Research Article

An Improved Neural Network Model Based on Inception-v3 for Oracle Bone Inscription Character Recognition

Ziyi Guo [iD],[1] **Zihan Zhou,**[1] **Bingshuai Liu** [iD],[1] **Longquan Li,**[2] **Qingju Jiao,**[3,4]
Chenxi Huang [iD],[1] **and Jianwei Zhang** [iD][5]

[1]*School of Informatics Xiamen University, Xiamen University, Xiamen 361000, Fujian, China*
[2]*School of Mathematical Sciences Xiamen University, Xiamen University, Xiamen 361000, Fujian, China*
[3]*School of Computer & Information Engineering, Anyang Normal University, Anyang 455000, China*
[4]*Key Laboratory of Oracle Information Processing in Henan Province, Anyang 455000, China*
[5]*College of Software Engineering, Zhengzhou University of Light Industry, Zhengzhou 475000, China*

Correspondence should be addressed to Jianwei Zhang; ing@zzuli.edu.cn

Oracle bone inscription is the ancestor of modern Chinese characters. Character recognition is an essential part of the research of oracle bone inscription. In this paper, we propose an improved neural network model based on Inception-v3 for oracle bone inscription character recognition. We replace the original convolution block and add the Contextual Transformer block and the Convolutional Block Attention Module. We conduct character recognition experiments with the improved model on two oracle bone inscription character image datasets, HWOBC and OBC306, and the results indicate that the improved model can still achieve excellent results in the cases of blurred, occluded, and mutilated characters. We also select AlexNet, VGG-19, and Inception-v3 neural network models for the same experiments, and the comparison result shows that the proposed model outperforms other models in three evaluation indicators, namely, Top-1 Accuracy, Top-3 Accuracy, and Top-5 Accuracy, which indicate the correctness and excellence of our proposed model.

1. Introduction

Oracle bone inscription is the earliest documental evidence found in China, engraved on tortoise nails or animal bones. It was used in divination in the late Shang Dynasty and is the earliest script form of Chinese characters [1]. Oracle bone inscription has been added to the Memory of the World Register of the United Nations Educational, Scientific and Cultural Organization (UNESCO). The script provides records of divinations and prayers of ancient people, showing the evolution of Chinese etymology. And it provides invaluable insight into the long-standing civilization and the social structure in early China. The study of oracle bone inscription is important not only for understanding the origin of Chinese characters but also for exploring the ancient history and cultural heritage of China and the world.

Character recognition is an essential part of the research of oracle bone inscription. Its purpose is to determine the categories of oracle bone inscription characters and, on this foundation, to decipher more variants of these characters by drawing on the existing results of decipherment. In addition, character recognition is also an essential task if we want to achieve rapid retrieval of oracle bone inscription characters. However, relying on oracle experts for manual annotation is a very time-consuming and human resource-consuming task, so many researchers are now resorting to neural networks and deep learning to achieve character recognition.

In this paper, we propose an improved neural network model based on Inception-v3. We replace the original convolution block and add the Contextual Transformer block and the Convolutional Block Attention Module. We apply the improved model to two oracle bone inscription

image datasets for character recognition and compare it with Inception-v3 and other classical neural network models such as VGG-19 and AlexNet. The results show that our model achieves the best performance, reaching an outstanding 98.171% Top-1, 99.837% Top-3, and 99.844% Top-5 Accuracy on the clearer dataset, and 87.732% Top-1, 94.847% Top-3, and 96.322% Top-5 Accuracy on the other dataset with noise, which are higher than other models, proving the correctness and excellence of our proposed model.

2. Literature Review

The basic idea of early traditional oracle bone inscription character recognition methods is to preprocess data first, then manually extract features around graph theory and topology, and then encode the features for matching and recognition. Li and Zhou et al. [2] regarded oracle bone inscription characters as an undirected graph composed of lines and points, so they extracted multilevel graph features based on graph theory, then recognized and classified them. Li et al. [3] abstracted oracle bone inscription characters and classified characters based on the graph isomorphism determination algorithm. S. Gu [4] considered that the topology of oracle bone inscription characters is more stable to some extent, and he used the minimum distance to judge the equivalence relationship of the topological structure coding of characters. These methods mainly focused on the font characteristics of oracle bone inscription characters and achieved meaningful results, but simple graph-theoretical characterizations with manual encoding are prone to underfitting in the case of a large number of data.

Many researchers have already implemented oracle bone inscription character recognition by neural networks and deep learning and achieved excellent results. Deep learning-based character recognition is supervised. It requires a large number of training data to enable deep neural networks to learn different patterns of oracle bone inscription characters and thus achieve automatic recognition of single character images. Lv et al. [5] proposed a curvature histogram-based Fourier descriptor to extract glyph features and then input the features to the classical support vector machine (SVM) model [6] for glyph classification. Guo et al. [7] proposed a multilevel oracle bone character representation method, combining sparse self-coding-based middle-level representation features and Gabor-based low-level representation features to describe oracle bone characters. Gao et al. [8] proposed a recognition method based on the Hopfield neural network with an analysis of the context for the recognition of fuzzy characters. Yongge Liu [9] extracted features by chunked histogram and introduced the classical SVM as their model for oracle bone inscription character classification. Meng et al. [10] extracted topological features by Hough transform as well as clustering and achieved recognition by calculating the distance between the actual image and the standard image. Liu et al. [11] created a convolutional neural network based on the classical SqueezeNet for the recognition of incomplete characters at the edges of oracle bones. Sun et al. [12] proposed a dual-view oracle bone character recognition system combining temporal-spatial psychovisual modulation (TPVM) and the character recognition algorithm. Zhang et al. [13] adopted an improved Siamese network to learn the similarity between an oracle bone character and the corresponding template typeset images. Fujikawa et al. [14] proposed a two-stage method that adopts the latest You Only Look Once (YOLO) model and MobileNet for character recognition. These methods introduced neural networks and deep learning, which make the model obtain a better ability of feature representation, so the accuracy of character recognition is improved significantly.

3. Materials

We collect and select two oracle bone character image datasets as our experimental data for training and testing. One is the handwriting oracle bone character recognition dataset (HWOBC), a handwritten character dataset oriented to offline recognition training of handwritten oracle bone characters [15]. The offline recognition of handwritten oracle bone characters is one of the essential steps in the digitization of handwritten oracle bone characters and literature. 22 oracle bone inscription researchers from different disciplines such as script, calligraphy, archaeology, history, and computer compiled this dataset by comparing the handwritten oracle bone character software with the standard oracle bone inscription character forms, so the image quality of this dataset is quite clear. A total of 83245 images are collected in this dataset and divided into 3881 categories according to the oracle bone inscription font code.

The other is a rubbing oracle bone character recognition dataset (OBC306). Huang et al. [16] first collected eight authoritative published works on oracle bone inscription as the material source of the dataset, scanned all pages of the works into digital images, retrieved all positions of characters on the rubbings with the help of dictionary tools, and then finally cut out each character instance manually as a single image of the character. Each character is a separate word class. During the clipping process, a rectangular box was used for frame selection, and it was close to the character instance as much as possible so that the clipped image does not contain too much redundant information. OBC306 contains 309551 oracle bone character images, covering 306 different oracle bone characters. It is the first public dataset with a large number of rubbings, as well as the first public dataset including a variety of published works on oracle bone inscription and a variety of oracle bone characters' special-shaped bodies.

Compared with HWOBC, OBC306 has a larger amount of data with diversity in size and shape. The number of samples is unevenly distributed in each class, showing the long tail effect, and they are all live images, which are blurrier than handwritten images. In addition, there are two other difficulties in this dataset from the perspective of character recognition [16], which are described in detail next.

The first difficulty lies in the existence of heterogeneous characters and the extreme irregularity of oracle bone characters. The different sizes, random directions, and scattered distribution of oracle bone characters increase the

difficulty of recognition. The oracle bone inscriptions are hieroglyphs, focusing more on describing specific features of things, so the relative positions of parts in the characters are not fixed, and the number of strokes fluctuated up and down. There was no unified writing standard at that time, so the writing style evolved.

Noise is another difficulty. Due to years of burial and corrosion, oracle bones are commonly damaged, so the images of dataset OBC306 are also affected by a variety of noise, which divides into the following three cases: first, due to burial and collision with sand and gravel during excavation, many oracle bone handwriting parts on the rubbings are broken, so that the characters on the corresponding images are partially covered by white noise, resulting in blurring of the characters. Second, since the ancient people used tortoise nails or animal bones for divination, they would sizzle them until they cracked, so the cracks would pass through some characters, resulting in the characters on images being covered by white striped areas. Third, many oracle bones are in fragments after excavation, so characters at the edge of the fragments are incomplete, resulting in large white areas that extend from the edge of the image covering the characters. Figure 1 shows some example images that are difficult to recognize. All the above factors indicate that this dataset is a more challenging one.

4. Methodology

In this paper, we propose an improved model based on Inception-v3 for oracle bone inscription character recognition. We replace the original convolution block and add the Contextual Transformer block and the Convolutional Block Attention Module. The main modules of our proposed model are described in detail below.

4.1. Model Structure. We replace the original simple convolution layer of the Inception-v3 model [17] with our new convolution block, then introduce the Contextual Transformer (CoT) block between the convolution and pooling layers to improve the recognition of nearest-neighbor spatial features; finally, we add the Convolutional Block Attention Module (CBAM) to the three main inception modules to enhance the performance of our network. Figure 2 shows the structure of the entire improved neural network model.

4.2. New Convolution Block. The original convolution block performs batch normalization after convolution, and then, the result is passed to the next layer by the activation function. We introduce an Inverted Bottleneck [18], which transforms a single convolution layer into a depthwise convolution layer and two pointwise convolution layers. We also refer to the property of the Multilayer Perceptron (MLP) block in the Transformer [19] that the hidden dimension is four times larger than the input dimension and design the pointwise convolution layer by setting the dimension of the middle hidden layer to four times the size of the input. At the same time, we set the size of the depthwise convolution layer to 7×7 to improve the accuracy. Since our operations

complicate a simple convolution layer into a 3-layer convolution and a summation, we replace the batch normalization with a simpler layer normalization [20] to reduce the complexity. Figure 3 shows the structure of the proposed new convolution block.

In the choice of activation function, we replace rectified linear unit (ReLU) with Gaussian error linear unit (GELU). The activation introduces the idea of random regularity, which is a probabilistic description of the neuronal input and is intuitively more in line with the natural understanding, and the experiment result is exactly better than ReLU. GELU can be computed using the Gauss error function as

$$\text{gelu}(x) = \frac{1}{2}x\left(1 + erf\left(\frac{x}{\sqrt{2}}\right)\right). \qquad (1)$$

where erf*(.)* is the Gauss error function.

4.3. Contextual Transformer Block. We introduce the Contextual Transformer (CoT) block [21], a novel Transformer-style module for visual recognition. There are a large number of heterogeneous characters in the oracle bone inscription, but they usually have the same features in some close positions. Convolution has no way to represent feature interactions at different spatial locations excellently, and a Transformer-style module is needed to improve this deficiency [22]. However, the conventional self-attention block ignores the rich contextual information among the nearest neighbors [23], while the CoT block can encode the context of the input keys by 3×3 convolution, which produces a static contextual representation of the input and better extracts the nearest-neighbor features. We place it after the initial 3-layer convolution block for feature learning of the nearest neighbor space. Figure 4 shows the structure of the Contextual Transformer block.

4.4. Convolutional Block Attention Module. We introduce the Convolutional Block Attention Module (CBAM) [24], which contains two modules of attention mechanism, the Channel Attention Module and the Spatial Attention Module.

The Channel Attention Module mainly processes the feature maps of different channels and tells the model in which maps should be paid more attention to. It first performs global MaxPooling or AveragePooling [25] for the feature maps on different channels and obtains a MaxPool Channel Attention vector and an AvgPool Channel Attention vector. Then input these two vectors into a weight-sharing multilayer perceptron (MLP) [26] with only one hidden layer to get two processed channel attention vectors. Finally, these two vectors are processed by element-wise summation and Sigmoid function, multiplied by the original feature map, and get a new feature map. It can be described as

$$M_c(F) = \sigma(MLP(\text{AvgPool}(F)) + MLP(\text{MaxPool}(F))). \qquad (2)$$

The Spatial Attention Module mainly processes the feature regions on the feature maps and tells the model in

(a)   (b)
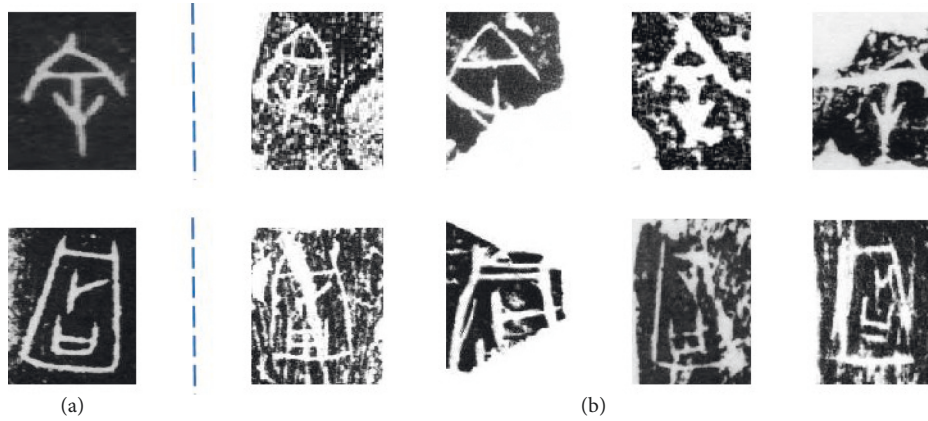
FIGURE 1: Two examples of characters from the OBC306. (a) On the left is a clear image of the character, and (b) on the right are various hard-to-recognize images of the character.
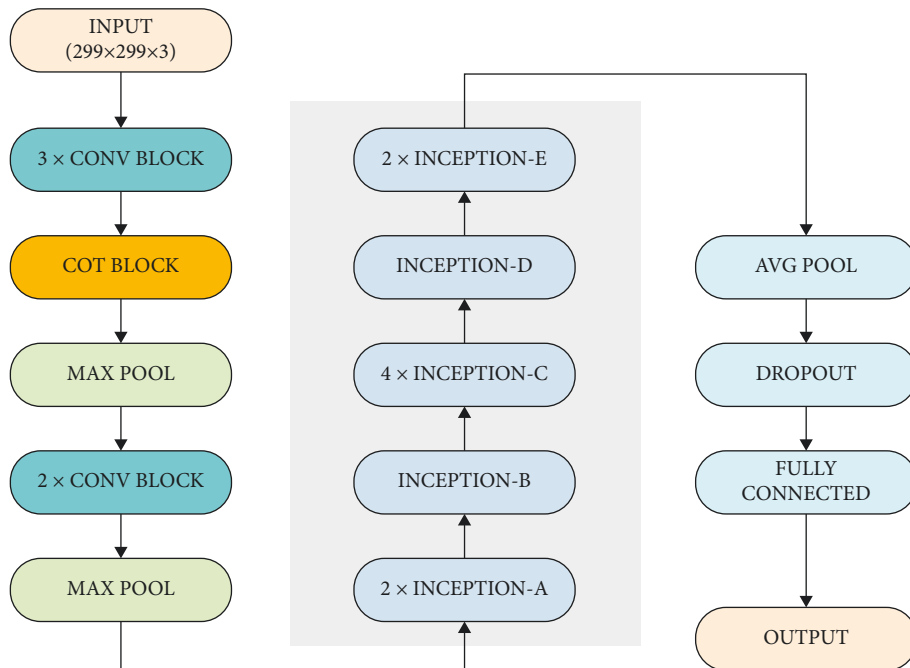

FIGURE 2: The structure of our improved neural network model.

which regions of the feature maps should be paid more attention to. It performs global MaxPooling and AvgPooling on the pixel values at the same locations on different feature maps in the axis direction, obtains two spatial attention maps, and concatenates them [27]. Then, the feature map passes through a 7×7 convolution and a sigmoid activation function to get a spatial attention matrix with the same dimension as the original feature map. Finally, the spatial attention matrix is multiplied by the original feature map and then outputs a new feature map. It can be described as

$$M_s(F) = \sigma\left(f^{7 \times 7}\left([\text{AvgPool}(F); \text{MaxPool}(F)]\right)\right). \quad (3)$$

The CBAM block will derive the attention map by two different dimensions sequentially and then multiply the input feature map by the attention map to achieve adaptive feature optimization. Figure 5 shows the structure of the CBAM. Using the attention mechanism, we can make our model more focused on the essential features and suppress unnecessary features [28]. We add CBAM blocks after the convolution layers in Inception-A, Inception-B, and Inception-D for extracting features in channel dimensions and spatial dimensions to improve the accuracy.

5. Results and Discussion

5.1. Experiments. We divide the two oracle character image datasets into a training set and a test set in the ratio of $7 : 3$, respectively, and then preprocess them after loading. Data loader loads the images and randomly crops them to different sizes and aspect ratios, resizes them to 299×299, and randomly rotates the images horizontally [29]. It enhances the diversity of the dataset, simulates the oracle characters
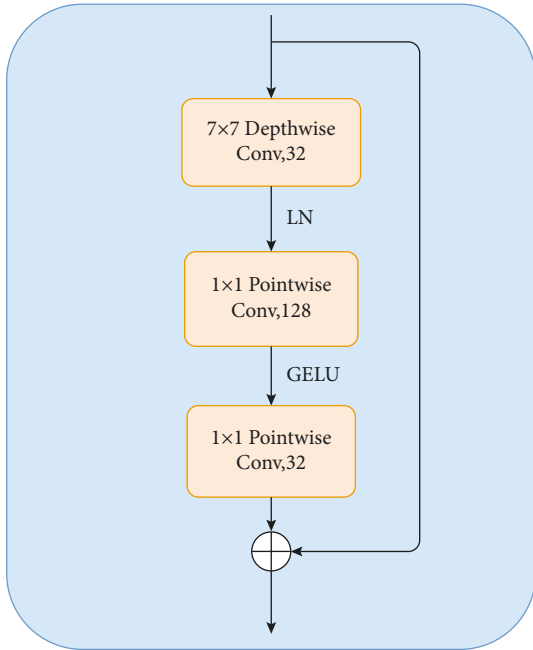
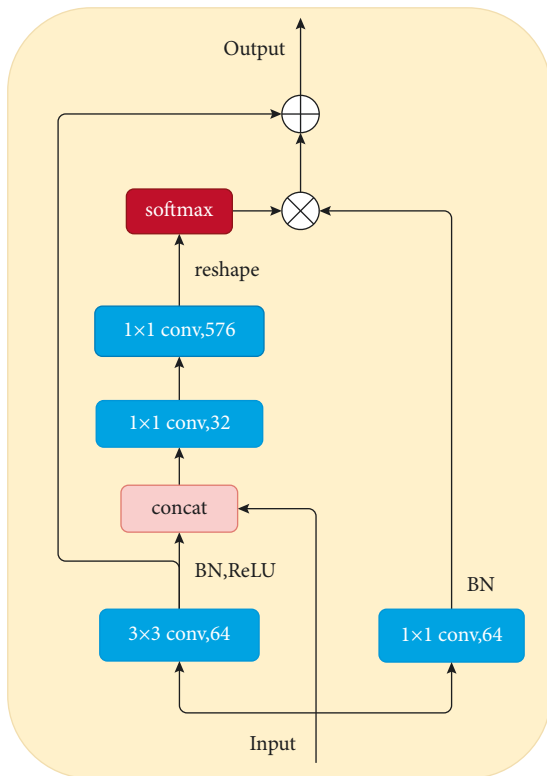FIGURE 3: The structure of the proposed new convolution block.

FIGURE 4: The structure of the Contextual Transformer block.

that appear in different situations, and tests the robustness of the model.

We choose AdamW [30] as the optimizer, which can effectively improve the generalization performance and better avoid the parameter overfitting problem by decoupling the weight decay from the gradient update. The loss function we choose is cross-entropy loss function [31], and the cross-entropy describes the closeness of the actual output to the expected output. The smaller the cross-entropy is, the smaller the closeness is. The formula of the cross-entropy is as follows:

$$
\begin{aligned}
\text{cross\_entropy} = - &\sum_x \left(p(x)\log q(x) \right. \\
&+ \left. (1 - p(x))\log(1 - q(x)) \right).
\end{aligned}
\tag{4}
$$

where p is the expected output and q is the actual output. There are three model accuracy evaluation indicators, Top-1 Accuracy, Top-3 Accuracy, and Top-5 Accuracy. They refer to the probability that the expected result is among the top n in the classification ranking of the actual output, and Top-1 Accuracy is the probability of complete correct identification.

Our experimental environment is PyTorch 1.10.0, Python 3.8, and Cuda 11.3, and the hardware configuration is Intel(R) Xeon(R) Gold 5320 @ 2.20 GHz CPU and 16 GB NVIDIA RTX A4000 GPU. The learning rate is set to decrease as the epoch increases so that the objective function is fast enough to reach the local optimum. We plot the change curves of Top-1 Accuracy and Loss during the experimental process of the OBC306 dataset in Figures 6 and 7. They indicate that Top-1 Accuracy gradually increases and loss gradually decreases, both converge after the 30th epoch, and there is no overfitting, which preliminarily proves the effectiveness of our model. Figure 8 shows the recognition results of some characters in the OBC306, which indicates that our model performs excellently on some character images that are difficult to recognize.

5.2. Comparison. To further demonstrate the excellence of our model, we chose Inception-v3, AlexNet [32], and VGG-19 [33] neural network models to conduct experiments under the same conditions and compared the results.

AlexNet innovatively applies rectified linear unit as the activation function and uses dropout to randomly ignore a portion of neurons during training to avoid overfitting. It also proposes a local response normalization (LRN) layer to create a competition mechanism for the activity of local neurons. It suppresses other neurons with smaller feedback, and the values with larger responses become larger, enhancing the generalization ability of the model, which is suitable for characters recognition.

VGG-19 has thirteen convolutional layers and three fully connected layers, each of which further extracts more complex features, so each layer can be seen as an extractor of multiple local features. It has smaller convolutional kernels (3×3) and smaller pooling kernels (2×2), which enhance the depth of the network and improve the recognition accuracy while ensuring that it has the same receptive field. VGG-19 is widely used in the field of image feature extraction and recognition.

The results on the two datasets are shown in Tables 1 and 2, which indicate that our improved model outperforms other models on both datasets, reaching 98.171% Top-1,
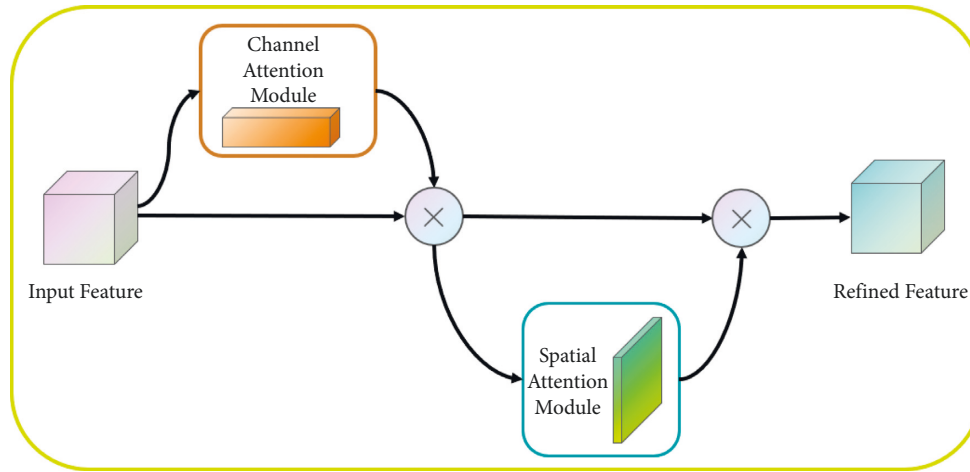
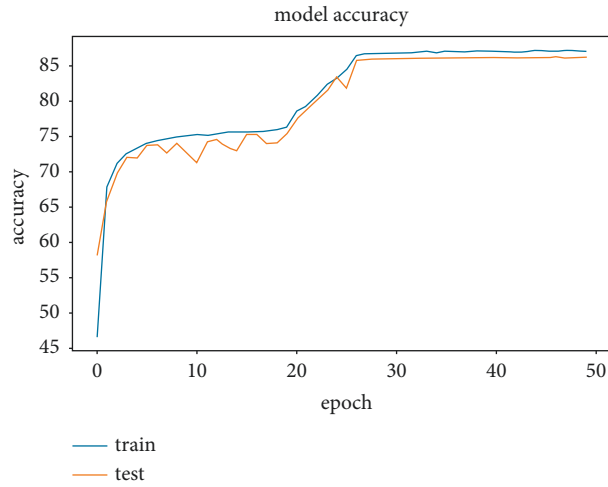FIGURE 5: The structure of the Convolutional Block Attention Module.

FIGURE 6: The Top-1 Accuracy curve of the model. The Top-1 Accuracy gradually increases to 87% and reaches convergence.
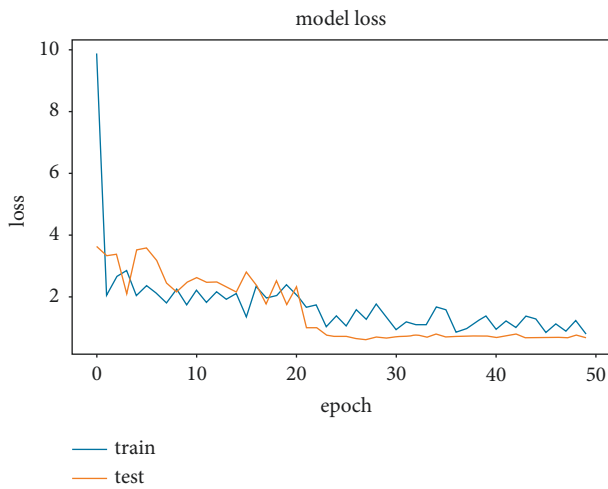

FIGURE 7: The loss curve of the model. The loss gradually decreases to 0.5 and reaches convergence.
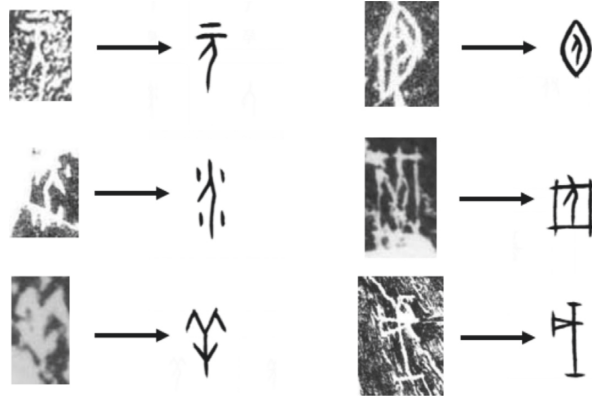
FIGURE 8: Correct recognition results for some challenging character images.

TABLE 1: Performance comparison of models on dataset HWOBC.

Method	Top-1 Accuracy (%)	Top-3 Accuracy (%)	Top-5 Accuracy (%)
AlexNet	97.145	99.707	99.779
VGG-19	97.183	99.646	99.749
Inception-v3	97.909	99.802	99.840
Proposed	**98.171**	**99.837**	**99.844**

The bold indicates the optimal value on this evaluation indicator.

TABLE 2: Performance comparison of models on dataset OBC306.

Method	Top-1 Accuracy (%)	Top-3 Accuracy (%)	Top-5 Accuracy (%)
AlexNet	81.688	91.360	93.856
VGG-19	82.841	92.207	94.425
Inception-v3	85.475	93.234	95.205
Proposed	87.732	94.847	96.322

The bold indicates the optimal value on this evaluation indicator.

99.837% Top-3, and 99.844% Top-5 Accuracy on HWOBC, and 87.732% Top-1, 94.847% Top-3, and 96.322% Top-5 Accuracy on OBC306. It proves the correctness and superiority of our proposed model.

6. Conclusions

In this paper, we propose an improved neural network model based on Inception-v3 for oracle bone inscription character recognition. We replace the original simple convolution layer with our new convolution block. We introduce the Contextual Transformer block between the convolution and pooling layers to improve the recognition ability of nearest-neighbor spatial features. We add the Convolutional Block Attention Module to three main inception modules to enhance the performance of character recognition. We apply the improved model to two oracle bone inscription character image datasets for character recognition and compare it with AlexNet, VGG-19, and Inception-v3 neural network models. The results indicate that our model achieves the best performance in both, reaching 98.171% Top-1, 99.837% Top-3, and 99.844% Top-5 Accuracy on dataset HWOBC, and 87.732% Top-1, 94.847%

Top-3, and 96.322% Top-5 Accuracy on dataset OBC306, which prove the correctness and excellence of our proposed model.

For further work in the future, we plan to improve the model for those characters that rarely appear. Because the number of sample images of these characters is small and their glyphs are usually complex, our model cannot be trained sufficiently to recognize them correctly. We will also try to improve the accuracy of recognition under noise interference as much as possible, such as adopting effective image preprocessing methods to reduce the effect of noise.

Data Availability

The two oracle bone inscription character image datasets used in this paper (HWOBC and OBC306) are both available from the website: https://jgw.aynu.edu.cn/ajaxpage/home2. 0/index.html.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] C. Cheung, "The Chinese history that is written in bone," *American Scientist*, vol. 106, no. 3, pp. 133–135, 2018.

[2] F. Li and X. Zhou, "A graph-theoretic approach to automatic oracle recognition," *Journal of Electronics and Information*, pp. 41–47, 1996.

[3] Q. S. Li, Y. X. Yang, and A. M. Wang, "A graph isomorphism approach for oracle recognition," *Computer Engineering and Applications*, vol. 47, no. 8, pp. 112–114, 2011.

[4] S. T. Gu, "A topological alignment-based method for oracle glyph recognition," *Computers and Digital Engineering*, vol. 44, no. 10, pp. 2001–2006, 2016.

[5] X. Q. Lv, M. N. Li, K. W. Cai, X. Wang, and Y.-M. Tang, "A graph recognition-based classification method for oracle bones," *Journal of Beijing University of Information Science and Technology*, pp. 92–96, 2010.

[6] T. Evgeniou and M. Pontil, "Support vector machines: theory and applications," *Advanced Course on Artificial Intelligence*, pp. 249–257, Springer, Berlin, Heidelberg, 1999.

[7] J. Guo, C. Wang, E. Roman-Rangel, and H. Y. Chao, "Building hierarchical representations for oracle character and sketch recognition," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 104–118, 2016.

[8] F. Gao, Q. Wu, Y. Liu, and X. Jing, "A semantic building block-based method for recognizing fuzzy oracle glyphs," *Science Technology and Engineering*, vol. 30, pp. 67–70, 2014.

[9] Y. G. Liu and G. Y. Liu, "SVM-based oracle bone character recognition," *Journal of Anyang Normal College*, vol. 2, pp. 54–56, 2017.

[10] L. Meng, "Recognition of oracle bone inscriptions by extracting line features on image processing," in *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods*, pp. 606–611, Porto, Portugal, Febraury 2017.

[11] M. Liu, G. Liu, G. Liu, Y. Liu, and Q. Jiao, "Oracle bone inscriptions recognition based on deep convolutional neural network," *Journal of Image and Graphics*, vol. 8, no. 4, pp. 114–119, 2020.

[12] W. Sun, G. Zhai, and Z. Gao, "Dual-view oracle bone script recognition system via temporal-spatial psychovisual modulation," in *Proceedings of the 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pp. 193–198, IEEE, Guangdong, China, August 2020.

[13] C. Zhang, R. Zong, S. Cao, M. Yi, and B. Mo, "AI-powered oracle bone inscriptions recognition and fragments rejoining," in *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pp. 5309–5311, Yokohama, Japan, January 2021.

[14] Y. Fujikawa, H. Li, X. Yue, and C. V. G. A. L. Aravinda, "Recognition of oracle bone inscriptions by using two deep learning models," *International Journal of Dental Hygiene*, 2022.

[15] B. Li, Q. Dai, F. Gao, W. Zhu, Q. Li, and Y. Liu, "HWOBC-a handwriting oracle bone character recognition database," *Journal of Physics: Conference Series*, IOP Publishing, vol. 1651, no. 1, , Article ID 012050, 2020.

[16] S. Huang, H. Wang, Y. Liu, X. Shi, and L. Jin, "Obc306: a large-scale oracle bone character recognition dataset," in *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*, pp. 681–688, IEEE, Australia, September 2019.

[17] S. Christian, V. Vincent, I. Sergey, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, Nevada, ND, USA, June 2016.

[18] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, https://arxiv.org/pdf/2201.03545, 2022.

[19] V. Ashish, N. Shazeer, N. Parmar et al., "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[20] J. Gu, Z. Wang, J. Kuen, and L. A. B. T. X. G. J. T. Ma, "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, 2018.

[21] Y. Li, T. Yao, Y. Pan, and T. Mei, "Contextual transformer networks for visual recognition," 2021, https://arxiv.org/pdf/2107.12292.

[22] C. Djork-Arné, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," 2015, https://arxiv.org/pdf/1511.07289.

[23] E. Hinton Geoffrey, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," 2012, https://arxiv.org/pdf/1207.0580.

[24] S. Woo, J. Park, and J.-Y. I. S. Lee, "Cbam: convolutional block attention module," *Computer Vision - ECCV 2018*, pp. 3–19, 2018.

[25] F. Kunihiko and S. Miyake, "Neocognitron: a self-organizing neural network model for a mechanism of visual pattern recognition," in *Competition and Cooperation in Neural Nets*, pp. 267–285, Springer, Berlin, Heidelberg, 1982.

[26] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, Las Vegas, NV, USA, July 2016.

[28] K. H. Kim, S. Hong, B. Roh, Y. Cheon, and M. Park, "Pvanet: deep but lightweight neural networks for real-time object detection," 2016, https://arxiv.org/abs/1608.08021.

[29] K. Hara, D. Saitoh, and H. Shouno, "Analysis of dropout learning regarded as ensemble learning," in *Proceedings of the ICANN 2016, International Conference on Artificial Neural Networks*, pp. 72–79, Springer, Barcelona, Sepetember 2016.

[30] D. P. Kingma and B. Jimmy, "Adam: a method for stochastic optimization," 2014, https://arxiv.org/abs/1412.6980.

[31] Z. Zhang and M. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[32] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, 2012.

[33] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, https://arxiv.org/abs/1409.1556.