The Eleventh International Conference on Advanced Computational Intelligence,
June 7-9, 2019, Guilin, China

Drawing Order Recovery based on deep learning

1st Rui Zhang
Guangxi Key Laboratory of Trused Software
Guilin University of Electronic Technology
Guilin, China
zr_hpu@163.com

2nd Jinlong Chen
Guangxi Key Laboratory of Cryptography & Information Security of Guilin University of Electronic Technology
Guilin, China

3rd Minghao Yang
National Laboratory of Pattern Recognition
Chinese Academy of Sciences
BeiJing, China

***Abstract*—Humans have the ability to recover the order from static handwritten images, after a large amount of data training, the machine may learn some patterns in the training data to imitate or learn a certain skill similar to humans. To overcome the problem of sequence recovery of static image strokes, this paper proposes a stroke recovery method based on deep convolutional neural network model. In the model training phase, by using the two-dimensional static handwritten image, the process of writing a font is convert into three channels includes strokes that have been written, possible positions of next strokes, and the completed font, and state of the input sample are quantified. In the recovery phase, the restored font is preprocessed to obtain the stroke segments of the font, and the trained model is used to evaluate the sequential combination of different stroke segments, so as to obtain the correct stroke order. With no more than one hundred of characters' writing experiences, the proposed method performs robustly and competitively among multi-writer handwriting DOR tasks.**

Keywords—handwriting, time series information, deep learning, order recovery, Convolutional Neural Networks

I. INTRODUCTION

Information management is an important way and method of social modernization, in which the information management of handwriting is an important application. At present, many Chinese character information management techniques are carried out for the specification of printed Chinese characters. In actual life and application scenarios, there are still a large number of handwritten materials. The fonts in the materials are not standardized printed fonts, but are written by different individuals, therefore, there is an urgent need for an effective handwritten material information management method. Handwriting recognition can be divided into offline handwriting recognition and online handwriting recognition due to different data collection methods[1]. The object processed by online recognition is a one-dimensional dynamic sequence recorded by a handwriting input device. From the aspect of information storage, one-dimensional online sequence data is easier to store and manage and analyze than offline two-dimensional static during writing. The generated dynamic information realizes the conversion of offline handwritten text data to online handwritten text data, improves the correct rate of handwritten text recognition, and effectively realizes the information management of handwritten characters. In addition, writing sequence recovery technology is also widely used in intelligent writing robots and intelligent robot arms[2,3].

At present, the dynamic sequence recovery methods are mainly based on graph models[4,5,6,7,8], writing rules, deep learning methods based on convolutional neural networks(CNN) and recurrence neural networks (RNN). The graph model based method usually first smooth and refines the image data, and then models the relationship between the font strokes. Generally, the strokes are regarded as nodes in the graph, so that the stroke recovery problem of the font is converted into searching for a path that meets certain conditions in the graph, and the found paths satisfy the maximization or minimization of an objective function. The design of the objective function is generally based on a minimum energy cost criterion on the assumption that people's writing process is always carried out in one of the most convenient and smoothest paths. The energy of the arm muscle spent on writing is minimal[9]. The writing rule-based recovery method utilizes the structure of the text, and the writing order of the strokes is obtained by describing and modeling the relationship of the text strokes in combination with the writing rules. Deep learning-based recovery methods are mainly implemented using convolutional neural networks (CNN) or cyclic neural networks (RNN), which mainly represent fonts in the form of sequences that can be trained by the network. A conditional generation model based on RNN is designed in [10]. This model uses RNN and character matrix index to model the writing process of the font, and uses the mixed Gaussian matrix (GMM) to model the specific position in writing process. The model can be used to generate the specified handwritten style of Chinese characters, but the model is only a Chinese character generation model, and there is no stroke order restoration for a specific handwritten Chinese character. A handwritten character recovery model that combining LSTM and CNN is designed by [11]. The model consists of two parts. The first part is the font sequence coding module. The module consists of convolutional neural network CNN and coding LSTM. The purpose is to encode handwriting into a sequence that can be used for LSTM training. The second part of the model is The decoding module of the sequence, the module is composed of decoding LSTM, and its main function is to convert the hidden layer vector outputted by the encoding module into sequence coordinates. The above CNN and RNN-based sequence generation or handwritten text recovery methods are mainly for handwriting in specific languages such as English or Hindi. These methods have achieved good recovery effects on certain data sets, however, this kind of method is not ideal for the recovery of Chinese, especially handwritten Chinese characters.

This paper proposes a stroke recovery method based on convolutional neural network. The method converts the process of handwritten font writing into different three-channel input samples, evaluates and then calculates these input samples, obtains the corresponding output labels of the samples, and finally uses the trained model for Handwritten recovery. At the time of recovery, the model is used to evaluate the order in which the stroke segments are written, and the order of strokes with the highest probability is obtained. The experiment proves that the model proposed in

978-1-5386-7732-2/19/$31.00 ©2019 IEEE

this paper has a better recovery effect on handwritten Chinese characters within 5 strokes .

II. HANDWRITTEN STROKE ORDER RECOVERY MODEL

A. *Pipeline of the Proposed Method*

The handwritten stroke recovery process based on convolutional neural network proposed in this paper is shown in Figure 1.The method is mainly divided into two parts: the training of stroke recovery network and the stroke restoration of handwritten Chinese characters. The stroke recovery network training part first uses the data in handwritten database to make input samples that meets the requirements of the network, then use the prepared samples to perform target optimization training on the network, and obtain a stroke prediction network with optimized weight parameters. The subsequent stroke recovery phase uses the trained network to predict the static image, and obtains the most likely stroke segment writing order, thereby restoring the drawing order of the handwritten font.

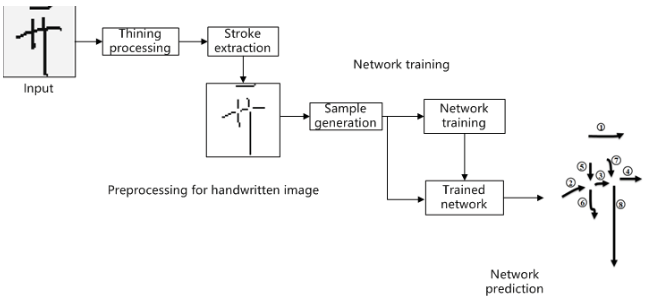

Fig. 1. The pipleline of the proposed model

In the training phase of stroke recovery network, it is divided into image data preprocessing, network training sample generation and sample corresponding label generation, network parameter training and optimization, etc. The data used by the stroke recovery network is a 48×48 binary image and a coordinate sequence when the font is written. In that image, the content with a pixel value of one is the handwritten pixel, the content with a pixel value of zero is regarded of background. Image data is preprocessed to convert a handwritten font with a width greater than 1 and a stroke intersections to a single-pixel stroke segment containing only the font skeleton. All stroke segments form the outline of the font, the restoration of the font writing order becomes the restoration of the writing order of all the stroke segments. The network divides the dynamic writing process of each time into a three-channel training sample containing the current writing state, the completed writing state and the complete font, each channel is 48×48 image data containing different stroke segments, The first channel is the stroke segment that has been completed, the second channel is the position of the endpoint of the stroke segment currently being written, and the third channel is the complete font. For each font in the database, first, fix the first channel, change the position of the end point of the stroke segment in the second channel, and generate different input samples and its labels represented by a number , according to the distance between the end point of the stroke segment and the correct position in the second channel, then update the stroke segments that have been written in the first channel and continue to change the contents of the second channel, resulting in different sample inputs and corresponding labels.

In handwritten stroke recovery stage, any given handwritten font image is preprocessed to obtain all the stroke segments of the font, and for each stroke in the font writing process, a corresponding three-channel input sample is produced. The trained network model determines the most likely writing position of the stroke. Then, in the case of the previous determination, the stroke determined by the previous stroke is used to continue to predict the position of the next stroke until the completion of the restoration of the font stroke order is completed.

B. *Handwritten image preprocessing and train sample*

The preprocessing of handwritten images includes the extraction of skeleton images, the extraction of stroke segments and the deletion of coordinate writing coordinate sequences. The image data of the handwriting obtained from the online handwritten font database and the coordinate sequence corresponding to the writing of the font. The image data is a binary image of 48×48 size, and the coordinate sequence is represented by the (1). The sequence of the coordinate sequence represents the order of writing of each stroke when the font is written.

$$[x_1, y_1], [x_2, y_2], \dots\dots, [x_n, y_n] \quad (1)$$

Since the data obtained from the online handwritten font database is the result of digitizing the original font written by the writer on the data collection device, as shown in 2(a), the average width of the font stroke is greater than one. In order to convert the two-dimensional image data into a dynamic time series, the width of the font stroke can be processed so that the width of the stroke is one pixel. The above object can be achieved by refining the handwritten image, and 2(b) is a result of refining the original image of the handwritten image using the skeleton extraction algorithm. In the refined handwritten font image, each pixel has at most 3 adjacent pixels in the surrounding 8 neighborhoods. Therefore, each stroke can find a unique starting point and endpoint except the intersection between the strokes. Thus, when restoring the handwritten font, as long as the order of each stroke end point is determined, the specific position and writing order of the line can be determined from the start point and the end point of the stroke. While refining the image, the corresponding coordinate sequence of the font is also processed, so that the refined image corresponds to the coordinate sequence of the font.

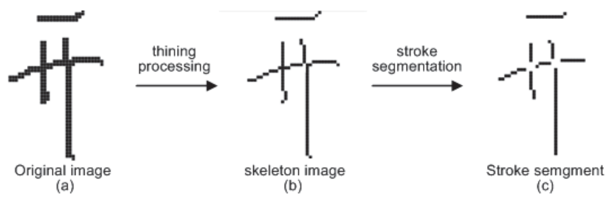

Fig. 2. Handwritten font image preprocessing

After the pre-processing of the handwritten image, the obtained segment of the stroke is shown in 2(c). In order to use the neural network to obtain the sequence of each stroke in the font writing process, the possible states and positions of the strokes at different times in the font writing process can be made into three-channel image data that the network can process. The first channel represents the stroke that has been written, the second channel represents the position of the next stroke, and the third channel represents the complete font. Fig. 3(a) is a fragment diagram of a stroke after image preprocessing. Fig. 3(b) is a three-channel input sample generated by the font and a corresponding fractional label, taking the word “开”, in Fig. 3(a), as an example, in the process of writing the second stroke of the font, the starting point of the second stroke may have different positions. Therefore, in Figure 3(b), the first channel is the first stroke that has been written, and the next possible position is selected in the second channel. The third channel is the complete character “开”, three channels together form a sample. The sample is evaluated and quantified, and the label of sample is assigned based on the distance between the next pen down position and the real position in the second channel. It can be seen that the sample is a characterization of a certain writing state of the font, and the sample can be used to train a network that predicts the next possible position of writing in a different writing time state.

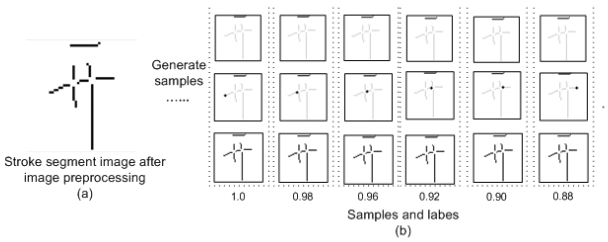

Fig. 3. Processing of training sample generation

C. Handwritten stroke recovery network

The handwritten stroke recovery network uses an improved VGG-16 convolutional neural network model. The structure of the entire network is shown in Fig .4. The input of the network is three-channel image data, and the image is a 48×48 binary image, wherein the font pixel is one, and the non-font portions is zero. The output of the network is the fractional value of the sample. (2) represents the training process of the network.

$$arg_{w,b}min\frac{1}{2|D|}\sum\left||F(x_i)-y_i|\right|_2^2 \quad (2)$$

In the above formula, w and b are the weight parameters of the network , D is the dataset of the whole training, F is the network model, x_i is the i-th three-channel input sample, y_i is the label corresponding to the input sample, and the network uses the minimum mean square error of all samples in dataset D to optimize the parameters of the entire model.

The entire network model consists of a series of convolutional layers, modified linear activation layers (ReLU), and fully connected layers. The first part of the network uses multi-level convolutional layer as the extraction of image features. There are 6 layers of convolutional layers. The number of convolution kernels of the first and second convolutional layers is 32, the size is 3×3 the convolution step is 1. The number of convolution kernels of the third layer is 64, and the size is 3×3, and the convolution stride is 2. The fourth layer convolution kernel number is 64, the size is 3×3, and the convolution stride is 2. The fifth layer convolution kernel number is 128, and the size is 3×3, the step is 2. The sixth layer convolution kernel number is 128, the size is 3×3, and the convolution step is 1. In order to increase the nonlinearity of the network, a nonlinear activation layer is used after each convolutional layer, where the activation layer uses a modified linear activation function (ReLU). After the first part of the convolution layer that extracts different levels of image features, the extracted features are combined using a fully connected layer to form a one-dimensional feature vector, so the seventh layer of the network model uses a full connection layers with 256 nodes. In order to prevent over-fitting of the network, add a dropout layer after the fully-connected layer. Finally, the eighth layer uses the fully connected layer with node 1 as the output of the network. The forward propagation of each layer of the network is as follows:

$$F_0(x_i)=x_i \quad (3)$$

$$F_l(x_i)=max(0,W_l*F_{l-1}(x_i)+b_l), \quad l=1\ldots,L-1 \quad (4)$$

$$F(x_i)=W_L*F_{L-1}(x_i)+b_L \quad (5)$$

where x_i is the *i-th* handwritten image sample, l is the number of layers of the network, w,b is the weight parameter of the network, F is the model of the network, F_0 is the input layer of the network, and F_i is the middle layer of the network.

III. Experiments

In order to verify the validity of the stroke order recovery method in this paper, the deep learning framework Keras is used to build the network model proposed in this paper. The train sample processing and network model training are completed. The experimental environment of this experiment is 64-bit Windows 10 system, Intel(R) Core(TM) i7-6700K CPU 4.00GHz, 16GB, NVIDIA GTX1080GPU .

A. Training data and model training

The data set used in this paper is the CASIA-OLHWDB handwritten database [12], which contains 1020 written online handwritten include both isolated characters and handwritten texts. The isolated characters samples is divided into three database: OLHWDB1.0-1.2; Among them, OLHWDB1.0 includes 3866 commonly used Chinese characters (including 3740 Chinese characters commonly used in GB2312) and 171 English numeric symbols, which are written by 420 people, a total of 420 sets, and 1680258 valid samples. OLHWDB1.1 includes 3755 GB2312 first-level Chinese characters, 171 characters, which are written by 300 people, 300 sets in total, and 1172907 valid samples. From the OLHWDB1.0 and OLHWDB1.1, 100 commonly used Chinese characters with higher frequency of writing were selected from 300 different people. The Chinese characters written by 240 people were randomly selected as the training set, and the remaining 60 Chinese characters were test set. Generate training samples for each handwriting using the method described in Section II-B. The training process of the network model takes the minimum mean square error of all sample in the training set as the objective function, and uses the back propagation algorithm to optimize and update the network weights, in order to make the network the network converge in a certain time, the random gradient learning algorithm with momentum is used to optimize the weight, The learning rate of the random gradient algorithm in training is set to 0.01, and the attenuation rate is 1e-6, set the momentum term of the random gradient learning algorithm to 0.9. Due to the large number of training samples, the weight update is performed by batch processing, which can accelerate the convergence process of the network to a certain extern. The size of the batch is set to 16.

$D_n U_n$. Table I. lists the correct rates for stroke recovery using the algorithm described.

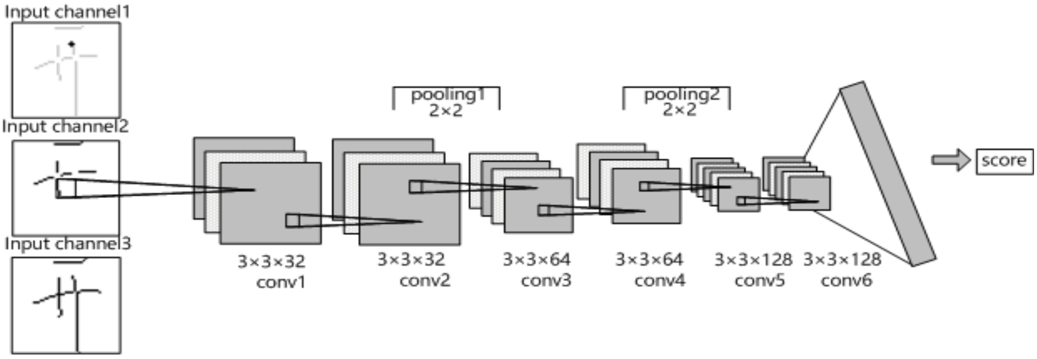

Fig. 4. The structure of handwritten stroke recovery neural network

B. *Experimental resuls and analysis*

When the network converges to a certain extent, the network loss rate will become relatively small and the network loss rate will be stable for a period of time, the network has reached convergence. The trained model is tested using test samples outside the training set. The recovery process of some fonts in the test set is shown in Fig. 6. The test data are labeled by categories according to whether the test fonts are *known* (D_y) or *unknown* (D_n) ,whether the font is written by someone in the training set(The writer's other fonts are include in the network training set as U_y, any font written by the writer is not include in the network training set as U_n) , we set the test set as $D_y U_n$,$D_n U_y$, $D_n U_n$.

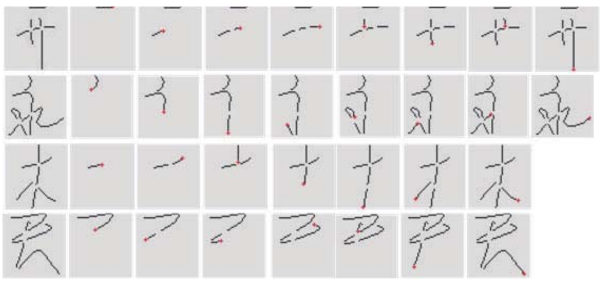

Fig. 5. Handwriting drawing order recovery for some Chinese characters

TABLE I. RESULTS OF THE RECOVERY OF 1000 HANDWRITTEN CHINESE CHARACTERS

	Number of strokes		
	0~5	***5~10***	***>10***
Quantity	289	607	104
$D_n U_n$	67.8%	53.6%	12.8%
$D_n U_y$	78.3%	60.9%	18.85
Average	73.0%	57.2%	15.8%

From the test set of OLHWDB1.0 and OLHWDB1.1, 1000 handwritten Chinese characters written by 60 writers were randomly selected as test samples to test the correct rate of network. Among the 60 selected writers, 30 writers as $D_n U_y$, and the words written by the other 30 writers set as $D_n U_n$. Table I. lists the correct rates for stroke recovery using the algorithm described.

From the Table I, we can see that the number of strokes in the 1000 test samples is the range of 5-10 strokes, there are 607, followed by the samples within 5 strokes, there are 289, and more than 10 strokes, there are 104 in total. For the above results, we can see that whether for the $D_n U_n$ type test sample or the $D_n U_y$ type test sample, the accuracy of stroke recovery gradually decreases with the increase of the number of strokes. For Chinese characters of 0-5 strokes, The average accuracy of stroke order recovery is 73.0%, the average accuracy of 5-10 strokes is 57.2%, and for more than 10 strokes, the average accuracy is only 15.8%, It shows that with the increase of the number of handwritten strokes, the difficulty of restoring the sequence of strokes is gradually increasing, and the accuracy of network recovery is also reduced. Especially when the number of strokes of handwritten fonts exceeds 10, the accuracy of network stroke recovery is drastically reduced. At the same time, it can ben seen that different types of test set samples also have an impact on the accuracy of network stroke order recovery. In different intervals of strokes, the $D_n U_y$ type of test sample are more accurate than the $D_n U_n$ type of test sample, it shows that personal writing style has a certain impact on the accuracy of the network., If the font of a particular writer is included in the training network, when the network restores the other fonts written by the writer, there is a high recovery accuracy relative to the same font written by a completely strange writer.

IV. CONCLUSION

Aiming at the problem of handwritten stroke recovery in handwriting, this paper proposes a handwritten stroke recovery method based on convolutional neural network, the method first performs refinement, stroke extraction and coordinate extraction on the handwritten image, then convert the writing process of the font at each moment into sample data with three channels. Each sample is a description and evaluation of the next possible position in the current writing state, using the generated samples, it can train a neural network model that predicts the next writing state, thus realizing the recovery of the handwritten stroke order. Through experiments, it can be concluded that the method can effectively recovery the stroke order of the simple handwriting within five strokes, but as the number of strokes increase, the accuracy of the network decrease, therefore, the effective recovery of the sequence of complex font strokes is the focus of our next study.

ACKNOWLEDGMENT

This research work is supported by the grant of Guangxi science and technology development project (No: AC16380124, AB18126053), the grant of Guilin science and technology development project (No: 20180107-4)， the grant of Guangxi Science Foundation(No: 2017GXNSFAA198226), the grant of Guangxi Key Laboratory of Trusted Software of Guilin University of Electronic Technology (No: KX201601), the grant of Guangxi Colleges and Universities Key Laboratory of Intelligent Processing of Computer Images and Graphics of Guilin University of Electronic Technology (No: GIIP201602,GIIP201702), the grant of Guangxi Cooperative Innovation Center of cloud computing and Big Data, Supported by Guangxi Colleges and Universities Key Laboratory of cloud computing and complex systems (No: YD16E11,YD17X06), the grant of Guangxi Key Laboratory of Cryptography & Information Security of Guilin University of Electronic Technology (No: GCIS201601, GCIS201602, GCIS201603), and the grant of Innovation Project of GUET Graduate Education (2017YJCX55).

REFERENCES

[1] L. Jin, Z. Zhuo, Z. Yang, W. Y, Z. Xie, J. Sun, " Applications of Deep learning for Handwritten Chinese Character Recognition: A review," in Acta Automatica Sinica, vol. 42, no. 8, pp. 1125-1141, Aug. 2016.

[2] F. Chao et al. " A robot calligraphy system: From simple to complex writing by human gestures, " in Engineering Applications of Artificial Intelligence, vol. 59, pp. 1-14,2017.

[3] J. Li, W. Sun, M. Zhou and X. Dai, "Teaching a calligraphy robot via a touch screen," 2014 IEEE International Conference on Automation Science and Engineering (CASE), Taipei, 2014, pp. 221-226.

[4] Yu Qiao, M. Nishiara and M. Yasuhara, "A Framework Toward Restoration of Writing Order from Single-Stroked Handwriting Image," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 11, pp. 1724-1737, Nov. 2006.

[5] G. Boccignone, A. Chianese, L.P. Cordella, A. Marcelli, "Recovering dynamic information from static handwriting," in Pattern Recognition,vol. 26,no. 3, pp. 409-418,1993.

[6] Yu Qiao and M. Yasuhara, "Recovering Drawing Order from Offline Handwritten Image Using Direction Context and Optimal Euler Path," 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, Toulouse, 2006, pp. 437-441.

[7] A. Sharma, "Recovery of drawing order in handwritten digit images," 2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013), Shimla, 2013, pp. 437-441.

[8] Yu Qiao and M. Yasuhara, "Recovering dynamic information from static handwritten images," Ninth International Workshop on Frontiers in Handwriting Recognition, Kokubunji, Tokyo, Japan, 2004, pp. 118-123.

[9] T. Nagoya , H. Fujioka , "Recovering Human-Like Drawing Order from Static Handwritten Images with Double-Traced Lines," Lecture Notes in Electrical Engineering, 2013, pp. 941-948.

[10] X. Zhang, F. Yin, Y. Zhang, C. Liu and Y. Bengio, "Drawing and Recognizing Chinese Characters with Recurrent Neural Network," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 4, pp. 849-862, 1 April 2018.

[11] A. Kumar Bhunia et al., "Handwriting Trajectory Recovery using End-to-End Deep Encoder-Decoder Network," 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, 2018, pp. 3639-3644.

[12] C. Liu, F. Yin, D. Wang and Q. Wang, "CASIA Online and Offline Chinese Handwriting Databases," 2011 International Conference on Document Analysis and Recognition, Beijing, 2011, pp. 37-41.