International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:14, No:2, 2020

A Character Detection Method for Ancient Yi Books Based on Connected Components and Regressive Character Segmentation

Xu Han, Shanxiong Chen, Shiyu Zhu, Xiaoyu Lin, Fujia Zhao, Dingwang Wang,

***Abstract*—**Character detection is an important issue for character recognition of ancient Yi books. The accuracy of detection directly affects the recognition effect of ancient Yi books. Considering the complex layout, the lack of standard typesetting and the mixed arrangement between images and texts, we propose a character detection method for ancient Yi books based on connected components and regressive character segmentation. First, the scanned images of ancient Yi books are preprocessed with nonlocal mean filtering, and then a modified local adaptive threshold binarization algorithm is used to obtain the binary images to segment the foreground and background for the images. Second, the non-text areas are removed by the method based on connected components. Finally, the single character in the ancient Yi books is segmented by our method. The experimental results show that the method can effectively separate the text areas and non-text areas for ancient Yi books and achieve higher accuracy and recall rate in the experiment of character detection, and effectively solve the problem of character detection and segmentation in character recognition of ancient books.

***Keywords*—**Computing methodologies, interest point, salient region detections, image segmentation.

I. Introduction

AMONG the many ethnic minorities in China, the Yi people are an outstanding nation with more than 2,000 years of history that have formed their own unique culture in long-term development [21]. According to statistics [22], there are more than 1 million people using Yi characters. Due to geographical differences, there are obvious differences in the Yi characters of different areas. Therefore, there are many characters in the Yi texts, and only the characters in the texts of the Yunnan-Sichuan-Guizhou dialect include as many as 80,000. As an important minority language, Yi characters are currently being used and have left many precious classical books throughout history. The ancient Yi books are an important carrier for recording the development of the Yi people for thousands of years, and the vast volumes of ancient Yi books are widely collected in libraries, research and translation institutions in China. In other countries, some institutions in the United Kingdom, Japan, France, and Switzerland also have large collections of Yi books. There are tens of thousands of classical Yi books. The content of Yi literature covers religion, history, philosophy, literature, language, medicine, astronomy, geography and agricultural technology. At present, the ancient Yi books are scattered among the people with more than 100,000 volumes. In China, Guizhou province is the birthplace of the ancient Yi characters, and the number and quality of the ancient Yi books are the highest in the country. The ancient Yi books in Guizhou account for two-thirds of the volumes in all of China, which contain a large number of stone inscriptions. The well-known ancient Yi books found in this area include the Southwest Yi records and Cuan Wen Cong Ke. The traditional Yi characters in Guizhou are also known as the ancient Yi characters [21]. The scanned images of the ancient Yi books in this paper are provided by the Research Institute of Yi Nationality Studies, Guizhou University of Engineering Science, Bijie City, Guizhou Province.

Over time, because of the weak concept of protection, most of the ancient Yi books have been seriously damaged and need to be protected and utilized by digital means. Therefore, how to locate and segment the ancient Yi characters in the ancient Yi books and obtain more information about the history and culture of the Yi nationality from the ancient books is the focus of the digitalization study of ancient Yi characters. The first problem facing the digitization of Yi characters is transforming the ancient books into a computer-readable file format and realizing the information processing and analysis of the Yi books. As the carrier of Yi books, stone carvings, cliff paintings, hibiscus, and paper books are often ambiguous or incomplete due to age, which creates great challenges to the detection of ancient Yi characters.

At present, there have been many studies [8]-[10], [13]-[17] on Chinese-English character detection methods in complex backgrounds. Most of these studies [13]-[17] used methods based on deep learning to detect and recognize the characters. However, these methods are not completely applicable to ancient Yi books with complex noise. First, few people can recognize and write Yi characters in China, and character labeling is very difficult. Second, compared with the detection of handwritten Chinese characters, the detection of the ancient Yi texts in the background of complex noise faces many problems, such as blurred images, serious pollution, and a messy writing format. Therefore, a simple and effective method for detecting characters is required. In recent years, there have been many studies [1]-[5] on the detection and recognition of minority nationality characters, but most of them only studied the segmentation and recognition of

Xu Han, Shanxiong Chen*, Xiaoyu Lin, Fujia Zhao, and Dingwang Wang are with the College of Computer &Information Science. Southwest University, Chongqing 400715, China (*Corresponding author, e-mail: csxpml@163.com).

Shiyu Zhu is with the Chongqing Institue of Engineering, Chongqing, 400056, China.

standard printed characters because handwriting is more casual than the printed text. The sampling and labeling work is time-consuming and labor-intensive and coupled with the damage and serious noise in some ancient books creates great difficulties in the sampling work.

In previous researches, some methods for detecting minority nationality characters have been proposed. Xiaodong completed initial sampling work by directly segmenting characters in the Yi character library [1]. This method is simple and fast, but only very standardized printed characters are applicable to this method and the difficulty of later recognition is easier than that of handwritten characters. Xiangdong et al. first used the OTSU algorithm to binarize the Mongolian ancient books images and then used the vertical projection information of the image to locate the text columns. Finally, the connected component analysis method was used to obtain the individual Mongolian characters[2]. However, the layout of the Mongolian ancient books involved in the study is relatively neat, and the image pollution and noise are also less. Halimurati et al. used the projection method to segment rows, and columns of Uighur scanned images and to detect the baseline, then set the average threshold to segment the adhesive Uighur characters. This method can segment the standard Uighur characters accurately [3]. Ming et al. applied the method of synthesizing horizontal projection and connected component analysis to realize line and word segmentation of Uyghur text and merged the cut characters by rules so that the accuracy of Uighur character cutting reached over 99% [4]. Xiaosong et al. detected and segmented oracle inscriptions written on bone fragments using a method based on connected components [5]. In addition, there have also been some studies on text detection in handwritten Chinese and Latin documents. Xiaohui et al. classified text and non-text areas in handwritten Chinese documents by means of a multilayer perceptron and convolutional neural network model based on a conditional random field and achieved ideal classification results [9]. Yue et al. effectively detected text lines and baselines in Latin manuscript documents using a multitask full-convolution neural network [10]. Zhang et al. used the method of stroke width transformation to detect candidate characters, and then adopted k-nearest neighbor algorithm to combine candidate characters to detect strings in any direction [8]. Therefore, how to locate and extract ancient Yi characters accurately under a complex noise background is the basis of the work of recognition and digitization of ancient Yi characters.

It can be seen that character segmentation is the precondition of character digitization in the above researches. Therefore, how to accurately locate and extract the ancient Yi characters in the complex noise background is the basis of ancient Yi character identification and digitization of ancient Yi books. In this paper, a method for locating and segmenting Yi characters in ancient Yi books under a complex noise background is proposed. First, the image is preprocessed, and the foreground and background pixels are segmented. In this process, the key is to process the image noise of ancient books. Because of the damage and pollution of ancient books, there is considerable noise in the image. Therefore, we processed a number of images in ancient Yi books and compared various preprocessing methods. The detailed process of preprocessing is described in detail in Section III *A*. After image preprocessing, a binary image with less noise is obtained. Then, the residual noise and some non-text areas are removed using the method based on the connected component. This process is described in detail in Section III *B*. Finally, a method based on connected component and regressive character segmentation is used to detect single characters in ancient Yi books. The implementation of this method is described in detail in Section III *C*, Section IV presents the experimental results and analysis of character detection, and Section V is the summary of the full text. The method proposed in this paper solves the problem of separation of text and non-text areas in the ancient Yi books and realizes the detection of the ancient Yi characters, which lays a solid foundation for the recognition of ancient Yi characters and the digitization of the ancient Yi books.

II. Reasons for not Adopting Deep Learning Methods

Currently, document analysis and text segmentation mostly adopt the method of deep learning, which greatly improves detection accuracy. In particular, complex text detection in natural scenes can effectively address multidistributed and multilayout documents using deep neural network structures.

Xiang et al. proposed using fully convolutional networks (FCN) to detect texts with different directions. Text/non-text areas can be generated by FCN. Both local and global cues are considered for localizing text lines in a coarse-to-fine procedure. The FCN classifier is used to predict the centroid of each character. Finally, the entire text line can be detected [13]. Renton et al. proposed a method of dilated convolutions based on FCN. Dilated convolutions allow us to never reduce the input resolution and produce pixel-level labeling. The FCN is trained to identify X-height labeling as a text line representation, which has many advantages for text recognition [14]. FCN is also used to analyze the layout structure of handwritten documents. The FCN is trained to segment the document image into different regions and detect the centerline of each text line by classifying pixels into different categories. By supervised learning on document images with pixelwise labels, the FCN can extract discriminative features and perform pixelwise classification accurately. After pixelwise classification, the noises are further reduced, wrong segmentations are corrected, and overlapping regions are identified [9]. Some researchers have proposed end-to-end, multimodal, FCN for extracting semantic structures from document images based on FCN. The document semantic structure extraction is considered a pixelwise segmentation task, This network can produce a correspondingly sized output of any size input through effective inference and learning [15]. Pengyuan et al. proposed an end-to-end trainable neural network model for scene text recognition based on Mask R-CNN, named Mask TextSpotter. Unlike the method of implementing text detection using an end-to-end trainable deep neural network, the masked text

detector used a simple and smooth end-to-end learning process to obtain accurate text detection and recognition through semantic segmentation. In addition, it is superior to previous methods in dealing with irregularly shaped text instances, such as curved text [16]. In 2016, a novel connectionist text proposal network (CTPN) was presented. The CTPN detects a text line in a sequence of fine-scale text proposals directly in convolutional feature maps and accurately localizes text lines in natural images. CTPN overcomes some of the limitations of the previous bottom-up approaches based on character detection, and until today, the framework has been a common method for text detection in OCR systems, greatly affecting the direction of subsequent text detection algorithms [17].

It is obvious that FCN, R-CNN, Fast-RNN and other deep learning frameworks have outstanding performance on text detection; however, there are some difficulties when those methods are directly used for the detection and recognition of ancient Yi texts.

1. The methods of deep learning inevitably require labeling of datasets. For the common text structure, as well as the mainstream language text, the labeling also requires higher labor costs, but for the minority texts of the ancient Yi, there are currently few people familiar with the language. Additionally, only a small number of experts are familiar with the layout and writing standards of ancient Yi.
2. Most of the ancient literature has been retained for a long time, even several hundred years. The characters may be sticky, so it is difficult to identify them. Under these conditions, labeling data becomes difficult.
3. In the literature, pictures and texts are mixed, and it is quite different from current document layouts. Labeling is difficult for mixed arrangements of text and non-text areas, especially text areas. The method of deep learning is powerless.

Therefore, we adopt a nondeep learning method to avoid labeling the data and use the inner characteristics of ancient Yi documents to detect the text. This is the difference between our paper and mainstream text detection.

III. Algorithm Description

A. Image Preprocessing

Because most of the ancient Yi books have a long history and are influenced by various environments, they all have yellowing, wrinkles, and stains. Image filtering can suppress the noise of the target image while retaining the details of the image features, which lays the foundation for subsequent binary processing. Additionally, it is important to revitalize the ancient books and documents, which is conducive to the preservation and dissemination of ancient books and documents. Through the denoising test and analysis of a large number of images, the ability and efficiency of denoising are synthesized. Finally, the original images are processed by nonlocal mean filtering, and then an improved local adaptive threshold binarization algorithm is used to binarize the processed images in the previous step.

1) Nonlocal Mean Filtering

Nonlocal mean filtering [12] considers the self-similarity of the image, which defines similar pixels as pixels with the same neighborhood pattern and uses the information in a fixed-size window around the pixel to represent the characteristics of the pixel, which is more reliable than the similarity information obtained from the information of a single pixel.

Given a discrete noisy image $v = \{v(i) \mid i \in I\}$, where i is the pixel in the image, the pixel value after nonlocal mean filtering $NL[v](i)$, for a pixel i, is computed as a weighted average of all the pixels in the image,

$$NL[v](i) = \sum_{j \in I} w(i,j) v(j) \tag{1}$$

where the family of weights $\{w(i,j)\}_j$ depends on the similarity between the pixels i and j and satisfies the usual conditions $0 \le w(i,j) \le 1$ and $\sum_j w(i,j) = 1$.

The similarity between two pixels i and j depends on the intensified gray level vectors $v(N_i)$ and $v(N_j)$, where N_k denotes a square neighborhood of fixed size and centered at a pixel k. This similarity is measured as a decreasing function of a Gaussian weighted Euclidean distance, $\|v(N_i) - v(N_j)\|_{2,a}^2$, where $a > 0$ is the standard deviation of the Gaussian kernel. This measure is much more adapted to any additive white noise such that a noise alters the distance between windows uniformly. Indeed,

$$E\|v(N_i) - v(N_j)\|_{2,a}^2 = \|u(N_i) - u(N_j)\|_{2,a}^2 + 2\sigma^2 \tag{2}$$

where u and v are the original and noisy images, respectively, and σ^2 is the noise variance. This equality shows that, in expectation, the Euclidean distance preserves the order of similarity between pixels. Therefore, the most similar pixels to i in v are also expected to be the most similar pixels to i in u. The weights associated with the quadratic distances are defined by

$$w(i,j) = \frac{1}{Z(i)} e^{-\frac{\|v(N_i) - v(N_j)\|_{2,a}^2}{h^2}} \tag{3}$$

where $Z(i)$ is the normalizing constant,

$$Z(i) = \sum_j e^{-\frac{\|v(N_i) - v(N_j)\|_{2,a}^2}{h^2}} \tag{4}$$

and the parameter h controls the decay of the exponential function, and therefore the decay of the weights, as a function of the Euclidean distances.

For computational purposes of the NL-means algorithm, we

can restrict the search of similar windows in a larger "search window" of size $S\times S$ pixels. In all the experiments, we fixed a search window of 21×21 pixels and a similarity square neighbor N_i of 7×7 pixels because the 7×7 similarity window is large enough to be robust to noise and small enough to address details and fine structure. We observed experimentally that parameter h can take a value of $12\times\sigma$, obtaining a high visual quality solution. Figs. 1(a) and (b), respectively, show the original scanned image of ancient Yi books and the effect of the image processed by nonlocal mean filtering.

(a)

(b)

Fig. 1 (a) The original scanned image of ancient Yi books. (b) The image processed by nonlocal mean filtering.

2) Local Adaptive Threshold Binarization

To further remove the smudges in the image and highlight the outline of the text area, it is necessary to binarize the image. In this paper, an improved local adaptive threshold binarization algorithm is proposed to binarize the image. Because ancient Yi books are stored for a long time, affected by uneven illumination and serious pollution, the traditional global threshold binarization method is not satisfactory for binary processing of ancient book. Therefore, a local adaptive threshold binarization algorithm is used to binarize the grayscale image processed by nonlocal mean filtering in the previous step. In this paper, Gaussian smoothing filtering is added on the basis of the traditional local adaptive threshold algorithm [11]. The reason that Gaussian filtering is introduced is that the pixels of the real image in the space are slowly changing, so the pixel variation in the adjacent points is not obvious, but there may be a large pixel difference between any two points. In other words, there is no significant correlation between noise in space. For this reason, Gaussian filtering can reduce noise while preserving the central pixel information and filter out noise for subsequent character segmentation.

Let x, y be the horizontal and vertical coordinates of the image, respectively. $v(x,y)$ denotes the gray value of the pixel at (x,y). Take (x,y) as the center of the square neighborhood F(x,y) whose size is $(2w+1)\times(2w+1)$, where w is a positive integer greater than 0, it can make the size of the square neighborhood size an odd number. $\tilde{v}(x,y)$ denotes the gray value smoothed by Gaussian filtering at (x,y), σ is the smoothing scale, k and l are the position parameters in the square neighborhood, and $b(x,y)$ denotes the result of binarization at (x,y). The improved algorithm is described as follows:

Input. Pixel gray value of a gray image at $v(x,y)$

$$\left(0\le v(x,y)\le 255\right).$$

Output. Pixel gray value of a binary image at $b(x,y)$.

$$\left(b(x,y)=0\vee b(x,y)=255\right)$$

Step1. Calculate the threshold $T_1(x,y)$ at $v(x,y)$.

$$T_1(x,y)=0.5\times\left[\max_{-w\le k,l\le w} v(x+k,y+l)+\min_{-w\le k,l\le w} v(x+k,y+l)\right] \quad (5)$$

Step2. The point $v(x,y)$ is filtered by Gaussian smoothing in a $(2w+1)\times(2w+1)$ window.

$$\tilde{v}(x,y)=\frac{1}{(2w+1)^2}\sum_{i,j\in F(x,y)} v(i,j)\times e^{-\left\{\frac{1}{2}\left[\left(\frac{i}{\sigma}\right)^2+\left(\frac{j}{\sigma}\right)^2\right]\right\}} \quad (6)$$

Step3. Calculate the threshold $T_2(x,y)$ of postwave $\tilde{v}(x,y)$.

$$T_2(x,y)=0.5\times\left[\max_{-w\le k,l\le w} \tilde{v}(x+k,y+l)+\min_{-w\le k,l\le w} \tilde{v}(x+k,y+l)\right] (7)$$

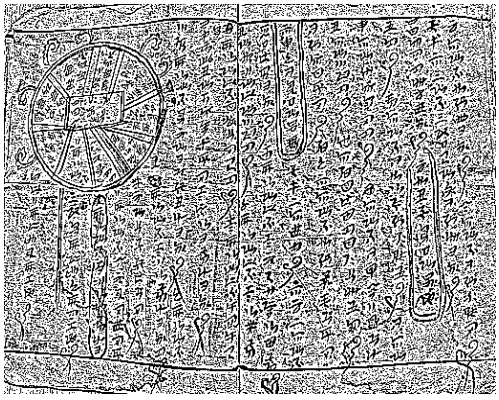
Step4. Assume $\alpha\in(0,1)$, and the point-by-point binarization of $v(x,y)$ is carried out.

$$b(x,y)=\begin{cases}0, if\ \ v(x,y)<(1-\alpha)T_1(x,y)+\alpha T_2(x,y)\\ 255, else\end{cases} \quad (8)$$

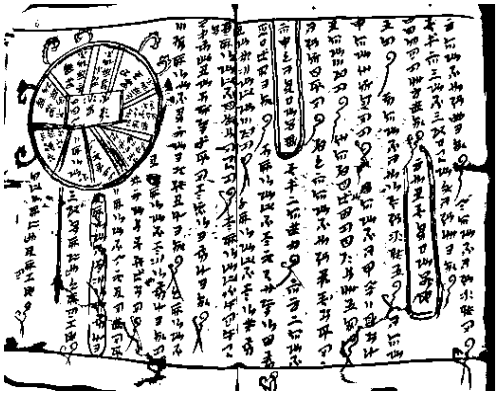
When α is 0, the algorithm is a traditional Bernsen algorithm; when $0<\alpha\le 1$, the algorithm is an improved algorithm. The w value is related to the running time of the algorithm and the generation of image artifacts: the larger w is, the longer the running time of the algorithm, and the fewer artifacts produced, and vice versa. The k and l parameters control the size of the operation window, which is an important parameter of the time consumed by the Bernsen algorithm, assuming that k is the length in the horizontal direction and l is the length in the vertical direction. If $k,l\neq 0$, the algorithm is based on a grid scan. If $k=0\vee l=0$, the algorithm is based on a line scan. Although grid scan can reduce the noise of binary images, it produces more artifacts, and the running time of the algorithm is greatly increased. Since the line scan only needs to scan from one direction, although more noise is generated, the shadow caused by the uneven illumination can be effectively removed, and the retention of the character detail information in the image is better than the grid scan. The improved algorithm proposed in this paper performs Gaussian smoothing on grayscale images while performing a line scan in the operation window.

In the process of the traditional Bernsen algorithm based on a line scan, the values of k and l always have a value of 0, so there is only one parameter, which is uniformly represented by w. The value of w is generally determined by the pixel size of the target information in the image. According to the analysis of the experimental image results, the value of w is generally between the minimum stroke width and the maximum stroke width of the character. At this time, the binarization effect on the text region is the best. Assuming that the target information in the image is the text region in the image, different values of w are selected to process the grayscale image. As shown in Fig. 2, when $w=1$, a large number of artifacts are generated in the image; when $w=25$, considerable noise is produced, causing loss of text area information. In this example, the target information is ancient Yi characters, their minimum stroke width is 6 pixels, and the maximum stroke width is 13 pixels. Therefore, the value of w should be chosen as 10, which keeps the information of the text area well and does not increase the running time of the algorithm.

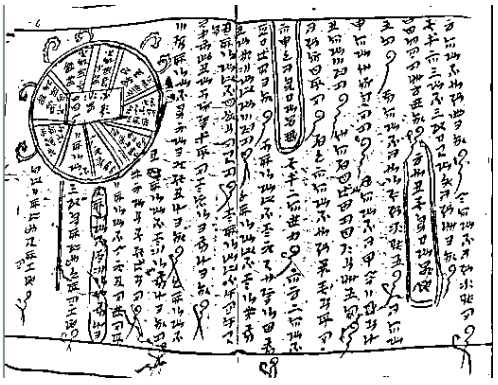
The value of α determines the relationship between noise smoothing and image target information retention. Adjusting the value of α can not only make the image adapt well to uneven illumination but also effectively remove noise in the image after smoothing. The larger the value of α, the more obvious the filtering effect, but the target information in the image is also filtered, and vice versa. Fig. 3 shows a comparison of image processing results when $w=10$ and α takes different values. When the value of α is 0.3, the image processing result not only saves the information of the text area well but also processes the noise more ideally.



(a) $w=1$



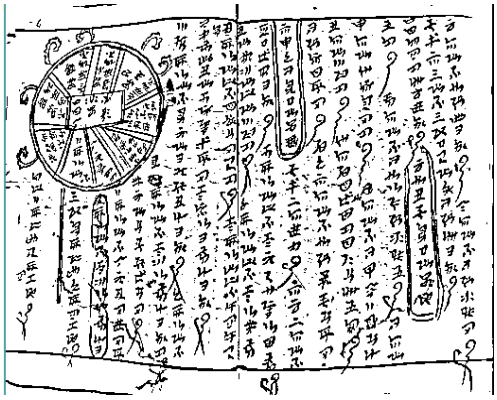
(b) $w=25$



(c) $w=10$

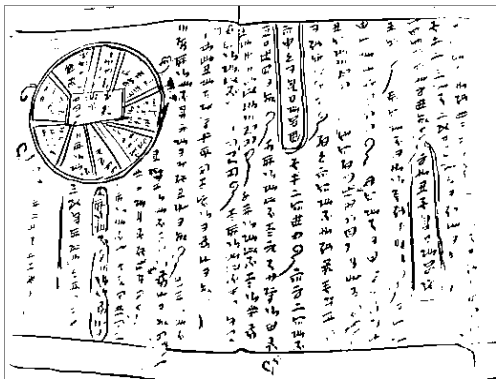
Fig. 2 Influence of the value of w on the processing effect of the traditional Bernsen algorithm

Due to uneven illumination and serious pollution, the traditional Bernese algorithm cannot perform good binarization of ancient books, but the improved algorithm in this paper can better adapt to uneven illumination and serious pollution.



(a) α=0

(b) α=0.3



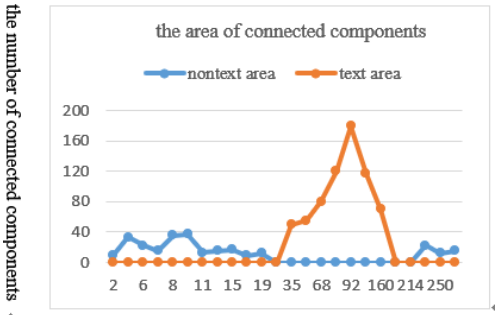
(c) α=0.7

Fig. 3 Comparison of image processing results when $w=10$ and α takes different values

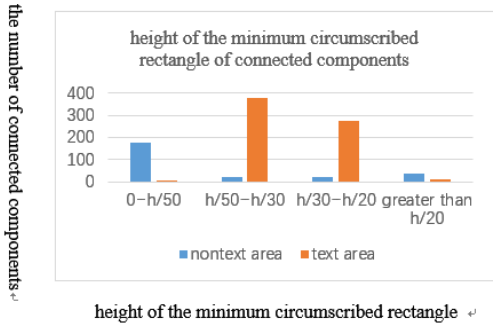
B. Nontext Removal Based on Connected Components

It can be observed that after preprocessing, there are still many nontext areas in the image (segment lines, punctuation marks, and picture decorations). Therefore, to further detect and segment the Yi characters, it is necessary to further remove these nontext areas. In [6], the author used the method based on connected components to detect the text area in the complex background image (house numbers, indicators, and advertising slogans), and obtained the ideal accuracy and recall rate. In this paper, based on the connected component method, the nontext area in the scanned image of the ancient Yi books in the complex noise background is effectively extracted and eliminated. The specific steps are as follows.

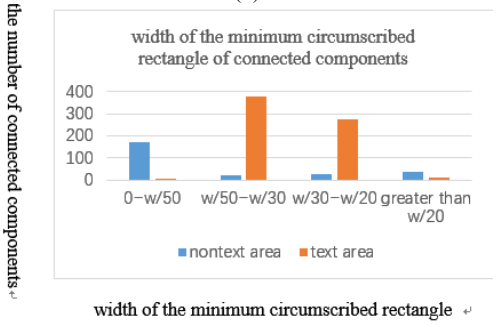
To analyze the related features of the connected component for text areas and nontext areas, we selected 672 text areas of single character and 258 nontext areas from 32 sample images of ancient Yi books to analyze their features of connected component, including the area of connected components; height, width and aspect ratio of the minimum circumscribed rectangle of connected components; and density of black pixels in the connected component. The statistical results are shown in Fig. 4.



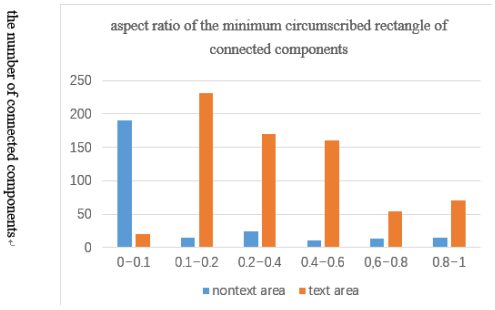
(a)



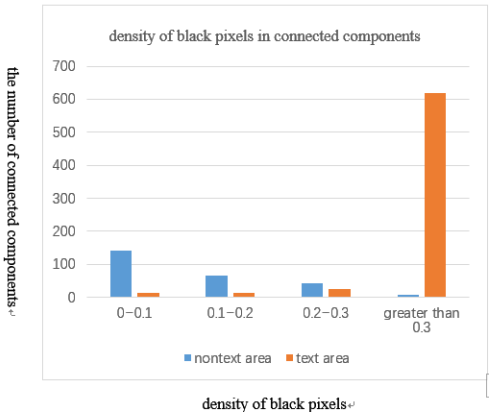
(b)



(c)



(d)



(e)

Fig. 4 Feature of connected components

Input: Binary image of ancient Yi books
Output: Results after removal of nontext areas

Step1. Extracting the small noise in the image. Since the characters in the ancient Yi books are soft pen handwriting, some small burrs exist at the edge of the character during the writing process. These small noises must be removed first. After the analysis and comparison of multiple images, it is found from Fig. 4 (a) that the area of the connected components for most nontext areas is less than 20 pixels. Therefore, we mark the connected component with an area of less than 20 pixels as small noise. Among them, the calculation method of the connected component area is the number of pixels in the minimum circumscribed rectangle of the connected component. Let x, y be the horizontal and vertical coordinates of the image, respectively. The function $f(x,y)$ denotes whether the pixel (x,y) in the binary image is the foreground pixel, as follows:

$$f(x,y)=\begin{cases}0, & \text{background}\\ 1, & \text{foreground}\end{cases} \tag{9}$$

If $f(x,y)=0$, then accumulate it as (10):

$$area=\sum_{x=0}^{\hat{w}-1}\sum_{y=0}^{\hat{h}-1} f(x,y) \tag{10}$$

where $\hat{w}, \hat{h}$ represent the width and height of the minimum circumscribed rectangle of the connected component, respectively.

Step2. Extract the larger segmentation lines, punctuation marks and picture decoration in the image. It is obvious to draw conclusions from Figs. 4 (b) and (c) that the height of the minimum circumscribed rectangle of the connected component for most text areas is less than $h/20$ and the width is less than $w/20$. Additionally, it is found from Fig. 4 (d) that the aspect ratio of the minimum circumscribed rectangle of the connected component for most nontext areas is less than 0.1, so we define a connected component as a nontext area that satisfies the following characteristics.

$$\hat{w} > w/20 \text{ or } \hat{h}>h/20 \tag{11}$$

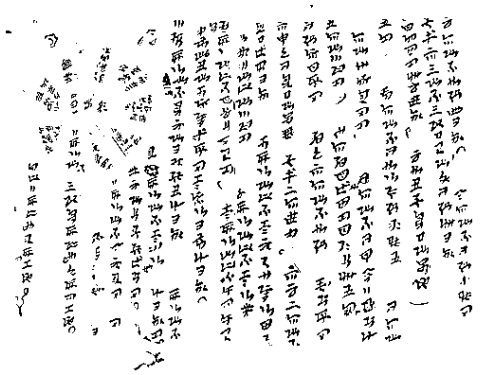
$$\frac{\min(\hat{w},\hat{h})}{\max(\hat{w},\hat{h})} < 0.1 \tag{12}$$

where w and h represent the width and height of the binary image, respectively, and $\hat{w}$ and $\hat{h}$ represent the width and height of the minimum circumscribed rectangle of the connected component, respectively.

Step3. Remove the nontext areas extracted in the previous two steps.

The test results show that this method can remove the noise and nontext areas in most ancient books. As shown in Figs. 5 and 6, we can see that the method based on the connected component has a good effect on the separation of text and nontext areas.

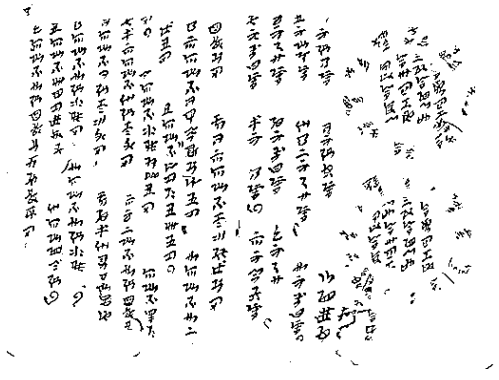
(a)



(b)

Fig. 5 (a) The original image of ancient books; (b) nontext filtered image

(a)



(b)

Fig. 6 (a) The original image of ancient books; (b) nontext filtered image

C. Detection of Single Characters in Ancient Yi Books

The processing of the first few steps can effectively remove most of the noise in the images of ancient books, and the separation of text and nontext area can be realized. On this basis, we further researched the detection of single characters in ancient Yi books. Because the writing layout of ancient Yi books is complex and disorderly, only the vertical direction has a relatively orderly arrangement, so only a single text detection method cannot achieve better detection accuracy. This paper proposes a method based on the combination of the connected components and regressive word segmentation to detect single characters in ancient books. The specific implementation steps are as follows:

Input: Binary image after removing nontext areas

Output: Detection results of single characters

Step1. When the characters are detected by the method based on the connected components, the characters whose writing structure is the left-right structure or the upper-lower structure are oversegmented, so the morphological corrosion operation is used to process the text area in the image before detection so that the character strokes have a small degree of adhesion.

Because the image resolution of the ancient Yi books processed in this paper is low, the image is convoluted with a small morphological structure. This paper selects a 1x5 rectangular structure.

Step2. The text area is detected by the method of the connected component method. After observing and analyzing Figs. 4 (b) and (c), it is found that the text areas in the image have the following characteristics:

$$w/50 < \hat{w} < w/20 \ and \ \mathrm{h}/50 < \hat{h} < h/20 \qquad (13)$$

where w and h represent the width and height of the image, respectively; $\hat{w}$ and $\hat{h}$ represent the width and height, respectively, of the minimum circumscribed rectangle of the connected component. In addition, some connected components of nontext areas have similar features with text areas, which cannot be distinguished only by the above features, but the density of black pixels in nontext areas is much lower than that in text areas. By observing Fig. 4 (e), it is found that the density of black pixels in the connected component for most text areas is greater than 0.3, so they can be distinguished and removed according to (14):

$$Den = \frac{N}{\hat{h} \times \hat{w}} \geq 0.3 \qquad (14)$$

where Den represents the density of black pixels in the rectangular frame, N represents the total number of black pixels in the rectangular frame, and $\hat{h}$ and $\hat{w}$ represent the height and width, respectively, of the minimum circumscribed rectangle of the connected component.

After many tests, some large rectangular frames will contain many small rectangular frames in the process of detecting the connected component. Therefore, the connected component should be merged, and the small rectangular frames should be removed. The parameters of connected component 1 are top 1, bottom 1, left 1, and right 1, and the parameters of connected component 2 are top 2, bottom 2, left 2, and right 2, where the parameters top and bottom represent the minimum and maximum values of the minimum circumscribed rectangle of

the connected component in the y-axis direction. Similarly, the left and right represent the minimum and maximum values, respectively, in the x-axis direction. Then, the connected component 1 includes connected component 2, which can be determined according to (15):

$$\begin{cases} top1 \le top2 \\ bottom1 \ge bottom2 \\ left1 \le left2 \\ right1 \ge right2 \end{cases} \tag{15}$$

Through the above rules, the text area is initially detected, but since the previous corrosion operation will cause some characters with short writing distances or characters that are originally stuck together to be recognized as single characters, then it is needed to perform secondary segmentation within these connected components.

Step3. The image projection method based on regressive word segmentation is used for secondary segmentation. The projection method accumulates pixel values in a certain direction of the image. For example, the projection value of images containing characters in the vertical and horizontal directions are p_x and p_y, respectively. Let x, y be the horizontal and vertical coordinates of the image, respectively, and the function $f(x,y)$ denote whether the pixel (x,y) in the binary image is the foreground pixel, as follows:

$$f(x,y) = \begin{cases} 0, \text{ background} \\ 1, \text{ foreground} \end{cases} \tag{16}$$

If $f(x,y)$=0, then accumulate it as (17):

$$p_x = \sum_y f(x,y), p_y = \sum_x f(x,y) \tag{17}$$

where p_x and p_y represent the cumulative values of foreground pixels along the x-axis and y-axis, respectively.

The method proposed in this paper is to perform a secondary segmentation of connected components that were not fully segmented before. For the need to perform secondary segmentation on a connected component, we specify that the maximum word width and maximum word height of the characters are $w/20$ and $h/20$, respectively, wherein, w and h represent the width and height of the entire image, respectively. If there is a connected component whose width or height is greater than the threshold, the connected component needs to conduct the secondary segmentation. Furthermore, it is only necessary to judge the width and height of the minimum circumscribed rectangle of the connected component to determine whether the connected component is vertically or horizontally projected. If the width of the minimum circumscribed rectangle of the connected component is greater than its height, the connected component is vertically projected, and vice versa. For the vertical projection example, if the foreground pixel value in the vertical direction is accumulated to 0, it can be used as the end or start of a character, the horizontal coordinate value of the line where the column is located can be obtained, and the line can be used as a character dividing line. Similarly, horizontal projection can be performed.

Since the characters in the ancient books are mostly handwritten, it is inevitable that there will be some adjacent characters with the adhesion of the strokes. In this case, the traditional projection method cannot completely segment the adhesive characters. This paper adopts the regressive character segmentation method to segment the adhesive characters. The following is an example of column segmentation to introduce the maximum width regressive character segmentation algorithm flow.

Let $L(x,y)$ be the lattice of a minimum circumscribed rectangle of connected components, wherein, x, y represent the horizontal and vertical coordinates, respectively, within the minimum circumscribed rectangle of the connected component. W_M is the maximum width of the character, (the value of W_M in this experiment is $w/20$, wherein w is the width of the entire image.), and the regressive range is represented by d (in step 3.3 of the experiment, the value of d is $w/60$) and assume that the starting position of the j th character is j_A, as shown in Fig. 7.

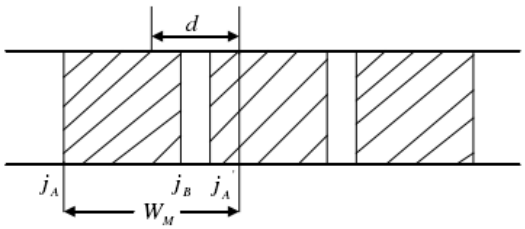


Fig. 7 Maximum width regressive segmentation

The flow of the regressive character segmentation method is described as follows:

- Step 3.1. Calculate the point of the first $\sum_{y=0}^{\hat{h}} L(x,y) = 0$ between $j_A \le x \le j_A + W_M$, which is set to j_B, and cut out the image between j_A and j_B, wherein $\hat{h}$ represents the height of the minimum circumscribed rectangle of the connected component.
- Step 3.2. If $j_B - j_A < \delta$ (δ is a constant, in this experiment, δ takes the minimum word width $w/50$), which is considered to be interference noise and negligible; otherwise, transfer to step 3.4.
- Step 3.3. Find the minimum value j_B of $\sum_{y=0}^{\hat{h}} L(x,y)$ in the range of $j_A + W_M - d \le x \le j_A + W_M$.
- Step 3.4. Make a vertical line from j_B as the dividing line

of the character. The width of the j th word is $j_B - j_A$.

- Step 3.5. Calculate $\sum_{y=0}^{\hat{h}} L(x,y)$ from j_B, when the value is not 0 (set to be j_A) and $j_A > j_B$, j_A is the left boundary of the $j+1$ th character, and then repeat the above steps.

IV. Experimental Results and Analysis of Character Detection

The scanned pictures of ancient books tested in this paper are provided by the Research Institute of Yi Nationality Studies, Guizhou University of Engineering Science, Bijie City, Guizhou Province. Among 3,052 scanned pictures, 47 pictures with complex backgrounds, high noise, and the most representative writing styles of ancient Yi characters were selected. We use the accuracy and recall rate defined by ICDAR2005 Robust Reading Assessment [7] to evaluate the performance of text area detection. The experimental environment of this paper is as follows: Windows operating system (Windows 10 Enterprise Edition), Intel (R) Core (TM) i7-7700 processor, 3.60 GHz main frequency, 8 GB memory, NVIDA GeForce GT710 graphics card, PyCharm 1.4 editor.

The accuracy rate is defined as the ratio of the number of text rectangles accurately retrieved to the number of all text rectangles retrieved, and the recall rate is defined as the ratio of the number of text rectangles accurately retrieved to the number of text rectangles that need to be accurately retrieved. Assume that the number of text rectangles accurately retrieved is m, the number of all text rectangles retrieved is m_a, and the number of text rectangles that need to be accurately retrieved is m_b, the formula for the accuracy rate pre and the formula for the recall rate rec are expressed as shown in (18):

$$\begin{aligned} pre &= m/m_a \\ rec &= m/m_b \end{aligned} \tag{18}$$

However, since the output text rectangle does not exactly coincide with the standard text rectangle, the ICDAR2005 robust evaluation reading team uses a matching value (the degree of matching between the two text rectangles) so that it can more accurately describe the accuracy of the positioning, and the matching value is defined as follows.

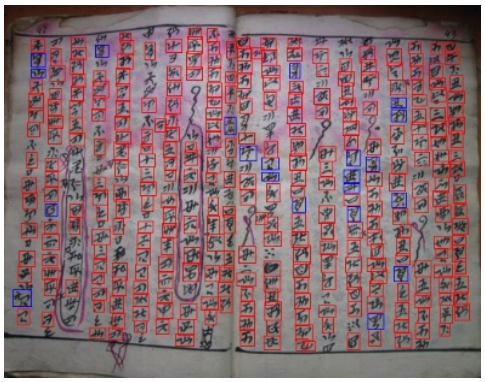
As shown in Fig. 8, in the game standard of ICDAR2005 [7], it is stipulated that a standard text rectangle is $R1$, and the rectangle obtained by the contestant is $R2$, the expression of the number of text rectangles accurately retrieved m, is shown in (19):

$$m = \frac{a_{R1 \cap R2}}{a_{R1 \cup R2}} \tag{19}$$

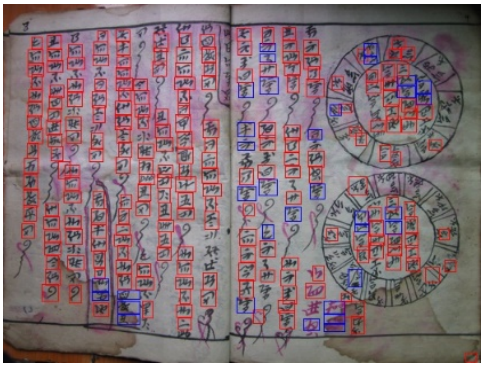
where a_R is the area of the rectangular frame R.

Fig. 8 When two rectangular frames are coincident, the calculation method of parameter m

Experiments show that the proposed method in this paper can segregate the text area and the complex background well and achieve high accuracy and recall rate in single character detection. The results of character detection in the experiment are shown in Fig. 9, where the red frame is marked as the single character detected by only using the method based on the connected component, and the blue frame is marked as the single character detected by using the regressive character segmentation projection in the minimum circumscribed rectangle of connected component. The detection results show that the method proposed in this paper combines the advantages of the method based on connected components and the method based on regressive character segmentation, which can effectively solve the problem of overlapping character regions and stroke adhesion. It can detect most characters in ancient book images with serious pollution and noise more accurately.



(a)



(b)

Fig. 9 Detection results of single characters

Table I shows the comparison between different methods on our dataset. We compare the proposed method with several traditional methods:

- The method based on connected components.
- The method based on traditional projection.
- The combining method based on traditional projection and connected components.
- The method based on maximally stable extremal region (MSER).
- The method based on stroke width transform (SWT).
- The combining the method based on MSER and SWT.

TABLE I
COMPARISON OF OVERALL PERFORMANCE BETWEEN THE PROPOSED METHOD AND OTHER TRADITIONAL DETECTION METHODS

Methods	accuracy rate/%	recall rate/%	average time /s
Method proposed in this paper	89	77	0.92
Only using method based on connected components [5]	73	65	0.74
Only using method based on traditional projection [3]	68	61	0.37
Method based on connected component+ method based on traditional projection [2]	81	70	0.89
Method based on maximally stable extremal region [18]	85	71	1.13
Method based on stroke width transform [19]	81	73	1.36
Method based on MSER and SWT [20]	83	76	1.58

It can be found by comparing the results of seven different detection methods that the detection time of the method based on the traditional projection is less, but the accuracy and recall rate are lower, mainly due to the disorder of the writing layout of ancient Yi books and the overlap of more character areas. The method based on connected components can deal with the overlap of character regions better, but it cannot achieve better detection results for the adhesive characters. After combining the two traditional methods, the accuracy and recall rate are improved to a certain extent. The method based on MSER and SWT achieves good results, but since many nontext areas and text areas have similar stroke widths, the accuracy of detection is reduced, and the detection time is longer. With the method proposed in this paper, the overlap of character areas and the adhesion of characters can be better handled. The accuracy and recall rate achieve the best results.

The main reasons why the accuracy and recall rate of the detection method used in this paper are lower than the ideal value are as follows: (1) The distance between the character and the dividing line in the ancient books is too close, and the character writing is relatively compact, which leads to the mistakes of removing them from the nontext area. (2) When a few characters are written, their strokes are too scattered, and they are mistakenly excluded during the detection of connected components. (3) The mixed layout of pictures and texts in ancient books is serious. A large number of pictures are interspersed between characters, resulting in a decrease in the accuracy of detection. As shown in Fig. 10, such ancient books with seriously mixed layouts of images and texts bring great difficulty to character detection.

(a)

(b)

Fig. 10 The seriously mixed layout of pictures and texts in ancient books

V. CONCLUSIONS

In this paper, a method for preprocessing and character detection of scanned images of ancient Yi characters in a complex noise background is implemented. First, the original image is preprocessed by nonlocal mean filtering and an improved local adaptive threshold binarization algorithm. Second, the method based on the connected component is used to remove the nontext areas. Finally, the single Yi character is detected by the method based on connected components and regressive character segmentation. Through the processing of the above steps, the experimental results show that the method proposed in this paper can achieve the highest accuracy and recall rate compared with traditional detection methods, but the effect of character detection in images with serious pictures-text mixed layout needs to be improved. How to improve the detection performance to recognize a single Yi

character and detect multidirection texts in a more complex background will be the main task in the next step.

ACKNOWLEDGMENT

This work is supported by the National Social Science Fund of China (19BYY171), the Open Projects Program of National Laboratory of Pattern Recognition, China Postdoctoral Science Foundation (2015M580765), and Chongqing Postdoctoral Science Foundation (Xm2016041), the Fundamental Research Funds for the Central Universities, China (XDJK2018B020, XDJK2019D017), Chongqing Natural Science Foundation (cstc2019jcyj-msxmX0130).

REFERENCES

[1] Xiaodong Jia, Wendong Gong and Jie Yuan, "Handwritten Yi character recognition with Density-based clustering algorithm and convolutional neural network," in Proceedings of CSE and EUC 2017, GuangZhou, China, 2017, pp. 337-341.
[2] Xiangdong Su and Guanglai Gao, "A knowledge-based recognition system for historical Mongolian documents," International Journal on Document Analysis and Recognition, vol. 19, pp. 221-235, 2016.
[3] Halmurat and Aziguli, "Research and development of a multifont printed uyghur character recognition system," Chinese Journal of Computers, vol. 27, pp. 1480-1482, 2002.
[4] Jianming Jin, Xiaoqing Ding, Liangrui Peng and Hua Wang, "Printed uyghur texts segmentation," Journal of Chinese Information Processing, vol. 18, pp. 76-83, 2004.
[5] Xiaosong Shi, Yongjie Huang and Yongge Liu, "Text on Oracle rubbing segmentation method based on connected domain," in Proceedings of IEEE IMCEC 2016, AnYang, China, 2016, pp. 414-418.
[6] Jinliang Yao, Lubin Weng and Xiaohua Wang, "A text region location method based on connected component," PR&AI, vol. 25, pp. 325-331, 2012.
[7] S. M. Lucas, "ICDAR 2005 text locating competition results," in Proceedings of ICDAR 2005, Seoul, korea, 2005, pp. 80-84.
[8] Y. Zhang, J. Lai and P. Yuen, "Text string detection for loosely constructed characters with arbitrary orientations," Neurocomputing, vol. 168, pp. 970-978, 2015.
[9] Yue Xu, Fei Yin, Zhaoxiang Zhang and Chenglin Liu, "Multi-task layout analysis for historical handwritten documents using fully convolutional networks," in Proceedings of IJCAI 2018, Stockholm, Sweden, 2018, pp. 1057-1063.
[10] Xiaohui Li, Fei Yin, Chenglin Liu, "Printed/Handwritten texts and graphics separation in complex documents using conditional random fields," in Proceedings of the 13th International Workshop on Document Analysis Systems. Vienna, Austria, 2018, pp. 145-150.
[11] J.Bensen, "Dynamic thresholding of grey-level images," in Proceedings of the 8th ICPR, Paros, France, 1986, pp. 1251-1255.
[12] A. Buades, B. Coll , J. M. Morel, "A non-local algorithm for image denoising," in Proceedings of CVPR 2005, San Diego, USA, 2005, pp. 60-65.
[13] Zheng Zhang, Chengquan Zhang, Wei Shen, et al, "Muti-oriented text detection with fully covolutional networks," in Proceedings of CVPR 2016, Las Vegas, USA, 2016, pp. 1-9.
[14] G. Renton, Y. Soullard, C. Chatelain, et al. "Fully convolutional network with dilated convolutions for handwritten text line segmentation," International Journal on Document Analysis and Recognition, vol. 21, pp. 177-186, 2018.
[15] J. Long, E. Shelhamer, T. Darrell, "Fully convolutionalnetworks for semantic segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, pp. 640-651, 2014.
[16] P. Lyu, M. Liao, C. Yao, et al. "Mask textspotter: an end-to-end trainable neural network for spotting text with arbitrary Shapes," in Proceedings of ECCV 2018, Munich, Gremany, 2018, pp. 1-17.
[17] Zhi Tian, Weilin Huang, Tong He, et al. "Detecting text in natural image with connectionist text proposal network," in Proceedings of ECCV 2016, Amsterdam, The Newtherlands, 2016, pp. 56-72.
[18] Yihua Fan, Dexiang Deng, Jia Yan, "Natural scene text detection based on maximally stable extremal region in color space," Journal of Computer Applications, vol. 38, pp. 264-269, 2018.
[19] S. Karthikeyan, V Jagadeesh, B. S. Manjunath, "Learning bottom-up text attention maps for text detection using stroke width transform," in Proceedings of IEEE International Conference on Image Processing (ICIP) 2013, Melbourne, VIC, Australia, 2013, pp. 1-5.
[20] Tongwei Lu, Renjun Liu, "Detecting text in natural scenes with multi-level MSER and SWT," in Proceedings of Ninth International Conference on Graphic and Image Processing (ICGIP) 2017, Qingdao, China, 2017, vol. 10615.
[21] Yan Wu, Jincheng Yin, "Guizhou Yi language information technology research overview," China Information Technology, vol. 8, pp. 63-65, 2017.
[22] Chengping Wang, "Design and implementation of ancient Yi input method based on Yunnan, Sichuan, Guizhou and Guilin Yi Character Sets," Computer and Information Technology, vol. 20, pp. 28-30, 2012.

Xu Han received a bachelor's degree in software engineering from Taiyuan University of Technology in 2017. Currently pursuing a master's degree in software engineering at Southwest University. His research interests include deep learning and image processing.

Shanxiong Chen received the Ph.D. degree in computer science and technology from the College of Computer, ChongQing University. He was a Visiting Scholar with South Australia University in 2014. He is currently an Associate Professor with the College of Computer and Information Science, Southwest University, China. He currently holds the post-doctoral position with Southwest University. His research interests include machine learning, pattern recogniton , and data mining.

Shiyu Zhu received the MA.Eng. degree in electronics and communication engineering from the College of Communication, ChongQing University.He is currently an Associate Professor with the College of Computer, Chongqing Engineering College, China.His research interests include machine learning, network security, and data mining.

Xiaoyu Lin received a bachelor's degree in computer science and technology from Yangtze Normal University in 2018. Currently pursuing a master's degree in software engineering at Southwest University. Her research interests include deep learning and image processing.