Pak. j. life soc. Sci. (2025), 23(2): 680-695 E-ISSN: 2221-7630;P-ISSN: 1727-4915

Pakistan Journal of Life and Social Sciences

www.pjlss.edu.pk

https://doi.org/10.57239/PJLSS-2025-23.2.00202

RESEARCH ARTICLE

Enhanced LeNet Model for Chinese Calligraphy Style Recognition

Luyao Sun[1], Noridayu Manshor[2]*, Fatimah Khalid[3]

[1,2]Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia (UPM), Serdang, Malaysia

[3]Department of Multimedia, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia (UPM), Serdang, Malaysia

ARTICLE INFO

Received: Jul 17, 2025

Accepted: Sep 8, 2025

Keywords

Chinese calligraphy
Concat Layer,
Deep Learning,
LeNet-5 Model,
Style Recognition,

***Corresponding Author:**

ayu@upm.edu.my

ABSTRACT

The development of Chinese characters over centuries has a lengthy history and is of great value for research. However, examining Chinese calligraphic characters can be challenging and requires a comprehensive understanding of Chinese history. This paper presents the CCSR (Chinese Calligraphy Style Recognition) model for recognizing calligraphic styles, which is based on the LeNet-5 model with a new Concat layer that includes weight mean and scaling operations. The proposed CCSR model effectively classifies calligraphic styles and outperforms traditional recognition methods in terms of accuracy and efficiency. Compared to conventional feature extraction methods, the CCSR model achieves a recognition accuracy of 98.35%, significantly higher than the best traditional methods. When compared with existing deep learning models, such as those incorporating Squeeze-and-Excitation (SE) and Convolutional Block Attention Module (CBAM) modules, the CCSR model not only achieves competitive or superior accuracy but also reduces training time by up to 28.5 minutes. Furthermore, the CCSR model demonstrates its robustness by successfully classifying calligraphic styles from a challenging dataset that contains low-resolution and poorly preserved images. The model maintains an accuracy of 90.79%, outperforming existing methods in difficult recognition conditions. These results highlight the CCSR model's superior performance in both accuracy and training efficiency, particularly in challenging scenarios involving low-quality data.

INTRODUCTION

Chinese calligraphy is more than just beautiful writing—it's a window into China's rich history and deep cultural roots. As shown in Figure. 1, Chinese calligraphy can generally be grouped into five main styles: seal script, clerical script, cursive script, running script, and regular script (Gang et al., 2023).

The seal script (Figure. 1(a)) is one of the oldest forms, often seen on ancient artifacts and seals, with intricate and symmetrical characters. The clerical script (Figure. 1(b)) emerged later, characterized by flattened strokes and distinctive end flares, making it more practical for daily use. Cursive script (Figure. 1(c)) is expressive and flowing but can be challenging to read at times. The running script (Figure. 1(d)) strikes a balance between fluidity and order, resulting in an elegant and readable format. Finally, the regular script (Figure. 1(e)) is the most common today, characterized by clear and precise strokes. Each style reflects the cultural mindset of its time, showing that Chinese calligraphy is both an art form and a legacy.

Figure 1. The Five Styles of "唐". (a) Seal script; (b) Clerical script; (c) Cursive script; (d) Running script, and (e) Regular script

In addition to the five major classifications of Chinese calligraphy, each calligrapher has their unique style. As shown in Figure. 2, even though the four famous Chinese calligraphers: Ouyang Xun, Yan Zhenqing, Liu Gongquan, and Zhao Mengfu all used regular scripts, they each had unique and distinctive styles.

Figure 2. The Four Styles of "为". (a) Ou Style; (b) Yan Style; (c) Liu Style, and (d) Zhao Style.

The China Academic Digital Associative Library (CADAL), the largest Chinese calligraphy database, houses over 40,000 historical calligraphy works (Jin and Huang, 2023). Calligraphy recognition is primarily approached using traditional methods or deep learning techniques, such as Convolutional Neural Networks (CNNs). Traditional methods focus on two stages: feature extraction, using techniques such as Gabor filters (Karunarathne et al., 2024, Wavelet Transform (Zhu et al., 2014), Generalized Search Tree (GIST) (Schoemans et al., 2024), Histogram of Oriented Gradients (HOG) (Admass et al., 2024), and Scale-Invariant Feature Transform (SIFT) (Weigang et al., 2025), and classification, which often uses Support Vector Machines (SVM) (Hou et al., 2021), Euclidean distance (Li et al., 2022), or SoftMax classifiers (Alwagdani et al., 2023).

Traditional recognition methods for Chinese calligraphy are complex, particularly in terms of feature extraction and classification. The future focus is on deep learning models, which can achieve effective classification by integrating feature extraction and recognition. While traditional methods require extensive technical support and are intricate, deep

learning models, like AlexNet and ResNet (Li, 2016), have gained popularity for improving efficiency. However, most existing models have complex structures, making training time-consuming (He and Sun, 2015). This paper aims to reduce model complexity while maintaining accuracy. Recent models, such as Squeeze-and-Excitation (SE) (Zhang et al., 2019a) and Convolutional Block Attention Module (CBAM) (Zhang et al., 2021), are effective but feature multiple network structures. Therefore, this paper focuses on using a simpler model for accurate Chinese calligraphy classification.

This study proposes a Chinese Calligraphy Style Recognition (CCSR) model based on the LeNet-5 deep learning architecture. The model features a simplified structure that improves both training and inference times while achieving a recognition accuracy of 98.6%. The introduction of a Concat layer enhances feature integration, making the model more precise and computationally efficient.

Our Key Contributions Include:

- This study proposes a novel Chinese Calligraphy Style Recognition (CCSR) model that integrates the LeNet-5 architecture with an innovative Concat layer. The introduction of the Concat layer enables more effective feature integration, significantly enhancing both recognition accuracy and training efficiency. Unlike existing methods, the Concat layer allows for more precise and efficient feature fusion, demonstrating the innovation of maintaining high performance while simplifying the model structure and reducing training time.
- A real-time calligraphy style recognition method is introduced, which dynamically adapts to varying data quality, including low-resolution and poorly preserved images. The system maintains high accuracy (90.79%) in challenging conditions, outperforming traditional methods and deep learning models such as SE and CBAM.
- The model reduces training time by up to 28.5 minutes compared to other deep learning models, providing a significant improvement in both training efficiency and recognition accuracy, which makes the CCSR model a more accessible and practical solution for real-world applications in Chinese calligraphy recognition.

The paper is structured as follows: Section "Related work" examines traditional and deep learning techniques for recognizing Chinese calligraphy. Section "Methodology" describes the CCSR model and the new Concat layer, while Section "Experiments and discussion" describes the experiments and their outcomes. Finally, Section Conclusion presents the study's conclusion.

Related Works

Traditional Chinese calligraphy recognition involves two main steps: feature extraction and classification, with a focus on feature recognition methods. Global features, such as Gabor (Karunarathne et al., 2024) and GIST (Zhang et al., 2013), and local features, including Wavelet transform (Raju, 2008), HOG (Chen et al., 2016), and SIFT (Burger W, and Burge M, 2022), are used for extraction. However, accuracy rarely exceeds 90%. A hybrid method combining GIST and SIFT (Zhang et al., 2019b) achieved an accuracy of 92%, but the process is complex, requiring multiple feature extraction steps and Principal Component Analysis (PCA) for dimensionality reduction. Most conventional methods employ SVM classifiers, which excel in two-class classification but struggle with multi-class classification, necessitating the use of additional classifiers and thereby increasing training time (Kurani et al., 2023). Despite advancements, the accuracy of traditional recognition methods remains under 90% (Weigang et al., 2025; Zhang et al., 2013; Chen et al., 2016).

Traditional recognition methods are cumbersome and complex, particularly in the recognition steps of feature extraction and classification. Traditional methods usually use manually designed feature representations, such as shape, texture, or brushstroke-based feature extraction. These feature representation methods cannot accurately represent all calligraphic style features, resulting in decreased classification accuracy. Traditional recognition methods also require manual parameter adjustments of the feature extraction methods, which can impact the accuracy of feature extraction (Zhang et al., 2013). In this case, applying a deep learning model in Chinese character recognition is the focus of future development.

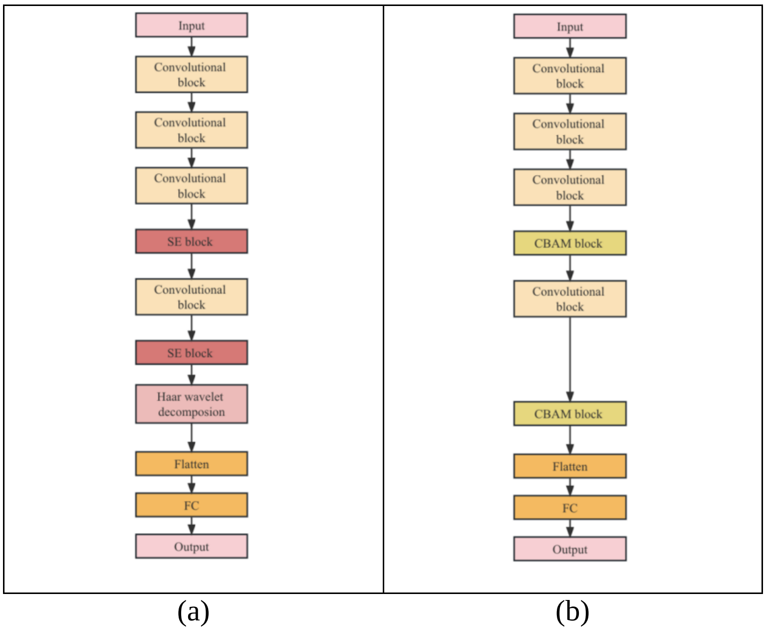

Figure 3. Structure of SE Model and CBAM Model. (a) SE model, (b) CBAM model.

In comparison to traditional models, deep learning models have the advantage of being able to classify data by learning its characteristics from a large amount of input data. Consequently, deep learning models can effectively increase recognition accuracy and decrease processing time. For example, the SE (Zhang et al., 2019a) and CBAM (Zhang et al., 2021) models have achieved recognition accuracies of over 97%. Compared with traditional recognition methods, the SE and CBAM models significantly improved recognition accuracy. Zhang, J. et al. added two SE modules, increasing the model depth by four layers, which in turn increased the computational requirements for both forward propagation and backward propagation, as well as the model complexity (Figure. 3(a)) (Zhang et al., 2019a). Zhang et al. added two CBAM modules, increasing the model depth by six layers, which also led to an increase in computational complexity (Figure. 3(b)) (Zhang et al., 2021). Therefore, employing multiple sets of SE modules and CBAM modules also increases the depth and complexity of the model, resulting in a longer path for the gradient to propagate through the network and increasing the computational effort, leading to a waste of training time.

Current deep learning models have complex model structures, which result in significant computational requirements and lengthy training and runtime times. However, when the model structure is simplified by reducing the number of layers, accuracy is again affected (He and Sun, 2015). Reducing training time and runtime while ensuring accuracy is now a key challenge in designing deep learning models.

The LeNet-5 model is the first deep learning model for classifying handwritten characters and the first and most effective model for classifying handwritten units (Nugraha et al., 2023). LeNet-5 consists of seven layers, including a convolutional layer, a sampling layer, a fully connected layer, and an output layer. Due to the lightweight nature of this network, both training time and the number of parameters are reduced, making the machine's classification easier. LeNet-5 is not only lightweight and effective but also applied to Chinese character recognition, where it achieves high accuracy in recognizing Chinese characters (Zaibi et al., 2021). Therefore, the LeNet-5 model is chosen as the base model in this study.

However, the LeNet-5 model also has some problems that can affect recognition accuracy. First, LeNet-5 does not include a data normalization operation, which may lead to significant changes in the weights and biases of the network, thereby affecting the stability of model training. The LeNet-5 model employs the Sigmoid activation function, which is an exponential operation with saturation. When the input value is very large or small, the output of the Sigmoid function is close to 1 or 0, resulting in a gradient close to zero, which will lead to the problem of gradient disappearance and slow down the parameter update during the model learning process, thus increasing the risk of over-fitting (Zaibi et al., 2021).

Based on the literature, this study proposes a CCSR model with a new Concat layer to reduce model training time while ensuring recognition accuracy. Section 3 describes the study's enhancements to LeNet-5 and the introduction of the new Concat layer.

METHODOLOGY

This section will discuss the improvements based on the LeNet-5 model and the operations included in the Concat layer. This section provides a detailed description of the CCSR model.

A. Improving the LeNet-5 for Calligraphy Style Recognition

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined.

Deep learning has a long history of development. The well-known deep learning expert LeCun proposed the pioneering LeNet-5 model, which is widely used to recognize handwritten digits (Zaibi et al., 2021). As shown in Figure. 4, there are a total of seven layers (not including the input layer), that is, C1 is a convolutional layer with six convolution kernels with a size of 5×5, S2 is a down-sampling layer (also called a pooling layer), and C3 is 16 convolutional layers of size 5×5, S4 is a down-sampling layer, C5, and F6 are two fully connected layers. The final output layer is a classification and recognition layer based on radial basis function (RBF).

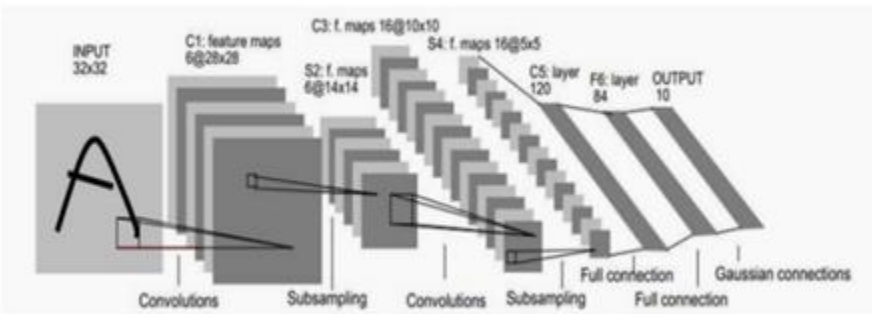

Figure 4. LeNet-5 Model (Zaibi et al., 2021)

Based on the LeNet-5 model, the proposed model will make several modifications and enhancements. As shown in Figure. 5, the Batch Normalization (BN) layer is carefully optimized, and the Rectified Linear Unit (ReLU) activation function replaces the LeNet-5 model's original Sigmoid activation function. The final classification output layer employs SoftMax, thereby generating many new convolutional neural network model structures, the BN-LeNet model.

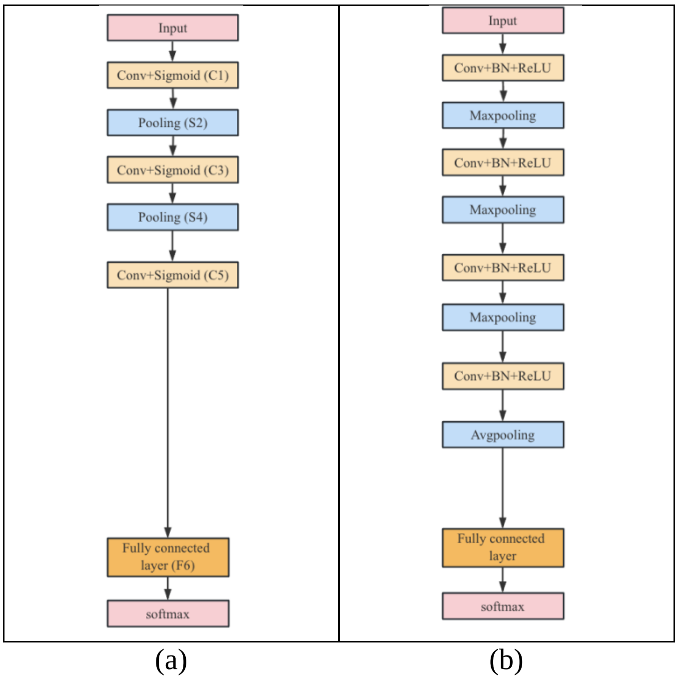

Figure 5. Structure Diagrams of the BN-LeNet and LeNet-5. (a) The Structure of the Original LeNet-5; (b) The Structure of the BN-LeNet model

The BN layer is added after the convolutional layer to prevent gradient dispersion during backpropagation of the loss function and to accelerate the convergence of the training model. After the BN, the nonlinear activation function Rectified Linear Unit (ReLU) is connected to enhance the characteristics of the nonlinear transformation (Luo et al., 2021). The complete BN algorithm operates as follows: At each Stochastic Gradient Descent (SGD) step, the activation is normalized by a mini-batch, ensuring the output has a mean of 0 and a standard deviation of 1. The final "scale and shift" operation guarantees that valuable data information is preserved (Segu et al., 2023).

The Entire BN-LeNet Algorithm Is As Follows, Where Ε Is A Constant:

Input: min-batch of Eigenvalue

$$\chi : B = \{X_1, X_2, X_3, \ldots, X_m\}$$

The parameters to be calculated are: γ , β

Output: $\{y_i = BN_{\gamma,\beta}(X_i)\}$

$$\mu_B = \frac{1}{m}\sum_{i=1}^{m} X_i \quad \text{// The mean value of min-batch}$$

$$\sigma_B^2 = \frac{1}{m}\sum_{i=1}^{m}(X_i - \mu_B)^2 \quad \text{//The variance of mini batch}$$

$$\widehat{X}_i \leftarrow \frac{X_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \quad \text{//Normalization}$$

$$y_i \leftarrow \gamma \widehat{X}_i + \beta : \equiv BN_{\gamma,\beta}(X_i) \quad \text{//Scales and shift}$$

In this paper, the activation functions used in the BN-LeNet model structure are all ReLU rather than the Sigmoid function used in the original LeNet-5 model. The ReLU function equation is denoted by Eq. (1).

$$f(x) = \max(0, x) \qquad (1)$$

The Sigmoid activation functions necessitate exponential computation, whereas ReLU only necessitates max computation. One of the benefits of the ReLU function lies in the fact that the calculation of this function is relatively simple compared to other activation functions and does not require a complex computation (Luo et al., 2021). As a result, the ReLU activation function can effectively accelerate the training process and shorten the time needed to achieve convergence (Hu et al., 2022)

SoftMax is typically used to solve multi-classification problems, whereas logistic regression is generally used to solve binary classification problems. The training data contains the category labels yi {1, 2..., k}, which are the values of the k categories. The prediction function of logistic regression can be extended to Eq. (2),

$$h_\theta(x_i) = \begin{bmatrix} p(y_i=1 \mid x_i;\theta) \\ p(y_i=2 \mid x_i;\theta) \\ \vdots \\ p(y_i=k \mid x_i;\theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{\theta_j^T(x_i)}} \begin{bmatrix} e^{\theta_1^T(x_i)} \\ e^{\theta_2^T(x_i)} \\ \vdots \\ e^{\theta_k^T(x_i)} \end{bmatrix} \qquad (2)$$

where θ1, θ2,..., θK are the model parameters to be trained, and $\frac{1}{\sum_{j=1}^{k} e^{\theta_j^T(x_i)}}$ is the normalization factor of the probability distribution to make the sum of all probabilities 1. To derive the loss function for SoftMax, define I {•} to denote that the rule for taking values is Eq. (3).

I {Expressions whose values are true} = 1

I {Expressions whose values are false} = 0 (3)

That is, for an input x, the probability that the classification category is j is Eq. (4).

$$p(y_i = j \mid x_i;\theta) = \frac{e^{\theta_j^T(x_i)}}{\sum_{j=1}^{k} e^{\theta_j^T(x_i)}} \qquad (4)$$

The gradient descent iterative algorithm is still used to solve the minimization problem for the SoftMax loss function, and the derivative of the parameter is given by Eq. (5).

$$\frac{\partial J(\theta)}{\partial \theta_j} = -\frac{1}{m}\sum_{i=1}^{m}\left[\left(Iy_i = j - p(y_i = j \mid x_i;\theta)\right)x_i\right] \qquad (5)$$

The gradient descent iteration can directly update the parameters to minimize the loss function according to Eq. (5), which is updated as follows (6).

$$\theta_k := \theta_k - \alpha\frac{\partial J(\theta)}{\partial \theta_j} \qquad (6)$$

B. Concat Layer

Based on the BN-LeNet structure, this section proposes a new Concat layer to improve the model's recognition accuracy after the last convolutional layer of the BN-LeNet model. Figure. 6 illustrates the architecture of the Concat layer. The output of the fourth convolutional layer is subjected to weighting means and scaling operations before the addition of the ReLU activation function. After scaling operations, the concat operation is then applied to the output features.

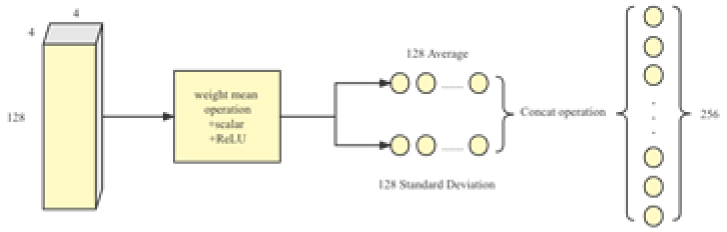

Figure 6. Structure of Concat Layer

The core concept of the Concat layer is inspired by the Mimic Norm, a novel normalization technique introduced by Fei et al. to enhance network convergence and training effectiveness. By utilizing the principles of Mimic Norm, the Concat layer introduces memory and productivity enhancements, contributing to the improved performance of the model (Fei et al., 2020).

The Concat layer initially performs the weight mean operation, as determined by the equations shown in Eqs. (7) and (8). S represents the initial weight value, "W" ˆW represents the weight mean value in Eq. (7). "M" "W" _"I,j" represents the number of input channels in Eq. (8). To meet the stability criteria, the weighting parameters are then scaled down by multiplying "1" /√("1-1/π") "≈1.2" (Fei et al., 2020). As the author removes the mean from each channel, the variance decreases, which this scaling compensates for

$$\widehat{W} = W - \mu W_{i,j} \tag{7}$$

where

$$\mu W_{i,j} = \frac{1}{I} \sum_{j=1}^{I} W_{i,j} \tag{8}$$

The output of the convolution layer (with weights mean+scalar+ReLU) is 128 x 4 x 4, indicating that 128 input channels of size 4 x 4 are used. Using Eqs. (9) and (10), the mean and standard deviation are computed in this phase. For each of the 128 input channels, the mean and standard deviation are independently calculated. The values are then recombined as the layer's output into a 256-dimensional vector feature, which is transmitted to the fully connected layer. By calculating the mean and standard deviation using the BN algorithm, the concat operation in this layer can improve recognition accuracy while reducing overfitting and accelerating training.

$$\mu = \frac{\sum_i \sum_j I(i,j)}{4*4} \tag{9}$$

$$\sigma = \sqrt{\frac{\sum_i \sum_j (I(i,j-\mu)^2}{4*4}} \tag{10}$$

C. CCSR Model Structure (BN-LeNet with Concat Layer)

The structure of the Chinese Calligraphy Style Recognition (CCSR) model is illustrated in Figure. 7. The network comprises a total of 10 layers. The first through seventh layers consist of four convolutional and three pooling layers. Except for the fourth convolutional layer (the seventh layer), each convolutional layer is followed by Batch Normalization (BN) processing and a nonlinear adjustment of the ReLU activation function. The eighth layer is the Concat layer, followed by the complete connection layer and the SoftMax classification layer.

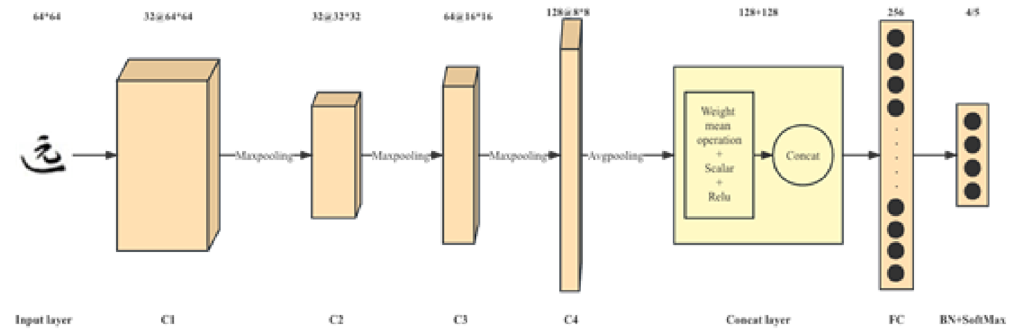

Figure 7. CCSR Model Structure

EXPERIMENTS AND DISCUSSION

In this section, experiments are conducted to validate the efficacy of the proposed CCSR model. First, three experimental datasets—D1, D2, and D3—are developed. Dataset D1 contains samples of four calligraphy styles by the renowned calligrapher Yan Zhenqing, spanning different historical periods (GXYB, DBT, MGXTJ, and YJMB). Dataset D2 includes four representative styles (Ou, Yan, Liu, and Zhao). Dataset D3 is a more challenging dataset comprising five fundamental script types: Seal (Zhuan), Clerical (Li), Regular (Kai), Running (Xing), and Cursive (Cao).

This section then presents three comparison experiments to assess the CCSR model's ability to recognize Chinese calligraphy styles. The first experiment compares the CCSR model with traditional recognition methods (HOG, GIST, PHOG, GIST+PHOG) using D1. The second experiment compares the proposed CCSR model with two existing deep learning models (SE and CBAM) using D2. The third experiment uses D3 to demonstrate the model's effectiveness even under poor image quality conditions.

D. Dataset

This section presents the Datasets D1, D2, and D3, which are Chinese calligraphy datasets used in the experiment. Each dataset is described in detail below.

1) D1: Four Styles of Yan Zhenqing (Zhang et al., 2019b)

In the experiment on style recognition of different fonts written by the same calligrapher, four calligraphic works by Yan from different periods are chosen to analyze the evolution of style by the same calligrapher. As shown in Figure. 8, from (a) to (d) are GuoXuji Tablet (GXYB), Duobaota Tablet (DBT), MaGu Tablet (MGXTJ), and Yanjiamiao Tablet (YJMB), of which GuoXuji Tablet and DBT are his early works, and MaGu Tablet and Yanjiamiao Tablet are his later works.

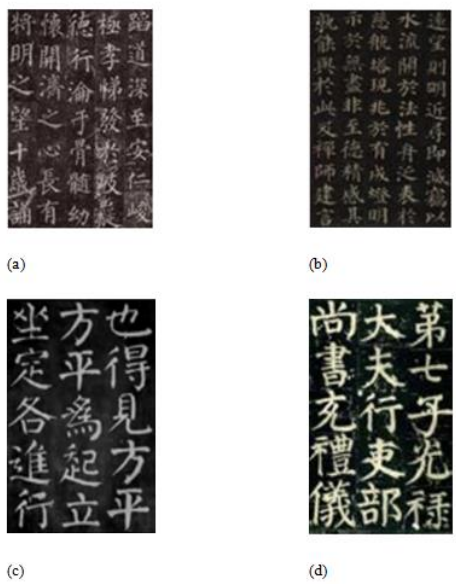

Figure 8. Four Calligraphic Works of Yan at Different Times. (a) GuoXuji Tablet (b) Duobaota Tablet (c) MaGu Tablet (d) Yanjiamiao Tablet

Totaling 3000 characters, GXYB has 800-character samples, DBT has 1000-character samples, MGXTJ and YJMB each have 600-character samples, and YJMB has 600-character samples.

2) D2: Four Calligraphers (Zhang et al., 2019a)

As shown in Figure. 9, this dataset includes four styles created by four calligraphers: Tang Dynasty Ouyang Xun (Ou), Yan Zhenqing (Yan), Liu Gongquan (Liu), and Yuan Dynasty Zhao Mengfu (Zhao). Each of the four styles of calligraphy contains 800 characters, totaling 3200 characters.

Figure 9. Four Famous Calligraphers Styles

3) D3: Five Types of Script (Challenging Dataset) (Wang et al., 2020)

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is "Heading 5". Use "Figureure caption" for your Figureure captions, and "table head" for your table title. Run-in heads, such as "Abstract", will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Datasets D1 and D2 are relatively simple to classify, as they are derived from clean, paper-based calligraphy works with minimal noise and preprocessing. To validate the model's efficacy, Dataset D3 was used, presenting a more challenging scenario. Derived from the MACT dataset (Wang et al., 2020), Dataset D3 includes severely damaged calligraphy from ancient books with worn characters, making recognition difficult. It contains five main scripts: Seal, Clerical, Regular, Running, and Cursive, and is challenging due to blurriness, noise, and imperfections. This dataset, comprising works from various scenes, presents a rigorous test case for the model, aiming to enhance recognition accuracy in low-resolution, noisy images and achieve optimal results in a shorter timeframe.

Figure 10. The Five Fundamental Scripts of Calligraphy

Specifically, Dataset D3 comprises a total of 9,116 samples of Chinese calligraphy characters. There are 1820-character samples of seal script (Zhuan), 1798-character samples of clerical script (Li), 1936-character samples of regular script (Kai), 1805-character samples of running script (Xing), and 1757-character samples of cursive script (Cao).

E. Experiment 1-Compare with Traditional Recognition Methods Using D1

The experiment utilized Dataset D1 to compare the CCSR model with traditional methods. The experiment was conducted by categorizing Yan Zhenqing's four styles. GXYB, DBT, MGXTJ, and YJMB are the four styles. HOG (Chen et al., 2016), GIST (Zhang et al., 2013), PHOG (Zhang et al., 2019b), and GIST+PHOG (Zhang et al., 2019b) were the traditional methods compared. The results are shown in Table 1.

Table 1. The Accuracy Compared with Traditional Methods

Style/ Method	HOG (Chen et al., 2016)	GIST (Zhang et al., 2013)	PHOG (Zhang et al., 2019b)	GIST+PHOG (Zhang et al., 2019b)	CCSR
GXYB	50.83%	86.67%	86.25%	94.58%	**99.10%**
DBT	69.17%	90.33%	88.67%	97.33%	**99.55%**
MGXTJ	59.17%	84.44%	81.67%	86.11%	**97.21%**
YJMB	65.00%	88.89%	84.44%	86.67%	**97.56%**
Overall	61.04%	87.58%	85.26%	91.17%	**98.36%**

In contrast to traditional recognition methods, the CCSR model achieves higher recognition accuracy. The experiments demonstrated that the CCSR model achieved a recognition accuracy of 98.36%. The experiment outcomes also demonstrated that the CCSR model outperformed the best traditional method (GIST + PHOG) by 7.19%. In summary, the CCSR model outperforms the conventional approach in terms of accuracy and recognition efficiency.

The results visualization of the five approaches is shown in Figure. 11. GIST+PHOG is the most effective traditional method for classifying styles, and it performs particularly well in classifying Yan Zhenqing's styles. GIST+PHOG achieves higher intra-class compactness. Figure. 11 also illustrates that the CCSR model has a greater gap between the different classes and that the data in each class is more compactly organized. As a result, the CCSR model achieves the best intra-class compactness and inter-class separability, demonstrating that it outperforms all existing recognition approaches.

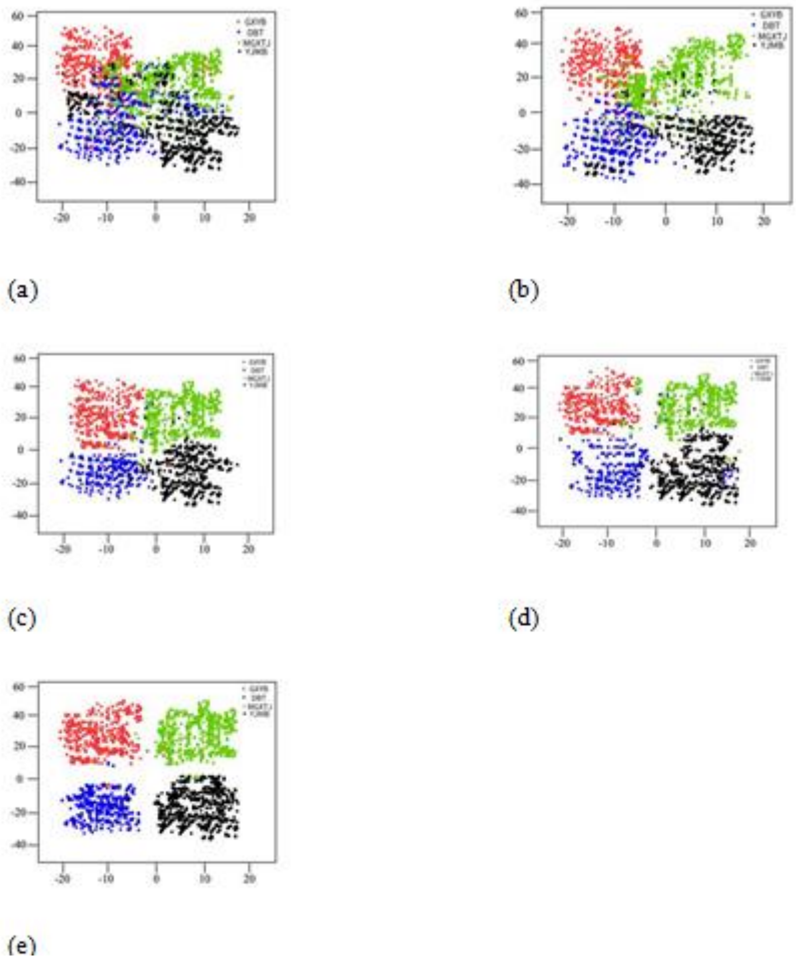

Figure 11. Data Visualization of Five Different Methods. (a) HOG (Chen et al., 2016), (b) GIST (Zhang et al., 2013), (c) PHOG (Zhang et al., 2019b), (d) GIST+PHOG (Zhang et al., 2019b), and (e) CCSR.

F. Experiment 2-Compare with Existing Deep Learning Models using D2

In this section, the CCSR model will be compared to two existing deep learning models for recognizing Chinese calligraphy styles. Models with SE (Zhang et al., 2019a) and CBAM (Zhang et al., 2021) were the two existing deep learning models for Chinese calligraphy style.

According to the models' respective structures, it is proven that the CCSR model has the most streamlined structure, with four convolutional blocks and a Concat layer. In the following section, experiments are designed to evaluate the recognition efficiency of the three models and the training time required to demonstrate that the CCSR model can minimize time consumption while maintaining accuracy.

Figure. 12 shows the relationship between training epochs and accuracy for the three models. The CCSR model quickly achieves and maintains the highest accuracy, outperforming the SE and CBAM models. All models show declining accuracy after 200 epochs, with no further improvement. Based on the training procedure, 200 epochs were determined to be the optimal training duration for all three models.

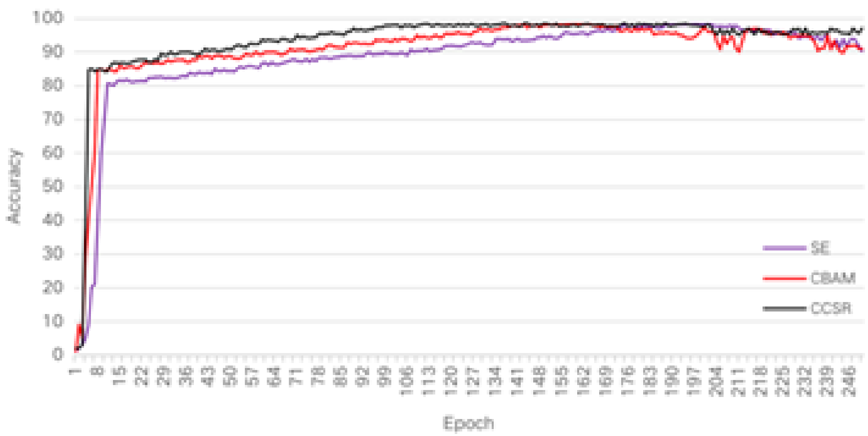

Figure 12. The Relationship between Epoch and Accuracy of Model Training

The experiment compared two deep learning models for Chinese calligraphy style recognition: the SE model (Zhang et al., 2019a) and the CBAM model (Zhang et al., 2021). Four calligraphy styles were represented in Dataset D2: Ou, Yan, Liu, and Zhao (3200 characters). The experiment analyzes explicitly the model complexity and recognition efficiency of the three models by displaying recognition accuracy, precision, recall, F1-scores, training time, and running time as experimental results.

Figures. 13, 14, and 15 illustrate the confusion matrix for the three models. Findings show that the recognition accuracy of all three models exceeded 96%, with the CBAM and CCSR models achieving over 98% across all subcategories. Particularly, the classification accuracy of the CCSR model for Ou, Yan, Liu, and Zhao exceeds 98%.

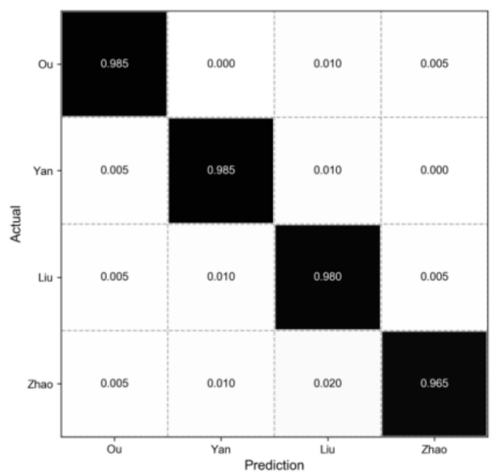

Figure 13. Confusion Matrix for the SE Model.

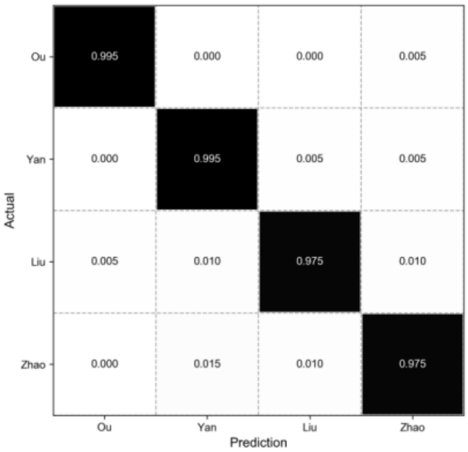

Figure 14. Confusion Matrix for the CBAM Model.

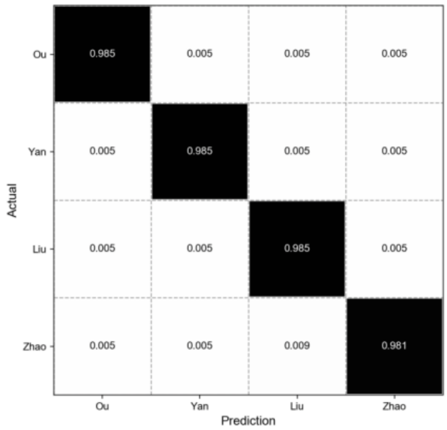

Figure 15. Confusion Matrix for the CCSR Model

In Table 2, the performance of three models is displayed. The three models comprise the SE model, the CBAM model, and the CCSR model. The CCSR model is only 0.1% less accurate than the CBAM model, but 0.6% more precise than the SE model. Simultaneously, each of the four CCSR model evaluation metrics was close to 98%, indicating the model's validity and stability. The CBAM model employs two CBAM modules and multiple convolutional blocks to focus on specific regions or features, enabling it to concentrate on particular areas or characteristics, which leads to high recognition efficiency but requires significant training resources. However, individual calligraphy character classification does not involve very complex feature information. The CCSR model can achieve effective style classification with only a single Concat layer for feature concatenation. Although the accuracy of the CCSR model is slightly lower than that of the CBAM model, its training efficiency is greatly improved, as it does not require complex mechanisms.

Table 2. The Results for the Three Models (Dataset D2)

Model	**SE (Zhang et al., 2019a)**	**CBAM (Zhang et al., 2021)**	**CCSR**
Precision	97.9%	98.3%	98.2%
Recall	97.8%	98.5%	98.4%
F1	97.8%	98.6%	98.3%
Accuracy	97.8%	98.5%	98.4%

This section reviews the model's training times, shown in Table 3. The CCSR model's training time was reduced by 16.68 minutes compared to SE and 28.48 minutes compared to CBAM. As a result, CCSR significantly reduces both training and recognition times while maintaining accuracy.

Table 3. 200 Epochs Training Time for the Three Models (Dataset D2)

Model	**SE (Zhang et al., 2019a)**	**CBAM (Zhang et al., 2021)**	**CCSR**
Training Time (min)	57.594	69.391	**40.911**

G. Experiment 3-Compare with Existing Deep Learning Models Using D3

To further validate the model's validity, Dataset D3 will be employed. In addition, the authors also utilize recognition accuracy, precision, recall, and F1-scores to determine model validity and training time to determine model complexity.

Figures. 16, 17, and 18 display the confusion matrices for the three models using Dataset D3. Compared to Figures. 13, 14, and 15, the accuracy of all three models decreased, indicating an increase in difficulty of recognition with Dataset D3. Figure. 18 also shows that not all the Kai scripts are classified as Cao scripts, so achieving high recognition accuracy may be possible if only the Kai and Cao scripts are classified.

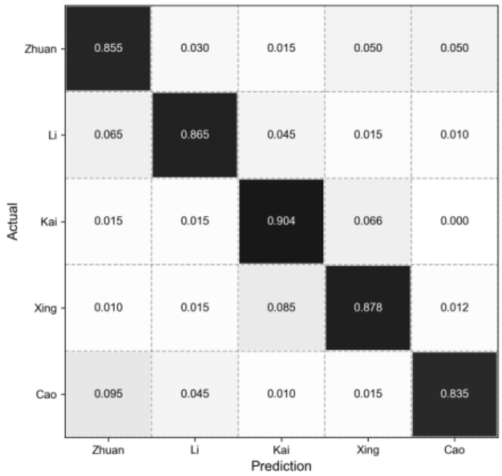

Figure 16. Confusion Matrix for the SE Model Using Dataset D3

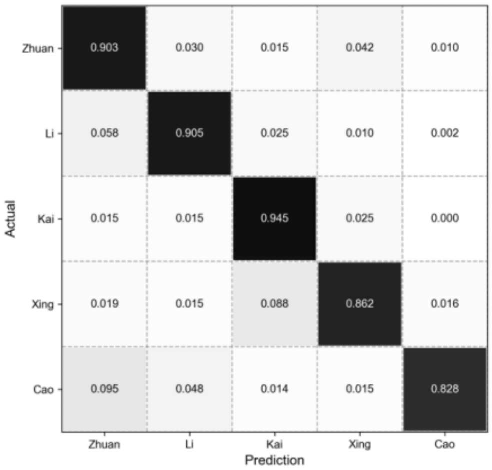

Figure 17. Confusion Matrix for the CBAM Model Using Dataset D3

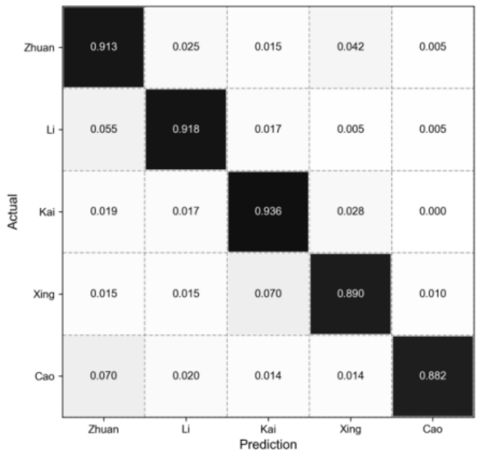

Figure 18. Confusion matrix for the CCSR model using Dataset D3

Table 4 shows the experimental results from Dataset D3, where performance metrics were lower than those for Dataset D2. The SE and CBAM models had precision, recall, F1-scores, and accuracy below 90%. In contrast, the CCSR model outperformed the others, with all four evaluation measures exceeding 90%. It achieved relatively high accuracy for the Zhuan, Li, and Kai scripts, with the Kai script reaching 93.63%. However, the accuracy for the Xing and Cao scripts was below 90%.

Table 4.The Results for the Three Models (Dataset D3)

Model	**SE (Zhang et al., 2019a)**	**CBAM (Zhang et al., 2021)**	**CCSR**
Precision	86.89%	88.50%	90.80%
Recall	86.77%	88.90%	90.79%
F_1	86.82%	88.69%	90.80%
Accuracy	86.77%	88.90%	90.79%

Table 5 discusses the recognition accuracy for each classification. The CCSR model demonstrated relatively high recognition accuracy for the Zhuan, Li, and Kai scripts, with the Kai script achieving a recognition accuracy of 93.63%. Moreover, the accuracy of the CCSR model was less than 90% for both the Xing and Cao scripts.

Table 6.The Accuracy of Three Models (Dataset D3)

Model	**Zhuan**	**Li**	**Kai**	**Xing**	**Cao**	**Overall**
SE (Zhang et al., 2019a)	85.55%	86.52%	90.39%	87.88%	83.51%	86.77%
CBAM (Zhang et al., 2021)	90.34%	90.57%	94.50%	86.21%	82.86%	88.90%
CCSR	91.32%	91.82%	93.63%	88.97%	88.21%	90.79%

Table 6 shows that the CCSR model's training time was reduced by 36.18 minutes compared to the SE model and 71.38 minutes compared to the CBAM model. While training times were significantly shorter, recognition accuracy on Dataset D3 decreased, and training time increased compared to Dataset D2, which suggests that lower image quality and more challenging recognition tasks lead to reduced model efficiency.

Table 6. 200 Epochs Training Time for the Three Models (Dataset D3)

Model	**SE (Zhang et al., 2019a)**	**CBAM (Zhang et al., 2021)**	**CCSR**
Training Time (min)	100.739	135.939	**64.558**

CONCLUSION

This paper improves the LeNet-5 model for enhanced Chinese calligraphy style recognition. The improved model includes four convolutional layers, each followed by a pooling and BN layer, with ReLU as the activation function. The result is a higher recognition rate for regular script from the four great calligraphers. A Concat layer placed before the fully connected layer further optimizes the model, resulting in the CCSR model, which has fewer parameters and faster training. The model achieved over 98% recognition accuracy on both datasets, meeting the goal of recognizing calligraphy style.

To test its real-world applicability, the study applied the CCSR model to the challenging Dataset D3, which includes blurry, noisy, and unclear calligraphy images from various scenes. The CCSR model achieved an impressive 90.65% accuracy. Compared to SE and CBAM models, it outperformed them in recognition accuracy while reducing training time by 36.18 and 71.38 minutes, respectively, demonstrating its efficiency and effectiveness.

The CCSR model has shown strong performance in recognizing individual Chinese calligraphy styles on single-character images. However, future research should focus on its application to multi-character calligraphy works, which present additional challenges due to complex arrangements and varying stroke thicknesses. While attention mechanisms like SE and CBAM focus on spatial and channel features, the CCSR model uses a concatenation layer to extract robust features. Investigating whether CCSR can maintain high accuracy in multi-character calligraphy and outperform attention-based models in this more complex scenario would be valuable.

Disclosure and Conflict of Interest

The author declares that there are no conflicts of interest related to this research. Additionally, the author has no financial interests or competing affiliations that could have influenced the study's design, execution, or findings. This manuscript is the author's original work and has not been previously published or submitted for review to any other journal or conference.

REFERENCES

Admass W. S., Munaye Y. Y., and Bogale G. A., 2024. Convolutional neural networks and histogram-oriented gradients: A hybrid approach for automatic mango disease detection and classification. International Journal of Information Technology, vol. 16, no. 2, pp. 817-829.

Alwagdani M. S., and Jaha E. S., 2023. Deep learning-based child handwritten Arabic character recognition and handwriting discrimination. Sensors, vol. 23, no. 15, p. 6774.

Burger W., and Burge M. J., 2022. Scale-invariant feature transform (SIFT). Digital Image Processing: An Algorithmic Introduction. Cham: Springer International Publishing, pp. 709-763.

Chen Y. S., Su G., and Li H., 2016. Machine learning for calligraphy styles recognition. CS229 Final Project, Stanford University, Palo Alto, California.

Fei W., Dai W., Li C., Zou J., and Xiong H., 2020. MimicNorm: Weight mean and last BN layer mimic the dynamic of batch normalization. arXiv preprint arXiv:2010.09278.

Gang Z. L., Luen L. C., and Cheong L. K., 2023. The Aesthetic Structure of Cursive Script. International Journal of Academic Research in Business and Social Sciences, vol. 13, no. 7, pp. 1159-1169.

He K., and Sun J., 2015. Convolutional neural networks at constrained time cost. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5353-5360.

Hou Q., Li C., Kang M., and Zhao X., 2021. Intelligent model for speech recognition based on SVM: A case study on English language. Journal of Intelligent & Fuzzy Systems, vol. 40, no. 2, pp. 2721-2731.

Hu H., Liu A., Guan Q., Qian H., Li X., Chen S., and Zhou Q., 2022. Adaptively customizing activation functions for various layers. IEEE Transactions on Neural Networks and Learning Systems, vol. 34, no. 9, pp. 6096-6107.

Jin J., and Huang C., 2023. China academic digital libraries over 20 years: From digital collaboration to digital knowledge service alliance. International Federation of Library Associations and Institutions (IFLA).

Karunarathne D., Wijesiriwardana C., Nishantha G., and Kumara M., 2024. Efficiency and accuracy in Sinhala handwritten character recognition: A Gabor-initialized CNN perspective. Sri Lanka Journal of Technology, vol. 5, no. 1, pp. 13-24.

Kurani A., Doshi P., Vakharia A., and Shah M., 2023. A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting. Annals of Data Science, vol. 10, no. 1, pp. 183-208.

Li B., 2016. Convolution Neural Network for Traditional Chinese Calligraphy Recognition. CS231N Final Project.

Li G., Fu L., Gao C., Fang W., Zhao G., Shi F., and Cui Y., 2022. Multi-class detection of kiwifruit flower and its distribution identification in orchard based on YOLOv5l and Euclidean distance. Computers and Electronics in Agriculture, vol. 201, p. 107342.

Luo X., Meng Q., Chen W., Wang Y., and Liu T. Y., 2021. Path-BN: Towards effective batch normalization in the Path Space for ReLU networks. Uncertainty in Artificial Intelligence, pp. 834-843.

Nugraha G. S., Darmawan M. I., and Dwiyansaputra R., 2023. Comparison of CNN's architecture GoogleNet, AlexNet, VGG-16, Lenet-5, Resnet-50 in Arabic handwriting pattern recognition. Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, vol. 8, no. 2.

Raju G., 2008. Wavelet transform and projection profiles in handwritten character recognition-A performance analysis. 2008 16th International Conference on Advanced Computing and Communications, pp. 309-314.

Schoemans M., Aref W. G., Zimányi E., and Sakr M., 2024. Multi-Entry Generalized Search Trees for Indexing Trajectories. Proceedings of the 32nd ACM International Conference on Advances in Geographic Information Systems, pp. 421-431.

Segu M., Tonioni A., and Tombari F., 2023. Batch normalization embeddings for deep domain generalization. Pattern Recognition, vol. 135, p. 109115.

Wang K., Yi Y., Tang Z., and Peng J., 2020. Multi-scene ancient Chinese text recognition with deep coupled alignments. Applied Soft Computing, vol. 108, p. 107475.

Weigang W., Ramos R. M., Brom P. C., and Li D. L., 2025. Threshold study for hanzi image recognition: Defining character and component limits in Chinese, Japanese, and Korean script processing. International Journal of Asian Language Processing, vol. 35, no. 1, p. 2450011.

Zaibi A., Ladgham A., and Sakly A., 2021. A lightweight model for traffic sign classification based on enhanced LeNet-5 network. Journal of Sensors, vol. 2021, no. 1, p. 8870529.

Zhang J., Guo M., and Fan J., 2019a. A novel CNN structure for fine-grained classification of Chinese calligraphy styles. International Journal on Document Analysis and Recognition (IJDAR), vol. 22, pp. 177-188.

Zhang J., Wang L., and Wen X., 2019b. Combination of GIST and PHOG features for calligraphy styles classification. 2019 4th International Conference on Multimedia Systems and Signal Processing, pp. 21-24.

Zhang J., Yu W., Wang Z., Li J., and Pan Z., 2021. Attention-enhanced CNN for Chinese calligraphy styles classification. 2021 IEEE 7th International Conference on Virtual Reality (ICVR), pp. 352-358.

Zhang Y., Liu Y., He J., and Zhang J., 2013. Recognition of calligraphy style based on global feature descriptor. 2013 IEEE International Conference on Multimedia and Expo (ICME), pp. 1-6.

Zhu C.-H., Gan H., and Wang J.-P., 2014. Research on classification and recognition of Chinese characters based on FSVM offline handwriting. Computer Engineering and Applications, vol. 23.