Intelligent Computing Technology and Automation
Z. Hou (Ed.)

doi:10.3233/ATDE231257

A Financial Accounting Voucher Recognition Strategy Based on Deep Learning

Yongpeng YANG[a,1], Hong LI[b], Xiangjun HE[a]
[a] *Beijing China-Power Information Technology Co., Ltd., Beijing, 100192, China*
[b] *Sate Grid XinJiang Information & Telecommunication Company, Urumqi, 830000, China*

Abstract. In order to improve the accuracy and efficiency of paper financial reimbursement vouchers, a deep learning based automatic recognition strategy for financial accounting vouchers is proposed. This article designs the basic process of deep learning for image classification and text recognition, and establishes invoice image datasets for training CNN networks. Next, parameter design is performed on the AlexNet network structure, and the trained network model is used to improve the digital recognition method. The experimental part of the text annotation system collected multiple invoice text recognition datasets, and the optimized model is subjected to network training and testing experiments on the dataset. The test results show that the detection accuracy of specific fields in the invoice image reaches over 95%, and the single item recognition rate of each field reaches about 95%. The model accuracy and inference speed can meet the performance requirements.

Keywords. deep learning; CNN; AlexNet; financial accounting voucher; recognition; invoice

1. Introduction

In the context of the information age, it has become a consensus to solve the efficiency issues of financial reimbursement in enterprises through internet technology. Through intelligent information technology means, enterprise financial operating costs can be reduced, unnecessary financial reimbursement environments can be reduced, and the accuracy of financial reimbursement can be improved. Since the transaction behavior between enterprises is becoming more complex, and the company's invoice business is growing at a high proportion. The procedures under administrative jurisdiction are more complex. How to use intelligent input methods to identify and input paper invoices is a future research hotspot The process of financial informatization in China is relatively slow, with a backward start. Currently, there is still a lack of financial management reimbursement systems that are suitable for Chinese characteristics, mainly reflected in two aspects: firstly, most state-owned enterprises and institutions in China still follow the principle of paper receipts and reimbursement [1,2].

The financial management method of signing has certain technical difficulties in image recognition of paper invoices, which cannot achieve 100% recognition; Secondly, large state-owned enterprise group companies and their subsidiaries are financially

[1] Corresponding Author: Yongpeng YANG; Beijing China-Power Information Technology Co., Ltd., Beijing, 100192, China; 18103693072@163.com

independent and have their own distinct financial reimbursement and management systems. These different systems cannot be interconnected, which poses significant obstacles to the sharing of financial data. With the rapid development of artificial intelligence technology, technologies such as deep learning and image processing are gradually being applied in various fields, such as face recognition and feature detection. The essence of deep learning is to train sample data by constructing multiple hidden layer neural networks, and the essence of training samples is to allow the network to autonomously learn the features of the samples. The development of related technologies in the field of deep learning has provided the possibility of further improving the efficiency of invoice management, and various invoice recognition systems have emerged.

This article first summarizes the current status and research methods of image recognition technology at home and abroad, with a focus on introducing deep learning methods based on CNN. Firstly, the basic process of intelligent recognition of financial accounting vouchers was designed based on demand analysis. Then preprocess the collected ticket images, including image size adjustment, graying, denoising, and other operations to improve the accuracy of the model. Use a deep learning framework to construct a CNN model for preliminary recognition of images in invoices. Finally, use an improved AlexNet network to train the AlexNet model using a preprocessed text image dataset. Finally, targeted improvements were made to the selected algorithm, and the improved algorithm was simulated and verified. The experimental results show that the improved CNN model achieves higher recognition accuracy with lower time costs, effectively improving the recognition rate of damaged invoices while avoiding overfitting, making it more suitable for the recognition task of financial invoices.

2. The Working Principle of Financial Accounting Voucher Recognition System

2.1 The overall Architecture of Accounting Voucher Recognition System

The overall architecture of the system adopts a form of front-end and back-end separation, consisting of the user end and the server end. The user end is responsible for interacting with the user to upload the target image and display the recognition results. The server is responsible for deploying deep learning algorithms, including image classification, text detection, text recognition, and text extraction modules. The overall architecture of the invoice recognition system is shown in Figure 1.

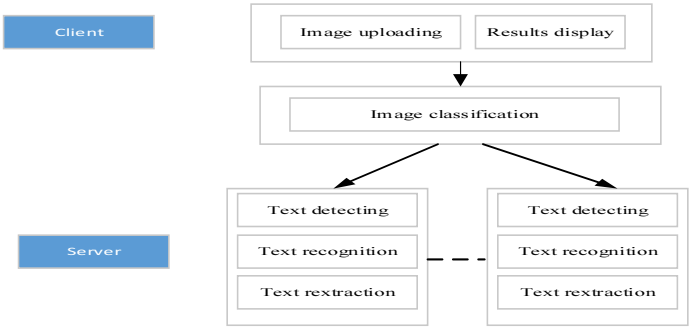


Figure 1. The overall architecture of the accounting voucher system.

Firstly, the image classification module will send images to different algorithm servers based on their classification results. Secondly, the algorithm server runs deep learning algorithms to detect text and extract the text box information and corresponding text information of the ticket [3]. Finally, the text extraction module matches key fields such as face amount and invoicing date based on the position of the text box and the recognition results of the text within it.

2.2 Preprocessing of Invoice Images

Due to the presence of certain noise in the original invoice image and the occurrence of jitter during image capture, preprocessing is carried out before image classification, mainly including filtering and tilt correction. Line preprocessing mainly includes two parts: filtering and tilt correction. Firstly, use a size of 5×5. The filtering window of 5 performs median filtering on the original image, greatly reducing noise in the image. Secondly, Hough transform is used to correct the tilt of the invoice image [4]. After determining the tilt angle of the invoice image, geometric transformation is needed to achieve tilt correction. The coordinates of the points in the output image after geometric transformation are usually approximated by polynomials. By analyzing the layout of tax invoices, a more accurate, accurate, and efficient positioning of the tax invoice information that needs to be extracted plays a crucial role in future tax invoice recognition.

$$y' = \sum_{r=0}^{m} \sum_{k=0}^{m-r} b_{kr} x^r y^k \tag{1}$$

where (x', y') is the point coordinate of output image; (x, y) is the point coordinate of original image; a_{kr}, b_{kr} are parameters.

Such change is linear for a_{kr}, b_{kr}. If the correspondence points (x, y) and (x', y') in two image are known, the value of a_{kr} and b_{kr} can be computed by the system of linear equations:

$$\begin{cases} x' = a_0 + a_1 x + a_2 y \\ y' = b_0 + b_1 x + b_2 y \end{cases} \tag{2}$$

During the rotation change, the Jacobian coefficient characterizing coordinate system changes is

$$J = \left| \frac{\partial(x', y')}{\partial(x, y)} \right| = \begin{vmatrix} \frac{\partial(x')}{\partial x} & \frac{\partial(x')}{\partial y} \\ \frac{\partial(y')}{\partial x} & \frac{\partial(y')}{\partial y} \end{vmatrix} \tag{3}$$

3. Recognition Algorithm Design

3.1 Image Recognition

Convolutional neural networks(CNN) have enormous advantages in the processing of two-dimensional images. The characters in the invoice information to be recognized are

essentially a two-dimensional image. Considering the powerful advantages of convolutional neural networks in image recognition, this article uses convolutional neural networks to complete the preliminary recognition of characters in invoices. The specific process is to simulate the printing effect that is similar to the invoice characters. This article preprocesses a portion of the images in the dataset to obtain an image dataset with simulated breakpoint effects. Then, the original dataset and the dataset with breakpoints are combined to form a brand new dataset, abbreviated as the self built dataset. This dataset is then used to train the CNN network. After the training is completed, the network model is saved locally. Finally, the saved network model is used to recognize the images in the invoice.

We design a relatively streamlined CNN network structure, using fewer convolutional kernels and smaller convolutional windows for feature extraction. You can choose fewer convolutional layers between 3-5 and use a smaller pooling window for downsampling to reduce the size of the feature map. Reducing the size of convolutional kernels and pooling windows: Smaller convolutional kernels and pooling windows can reduce computational complexity and accelerate recognition speed. Using Leaky ReLU, the slope of negative values can be controlled by setting the negative slope parameter, which is set to 0.1 in this example. The entire network is built using Python language and Python deep learning framework, with some core codes described as follows:

```
# Assuming the training dataset train is ready_ Loader and test dataset test_
Loader
for epoch in range(num_epochs):
    running_loss = 0.0
    for i, data in enumerate(train_loader, 0):
        inputs, labels = data
        optimizer.zero_grad()
        outputs = net(inputs)
        loss = criterion(outputs, labels)
        loss.backward()
        optimizer.step()
        running_loss += loss.item()
        if i % 100 == 99:
            print('[%d, %5d] loss: %.3f' % (epoch + 1, i + 1, running_loss /
100))
            running_loss = 0.0
```

3.2 Text Recognition

The data information on financial invoices may be affected by noise. Noise can be random errors in imperfect data collection processes caused by scanning, printing, transmission, or other factors. These noises may cause errors or inaccuracies in the data on the invoice. Before data recognition, preprocess the invoice image, such as denoising, image enhancement, image smoothing, etc., to reduce the impact of noise. However, if only one feature extraction method is used for recognition processing, the results will not be ideal. Deep learning, on the other hand, does not care about the characteristics of the data. It can automatically learn from massive amounts of data and find features in the data. Users only need to understand the network structure and do not have to rely on the characteristics of the data like feature extraction. Therefore,

deep learning using a large amount of noisy invoice data for model training can help the model learn and adapt to the presence of noise. The AlexNet model with a 5-layer convolutional structure has demonstrated the effectiveness of convolutional neural networks in complex models, and introduced GPU training into the research field, resulting in a milestone in shortening big data training time. It is widely used in the field of image recognition and has outstanding effects. Its specific structure is shown in Figure 2. By performing enhancement operations on training data, such as rotation, translation, scaling, etc., the diversity of the data can be increased, helping the model better adapt to noise. The Softmax layer performs corresponding function processing on the output of fully connected layer 3, which can output 1000 categories

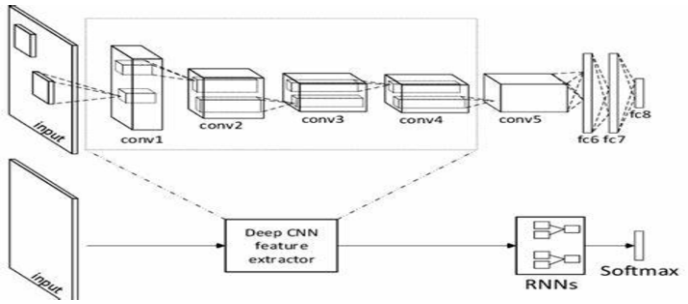


Figure 2. AlexNet network structure graph.

Previous studies have found that when using multiple single character images to train a network model, due to the large sample size, it requires a lot of time and computer resources to train a model well [5-7]. Although distributed parallel computing can accelerate model training, the computer resources occupied have not decreased. In order to solve the problems of the Alexnet network, this article made a series of adjustments to the original Alexnet network:

The improvement of Alexnet network model includes the follows:

(1) The convolutional kernel size used by AlexNet is 11x11, and smaller convolutional kernels such as 5x5 can be used;

(2) Introduce more convolutional layers and fully connected layers to improve the expression ability and accuracy of the model;

(3) Using batch normalization to accelerate the convergence speed of the network and improve the stability and accuracy of the model;

(4) Introducing residual connections to solve the problems of vanishing and exploding gradients, making the network easier to train and optimize;

(5) Use data augmentation techniques to increase the diversity of training data, while using regularization techniques Dropout to reduce overfitting.

To avoid the problem of ambiguity in average pooling, a local response normalization layer similar to the lateral inhibition mechanism of biological neural activity is proposed. For each neuron's output, calculate its response activity in the local neighborhood and compare it with the responses of other neurons in the neighborhood. Then, for neurons with larger responses, a positive stimulus is applied to enhance their corresponding feature representation. In order to improve the LRN layer in the AlexNet network, it can be considered to increase the degree of positive excitation, so as to exert stronger positive effects on responsive neurons. This can further enhance the model's learning ability for important features, thereby improving the model's performance [8].

The specific expression for local normalization response is:

$$b_{x,y}^{i} = a_{x,y}^{i}\left[k + \alpha \sum_{j=\max(0,i-n/2)}^{\min(N-1,i+n/2)} (a_{x,y}^{i})^{2}\right] \quad (4)$$

The actual put is $a_{x,y}^{i}$ in the i_{th} Convolutional Layer, N is the total number of thoroughfare, k, α, β are hyper-parameters which are usually set as 2, 000001 and 0.25.

4. Experimental Analysis

Firstly, the original images of invoices in the dataset are used to train the invoice classification model. Collect raw images containing different types of invoices and organize them into a dataset [9]. Preprocess the original image, including image scaling, cropping, rotation, graying, and other operations to ensure that the image has consistent size and format. Label data: Add labels to each image, indicating the type of ticket they belong to. We can manually annotate or use automated tools to annotate the entire dataset into training, validation, and testing sets. Usually, the training set is used for model training, the validation set is used to adjust the hyperparameters of the model, and the testing set is used to evaluate the performance of the model. The first type of dataset collected 269 value-added tax invoices to test the accuracy of the system's digital recognition: 4012 in the buyer's taxpayer identification number area and 4238 in the seller's taxpayer identification number area. Six invoice images were randomly selected from the test set, and the corresponding invoice type prediction results were provided by the computer. The visualization effect of the invoice type recognition experiment is shown in Figure 3. Figure 4 shows the segmentation result based on the improved CNN algorithm: taking the minimum bounding rectangle of the target area. Eliminate enclosed contours and eliminate rectangles of unreasonable size; Use SVM to determine whether the images inside the rectangular box are numeric, and save the sequence numbers in order from left to right. It can be seen that this scheme can perform structured and precise identification, obtain multiple reference boundary lines in the approximate area, and determine the actual area of the invoice based on the multiple reference boundary lines.

Figure 3. Invoice image recognition results

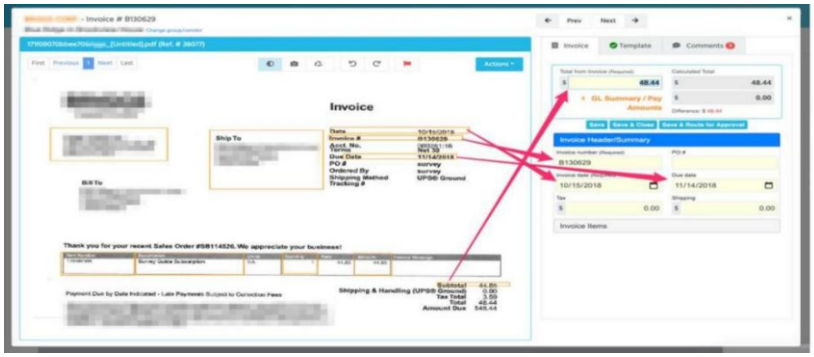
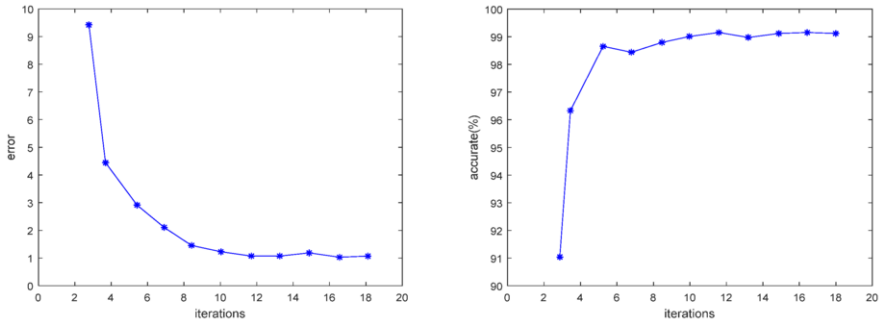


Figure 4. Invoice area and text recognition results

We have designed and generated a second dataset for network training based on the multi region localization method designed in this article. The content included on the voucher includes the name of the goods or taxable services, the name of the goods that need to be reflected on the ticket, the unit of measurement, and the specifications and models: the above can be filled in according to business requirements. For each invoice, the preset invoice recognition model is called to identify the invoice content within the actual area of the invoice. Optionally, the rough area of the invoice in the image is processed to obtain multiple reference boundary lines in the rough area, including: processing the rough area of the invoice in the image to obtain a line map of the grayscale contour in the rough area of the invoice. Based on the information of the invoice to be identified, locate and segment the area of the invoice image where it is located to obtain the image of the invoice information to be identified; Identify the text content of the invoice information image to be identified and obtain the invoice information to be identified. The second type of dataset generated in this experimental design consists of 3000 sheets Build an AlexNet network model using Python language and Python deep learning framework, set the number of training epochs to 20, and the batch size to 128. The training results are shown in Figures 5. It can be seen that the improved algorithm can indiscriminately recognize every character in the invoice image, and the recognition accuracy remains above 95%.



(a) training error (b) accuracy of testing set

Figure 5. Invoice area and text recognition results

According to Table 1, the overall recognition rate of numbers has reached 98.68%. Deep learning indeed has strong self-learning ability, which can be trained through a large amount of data to improve the system's recognition ability. For invoice recognition, deep learning can automatically extract key information such as invoice

number, amount, date, etc. by learning the features and patterns of invoices. In addition, the concise layout of invoices can also help improve the system's recognition rate, as a concise layout can reduce the possibility of interference and misreading, making it easier for the system to accurately identify invoice content The recognition rate of taxpayer identification number and invoice number digits is usually higher than the amount and tax amount. This is because the taxpayer identification number and invoice number usually have a fixed format and length, and are not easily affected by the layout arrangement. In contrast, there may be multiple representations of amounts and taxes, such as the position of decimal points, differences in currency symbols, etc. These factors may increase the difficulty of identifying amounts and taxes. However, through deep learning training, the system can learn common patterns and features of amounts and taxes, thereby improving its recognition rate. Although the recognition rate of amounts and taxes may be relatively low, the recognition ability of the system will continue to improve with more data training and algorithm improvements.

Table1. Digital identification of each information area

	Character number to be recognized	**Correct number**	**Wrong number**	**Recognition rate(%)**
amount of money	2351	2319	32	98.63
Tax amount	2678	2567	111	95.85
Invoice number	1885	1850	35	98.14
Taxpayer identification number of the purchaser	3920	3913	7	99.82
Taxpayer identification number of the seller	3885	3879	6	99.84

5. Conclusion

To address the cumbersome and complex process of voucher entry in the financial reimbursement process, this paper proposes an intelligent recognition scheme for financial reimbursement vouchers based on image recognition technology. We use computer vision and machine learning algorithms to automatically recognize and extract information from financial reimbursement vouchers. In image preprocessing, the invoice image is first used for a series of preprocessing such as filtering and adaptive threshold to obtain a binary image. Then, a CNN image recognition model is used to preprocess the image, such as filtering and tilt correction. After the training is completed, the model is evaluated and optimized, using a portion of data that has not participated in the training to evaluate and calculate the accuracy, recall, and other indicators of the model. The experimental results show that the proposed image classification has better accuracy and significantly reduces the average time for image classification. The overall results of image classification have been effectively improved, with an overall recognition rate of over 95%. The scheme not only effectively overcomes the shortcomings of current image classification methods, but also has high practical application value. This technology can help businesses and individuals quickly and accurately process a large number of financial reimbursement vouchers, improving work efficiency and accuracy.

References

[1] A Preliminary Research on the Preprocessing and Entry System Design of Financial Reimbursement Vouchers Based on Image Recognition，2021, 37(12): 149-151.
[2] As, Imdat, S. Pal, and P. Basu. Artificial intelligence in architecture: Generating conceptual design via deep learning. International Journal of Architectural Computing 16.4(2018):306-327.
[3] CHEN Yanlan. Design of a Deep Learning-Based Invoice Recognition System. Information & Computer, 2023, 4: 176-180
[4] ZHANG Zhen, NI Hong-jun. Classification of Invoice Image Based on Deep Learning. JOURNAL OF NANTONG VOCATIONAL UNIVERSITY, 2020, 34(2): 79-83.
[5] Choi, Jin Sol, et al. Visual Speech Recognition System with Deep Neural Networks. International Journal of Applied Engineering Research, 2018 15:13.
[6] Lakshmi S V, Ahilan A, Rejula M A, et al. Recognition of brain stroke shape using multiscale morphological image processing. The Imaging Science Journal, 2021, 69(1-4):28-37.
[7] Aouani, Hadhami, and Y. B. Ayed. Speech Emotion Recognition with deep learning. Procedia Computer Science, 2020, 176 :251-260.
[8] Lakshmi, S. Venkata, et al. Recognition of brain stroke shape using multiscale morphological image processing. The Imaging Science Journal, 2021,69(4):28-37.
[9] Mishra S R, Mishra T K, Sanyal G, et al. Real time human action recognition using triggered frame extraction and a typical CNN heuristic. Pattern Recognition Letters, 2020, 135:329-336.