VGG-16 Framework with Dropout for Oracle Bone Inscription Recognition Using Transfer Learning

Zixin Wang *

School of Communication Engineering, Jilin University, Changchun, China, 130025

* Corresponding Author Email: wangzx2022@mails.jlu.edu.cn

Abstract. Oracle Bone Inscription (OBI) is a treasure of Chinese culture, as the research on it progresses, higher requirements are placed on its recognition. Due to the long history of oracle bone inscription and the high cost of labelling, its recognition belongs to the task of Few-Shot learning. This paper proposes a recognition network model based on VGG-16 framework with dropout. Transfer learning strategy using similar feature between oracle bone inscription and modern Chinese character is used to solve the issue of lack of training data. Experimental results on OBI376 dataset shows that the transfer strategy proposed is better than a direct transfer from ImageNet weights and a new training. After the feasibility analysis, different fine-tuning strategies are tried to pursue highest accuracy on test dataset on OBI376. Then two dropout operations are added between the three fully connected layers in the model trained through the fine-tuning strategies with highest accuracy. Comprehensive parameter adjustment for the dropout rate is conducted to obtain the optimal model under this framework. Finally, a recognition model with 98.22% accuracy on test dataset is obtained in this paper, which shows that it has good generalization ability and the transfer strategy proposed is indeed useful for oracle bone inscription recognition.

Keywords: Oracle Bone Inscription, VGG16, Transfer Learning, Dropout Technique.

1. Introduction

Oracle bone inscription is the earliest Chinese character with complete system discovered so far, which has profound historical and cultural significance. In 2007, Oracle bone inscription was selected into the UNESCO's "Memory of the World" register, and its research value was recognized by the world. However, as research progresses, the current Oracle recognition work relies heavily on experts, resulting in low recognition accuracy and efficiency, which hinders further research.

The current Oracle bone inscription recognition technology is still in the initial stage. The mainstream method is to use convolutional neural networks in deep learning to automatically extract topological features and perform self-iterative training. However, not only is it quite difficult to obtain rubbings of oracle bone inscriptions, but also labelling oracle bone inscription image datasets manually is time-consuming, labor-intensive, and costly, ultimately leading to problems such as lack and unbalanced distribution of training samples, scarcity of label corpus, etc.

Above all, oracle Bone Inscription recognition task belong to Few-shot learning task. The traditional deep learning methods with excellent performance in the field of character recognition usually have high degree of dependence on a large scale of samples to train, directly applying these deep network architectures to the oracle bone inscription recognition task may lead to poor generalization of the model.

To enhance model's generalization, this paper utilizes the similar features between oracle bone inscriptions and modern Chinese characters, and proposes a transfer learning strategy from modern Chinese character image dataset (source domain) to oracle bone inscription image (target domain) based on the mature recognition technology of modern Chinese characters. The current method of directly training on the oracle bone inscription image dataset with deep learning networks are abandoned. The major contributions of this paper are:

- Detailed experimental analysis that identifies the effectiveness of the 8 developed models on OBIs recognition task and selecting the most suitable framework for transfer learning.
- A feasibility analysis which proves that the transfer based on modern Chinese character dataset

has better performance than a transfer based on ImageNet weights, and the transfer strategy proposed could indeed enhance accuracy.

• A model after micromesh fine tuning which has excellent generalization on test dataset.

2. Related works

With the continuous integration of digital technology and humanities research, the oracle bone inscription recognition technology, as an interdisciplinary field, has made great progress in recent years. Present oracle bone inscription recognition technology has undergone a development process from machine learning to deep learning. As a representative algorithm in the field of machine learning, Support Vector Machine (SVM) has been widely used in image classification tasks. After extracting the image features of oracle bone inscriptions, some scholars used SVM for classification. Since machine learning methods require human-provided features, the efficiency of machine learning technology in oracle bone inscription recognition is relatively low, and it relies heavily on experts.

Thanks to the advancement of deep learning technology, the current mainstream recognition techniques utilize convolutional layers within deep networks to automatically extract features, significantly lowering the technical barrier and reducing the dependency on experts. Many scholars have already got powerful model for oracle bone inscription recognition using deep neural networks. Zhen et al. [1] uses improved YOLOV8 to detect oracle bone inscriptions on rubbing images, which added a small target detection head, modified the loss function, and embedded a CBAM. Although their model was used to detect instead of classifying, their work shows that deep networks can extract features from OBI. Liu et al. [2] designed a specific convolutional neural network according to characteristics of OBIs, which achieved 84.45% accuracy. Yuan et al. [3] proposed a novel recurrent graph neural network (R-GNN) for the automatic recognition of oracle bone inscription fonts, which used convolutional neural networks (CNNs) and graph neural networks (GNNs) to extract local feature and global contextual information respectively and combined them effectively using recurrent neural networks (RNNs). It finally achieved a Top-1 accuracy of 88.2%. Fujikawa et al. [4] propose a new two stage method using YOLOv4 and MobileNet to recognize OBIs. Gao et al. [5] propose an image translation method based on Generative Adversarial Networks (GANs) from oracle bone inscriptions to modern Chinese characters, which also inspires our work. HAN et al. [6] integrate the idea of self-supervised learning into data augmentation, achieving high performance in recognizing oracle bone inscriptions that are rarely photographed.

However, data dependency is one of the issues faced by deep learning. Firstly, in terms of oracle bone inscription dataset, some oracle bone inscriptions have only one or a few basic samples, and the annotated corpus is scarce and unevenly distributed. At the same time, some rubbings of oracle bone inscriptions are affected by factors such as long history and poor preservation conditions, resulting in varying degrees of damage to the surface, which makes most of the images of oracle bone inscriptions incomplete and noisy. What's more, traditional data enhancement technology such as rotation and mirroring may lead to meaning change of the OBIs. All those problems have led to a small number of oracle bone inscription samples and an insufficient dataset. Deep learning algorithms can only fully exert their performance when the training set samples are large enough, so models above usually fail to achieve ideal generalization performance. In terms of algorithm improvement, given that directly using deep learning methods cannot effectively recognize fonts with small samples, Li et al. [7] proposed a confusion strategy that utilizes a method of mixing majority and minority classes to increase samples and employs a triplet loss function to overcome the issue of imbalanced distribution. However, due to the high complexity of the algorithm, its efficiency and practicality are not entirely satisfactory.

Transfer learning is a subject that focuses on how to build an efficient and reliable model with a relatively small dataset, its core idea is to transfer a trained model to another problem's model in some ways, while retaining the knowledge intersected by the two problems, so as to utilize the knowledge that the model has learned in one task to solve another task. The review paper [8] from Pan

systematically expounds the research process of transfer learning, and puts forward the formal definition and classification of transfer learning. Transfer learning pre-trains a model by utilizing abundant data from related domains or tasks, and then transfers this pre-trained model to a Few-shot target task, thereby fully leveraging existing knowledge and experience to overcome the issue of data scarcity. The previous task before transfer is called source domain while the Few-shot task waited to be solved is called target domain. Transfer learning has gradually developed into an important method to improve the generalization ability of models gradually. Nidhal Azawi [9] combined many deep CNNs with a focus on maximizing the accuracy for each class, rather than just overall accuracy using transfer learning based on ImageNet weights, which made models achieve high accuracy within only a few epochs. Kanishk Bansal and Amar Singh [10] used a VGG-16 model trained on ImageNet dataset to transfer knowledge for geographical landmark recognition. Most researchers used ImageNet weights to do transfer, which is fast and convenient indeed but may has great difference to the target domain. This paper introduces a more powerful transfer from modern Chinese character to OBI using the similar features between two characters sufficiently.

3. Materials

3.1. OBIs dataset

The OBIs dataset used in this paper is from HWOBC dataset provided by 'Yinqi Wenyuan' platform. The "Yinqi Wenyuan" oracle bone inscription big data platform is an oracle bone inscription knowledge-sharing platform that integrates OBI databases, bibliography databases, and literature databases. It was designed, data collected, and collated by the Key Laboratory of Oracle Bone Inscription Information Processing at Anyang Normal University, and is freely accessible to the world, changing the previous difficulties in Oracle Bone Inscription research. 376 kinds of deciphered oracle bone inscriptions have been selected and retained to form the OBI376 dataset, which is split into training set and test set. Train set contains 62589 images belonging to 376 classes, about 166 images per class averagely. Test set contains 3988 images belonging to 376 classes, about 10 images per class averagely.

3.2. Modern Chinese characters dataset

In terms of modern Chinese character image dataset, this paper uses HWDB dataset. The HWDB (Handwritten Chinese Character Database) dataset is a handwritten Chinese character recognition dataset provided by the Institute of Automation, Chinese Academy of Sciences (CASIA). This dataset is widely used in research and development in the field of handwritten Chinese character recognition, covering various data types from single character recognition to text line recognition. This paper adopts the 1.1 version and the offline single character type of data. 1000 kinds of modern Chinese characters have been selected and retained to form the HWDB1000 dataset. All images in this dataset are used to train. There are 239038 images belonging to 1000 classes, about 239 images per class averagely.

4. Methodology

4.1. Gaussian Blur for the noise problem

The oracle bone inscriptions images in the OBI376 dataset may contain considerable noise. How to denoise the images to improve recognition accuracy and efficiency is an issue that requires in-depth research and resolution. In response to this, the Gaussian Blur algorithm is used to denoise the images. Gaussian Blur employs a sliding Gaussian kernel, which assigns a weight to each pixel within the kernel based on the Gaussian distribution, excluding the kernel's central pixel. The RGB values of these weighted pixels are then averaged and assigned to the central pixel of the kernel, ultimately achieving a blurring and smoothing effect. Gaussian filtering is straightforward and easy to implement,

effectively eliminating Gaussian noise in images. Fig. 1. shows the OBI images before and after Gaussian Blur.

(a) OBI images before Gaussian Blur (b) OBI images before Gaussian Blur

Fig. 1. OBI images before and after Gaussian Blur

4.2. Model's selection and fine-tuning

In order to select the framework which is suitable for OBI recognition task, eight famous and powerful models are taken into consideration. Through directly training these models on the OBI376 dataset, the model with highest accuracy on test dataset could be selected. Based on the framework selected, a 1000 classification model is pre-trained on HWDB1000 dataset for 20 epochs. For the pre-trained model, set different proportions of trainable/non-trainable layers, modify the fully connected layers of the model, and use the OBI dataset for secondary training for 10 epochs, which is the transfer learning process. After the secondary training, the transfer proportion with highest accuracy on test dataset could be selected. Under the optimal framework and the optimal proportion, dropout operation is added into the model in order to avoid overfitting. Through adjusting the dropout ratio, an optimal model with highest accuracy can be obtained. All models are set up under TensorFlow frame, loss function uses cross-entropy loss, optimizer uses Adam optimizer and the class mode of dataset label is categorical. The procedural steps for model's selection and fine-tuning are as follows. The illustration of model's fine-tuning is shown in Fig. 2.

Step 1: Select the most suitable framework for OBI recognition task.

Step 2: Select the optimal transfer proportion of trainable/non-trainable layers on pre-trained model.

Step 3: Select the optimal dropout ratio and try to increase training epochs.

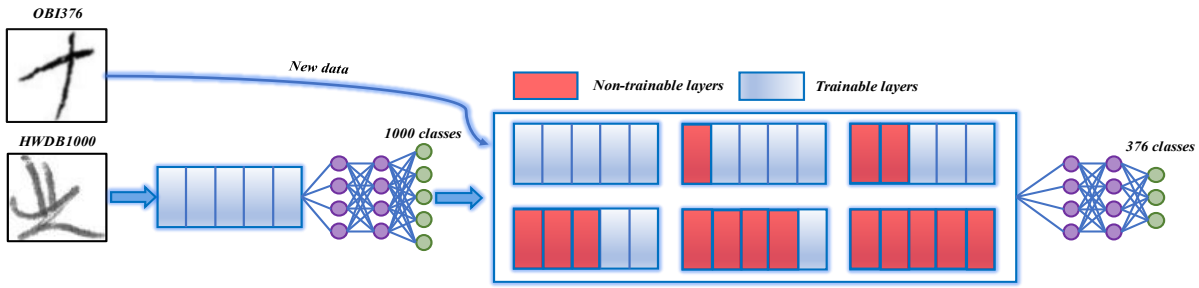


Fig. 2. Illustration of model's fine-tuning during transfer process

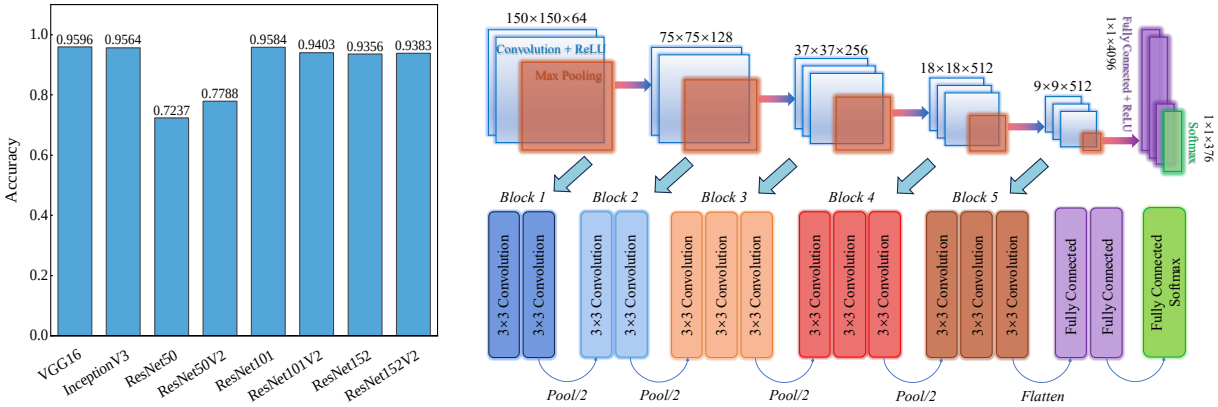
5. Experimental results

5.1. Framework selection

To find the most suitable neural network framework, 8 models are taken into consideration, they are: VGG16, InceptionV3, ResNet50, ResNet50V2, ResNet101, ResNet101V2, ResNet152, ResNet152V2. Each model is trained for 10 epochs using Adam optimizer. Figure (a) in Fig. 3 shows their performance on test dataset which indicates generalization. It is obvious that the VGG16 framework is the most suitable one, it gets 95.96% accuracy which is the highest among all models.

Besides, what's better is that the VGG16 framework is more convenience to fine tuning than other models.

The VGG16 framework was proposed by the Visual Geometry Group from Oxford in 2014, which extensively utilizes small 3x3 convolutional kernels and constructs a very deep network structure. The VGG architecture is straightforward and clear, demonstrating the importance of depth in enhancing network performance. It has shown remarkable generalization capabilities across multiple vision tasks. Its structure is shown in figure (b) in Fig. 3.

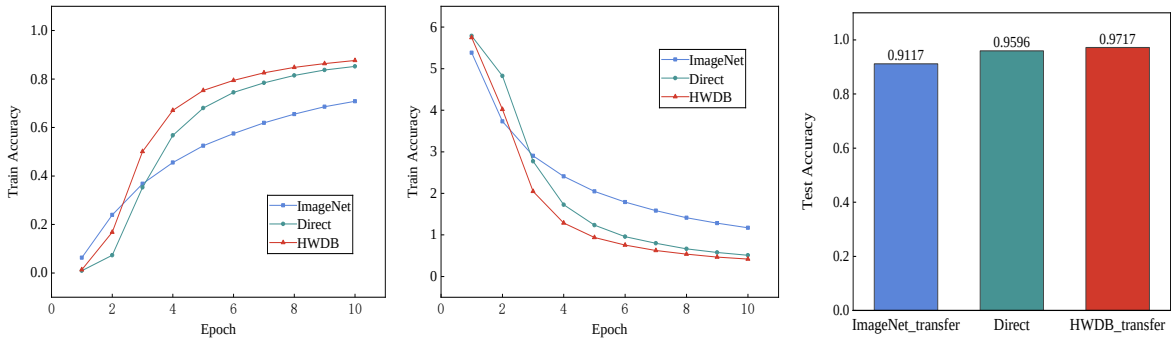


(a) Comparison results (b) Illustration of VGG16

Fig. 3. Comparison among all 8 models and illustration of VGG16

5.2. Feasibility analysis about the transfer

In transfer learning, when two fields are closer, the more knowledge is shared, and the easier the transfer is. Researchers usually only focus on the convenience that transferring directly from ImageNet weights, but may ignore the huge difference between ImageNet weights and the target tasks. Transferring from ImageNet weights could indeed accelerate train process, but may not be beneficial to enhance model's generation ability. Given that, this paper proposes the transfer from HWDB1000 dataset. Feasibility experiment takes three cases into consideration, they are a transfer from ImageNet weights, a transfer from pre-trained model trained on HWDB1000 dataset and a direct learning on OBI376 dataset. As for the transfer process in this part, all convolution layers and polling layers before the fully connected layers are frozen to be non-trainable. Only the fully connected layers could be trained. Through this way, the pre-trained model no longer has ability to extract features from OBI images. The train accuracy and loss are shown in figure (a) and figure (b) in Fig. 4. It is obvious that the transfer from HWDB1000 dataset performs better than the transfer from ImageNet weights. That's because modern Chinese character and OBIs have many similar features. Figure (c) shows the accuracy on test dataset of three cases, the transfer from HWDB1000 dataset has the better generalization ability than both the other two cases, which proves the effectiveness of proposed transfer strategy.



(a) Train accuracy (b) Train loss (c) Test accuracy

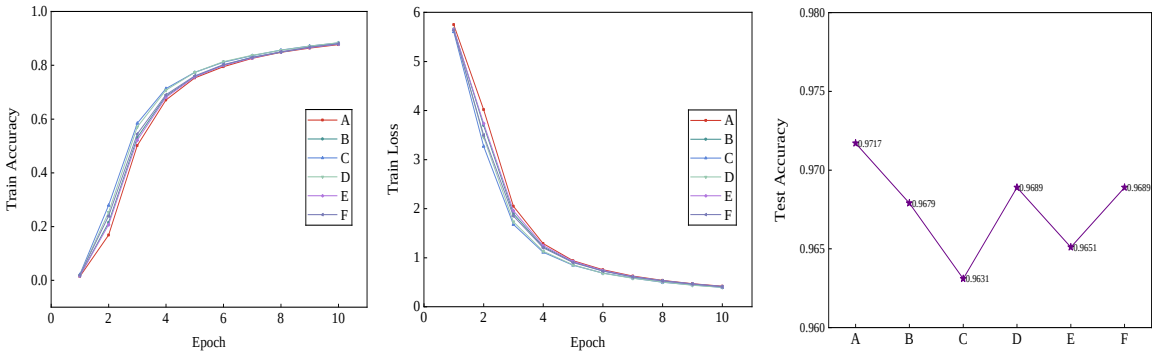
Fig. 4. Comparison between two transfer strategies and direct learning

5.3. Fine-tuning for the pre-trained model

In feasibility analysis experiment above, only fully connected layers are unfrozen to be trainable. In this part, more layers are unfrozen to be trained in order to find the optimal fine-tuning strategies. When more layers are unfrozen, the pre-trained model will extract more features of OBI images during secondary training and get higher training rate. However, more unfrozen layers may also lead to overfitting on OBI376 dataset and decreasing of generalization. So, it is a significant task to find a balance between trainable layers and non-trainable layers. In order to improve efficiency, gradually unfreeze the pre-trained VGG16 model in blocks. For convenience, assign a letter to each fine-tuning case to mark, which is shown in Table 1. Experimental results are shown in Fig. 5. When there are only fully connected layers unfrozen, the pre-trained model learns slower than other cases in previous epochs, but they can get almost the same accuracy and loss on the training set. As for the test set, fine-tuning A with only fully connected layers unfrozen can get the highest accuracy (97.17%). Due to the two domains are too similar and VGG16 framework have sufficient neurons in fully connected layers, only training fully connected layers is enough, extracting features of OBI using convolution blocks may be superfluous.

Table 1. Different transfer strategies

NUM	Trainable layers
A	Fully Connected
B	Fully Connected + Block5
C	Fully Connected + Block5 + Block4
D	Fully Connected + Block5 + Block4 + Block3
E	Fully Connected + Block5 + Block4 + Block3 + Block2
F	Fully Connected + Block5 + Block4 + Block3 + Block2 + Block1



(a) Train accuracy (b) Train loss (c) Test accuracy

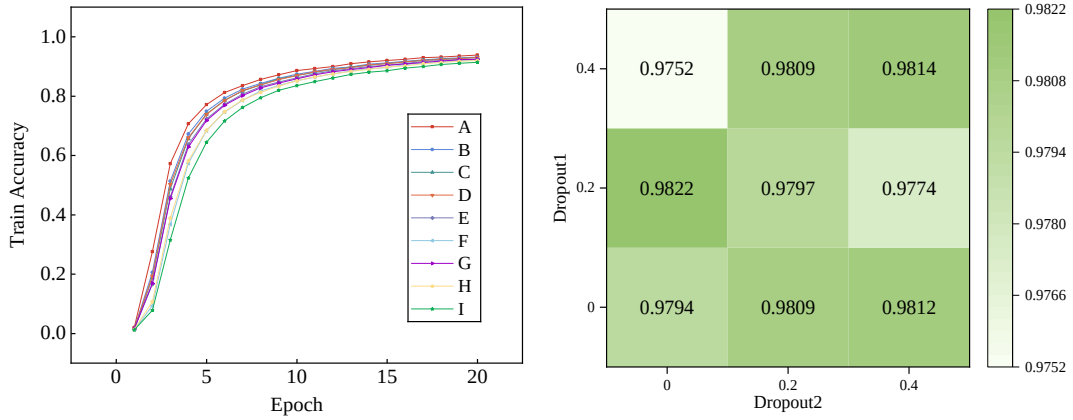
Fig. 5. Comparison between all fine-tuning strategies

5.4. Add dropout technique and adjust the dropout rate

Dropout is a simple yet effective regularization technique that reduces overfitting in neural networks and improves the model's generalization ability by randomly dropping neurons. Additionally, Dropout can reduce the amount of training data required, ultimately enhancing the overall training efficiency. When using dropout technology, the dropout rate is an important hyperparameter that needs to be adjusted according to specific tasks and network structures. Therefore, selecting an appropriate dropout rate is very important in the process of debugging the model. Since dropout is usually used only in fully connected layers, two dropout layers are added between three fully connected layers in this part. The first dropout layer is added after the first fully connected layer and the second is added after the second connected layer. The former is called as *dropout1* and the latter is called as *dropout2*. Table 2 shows different combination of dropout1 rate and dropout2 rate. Fig. 6 shows different results using different dropout rate combination under the above transfer strategy.

Table 2. Different combination of dropout rate

NUM	Dropout1	Dropout2
A	0.0	0.0
B	0.0	0.2
C	0.0	0.4
D	0.2	0.0
E	0.2	0.2
F	0.2	0.4
G	0.4	0.0
H	0.4	0.2
I	0.4	0.4



(a) Train accuracy (b) Test accuracy

Fig. 6. Dropout technique used in transferred model

6. Conclusion and future work

In this paper, transfer learning strategy is introduced to OBI recognition in order to enhance model's training speed and generalization ability. A transfer from HWDB dataset replaces the traditional transfer from ImageNet weights. First, 8 different and powerful models are used to deal with OBI recognition task and VGG16 is the most suitable framework. Second, feasibility analysis proves that the proposed transfer performs better than direct learning and the transfer from ImageNet weights. Then different transfer strategies about the proportion between trainable layers and non-trainable layers are used to second training on the pre-trained model. Through contrast experiment, only making fully connected layers trainable is the optimal fine-tuning strategy which can lead to 97.17% accuracy on test dataset. Finally, dropout technique is added into fully connected layers and a model with 98.22% accuracy is obtained through adjusting dropout rate.

This paper obviously proves transfer learning strategy proposed is useful in OBIs recognition tasks, which reflects in both training speed and generalization ability. In transfer learning, the closer the two domains are, the easier the transfer learning will be. However, the difference in data distribution across domains becomes a major obstacle to model transfer. Domain adaptation is currently one of the frontier research directions in the field of computer vision, aiming to learn a model that enables the knowledge acquired in the auxiliary domain to be generalized well in the target domain. Although modern Chinese character and oracle bone inscriptions are similar enough, domain adaptation algorithms can be added into the transfer process to enhance model's generalization ability further.

Acknowledgements

The authors gratefully acknowledge the financial support from China Undergraduate Innovative Entrepreneurial Training Program funds.

References

[1] Zhen Q, Wu L, Liu G. An Oracle Bone Inscriptions Detection Algorithm Based on Improved YOLOv8[J]. Algorithms, 2024, 17(5): 174-.

[2] Liu M, Gao F, Li B, et al. Application of Neural Network in Oracle Bone Inscriptions [J]. Springer, Cham, 2021.

[3] Jiang Y, Shanxiong C, Bofeng M, et al.R-GNN: recurrent graph neural networks for font classification of oracle bone inscriptions [J]. Heritage Science, 2024, 12(1):

[4] Yoshiyuki F, Hengyi L, Xuebin Y, et al.Recognition of Oracle Bone Inscriptions by using Two Deep Learning Models [J]. International Journal of Digital Humanities, 2022, 5(2-3): 65-79.

[5] Gao F, Zhang J, Liu Y G, et al. Image Translation for Oracle Bone Character Interpretation [J]. Symmetry, 2022, 14: 743.

[6] Han W, Ren X, Lin H, et al. Self-supervised Learning of Orc-Bert Augmentor for Recognizing Few-Shot Oracle Characters [M]. 2021.

[7] Li J, Wang Q F, Zhang R, et al. Mix-Up Augmentation for Oracle Character Recognition with Imbalanced Data Distribution [J]. Springer, Cham, 2021.

[8] Pan S J, Yang Q. A Survey on Transfer Learning [J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345-1359.

[9] Azawi N. Handwritten digits recognition using transfer learning [J]. Computers & Electrical Engineering, 2023, 106: 108604-.

[10] Bansal K, Singh A. Development of VGG-16 transfer learning framework for geographical landmark recognition [J]. Intelligent decision technologies: An international journal, 2023, 17(3): 799-810.