer,¹ an open-source library relating to Chinese characters and strokes. Based on this dataset, we conducted a simulation experiment of font style fusion and font generation.

The hardware platform the experiment conducted on has Intel CPU E5–2650, 64G memory, GPU Quadro K4200, and the software environment is Ubuntu 14.04 with CUDA v8.0 installed. And we use Tensorflow v1.0 as the deep network framework.

4.2 Stroke extraction

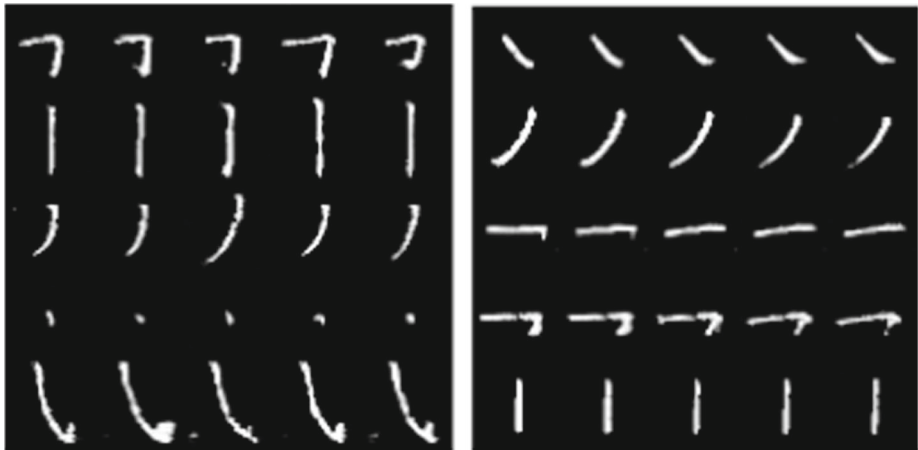
In order to verify the accuracy of the stroke extraction algorithm based on the improved coherent point drift algorithm, this paper carries out the simulation experiment of stroke extraction on 3500characters of Kai, Hei and Song font library. After the extraction, the k-means clustering algorithm is used to classify the strokes of each type, and finally the 200 fine-grained stroke categories are obtained. Then input them into the algorithm as primitive strokes. Stroke extraction experiment adopts improved coherent point drift algorithm, and the experimental results are shown in Table 1:

The results in Table 1 show that the stroke extraction algorithm proposed in this paper is effective.

4.3 Font generation by style fusion

To indicate the effectiveness and robustness of the LSCGAN proposed in this paper, we perform the font style fusion and generation experiments. In these experiments, the Hei typeface and the Kai typeface are used to verify the model. In our experiment, 3500 Chinese characters with Song typeface and a few Chinese characters with Hei typeface and Kai typeface are used as training data. And the outputs of our

¹ <https://github.com/chanind/hanzi-writer>



(a) Variety of strokes with same type and style per row (b) Variation of strokes showing style transfer

Fig. 6 The results of Chinese character strokes generation

experiments are the characters with Hei typeface, the characters with Kai typeface and the characters with fusion style.

Figure 6 shows the results of strokes generated by LSCGAN. Here the (a) of Fig. 6 is the output strokes with the same input of stroke type and font style. The results indicate the stroke type is controllable and the stroke style will change with the input. The (b) is the output strokes of fusing Hei typeface and Kai typeface together.

In order to further verify the effectiveness of fusion strategy, the one-hot vector of the font style is generated by interpolation. Suppose v_1 and v_2 denotes the one-hot vectors of the two font styles and v denotes the one-hot vector after fused, and α is the fusion coefficient.

$$v = \alpha * v_1 + (1-\alpha) * v_2 \quad (4)$$

In the experiments, the condition information change with the fusion coefficient α . And 3500 characters will be generated with a variety of condition information. Part of experimental results is presented in Fig. 7.

The results in Fig. 7 shows that the LSCGAN can not only generate the font with specific style stably, but also fuse two fonts style together to generate the new font.

4.4 Discussion

In this paper, the fusion of font styles wasn't performed on the whole Chinese character but the strokes. So the correctness of the Chinese character structure is guaranteed. Our method wasn't good for the font styles without clear strokes, such as "cursive font". And the methods proposed by Pengyuan Lyu et al. [15] and Chuang Yu-Neng et al. [3] will work a lot better for "cursive font". Both of their methods generated font style based on character image. And the image based model is easy to generate fonts which will look pretty good, especially those fonts with unclear strokes, but the structure of the Chinese character may be incorrect.

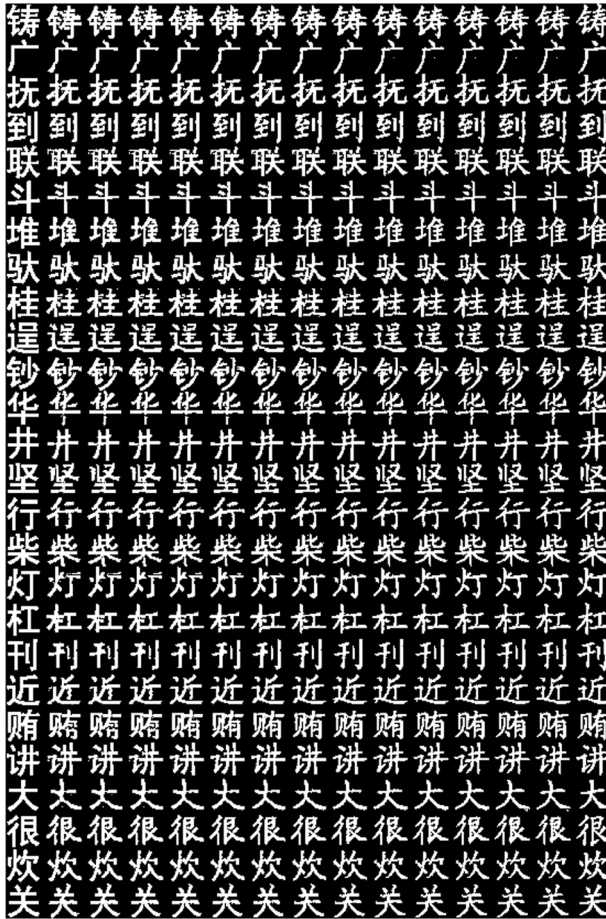


Fig. 7 The results of font style fusion and generation. The first column is the standard Hei typeface, the second column is the Hei typeface generated by our LSCGAN. The columns from the third to thirteenth are the fonts with fusion style, and the fusion coefficient $\alpha = 0.05, 0.1, 0.2, \dots, 0.9, 0.95$

5 Conclusions

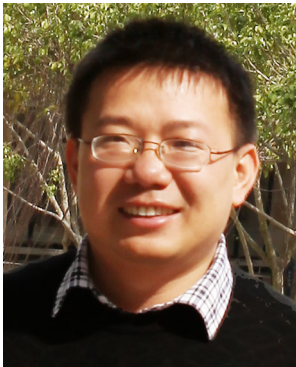
We propose an automatic Chinese font generation method based on least squares condition generative adversarial net (LSCGAN). The LSCGAN adopt the least squares loss function for the discriminator based on the conditional generative adversarial nets. To enhance the controllability of the GAN, a condition is attached to the input. In the meantime, the least squares loss function can speed up convergence. In our method, the strokes are treated as one of the key features of Chinese character, and a new stroke can be generated based on the fusion of two different fonts via LSCGAN. Numerous experiments are conducted and the results demonstrate that LSCGAN perform more stable and diversification with simply training process and no need for priori information. The future work includes learning the structure information of the font and improving the LSCGAN to complement the Chinese character font generation framework presented here.

Acknowledgements This work is supported by the National Key Technology R&D Program (No. 2017YFC011300, No. 2016YFB1001503), the Nature Science Foundation of China (No. 61422210, No. 61373076, No. 61402388, and No. 61572410), the Nature Science Foundation of Fujian Province, China (No. 2017 J01125).

References

1. Arjovsky M, Chintala S, Bottou L (2017) Wasserstein GAN. arXiv:1701.07875v2
2. Chen X, Duan Y, Houthoofd R et al (2016) InfoGAN: interpretable representation learning by information maximizing generative adversarial nets. *Neural Information Processing Systems (NIPS)*
3. Chuang Y-N, Huang Z-Y, Tsai Y-L (2017) Variational grid setting network. arXiv preprint arXiv:1710.01255
4. Denton E, Chintala S, Szlam A et al (2015) Deep generative image models using a Laplacian pyramid of adversarial networks. arXiv:1506.05751
5. Gatys LA, Ecker AS, Bethge M (2015) A neural algorithm of artistic style. arXiv:1508.06576v2
6. Goodfellow IJ, Pougetabadié J, Mirza M et al (2014) Generative Adversarial Nets. *Adv Neural Inf Process Syst* 3:2672–2680
7. Gregor K, Danihelka I, Graves A et al. (2015) DRAW: a recurrent neural network for image generation. arXiv:1502.04623v2
8. He K, Zhang X, Ren S et al (2016) Deep residual learning for image recognition. 2016 IEEE conference on computer vision and pattern recognition (CVPR), pp 770–778
9. Isola P, Zhu JY, Zhou T et al (2016) Image-to-image translation with conditional adversarial networks. arXiv:1611.07004v1
10. Johnson J, Alahi A, Li FF (2016) Perceptual losses for real-time style transfer and super-resolution. arXiv:1603.08155v1
11. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. *International conference on neural information processing systems*. Curran associates Inc., pp 1097–1105
12. Li C, Jiangqing W, Bo L et al (2013) Algorithm on strokes separation for Chinese characters based on edge. *Computer. Science* 40(7):307–311
13. Lian Z, Zhao B, Xiao J (2016) Automatic generation of large-scale handwriting fonts via style learning[C]// SIGGRAPH ASIA 2016 technical briefs. ACM, pp 12
14. Liu MY, Tuzel O (2016) Coupled generative adversarial networks. arXiv:1606.07536v2
15. Lyu P et al (2017) Auto-encoder guided GAN for Chinese calligraphy synthesis. arXiv preprint arXiv:1706.08789
16. Mao X, Li Q, Xie H et al (2017) Least squares generative adversarial networks. arXiv:1611.04076v3
17. Mirza M, Osindero S (2014) Conditional generative adversarial nets. arXiv:1411.1784
18. Mouchère H et al (2013) A dynamic time warping algorithm for recognition of multi-stroke on-line Handwritten characters. *Journal of South China University of Technology* 41(7). <https://doi.org/10.3969/j.issn.1000-565X.2013.07.000>
19. Myronenko A, Song X (2010) Point set registration: coherent point drift. *IEEE Trans Pattern Anal Mach Intell* 32(12):2262–2275
20. Odena A (2016) Semi-supervised learning with generative adversarial networks. *International conference on machine learning (ICML)*
21. Odena A, Olah C, Shlens J (2016) Conditional image synthesis with auxiliary classifier GANs. arXiv:1610.09585
22. Radford A, Metz L, Chintala S (2015) Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv:1511.06434v2
23. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556
24. Springenberg JT (2015) Unsupervised and semi-supervised learning with categorical generative adversarial networks. arXiv:1511.06390v2
25. Szegedy C, Liu W, Jia Y et al (2015) Going deeper with convolutions. 2015 IEEE conference on computer vision and pattern recognition (CVPR), pp 1–9
26. Taigman Y, Polyak A, Wolf L (2016) Unsupervised cross-domain image generation. arXiv:1611.02200v1
27. Tan CL, Cao R (2000) A model of stroke extraction from Chinese character images. In: *Proceedings 15th International Conference on Pattern Recognition, Barcelona, Spain, vol 4*, pp 368–371
28. Wu L, Xia Y, Zhao L et al (2017) Adversarial neural machine translation. arXiv:1704.06933v3
29. Xiafe Z, Jiayan L et al (2016) Extracting Chinese calligraphy strokes using stroke crawler. *Journal of Computer-Aided Design & Computer Graphics* 28(2):301–309
30. Xiaohu M, Yulong L, Zhigeng P et al (1999) The automatic generation for Chinese outline font and It's transformation method. *Journal of Chinese Information Processing* 13(2):46–50

31. Xu S, Jin T, Jiang H et al (2009) Automatic Generation of Personal Chinese Handwriting by Capturing the Characteristics of Personal Handwriting. Conference on Innovative Applications of Artificial Intelligence, July 14–16, 2009, Pasadena, California, USA. DBLP
32. Zhang XY, Yin F, Zhang YM et al (2017) Drawing and recognizing Chinese characters with recurrent neural network. TPAMI <https://doi.org/10.1109/TPAMI.2017.2695539>
33. Zhao S et al (1949) Predicting personalized image emotion perceptions in social networks. IEEE Trans Affect Comput 99:1–1
34. Zhao S, Yao H, Sun X (2011) Affective video classification based on Spatio-temporal feature fusion. Sixth international conference on image and graphics IEEE computer. Society:795–800
35. Zhao S et al (2015) Strategy for dynamic 3D depth data matching towards robust action retrieval. Neurocomputing 151:533–543
36. Zhao S et al (2016) Predicting personalized emotion perceptions of social images. ACM on multimedia conference ACM, pp 1385–1394
37. Zhao S et al (2017) Continuous probability distribution prediction of image emotions via multi-task shared sparse regression. IEEE Trans Multimedia 99:1–1
38. Zong A, Zhu Y (2014) StrokeBank: automating personalized chinese handwriting generation twenty-eighth AAAI conference on artificial intelligence. AAAI Press, Palo Alto, pp 3024–3029



Xianming Lin received the Ph.D. degree in Intelligent Multimedia Information Processing from the School of Information Science and Engineering, Xiamen University, Xiamen, China, in 2014. He is an Assistant Professor with Xiamen University. His current research interests include mobile visual retrieval, computer vision and machine learning.



Jie Li received the B.E. degree from the school of Information Science and Engineering, Xiamen University, Xiamen, China, in 2017. And from 2017, as a Master Student, he focuses on the area of deep learning and computer vision.



Hualin Zeng received the Ph.D. degree in Intelligent Multimedia Information Processing from the School of Information Science and Engineering, Xiamen University, Xiamen, China, in 2016. She is an Assistant Professor with Xiamen University. Her current research interests include natural language processing, computer vision and machine learning.



Rongrong Ji serves as the Professor of Xiamen University, where he directs the Intelligent Multimedia Technology Laboratory (<http://imt.xmu.edu.cn>) and serves as a Dean Assistant in the School of Information Science and Engineering. He has been a Postdoc research fellow in the Department of Electrical Engineering, Columbia University from 2010 to 2013, worked with Professor Shih-Fu Chang. He obtained his Ph.D. degree in Computer Science from Harbin Institute of Technology, graduated with a Best Thesis Award at HIT. He had been a visiting student at University of Texas of San Antonio worked with Professor Qi Tian, and a research assistant at Peking University worked with Professor Wen Gao in 2010, a research intern at Microsoft Research Asia, worked with Dr. Xing Xie from 2007 to 2008. He is the author of over 40 peer-reviewed journals and conferences including IJCV, TIP, TMM, ICCV, CVPR, IJCAI, AAAI, and ACM Multimedia. His research interests include image and video search, content understanding, mobile visual search, and social multimedia analytics. Dr. Ji is the recipient of the Best Paper Award at ACM Multimedia 2011 and Microsoft Fellowship 2007. He is a guest editor for IEEE Multimedia Magazine, Neurocomputing, and ACM Multimedia Systems Journal. He has been a special session chair of MMM 2014, VCIP 2013, MMM 2013 and PCM 2012, would be a program chair of ICIMCS 2016, Local Arrangement Chair of MMSP 2015. He serves as reviewers for IEEE TPAMI, IJCV, TIP, TMM, CSVT, TSMC A\B\C and IEEE Signal Processing Magazine, etc. He is in the program committees of over 10 top conferences including CVPR 2013, ICCV 2013, ECCV 2012, ACM Multimedia 2013–2010, etc.