

MBE, 17(4): 3203–3223. DOI: 10.3934/mbe.2020182 Received: 19 December 2019 Accepted: 12 April 2020 Published: 23 April 2020

http://www.aimspress.com/journal/MBE

## Research article

# Robust table recognition for printed document images

# Qiaokang Liang<sup>1,2,\*</sup>, Jianzhong Peng<sup>1,2</sup>, Zhengwei Li<sup>3</sup>, Daqi Xie<sup>1,2</sup>, Wei Sun<sup>1,2</sup>, Yaonan Wang<sup>1,2</sup> and Dan Zhang<sup>4,\*</sup>

- <sup>1</sup> College of Electrical and Information Engineering, Hunan University, Changsha 410082, China
- <sup>2</sup> National Engineering Laboratory for Robot Vision Perception and Control, Hunan University, Changsha 410082, China;
- <sup>3</sup> Department of Mechanical Engineering, University of Alberta, Edmonton, AB T6G 2R3, Canada
- <sup>4</sup> Department of Mechanical Engineering, York University, Toronto, ON M3J 1P3, Canada.
- \* Correspondence: Email: qiaokang@mail.ustc.edu.cn, dan.zhang@uoit.ca; Tel: 86073188822224.

**Abstract:** The recognition and analysis of tables on printed document images is a popular research field of the pattern recognition and image processing. Existing table recognition methods usually require high degree of regularity, and the robustness still needs significant improvement. This paper focuses on a robust table recognition system that mainly consists of three parts: Image preprocessing, cell location based on contour mutual exclusion, and recognition of printed Chinese characters based on deep learning network. A table recognition app has been developed based on these proposed algorithms, which can transform the captured images to editable text in real time. The effectiveness of the table recognition app has been verified by testing a dataset of 105 images. The corresponding test results show that it could well identify high-quality tables, and the recognition rate of low-quality tables with distortion and blur reaches 81%, which is considerably higher than those of the existing methods. The work in this paper could give insights into the application of the table recognition and analysis algorithms.

**Keywords:** table image recognition; binarization algorithm; deep learning; recurrent neural network; character recognition

#### 1. Introduction

Tables in documents such as product catalogues, balance sheets, and financial reports are important expressive objects that present statistical and relational information. In the past several decades, Optical Character Recognition (OCR) is widely implemented in various applications by converting printed text into editable text, such as archival literature, office automation and license plate recognition [1]. This advanced technology integrates the digital image processing, computer vision and other disciplines. The rapid development of OCR has promoted the transformation of many industries, since it can significantly save the working hours, as well as the labor costs. However, the printed document recognition remains challenging. For example, images obtained by photographing or scanning contain a lot of complicated information, i.e., tables, formulas, images, and a large number of Chinese characters.

The general document image character recognition is mainly accomplished by the following steps 2. First, we obtain the information of documents in a real scene by photographing or scanning the original paper documents stored in the form of image. Secondly, we apply the knowledge of the image to analyze the layout of the image, then separate the corresponding modules and send each module to the corresponding processor for processing. Thirdly, employ different functions of different modules in the related technology of document character recognition to distinguish and identify the characters in each section. The last two steps play a vital role in the document image recognition. Unlike other document recognition technologies, the table recognition requires not only extracting the frame and lines of the table, but also obtaining useful information contained in the table, such as numbers, characters, and formulas.

In the printed documents, the form of tables can be mainly divided into two categories. One is the mixed type document including pictures, characters, tables, etc. The other one is only composed of tables, e.g. financial statement, transcripts, and other single structured tables. The latter one is less challenging since the structure and information of the table can be directly extracted and identified after analyzing the table. For the former, it is more complicated because it has to preprocess the document image to minimize its own noise interference, extract the table parts, and use algorithms for identification and analysis. This paper focuses on the recognition of the commonly used tables of the former type formed by rectangular elements.

There are a large number of recognition approaches in the field of image processing for various recognition tasks. Ranka et al. [3] tackled the problem of table detection and retention by proposing a bi-modular approach based on structural information of tables includes bounding lines, row/column separators, spaces between columns. Experiments on a dataset of above 600 images consisting of more than 829 tables have detected 90% of the table correctly. Kasar et al. [4] presented a query-based approach to selectively extract tabular information and recognize the table structure from scanned documents. The query pattern is first transformed into an attributed relational graph and a fast graph matching technique was then used to retrieve other similar graphs from the document images. Cuevas [5] presented the Block-matching algorithm based on harmony search optimization for motion estimation which could be viewed as an optimization problem whose goal was to find the bestmatching block within a search space. The average number of search points visited by the HS-BM algorithm ranges from 9.2 to 17.3, representing 4 and 7.4% respectively in comparison to the FSA method. Sage et al. [6] proposed a generic method for end-to-end table field extraction that started with the sequence of document tokens segmented by an OCR engine. The proposed method outperformed the feedforward network with a token level recurrent neural network combining spatial and textual features.

As the initial approach in table recognition, the commonly used preprocessing algorithms generally include denoise, image binarization, tilt correction and perspective correction. In the first step, denoise can make the tables and character information in the images more prominent. The binarization algorithm used in the second step is crucial to the recognition result since it can enhance

the foreground components and weaken the background components. Based on the focused level of information, binarization algorithms can be generally categorized into the global algorithms and the local algorithms [7]. The global binarization algorithms select a single intensity threshold that separate pixels into two classes, the foreground and the background, by maximizing intra-class intensity variance for the entire grayscale image. The typical global binarization algorithms are Otsu algorithm [8] and iterative method [9]. By contrast, local binarization algorithms divide the image into small block units, and estimate different thresholds for every pixel according to the grayscale information of its neighboring pixels. Several local binarization methods have been proposed such as Niblack [10], Sauvola [11,12], and Bernsen [13] algorithms. Generally, global binarization algorithms perform well with high efficiency for typical scanned table images, while local binarization methods can deal with table images with high computational complexity. It should be mentioned that binarization algorithms have been constantly optimized to adapt to various light conditions [14]. In the third step, to address the issues caused by the tilted or deformed table images, the tilt correction algorithms or the perspective correction algorithms have been utilized respectively, such as the projection-based and Hough transform-based methods [15]. Besides, layout analysis of document images is also implemented in certain scenarios by using the top-down method and the bottom-up method [16,17].

Significant efforts also have been made to develop methods and algorithms for table recognition after the preprocessing of the table images. The extraction methods are proposed to identify the geometric structures of the tables based on the different logical relations [18]. One of the most extensively used extraction method is the projection method, which projects the table in the horizontal and vertical directions respectively and obtains the horizontal and vertical line segments. In addition, a model-based approach is formed to obtain the characteristics of the tables with the topological relationship between table cells [19], which makes the result of this approach more accurate and flexible than that of the extraction method. Methods based on formal description languages have been developed. For example, some methods utilize the Latex typesetting system as a language description module to represent the table. The Latex typesetting system describes tables by means of table description language, and the structure of the table is parsed and saved.

The early classification networks for character recognition are mainly built based on the AlexNet and ResNet networks designed by Microsoft [20,21]. With the rapid development of deep learning, especially the deep learning frameworks such as convolutional neural network (CNN) [22], it becomes possible to develop the end-to-end character recognition systems. Compared with the early classification networks, the recognition algorithms based on deep convolutional networks have strong fault tolerance and classification ability, and do not require complicated pre-processing and feature extracting, which significantly reduces the recognition complexity and obtains a higher recognition accuracy. Table recognition workflow comprises table image pre-processing, table detection, and text character recognition.

It is, therefore, the objective of this paper is to develop an efficient recognition system for table images by integrating the advanced algorithms, especially the deep learning framework. The main processing procedures of the proposed table image recognition have been outlined in Figure1 and the remainder of the paper is organized as follows. Section 2 illustrates the algorithms implemented in the proposed system to preprocess the table images. In Section 3, the approaches utilized to extract the table lines and locate the table cells have been demonstrated. After that, the deep learning framework is proposed in Section 4 to recognize the printed characters. By integrating the proposed algorithms and approaches, an android-based application for table image recognition has been developed in

Section 5 and its effectiveness has been verified with practical tests. Finally, the main conclusions are summarized in Section 6.



Figure 1. Flow chart of the overall table recognition system.

#### 2. Image preprocessing

To enhance the features of table images for further processing, a preprocessing algorithm has been proposed for table recognition with its overall procedures illustrated in Figure 2. As shown in Figure 2, the table image is firstly denoised [23], and then the grayscale image is obtained to make the table outline and content clearer. Image binarization method is optimized and its robustness is enhanced, leading to significant decrease in the amount of information that is subjected to further processing. Subsequently, edge detection on the binary image is performed to obtain peripheral contour information. Then, tilt correction and perspective correction are applied for regular and unregular rectangles, respectively.



Figure 2. Flow chart of the image pre-processing.

#### 2.1. Denoise

There are many noise reduction algorithms developed for table image preprocessing, such as mean filtering, mask smooth filtering, and median filtering. Experiments have been conducted to show their efficiencies of reducing noise in the images with the original image shown in Figure 3(a). Among these existing noise reduction algorithms, the mean filtering can be defined as

$$A(i,j) = \sum \frac{B(i,j)}{N}, \quad (i,j) \in P,$$
(1)

where B(i, j) represents the gray value of a point in the image before processing, A(i, j) represents the gray value of the point after mean filtering, N denotes the sum of pixels in the neighborhood of the point, and P stands for the set of coordinates of these points in the neighborhood. The mean filtering algorithm is easier to operate with higher processing speed. The experimental result of the mean filtering has been shown in Figure 3(b), where the gray distribution of the local pixels in the image is relatively average and the overall image looks smoother.



**Figure 3.** Denoised images generated by noise reduction algorithms: (a) The original image, (b) the result of the mean filtering, (c) the result of the mask smooth filtering, (d) the result of the median filtering.

The mask smoothing, referring to the common local smoothing algorithm, is frequently utilized to improve the overall quality of the image, which makes the brightness of the image more even. This algorithm can detect the edge in the image by solving the variance and the average value, according to the difference between the foreground and background characteristics. Then, the template is implemented to calculate the smoothing effect. The experimental result of the mask smooth filtering is shown in Figure 3(c).

The median filtering algorithm, as an effective method widely utilized in image preprocessing, processes the image with a sliding window. Its general process consists of two steps, setting the sliding window with the target pixel as the center, and replacing the target pixel value with the median of all the pixels in the window. The image after pre-processing with median filtering algorithm is illustrated in Figure 3(d) for comparison.

It can be observed from Figure 3 that the mean filtered image is smoother, but it is easy to lose information and the filtering effect on salt and pepper noise is limited. Compared with mean filtering algorithm, the smooth and median filtering algorithms can better reduce the noise in the image. However, the calculation time of smooth filtering is much higher than that of median filtering. Based on the comparison involving the above characteristics, the median filtering with the  $5 \times 5$  template will be utilized to process the table image.

#### 2.2. Binarization

Binarization algorithms, which can enhance the foreground component and weaken the background component, also play a vital role in table image preprocessing. The detection accuracy of edge information and characters in table recognition depends on the effect of binarization. After binarization, the table images retain only the main information such as table lines and text characters. Many global and local binarization algorithms have been proposed, such as the Otsu algorithm and the Bernsen algorithm. Among these binarization algorithms, the Sauvola algorithm is a classical local binarization algorithm proposed for images suffering from poor or uneven illumination. Specifically, this algorithm takes the local mean as the benchmark, makes a fine-tuning according to the standard

deviation, and then uses the integral graph method. It takes a sliding window with the size is  $w^*w$  and solves the threshold T(x, y) in the window. Accordingly, this algorithm can be described as

$$T(x,y) = m(x,y) \left[ 1 + k \left( \frac{s(x,y)}{R} - 1 \right) \right],$$
(2)

with R, m(x, y) and s(x, y) being the maximum standard deviation of gray scale, the average gray value, the standard deviation in the sliding window, respectively. Here, k is a correction factor ranging from 0 to 1. It should be mentioned that the value 128 is usually assigned to R during the calculation.

The Sauvola algorithm can deal with the images suffering from uneven illumination, but it is not robust enough for low-contrast areas and often results in loss of detail. In this paper, an improved robust binarization algorithm based on the Sauvola algorithm is proposed for degraded table images with low contrast, uniform background, and uneven illumination. To enhance the robustness of the contrast region and reduce the loss of detail in the low contrast region, m(x, y) representing the mean of the pixel values within the window in the Sauvola algorithm, is replaced by the geometric mean of m(x, y) and maximum value max(x, y) in the window, which can be illustrated as

$$m'(x, y) = \sqrt{m(x, y) * \max(x, y)}$$
 (3)

For the table image under uneven illumination in Figure 4(a), the Otsu algorithm has poor performance, and many black blocks appear in the uneven portion, which causes a greater impact on the subsequent recognition, which could be seen in Figure 4(b). The Sauvola algorithm and the proposed algorithm have better performance in areas under uneven illumination. However, as illustrated in Figure 4(c), the Sauvola algorithm is prone to losing details in processing areas with low contrast. It can be clearly observed from Figure 4(d) that the proposed algorithm performs better under uneven illumination with stronger robustness and less information loss in low contrast regions.



**Figure 4.** Binarization results by different algorithms: (a) Original image, (b) binarization result with the Otsu algorithm, (c) binarization result with the Sauvola algorithm, and (d) binarization with the proposed algorithm.

#### 2.3. Correction

The table images obtained could be tilted and deformed due to the placement of the target or the angle of the camera. In what follows, it is crucial to implement image correction in table image preprocessing. Before correcting the table image, whether the table image constitutes a regular rectangle should be firstly determined. After that, the distances between the center and the four endpoints in the table image are calculated and then used as a criterion for correction. Common correction methods include tilt correction [24], correction method based on projection [25], and correction method based on contour extraction [26].



Figure 5. Tilt correction: (a) Original image, (b) corrected image.

In this paper, a tilt correction method based on the modified Hough transform is adopted when the table in the image is regular rectangle. Table characteristics such as horizontal and vertical lines are obtained by the modified Hough transform. Then, the detected longest line segment is used to calculate tilt angle, which can be obtained according to the following relation,

$$\beta = \begin{cases} \alpha, & -\frac{\pi}{4} + k\pi < \alpha \le \frac{\pi}{4} + k\pi \\ \alpha - \frac{\pi}{2}, & \frac{\pi}{4} + k\pi < \alpha \le \frac{3\pi}{4} + k\pi \end{cases}$$
(4)

where  $\alpha$  represents the tilt angle of the image and  $\beta$  refers to the tilt angle of the line segment. The original image and the image processed by tilt correction have been shown in Figure 5(a) and (b) respectively, which show the effectiveness of the proposed method.

If the table in the image is not a regular rectangle, perspective correction technology could be implemented to process this image [27]. Perspective correction projects the original image onto a new visual plane with the perspective transformation [28]. Specifically, the coordinate system (u, v) in the original image is converted to a new coordinate system (x, y). The common transformation formula for the perspective transformation can be generally expressed as

$$\begin{bmatrix} x' & y' & w' \end{bmatrix} = \begin{bmatrix} u & v & w \end{bmatrix} * \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix},$$
(5)

where x = x'/w', y = y'/w'. Also, the transformed matrix can be described as

$$Transform = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} T_1 & T_2 \\ T_3 & a_{33} \end{bmatrix},$$
(6)

where  $T_1$  indicates that the image is linearly transformed,  $T_2$  is used to produce a perspective transformation, and  $T_3$  represents the translation of the image. The four endpoints of the quadrilateral in the table image are selected as reference points, which are used to obtain the transformation matrix of perspective transformation. In order to show the effectiveness of perspective correction, experiments

have been performed on the original image and the perspective-corrected image respectively, as shown in Figure 6.

#### 3. Table lines extraction and cells location

#### 3.1. Table lines extraction

Common methods for detecting table lines include projection-based detection algorithms, detection algorithms based on directed single-link chains, and morphological-based detection algorithms. Among these existing methods, the morphological method [29] is one of the most widely used techniques to extract image components that contain useful information for expressing and depicting the shape of the region in the image. The subsequent recognition work can obtain the most essential shape features of the table such as boundaries and connected areas, where erosion and dilation are the most basic operations.

The erosion operation with a structure element S shrinks the subset S[w] of the image W that is congruent with the structure element to point w. The shrinking process is called erosion, which is defined as

$$A\Theta B = \left\{ w \mid S[w] \in W \right\}.$$
<sup>(7)</sup>

Erosion eliminates small-scale details such as burrs from a table image with the targets and nontargets separated by selecting structural elements with different specifications for specific situations.

Different from the erosion operation, the dilation operation is utilized to expand the subset S[w] of the image W that is congruent with the structure element S to point w. This expanding process is called dilation and can be described as

$$A \oplus B = \left\{ w \mid S[w] \cap w \neq \emptyset \right\}.$$
(8)

It should be mentioned that the dilation of an image can expand both the inner and outer boundaries of regions, supplement the cavity and connect similar objects.

In this paper, the horizontal and vertical structure elements have been selected to detect the horizontal and vertical lines of the table. The extraction algorithm uses an open operation on the binarized table image, i.e., the table image is eroded and then expanded, which can effectively eliminate edge burrs. As shown in Figure 7, it can be observed that the outlines of the tables have been extracted.



**Figure 6.** Perspective correction for table image: (a) The original gray image, (b) the corrected image.

Due to the complexity of the natural scene, the table lines detected by the morphological operation are particularly prone to uneven thickness. The thinning algorithm should be further applied to delete meaningless contour points in the image and keep only the bone points, which makes the table contours evenly distributed [30]. Refinement algorithms, such as Hilditch algorithm [31], OPTA algorithm [32], Zhang fast parallel algorithm [34], and morphological methods, have been widely utilized in image processing. This paper adopts the morphological method to refine the table image and obtains a satisfactory skeleton diagram. As shown in Figure 8, the outlines of the table lines are clear and evenly distributed.



**Figure 7.** Table line extraction: (a) The original image, (b) the binary image, (c) the extraction of vertical table lines, (d) the extraction of horizontal table lines, (e) the combined table.



Figure 8. Refinement of table lines: (a) The original table, (b) the refined table.

投资 <mark>后动</mark> 产生的现金 <sup>流</sup> 量序额	-8, 607, 666, 69	-22, 928, 266, 78	10.01
筹资 <mark>活动产生的现</mark> 金流量序额	156, 808, 819, 58	0 <b>19, 172, 0</b> 86, <b>7</b> 3	<u>50</u> . <u>85</u>
	-08, 297, 290. 67	-2,908,648,50	-357. [6
12 12 12 12 12 12 12 12 12 12 12 12 12 1	-10, 720, 564, 26	-2,034,542,26	-405. 03
<u>淨困</u> 国	-12, 500, 809, 95	-8, 152, 627, 90	-142. 51
基本每股收益的稀释每股收益	-0 <mark>.</mark> 0230	-0 0095	-142. 11

Figure 9. Connected domains of a table image.

#### 3.2. Table cells location

A method to estimate the line height of characters based on the connected domain has been proposed to solve the connected domain on the separated table illustrated in Figure 9 [34], which makes it easy to get the height and width histograms in the connected domain. Projection of the table image along the y direction is utilized to calculate the height of table cells, as shown in Figure 10.

Each table cell can be obtained based on the mutually exclusive relationship between the contours. The table image is composed of rectangular boxes, and each rectangular box is a relatively independent contour. Specifically, estimate the average line height of the characters. According to the filling characteristic of the table characters, the average line height of the characters is labelled as *hchar*, the height of a contour is *h*, and the width is *w*. The following restrictions are imposed

$$h > hchar*1.5 \text{ and } w > hchar*3.$$
(9)

Then, add the filter conditions to the contours obtained in the previous step to filter out all the rectangular contours. Iterate through all the rectangular contours and compare each contour with the rest of the contours with the method based on mutual exclusion. The main code is listed as following

```
for (size_t i = 0; i < cSize; i++)

{

Rect r1 = boundingRect(contours2[i]);

for (size_t j = 0; j < cSize; j++)

{

if (j == i)

continue;

Rect r2= boundingRect(contours2[j]);

if (r1 == (r2 & r1))

flag[j] = false;

if (r2 == (r2 & r1))

flag[i] = false;

}
```

Therefore, the rectangular contours filtered by the algorithm based on mutual exclusion can be obtained as the target cells, as shown in Figure 11.

实例	种属	重量(平均)	颜色(主)	生命周期/保质期
非洲象	动物	5t	土灰色	70年
大白鲨	动物	3.2t	灰白色	70年
苹果	植物	250g	红色	5~15天
梨	植物	300g	黄色	5~10天

Figure 10. Projection of the table image along the y direction.

			()	
实 例	种属	重量(平均)	颜色(主)	生命周期/保质期
非洲象	动物	5t	土灰色	70年
大白鲨	动物	3.2t	灰白色	70年
苹果	植物	250g	红色	5~15天
梨	植物	300g	黄色	5~10天
	· · · · · ·			

(a)

(b)

Figure 11. Locating table cells: (a) The binary image, (b) the located cells.

Locating the cells of the table image, labeling and sorting them, and then entering them into the character recognition module for identification are crucial steps in the whole system. We found through experiments that the previous image preprocessing are actually indispensable due to the fact that these tasks could affect the final accuracy of the table recognition and the robustness of the system. First, a large amount of obvious noise in the image can be filtered by denoising, which is the most basic operation and is beneficial to subsequent operations. The image binarization algorithm ensure clearer outlines and more prominent characters of table images. Different correction algorithms are utilized to obtain a relatively clear and flat table images. It is considerably conducive to applying morphological operations to position the table outline and accurately locate the table cells to improve the recognition accuracy of cell texts. In summary, image preprocessing is the basis of the table recognition system. Better preprocessing algorithms could improve the positioning rate of the table and the recognition accuracy of cell texts.

### 4. Recognition of printed characters

Recognition of printed characters is a very important step in the table recognition. Traditional methods of character recognition include the methods based on statistical pattern [35] and structural

pattern [36]. In the past several decades, deep learning has achieved great success in the field of image processing. Since the birth of the AlexNet network [20], various deep convolutional networks have continually updated the accuracy of various classification tasks. From AlexNet to ZF Net [37], VGGNet [38], GoogLeNet [39], ResNet [21] and DenseNet [40], CNNs are growing faster and faster. In addition, Long Short-Term Memory (LSTM) 41, GRU 42, hierarchical multiscale recurrent neural network (RNN) 43 and bidirectional LSTM 44, which focus on processing time series data, also have been employed in the field of speech processing and natural language processing. RNN has a memory function for the information of the past moment 41, that is, the input of the current hidden layer of RNN not only comes from the output of the input layer, but also from the output of the previous hidden layer. In this paper, the CNN+LSTM network is employed to identify characters.



Figure 12. The architecture of the network model for Recognition of printed characters.

The architecture of a convolutional neural network for identifying printed characters has been illustrated in Figure 12, where CNN adopts the classic VGG16 network. The 3rd and 4th max pooling layers of VGG16 take a  $1 \times 2$  rectangular pooling window due to the fact that the text images are mostly shorter and wider. The feature sequence of the input image obtained through the deep convolutional network is further processed by the LSTM algorithm, which is a special form of RNN and often used as a solution to sequence data. The LSTM network subtly adds the input, output and forget gates, and the self-looping weight of the network is constantly changing. According to the network structure of LSTM, the output can be obtained as following

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f),$$
(10)

$$i_t = \sigma(wW_i[h_{t-1}, x_t] + b_i)$$
 (11)

$$\hat{C}_{t} = \tanh(W_{c}[h_{t-1}, x_{t}] + b_{c}), \qquad (12)$$

$$C_{t} = f_{t} * C_{t-1} + i_{t} * \tilde{C}_{t}, \qquad (13)$$

$$O_t = \sigma(W_0[h_{t-1}, x_t] + b_0), \qquad (14)$$

$$h_t = O_t * \tan_{\rm h}(C_t), \tag{15}$$

where  $i_t$  represents the input threshold,  $f_t$  represents the forgotten threshold,  $\tilde{C}_t$  refers to the state of the cell at the previous moment,  $C_t$  refers to the state of the cell at this time,  $O_t$  represents the output threshold,  $h_t$  stands for the output at this time and  $h_{t-1}$  stands for the output at the previous moment. The symbols  $b_f$ ,  $b_i$ ,  $b_c$ , and  $b_o$  indicate the activation result of the corresponding value in the current

The large-scale dataset used in this paper was generated by the large database and image processing technology published on GitHub, which covers multiple fonts and various printed characters such as Chinese characters, English letters, numbers, and punctuation marks. A large multi-character image data is generated for training duo to the consideration that the cell size of the table and the length of the characters in the table cell are not fixed. Specifically, there are  $30000 \times 100$  images for training and 30000 images for testing. Data augmentation was performed with OpenCV 29 in order to further enhance the generalization of the training model. Specifically, data enhancement methods include text blurring, text tilting, text rotating, font stretching, adding different backgrounds and noises, stroke adhesion, and stroke break were applied. Figure 14 shows some pictures for training and testing from the augmented dataset containing the images with different character types and lengths.



Figure 13. (a) The loss curve, and (b) the accuracy curve during training.

In the training phase, the algorithm has been developed on ubuntu 16.04 with a single RTX 2080TI graphics card by using Python and Pytorch [46]. Also, the root mean square prop (RMSprop) algorithm 47 is utilized to optimize the network parameters, where the batch size is 64 and the initial learning rate is 0.0001. In the development environment of Ubuntu16.04 + Anaconda, the network model is implemented by using the development framework of pytorch and the training of the model is completed on the server of the laboratory. Figure 13(a,b) show the curve of loss reduction and accuracy curve during training respectively, which indicates that the loss of the train dataset goes to

state, respectively.

0.4 and the loss of the test dataset approaches 0.45 after the model iterates 50 epochs. At this time, the corresponding train accuracy becomes stable at 96.7% and the test accuracy is stable at 96.3%. The experiment finally obtained a depth network model with a recognition accuracy of 96.7% to identify the character data in the table.

In order to verify the validity of the proposed recognition algorithm, 200 cell text images have been selected as experimental data for the comparison with the Resnet network. As listed in Table 1, the experimental results clearly show that the recognition accuracy and speed of the CNN + LSTM algorithm used in this paper are better than those of the Resnet network.

姻缘也许在以前就已经	真的被SAP收到麾下 了	采用Google搜	刑事侦查学(0358	heopeningp
时间长了, 嚷: Con	1948年11月4日	1;if(form.	疗可用5% 硫磺霜和き	这些学校在当地招得多
年科技创新大赛(含全	220元/人,"Wh	第21题错选C的较多	种情况,从2003年	otivated;a
限公司(0751-H	earstoyour	前价格的()A、20	Female(女性)	07centLeft
学院00级2+2班一	十分俊雅。战场中的诸	这是经中央军委批准的	8. 88亿美元。把思	不管是友好的还是敌对
并嘱咐其多练习、模仿	方相比以往的硬碰硬的假	2 己的领导风格必然难以	座历史上曾深陷战争的	着:比如法语区的干酪
许多民族企业越来越意	额为6280万元)时	届黄山国际旅游节暨徽	盖18个省,我何生矣	35:48]境况难堪
凯的小怪楼就建在此地	Inspiteoft	得更为安全的另一个	原百金,人家没有想到的	5 慎大盘回调。在他心里
美女妇云的班级排到第	不传也。蘸出瓶中甘	<b>常</b> 果至一条就是自己自我	不育,记者也采访了4	移 资料图:2004年底
次股改的成功,《长城	在孩子们看来叫人兴奋	旗下, third…;	OnDemand软件	inaccurate
的议题都是不可以轻率	从而改善一下饮食习惯	市、自治区)招办批准	2不敌青岛的那场比赛	海上任何德国船只发起
餐不可叠加申请还没有	着隧道游进了地下工厂	莱之类的食品患胃癌的	正的"国宝"是"人才	制中的AMS-02号
很好的性教育教材。	晋"既见君子,报告显示	学校录取新生的主要依	通股股东每10凉了成	"哥哥,"工程"的提
曾当过家具制造工的人	平均每人掌握了3门专	本场关系到客队夺冠问	而效能则来自于单纯。	听得到的信息发布方式
真正能够获得高票房的	今日怎生区处?"行者	违者将追究其法律责任	), left='+x	吗?这不是反对肃反运
, [NIT网站200	都不满意一个下午	-CNN的新闻、电视剧	制。应当在定期报告中热	上一页[1][2][

(a)



(0)

Figure 14. (a) Sample pictures for training, and (b) sample pictures for testing.

Recognition network The number of the effective recognition		Recognition rate	Recognition time of each piece (s)
Resnet	189	94.5%	2.2
CNN + LSTM	193	96.5%	0.05

 Table 1. Comparative experimental results of different methods.

#### 5. Design of recognition system

In this paper, an android-based app has been developed by integrating the proposed recognition algorithm with its interfaces shown in Figure 15. In order to verify the recognition performance of the proposed recognition algorithm, 105 table document images obtained in the natural scene with different shapes and backgrounds have been collected. Performance comparison was carried out with the common software system on the market and some existing table detection methods, as shown in Table 2. It can be observed that the proposed recognition system and the existing recognition systems in the market both have a high recognition rate for the normal printed tables, while the former outperforms other methods on recognition rate for table images in the natural scene with the noise, blur, distortion and other conditions.

中国联通常		© IQI IIID ≠ 08:51	中国政语 54 👷 🖬	l.	₩ 🗇 IIID # 21:31		A	В	С
RX			照片			1	小组实验测出,	人在呼吸中各种	气体的体积分数
000	1 pro 1 1 1 1/2					2	气体	吸入气体	呼出气体
	HUNAN UNIVERSITY		某科学兴博术组定验	御井 人大部門内		3	X	78%	75%
			气体	吸入气体	<b>在</b> 市气体的体积分数加下来 算出与这	4	Y	21%	15%
			X	78%	75%	5	COz	0.03%	3.68%
			Y	21%	15%	6	HzO	0.06%	5.44%
			H <sub>2</sub> 0	0.03%	3.68%	7	其他	0.91%	0.88%
	X		其他	0.91%	0.88%	8			
			山诸你判断:X是	14 开始识别	填化学式)。 約气信音	9			
	拍照识别表	8				10			
	JEARCOLDSAC					11			
						12			
	批量订別。					13			
	110.00.00101-					14			
						15			
						16			
						17			
首市	→ 仕生					18			
	~10								

Figure 15. Table recognition APP.

Table 2. Performance co	mparison	between the	proposed	approach and	l the existing r	nethods.

Method	Inspection object	Description	Accuracy	Application
Fan et al. 48 (2015)	Table on PDF files	Apache PDFBox & Stanford NLP toolkit & classifiers (Naive Bayes, Logistic Regression and Support Vector Machine)	0.7948	PDF file
Gilani et al. 49 (2017)	Table on document images with varying layouts	Image transformation & deep learning	0.8629	Document & research paper & magazine
Koci et al. 50 (2017)	Table structure in spreadsheets	Heuristics-based method	0.78	Spreadsheet
Arif et al. 51 (2018)	Tabular regions from document images	Color coding or coloration & Faster R-CNN	0.8964	Document image
*****FineReader	Text document	online recognition-server	0.6850	Text document & document image
The proposed	Table and character on phone images	Image processing & CNN & RNN	0.8667	Natural or unnatural scene image taken by mobile phone

In conclusion, a novel recognition system for table images in natural backgrounds has been

proposed. In the process of table image recognition, image pre-processing methods including denoise, binarization, tilt correction, and perspective correction are first performed. Morphological operations, refinement, and a contour mutual exclusion algorithm are then utilized to locate the table. The CNN + LSTM network is employed to identify characters in cells with an accuracy of 96.5%. A table recognition APP is developed based on the proposed system with effective frames and optimal algorithms, which is more effective than other table recognition methods. Furthermore, the results of comparative experiments have demonstrated that the proposed method show higher accuracy than the existing commercial recognition system and methods. It is hoped that the proposed solution could be of significance for the table recognition. Further research will focus on improving light-weight architecture and recognition accuracy, and further applying it to more complex recognition tasks, such as the variety of tables with severe distortions, etc.

#### Acknowledgement

This work is supported in part by the National Natural Science Foundation of China (NSFC 61673163), Chang-Zhu-Tan National Indigenous Innovation Demonstration Zone Project (2017XK2102).

#### **Conflict of Interests**

The authors declare that there is no conflict of interest regarding the publication of this paper.

### References

- H. Singh, A. Sachan, A Proposed Approach for Character Recognition Using Document Analysis 1. with OCR, 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), 190-195. Available from: 2018, https://ieeexplore.ieee.org/abstract/document/8663011.
- 2. A. M. Sabu, A. S. Das, A Survey on various Optical Character Recognition Techniques, 2018 Conference on Emerging Devices and Smart Systems (ICEDSS), 2018, 152–155. Available from: https://ieeexplore.ieee.org/abstract/document/8544323.
- V. Ranka, S. Patil, S. Patni, T. Raut, K. Mehrotra, M. K. Gupta, Automatic Table Detection and 3. Retention from Scanned Document Images via Analysis of Structural Information, 2017 Fourth International Conference on Image Information Processing (ICIIP), 2017, 244–249. Available from: https://ieeexplore.ieee.org/abstract/document/8313719/.
- 4. T. Kasar, T. K. Bhowmik, A. Belaïd, Table information extraction and structure recognition using query patterns, 2015 13th International Conference on Document Analysis and Recognition(ICDAR), 2015, 1086-1090. Available from: https://ieeexplore.ieee.org/abstract/document/7333928.
- 5. E. Cuevas, Block-matching algorithm based on harmony search optimization for motion estimation, Appl. Intell., 39 (2013), 165–183.

- C. Sage, A. Aussem, H. Elghazel, V. Eglin, J. Espinas, *Recurrent Neural Network Approach for Table Field Extraction in Business Documents*, International Conference on Document Analysis and Recognition(ICDAR), 2019. Available from: https://hal.archives-ouvertes.fr/hal-02156269/.
- A. Shrivastava, D. K. Srivastava, A Review on Pixel-Based Binarization of Gray Images, Proceedings of the International Congress on Information and Communication Technology, 2016, 357–364. Available from: https://link.springer.com/chapter/10.1007/978-981-10-0755-2\_38.
- 8. A. K. Khambampati, D. Liu, S. K. Konki ; K. Y. Kim, An Automatic Detection of the ROI Using Otsu Thresholding in Nonlinear Difference EIT Imaging, *IEEE Sens. J.*, **18** (2018), 5133–5142.
- 9. M. Valizadeh, E. Kabir. Partitioning of feature space by iterative classification for degraded document image binarization, *IET image Process.*, **6** (2012), 804–812.
- 10. L. P. Saxena, Niblack's binarization method and its modifications to real-time applications: A review, *Artif. Intell. Rev.*, **51** (2019), 673–705.
- M. Kiran, I. Ahmed, N. Khan, A. G. Reddy, Chest X-ray segmentation using Sauvola thresholding and Gaussian derivatives responses, *J. Ambient Intell. Humanized Comput.*, **10** (2019), 4179– 4195.
- Z. Hadjadj, A. Meziane, Y. Cherfa, M. Cheriet, I. Setitra, *ISauvola: Improved Sauvola's Algorithm for Document Image Binarization*, International Conference on Image Analysis and Recognition, 2016, 737–745. Available from: https://link.springer.com/chapter/10.1007/978-3-319-41501-7\_82.
- L. Yang, Q. Feng. The Improvement of Bernsen Binarization Algorithm for QR Code Image, 2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS), 2018, 931–934. Available from: https://ieeexplore.ieee.org/abstract/document/8691255.
- I. Pratikakis, K. Zagoris, G. Barlas, B. Gatos, ICFHR2016 Handwritten Document Image Binarization Contest (H-DIBCO 2016), 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2016. Available from: https://ieeexplore.ieee.org/abstract/document/7814134.
- 15. O. Boudraa, W. K. Hidouci, D. Michelucci, Using skeleton and Hough transform variant to correct skew in historical documents, *Math. Comput. Simul.*, **167** (2020), 389–403.
- 16. T. A. Tran, K Oh, I. S. Na, G. S. Lee, H. J. Yang, S. H. Kim, A robust system for document layout analysis using multilevel homogeneity structure, *Expert Syst. Appl.*, **85** (2017), 99–113.
- 17. J. Ryu, H. I. Koo, N. I. Cho, Word Segmentation Method for Handwritten Documents based on Structured Learning, *IEEE Signal Process. Lett.*, **22** (2015), 1161–1165.
- A. Riad, C. Sporer, S. S. Bukhari, A. Dengel, *Classification and Information Extraction for Complex and Nested Tabular Structures in Images*, 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), 2017, 1156–1161. Available from: https://ieeexplore.ieee.org/abstract/document/8270122.
- H. T. Tran, T. A. Tran, I. S. Na, S. H. Kim, Cell decomposition for the table in document image based on analysis of texts and lines distribution, 2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN), 2016, 736–738. Available from: https://ieeexplore.ieee.org/abstract/document/7537135.
- A. Krizhevsky, I. Sutskever, G. E. Hinton, *ImageNet classification with deep convolutional neural networks*, Advances in Neural Information Processing Systems 25 (NIPS 2012), 2012, 1097–1105. Available from: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networ.

- K. He, X. Zhang, S. Ren, J. Sun, *Deep residual learning for image recognition*, In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, 770–778. Available form: http://openaccess.thecvf.com/content\_cvpr\_2016/html/He\_Deep\_Residual\_Learning\_CVPR\_20 16\_paper.html.
- 22. Y. Wei, Y. Zhao, C. Lu, S. Wei, L. Liu, Z. Zhu, et al. Cross-Modal Retrieval with CNN Visual Features: A New Baseline, *IEEE Trans. Cybern.*, **47** (2017), 449–460.
- 23. C. Tian, Y. Xu, W. Zuo, Image denoising using deep CNN with batch renormalization, *Neural Networks*, **121** (2020), 461–473.
- 24. D. Yang, H. Zhou, L. Tang, S. Chen, S. Liu, A License Plate Tilt Correction Algorithm Based on the Character Median Line Algorithm de correction d's inclinaison de plaque d's immatriculation base sur la ligne mediane du character, *Can. J. Electr. Computer Eng.*, **41** (2018), 145–150.
- Q. An, J. Shi, J. Li, F. Cai, *Elevator button recognition using auto-slant correction and projection histogram*, 2017 10 th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2017. Available from: https://ieeexplore.ieee.org/abstract/document/8302054.
- R. Baran, A. Dziech, J. Wassermann, Contour Extraction and Compression Scheme Utilizing Both the Transform and Spatial Image Domains, International Conference on Multimedia Communications, Services and Security, 1–15. Available from: https://link.springer.com/chapter/10.1007/978-3-319-69911-0\_1.
- J. Tang, H, Huang, L. Shi, Z. Chen, Y. Lu, H. Chen, *An Improved Perspective Transform for Image Distortion Correction*, 2018 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), 2018. Available from: https://ieeexplore.ieee.org/abstract/document/8448538/.
- 28. Q. Vien, H. X. Nguyen, B. Barn, X. Tran, On the Perspective Transformation for Efficient Relay Placement in Wireless Multicast Networks, *IEEE Commun. Lett.*, **19** (2015), 275–278.
- 29. A. C. Jalba, M. H. F. Wilkinson, J. B. T. M. Roerdink, Shape representation and recognition through morphological curvature scale spaces, *IEEE Trans. Image Process.*, **15** (2006), 331–341.
- 30. Y. Li, H. Zheng, Z. Yan, L. Chen. Detail preservation and feature refinement for object detection, *Neurocomputing*, **359** (2019), 209–218.
- 31. M. Naseri, S. Heidari, R. Gheibi, L. Gong, M. A. Raiji, A. Sadri, A novel quantum binary images thinning algorithm: A quantum version of the Hilditch's algorithm, *Optik*, **131** (2017), 678–686.
- C. Zhang, W. Zhong, C. Zhang, X. Qin, *Simulation Design of Improved OPTA Thinnin Algorithm*, International Conference on Mechatronics and Intelligence Roboyics (ICMIR), 2017, 105–114. Available from: https://link.springer.com/chapter/10.1007/978-3-319-70990-1\_15.
- A. K. J. Saudagar, H. V. Mohammed, OpenCV Based Implementation of Zhang-Suen Thinning Algorithm Using Java for Arabic Text Recognition, Information Systems Design and Intelligent Applications, 2016, 265–271. Available from: https://link.springer.com/chapter/10.1007/978-81-322-2757-1\_27.
- 34. X. Shi, Y. Huang, Y. Liu, *Text on Oracle rubbing segmentation method based on connected domain*, 2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), 2016: 414–418. Available from: https://ieeexplore.ieee.org/abstract/document/7867245.
- 35. Y. Sun, Z. Guo, W. Qiu, *Research on the Handwriting Character Recognition Technology Based on the Image Statistical Characteristics*, International Conference on Geo-Spatial Knowledge and

Intelligence, 2018, 13-20. Available from: https://link.springer.com/chapter/10.1007/978-981-13-0896-3 2.

- A. K. Sharma, P. Thakkar, D. M. Adhyaru, T. H. Zaveri, Handwritten Gujarati Character Recognition Using Structural Decomposition Technique, *Pattern Recognit. Image Anal.*, 29 (2019), 325–338.
- M. D. Zeiler, R. Fergus, *Visualizing and understanding convolutional networks*, European Conference on Computer Vision. Cham, Switzerland: Springer International Publishing AG, 2014, 818–833. Available from: https://link.springer.com/chapter/10.1007/978-3-319-10590-1\_53.
- 38. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556*, 2014.
- 39. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, *Going deeper with convolutions*, In Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, 1–9. Available from: https://www.cvfoundation.org/openaccess/content\_cvpr\_2015/html/Szegedy\_Going\_Deeper\_With\_2015\_CVP R\_paper.html.
- 40. G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, *Densely connected convolutional networks*, In Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, 4700–4708. Available from: http://openaccess.thecvf.com/content\_cvpr\_2017/html/Huang\_Densely\_Connected\_Convolution al\_CVPR\_2017\_paper.html.
- 41. N. K. Manaswi, Deep Learning with Applications Using Python, Springer, (2018), 115-126.
- 42. J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, *arXiv:1412.3555*, 2014.
- 43. J. Chung, S. Ahn, Y. Bengio, Hierarchical multiscale recurrent neural networks, *arXiv:1609.01704*, 2016.
- 44. G. Liu, J. Guo, Bidirectional LSTM with attention mechanism and convolutional layer for text classification, *Neurocomputing*, **337** (2019), 325–338.
- 45. Y. Bengio, P. Simard, P. Frasconi, Learning long-term dependencies with gradient descent is difficult, IEEE Trans. Neural Networks, **5** (1994), 157–166.
- 46. (CRNN) Chinese Characters Recognition, 2020. Available from: https://github.com/Sierkinhane/crnn\_chinese\_characters\_rec.
- 47. S. Ruder, *An overview of gradient descent optimization algorithms*, 2016. Available from: http://sebastianruder.com/optimizing-gradient-descent/index.html .
- 48. M. Fan, D. S. Kim, Detecting Table Region in PDF Documents Using Distant Supervision, *arXiv:1506.08891*, 2015.
- 49. A. Gilani, S. R. Qasim, I. Malik, F. Shafait, *Table Detection Using Deep Learning*, 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), 2017, 771–776. Available from: https://ieeexplore.ieee.org/abstract/document/8270062.
- E. Koci, M. Thiele, O. Romero, W. Lehner, *Table Identification and Reconstruction in Spreadsheets*, International Conference on Advanced Information Systems Engineering (CAiSE), 2017, 527–541, Available from: https://link.springer.com/chapter/10.1007/978-3-319-59536-8\_33.

51. S. Arif, F. Shafait, *Table Detection in Document Images using Foreground and Background Features*, Digital Image Computing: Techniques and Applications (DICTA), 2018. Available from: https://ieeexplore.ieee.org/abstract/document/8615795.



©2020 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0)