Chinese Journal of Electronics
Vol.32, No.1, Jan. 2023

Ancient Character Recognition: A Novel Image Dataset of Shui Manuscript Characters and Classification Model

TANG Minli[1,2,3], XIE Shaomin[1,3], and LIU Xiangrong[1,3]

(1. *Department of Computer Science and Technology, School of Informatics, Xiamen University, Xiamen 361005, China*)

(2. *School of Big Data Engineering, KaiLi University, KaiLi 556011, China*)

(3. *Key Laboratory of Digital Protection and Intelligent Processing of Intangible Cultural Heritage of Fujian and Taiwan, Ministry of Culture and Tourism, Xiamen University, Xiamen 361005, China*)

Abstract — Shui manuscripts are part of the national intangible cultural heritage of China. Owing to the particularity of text reading, the level of informatization and intelligence in the protection of Shui manuscript culture is not adequate. To address this issue, this study created Shuishu_C, the largest image dataset of Shui manuscript characters that has been reported. Furthermore, after extensive experimental validation, we proposed ShuiNet-A, a lightweight artificial neural network model based on the attention mechanism, which combines channel and spatial dimensions to extract key features and finally recognize Shui manuscript characters. The effectiveness and stability of ShuiNet-A were verified through multiple sets of experiments. Our results showed that, on the Shui manuscript dataset with 113 categories, the accuracy of ShuiNet-A was 99.8%, which is 1.5% higher than those of similar studies. The proposed model could contribute to the classification accuracy and protection of ancient Shui manuscript characters.

Key words — Shui manuscript characters, ShuiNet-A, Artificial neural network, Handwritten character recognition.

I. Introduction

Approximately 500,000 people in China and Vietnam have Shui ethnic group. Shui manuscripts are the ancient texts and books of Shui ethnic group. They contain ancient written symbols similar to oracle bone inscriptions that depict cultural information on ancient astronomy, folklore, ethics, and law of Shui ethnic group. In 2006, Shui manuscripts were included in the first intangible cultural heritage protection list of China. However, Shui manuscripts are known only to hundreds because of the unique way of inheritance. Recently, Shui manuscripts have attracted the attention of the linguistic community owing to their unique charm and high research value. However, although local governments have focused on the inheritance of Shui manuscripts, conservation efforts have been limited. Fig.1 shows the "fish and phoenix" totem bronze coin with Shui manuscript characters that was cast by the ancient Shui people. In the last decades, advancements in the field of computer vision have facilitated the study of ancient writing systems, especially the automatic recognition of characters. At present, the digitization of information has become an essential means of protecting cultural heritage; in particular, artificial intelligence (AI) can lead the way in preserving our distinct cultural heritage for the next generation. The use of deep learning techniques, specifically the convolutional neural networks (CNNs) for the digital transformation of cultural heritage, is gradually gaining popularity.

Fig. 1. The "fish and phoenix" totem bronze coin.

Manuscript Received Apr. 7, 2022; Accepted July 5, 2022.
 DOI:10.23919/cje.2022.00.077

In recent years, deep learning has made great progress in handwritten character recognition tasks, but the studies on recognizing Shui manuscript characters are few. Moreover, the datasets presented in these studies are small, i.e., few samples and categories. Such a dataset cannot meet the data size requirements of modern deep learning. Data, computing power, and algorithms are the three major elements of AI research. Large-scale datasets help to improve the accuracy of algorithms. Approximately 500 Shui manuscript characters have been deciphered by experts until now, but there are no publicly available datasets. Therefore, two questions arise: Can more categories of Shui manuscript characters be classified? Can the classification accuracy of Shui manuscript characters be improved?

Our study aimed to answer these questions. The main contributions of our study are:

1) In response to the lack of a publicly available dataset of Shui manuscript characters, we created Shuishu_C, a novel image dataset of Shui manuscript characters. "Shuishu" is derived from the spelling of the Chinese word "水书," and "C" stands for "character"; this represents that each image contained in the dataset is a Shui manuscript character. This dataset was larger and contained more samples than the existing counterparts.

2) We proposed ShuiNet-A, an attention-based model for the recognition of ancient characters in Shui manuscripts, in which "A" stands for "attention." The model combined the channel and spatial dimensions to extract the key features of target samples. Experiments verified that the obtained classification accuracy was higher than those of similar studies.

II. Related Works

Handwritten character recognition has been researched for many years. Before 2011, the traditional handwritten character recognition method was quite complicated; it required preprocessing, feature extraction, feature dimension reduction, and classifier design. Over the last decade, deep learning methods have been applied to numerous fields, especially convolutional neural network-based methods, which have been very successful in bioinformatics, medical big data, etc. [1]–[4]. Meanwhile, convolutional neural networks (CNN) have been considerably influential in overcoming many challenges related to computer vision and pattern recognition, and has been successfully applied in handwritten character recognition in different languages. The authors in [5]–[9] presented studies on the recognition of Tamil characters using CNN The difference lies in the improvement of the CNN structure. The authors in [10]–[12] presented studies on Arabic character recognition, which mainly use CNNs to combine feature selection methods in machine learning, and obtain satisfactory results for Arabic character recognition. In [13]–[15], the authors combined deep learning methods with traditional methods of feature extraction or some preprocessing methods to improve the recognition accuracy of handwritten Persian characters. In [16], [17], the authors used deep neural networks to recognize handwritten Devanagari characters and have demonstrated that deep neural networks are capable of achieving the highest level of accuracy in the recognition of Devanagari characters. The authors in [18] proposed a novel radical analysis network with densely connected architecture to analyze Chinese character radicals and its two-dimensional structures simultaneously. Evaluated on ICDAR-2013 competition database, the proposed approach significantly outperforms the whole-character modeling approach with a relative character error rate reduction of 18.54%. The authors in [19] proposed an end-to-end neural network model for unconstrained text recognition. The architecture of the model is a fully convolutional network without any recurrent connections trained with the connectionist temporal classification loss function. The model has won the ICFHR2018 Competition on Automated Text Recognition on a READ dataset. The authors in [20] proposed a new architecture of a deep CNN with high recognition performance that is capable of learning deep features for visualization. According to the evaluation on the ICDAR-2013 offline HCCR competition dataset, the model has a relative 0.83% error reduction while having 49% fewer parameters and the same computational cost compared to the current state-of-the-art single-network method trained only on handwritten data. The authors in [21] proposed a four-layer CNN for the recognition of handwritten characters in two Indic scripts, Bangla and Meitei Mayek. In addition, they validated the proposed Manipuri character dataset, called "Mayek27," with the same model. The authors in [22] identified handwritten Farsi digits written with different handwritten styles by using a new combination of CNN layers.

Most of the studies on the recognition of other ethnic minority scripts in China predate the studies on the recognition of Shui manuscripts. For example, relatively well-developed datasets have been established for Tibetan, Mongolian, and Yi scripts. In 2008, the authors in [23] conducted domestic research on the recognition of offline handwritten Tibetan script, and handwritten Tibetan script datasets were established. On December 3, 2018, Jushen AI Technology, a team of college students from Minzu University of China, released the world's first set of Tibetan handwrit-

ten digital dataset TibetanMNIST. In [24], Zhu researched the offline recognition of handwritten Yi scripts in China in 2010 and established an offline handwritten Yi script dataset. In 2014, the authors in [25] researched the recognition of Mongolian scripts in China. In 2018, the authors in [26] proposed MHW, a Mongolian offline handwritten dataset. Compared to the aforementioned research on the script recognition of ethnic minorities, the research on the recognition of Shui manuscripts started late. The authors in [27] proposed a neural network with four convolutional layers, which has a dataset of 50,000 images of Shui manuscript characters. They achieved a classification accuracy of 93.3%. In [28], Xia proposed an 11-layer CNN with a dataset of 60,000 samples from 80 categories for the recognition of Shui manuscript characters and achieved a classification accuracy of 98.3%. The authors in [29] focused on three key technologies for the recognition of Shui manuscript characters in ancient books in China: super-resolution image generation, image category labeling, and handwritten character recognition. They used the feedback from the CNN to determine the algorithmic model of the hyperparameters for cluster labeling and conducted experiments on their own dataset of 6,230 samples. Although the algorithm mentioned in their paper can play a role in the recognition of ancient characters, there are a certain number of errors originating from the automatic labeling. In [30], Ding implemented a feature extraction and classification algorithm based on multilayer perceptron and CNN models in the MATLAB platform for Shui manuscript characters. A total of 1,800 samples from six categories were input into a three-layer neural network, and a classification accuracy of 90.4% was achieved. In addition, 8,500 samples from 17 categories were input into a one-layer CNN, and a classification accuracy of 93.74% was achieved. The authors in [31] proposed a method based on adaptive image enhancement and region detection and segmentation, and they applied it to images of Shui manuscript characters, eventually extracting relatively complete Shui manuscript characters.

The above-mentioned studies report on the progress of text recognition work, and they rely on CNN methods for text recognition. Furthermore, studies on the automatic recognition of Shui manuscript characters by AI techniques are rare. Hence, our work contributes to the intelligent development of Shui manuscript culture.

III. Materials and Methods

1. Data preparation

To the best of our knowledge, there is no publicly available dataset of Shui manuscript characters. In response to this, we developed such a dataset. Fig.2 shows the detailed construction process. First, images of Shui manuscripts were collected and then preprocessed, mainly by alignment, noise removal, and character segmentation. Subsequently, the objects in the images were manually labeled. In the case of imbalanced datasets, data augmentation was performed. After these operations, the dataset of Shui manuscript characters was finally constructed.

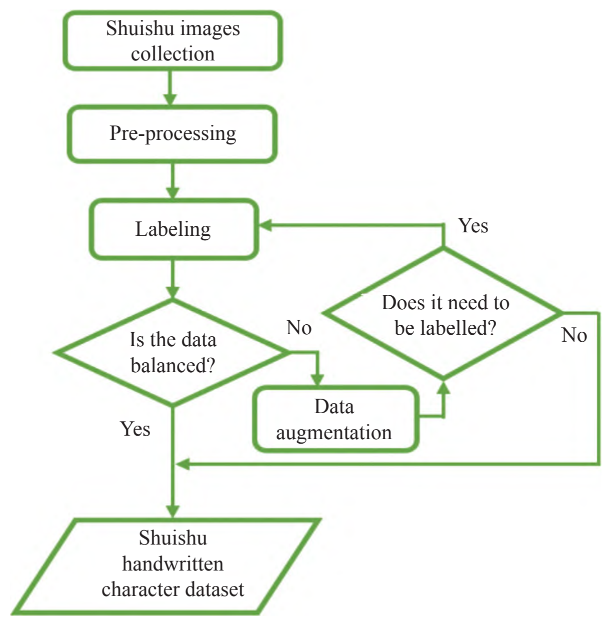


Fig. 2. Method of dataset construction.

Because the ancient books of Shui manuscripts are very precious, most of them are collected and maintained by government agencies. During the image collection, we visited the city that was inhabited by the Shui people; we visited the local library, museum, and Research Institute of Shui manuscripts and gathered information from experts. In the end, we collected 1,789 Shui manuscript images and preprocessed them mainly by filtering, correction, binarization, and noise reduction. Thereafter, we stored them in JPG format, with an average resolution of 1,943×2,924. The Shui manuscript images were manually labeled, and an example of the label dictionary is shown in Table 1.

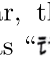
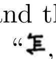
Finally, we obtained 1,789 labeled images that contained 113 categories and 91,336 characters. Statistics on the labeled images of Shui manuscripts showed that the number of samples across different categories varied enormously, as shown in Fig.3. In particular, the category with the smallest number of samples was “  ,” with only one sample, and the category with the largest number of samples was “  ,” with 7,524 samples, there were 33 categories with a sample size of less than 100, accounting for 29%, 32 categories with a sample size of

Table 1. Example of Shui manuscripts characters labeling dictionary

Category	Glyph structure	Meaning	Category	Glyph structure	Meaning
ba		eight	ceng		floor
daos		knife or kill	bix		star
deng		wait	chuang		window
di		land	chunt		spring
dif		place	gsui		cereals
hu		tiger	hua		flower

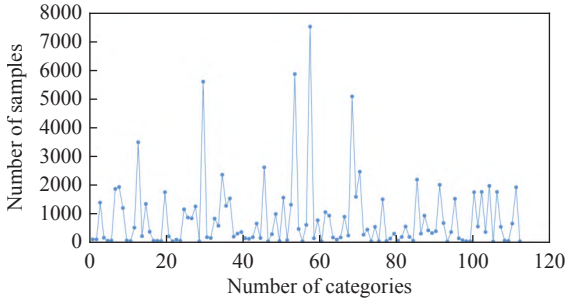


Fig. 3. Statistical analysis of sample size.

100 to 500, accounting for 28%, 17 categories with a sample size of 500 to 1,000, accounting for 15%, and 31 categories with more than 1,000 samples, accounting for 28%.

Every ethnic script contains not only commonly used but also rare characters; thus, imbalanced datasets are a common problem. In the recognition of Shui manuscript characters, the character samples are too few to support model training, and data augmentation is required for categories with few samples. Therefore, we performed data augmentation on the categories with less than 500 samples. With the help of 20 volunteers, 20, 487 character samples were created by handwriting and added in 65 categories. Fig.4 shows the augmented images of Shui manuscripts.

Fig. 4. Shui manuscripts text image added by handwriting.

The key to creating an image of a Shui manuscript character is to crop the target character from the text image. In this study, the character cropping was completed by segmenting the labeled images. Each labeled image had a corresponding file in XML format; the pixels in the corresponding range were extracted, resulting in 111,614 images of the Shui manuscript characters using the lower left and upper right coordinates of each character recorded in the XML file. To deal with complex recognition scenarios and improve the generalization capability of the model, 113 categories of Shui manuscript characters were subjected to data augmentation to increase the diversity of the samples by scaling, rotating, flipping, and adding noise. Ultimately, the original Shui manuscript dataset was increased significantly. To balance the data size, we made additions and deletions so that each category contained 2,000 samples.

Finally, Shuishu_C contained 113 categories with 226,000 labeled images of Shui manuscript characters with an average resolution of 124×117 (Fig.5).

Fig. 5. The sample images of Shui manuscripts characters.

The characteristics of Shuishu_C are as follows:

1) Increased number of categories and dataset of a larger scale. It was found that Shuishu_C was larger and had more categories than similar datasets, as shown in Table 2.

2) Image diversity. The sample images in

Table 2. Comparison with other Shui manuscripts characters recognition results

Dataset	Number of categories	Sample size
Y. Weng and C. Xia [27]	Unknown	60,000
H. Zhao *et al.* [29]	Unknown	6230
Q. Ding [30]	17	8500
Shuishu_C	113	226,000

Shuishu_C contained a variety of scenes. Fig.6 shows the diversity of character "," with samples covering different character orientations and various noises added. The diversity of the data was ensured to improve the robustness of the model and avoid overfitting. Fig.7 shows the visualization results of the data distribution before and after data augmentation for the character "," and "," with the original data on the left and the "original data + augmented data" on the right, i.e., the sum of the samples of that category obtained after data augmentation, showing that data augmentation could compensate for whitespace in sample style transitions.

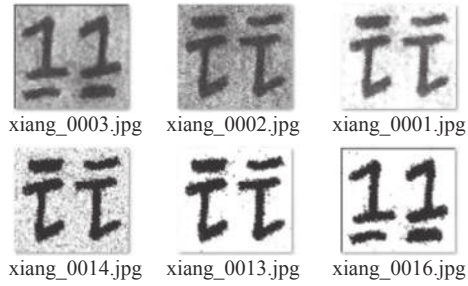


Fig. 6. Examples of image diversity.

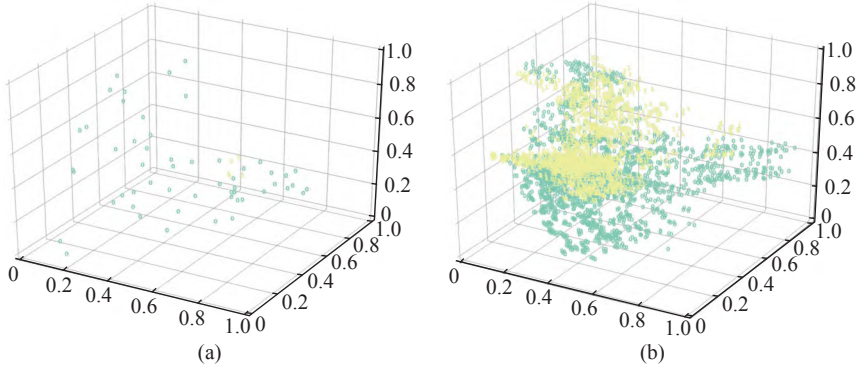


Fig. 7. Visualisation of data distribution before and after augmentation. (a) The visualization results of the original data distribution; (b) The visualization results of the data distribution after data augmentation.

3) Balanced sample size. Shuishu_C contained 113 categories with 2,000 labeled images of characters and a balanced sample size. The increase in the categories and sample sizes of Shui manuscript will be studied in the future.

2. CNN

A CNN is an artificial neural network comprising multiple hidden layers, is extremely powerful in extracting information, and is used in a wide range of applications. The input layer is used to pass the incoming information; the convolutional, pooling, and fully connected layers are collectively called the hidden layers and are responsible for processing and handling the incoming information; the output layer is used to output the result. Fig.8 shows the typical structure of a CNN. The most common CNNs are LeNet [32], AlexNet [33], VGG [34], and ResNet [35].

3. Attention mechanism

The attention mechanism has been proposed by Bahdanau *et al.* [36]. Because of its success in solving

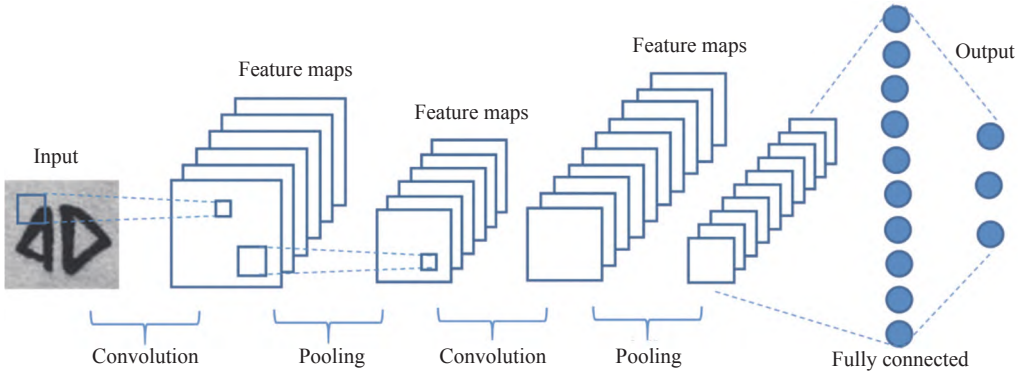


Fig. 8. Typical CNN architecture.

machine translation problems, it has been widely used in various fields, such as natural language processing and computer vision. The core idea of the attention mechanism is to allow the neural network to select some information as input. It pays attention only to the critical information of the input so that the efficiency of the neural network can be improved and more complex models can be fitted. The calculations involved in the attention mechanism are as follows:

$$t_i = \tanh\left(W \times h_{i-1} + b\right) \tag{1}$$

$$\alpha_i = \frac{\exp\left(t_i\right)}{\sum_{i=1}^{T} \exp\left(t_i\right)} \tag{2}$$

$$c = \sum_{i=1}^{T} \alpha_i h_i \tag{3}$$

where W denotes the vector of weights, h_{i-1} the output of the previous layer, b the bias term, T the length of the output of the previous layer, and c the final vector obtained after the weighted combination of h_i and α_i.

In recent years, there have been many improved attention mechanism models for classification tasks, such as convolutional block attention module (CBAM) [37], SE attention [38], SK attention [39], Coordinate attention [40], Triplet attention [41], Swin transformer [42], etc. These are all excellent methods proposed in international top conference on computer vision, hence, we have embedded these attention modules into the model for the classification of Shuishu handwritten characters, and introduce the experimental details later in Section V. After comparison, we found that CBAM is the most suitable for the classification task of Shuishu handwritten characters. As a result, we will introduce CBAM in more detail.

4. Convolutional block attention module

The CBAM was proposed in 2018 by Woo *et al.* as an attention mechanism that combines the channel and spatial dimensions. As shown in Fig.9, the CBAM consists of two independent submodules, the channel attention module (CAM) and the spatial attention module (SAM). CAM and SAM are combined in series to extract the key information on the channel and space. The CAM first performs both global maxpooling and global average pooling on the feature map and then combines the output into a single-pixel feature map of n channels through a two-layer fully connected network; subsequently, it extracts the focused features through the activation function to obtain a $1 \times 1 \times n$ output. Thereafter, SAM takes the output of CAM as its input. After global maxpooling and global average pooling, the features of n channels are obtained as two single-channel outputs of size $w \times h$. Then, the two output features are combined, a convolution operation is performed, and the activation function is used to get the output of size $w \times h \times 1$. Finally, the output of the features from the two modules is multiplied to obtain the final feature.

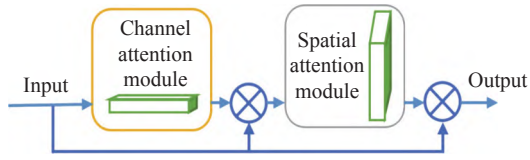


Fig. 9. Structure of CBAM.

CBAM combines a channel attention mechanism with a spatial attention mechanism, which reduces computational overhead compared to other attention mechanisms and allows for better concentration of attention regions around glyphs, resulting in better extraction of features of glyphs and improved recognition accuracy. We therefore incorporate CBAM into the design of the Shui manuscript character recognition model.

5. Evaluation indicators

In this study, some useful statistical metrics were used to evaluate the classification model, including Accuracy, Precision, Recall, and F1-score, which were calculated as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \tag{4}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{6}$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{7}$$

where TP indicates actual positive samples and model predictions are positive; TN indicates actual negative samples and model predictions are negative; FP indicates actual negative samples but model predictions are positive; and FN indicates actual positive samples but model predictions are negative.

IV. ShuiNet-A

1. Structure of ShuiNet-A

ShuiNet-A consists of a convolutional layer, four blocks, and a fully connected layer. Each block is used as a reference for the structure of a residual block, CBAM is added to the input of each block. Each residual block contains three convolutional layers and a pooling layer. Because the text in Shui manuscript is mainly composed of line strokes, only line features need to be extracted. Moreover, the pixel point with the largest value represents the adjacent feature; thus, max-

pooling is used in ShuiNet-A.

In a previous study, we compared networks with depths of 17, 26, 35, 44, and 53 hidden layers and found that the networks with a depth of 17 hidden layers had the best results; hence, such a structure was used for ShuiNet-A. We also found that, among the sigmoid, tanh, ReLU, leaky ReLU, RReLU, PReLU, and ELU functions, the best results were obtained by using leaky ReLU as the activation function. For this reason, we used leaky ReLU and batch normalization (BN) regularization after each convolution in ShuiNet. Using BN ensures that the input value of each layer is the standard normal distribution.

For the consideration of input sizes, we used 32×32, 64×64, 124×117, 144×137, 140×140, and 160×160 as the input sizes for the preliminary experiments, where the sizes 124×117 and 144×137 were the average width and height values for a single category of 1,000 or 2,000 images, respectively.

Table 3 shows the structure details of ShuiNet-A.

Table 3. Structure details of ShuiNet-A

Name	Size-out	Layer
Input	128×128×3	
conv_1	64×64×32	Kernel: 7×7, 32, stride = 1, padding = 3
		MaxPooling 3×3, stride = 2
block_1	32×32×64	CBAM: channel = 32, reduction = 16, 3×3
		Kernel: 1×1, 16
		Kernel: 5×5, 16, padding = 2
		Kernel: 1×1, 64
		MaxPooling 3×3, stride = 2
block_2	16×16×128	CBAM: channel = 64, reduction = 16, 3×3
		Kernel: 1×1, 32
		Kernel: 5×5, 32, padding = 2
		Kernel: 1×1, 128
		MaxPooling 3×3, stride = 2
block_3	8×8×256	CBAM: channel = 128, reduction = 16, 3×3
		Kernel: 1×1, 64
		Kernel: 5×5, 64, padding = 2
		Kernel: 1×1, 256
		MaxPooling 3×3, stride = 2
block_4	4×4×512	CBAM: channel = 256, reduction = 16, 3×3
		Kernel: 1×1, 128
		Kernel: 5×5, 128, padding = 2
		Kernel: 1×1, 512
		MaxPooling 3×3, stride = 2
block_5	2×2×1024	CBAM: channel = 512, reduction = 16, 3×3
		Kernel: 1×1, 256
		Kernel: 5×5, 256, padding = 2
		Kernel: 1×1, 1024
		MaxPooling 3×3, stride = 2
Classification		MaxPooling 2×2, stride = 1
		Fully connected

2. Feature extraction strategy

In this section, the feature extraction strategy will be thoroughly described from the input phase, focusing on the structure of block_1 as an example. The image has a size of 128×128 and three channels, and thus the input size is 128×128×3. First, 32 convolution kernels of size 7×7 were used in the first layer to convolve the input image with a stride of two and a padding of three. Next, a 3×3 maxpooling was performed to obtain an output of size 64×64×32. Five residual blocks, block_1 to block_5, were connected thereupon. Fig.10 shows the overall structure of ShuiNet-A and its residual block. In block_1, the feature map obtained from the previous layer passes through the CBAM and enters the CAM to extract the key feature information in the channel dimension, setting the number of channels to 32 and the reduction to 16. Subsequently, the output feature map enters the SAM, where the size of the convolution kernel is 3×3, to extract the key feature information in the spatial dimension of the feature map. The SAM has a convolution kernel size of 3×3 and extracts key feature information in the spatial dimension. The output is then subjected to a three-layer convolution operation with kernels of 1×1, 5×5, and 1×1 in the three convolutional layers, and it is finally subjected to a maxpooling operation to obtain the input of the next block. The above operation is repeated for block_2 to block_5 with different numbers of convolution kernels, and the output feature maps are 16×16×128, 8×8×256, 4×4×512, and 2×2×1024. After maxpooling with a kernel of 2×2 and a stride of one, a fully connected layer and softmax activation are finally performed to obtain the classification result

V. Method Comparison Experiments

1. Experimental environment

Experiments and evaluations were performed in a deep learning environment using Pytorch as the backend, on a GPU with two NVIDIA GeForce GTX 1080 Ti graphics cards. The experimental codes were all written in Python and the same environment was used for all of the experiments.

2. Experimental settings

In the experimental part, we carried out several sets of comparative experiments, the details are as follows:

1) The first set of experiments, we applied CBAM to different positions for ablation experiments. Used Shuishu_C as input.

2) The second set of experiments, we embeded CBAM, SE attention, SK attention, Coordinate attention, Triplet attention and Swin transformer into the model respectively for classification experiments on the

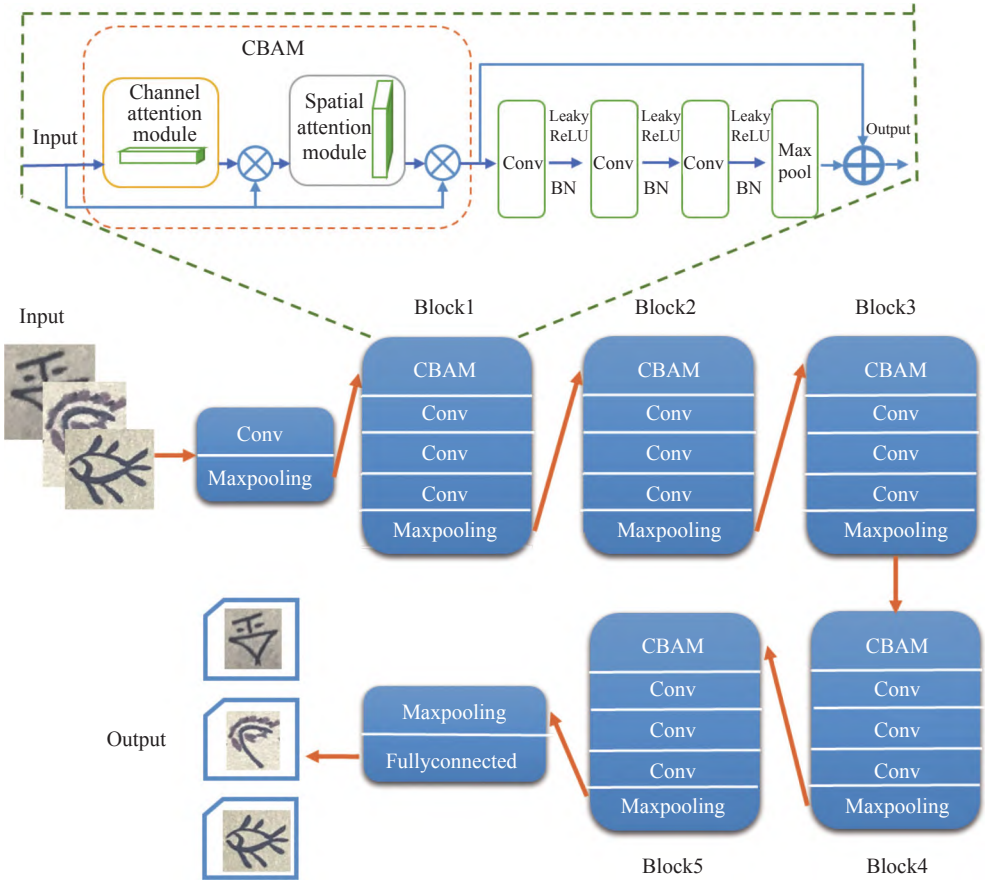


Fig. 10. Overall structure of ShuiNet-A and its residual block.

Shuishu_C dataset.

3) The third set of experiments, which aimed to find the best experimental parameters, adopted ShuiNet-A and Shuishu_C. We set different learning rates and different batch sizes to find the parameter combination with the highest recognition accuracy through comparative analysis.

4) The purpose of the fourth set of experiments was to compare ShuiNet-A with other CNNs, such as LeNet, AlexNet, VGG, and ResNet, all of which use Shuishu_C as input.

We have an ample amount of data, with 2,000 samples for each class of characters, for a total of 226,000 samples. In our preliminary experiments, we found that more than 1,000 samples were sufficient for the shui character recognition task to converge. We wanted the trained model to have better performance on unknown data, hence, in each set of experiments, we divided Shuishu_C into a training set, a validation set and a test set in the ratio of 6:2:2.

3. Results and discussion

In the first set of experiments, we applied CBAM to different positions for ablation experiments, and the results are shown in Table 4. The results show that the highest accuracy can be obtained when CBAM is placed at the position proposed in this paper.

In the second set of experiments, to compare the effects with other attention methods on the model. we

Table 4. Comparison results of CBAM in different locations

CBAM location	Reasoning time (s)	Accuracy (%)	F1-score
In front of convolution 2 in each block	0.009	99.77	0.9977
In front of convolution 3 in each block	0.009	99.57	0.9957
In front of block1	0.006	99.75	0.9975
In front of block2	0.007	99.47	0.9947
In front of block3	0.007	97.66	0.9766
In front of block4	0.007	99.55	0.9955
In front of block5	0.007	99.76	0.9976
Behind block5	0.005	99.76	0.9976
In front of convolution 1 in each block (The location of our model)	0.009	**99.78**	**0.9978**

embedded CBAM, SE attention, SK attention, Coordinate attention, Triplet attention, and Swin transformer into the model respectively, for the Shuishu_C dataset classification experiment. The experimental results are shown in Table 5. It exhibits that the model using CBAM achieves the highest accuracy, demonstrating that CBAM is best suited for the classification task of Shuishu handwritten characters.

Table 5. Comparing the results with other attention methods on Shuishu_C dataset

Method	Reasoning time (s)	Accuracy (%)	F1-score
CBAM	0.009	**99.78**	**0.9978**
SE attention	0.007	99.73	0.9973
SK attention	0.013	99.75	0.9975
Coordinate attention	0.010	99.73	0.9973
Triplet attention	0.012	99.53	0.9953
Swin transformer	0.011	96.80	0.9081

In the third set of experiments, the learning rate was set to 0.01 and 0.001, and the batch size was set to 4, 8, 16, 32, and 64. According to the results shown in Table 6, the highest recognition accuracy was obtained when the learning rate was set to 0.001 and the batch size was set to 8.

Table 6. Comparison results of different parameters of ShuiNet-A

Batch_size	Accuracy	
	Learning rate=0.01	Learning rate=0.001
4	0.9970	0.9976
8	0.9975	**0.9978**
16	0.9974	0.9974
32	0.9977	0.9970
64	0.9974	0.9966

In the fourth set of experiments, Shuishu_C was input into classic CNN models, such as LeNet, AlexNet, VGG, ResNet, and DenseNet, for training and testing. In addition, ShuiNet-A without CBAM is referred to as ShuiNet, and ShuiNet is also included in the comparative experiment. The SGD optimizer was used, the momentum value was set to 0.9, and the cross-entropy loss function was used. According to the experimental results shown in Table 7, all models had good classification results on the dataset, and the classification accuracy of each model was slightly different. The accuracy of the abbreviation ShuiNet was 0.03 percentage points lower than that of ShuiNet-A, indicating that the addition of CBAM helped to improve the performance of the model. The classification accuracy of ShuiNet-A was 99.78%, which was higher than those of LeNet and AlexNet, comparable to that of Vgg16BN, and slightly lower than that of ResNet101. However, ShuiNet-A performed better when factors such as the model size, inference time, and classification accuracy were considered. Our goal is to develop Apps or applets that drive the spread of this culture, therefore the model to be deployed must be lightweight, efficient, and should provide a high level of accuracy. The ShuiNet-A model meets these requirements.

Table 7. Comparing the results with other CNNs on Shuishu_C dataset

Method	Size (M)	Reasoning time (s)	Accuracy (%)	F1-score
ResNet101	163.61	0.072	99.84	0.9984
AlexNet	219.22	0.004	99.72	0.9972
Vgg16BN	514.02	0.006	99.78	0.9978
LeNet	0.279	0.001	95.82	0.9582
ShuiNet (ours)	13.42	0.005	99.75	0.9975
ShuiNet-A (ours)	**13.74**	**0.009**	**99.78**	**0.9978**

VI. Dataset Validation Experiment

1. Experimental environment

Experiments and evaluations were performed in a deep learning environment using Pytorch as the backend, on a GPU with two NVIDIA GeForce GTX 1080 Ti graphics cards. The experimental codes were all written in Python and the same environment was used for all of the experiments.

2. Experimental settings

We conducted several sets of comparative experiments, the details are as follows:

1) The purpose of the first set of experiments was to verify the effect of data augmentation on the recognition performance of the model. Two sets of Shuishu handwritten single-character image datasets were used as the input of ShuiNet-A: the original dataset without data augmentation and Shuishu_C with a balanced sample size.

2) The purpose of the second set of experiments was to verify the stability of ShuiNet-A by adding other types of handwritten character datasets as input. The datasets used in the experiment were Letters, HWDB1, and MNIST. Letters is a dataset of handwritten English letters that contains 18,726 samples of handwritten images of ten English letters from A to J; HWDB1 is a dataset of handwritten Chinese characters that contains 30,074 samples from 101 categories; MNIST is a dataset of handwritten digits that contains 70,000 samples of ten digits, from 0 to 9.

We also divided Shuishu_C into a training set, a validation set and a test set in the ratio of 6:2:2.

3. Results and discussion

In the first set of experiments, we set the learning rate to 0.001, used the stochastic gradient descent

(SGD) optimizer, set the momentum value to 0.9, used the cross-entropy loss function, set the batch size to 128, and the number of training epochs to 120. According to the experimental results shown in Fig.11, the model trained with the original dataset had high training accuracy but low testing accuracy and frequent oscillations, which is an apparent overfitting phenomenon. The model trained with Shuishu_C, which had a balanced sample size, had a good fit and the model could converge quickly.

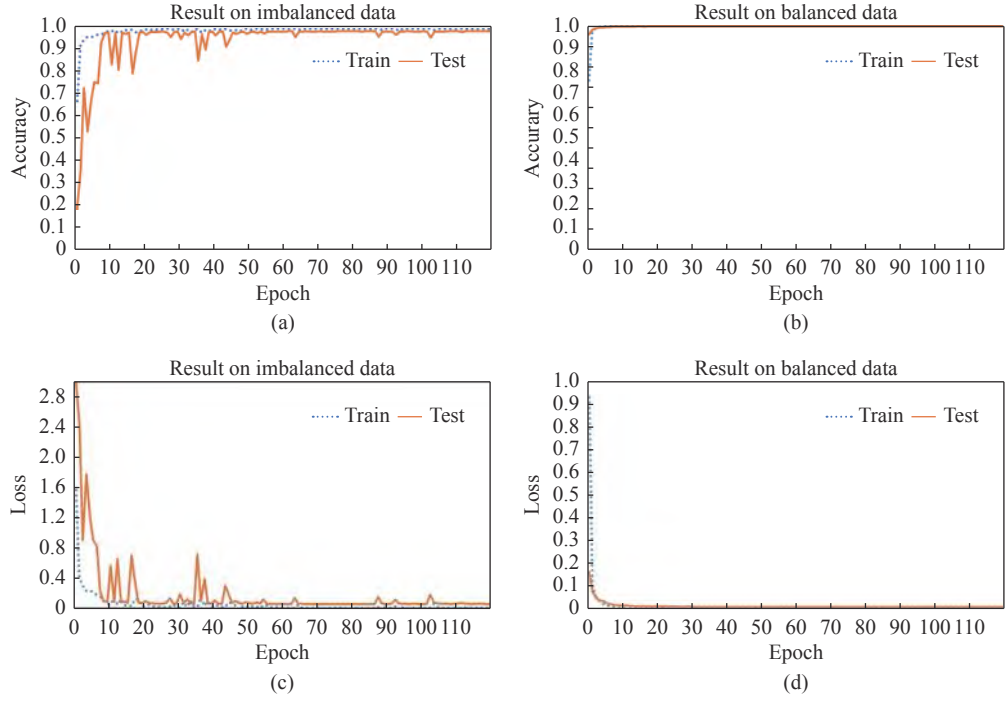


Fig. 11. Comparison of training and testing result. (a) The accuracy of imbalanced data; (b) The accuracy of balanced data; (c) The loss of imbalanced data; (d) The loss of balanced data.

Table 8 shows the comparison of precision, recall, and F1-score obtained in the first set of experiments. According to the experimental results, a dataset with a balanced sample size obtained higher accuracy on ShuiNet-A than on the original dataset, and thus the data augmentation method proved necessary and effective in this study.

Table 8. Comparison results of balanced and unbalanced datasets on ShuiNet-A

Dataset	Precision	Recall	F1-score
Imbalanced dataset	0.9277	0.9049	0.9130
Balanced dataset	**0.9953**	**0.9953**	**0.9953**

In the fourth set of experiments, each dataset of handwritten characters was trained and tested with ShuiNet-A and compared with the results of Shuishu_C. Under the same environment settings, the datasets of handwritten Chinese characters on the one hand, and the digital datasets Letters, HWDB1, and MNIST on the other hand, were input into ShuiNet-A for training and testing. The training loss of each dataset in ShuiNet-A can quickly converge and finally achieve a high classification accuracy. The classification accuracy of the handwritten characters in the water script is still the highest, as shown in Table 9. A possible reason could be that ShuiNet-A was originally built to recognize handwritten characters in Shui manuscripts and the parameter settings were based on the requirements of the handwritten characters dataset in Shui manuscripts; thus, the classification accuracy obtained by Shuishu_C was highest. In addition, the samples in Letters, HWDB1, and MNIST were less than those in Shuishu_C to varying degrees, which also led to a slightly lower classification accuracy than that of the Shui manuscript character dataset.

The comparison of the experimental results of

Table 9. Comparison results of different datasets on ShuiNet-A

Dataset	Size	Precision	Recall	Accuracy (%)
Letters	18,726 samples in 10 classes	0.9457	0.9454	94.5
HWDB1	30,074 samples in 101 classes	0.9609	0.9592	95.9
MNIST	70,000 samples in 10 classes	0.9936	0.9934	99.4
Shuishu_C	226,000 samples in 113 classes	**0.9978**	**0.9978**	**99.8**

ShuiNet-A with those of other Shui manuscript character recognition studies is shown in Table 10. The comparison results show that the accuracy of ShuiNet-A surpasses other similar studies.

Table 10. Comparing ShuiNet-A with the results of known Shui manuscripts characters recognition studies

Method	Dataset size	Accuracy (%)
Y. Weng and C. Xia [27]	50000 samples	93.3
C. L. Xia [28]	60,000 samples in 80 classes	98.3
Q. Ding [30]	8,500 samples in 17 classes	93.7
ShuiNet-A (ours)	226,000 samples in 113 classes	**99.8**

VII. Conclusions

At present, the degree of informatization and intelligence in the inheritance and protection of the Shui manuscript culture is not high, and the achievements are scarce. To address this problem, this study constructed Shuishu_C, a dataset of Shui manuscript characters, introduced the attention mechanism into the Shui manuscript recognition, designed ShuiNet-A, a lightweight network model, and finally verified the effectiveness of the model through experiments. The experimental results revealed that the training and testing of Shuishu_C by using ShuiNet-A achieved a classification accuracy of 99.8% on the Shui manuscript dataset with 113 categories, which is higher than that of other models. We believe that this study is of great significance because it enhances the preservation and transmission of Shui manuscript culture and further promotes the research on Shui Shu. In the future, we will investigate speech recognition and machine translation pertaining to Shui manuscripts.

References

[1] T. Y. Wang, M. Endo, Y. Ohno, *et al.*, "Convolutional neural network-based computer-aided diagnosis in Hiesho (cold sensation)," *Computers in Biology and Medicine*, vol.145, article no.105411, 2022.

[2] S. Zou, C. Li, H. Sun, *et al.*, "TOD-CNN: An effective convolutional neural network for tiny object detection in sperm videos," *Computers in Biology and Medicine*, vol.146, article no.105543, 2022.

[3] P. S. Glaret and P. Muthukannan, "Optimized convolution neural network based multiple eye disease detection," *Computers in Biology and Medicine*, vol.146, article no.105648, 2022.

[4] H. Shin, "Deep convolutional neural network-based signal quality assessment for photoplethysmogram," *Computers in Biology and Medicine*, vol.145, article no.105430, 2022.

[5] M. Sornam and C. V. Priya. "Deep convolutional neural network for handwritten tamil character recognition using principal component analysis," in *Proceedings of International Conference on Next Generation Computing Technologies*, Springer, Singapore, pp.778–787, 2018.

[6] M. A. Pragathi, K. Priyadarshini, S. Saveetha, *et al.*, "Handwritten tamil character recognition using deep learning," in Proceedings of *IEEE International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN)*, Vellore, India, pp.1–5,2019.

[7] N. Ulaganathan, J. Rohith, A. S. Abhinav, *et al.*, "Isolated handwritten Tamil character recognition using convolutional neural networks," in *Proceedings of IEEE 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, Thoothukudi, India, pp.383–390, 2020.

[8] C. Vinotheni, S. L. Pandian, and G. Lakshmi, "Modified convolutional neural network of Tamil character recognition," in *Advances in Distributed Computing and Machine Learning, Lecture Notes in Networks and Systems, vol.127*, Springer, Singapore, pp.469–480, 2021.

[9] R. Jayakanthan, A.H. Kumar, N. Sankarram, *et al.*, "Handwritten Tamil character recognition using ResNet," *International Journal of Research in Engineering, Science and Management*, vol.3, no.3, pp.133–137, 2020.

[10] C. Boufenar, A. Kerboua, and M. Batouche, "Investigation on deep learning for off-line handwritten Arabic character recognition," *Cognitive Systems Research*, vol.50, pp.180–195, 2018.

[11] M. Elkhayati and Y. Elkettani, "Towards directing convolutional neural networks using computational geometry algorithms: Application to handwritten Arabic character recognition," *Advances in Science, Technology and Engineering Systems Journal*, vol.5, no.5, pp.137–147, 2020.

[12] A. Qaroush, A. Awad, M. Modallal, *et al.*, "Segmentation-based, omnifont printed Arabic character recognition without font identification," *Journal of King Saud University-Computer and Information Sciences*, vol.34, no.6, pp.3025–3039, 2022.

[13] F. Sarvaramini, A. Nasrollahzadeh, and M.Soryani, "Persian handwritten character recognition using convolutional neural network," in *Proceedings of IEEE Iranian Conference on Electrical Engineering*, Mashhad, Iran, pp.1676–1680, 2018.

[14] M. Rahmati, M. Fateh, M. Rezvani, *et al.*, "Printed Persian OCR system using deep learning," *IET Image Processing*, vol.14, no.15, pp.3920–3931, 2020.

[15] A. M. M. H and A. Bossaghzadeh, "Improving persian digit recognition by combining deep neural networks and SVM and using PCA," in *Proceedings of IEEE International Conference on Machine Vision and Image Processing (MVIP)*, Tehran, Iran, pp.1–5, 2020.

[16] M. Jangid and S. Srivastava, "Handwritten devanagari character recognition using layer-wise training of deep convolutional neural networks and adaptive gradient methods," *Journal of Imaging*, vol.4, no.2, article no.41, 2018.

[17] R. Guha, N. Das, M. Kundu, *et al.*, "DevNet: An efficient cnn architecture for handwritten devanagari character recognition," *International Journal of Pattern Recognition and Artificial Intelligence*, vol.34, no.12, article no.2052009, 2020.

[18] W. Wang, J. Zhang, J. Du, *et al.*, "DenseRAN for offline handwritten chinese character recognition," in *Proceedings of IEEE 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, Niagara Falls, NY, USA, pp.104–109, 2018.

[19] M. S. Yousef, K. F. Hussain, and U. S. Mohammed, "Accurate, data-efficient, unconstrained text recognition with con-

volutional neural networks," *Pattern Recognition*, vol.108, article no.107482, 2020.
[20] P. Melnyk, Z. You, and K. Li, "A high-performance CNN method for offline handwritten Chinese character recognition and visualization," *Soft Computing*, vol.24, no.11, pp.7977–7987, 2020.
[21] A. Hazra, P. Choudhary, S. Inunganbi, *et al.*, "Bangla-Meitei Mayek scripts handwritten character recognition using convolutional neural network," *Applied Intelligence*, vol.51, no.4, pp.2291–2311, 2021.
[22] Y. A. Nanehkaran, D. Zhang, S. Salimi, *et al.*, "Analysis and comparison of machine learning classifiers and deep neural networks techniques for recognition of Farsi handwritten digits," *The Journal of Supercomputing*, vol.77, no.4, pp.3193–3222, 2021.
[23] D. X. Zhao and C. X. Zhao, "Feature extraction of offline handwritten Tibetan script," *Gansu Science and Technology*, vol.24, no.5, pp.48–49, 2008. (in Chinese)
[24] L. H. Zhu, "Offline handwritten Yi character recognition based on combination features and multi-classifier integration, " *Journal of Yunnan Minzu University : Natural Science Edition*, vol.19, no.5, pp.329–333, 2010. (in Chinese)
[25] H. Chun, "Research of printed Mongolian character recognition," *Journal of Inner Mongolia Minzu University (Natural Sciences)*, vol.29, no.06, pp.627–628, 2014. (in Chinese)
[26] D. E. J. Fan, G. L. Gao, and H. J. Wu, "MHW Mongolian offline handwriting database and its application," *Journal of Chinese Information Processing*, vol.32, no.1, pp.89–95, 2018. (in Chinese)
[27] Y. Weng and C. Xia, "A new deep learning-based handwritten character recognition system on mobile computing devices," *Mobile Networks and Applications*, vol.25, no.2, pp.402–411, 2020.
[28] C. L. Xia, "Research and application of water book image recognition algorithm based on deep learning," *Ph.D.Thesis*, Minzu University of China, China, pp. 26–41, 2019. (in Chinese)
[29] H. Zhao, H. Chu, Y. Zhang,*et al.*, "Improvement of ancient shui character recognition model based on convolutional neural network," *IEEE Access*, vol.8, pp.33080–33087, 2020.
[30] Q. Ding, "Research on the feature extraction and classification method of Shui script under the Matlab platform," *Electronic Technology and Software Engineering*, vol.14, pp.155–157, 2020. (in Chinese)
[31] X. Z. Yang, S. Wu, H. Xia, *et al.*, "Research on Shui characters extraction and recognition based on adaptive image enhancement technology," *Computer Science*, vol.48, no.6A, pp.74–79, 2021. (in Chinese)
[32] Y. LeCun, L. Bottou, Y. Bengio, *et al.*, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol.86, no.11, pp.2278–2324, 1998.
[33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol.60, no.6, pp.84–90, 2017.
[34] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint*, arXiv:1409.1556, 2014.
[35] K. He, X. Zhang, S. Ren, *et al.*, "Deep residual learning for image recognition," *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA, pp.770–778, 2016.
[36] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint*, arXiv:1409.0473, 2014.
[37] S. Woo, J. Park, J. Y. Lee, *et al.*, "CBAM: Convolutional block attention module," in *Proceedings of European Conference on Computer Vision (ECCV)*, Munich, Germany, pp.3–19, 2018.
[38] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, USA, pp.7132–7141, 2018.
[39] X. Li, W. Wang, X. Hu, *et al.*, "Selective kernel networks," in *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, USA, pp.510–519, 2019.
[40] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, TN, USA, pp.13713–13722, 2021.
[41] D. Misra, T. Nalamada, A. U. Arasanipalai, *et al*, "Rotate to attend: Convolutional triplet attention module," in *Proceedings of IEEE/CVF Winter Conference on Applications of Computer Vision*, Waikoloa, HI, USA, pp.3139–3148, 2021.
[42] Z. Liu, Y. Lin, Y. Cao, *et al.*, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of IEEE/CVF International Conference on Computer Vision*, Montreal, Canada, pp.9992–10002, 2021.

TANG Minli was born in Guizhou Province, China, in 1982. She is an Associate Professor at KaiLi University. She is currently studying for her Ph.D. degree at Xiamen University. Her research interests include computational intelligence, pattern recognition, and computer vision.
(Email: tangml@stu.xmu.edu.cn)

XIE Shaomin was born in Guangdong Province, China, in 1998. He is studying for a master's degree at Xiamen University. His research interests include computational intelligence and pattern recognition.
(Email: xsmin@stu.xmu.edu.cn)

LIU Xiangrong (corresponding author) was born in Hunan Province, China, in 1978. He is a Professor and Ph.D. Supervisor at Xiamen University. His research interests include computational intelligence, data mining, and computational theory.
(Email: xrliu@xmu.edu.cn)