Industry Science and Engineering Vol. 1 No. 8, 2024

Date Recognition of Handwritten Chinese Documents Based on Object Detection and Character Classification

Zhenyu Liu, Jie Zhang*
School of Automation, Nanjing University of Science and Technology, Nanjing, China
**Corresponding Author.*

Abstract: With the increasing degree of informatization in human society, paper documents are increasingly stored and processed in the form of electronic documents. People often want to extract key information such as dates in the process of digitizing paper documents. However, for text images with severe linking, extracting key information is very difficult. This paper proposes a date recognition framework for handwritten Chinese documents. First, YOLOv9 is trained to detect dates in images. Then single characters are segmented according to the pixels and character edges of the date area. And finally, Support Vector Machine (SVM) and Convolutional Neural Network (CNN) are used to classify handwritten Arabic numerals and handwritten Chinese characters and recognize handwritten numerals. At the same time, the date area in the text is located and recognized according to the date writing habits in Chinese documents. The method in this paper was experimented on 150 different handwritten Chinese text images with dates, and the recognition accuracy of dates in images reached 91.3%. The method proposed in this paper effectively achieves the localization and recognition of dates in handwritten Chinese character documents.

Keywords: Handwritten Chinese Characters; Object Detection; Character Segmentation; SVM; CNN; Date Recognition

1. Introduction

Preserving handwritten documents in the form of digital images can effectively overcome the defects of paper documents, such as the aging of paper and the blurring of the contents of old documents. Handwritten documents generally contain key information such as date and title composed of characters. In order to extract character information from document images, Optical Character Recognition (OCR) [1] came into being. The emergence of OCR solves the important problem of transforming information from images into information that can be recognized by computers. However, unlike machine-printed characters, the characters in handwritten documents usually do not have unified image features. This determines that OCR is difficult to recognize handwritten text and the model is complex, especially for handwritten Chinese character recognition (HCCR) [2,3], which has many types of characters and complex character structures. At the same time, key information such as the date of document formation often contains more than one kind of characters, which brings challenges to the extraction of key information from handwritten Chinese document images.

In the past decade, text recognition technology based on target detection algorithms has been successfully applied to text detection in natural images and has achieved good detection results [4-6]. For example, tasks such as license plate recognition [7,8] and text recognition in signboards [4-6]. However, text in natural images usually has standardized features, which leads to certain limitations in extracting information from messy handwritten documents. It is undeniable that although OCR technology is becoming more and more mature [9], recognizing and extracting key information from the entire text through OCR requires accurate character segmentation and complex character recognition models and even natural language processing models (NLP) [10,11]. Therefore, when the task focuses on recognizing certain key information rather than the entire text, target detection algorithms can play a very important role. In addition, due to the high complexity of handwritten Chinese text images and the difficulty in effectively segmenting characters [12], although there are

many HCCR methods [3], they also have certain limitations in application. How to extract key information such as date from handwritten Chinese texts through a more accurate, stable and universal method is a very practical issue.

To solve the above problems, this paper proposes a date recognition method for handwritten Chinese document images based on YOLOv9 [13], which combines SVM and CNN. The method is divided into three parts. In the first part, the target detection algorithm is innovatively combined with text recognition. The more prominent features of date in document images are used to locate the date in the handwritten Chinese document through the currently more advanced target detection algorithm YOLOv9, and the image where the located date part is output. In the second part, the image where the located date part is preprocessed, and a text character segmentation algorithm for handwritten Chinese characters is designed. The algorithm achieves effective character segmentation by performing pixel accumulation and edge detection on the characters in the images. At the same time, the feature vectors of the characters are generated according to the geometric features of the character image and the accumulation of binary image pixels. Finally, in the third part, SVM is used to classify the character feature vectors to determine whether the characters are Arabic numerals or Chinese characters, and make corresponding labels. The labeled characters are processed through the recognition model based on LeNet-5. And the recognized date is finally output.

The method proposed in this paper has three main advantages. First, this method takes advantage of the target detection algorithm and is more efficient in date recognition than global HCCR. Second, this method only performs character segmentation and recognition on the date area, and date recognition combined with SVM classification has better results. Third, since it does not rely on a unified network model, this method is relatively flexible and has better universality to a certain extent. The method in this paper does not involve image preprocessing issues, and intends to perform date recognition in handwritten Chinese documents that have been preprocessed. We used the method in this paper to test 150 handwritten and computer-generated Chinese document images and analyzed their recognition accuracy. The specific contributions of this paper are as follows:

- YOLOv9 is used to train a handwritten Chinese document dataset that has been marked with date labels. And a date detection model for handwritten documents based on the target detection algorithm is established, which provides a new idea for extracting key information from handwritten document images;
- A character segmentation algorithm for handwritten Chinese date is designed. The image is segmented initially based on the edge detection algorithm. And then the image is binarized and pixel accumulation is performed. The original image is further segmented by combining the pixel accumulation statistics and edge segmentation effect to improve the segmentation effect of the character segmentation algorithm on the adhesion font;
- A handwritten Chinese date recognition framework based on SVM character feature classification and CNN handwritten numeral recognition is proposed. The single-character image features generated by segmentation are used to distinguish handwritten numerals from handwritten Chinese characters. The trained CNN model is used to perform numeral recognition on characters classified as numerals. The character classification and numeral recognition models are separated to improve the recognition effect and flexibility of the algorithm.

The rest of this paper is organized as follows: Section II summarizes the research results achieved in current related work. Section III introduces the proposed method in detail. Section IV shows the experimental results and performance of the proposed method. Section V presents our conclusions.

2. Related Work

The concept of OCR can be traced back to the 1920s [1]. Early OCR technology used algorithms to analyze the stroke features of English letters in character images and then matched them with character templates to recognize optical characters. With the continuous maturity of computer technology and neural networks, the current OCR tasks in natural images and text images are increasingly dependent on complex neural network frameworks. These frameworks no longer rely

on character segmentation algorithms, but directly recognize image features through neural networks, such as the neural network framework of RCNN combined with CTPN [14]. As a branch of OCR technology, HCCR, a technology for recognizing handwritten Chinese characters, has also made great progress in the past few decades [2]. Zhong et al.[15] used the structural features of the GoogleNet deep network model to design the HCCR-GoogleNet model for handwritten Chinese characters. Using the character directional feature map extracted by Gabor transform as input, they achieved a recognition rate of 96.74% on the CASIA-HWDB 1.0 dataset. Xiao et al.[16] used a lightweight CNN model to achieve fast offline handwritten Chinese character recognition, with a recognition rate of 98.33% and a recognition time of only 9.8 ms per character. However, this work requires a lot of model training. Shi et al. [17] designed a text adaptive segmentation algorithm based on the morphology of characters and achieved good handwritten Chinese character recognition results through a transfer learning model.

However, OCR tasks for different images have different analysis strategies and methods. Text recognition in natural images usually requires first segmenting the image instances, and then performing character segmentation and recognition on the instance regions containing text information. Wang et al. [18] designed the R-YOLO recognition framework, which is not limited to horizontal text, but can be used to detect text in any direction in natural images. Shashidhar et al. [19] proposed a vehicle license plate detection method based on YOLOv3, which detects the license plate region through the target detection algorithm. And then recognizes the characters through the trained CNN model. The model achieved a recognition accuracy of 91.5% on the constructed license plate test set. Since the characters in natural images are often machine-printed or artistic characters with obvious features, and the adhesion and coverage between characters are not high, the key to natural image text recognition is to establish an effective and stable image segmentation and character recognition model. For text images, the OCR tasks mainly target three types of objects, fully handwritten text, semi-machine typed text , and fully machine typed text. Among them, all machine-printed text and machine-printed text in handwritten document can be easily recognized by open sourceOCR libraries such as Tesseract, and the effect is good [5]. For the text in handwritten images, when building the recognition model, it is necessary to focus on issues such as text line segmentation, adhesion between characters, and character recognition accuracy [20].

As mentioned above, the OCR methods of text images can be generally divided into two categories: traditional methods based on character segmentation and methods based on neural network models. Among them, the methods based on character segmentation require text line segmentation and character segmentation of the global text, and the recognition effect depends largely on the effect of character segmentation [21]. Wu et al. proposed a semantic segmentation model based on label encoding for character segmentation, which achieved good results in both natural images and text images. Kohli et al. [22] performed pixel-level path search for characters in the images. This method does not rely on deep learning networks, but directly segments characters based on their structural features. Before the rise of neural networks, character recognition usually required extracting features such as strokes from segmented character images, and then matching them with standard character templates to achieve the purpose of recognition [1]. In 2013, the establishment of the MNIST dataset ushered in the era of deep learning for OCR. An et al. [23] trained a deep CNN model and achieved a test accuracy of more than 99% on the MNIST dataset. Not only handwritten digits, but also many categories of character recognition have begun to rely on classification and recognition methods based on neural networks. Wei et al.[24] applied the transfer learning framework to character recognition and used the Inception v3 deep neural network to build an OCR system based on transfer learning. After 50000 iterations of training, their research results achieved a recognition accuracy of up to 95% in high-quality images and 78% in low-quality images. Parthiban et al.[25] used a recurrent neural network (RNN) to classify and predict handwritten English letters, relying on the memory characteristics of the RNN to achieve a recognition accuracy of more than 90%. Rawls et al.[26] combined the

CNN used for character classification with a one-dimensional long short-term memory (LSTM) recurrent network to achieve non-segmented OCR of text content. Garg et al [27] combined the YOLO and Connectionist Temporal Classification (CTC) algorithms to achieve word localization of English handwritten text in the form of target detection, and used HTR and CTC to achieve good word recognition results.

In addition, in order to make the recognition effect more accurate and stable, some studies have focused on efficient character segmentation through neural networks [21]. These studies accurately segment all characters in text images, and combined with NLP, they can effectively extract key information from text images and even understand the text semantics. Given the complexity and irregularity of handwritten text images, the accuracy and efficiency of global character segmentation in images when extracting key text information still need to be improved. Therefore, considering the excellent effect of target detection algorithms in natural images, this paper applies target detection to date recognition in handwritten Chinese document images, and obtains a more accurate and fast key text information positioning method. At the same time, the addition of target detection algorithms greatly reduces the text content that needs to be processed and recognized. For OCR of small amounts of text, there is no need to use complex neural network models. Traditional character segmentation and recognition models can achieve high-precision date recognition effects.

3. Proposed Method

In this paper, the proposed method mainly includes three parts: date location, character segmentation and character recognition. YOLOv9 is used to locate the date text information in the handwritten Chinese document, and SVM and CNN are used to classify the segmented characters and recognize Arabic numerals respectively. The algorithm framework proposed in this paper is shown in Figure 1.

YOLOv9 is trained on a self-made handwritten Chinese text dataset with date information to locate dates in text images. Character segmentation and character recognition are the key issues in this paper. Specifically, this paper proposes a text segmentation algorithm based on edge detection and pixel accumulation, which can effectively segment characters in the date. For the segmented characters, a SVM classifier based on character image features is designed to classify date area characters into numerals and non-numeral characters, so that the CNN model works only on handwritten Arabic numerals recognition, improving recognition accuracy.

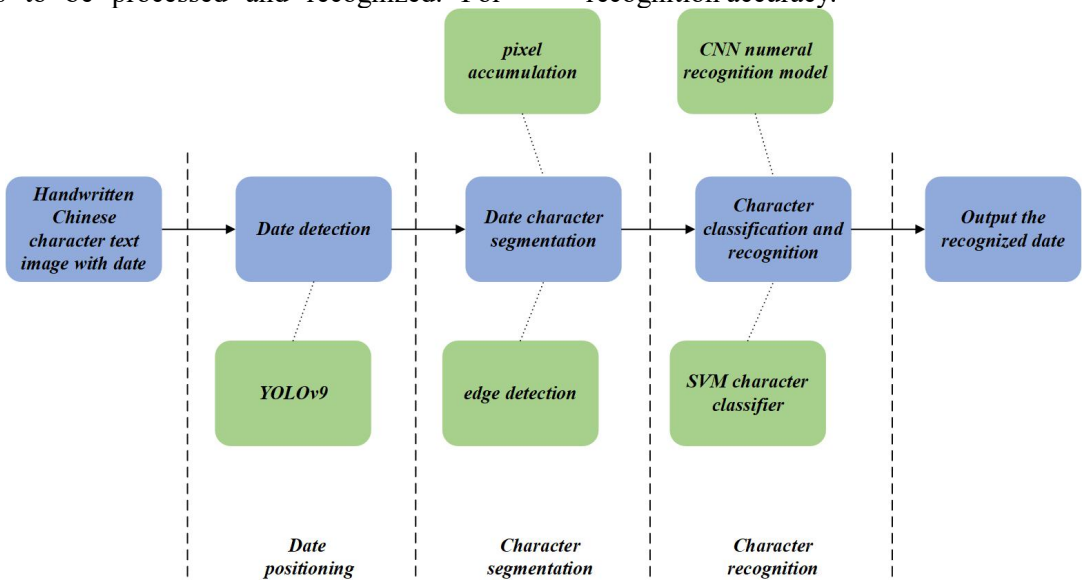


Figure 1. Handwritten Chinese Document Date Recognition Algorithm Framework

3.1 Date Detection Based on YOLOv9

The YOLO series of algorithms are widely used in various instance-based target detection tasks. YOLOv9 has been released in February 2024 and has made significant improvements compared to previous versions. It introduced technologies such as PGI and GELAN, which effectively solved the problem of information

loss during the forward pass of the neural network and improved the accuracy of detection [13]. This paper uses YOLOv9 to detect date information in handwritten Chinese document. We train it on a dataset of 1000 dates that are labeled. These datasets include some open archive images, handwritten text images, and handwritten text images generated by computers with date information.

The most important purpose of using YOLOv9 to detect handwritten Chinese document is to directly obtain the date area in the images through the target detection algorithm, rather than segmenting all characters and then locating the date through character recognition. Theoretically, the advantage of doing so is that it avoids the low efficiency and low precision of global character segmentation and text recognition, and greatly reduces the difficulty of subsequent segmentation and recognition of date text. This can greatly improve the efficiency of the algorithm. We will introduce the training effect of the above model and the recognition accuracy of the model on the test set in detail in the experimental section.

3.2 Character Segmentation Algorithm Based on Edge Detection and Pixel Accumulation

This paper uses a combination of Canny edge detection and pixel accumulation to achieve the segmentation of date characters. The pixel accumulation segmentation method based on a specific threshold enables the algorithm to have a good segmentation ability for contiguous characters. The date text image is divided into a surface original image and an underlying skeleton image for two segmentations. The Canny algorithm is used to perform preliminary segmentation on the surface original image to generate the first segmented character image. The Zhang-Suen algorithm [28] is used to extract the skeleton of the original image to obtain the stroke map of the character image. The stroke skeleton map of the generated character is obtained based on the preliminary segmentation. And the second step of segmentation is performed based on the pixel accumulation of the stroke character on the horizontal axis. The framework of the segmentation algorithm is shown in Figure 2.

Canny edge detection requires removing noise from the image through Gaussian filtering [29] to obtain a blurred grayscale image that preserves the image edge details. The Sobel operator is used to detect gradients on blurred images. The Sobel operator calculates the gradients of the image in the horizontal and vertical directions, and uses equation (1) to find the gradient and direction of the boundary in the gradient image. In Equation (1), $Edge\_Gradient$ is the image boundary gradient. $Angle$ is the boundary gradient direction. And G_x and G_y are the gradient maps calculated by the Sobel operator on the horizontal and vertical axes of the original image respectively.

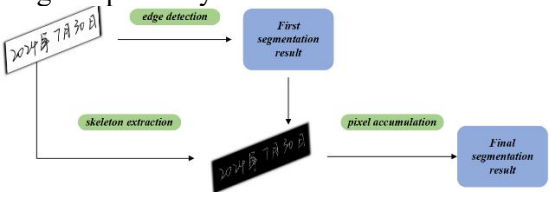


Figure 2. Character Segmentation Algorithm Framework Combining Edge Segmentation and Pixel Accumulation

$$Edge\_Gradient(G) = \sqrt{G_x^2 + G_y^2}$$
$$Angle(\theta) = \tan^{-1}\left(\frac{G_x}{G_y}\right) \quad (1)$$

After obtaining the gradient and direction of the image, the entire gradient map is traversed. The character edge points are retained and non-edge points are suppressed by the non-maximum suppression (NMS) method. NMS compares the gradient of the traversed pixel point with the gradient size of the pixel points with the same direction in the neighborhood. If the gradient is the maximum value in the neighborhood, it is determined to be an edge point, otherwise it is a non-edge point. The NMS determination conditions are as shown in Equation (2), where $Edge\_Point$ indicates that it is determined to be an edge point. $Deedge\_Point$ indicates that it is determined to be a non-edge point. P_G is the traversed pixel point. G_G is the gradient of the traversed pixel point. And P_R is the traversed pixel point in the gradient map. Other pixels in the neighborhoo d of the pixel that have the same direction as P_G (including P_G), G_R is the gradient size of P_R.

$$if\ G_G = Max(G_R\ in\ P_R)\ then\ P_G\ is\ Edge\_Point$$
$$if\ G_G < Max(G_R\ in\ P_R)\ then\ P_G\ is\ Deedge\_Point \quad (2)$$

After NMS, the edge points of the characters in the grayscale character image can be obtained. The edge points are connected to construct the closed curve segment image of the character

edge, and then the character image $\{I_O\}$ is segmented according to the maximum and minimum values of the coordinates of each edge curve segment in the image. The character stroke image $\{I_B\}$ can be extracted from the stroke map based on the image coordinates of these preliminary segmented characters. Taking every 4 pixels on the horizontal axis as the minimum statistical unit, the pixels of each image in $\{I_B\}$ are summed along the vertical axis to obtain the pixel statistics map $\{T_B\}$ about the horizontal axis. Take an element in $\{I_B\}$ as an example:

$$T_B = Sum(I_B)_{Asix=y\ and\ Step=4} \quad (3)$$

All images in $\{I_B\}$ are stroke skeleton images, and any point on all strokes is a single pixel point. Therefore, the parts where characters are connected in $\{T_B\}$ will have obvious low statistical values. Based on this, we set the secondary segmentation threshold ρ according to the minimum statistical unit. The intervals less than ρ in each statistical graph in $\{T_B\}$ are determined as segmentation points. The coordinates of the point with the smallest pixel cumulative value in the interval are used as the segmentation point σ to perform secondary segmentation on $\{I_O\}$, and the final segmented single-character image set $\{I_F\}$ is obtained.

$$I_F = Insert(\sigma|I_O)_{T_B} \quad (4)$$

3.3 Character Classifier Based on SVM

In the structure of Chinese text , the characters such as “年” , “月” and “日” contained in the date information have a relatively simple font structure , especially handwritten Chinese characters. The convolution features of these three characters are highly similar to Arabic numerals. Directly training the CNN classifier with these characters and ten Arabic numerals image samples will result in feature overlap, making it impossible to effectively extract date information from the image. Therefore, this paper constructs feature vectors based on the pixel distribution, density and other features that are significantly different between numerals and single-character images of Chinese characters. Classify handwritten numerals and Chinese characters in high-dimensional space through SVM.

As shown in Figure 3, according to the writing style of most people, the significant difference between Arabic numerals and Chinese characters in the same handwritten Chinese document is that Arabic numerals have fewer strokes. And the most complex numbers can be written with two strokes, such as “4” and “5”. At the same time, the cumulative distribution of pixels of Arabic numerals on the horizontal axis tends to be centered, and the character shape is relatively slender. These shape characteristics of character images provide a clear idea for designing feature vectors to distinguish Arabic numerals from Chinese characters.

(a) Chinese character “我”
Number of strokes: 7
Character width ratio: 56%
Character pixels: 1848

(b) Chinese character “年”
Number of strokes: 6
Character width ratio: 48%
Character pixels: 1540

(c) Chinese character “能”
Number of strokes: 10
Character width ratio: 51%
Character pixels: 1877

(d) Arabic numeral “6”
Number of strokes: 1
Character width ratio: 28%
Character pixels: 759

(e) Arabic numeral “4”
Number of strokes: 2
Character width ratio: 34%
Character pixels: 779

(f) Arabic numeral “5”
Number of strokes: 2
Character width ratio: 29%
Character pixels: 596

Figure 3. Example of Shape Feature Comparison Between Handwritten Arabic Numerals and Handwritten Chinese Characters

In order to describe the pixel distribution features of the image more accurately, preprocessing is first performed before extracting the character image feature vectors, including image grayscale and binarization. And then the binary image size is normalized. According to the edge of the binary image, a rectangular image containing characters is intercepted. The rectangle is expanded into a square based on the longer side of the rectangle. The expanded image is adjusted to 24×24 pixels, and 4 background pixels are expanded in four directions respectively to ensure that the character pixels are located in the center of the image. Finally, the character skeleton is extracted and the pixels are normalized. An example of the image preprocessing process is shown in Figure 4.

Figure 4. Character Image Preprocessing

The statistical graph of the character image is calculated according to Equation (3). The cumulative value of the pixels in each interval is ω_{Ti}. The feature vector ω_1 is extracted according to Equation (5). Among them, ω_i is the i-th element of the feature vector ω. φ is the pixel threshold. And ω_{1i} is set to 1 when ω_{Ti} is greater than φ in the interval, otherwise it is set to 0. In this way, a 7-element feature vector is generated for each 28×28 character image. The vector describes the density and distribution information by the distribution of pixels in the image on the horizontal axis. We perform statistics on 1000 preprocessed handwritten Chinese characters and 1000 handwritten numeral images, and calculate the average value of the pixels counted in all intervals as φ.

$$\omega_1 = [\omega_{1i}]_{\omega_{Ti}>\varphi,\ \omega_{1i}=1 | \omega_{Ti}<\varphi,\ \omega_{1i}=0},\ i = 1, \dots, 7 \quad (5)$$

As mentioned above, another significant difference between handwritten Chinese characters and numerals is the number of strokes, which can be simplified to the density of the character image in the entire 28×28 pixels. For the same 2000 data set mentioned above, after preprocessing, the proportion of stroke pixels in the entire image ω_2 is calculated. And ω_1 and ω_2 are used as the 8-dimensional feature vector ω for character classification, as shown in Equation (6).

$$\omega = [\omega_1\ \omega_2] \quad (6)$$

15000 images are selected from the MNIST handwritten numeral dataset and the HWDB1.0 handwritten Chinese character dataset for processing and feature vector ω is extracted. These images are sent to SVM for classification training. When training the SVM model, we use the Sigmoid kernel function to classify the feature vectors. Sigmoid maps data to the range of -1 to 1 in high-dimensional space to classify the data. And it has good processing capabilities for nonlinearly separable data.

$$K(x, y) = \tanh(ax^T y + b) \quad (7)$$

The expression K of the Sigmoid kernel function is shown in Equation (7), where a and b are the built-in parameters of the Sigmoid kernel function, x and y are the input feature vectors. The feature vector is mapped to a new feature space through the Sigmoid kernel function, so that the inseparable data becomes linearly separable in the new feature space to achieve the purpose of classification.

3.4 Handwritten Numeral Recognition Based on LeNet-5

This paper uses the LeNet-5 convolutional neural network model to classify and recognize handwritten numerals. LeNet-5 is a typical CNN model, and its network structure is shown in Figure 5. The input images we expect are character images of size 28×28, so the input layer in Figure 5 has been modified accordingly. At the same time, in order to reduce the complexity of the network and increase the training speed, the C5 convolution layer in the original LeNet-5 is deleted. And the features of the pooling layer are directly connected by the fully connected layer.

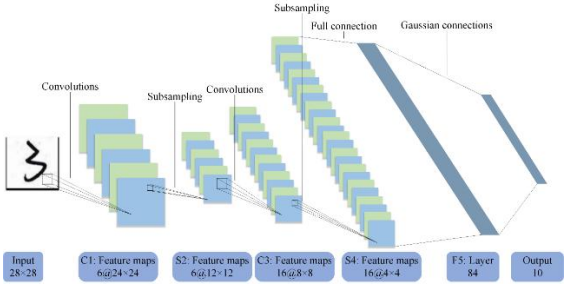


Figure 5. LeNet-5 Neural Network Structure Suitable for Handwritten Digit Recognition and Classification

In the modified LeNet-5, C1 to F5 are regarded as hidden layers. C1 is the first convolutional layer, consisting of 6 convolution kernels of size 5×5, and outputs 6 feature maps of size 24×24 by convolving the original image with a step size of 1. S2 is the first maximum pooling layer, with a pooling area size of 2×2, a step size of 2, and outputs 6 feature maps of size 12×12. C3 is the second convolution layer, which consists of 16 convolution kernels of size 5×5, a convolution step size of 1, and outputs 16 feature maps of size 8×8. S4 is the second maximum pooling layer, with a pooling area size of 2×2, a step size of 2, and outputs 16 feature maps of size 4×4. F5 is the fully connected layer, which connects the feature maps output by the S4 layer and uses the Sigmoid function as the activation function. The final output layer is the Softmax connection layer composed of 10 neurons to perform the classification task of ten Arabic numerals.

$$Softmax(x_j) = \frac{\exp(x_j)}{\sum_{i=1}^{n} \exp(x_i)}, j = 1, \dots, n \quad (8)$$

The Softmax activation function is shown in Equation (8). Where x_j is the input signal. And the denominator represents the exponential function sum of all input signals. After the input image is convolved and passed through the Softmax activation layer, the predicted probabilities of all classifications are obtained. The more features of a certain category the image contains, the larger the $Softmax$ value of the category it obtains, that is, the greater the probability that the image is predicted to be of this category.

The training process of LeNet-5 is divided into two stages: forward propagation and back propagation. In the forward propagation process, the training samples are output through each layer of the network. In the back propagation process, the mean square error between the output obtained by the forward propagation and the training sample labels is first calculated. And then the partial derivative of the error with respect to the weight in the network is calculated by the gradient descent method, and the weight and bias are updated. The training of the network model is completed by iteration until the preset parameters of the training are met. The character images are classified and predicted through the trained model. The category with the largest $Softmax$ value is taken as the recognition result to realize numeral recognition.

3.5 Generation of Handwritten Chinese Text Dates

Different from the date writing habits of other languages, Chinese dates are composed of Arabic numerals and Chinese characters alternately. We summarize Chinese dates into the following forms:

- "$xxxx$年x月x日"
- "$xxxx$年xx月x日"
- "$xxxx$年x月xx日"
- "$xxxx$年xx月xx日"

After the area of the date text is identified, the characters are segmented using the method proposed above. The SVM character classifier is used to distinguish between Arabic numerals and Chinese characters in the segmented characters of the date text. At the same time, the text is judged again based on the generated character classification labels and the above date writing forms to ensure that the text contains date information. The CNN model recognizes the numeral characters obtained by SVM classification. The recognized numbers are brought into the above-mentioned four date writing formats, and finally the recognized date is output.

4. Experiments

4.1 Handwritten Chinese Character Date Detection Experiment Based on YOLOv9

A dataset of 1000 handwritten Chinese text images is sent to YOLOv9 for training. These images included public archive images we searched online, artificial handwritten Chinese text images, and pseudo handwritten Chinese text images obtained by handwritten text generation software. They are divided into training set, validation set, and test set in a ratio of 7:2:1. We have trained the YOLOv9 model on an NVDIA GTX 4060 GPU with 16GB of RAM, setting the epoch to 300 and the batch_size to 8. The best training model curve output from YOLOv9 is shown in Figure 6.

The test set recognition results returned during the training process are shown in Figure 7.

We manually verified the detection results of 100 test sets, and have found that there are 3 images containing unrecognized dates and misrecognition situations similar to the last

image in Figure 7. That is, the recognition accuracy of the model on the dataset used in this article is 97%. And the confidence of all recognized date areas is above 0.75 on average.

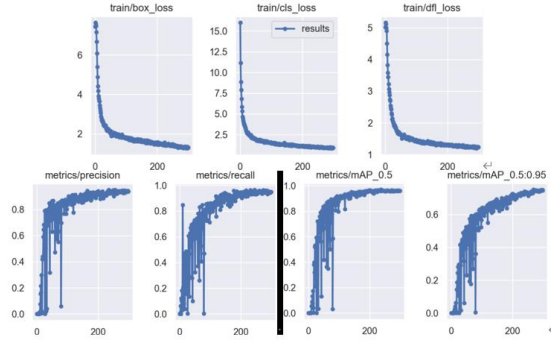
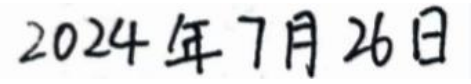


Figure 6. YOLOv9 Training and Test Result Curves

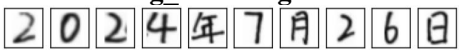
Figure 7. The Results of Test Set Output from YOLOv9



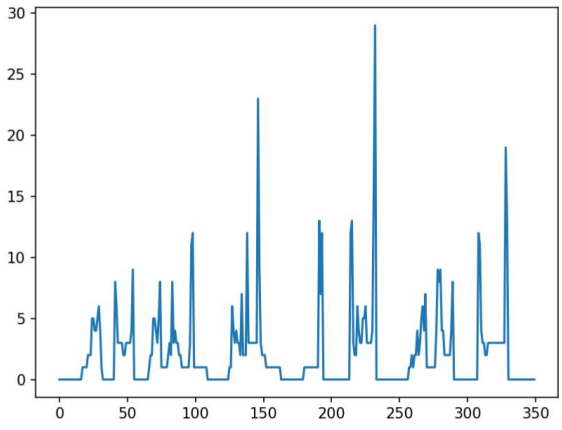
(a) Original Image of Handwritten Date in Chinese

(b) First Segmentation Image Using Canny Edge Detection

(c) Binary Skeleton Image Processed by Zhang_Suen Algorithm



(e) Final Segmented Image



(d) Skeleton Pixel Statistical Image

Figure 8. Character Segmentation Algorithm Processing and Result Image

Table 1. Comparison of the Accuracy of Three Character Segmentation Methods

	Number of mis-segmented images	Segmentation accuracy
Edge detection method	21	62.5%
Pixel accumulation method	18	67.8%
This paper's method	2	96.4%

4.2 DateText Character Segmentation Experiment

In order to verify the effectiveness of the character segmentation algorithm, several handwritten Chinese date images have been generated by handwriting and software. These images are processed using the proposed character segmentation algorithm. All handwritten date images are processed into 350×50 pixels in size to unify the parameters of the segmentation algorithm. Taking the date image in Figure 8(a) as an example, Figure 8 shows the processing of the character segmentation algorithm.

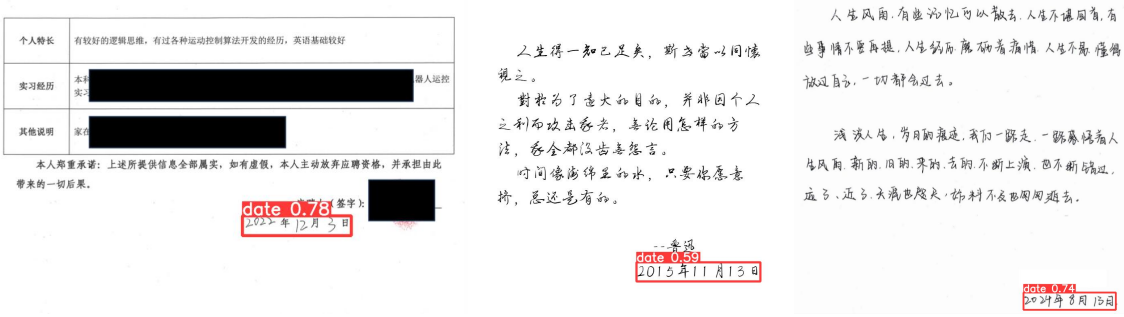
Figure 8(d) is the pixel cumulative statistics extracted from Figure 8(c). When the accumulated pixel value of a point is less than 3the segmentation point is set. The image obtained by the first segmentation is processed in stages. And a smaller erosion kernel is selected for more detailed edge detection. The final segmentation result is generated by combining the above processing, as shown in Figure 8(e).

We segmented 56 handwritten date images using edge detection-based character segmentation, pixel-based character segmentation, and the character segmentation algorithm proposed in this paper. The accuracy comparison is shown in **Table 1**.

4.3 Handwritten Chinese Character Date Recognition Experiment

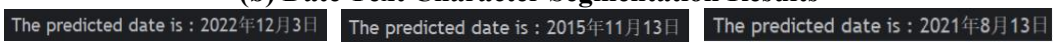
First, the performance of the SVM-based character classifier is tested. 2000 character images are selected from the MNIST handwritten digital data set and the HWDB1.0 handwritten Chinese character data set, respectively. According to the image

requirements of the method proposed in this paper, the preprocessing is performed. And the labels are divided into "digital characters" and "non-digital characters". The preprocessed character images are sent to the classifier for classification, and a classification accuracy of 92.8% is obtained. Secondly, the recognition accuracy of handwritten numeral characters is tested based on the training result of LeNet-5. During the training process, 2000 test sets are divided. And an accuracy of more than 97% is obtained on these 2000 test sets.

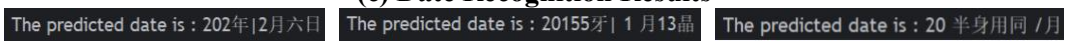


(a) YOLOv9 Detection Results

(b) Date Text Character Segmentation Results



(c) Date Recognition Results



(d) Tesseract Recognition Results

Figure 9. Example of Running the Date Recognition Algorithm

We conducted experiments on 150 self-made data sets, most of which are handwritten Chinese texts with dates generated by software, and also include some real handwritten Chinese texts and archive samples with dates. The recognition results of some samples are shown in Figure 9, and the accuracy of the results of each step is given in Table 2. In addition, we also use the open source, OCR library, Pytesseract to recognize the date recognized by YOLOv9, and the recognition effect is shown in Figure 9(d). From the comparison of the two recognition effect diagrams, it can be analyzed that Pytesseract is not ideal for the segmentation of text lines, resulting in low recognition accuracy. at the same time, when performing character recognition, the digital and non-digital characters are not classified like the our method, resulting in the situation in which the number "1" is recognized as the character "|" in the figure 9(d).

Table 2 The Accuracy of the Date Recognition Algorithm in Each Process

	Accuracy
Date text detection	95.3%
Date character segmentation	90.6%
Date recognition	91.3%

In order to further compare the recognition effect, the 1073 valid digital character images segmented above are recognized by the digital recognition model in this article and Pytesseract respectively. The recognition results of the two are shown in Table 3.

Table 3 Comparison of Recognition Results

	Number of misrecognized characters	Recognition accuracy
Pytesseract	79	92.6%
Numeral recognition model based on LeNet-5	31	97.1%

5. Conclusions

This paper designs a date recognition method for handwritten Chinese documents. We innovatively applies, the target detection algorithm YOLOv9 to the key information recognition task of text images. We proposes character segmentation method based on edge detection and pixel accumulation to segment the characters in the date text. SVM and LeNet-5 are used in the classification and recognition tasks of handwritten numbers and Chinese characters. Finally, the algorithm outputs the date according to the writing standard in Chinese characters. The proposed method is tested on a self-made dataset and obtained good recognition results. The work in this paper can be extended to other key

information extraction tasks in handwritten documents, such as document names, page numbers, and basic information in table documents. It will have a very broad application prospect when combined with NLP technology.

References

[1] Mori S, Suen C Y, Yamamoto K. Historical review of OCR research and development. Proceedings of the IEEE, 1992, 80(7): 1029-1058.

[2] Zhang X Y, Bengio Y, Liu C L. Online and offline handwritten Chinese character recognition: A comprehensive study and new benchmark. Pattern Recognition, 2017, 61: 348-360.

[3] Dai R, Liu C, Xiao B. Chinese character recognition: history, status and prospects. Frontiers of Computer Science in China, 2007, 1: 126-136.

[4] Haifeng D, Siqi H. Natural scene text detection based on YOLO V2 network model journal of physics: conference series. IOP Publishing, 2020, 1634(1): 012013.

[5] Chaitra Y L, Dinesh R, Jeevan M, et al. An impact of YOLOv5 on text detection and recognition system using Tesseract OCR in images/video frames. 2022 IEEE International Conference on Data Science and Information System (ICDSIS). IEEE, 2022: 1-6.

[6] Gao X, Han S, Luo C. A detection and verification model based on SSD and encoder-decoder network for scene text detection. IEEE Access, 2019, 7: 71299-71310.

[7] Laroca R, Severo E, Zanlorensi L A, et al. A robust real-time automatic license plate recognition based on the YOLO detector. 2018 international joint conference on neural networks (ijcnn). IEEE, 2018: 1-10.

[8] Chen R C. Automatic License Plate Recognition via sliding-window darknet-YOLO deep learning. Image and Vision Computing, 2019, 87: 47-56.

[9] Memon J, Sami M, Khan R A. Handwritten optical character recognition (OCR): A comprehensive systematic literature review (SLR). IEEE access, 2020, 8: 142642-142668.

[10] Nguyen T T H, Jatowt A, Coustaty M. Survey of post-OCR processing approaches. ACM Computing Surveys (CSUR), 2021, 54(6): 1-37.

[11] Lopresti D. Optical character recognition errors and their effects on natural language processing. Proceedings of the second workshop on Analytics for Noisy Unstructured Text Data. 2008: 9-16.

[12] Fujisawa H, Nakano Y, Kurino K. Segmentation methods for character recognition: from segmentation to document structure analysis. Proceedings of the IEEE, 1992, 80(7): 1079-1092.

[13] Wang C Y, Yeh I H, Liao H Y M. Yolov9: Learning what you want to learn using programmable gradient information. arxiv preprint arxiv:2402.13616, 2024.

[14] Liu Y. Sequence Recognition of Scene Text Based on CRNN and CTPN Models. Proceedings of the 2022 6th International Conference on Electronic Information Technology and Computer Engineering. 2022: 196-200.

[15] Zhong Z, Jin L,Xie, Z. High performance offline handwritten chinese character recognition using googlenet, and directional feature maps. 2015 13th international conference on document analysis and recognition (ICDAR). IEEE, 2015: 846-850.

[16] Xiao X, Jin L, Yang Y. Building fast and compact convolutional neural networks for offline handwritten Chinese character recognition. Pattern Recognition, 2017, 72: 72-81.

[17] Shi P, Lou Y, Xia R. Handwritten Chinese character recognition based on morphology and transfer learning. 2023 International Conference on Intelligent Perception and Computer Vision (CIPCV). IEEE, 2023.

[18] Wang X, Zheng S, Zhang C. R-YOLO: A real-time text detector for natural scenes with arbitrary rotation. Sensors, 2021, 21(3): 888.

[19] Shashidhar R, Manjunath A S, Kumar R S. Vehicle number plate detection and recognition using yolo-v3 and ocr method 2021 IEEE International Conference on Mobile. Networks and Wireless Communications (ICMNWC). IEEE, 2021: 1-5.

[20] Li X, Zhang X, Yang B. Character segmentation in text line via convolutional neural network. 2017 4th International Conference on Systems and Informatics (ICSAI). IEEE, 2017: 1175-1180.

[21] Wu X, Chen Q, Xiao Y. LCSegNet: An efficient semantic segmentation network for large-scale complex Chinese character recognition IEEE Transactions on Multimedia 23 (2020): 3427-3440.
[22] Kohli M, Kumar S. Segmentation of handwritten words into characters. Multimedia Tools and Applications 2021, 80: 22121-22133.
[23] An S, Lee M, Park S. An ensemble of simple convolutional neural network models for mnist digit recognition. arxiv preprint arxiv:2008.10400 2020.
[24] Wei T C, Sheikh U U, Ab Rahman A A H. Improved optical character recognition with deep neural network. 2018 IEEE 14th international colloquium on signal processing & its applications (CSPA). IEEE, 2018: 245-249.
[25] Parthiban R, Ezhilarasi R, Saravanan D. Optical character recognition for English handwritten text using recurrent neural network. 2020 International Conference on System, Computation, Automation and Networking (ICSCAN). IEEE, 2020: 1-5.
[26] Rawls S, Cao H, Kumar S. Combining convolutional neural networks and lstms for segmentation-free ocr2017 14th IAPR international conference on document analysis and recognition (ICDAR). IEEE, 2017, 1: 155-160.
[27] Garg N, Sharma N, Jain G. Handwriting Recognition System Using YOLO and CTC 2023 International Conference on Advanced Computing & Communication Technologies (ICACCTech). IEEE, 2023: 496-502.
[28] Chen W, Sui L, Xu Z. Improved Zhang-Suen thinning algorithm in binary line drawing applications 2012 International Conference on Systems and Informatics (ICSAI2012). IEEE, 2012: 1947-1950.
[29] Deng G, Cahill L W. An adaptive Gaussian filter for noise reduction and edge detection 1993 IEEE conference record nuclear science symposium and medical imaging conference. IEEE, 1993: 1615-1619.