ICIC Express Letters
Volume **15**, Number **11**, November **2021**
 ISSN 1881-803X
pp. **1187–1195**

CHINESE CALLIGRAPHY RECOGNITION SYSTEM BASED ON CONVOLUTIONAL NEURAL NETWORK

Wenyi Cui and Kohei Inoue*

Faculty of Design
Kyushu University
4-9-1, Shiobaru, Minami-ku, Fukuoka 815-8540, Japan
cuiwenyi1996@outlook.com; *Corresponding author: k-inoue@design.kyushu-u.ac.jp

Received April 2021; accepted June 2021

Abstract. *This paper presents a Chinese calligraphy recognition system which can extract text from an image of Chinese calligraphy and recognize its style of calligraphy fonts. A multi-label convolutional neural network (CNN) recognition model is created and trained to recognize both textual content and font of single Chinese character at the same time. A large number of calligraphy images of single Chinese character are collected and prepossessed to form a dataset for training the model. Several images of calligraphy works of ancient Chinese calligraphers are used to evaluate the performance of the proposed system, and the experimental results showed the capability of the proposed system to recognize Chinese calligraphy.*
Keywords: Chinese calligraphy recognition, Handwritten character recognition, Convolutional neural network (CNN), Deep learning, Chinese character, Seal script, Official script, Regular script, Running script, Cursive script

1. **Introduction.** As image digitization develops, character recognition has become a research hotpot in the field of computer vision, which shows great value in data entry of paper documents. Compared with printed document, the recognition of handwriting is more difficult as the shape of handwriting characters is more irregular.

As a kind of handwriting art form, there are basically 5 types of fonts of Chinese calligraphy: seal, official, regular, running and cursive scripts. However, the shape of characters in Chinese calligraphy varies greatly among different calligraphist and differs a lot from daily use standard font, which cause considerable challenge for people to recognize the content of calligraphy work immediately.

Therefore, a real-time calligraphy recognition system can help calligraphy amateurs better understand calligraphy works by present font and textual content of input calligraphy image. The system can also be used to digitize calligraphy works by simply input the image of calligraphy work instead of typing out the text manually.

In this paper, we designed and implemented a calligraphy recognition system based on convolutional neural network. The system can recognize both font and textual content with greater correct rates than previous studies. We established a calligraphy character data set to train the network and used images of different calligraphy works to test the system feasibility.

The paper is divided into 6 sections as follows.

Section 1 is introduction. In this section, we introduced the background of character recognition and the research significance of Chinese calligraphy recognition.

Section 2 is related work. In this section, we summarized research status of Chinese calligraphy recognition and compared them on correct rates.

DOI: 10.24507/icicel.15.11.1187

Section 3 is the proposed CNN model. In this section, we described operating principles of each layer of the proposed CNN model. Then, we analyzed recognition results for the proposed model on training set and testing set. Finally, we compared correct rates of the proposed model with related research and confirmed the advantage of the proposed CNN model.

Section 4 is system architecture. In this section, we described the detailed architecture of the proposed system. The system is divided into four modules based on functionality: machine learning training, character segmentation, character recognition and user interface modules. We also introduced working process of the system.

Section 5 is example of system operations. In this section, we presented pages of the proposed system by different input images.

Section 6 is conclusion. In this section, we analyzed execution result of the system and listed future improvements.

2. **Related Work.** There are two main directions for the existing studies of Chinese calligraphy: recognition of font and recognition of textual content.

In the field of textual content recognition, Li [1] proposed a method based on support vector machine in 2013 with 96% recognition rate of official script and regular script; Lin [2] proposed a method based on locality sensitive hashing in 2014 with 81%, 90%, 100%, 81%, 63% recognition rate of seal, official, regular, running and cursive scripts, respectively.

In the field of font recognition, Mao [3] proposed a method based on using Gabor filters as texture discriminator in 2014 with 99%, 98%, 100%, 51%, 71% recognition rate of seal, official, regular, running and cursive scripts; Wang et al. [4] proposed a method based on principal component analysis in 2016 with 99%, 96%, 91%, 73%, 24% recognition rate of seal, official, regular, running and cursive scripts, Yan [5] proposed a method based on local pixel probability pooling in 2018 with 92.4% recognition rate of official, regular and running scripts.

Convolutional neural network (CNN) [6] is a class of deep neural networks [7], which is based on the shared-weight architecture of the convolution kernels that shift over input features and provide translation invariant responses. It is widely used in image recognition systems due to its excellent characteristics.

VGG-net [8] is one kind of CNN invented by a study group of University of Oxford, beat the GoogLeNet [9] and won the localization task in ILSVRC 2014 (ImageNet Large Scale Visual Recognition Challenge 2014)[1]. Network configurations evaluated in [8] are outlined in Figure 1.

VGG-net can provide high performance with simple network structure. Using 3×3 small-sized filter, the computational time can be reduced greatly. Convolutional layers will be repeated for several times before turning to pooling layer, which can help to raise the accuracy of recognition.

Li [10] used CNN to recognize different traditional Chinese calligraphy styles, and achieved the test accuracy of 88.6%. Zou et al. [11] improved the performance of CNN on handwritten Chinese character recognition by combining cross entropy with similarity ranking function and using it as loss function. Wen and Sigüenza [12] proposed a CNN-based method for Chinese calligraphy style recognition based on full-page document.

To date, there is no study which can meet the request to recognize over 90% correct rate in both font recognition and textual content recognition. To overcome this issue, in this paper, we propose a Chinese calligraphy recognition system which recognizes both the font styles and textual contents of Chinese calligraphy artworks.

[1]http://www.image-net.org/challenges/LSVRC/2014/

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

FIGURE 1. Network configuration of VGG-nets, one per column. The depth of the configurations increases from the left (A) to the right (E). As more layers are added, the convolutional layer patrameters are denoted as "conv [receiptive field size]-[number of channels]". For more details, please refer to the original paper [8].

3. **Proposed CNN Model.** Inspired by the above VGG-net, in this paper, we propose the network structure of VGG-net shown in Figure 2. It has 10 weight layers consisting of 5 convolutional layers with 3 pooling layers and 2 fully connected layers.

The first layer is a convolutional layer with $3 * 3$ filter, and uses 64 filters that results in $96 * 96 * 64$ volume. After this, pooling layer is used with max-pool of $3 * 3$ size and stride 3 which reduces height and width of a volume from $96 * 96 * 64$ to $32 * 32 * 64$.

This is followed by 2 more convolution layers with 128 filters. This results in the new dimension of $32 * 32 * 128$. After pooling layer is used, volume is reduced to $16 * 16 * 128$.

Two more convolution layers are added with 256 filters each followed by down sampling layer that reduces the size to $16 * 16 * 256$.

After the final pooling layer, $8 * 8 * 256$ volume is flattened into two fully connected layers with 4096 and 205 channels and resulting output possibilities of 205 labels including 200 character contents and 5 font styles. Two labels with the highest possibility within 200 character content labels and 5 font style labels will be selected respectively as the final recognition result.

Convolutional layers obtain a feature map of the image through multiplication of convolutional kernel with the input. The input image sizes are standardized to $86 * 86$ and all kernel sizes are fixed to $3 * 3$. The function of convolutional layer is as follows. Let w_i and b be the parameters of a convolutional kernel, which are obtained through backpropagation. Then for some input signals x_i, the convolutional layer computes the following

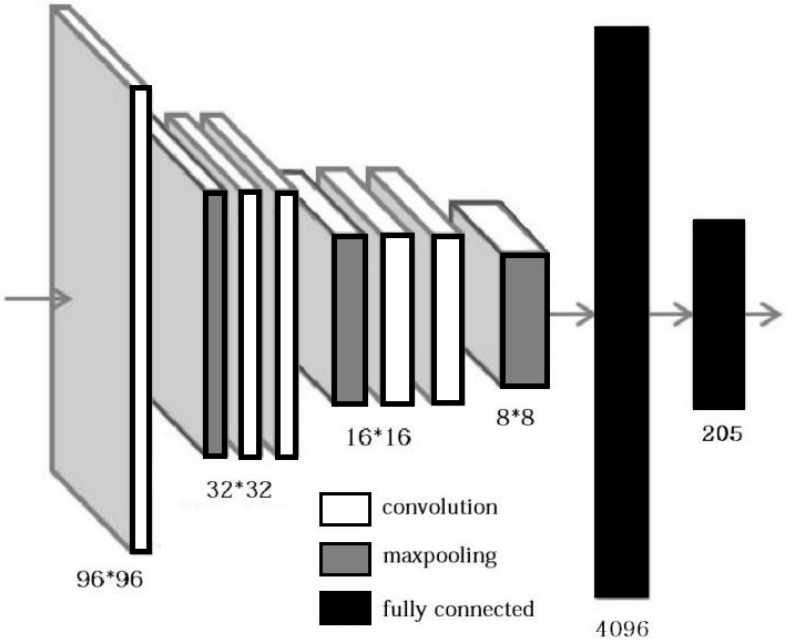
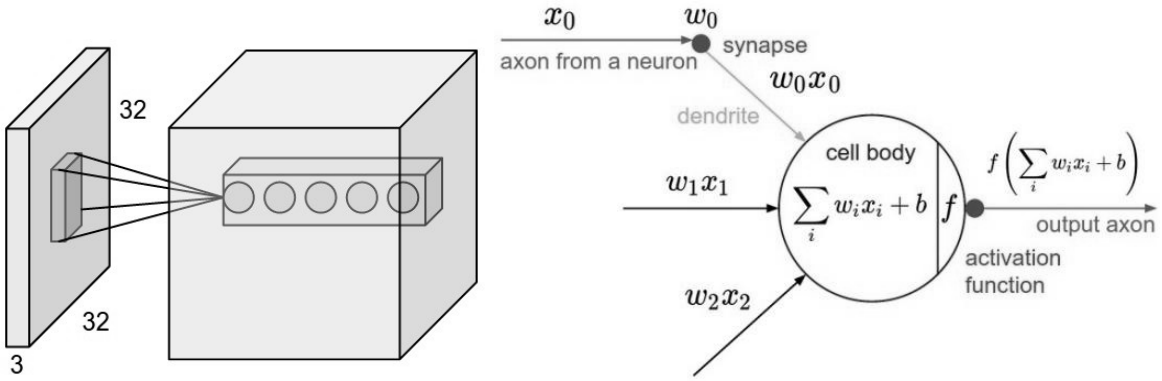


FIGURE 2. Network structure of VGG-net



(a) Process of multiplication (b) Artificial neuron

FIGURE 3. Structure of convolutional layer

function:

$$f(x) = \sum_i w_i x_i + b. \tag{1}$$

Figure 3(a) shows the process of multiplication between the convolutional kernels (the number of channels is 5, which represents there are 5 kernels in all in this layer) and the input image with size $32 * 32 * 3$ (RGB image), where the depth of the output layer is determined by the number of kernels.

Figure 3(b) shows the biological inspiration of CNN. The entrances of the neuron is called dendrites, and the exits are called axons. Each neuron receives electrochemical impulses from other neurons through their axons (outputs) and the dendrites (inputs) of the receptor.

Pooling layers simplify the model through non-linear down-sampling functions. This paper uses max pooling which partitions the input image into a set of rectangles with a stride of 2, and outputs the maximum for each such sub-region. The function of pooling layer is as follows.

$$f_{X,Y}(x) = \max_{a,b\in\{0,1\}} S_{2X+a,2Y+b}, \tag{2}$$

where X and Y denote coordinates of the point in the output image, and $S_{2X+a,2Y+b}$ denotes value of point $(2X+a, 2Y+b)$ in the input image.

Figure 4 shows the process of max pooling with $2*2$ filters and stride 2, where the image size decreases from $224*224*64$ to $112*112*64$.

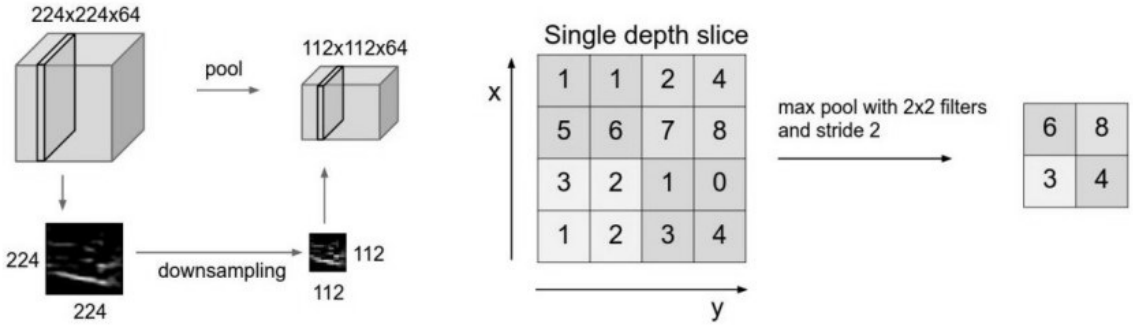


FIGURE 4. Structure of pooling layer

Fully connected layers complete final classification through matrix multiplication followed by a bias offset (vector addition of a learned or fixed bias term). The first fully connected layer has 4096 channels followed by another fully connected layer with 205 channels to predict 205 labels. The function of fully connected layer is as follows:

$$f(x) = xW, \tag{3}$$

where x denotes an input vector, to which a weight matrix W is multiplied from right. Figure 5 shows the structure of a fully connected layer, where a three-dimensional input vector is multiplied by a matrix W_0 from right, that produces a four-dimensional vector.

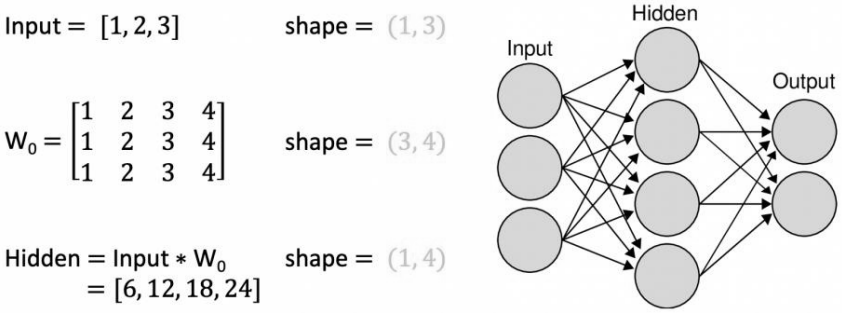


FIGURE 5. Structure of fully connected layer

An image database is constructed in advance for model training. The database is constituted with more than 50,000 pictures of single Chinese characters. There are 200 different characters, with 5 different fonts for each character and 50 images for each font. Each image is given with 2 labels which represent textual content and font of the character. The 80 percent of the database is used as training set and 20 percent of the data set is used as testing set.

We train the model with the database created before, and obtain adjusted parameters which bring better results. The training loss and accuracy are shown in Figure 6, where the vertical axes denote the loss/accuracy defined as

$$Loss = -\ln p_c, \quad Accuracy = \frac{n_{true}}{n_{all}}, \tag{4}$$

where p_c denotes probability of class c, n_{true} denotes number of classes when predict class agrees with the actual class, and n_{all} denotes total number of classes.

The horizontal axes denote the number of epochs. The solid, dash-dotted, dotted, and dashed lines denote the train loss, validation loss, train accuracy and validation accuracy, respectively. The training loss (solid line) smoothly decreases with the progress of the epoch, and the validation loss (dash-dotted line) also decreases with exceptional rises. On the other hand, the training accuracy (dotted line) smoothly increases with the progress of the epoch, and the validation accuracy (dashed line) also increases with exceptional falls.

The confusion matrices of the recognition results for the trained model on training set and testing set are shown in Table 1 and Table 2.

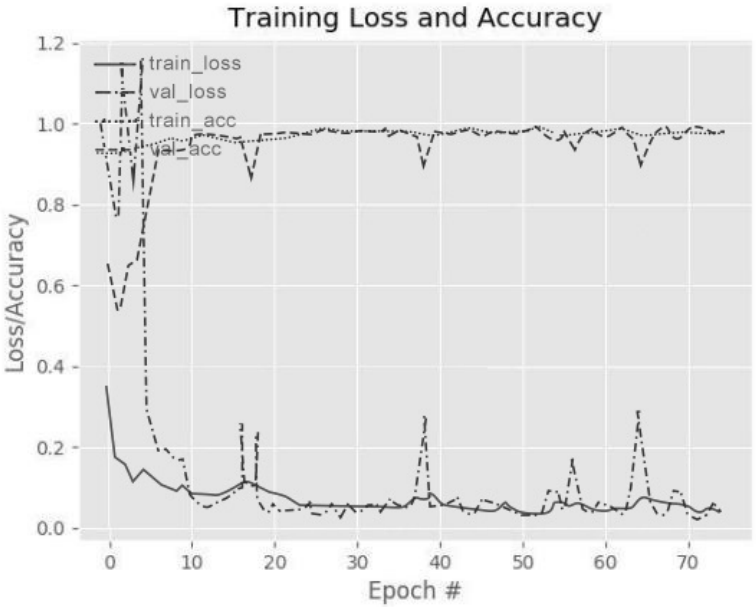


FIGURE 6. Loss & accuracy

TABLE 1. Confusion matrix for the recognition of training set

		Projected					Total
		Seal	Official	Regular	Running	Cursive	
Actual	Seal	**4721**	27	0	11	3	4762
	Official	5	**4145**	3	10	2	4165
	Regular	1	22	**10524**	247	89	10883
	Running	5	31	34	**13328**	507	13905
	Cursive	2	6	8	1070	**6460**	7546
Total		4734	4231	10569	14666	7061	41261

TABLE 2. Confusion matrix for the recognition of testing set

		Projected					Total
		Seal	Official	Regular	Running	Cursive	
Actual	Seal	**1993**	3	1	2	1	2000
	Official	3	**1992**	0	2	1	1998
	Regular	0	1	**1990**	35	30	2056
	Running	1	2	7	**1808**	129	1947
	Cursive	3	2	2	153	**1839**	1999
Total		2000	2000	2000	2000	2000	10000

The precisions of seal, official, regular, running and cursive scripts of training set are 4721/4734 = 99.7%, 4145/4231 = 97.9%, 10524/10569 = 99.5%, 13328/14666 = 90.8%, 6460/7061 = 91.5%, respectively.

The precisions of seal, official, regular, running and cursive scripts of testing set are 1993/2000 = 99.6%, 1992/2000 = 99.6%, 1990/2000 = 99.5%, 1808/2000 = 90.4%, 1839/2000 = 91.9%, respectively.

The above recognition results show that the proposed model meets the request to recognize over 90% correct rate in both font recognition and text content recognition.

We can also see that running and cursive scripts are more difficult to recognize since they are commonly written continuously without pausing and have irregular forms.

4. **System Architecture.** In this section, we describe the detailed architecture of the proposed system. The system is implemented in Python programming language, and divided into four modules: machine learning training, character segmentation, character recognition and user interface modules. System architecture is shown in Figure 7.

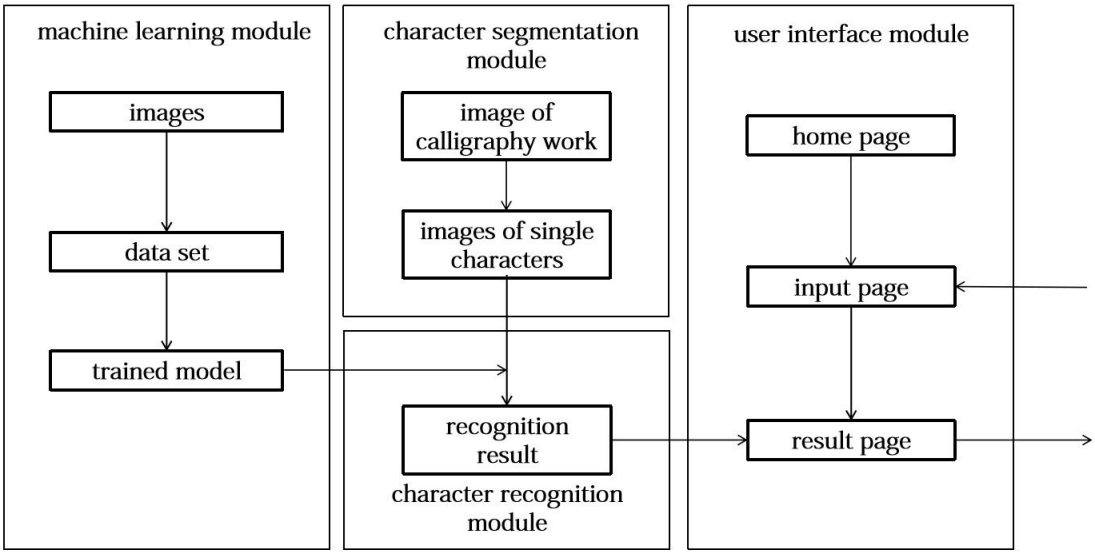


FIGURE 7. System architecture

In machine learning training module, a multi-label classical CNN model is created to recognize single Chinese characters. The CNN model is trained using database, and the trained model is saved for further progress.

We have created a GUI for Chinese calligraphy recognition. The GUI consists of three pages of the system: home page, input page and result page. Home page is displayed when the program starts. User can enter input image into the input page by clicking the button on the home page through text guidance. User can input an image of calligraphy in the input page, and the input image will be passed to character segmentation module.

In character segmentation module, a monochrome is obtained through prepossessing such as contrast enhancement and image binarization. The monochrome is cut by column through projection distribution. Background noise is removed, and the outline of characters is emphasized through processing such as shrinkage processing, median filter, and closing processing. Images of columns are segmented into images of single characters through contour detection function.

Then, images of single characters will be passed to character recognition module and recognized by the trained CNN model.

Recognition result, including font and textual content by column will be returned to user interface model. The system will turn to result page to present recognition result to the user.

5. **Example of System Operations.** Run the system by using images of work 'Zhencao Qianziwen' of Chinese calligrapher Zhao Mengfu as input images. Figure 8(a) shows home page displayed when program starts. User can click the start button to select image file to be input. The system will turn to input page when image file is selected. Figure 8(b) shows input page of the system. The selected image is displayed, and the procedure of recognition is performed through clicking the start button. The system turns to result page when recognition is completed. Figure 8(c) shows result page of the system. The font and textual contents are displayed on the adjacent columns. By clicking home button, the system returns to home page.

(a) Home page (b) Input page (c) Result page

FIGURE 8. Test results

Next, we run the system by using images of work 'The preface of LanTingJi' and 'the monument of XiaoNvCaoE' of famous Chinese calligrapher Wang Xizhi as input images. The input image and result page are shown in Figures 9 and Figure 10. For both images, we obtained correct results of both font recognition and textual content recognition.

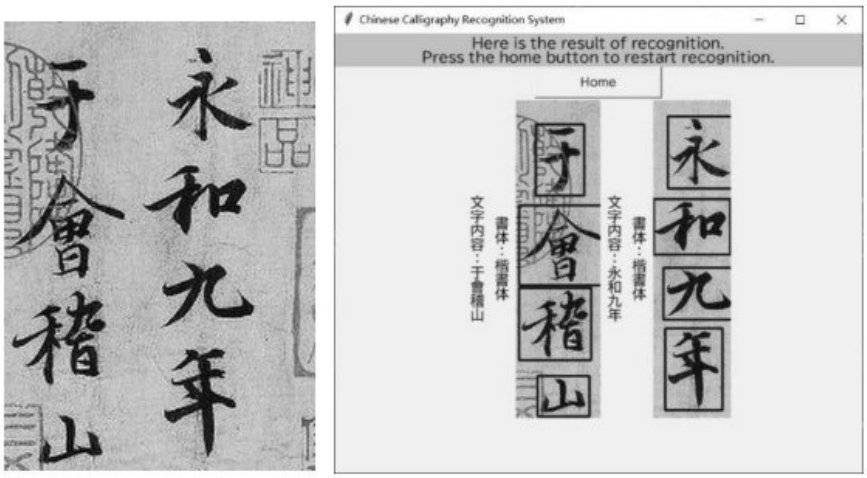


FIGURE 9. Input image and result page of work 'The preface of LanTingJi'

6. **Conclusion.** The application of CNN to Chinese calligraphy recognition is studied in this paper. VGG-net is used to enhance the effectiveness of the system. The system is proved to be capable of recognizing Chinese calligraphy by using images of different calligraphy works. For the future, the system will be improved by adding more functions such as dictionary function by connecting the system to other databases.

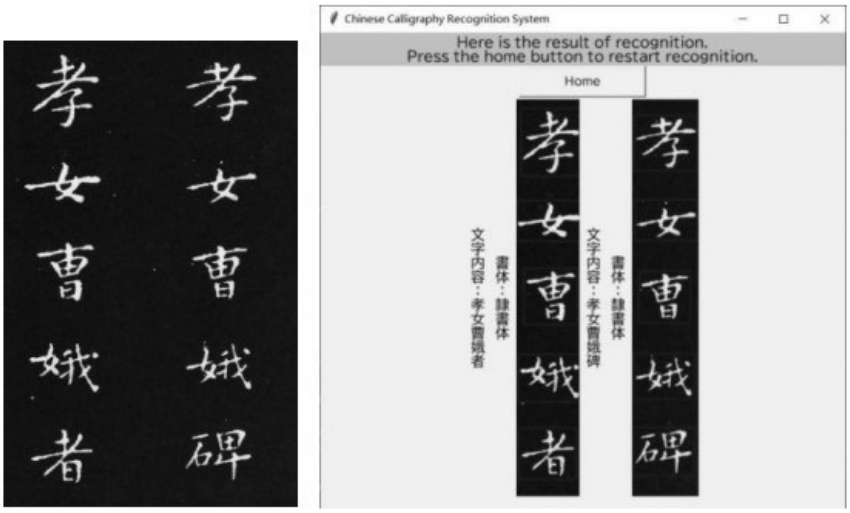


FIGURE 10. Input image and result page of work 'the monument of XiaoNvCaoE'

Acknowledgment. This work was supported by JSPS KAKENHI Grant Number JP21K 11964.

REFERENCES

[1] W. Li, *Research on Key Technologies of Chinese Calligraphy Synthesis and Recognition for Chinese Character of Video*, Ph.D. Thesis, School of Information Science and Engineering, Xiamen University, Xiamen, China, 2013.

[2] Y. Lin, *Research and Application of Chinese Calligraphic Character Recognition*, Ph.D. Thesis, College of Computer Science, Zhejiang University, Hangzhou, China, 2014.

[3] T. J. Mao, *Calligraphy Writing Style Recognition*, Ph.D. Thesis, College of Computer Science, Zhejiang University, Hangzhou, China, 2014.

[4] X. Wang, X. F. Zhang and D. Z. Han, Calligraphy style identification based on visual features, *Modern Computer*, vol.21, pp.39-46, 2016.

[5] Y. F. Yan, *Calligraphy Style Recognition Based on CNN*, Ph.D. Thesis, College of Information and Computer, Taiyuan University of Technology, Taiyuan, China, 2018.

[6] A. Krizhevsky, I. Sutskever and G. E. Hinton, ImageNet classification with deep convolutional neural networks, *Proc. of the 25th International Conference on Neural Information Processing Systems (NIPS'12)*, vol.1, pp.1097-1105, 2012.

[7] Y. LeCun, Y. Bengio and G. Hinton, Deep learning, *Nature*, vol.521, pp.436-444, DOI: 10.1038/nature14539, 2015.

[8] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, *International Conference on Learning Representations*, 2015.

[9] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, Going deeper with convolutions, *arXiv.org*, arXiv: 1409.4842, 2015.

[10] B. Li, *Convolution Neural Network for Traditional Chinese Calligraphy Recognition*, CS231N Final Project, http://cs231n.stanford.edu/reports/2016/pdfs/257_Report.pdf, 2016.

[11] J. Zou, J. Zhang and L. Wang, Handwritten Chinese character recognition by convolutional neural network and similarity ranking, *arXiv.org*, arXiv: 1908.11550, 2019.

[12] Y. Wen and J. S. Sigüenza, Chinese calligraphy: Character style recognition based on full-page document, *Proc. of the 2019 8th International Conference on Computing and Pattern Recognition (ICCPR'19)*, pp.390-394, DOI: 10.1145/3373509.3373512, 2019.

[13] X. Wang, Y. Sheng, H. Deng and Z. Zhao, CharCNN-SVM for Chinese text datasets sentiment classification with data augmentation, *International Journal of Innovative Computing, Information and Control*, vol.15, no.1, pp.227-246, 2019.

[14] M. Nagano and T. Fukami, Development of a skin texture evaluation system using a convolutional neural network, *International Journal of Innovative Computing, Information and Control*, vol.16, no.5, pp.1821-1827, 2020.