Scholar Publishing Group
International Journal of Neural Network
https://doi.org/10.38007/NN.2020.010101
ISSN 2790-2692 Vol. 1, Issue 1: 1-8

Invoice Recognition System based on Neural Network

Zhenhao Wan*

Wuhan University of Technology, Wuhan, China

**corresponding author*

Keywords: Neural Network, YOLOv3 Model, Invoice Recognition, Recognition System

Abstract: The reimbursement process of invoices is very complicated, which requires manual input of key information in invoices, which wastes a lot of manpower and time. Therefore, it is particularly important to design an algorithm for intelligent identification of invoice information. This paper mainly carries on the invoice recognition system based on neural network. This paper firstly preprocessed the image and improved the Hough transform to detect the tilt Angle of the invoice image by taking the long horizontal line in the invoice layout as the target. After that, the stamp of the invoice image is removed to reduce the interference of text detection and recognition. Secondly, this paper improves invoice recognition based on YOLOv3 detection algorithm. In this paper, the invoice recognition system is constructed and compared with the other two systems. Through the experimental comparison results, it can be known that the system improves the efficiency of the staff in processing paper invoices and reduces the workload of their later registration and verification of invoices.

1. Introduction

Nowadays, the application of invoice is everywhere in life. It is not only a kind of legal certificate in the purchase or sale of enterprises and individuals, but also an important basis for the financial department of each company to proofread funds. The traditional bill reimbursement process is manually completed, which is not only inefficient and requires a large amount of human resources, but also difficult to ensure the accuracy and confidentiality of bill data in the process of manual operation [1-2]. With the development and progress of computer related technology, nowadays, computer processing images faster and faster, and the rapid development of deep learning technology, which makes use of computer technology to bill for reimbursement to become a reality, this paper submit an expense account approach can greatly improve the working efficiency of the financial staff, at the same time, also can ensure the accuracy of the instrument data and confidentiality, It has a high application value [3-4]. For the reimbursement of invoices, each enterprise has its different reimbursement methods, although some enterprises are carrying out automatic identification of bills related business, but the vast majority of enterprises still use the manual audit method, which has the following shortcomings: Due to the use of manual verification of the reimbursement method, all the valid information in the ticket need to be manually entered into the system by the staff. In this process, the correct rate of the staff's input information cannot be effectively guaranteed. When the amount of bills is too large, the staffs are prone to errors in the input information when they work for a long time due to the high working pressure [5]. In order to ensure the accuracy of effective information of bills, financial departments of enterprises often need to proofread the information entered by staff repeatedly, which makes the review period of bill reimbursement longer and the reimbursement efficiency lower [6].

In recent years, invoice text recognition methods based on deep convolutional neural networks have become dominant. Among them, recurrent neural network is outstanding in the recognition of invoice text with dependent context serialization information. Because this kind of recognition method depends on the character segmentation or which is different from traditional single character recognition method, which can completely get rid of the limitation of character segmentation, and predicting the memory and context information and learn to the serialization of semantic information, and then the input data serialization mapping for variable scale sequence of text output [7]. Among them, the recognition schemes containing memory function mechanism mainly adopt the framework of "code-encode" that is the framework structure of encoding and decoding. Generally speaking, the encoder uses CNN or RNN to convert the input text image into the text sequence information to be recognized, while the decoder is based on the generation label of the target text and vectorization of the text sequence information, and finally obtains the target text to be recognized through the cyclic network layer [8]. In the recognition model based on CTC loss function, CNN is first used to extract text features, then RNN is used to model the serialization dependence information of text features, and finally CTC loss function is used to predict text sequences [9]. This sequence-to-sequence method can better combine the text information of the context, reasonably predict the subsequent text lines, and filter the predicted value of each paragraph to leave the corresponding text line information, so that the accuracy of the transferred text sequence is higher. However, such recognition methods may also suffer from slow training, overfitting or gradient explosion [10].

The realization of automatic bill identification system can not only simplify the reimbursement process of bills, improve the security and confidentiality of bill data, but also improve the efficiency of financial personnel, reduce the waste of resources, and enhance the competitiveness of enterprises.

2. Invoice Detection based on YOLOv3 Model

2.1. Image Preprocessing

Due to the shooting light, shooting background, shooting Angle and the lack of cleanliness of the invoice itself, the invoice image will have strong noise and even the information on the invoice cannot be recognized [11]. Direct use of the original invoice image to identify, will cause a great error in the identification results, resulting in poor identification results. In order to prevent the previous image from having a negative impact on invoice recognition and improve the confidence of invoice recognition, it is imperative to preprocess invoices, including image gray level, denoising, image skewness correction and character cutting [12].

Firstly, the invoice image in the dataset is gray-scaled, which preserves the brightness and contrast of the whole image on the basis of retaining the current image information [13].

Weighted average method is based on some indicators and actual requirements, set the size of weight to distinguish the importance of image information. The specific operation process is to weight and average the three components of R, G and B, and then use the calculated value as the gray value of R, G and B corresponding to the original image [14]. Since human eyes have the highest sensitivity to green, followed by red and the lowest sensitivity to blue, non-important information such as border in the invoice image is red, and the information of invoice number is blue. Combining these two factors, the coefficient of red is 0.3, the coefficient of green is 0.6, and the coefficient of blue is O.1 when applying the weighted average method. The specific formula is shown in Formula (1):

$$f(i,j) = 0.3R(i,j) + 0.6G(i,j) + 0.1B(i,j) \tag{1}$$

When collecting invoices, there may be tilted placement of invoices, defects in collecting hardware scanning equipment, or some other reasons that lead to skewed images. The tilted image has a great influence on the realization of the following information area positioning and character cutting. In order to handle the following steps smoothly, the tilted image should be corrected first. In this paper, the Hough transform is used to correct the image skew [15].

The presence of noise in the picture will lead to the reduction of the SNR of the image, which will have an impact on the accuracy of subsequent invoice recognition and the speed of network recognition. Therefore, image denoising must be completed before character cutting [16]. At present, the commonly used denoising method refers to the direct operation on the pixels of the image, mainly including median filtering neighborhood average method is also known as mean filtering. After studying the advantages and disadvantages of several methods, this topic adopts median filtering. Median filtering can save the edge details of the image. The principle of median filtering is to calculate the median of each pixel in the image and replace the original value of each pixel in the image with the median of all pixels [17]. Let the input original invoice image size be 7*7, with 3*3 filtering window filtering.

After determining the specific coordinate position of the invoice, it is necessary to locate and extract the key information character area in the invoice. Because the format of the VAT invoice is fixed and the printing color is fixed, the binary image with the rectangular box of the invoice can be obtained through the color space information [18]. Outside the rectangular box, there is invoice title and two-dimensional code information; The inside of the rectangular box basically contains all the key information, but these items are distributed around the grid, which is easy to affect each other when positioning. Due to the fixed position of each part of the invoice and the key information inside the rectangular box, the invoice positioning is divided into two parts: coordinate systems are established for the inner and outer areas of the rectangular box, and the relative position relationship between each part and the origin of coordinates is used for positioning.

The number of characters is the most obvious attribute, so the quality of the character segmentation is directly related to the recognition rate of the automatic recognition of the final invoice contents. In the case of a VAT invoice, split a single character in the string field. In view of the limitations of other methods of invoice character segmentation, combined with the fixed parts of the invoice layout separated by grids, the normal characters, hyphenated characters and disconnected characters in the invoice are segmented based on projection feature method.

There are two main methods of image normalization: linear and nonlinear. The linear normalization method can adjust the image while keeping the proportion of each information part of the original image unchanged, while the nonlinear normalization method can not keep the original image unchanged and will affect the retention of image information. This paper mainly adjusts the size of the original image of VAT invoice, standardizes its size uniformly, and uses bilinear interpolation to size the invoice image into 230×230.

2.2. Improved YOLOv3 Model

The text image dataset in this paper is quite different from the traditional image dataset, which contains mostly small and medium targets, and

The size of some fields occupies a very small proportion compared with the whole picture. If the three features in the original YOLOv3 model are used to detect the size division, it is obviously inappropriate and prone to error detection. Therefore, a larger feature map is needed to detect the text target with a smaller size in the invoice image.

In this paper, based on the original YOLOv3 model, a fourth feature scale of 104×104 is added to it, and the large-scale feature detection of the original YOLOv3 model is trimmed. That is, firstly, the Convolutional Set of the detection layer whose feature map size is 52×52 is convoluted by 1×1, and then the upsampling operation is performed by twice, and the output feature map scale is increased from 52×52 to 104×104. Then, the 109th layer and the 11th layer of the feature extraction network are fused by Route layer to make full use of deep features and shallow features. In this way, the underlying features are integrated. The three characteristic scales are 104×104, 52×52 and 26×26, respectively. The new network structure adds larger scale feature detection and more subtle image feature information. In addition, this method only adds additional cross-layer connections to the original network structure, and removes the large-scale feature detection branch of the original model, without increasing the number of parameters and detection time of the original model, so as to strengthen the detection effect of small targets in the invoice image.

The traditional YOLOv3 algorithm uses K-means clustering algorithm to obtain the parameters of anchor frame scale, and the effect of this algorithm depends on clustering

Numerical selection of class cluster K. In the k-means algorithm, the initial cluster center needs to be selected by human intervention, but different choices will produce different results, so once a poor initial value is selected, the clustering effect will be unsatisfactory.

Aiming at the defects of the above k-means algorithm, this paper adopts the K-means++ clustering algorithm for the initial candidate of YOLOv3

Box clustering, k-means ++ clustering algorithm is the perfect algorithm of K-means algorithm. The core principle of K-means++ clustering algorithm is to make the interval of the initial clustering center as large as possible. The specific steps are as follows:

Initialize any point in the data set as the clustering center point of the first cluster;

Obtain the nearest distance F(x) from the rest of the sample points to the cluster center;

The probability of each sample being selected as the next cluster center is defined as Qx, and the size of F(x) directly affects the probability of being selected as the next cluster center. According to the core idea of the algorithm, the larger the distance, the more likely it is to be selected. Qx calculation formula is shown in (2) :

$$Q_x = \frac{F(x)^2}{\sum_{x \in X} F(x)^2} \tag{2}$$

The roulette wheel algorithm was used to determine the new cluster center point.

Repeat steps 2, 3 and 4 until K cluster centers are selected;

The K initial clustering centers obtained in the fifth step are applied to the k-means algorithm for clustering.

In order to avoid the problem that the model loading time is too long and the computer memory loss is serious, this paper uses the model pruning method to compress the trained network model. The steps of model pruning are as follows: the activation degree of neurons in the pre-trained model is measured; The less activated neurons in the model were removed; The pruning may reduce the performance of the model, so it is necessary to fine-tune the model to complete the model repair. The effect of the fine-tuned model was evaluated to judge whether the pruned model met the requirements.

3. Set up Invoice Identification System

FIG. 1 shows the overall block diagram of the system. Firstly, the invoice image taken by the handheld device is preprocessed, including denoising filter, gray level, etc. Then, for the location of invoice data area, the method of template matching with special symbols of each area was used to roughly locate the data area, and then the vertical horizontal projection segmentation method was used to remove the influence of the upper and lower borders to accurately locate the data area. Then, it locates to the data area and uses the improved drip algorithm to segment it. Firstly, it uses the vertical projection method to segment it and sets the character threshold. If the value is larger than the threshold, the drip algorithm is used to segment the adhesive characters. Then, a single character is made into a character set, and LENet-5 network is built through TesorFlow framework, and the final recognition model result is obtained by training. Finally, the invoice code will be identified through HTTP protocol through the verification website to verify the invoice.

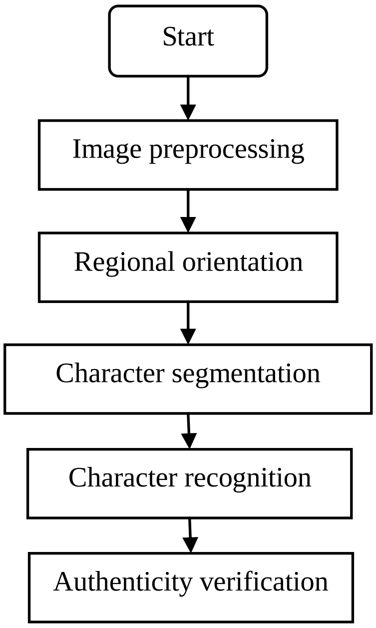

Figure 1. System block diagram

4. System Experiment and Result Analysis

In this experiment, 100 invoice images taken by handheld devices were identified. With the increasing use of value-added tax invoices, automatic recognition technology of value-added tax invoices has been attached great importance by many companies, such as Wentong, Hanvon, etc., and many open source text recognition engines, such as tesseract-OCR, etc. In this paper, tesseract-OCR and Hanvon VAT invoice recognition software are used to conduct a comparative experiment on the VAT invoices collected in this paper.

Table 1. Comparison of regional recognition accuracy of buyers

	Number of characters	Identification number	Correct
Tesseract-OCR	1759	1613	91.70%
Hanvon	1759	1684	95.74%
Our system	1759	1712	97.32%

As shown in Table 1, the recognition accuracy of the proposed algorithm is the highest in the buyer region, leading tesseract-OCR by 5.62% and Hanvon by 1.58%.

Table 2. Comparison of seller's regional accuracy

	Number of characters	Identification number	Correct
Tesseract-OCR	3625	3377	93.16%
Hanvon	3625	3525	97.24%
Our system	3625	3587	98.96%

As shown in Table 2, in the seller area, the recognition accuracy of the proposed algorithm is also the highest, leading tesseract-OCR by 5.8% and Hanvon by 1.72%.

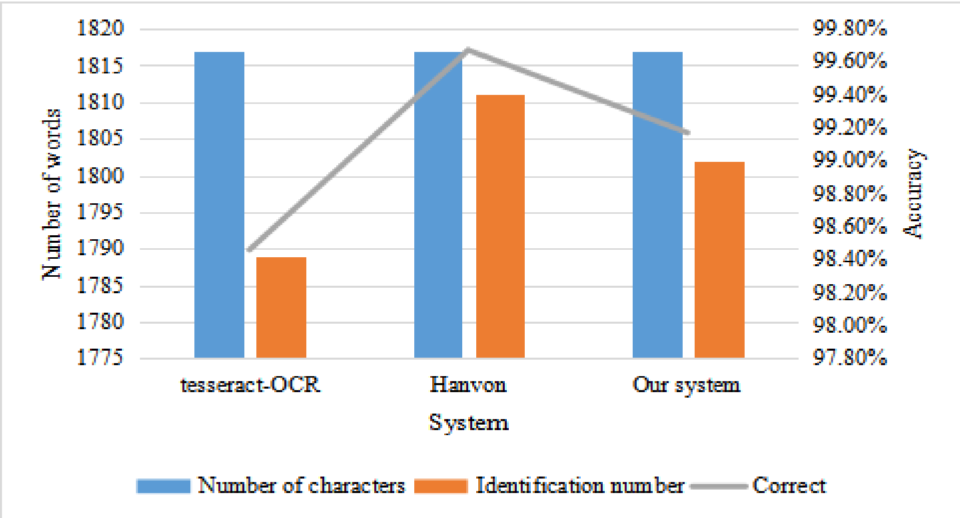

Figure 2. Code, number, date region recognition accuracy comparison

As shown in Figure 2, in the area of invoice code, number and date, the recognition effect of the algorithm in this paper is not as good as that of Hanvon which is 0.5 percentage points lower than that of Hanvon but still 0.71 percentage points higher than that of tesseract-OCR.

Table 3. Comparison of regional recognition accuracy of amount and tax amount

	Number of characters	Identification number	Correct
Tesseract-OCR	774	744	96.13%
Hanvon	774	750	96.90%
Our system	774	752	97.16%

As shown in Table 3, in the area of amount and tax, the recognition accuracy of the proposed algorithm is still higher than that of the other two systems, leading tesseract-OCR by 1.03% and Hanvon by 0.26%.

As shown in Figure 3, the overall character recognition accuracy of the system in this paper is the highest, which is 1.04 percentage points ahead of Hanvon's method and 4.14 percentage points ahead of tesseract-OCR.

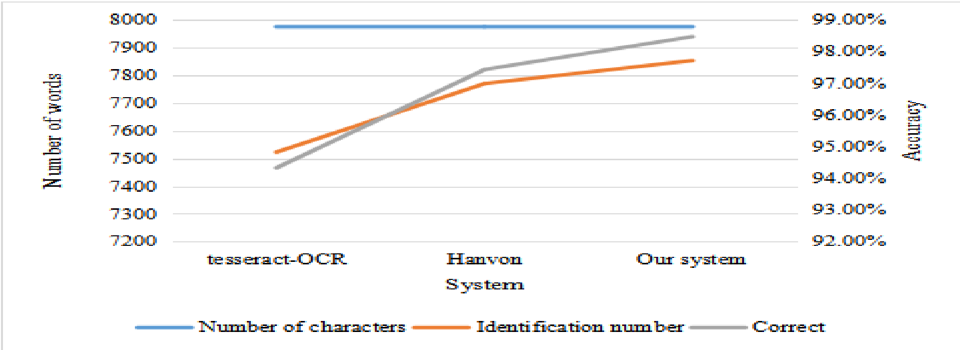

Figure 3. Comparison of total character recognition accuracy

5. Conclusion

With the maturity of information technology and the rapid development of intelligent office, invoice recognition has become the mainstream direction in the research field, which predicts the actual information through the prior information provided by training samples. How to identify quickly and efficiently is also a hot research direction. The scene of the image is different, and the method used is also different. In this paper, the target detection network of YOLOv3 is used to optimize the characteristics of the invoice data set and construct the invoice detection system. Paper value-added tax invoice automatic identification system research, the author of this paper made a certain achievement, but if you want to put the invoice automatic identification system as a commodity software, the use of the enterprise and supply, this system also has many shortcomings, confined to the level of the individual as well as the time is limited, this article not to print all the information on the VAT invoice do identification, Because the information contains not only Chinese characters, but also two-dimensional code and other symbols, simple recognition technology can not complete the recognition of two-dimensional code.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Butaev M M, Babich M Y, Salnikovq I I, et al. Neural Network for Handwriting Recognition. Nexo Revista Cient fica, 2020, 33(2):623-637. https://doi.org/10.5377/nexo.v33i02.10798
[2] Eckert A. Identification and Measurement of Hidden Costs in a Brazilian Turbine Manufacturer Company. Revista Gestão Industrial, 2018, 14(1):199-234.
[3] Bagwell J. Fuel Management: Key Areas to Watch. NAFA Fleet Executive, 2018, 11(6):9-9.
[4] Kosasi M H, Andi A, Arijanto L. Analisa Aspek-Aspek Dalam Prosedur Change Order Pada Proyek Konstruksi. Dimensi Utama Teknik Sipil, 2019, 6(1):9-16.

https://doi.org/10.9744/duts.6.1.9-16

[5] Wiseman Y. Vehicle Identification by OCR, RFID and Bluetooth for Toll Roads. International Journal of Control and Automation, 2018, 11(9):67-76. https://doi.org/10.14257/ijca.2018.11.9.07

[6] Bulan O, Kozitsky V, Ramesh P, et al. Segmentation- and Annotation-Free License Plate Recognition With Deep Localization and Failure Identification. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(9):2351-2363. https://doi.org/10.1109/TITS.2016.2639020

[7] Obaidullah S M, Santosh K C, Das N, et al. Handwritten Indic Script Identification in Multi-Script Document Images: A Survey. International Journal of Pattern Recognition & Artificial Intelligence, 2018, 32(10):1856012.1-1856012.26. https://doi.org/10.1142/S0218001418560128

[8] Singh P K, Sarkar R, Das N, et al. Benchmark databases of handwritten Bangla-Roman and Devanagari-Roman mixed-script document images. Multimedia Tools and Applications, 2018, 77(7):8441-8473. https://doi.org/10.1007/s11042-017-4745-3

[9] Erdal, Erguzen, Atilla. Medical Image Archiving System Implementation with Lossless Region of Interest and Optical Character Recognition. Journal of Medical Imaging and Health Informatics, 2017, 7(6):1246-1252. https://doi.org/10.1166/jmihi.2017.2156

[10] Garg S, Kumar K, Prabhakar N, et al. Optical Character Recognition using Artificial Intelligence. International Journal of Computer Applications, 2018, 179(31):14-20. https://doi.org/10.5120/ijca2018916390

[11] Ganesh S, Bhatti P T, Alkhalaf M, et al. Combining Optical Character Recognition with Paper ECG Digitization. IEEE Journal of Translational Engineering in Health and Medicine, 2020, PP(99):1-1. https://doi.org/10.1109/JTEHM.2020.3083482

[12] Rizvi S, Sagheer A, Adnan K, et al. Optical Character Recognition System for Nastalique Urdu-Like Script Languages Using Supervised Learning. International Journal of Pattern Recognition and Artificial Intelligence, 2019, 33(10):239-245. https://doi.org/10.1142/S0218001419530045

[13] Konovalenko I A, Kokhan V V, Nikolaev D P. Optimal affine image normalization approach for optical character recognition. Computer Optics, 2020, 45(1):90-100. https://doi.org/10.18287/2412-6179-CO-759

[14] Mahesha C. Telugu Optical Character Recognition Using Cloud Computing and Python. International Journal of Computer Science Engineering and Information Technology Research, 2020, 11(1):95-106. https://doi.org/10.24247/ijcseitrjun20209

[15] Hubert, Phoenix P, Sudaryono R, et al. Classifying Promotion Images Using Optical Character Recognition and Nave Bayes Classifier. Procedia Computer Science, 2020, 179(1):498-506. https://doi.org/10.1016/j.procs.2020.01.033

[16] Santoni M M, Chamidah N, Prasvita D S, et al. Penerjemahan Bahasa Indonesia ke Bahasa Minang dari Optical Character Recognition dengan Menggunakan Algoritme Edit Distance. Jurnal Ilmu Komputer dan Agri-Informatika, 2020, 7(2):105-113. https://doi.org/10.29244/jika.7.2.105-113

[17] Rajbongshi A, Islam M I, Rahman M M, et al. Bangla Optical Character Recognition and Text-to-Speech Conversion using Raspberry Pi. International Journal of Advanced Computer Science and Applications, 2020, 11(6):274-278. https://doi.org/10.14569/IJACSA.2020.0110636

[18] Rismanto R, Prasetyo A, Irawati D A. Optimalisasi Image Thresholding Pada Optical Character Recognition Pada Sistem Digitalisasi dan Pencarian Dokumen. Petir, 2020, 13(1):1-11. https://doi.org/10.33322/petir.v13i1.659