2024, 21(2 Special Issue): 0696-0705
https://doi.org/10.21123/bsj.2024.9768
P-ISSN: 2078-8665 - E-ISSN: 2411-7986

Baghdad Science Journal

Qin Seal Script Character Recognition with Fuzzy and Incomplete Information

Yun Ou*, Zhen-Jie Zhou, Di-Wen Kang, Pan Zhou, Xue-Wei Liu

School of Communication and Electronic Engineering, Jishou University, Jishou, 416000, Hunan, China.
*Corresponding Author.

ICAC2023: The 4th International Conference on Applied Computing 2023.

Received 30/09/2023, Revised 10/02/2024, Accepted 12/02/2024, Published 25/02/2024

Abstract

The dependable and efficient identification of Qin seal script characters is pivotal in the discovery, preservation, and inheritance of the distinctive cultural values embodied by these artifacts. This paper uses image histograms of oriented gradients (HOG) features and an SVM model to discuss a character recognition model for identifying partial and blurred Qin seal script characters. The model achieves accurate recognition on a small, imbalanced dataset. Firstly, a dataset of Qin seal script image samples is established, and Gaussian filtering is employed to remove image noise. Subsequently, the gamma transformation algorithm adjusts the image brightness and enhances the contrast between font structures and image backgrounds. After a series of preprocessing operations, the oriented gradient histograms (HOG) features are extracted from the images. During model training, different weights are assigned to classes with varying sample quantities to address the issue of class imbalance and improve the model's classification accuracy. Results show that the model achieves an accuracy of 95.30%. This research can help historians quickly identify and extract the text content on newly discovered Qin slip cultural relics, shortening the cycle of building a historical database.

Keywords: Ancient character recognition, HOG features, imbalanced data, image enhancement, SVM model.

Introduction

Qin seal script, also known as Qinjian, refer to the bamboo slips left by the Qin State during the warring states period and the late Qin Dynasty in China. They are essential cultural relics for studying Qin culture. The content recorded on Qinjian is important to archaeology, history, and linguistics research. Due to their ancient origin and factors such as natural erosion and improper preservation, some Qinjian have suffered from corrosion and dehydration, resulting in blurred and missing characters, making them difficult to identify. Identifying ancient Qin characters relies on knowledgeable and extensively trained domain experts. It is a considerable challenge to recognize the Qin characters efficiently and accurately because of some existing bottlenecks, such as the complexity of Qin characters, extensive character sets, the high similarity between characters, and the diverse writing styles.

The Qinjian images have various handwriting styles, exhibiting significant irregularity and uniqueness. Traditional template-matching algorithms need to perform better on such tasks. A deep neural network(DNN) is the ideal way to address the problem of handwritten font recognition. Lehan et al. first binarized the original ancient book image, cut out the text through the density and layout of the white pixels representing the text, and used the AlexNet model to complete the recognition task, and the recognition accuracy reached about 40%[1]. However, effective DNNs require much image data for parameter learning. Due to the scarcity of Qinjian image datasets at this stage, applying DNNs for the Qinjian image classification seems complicated.

Although there is limited research on Qinjian recognition, other studies on recognizing other objects or ancient scripts could serve as references. Llorca et al. addressed the problem of vehicle logo recognition in intelligent transportation systems. They utilized Local Binary Patterns (LBP), Scale-Invariant Feature Transform (SIFT), and Histograms of Oriented Gradients (HOG) as features for vehicle logos. They employed an SVM model to classify the content of sliding windows and estimated the classification of vehicle logos using majority voting[2]. Sonika et al. comprehensively considered multiple feature extraction and classification techniques. They achieved recognition accuracy of 88.95% by combining multiple classifiers using the majority voting strategy on a dataset of samples from ancient Sanskrit manuscripts[3].

Method Overview

A Five-Part Composition of the Task of Recognizing Partial and Blurred Qinjian:

- Dataset preparation;
- Data preprocessing;
- Feature extraction;
- Hyperparameter settings and model training.

Task Flowchart as shown in Fig. 1.

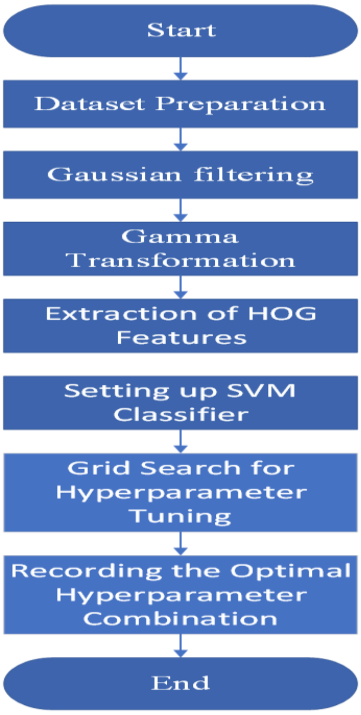

Start
Dataset Preparation
Gaussian filtering
Gamma Transformation
Extraction of HOG Features
Setting up SVM Classifier
Grid Search for Hyperparameter Tuning
Recording the Optimal Hyperparameter Combination
End

Figure 1. Task flow chart.

This study collected 681 digital images of handwritten characters and numbers from Qinjian slips, and the data annotation process was completed. The data was sourced from "The Qinjian in the Collection of Yuelu Academy," published by the Shanghai Lexicographical Publishing House in China. The text content in the image data is classified into fourteen categories, with some fonts experiencing stroke incompleteness. Fig. 2 shows the image samples from each category in the dataset and their corresponding Simplified Chinese characters.

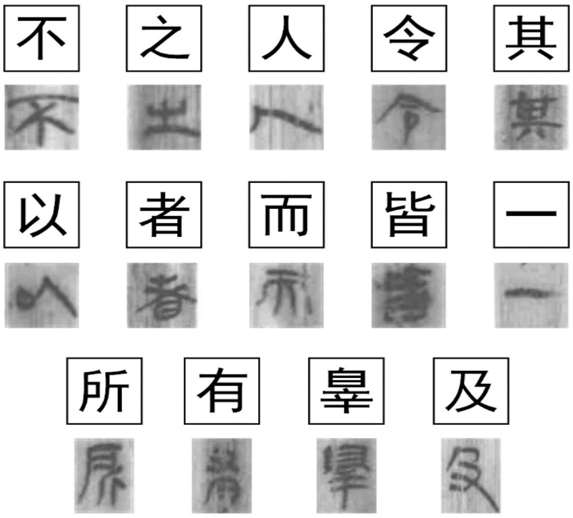


Figure 2. A Categorization of Text Content in Image Data and Corresponding Simplified Chinese Characters.

By processing and cutting the books in HD PDF format that recorded the Qin simplified images, the Qin simplified text images of different sizes were obtained. The distribution of categories in the data set is shown in Tab 1.

Table 1. The Distribution of classes For The Dataset.

Categories	Number of Classes
者	109
令	87
其	73
之	69
以	57
人	52
及	50
不	46
皆	42
而	42
一	40
有	39
皐	36
所	35

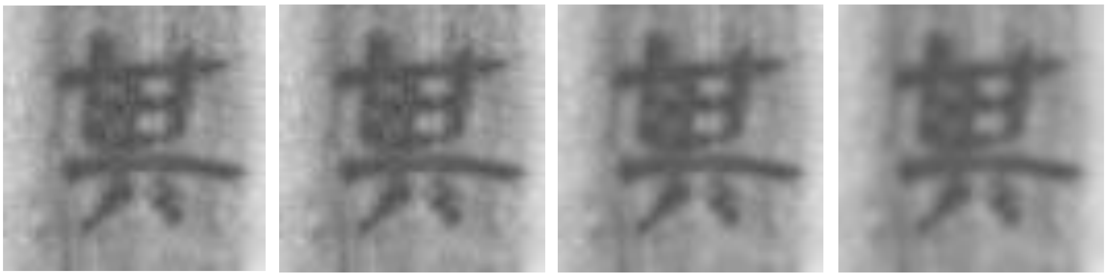
Preprocessing

The existing Qinjian images are limited by the image acquisition equipment and environmental conditions during the excavation of Qin bamboo cultural relics, resulting in poor image quality. Therefore, image preprocessing is required before extracting HOG features from the images.

After converting the image to grayscale, the Gaussian filtering algorithm removes image noise. Gaussian filtering is a simple and fast linear smoothing denoising algorithm in image enhancement technology, which is widely employed in the denoising process of image processing[4,5]. According to the selected fixed window size, each pixel in the image is scanned, and the weighted value of pixels in the neighborhood determined by the window is used to replace the value of the central pixel in the window. The value of the central pixel in a window is the weighted sum of all the pixel values within the template, with weights following a Gaussian distribution. The formula for computing each pixel value in an image is given by Eq. 1.

$$\begin{cases} P(x,y) = \sum_{j=y-n}^{y+n} \sum_{i=x-n}^{x+n} W(i,j)^{*} P(i,j) \\ Wi,j = \frac{w(i,j)}{\text{sum}} \\ \text{sum} = \sum_{j=0}^{2n} \sum_{i=0}^{2n} w(i,j) \\ w(i,j) = \frac{1}{2\pi\sigma^2} e^{-\frac{(i-n)^2+(j-n)^2}{2\sigma^2}} \end{cases} \quad 1$$

$W(i,j)$ represents the normalized weight value, while $W(i,j)$ denotes the original weight value. Sigma (σ) refers to the variance of x and y. The filtering effect is influenced by both σ and the size of the window. Specifically, a larger σ value leads to a better filtering effect but also a more blurred image. The effects for different σ values are compared in Fig. 3. Furthermore, a larger window size also leads to a better filtering effect but increases the blurriness in the image. This effect is compared for different window sizes in Fig. 4.



The Original Image

n = 3, σ =0.1

n = 3, σ =0.5

n = 3, σ =1.0

Figure 3. The effect of Gaussian filtering using different σ values (where n is the size of the convolution window and σ is the variance of the Gaussian function)

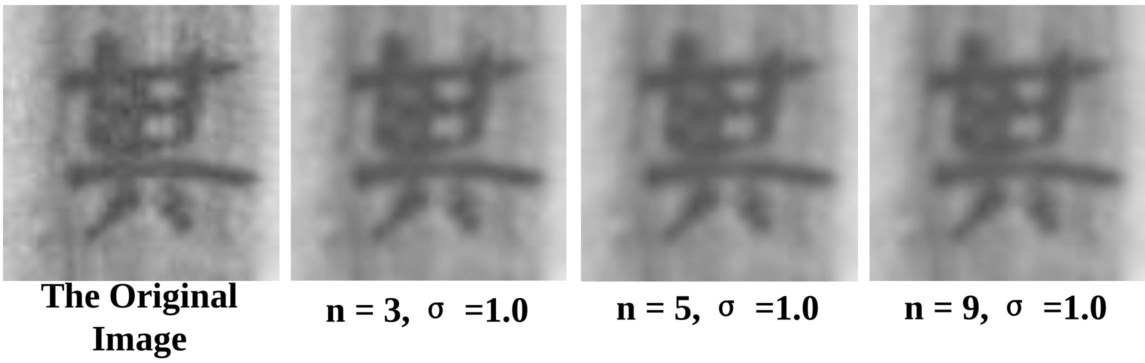


Figure 4. The effectiveness of Gaussian filtering can be adjusted by using windows of different sizes, where n represents the size of the convolution window and σ is the variance of the Gaussian function.

Intuitive observation of the image shows that the Gaussian filtering effectively reduces image noise. However, the image becomes somewhat blurred while increasing variance and window size. After removing image noise, gamma transformation is next used to adjust the image's gray values. Prior to computing the image's HOG features, it is necessary to reduce the effect of factors such as illumination disparities and local shadows on the image. The Gamma transformation is a simple and widely applicable method for image enhancement[6,7].The output image brightness is controlled by two parameters, γ and E. The definition of gamma transformation is given by Eq. 2:

$$I_{out} = EI_{\text{in}}^{\gamma} \qquad 2$$

Where, I_{in} represents the input image intensity while I_{out} represents the output image intensity. Whe E n is equal to 1, if $\gamma < 1$, then I_{out} will be darker than I_{in}; if $\gamma < 1$, then I_{out} will be brighter than I_{in} . The effect of gamma transformation applied to the Qinjian image is shown in Fig. 5.

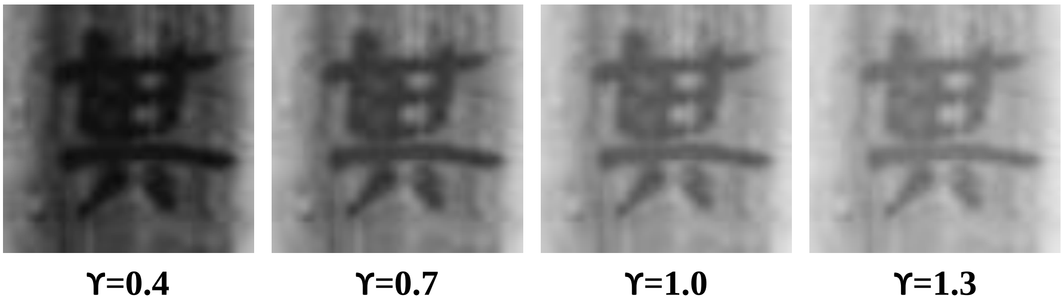


Figure 5. Effect of the gamma transform with different sizes of γ.

As can be seen from the effect image, the gamma transformation on the Qin bamboo manuscript image enhances the brightness and contrast, making the text stand out more prominently. When the value of γ is less than 1, the dark areas remain unchanged while the bright areas become darker as γ decreases. Conversely, when the value of γ is greater than 1, the bright areas remain the same while the dark areas become brighter as γ increases.

Feature Extraction

After completing the preprocessing operations on the images, extracting the HOG features of the images and mining highly discriminative feature vectors from the pixel matrices of the images are the next key operations.

To begin with, the Sobel operator is used to calculate the gradient information of the image. The Sobel operator is a type of discrete differential operator that performs weighted averaging and differentiation calculations on the pixel values of the image, resulting in an approximate value of the image gradient[8]. The operator consists of two sets of matrices, namely the horizontal and vertical matrices, which are convolved with the image in the x and y directions, respectively, to calculate the approximate values of the horizontal and vertical gradients. This is defined as Eq.3.

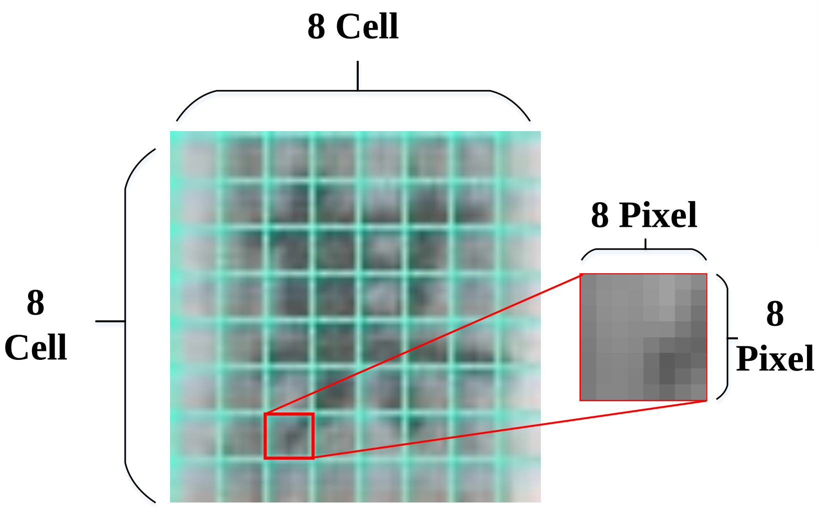
$$\begin{cases} G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * A, \\ G_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * A, \\ G = \sqrt{G_x^2 + G_y^2}, \\ \theta = \arctan\left(\frac{G_y}{G_x}\right). \end{cases} \qquad 3$$

In the equation, I_{in} represents the input image, G_x represents the horizontal gradient, G_y represents the vertical gradient, G represents the approximate gradient magnitude of the pixel, and θ represents the gradient direction.

After calculating the gradient values of the image, the next one is to divide the entire image into multiple cells and perform histogram statistics on all pixels within each cell. The interval [0,180] is uniformly divided into 9 parts to serve as channels for histogram statistics. Each pixel in the cell contributes to the histogram channel corresponding to its gradient direction through weighted voting where the weight is determined by the amplitude of the gradient. If the gradient direction of a pixel falls between two adjacent intervals, its amplitude is proportionally distributed to the channels of these two intervals[9]. .During voting, the image gradient direction is considered "orientation-free", meaning that two pixels with opposite gradient directions will contribute to the same channel. Taking the character in the image sample of size as an example, the cell division method and histogram calculation approach are shown in Fig. 6.

After computing the gradient histograms, each region of the image is normalized to mitigate the impact of brightness variations and other factors on feature extraction. The normalization process involves using a sliding window approach with a window size of four cells. The regions obtained by sliding the window are referred to as blocks. The gradient histograms of the four cells within each block are concatenated into a feature vector of length 36, which is then normalized.

The sliding window has a stride of 8 pixels in both the horizontal and vertical directions. It progressively moves in each direction, generating a feature vector of length 36 with each slide. For example, if it considers an image with a size of 64×64, the image can be divided into a total of 49 blocks by horizontally dividing it into 8 cells and vertically dividing it into 8 cells. Concatenating the histograms results in a HOG feature vector of length 1764.



12	13	4	8	14	2	25	39
12	14	1	5	15	10	38	43
13	15	5	6	17	25	43	39
13	16	6	7	23	44	41	27
13	16	6	15	35	50	26	1
11	13	5	24	43	24	16	28
10	12	3	23	36	16	36	32
10	11	5	16	28	34	38	23

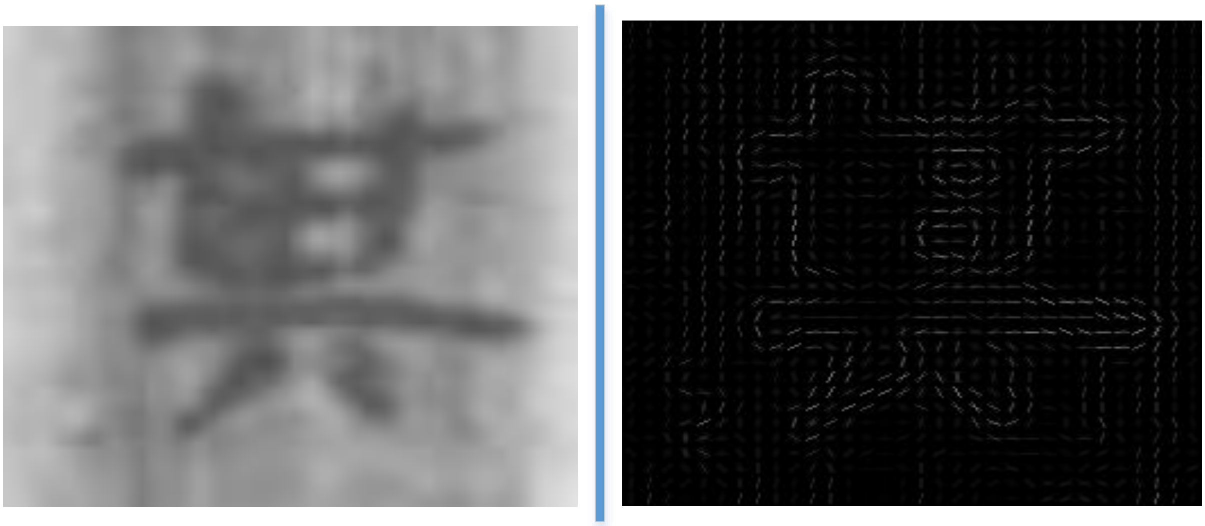
Gradient Magnitude

80	10	14	83	91	44	30	42
94	0	0	73	85	33	24	30
77	85	25	40	54	59	28	25
77	82	14	66	11	67	44	33
82	82	14	26	49	67	63	64
93	86	2	16	18	55	7	44
5	4	18	7	170	100	40	53
21	19	149	169	140	83	48	64

Gradient Direction

Figure 6. Cell division method and gradient information.

The visualization of the feature vector is depicted in Fig. 7, where the short lines indicate the gradient directions. From the visualization of HOG features, it can be observed that the gradient features corresponding to regions with more pronounced variations in grayscale values within the characters are extracted and preserved in the feature vectors, while other redundant information in the image is ignored.



A.Example Image

B.Feature Vector Visualization

Figure 7. Cell division method and gradient information.

SVM Classifier

The Qinjian images have various handwriting styles, exhibiting significant irregularity and uniqueness. This paper selects Support Vector Machines (SVM) as the core model to accomplish the target task.

Currently, the most common approaches to extend SVM for multi-class classification are one-against-one and one-against-all. When dealing with feature vectors extracted from Qinjian images, the key to achieving efficient and accurate classification using SVM lies in the choice of kernel function and the multi-class algorithm[10]. This study encounters two issues with the dataset: (1) scarcity of annotated data, allowing for limited data during model learning, and (2) severe class imbalance due to variations in character frequency, resulting in significantly different quantities of images for each class. It further indicates that the dataset used in this task suffers from severe data imbalance. Therefore, this study employs an algorithmic mechanism during the model training process, assigning a weight to each class to ensure that the model can focus more on the minority class and improve overall performance[11]. The specific calculation method for the weights of each class is as follows: the weight for each class is determined by taking the reciprocal of the frequency of occurrence for that class, as indicated in Eq. 4.

$$W_i = \frac{N}{N_i} \tag{4}$$

Here, N represents the total number of samples, N_i represents the number of samples in class i, and W_i represents the weight of each class.

The influence of category weights on the model training process is achieved by controlling the regularization parameters of different categories. The regularization parameter for the i-th category is represented by Eq. 5

$$C_i = W_i \times C \tag{5}$$

Here, C is the regularization hyperparameter set for the overall model.

Hyperparameter Setting

Data set partitioning is a crucial step in machine learning as it allows for evaluating model performance and tuning hyperparameters. In the case of image data sets with limited data, considering the small sample size, an employ the 20-fold cross-validation method for data set partitioning to conduct cross-validation and adjust model hyperparameters.

The text images cropped from the Qinjian exhibit varying sizes, requiring resizing the images to a uniform size of 64x64 before performing feature extraction. The gradient features of the images are sensitive to lighting conditions and other factors. Therefore, during the preprocessing stage, it is essential to preserve the gradient features of the images while emphasizing the textual content. A Gaussian filter with a window size of 3 and a mean of 1 is applied to the images to reduce noise. Additionally, a gamma transformation with a gamma value of 0.7 is utilized to enhance the contrast between the light and dark areas of the image. For Qinjian images of size 64x64, extracting the Histogram of Oriented Gradients (HOG) features yields a feature descriptor of length 1760. The parameters used for HOG feature extraction are set according to Eq.6.

$$\begin{cases} \text{bins} = 9 \\ \text{Cell} = 8 \times 8 \\ \text{Block} = 2 \times 2 \end{cases} \qquad 6$$

Results and Discussion

During the model performance evaluation phase, this study selects the following metrics to assess the model:

- Precision;
- Recall;
- F1 Score;
- Weighted Average Accuracy;
- Confusion Matrix.

The evaluation metric values calculated using cross-validation are presented in Tab 2, and the confusion matrix is shown in Fig .8.

For SVM multi-class models, the primary hyperparameters that need adjustment are[12]:

- Regularization parameter;
- Kernel function;
- Kernel function parameters.

To select the optimal hyperparameters for the model, this research utilizes grid search to search through the combinations of hyperparameters. The search range of the finally selected parameters is shown in Eq. 8.

$$\begin{cases} Regularizatoin\ Parameter = [0,\ 5] \\ Kernel\ Function = poly\ or\ rbf \\ Polynomial\ Degree = [2, 6] \end{cases} \qquad 8$$

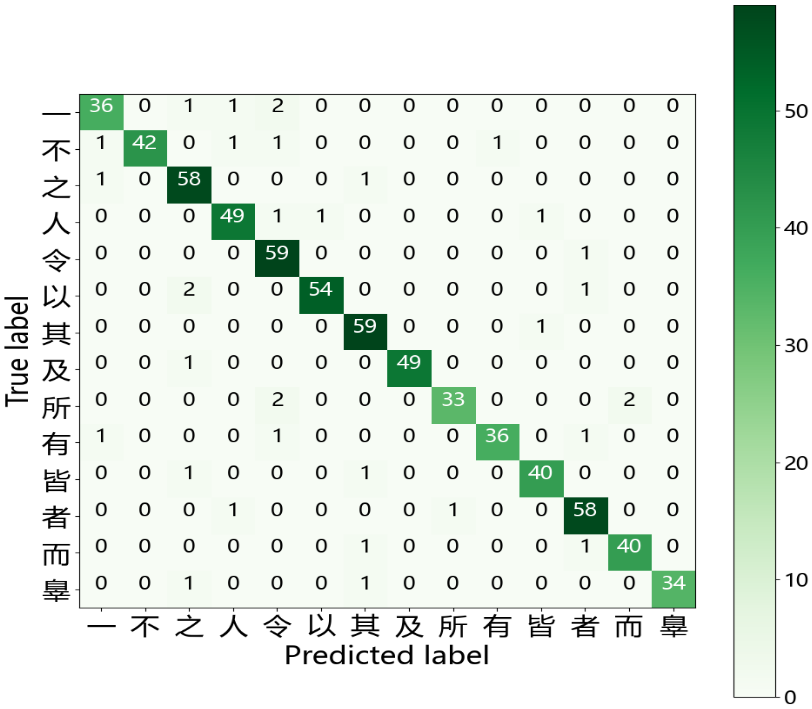
This paper exhaustively evaluate each combination to find the best hyperparameter values that maximize the model's performance. The model parameters are set according to Eq. (9)

$$\begin{cases} Regularizatoin\ Parameter = 0.1 \\ Kernel\ Function = poly \\ Polynomial\ Degree = 2 \end{cases} \qquad 9$$

Select the optimal hyperparameter combination with the highest classification accuracy on the validation set.

Table 2. Model Evaluation Metrics.

Categories	**Precision**	**Recall**	**F1 Score**
一	0.92	0.90	0.91
不	1.00	0.91	0.95
之	0.91	0.97	0.94
人	0.94	0.94	0.94
令	0.89	0.98	0.94
以	0.98	0.95	0.96
其	0.94	0.98	0.96
及	1.00	0.98	0.99
所	0.97	0.89	0.93
有	0.97	0.92	0.95
皆	0.95	0.95	0.95
者	0.94	0.97	0.95
而	0.95	0.95	0.95
皋	1.00	0.94	0.97
Weighted Average Accuracy	0.95	0.95	0.95



	一	不	之	人	令	以	其	及	所	有	皆	者	而	皋
一	36	0	1	1	2	0	0	0	0	0	0	0	0	0
不	1	42	0	1	1	0	0	0	0	1	0	0	0	0
之	1	0	58	0	0	0	1	0	0	0	0	0	0	0
人	0	0	0	49	1	1	0	0	0	0	1	0	0	0
令	0	0	0	0	59	0	0	0	0	0	0	1	0	0
以	0	0	2	0	0	54	0	0	0	0	0	1	0	0
其	0	0	0	0	0	0	59	0	0	0	1	0	0	0
及	0	0	1	0	0	0	0	49	0	0	0	0	0	0
所	0	0	0	0	2	0	0	0	33	0	0	0	2	0
有	1	0	0	0	1	0	0	0	0	36	0	1	0	0
皆	0	0	1	0	0	0	1	0	0	0	40	0	0	0
者	0	0	0	1	0	0	0	0	1	0	0	58	0	0
而	0	0	0	0	0	0	1	0	0	0	0	1	40	0
皋	0	0	1	0	0	0	1	0	0	0	0	0	0	34

Figure 8. Confusion Matrix.

Conclusion

As part of the digital processing practice for China's intangible cultural heritage, this project aims to promote accurate recognition of Qinjian manuscript handwritten characters. Given the limited amount and low quality of Qinjian character image data, an appropriate algorithmic model is adopted to recognize the complex and diverse handwritten ancient Chinese characters on the Qinjian while ensuring model stability and recognition accuracy efficiently and intelligently.

The rapid and accurate identification of Qin bamboo slips holds significant importance for the preservation and scholarly research of cultural relics. By employing automated methods to swiftly identify and extract pertinent information from Qin bamboo slips, the workload of archaeologists can be reduced, research cycles can be shortened, and the overall safeguarding of cultural relics can be promoted. The research outcomes can effectively address the fast and accurate recognition of Qinjian handwritten characters, assisting scholars in establishing a Qin historical database. In the future, this project can be extended to annotate Qinjian image data and optimize the models, leading to the construction of a Qinjian character database. The implementation approach can also be transplanted into digital recognition research of other rare languages or inscriptions on cultural relics.

For future research, efforts will be made to accelerate dataset expansion by collecting more sample data. In terms of model optimization, ensemble learning techniques can be employed to combine multiple models, enhancing the character categories while maintaining the high accuracy of the models. For future research, efforts will be made to accelerate dataset expansion by collecting more sample data. In terms of model optimization, ensemble learning techniques can be employed to combine multiple models, enhancing the character categories while maintaining the high accuracy of the models.

Acknowledgment

Project supported by the National Natural Science Foundation of China (62066016), the General project of Hunan Provincial Department of Education (22C0289), College Students in Hunan

Province and Scientific Research Project of Education Department of Hunan Province (Nos. 2958 and S202310531060), the College Students' innovation and entrepreneurship training program of Hunan Province, China([2021]197-2958).

Authors' Declaration

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Furthermore, any Figures and images, that are not ours, have been included with the necessary permission for re-publication, which is attached to the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee in Jishou University.

Authors' Contribution Statement

Write Z. J. Z. and Y. O. designed and directed the project. D.W. K. and P. Z. performed the experiments. All authors contributed to the final manuscript.

References

1. Chen L, Lyu B, Tomiyama H, Meng L. A method of Japanese ancient text recognition by deep learning. Procedia Computer Science. 2020 Jan 1;174:276-9. https://doi.org/10.1016/j.procs.2020.06.084.
2. Nagane AS, Patil CH, Mali SM. Classification of Brahmi script characters using HOG features and multiclass error-correcting output codes (ECOC) model containing SVM binary learners. In2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE) 2023 Jan 27 (pp. 448-451). IEEE.https://doi.org/10.1109/IITCEE57236.2023.10091084
3. Narang S, Jindal MK, Kumar M. Devanagari ancient documents recognition using statistical feature extraction techniques. Sādhanā. 2019 Jun;44:1-8. https://doi.org/10.1007/s12046-019-1126-9.
4. Suryanarayana G, Chandran K, Khalaf OI, Alotaibi Y, Alsufyani A, Alghamdi SA. Accurate magnetic resonance image super-resolution using deep networks and Gaussian filtering in the stationary wavelet domain. IEEE Access. 2021 May 5;9:71406-17. https://doi.org/10.1109/ACCESS.2021.3077611.
5. Zhang L, Wang X, Dong X, Sun L, Cai W, Ning X. Finger vein image enhancement based on guided tri-Gaussian filters. ASP Transactions on Pattern Recognition and Intelligent Systems. 2021 Apr 27;1(1):17-23. http://dx.doi.org/10.52810/TPRIS.2021.100012.
6. Shi Z, Feng Y, Zhao M, Zhang E, He L. Normalised gamma transformation - based contrast - limited adaptive histogram equalisation with colour correction for sand–dust image enhancement. IET Image Processing. 2020 Mar;14(4):747-56. https://doi.org/10.1049/iet-ipr.2019.0992.
7. Li G, Yang Y, Qu X, Cao D, Li K. A deep learning based image enhancement approach for autonomous driving at night. Knowledge-Based Systems. 2021 Feb 15;213:106617. https://doi.org/10.1016/j.knosys.2020.106617.
8. Zhou RG, Liu DQ. Quantum image edge extraction based on improved sobel operator. International Journal of Theoretical Physics. 2019 Sep 15;58:2969-85. https://doi.org/10.1007/s10773-019-04177-6.
9. Pang Y, Yuan Y, Li X, Pan J. Efficient HOG human detection. Signal processing. 2011 Apr 1;91(4):773-81.
10. Abdulmajeed AA, Tawfeeq TM, Al-jawaherry MA. Constructing a software tool for detecting face mask-wearing by machine learning. Baghdad Science Journal. 2022 Jun 1;19(3):0642-. https://doi.org/10.21123/bsj.2022.19.3.0642.
11. He H, Garcia EA. Learning from imbalanced data. IEEE Transactions on knowledge and data engineering. 2009 Jun 26;21(9):1263-84. https://doi.org/10.1109/TKDE.2008.239.
12. Khanday AM, Khan QR, Rabani ST. Detecting textual propaganda using machine learning techniques. Baghdad Science Journal. 2021 Mar 10;18(1):0199-. https://doi.org/10.21123/bsj.2021.18.1.0199.

التعرف على أحرف Qin Seal Script بمعلومات غامضة وغير كاملة

يون وي، زين جي زو، دي وين كانج، بان زو، زو وي لو

كلية الاتصالات والهندسة الإلكترونية، جامعة جيشو، جيشو، 416000، هونان، الصين.

الخلاصة

يعد التحديد الموثوق والفعال لأحرف نص ختم تشين أمرًا محوريًا في اكتشاف القيم الثقافية المميزة التي تجسدها هذه القطع الأثرية والحفاظ عليها ووراثتها. تستخدم هذه الورقة رسومًا بيانية للصور لميزات التدرجات الموجهة (HOG) ونموذج SVM لمناقشة نموذج التعرف على الأحرف لتحديد أحرف نص ختم Qin الجزئية وغير الواضحة. يحقق النموذج التعرف الدقيق على مجموعة بيانات صغيرة وغير متوازنة. أولاً، تم إنشاء مجموعة بيانات لعينات صور البرنامج النصي لختم تشين، وتم استخدام التصفية الغوسية لإزالة ضوضاء الصورة. وبعد ذلك، تقوم خوارزمية تحويل جاما بضبط سطوع الصورة وتعزيز التباين بين بنيات الخطوط وخلفيات الصورة. بعد سلسلة من عمليات المعالجة المسبقة، يتم استخراج ميزات الرسوم البيانية التدرجية الموجهة (HOG) من الصور. أثناء التدريب على النموذج، يتم تعيين أوزان مختلفة للفئات ذات كميات عينات مختلفة لمعالجة مشكلة عدم التوازن الطبقي وتحسين دقة تصنيف النموذج. أظهرت النتائج أن النموذج حقق دقة قدرها 95.30%. يمكن أن يساعد هذا البحث المؤرخين في التعرف بسرعة على محتوى النص واستخراجه من آثار تشين سليب الثقافية المكتشفة حديثًا، مما يختصر دورة بناء قاعدة بيانات تاريخية.

الكلمات المفتاحية: التعرف على الأحرف القديمة، وميزات HOG، والبيانات غير المتوازنة، ونموذج SVM.