Oracle Bone Inscriptions Recognition Based on Deep Convolutional Neural Network

Mengting Liu[1,2], Guoying Liu[1,2,3], Yongge Liu[1,2,3], and Qingju Jiao[2,3]
[1] School of Information Engineering, Zhengzhou University, Zhengzhou, Henan Province, China
[2] Key Laboratory of Oracle Bone Inscriptions Information Processing, Ministry of Education of China, Anyang, Henan Province, China
[3] School of Computer & Information Engineering, Anyang Normal University, Anyang, Henan Province, China
Email: {ai15037339826, ay_liuyongge}@163.com

***Abstract*—In this paper, we describe a new deep learning model for Oracle Bone Inscriptions recognition (OBIs). OBIs are early hieroglyphs in China, which to realize the rapid and accurate image retrieval of large scale oracle bones datasets and break through the limitations of current conventional retrieval methods. With the collected images of the oracle-like characters as input, the model extracts the image features by itself and implicitly learns from the training data to automatically recognize the oracle-like characters. Experimental results show that the proposed method can achieve a high recognition rate. In particular, the accuracy rate of the first 5 prediction categories (top-5) reaches 94.2%, which significantly decreases the search speculation space of archaeological researchers and improves the efficiency and accuracy of OBIs recognition.**

***Index Terms*—convolutional neural network, oracle bone inscriptions, recognition**

I. Introduction

Oracle Bone Inscriptions (OBIs) refer to the ancient Chinese characters used for divination on tortoise shells and animal bones. They are great significance to inherit and carry forward the great traditions of the Chinese nation by making full use of OBIs to study the social history and culture of the Shang dynasty. Using various methods to understand the OBIs is called OBIs interpretation, to solve the contents recorded in OBIs, give play to the role of OBIs in various fields. After the discovery of oracle-bones, many fragments of the oracle-bone were broken up due to their long-term burial in the ground and later excavations. Therefore, the fragments needed to be conjugated to complete the written sentences on the oracle bones. At present, the proposed conjugation conditions include age, handwriting, bone plate, fragment, word, edge six items. There are few oracle-bone slices with their edge characteristics intact, and the efficiency of binding oracle-bone slices according to the edge contour is low. Therefore, it is a feasible strategy to conjugate oracle-bone slices according to the incomplete oracle-bone character recognition on the edge of rubbings. With the rapid development of computer technology, primarily image recognition technology, it is possible to use image recognition [1] technology to assist OBIs experts in the process of oracle bone fragments conjugations, among which the identification of the incomplete OBIs has become the key. Therefore, this paper proposes a strategy for the identification of the incomplete OBIs characters at the edges of the oracle-bone tablets by deep convolutional neural network, in the hope of more accurate and rapid identification of the incomplete OBIs characters or obtaining the maximum similarity of the incomplete OBIs characters, so as to provide data support for the examination of the partial oracle-bone tablets and the interpretation of the oracle-bone characters.

Manuscript received June 10, 2020; revised October 16, 2020.

II. Related Research

As an early hieroglyph in China, OBIs have been recognized by many scholars with different methods. Reference [2] proposed a method to identify the shape of OBIs by using the features of graph theory and strokes. [3] proposed a method to identify the shape of OBIs by using the principle of figure features. [4] proposed to use the method of map isomorphism to identify the shape of OBIs. [5] proposed the support vector machine classification technology to study the recognition of OBIs. Shaotong Gu proposed the recognition method based on topological registration [6] and the fractal geometry method [7] to identify the oracle-bone glyph. At present, the recognition of the incomplete OBIs is still in the exploration stage. In 2012, Hinton *et al.* won the first place in ImageNet image classification challenge by using the famous AlexNet in [8], with an accuracy rate 10% higher than the second place in the traditional method, and deep learning once again attracted full attention. Since then, the deep learning algorithm in [9]-[12] based on the convolutional neural network has continuously made breakthroughs in large-scale competitions in the field of image classification and recognition and even exceeded the recognition accuracy of human beings in some fields. Based on the dominant feature extraction ability of deep learning, this paper proposes a method for recognition of incomplete OBIs based on deep convolutional neural networks. Fine-tuning in [13] was performed for classical deep convolutional neural networks (such as AlexNet, VGG19 in [11],

©2020 Journal of Image and Graphics
doi: 10.18178/joig.8.4.114-119

ResNet in [14], and so on.), and the number of classifications of softmax classifier was adjusted. All training parameters other than the full connection layer of the last layer of the network were retained, thereby achieving rapid training convergence. Experimental results show that this method can achieve a better recognition effect for the incomplete OBIs dataset with a small number of samples.

III. Recognition Framework

A. OBIs Recognition Based on Convolutional Neural Network

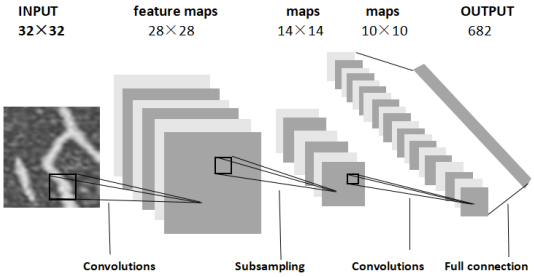

Figure 1. OBIs recognition framework based on convolutional neural network.

Convolutional Neural Network (CNN) is the most widely used in deep learning technology, which has excellent performance in image classification, image retrieval, object detection, and other tasks. Fig. 1 is the OBIs recognition framework based on the convolutional neural network, which is composed of the convolution layer, pooling layer, and full connection layer. In this network, images of broken bones and bones are input, and the convolution layer convolutes the input images to extract the features of the images. A convolution layer is usually composed of multiple convolution kernels. Commonly used convolution kernels have sizes of 3×3, 5×5 and 7×7. Convolution kernels of different sizes can extract different image features. The pooling layer is placed behind the convolution layer; that is, the input of the pooling layer is the output of the convolution layer. Pooling is an operation similar to convolution. All parameters of the pooling layer are super parameters, which can be obtained without learning. Pooling can be divided into two methods: maximum pooling and average pooling. The maximum pooling is to select the maximum value in the pixel matrix of convolution kernel size as the eigenvalue, and the average pooling is to obtain the average value in the pixel matrix of the convolution kernel size as the eigenvalue. In a convolutional neural network, one or more full connection layers are connected after multiple convolutional layers and pooling layers. The full connection layer ACTS as a "classifier" in the entire convolutional neural network. Each neuron is fully connected to all the neurons in the previous layer, and local information with category differentiation in the convolution layer or pooling layer can be integrated.

B. Loss Function

In machine learning, the objective function is the loss function of data samples, and the loss function used in the experiment is the multi-classification cross-entropy loss function shown in (1):

$$loss(x, label) = -(x\log(label) + (1-x)\log(1-label)) \quad (1)$$

In equation (1), the function loss(x, label) is the loss function of a single sample, $x \in R^N$. The data are grouped into batches for training. The output of the neural network is an m×n two-dimensional matrix, where m is the number of batches and n is the number of classifications. The corresponding label is also a two-dimensional matrix, and the result of the loss function is a column vector. The experiment eventually requires a probability value of a different category, and if the output becomes a probability distribution, the softmax layer is needed. Suppose the original output of the neural network is label_1, label_2, label_3..., label_n, the output after softmax regression processing is shown in (2):

$$soft\max(label\_i) = label\_i' = \frac{e^{label\_i}}{\sum_{i=1}^{n} e^{label\_i}} \quad (2)$$

Clearly, $\sum_{i=1}^{n} label\_i' = 1$. The output of a single node becomes a probability value, which is processed by softmax as the output of the neural network. To reduce the error between the model output and the real output, the method of Gradient Descent (GD) in [15] is usually used to iterate the model structure.GD is a search-based optimization method that finds a new iteration point through the gradient direction of the current point, moves from the current point to the new iteration point, and then continues to search for the new iteration point until the optimal solution is found, minimizing the error between the model output and the real output. In this experiment, the convolutional neural network is using Stochastic Gradient Descent (SGD) [16] for model parameter learning, that is, the model weight is updated by randomly selecting one sample at a time, to obtain approximate gradient descent search and improve the iteration speed.

TABLE I. Structure of AlexNet, VGG, Inception-V3, ResNet, and SqueezeNet Model

Model Name	AlexNet	SqueezeNet	VGG19	Inception-V3	ResNet
Year	2012	2016	2014	2015	2015
Image Size	224×224	224×224	224×224	229×229	224×224
Depth	8	10	19	42	151
Filter Size	11,5,3	7,3,1	3	3,1	7,1,3

C. The Best Model Structure

Studies show that with the increase of network depth, the learning ability and pattern expression ability of a convolutional neural network will be strengthened. Since AlexNet directly refreshed the recognition rate of ImageNet in 2012, the network depth of the subsequent VGG, Resnet and other models gradually increased, and the performance was proved to be better and better in ImageNet dataset. In this paper, several groups of comparative experiments are conducted to verify the effectiveness of deep convolutional neural network for the recognition of incomplete OBIs and explore the influence of different depths on the recognition effect. Experiment in AlexNet, VGG, ResNet which are three classical convolution neural network with incomplete OBIs dataset, all will be the last layer network connection layer categories parameters changed to 682, using the incomplete OBIs features of each type of OBIs characters were extracted from training set and its training into new recognition network, finally it has been trained to identify network to incomplete OBIs in the testing set. Since the accuracy difference of top-5 in the three networks is relatively large, which is supposed to be caused by the significant difference in network depth. Therefore, SqueezeNet in [17] and Inception-V3 in [18] are added to the experiment, making the depth difference between the networks relatively small. The model structure of the five convolutional neural networks used in the experiment is shown in Table I.

Table I is the structure of AlexNet, VGG, Inception-V3, ResNet, SqueezeNet model. From the table, we can see the appearance year, input image size, network depth, and convolution kernel size of the five networks. Among them, the input image size of the five networks is 224×224, except for the Inception-V3, whose input image size is 229×229. The depth of the AlexNet network is the smallest, and the depth of SqueezeNet, VGG, Inception-V3, ResNet network gradually increases. Experimental results show that SqueezeNet has the highest accuracy of top-5 on the data set of mutilated tortoise characters, which verifies the conjecture that network depth has a significant influence on the recognition result.

SqueezeNet is an effective and lightweight CNN model, emphasizes design a more efficient way of network computing, reduce network parameters, and no loss of network performance. The core of SqueezeNet is the Fire module shown in Fig. 2, which consists of the Squeeze layer of 1×1 convolution kernel and mixing of two convolution layers of 1×1 and 3×3 convolution kernel. After expanding layer, feature maps obtained from 1×1 and 3×3 are pieced together in the channel dimension. The design concept of SqueezeNet has three points: (1) use the convolution kernel of 1×1 to replace the convolution kernel of 3×3. Under a given amount of convolution kernel budget, most sizes of these convolution cores are set as 1×1. Because the number of convolution kernel parameters of 1×1 is 9 times less than that of 3×3, 1×1 is used to partially replace the 3×3 convolution kernel on the condition that the identification accuracy of the network is not affected. (2) Reduce the number of feature maps of the 3×3 convolution kernel. Since the 1×1 convolution kernel has the function of raising dimension and decreasing dimension, the 1×1 convolution kernel is used for transition, and then it is connected to the 3×3 convolution kernel to achieve this goal. In the Squeeze layer, there are all convolution kernels of 1x1, and the quantity is denoted as S1.In the Expand layer, there are convolution kernels of 1x1 and 3x3, and the quantity is denoted as E1 and E3, respectively. S1 is required to be less than E1+E3, which can reduce the number of feature maps of 3×3 convolution kernel. (3) delay the sampling time of the network so that the convolution layer can obtain more massive feature graphs.

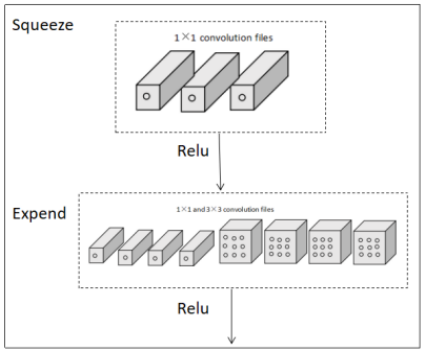

Figure 2. Fire module structure.

IV. Experiment and Result Analysis

A. Experimental Data

The data set of the incomplete oracle bones used in the experiment is the incomplete OBIs appearing on all the rubbings in the ten records, a total of 682 kinds of oracle bone characters, broken bones, word picture as shown in Fig. 3, The first line is four incomplete pictures of the word “Ren” of oracle, the second line is four incomplete pictures of the word “Lei” of oracle, and the third line is four incomplete pictures of the word “Zi” of oracle.

Figure 3. Picture of damaged oracle bones on rubbings.

The dataset consists of training dataset and testing dataset. Rotation and deformation are used for data enhancement to equalize the data samples in each class. The training dataset uses this data enhancement method to make 350 images in each class, and the testing dataset uses this data enhancement method to make 50 images in

each class. The size of the data images in the data set varies, and the size of the original images remains when data enhancement is carried out. Fig. 4 is the effect diagram of the small rotation operation of the data with an unfixed Angle, and the maximum rotation Angle of both left and right is 10. Fig. 5 is the effect diagram of the perspective deformation-oblique quadrangle transformation operation of the data.

Figure 4. Operation effect diagram of small rotation of data with no fixed angle.

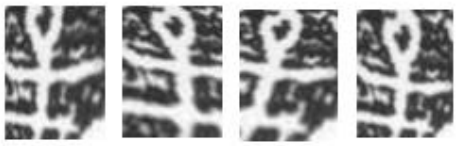

Figure 5. Perspective deformation-oblique quadrangle deformation operation diagram of data.

B. Experimental Environment and Configuration

The computer hardware environment used in this experiment is Intel ® to e5-2683v3 with 16GB memory, GPU is NVIDIA TITAN Xp, and the software environment is python 3.6.8, PyTorch 1.0.1, post2. The training was conducted in Batch mode, with the Batch value of 64. Gradient descent method was used to learn model parameters, and the learning rate was 0.001. The loss function USES the cross-entropy loss function. Due to the different sizes of the images in the data set, when the images are sent to the convolutional neural network, the smaller edges of the rectangular images need to be adjusted to the size required by the network, and then the square images needed by the network need to be cut from the center are sent to the network. The network and training parameters of the training set were saved after 40 epochs to learn iteratively, and then the saved parameters were used for identification on the test set to observe and analyze its top-5 accuracy.

C. Evaluation Indexes

Image recognition is to judge the performance of the recognition algorithm by the recognition accuracy of the image. In this experiment, top-5 accuracy is adopted to measure the recognition accuracy. The calculation formula of top-5 accuracy is shown in (3):

$$P(top-5)=\frac{m}{N} \tag{3}$$

D. Experimental Results and Analysis

According to the experimental setting, the top-5 Accuracy of AlexNet, VGG19, Inception-V3, ResNet, and SqueezeNet is shown in Fig. 6. We can see that the top-5 accuracy curve of the training set of the five networks converged when the epoch was 20, among which the top-5 accuracy of the Inception V3 network remained around 80%, and the top-5 accuracy of the five networks was the lowest. Meanwhile, the accuracy of top-5 of SqueezeNet and AlexNet was relatively high, reaching 90% when the epoch was 10. In the deep convolutional neural network, although increasing the network depth can learn more abstract features, the problem of gradient attenuation should be considered.Inception-V3 has a network depth of 47, but the accuracy of the top-5 test set is 6% lower than that of VGG19 with a depth of 19, indicating that too much network depth causes the problem of gradient disappearance or explosion, which leads to network performance degradation. Moreover, a depth of 152 ResNet recognition effect VGG19 and ResNet is on the basis of VGG network, considering the depth of the network is not easy to train, but the deep web can increase their ability to learn, so as to put forward to keep the complexity of the network layer, increase the short circuit between every two-layer mechanism to form a residual learning to solve the problem.

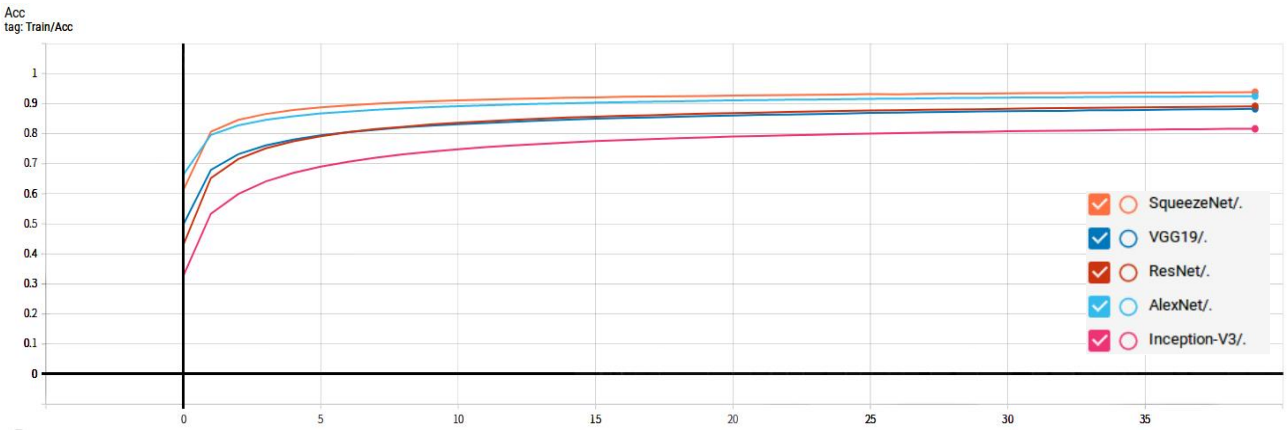

Figure 6. Training top-5 accuracy curve of five networks.

Table II is the top-5 accuracy results of the test set of this experiment on five convolutional neural network frameworks, including AlexNet, VGG19, Inception-V3, ResNet, and SqueezeNet. According to the experimental results, the model has effectively learned the consistency characteristics of each type of oracle letters, and SqueezeNet and AlexNet networks have high accuracy in the test set top-5. According to Table I, the depth of SqueezeNet network is 10, and that of AlexNet network is 8, both of which have relatively small depths and relatively high recognition rate, indicating that to build a network suitable for the characteristics of its data set, the more significant the network depth, the better the recognition effect. The features of the oracle bones data set are relatively simple, and it can be seen from Fig. 1 that there is background noise on the data set, while the convolutional neural network with a greater depth has lower recognition accuracy.

TABLE II. TOP-5 ACCURACY RESULTS OF TEST SETS OF FIVE NETWORKS

Network	Accuracy (%)
AlexNet	93.68
VGG19	89.55
SqueezeNet	94.38
ResNet	89.75
Inception-V3	83.20

V. CONCLUSION

In this paper, a method based on deep convolutional neural network for the recognition of incomplete OBIs is proposed, which can obtain the predicted recognition results with the accuracy of top-5 of 94.38%. Experimental results show that the deep convolutional neural network can effectively assist in the work of thyroid conjugation. This method can extract features from the data set of incomplete OBIs with few samples, to identify the incomplete OBIs on different rubbings.

It can be seen from the analysis of the experimental results that the structure of oracle bones is relatively simple, and it is not necessary to create an intense network structure. The next step is to create a convolutional neural network which is more suitable to extract features of oracle bone character data set based on the structural characteristics of SqueezeNet network based on this experiment. At the same time, considering the severe background noise such as scratches and corrosion on the rubbings, the denoising operation should be added in the recognition to improve the recognition accuracy.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Yongge Liu and Guoying Liu conducted the research; Qingju Jiao analyzed the data; Mengting Liu wrote the paper; all authors had approved the final version.

ACKNOWLEDGMENT

This work is supported by The Oracle Bone Inscriptions Research and Application Special Project of Ministry of Education and National Language Committee of China (Grant number: YWZ-J010, YWZ-J023), the National Natural Science Foundation of China (Grant number: U1804153, 61806007), the National Social Science Fund Major Entrusted Project of China (Grant number: 16@ZH017A3), Program for Changjiang Scholars and Innovative Research Team in University (Grant number: 2017PT35).

REFERENCES

[1] W. Li, L. Duan, D. Xu, and I. W. Tsang, "Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 6, pp. 1134-1148, 2013.

[2] X. Zhou and F. Li, "Study on identification methods of oracle bone inscriptions," *Journal of Fudan University: Natural Science Edition*, vol. 35, no. 5, pp. 481-486, 1996.

[3] F. Li and X. Zhou, "Graph theory method of oracle bone inscriptions automatic recognition," *Journal of Electronics and Information*, pp. 41-47, 1996.

[4] Q. Li, Y. Yang, and A. Wang, "Graph isomorphism method for oracle bone inscriptions recognition," *Computer Engineering and Application*, vol. 47, no. 8, pp. 112-114, 2011.

[5] Y. Liu and G. Liu, "Oracle bone recognition based on SVM," *Journal of Anyang Normal University*, vol. 2, pp. 54-56, 2017.

[6] S. Gu, "Identification method of oracle bone inscriptions based on topological registration," *Computer and Digital Engineering*, vol. 44, no. 10, pp. 2001-2006, 2016.

[7] S. Gu, "Identification method of oracle bone inscriptions based on fractal geometry," *Chinese Journal of Information Technology*, vol. 32, no. 10, pp. 138-142, 2018.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Advances in Neural Information Processing Systems*, 2012, pp. 1097-1105.

[9] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 142-158, 2015.

[10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2015. pp. 1-9.

[11] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Advances in Neural Information Processing Systems*, 2012, pp. 1097-1105.

[13] Z. Zhou, J. Shin, L. Zhang, S. Gurudu, M. Gotway, and J. Liang, "Fine-tuning convolutional neural networks for biomedical image analysis: Actively and incrementally," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7340-7351.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778.

[15] S. Ruder, "An overview of gradient descent optimization algorithms," arXiv preprint arXiv:1609.04747, 2016.

[16] G. B. Orr and K. R. Müller, *Neural Networks: Tricks of the Trade*, Springer, 2003.

[17] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size," arXiv preprint arXiv:1602.07360, 2016.

[18] T. Rui, F. Xiao, J. Tang, F. Zhang, C. Yang, and M. Liu, "Research on multitask deep learning network for semantic segmentation and object detection," in *Proc. Pacific Rim Conference on Multimedia*, 2018, pp. 708-718.

Copyright © 2020 by the authors. This is an open access article distributed under the Creative Commons Attribution License (CC BY-NC-ND 4.0), which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

Yongge Liu was born in September, 1966. He received a master's in computer software and theory of computer science in Northwestern Polytechnical University. He is a master's tutor. He is the director of the Key Laboratory of Oracle information processing, Ministry of Education. He is a professor at school of computer and information engineering, Anyang Normal University.

He has published more than 30 papers in the field of handwritten character recognition, pattern recognition, machine learning, image processing and intelligent system. He visited University of California at Los Angeles in the United States as a visiting scholar in 2012. He has presided over two of the National Natural Science Foundation of China .He obtained one scientific and technological progress award of Henan Province in 2010. He obtained 3 patents of china.

Mengting Liu was born on June 5, 1995. She is a postgraduate student studying at a master's degree in School of Information Engineering, Zhengzhou University. Her research interests mainly focus on machine learning and computer vision, in particular, pattern recognition in digital images. She is currently a member of Key Laboratory of Oracle Bone Inscriptions Information Processing, Ministry of Education of China.

Guoying Liu was born on July, 1979. He received the Ph.D. degrees in geoscience and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, China, in 2009. His research interests mainly focus on machine learning and computer vision, in particular, pattern recognition in digital images. He is currently a member of Key Laboratory of Oracle Bone Inscriptions Information Processing, Ministry of Education of China.

Qingju Jiao was born on November, 1983. He received the Ph.D. degree in Control Science and Engineering from the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, China, in 2015. His research interests mainly focus on pattern recognition, complex network, Oracle Bone Inscriptions information processing. He is currently a member of Key Laboratory of Oracle Bone Inscriptions Information Processing, Ministry of Education of China.