Research on Tickets Classification Method Based on Convolutional Neural Network

Zhuang Chen¹, Quanxing Xiang²,*

¹School of Computer Science and Engineering, Chongqing University of Technology, Chongqing, 400065, China
²School of Software Engineering, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China

*Corresponding author e-mail: s181231007@stu.cqupt.edu.cn

Abstract. The traditional process of tickets reimbursement is cumbersome, which consumes a lot of human and material resources. Before tickets reimbursement, tickets need to be classified. This research mainly introduces a classification method of tickets based on convolutional neural network, which can complete the classification of various tickets.

Keywords: Tickets Reimbursement, Tickets Classification, Convolutional Neural Networks

1. Introduction

As a proof of payment, tickets are used increasingly widespread in real life. The traditional methods of tickets reimbursement are inefficient, which requires consume a lot of human and material resources[1]. To solve the problem of low efficiency in tickets reimbursement, in recent years, automatic tickets recognition technology has been extensively studied.

Before completing the tickets identification, the tickets need to be classified. There are two common tickets types: electronic tickets and paper tickets. For electronic tickets, the layout is neat, clear and easy to distinguish. For paper invoices, the image quality is easily affected by external factors, such as the quality of the paper and printing, the pixels of the photographer’s mobile phone, and so on. Therefore, the classification and identification of paper tickets are more difficult.

For the problem of image classification, which can use the convolutional neural network model to train the input images to achieve the purpose of classification[2]. In such problems, the commonly used convolutional neural network models are as follows: Alex Net[3], VGG-16[4], Res Net[5] and so on.

Generally speaking, the performance of a convolutional neural network model is positively related to its depth[6], but as the number of model layers increases, its parameters will increase, too. which
means that the difficulty of model learning will increase[7]. As a result, the phenomenon of gradient disappearance or overfitting occurs. Although the residual unit of the Res Net network model has a good performance in solving the problem of gradient disappearance, Res Net is also helpless for the overfitting phenomenon that is common in deep neural networks[8].

In practical applications, the smaller size of model, the smaller the amount of calculation and memory occupied[9]. The Res Net network model has a large number of layers and a large number of weight parameters[10]. For this research, if you choose the Res Net network model, it will take up a lot of storage space. Therefore, we prefer to choose a simpler network model.

This research designed a method of bill classification based on deep learning, which was improved on the basis of the Alex Net network model, which can complete the automatic classification of various bills.

The paper recall is structured as follows. The data set will be introduced in Section 2. The model structure and its differentiation from the Alex Net will be explained in Section 3. In order to check the superiority of the proposed model, experimental result and analysis will be discussed in Section 4. Finally, Section 5 presents the paper’s conclusions.

2. Introduction to the Data Set

The data set used in this article is shown in table 1. The total number of samples is 1650, and there are a total of 6 categories. Each category has 275 pictures, namely: air tickets, lottery tickets, electronic invoices, paper invoices, train tickets, and shopping receipts. Each category randomly selects 25 pictures as the test set, and the remaining 250 pictures are divided into training set and verification set according to the ratio of 7:3.

| Data Set     | Samples | Images format |
|--------------|---------|---------------|
| Training set | 1050    |               |
| Validation set | 450    | JPG           |
| Testing set  | 150     |               |

3. Structural Analysis of the Network Model

The structure of the Alex Net network is shown in the left picture of Figure 1, it has 5 convolutional layers, 3 pooling layers and 3 fully connected layers. A convolution kernel of size $11 \times 11$ is used in the first convolution layer, and the step size is set to 4. The second convolutional layer uses a $5 \times 5$ convolution kernel, and set the step size to 1. Due to the Alex Net network a large number of parameters, it is not suitable for the data set in this research. Therefore, the structure of the Alex Net network needs to be simplified and improved.

This research draws on the advantages of the Alex Net network model and designs a convolutional neural network model. As shown in the right picture of Figure 2, the model has 6 convolutional layers, and each 2 convolutional layers as a group. The size of the convolution kernel of the first group is $5 \times 5$, $5 \times 11$, and $11 \times 11$.
and the size of the convolution kernel of the remaining groups is $3 \times 3$. The number of convolution kernels in each group is set to 16, 32, and 64 for extracting image features.

![Diagram of AlexNet and Ours](image)

**Figure 1.** Structure of Alex Net and ours

When the input image passes through the convolutional layer, filter traverses the image in accordance with the step size and extracts the image features to create a new image. The specific operations are as follows[11]:

$$\text{output} = \frac{I - f + 2 \times p}{s} + 1$$

(1)

Where output represents the size of the picture after the convolution operation, $I$ denote the size of the input picture, $f$ is filter size, $p$ denote padding size, and $s$ represents the convolution step size.
Figure 2. First convolutional layer of Alex Net

As shown in Figure 2, when the input image passes through the first convolutional layer of Alex Net, it will be convolved with the convolution kernel of size $11 \times 11$. When the step size is set to 4, the output feature map size is $55 \times 55$. There are 34944 parameters in this process.

As shown in Figure 3, by changing the network structure, the size of output feature map is $54 \times 54$, but the number of parameters only have 7632.

Figure 3. The improved first set of convolutional layers

Introducing the maximum pooling layer can effectively reduce the amount of calculation of the model. This paper sets the pooling core size to $2 \times 2$, which means that the length and width of the feature map are reduced to half of the original.

The fully connected layer can fuse the features extracted by the convolutional layer to perform classification tasks. In this paper, there are 3 fully connected layers, The number of neurons are 32, 16, and 6, which greatly reduces the amount of calculation.

Since the convolutional layer is linear, the activation function is used to add nonlinearity to the convolutional neural network, thereby improving the fitting ability of the model. After each convolutional layer, the corresponding ReLU activation function is connected; in the last layer, the Softmax function is used for classification to complete the classification task of the note.
In order to ensure the final convergence of the model, 200 rounds of model training were set, with 32 samples in each batch. The learning rate is set to 0.0005. In order to prevent the occurrence of over-fitting to a certain extent, the discard rate is set to 0.2, so that each neuron fails with a probability of 0.2 in each training batch, which reduces the interdependence and connection between neurons. Enhance the generalization ability of the model.

4. Experimental Results and Analysis

In the problem of bill image classification, accuracy is the most important evaluation index. In order to verify the accuracy of model classification, three models were trained and compared, including Alex Net, VGG-16 and the model designed in this chapter. Finally, we conducted experiments on the test set, the experimental results are shown in Figure 4, There are 150 samples in the testing set, and the blue squares indicate the number of samples that are correctly predicted. Through calculation, we can get the experimental results in Table 2.

As show in the table 2, the model proposed in this study has the highest classification accuracy rate, which is 94.66%. Although the VGG-16 network model has the highest complexity, its accuracy rate is not the highest. At the same time, In the training phase, the model proposed in this paper takes the shortest time, which is 1798 seconds. When testing 150 images, the time used by the three network
models is also different. The network model in this study takes the shortest time, which is 10 seconds, mainly because of the small scale of the model. Therefore, in terms of comprehensive accuracy, training times and testing time, the performance of the model proposed in this study is better.

Table 2. Comparison of models

| Models | Accuracy(%) | Training times(s) | Testing times(s) |
|--------|-------------|-------------------|-----------------|
| AlexNet | 93.33%      | 1804              | 10              |
| VGG-16 | 93.33%      | 2406              | 12              |
| Ours   | 94.66%      | 1798              | 10              |

5. Conclusion

This article designs a convolutional neural network model to implement a variety of ticket classification tasks. The model mainly includes six convolutional layers, three pooling layers, and three fully connected layers. Experimental results show that compared with other models, the model proposed in this study has higher classification accuracy and faster running speed.

References

[1] D. Zhang, X. Han, and C. Deng, “Review on the research and practice of deep learning and reinforcement learning in smart grids,” CSEE J. Power Energy Syst., vol. 4, no. 3, pp. 362–370, Sep. 2018.

[2] K. Azhar, F. Murtaza, M. H. Yousaf, and H. A. Habib, “Computer vision based detection and localization of potholes in asphalt pavement images,” in IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), Vancouver, BC, Canada, 2016, pp. 1–5.

[3] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, “Deep feature extraction and classification of hyperspectral images based on convolutional neural networks,” IEEE Trans. Geosci. Remote Sens., vol. 54, no. 10, pp. 6232–6251, Oct. 2016.

[4] Z. Ma, Z. H. Tan, and J. Guo, “Feature selection for neutral vector in EEG signal classification,” Neurocomputing, vol. 174, no. PB, pp. 937–945, Jan. 2016.

[5] T. Zhou, S. Yang, L. Wang, J. Yao, and G. Gui, “Improved cross-label suppression dictionary learning for face recognition,” IEEE Access, vol. 6, pp. 48716–48725, Aug. 2018.

[6] J. Pan, Y. Yin, J. Xiong, W. Luo, G. Gui, and H. Sari, “Deep learning based unmanned surveillance systems for observing water levels,” IEEE Access, vol. 6, no. 1, pp. 73561–73571, 2018.

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 779–788.

[8] Z. Ma, J. Xie, H. Li, Q. Sun, Z. Si, J. Zhang, and J. Guo, “The role of data analysis in the development of intelligent energy networks,” IEEE Netw., vol. 31, no. 5, pp. 88–95, Sep.
2017.

[9] M. Liu, J. Yang, T. Song, J. Hu, and G. Gui, “Deep learning-inspired message passing algorithm for efficient resource allocation in cognitive radio networks,” IEEE Trans. Veh. Technol., pp. 1-13, Nov. 2018, doi: 10.1109/TVT.2018.2883669.

[10] Z. Ma, H. Yu, W. Chen, and J. Guo, “Short utterance based speech language identification in intelligent vehicles with time-scale modifications and deep bottleneck features,” IEEE Trans. Veh. Technol., vol. 68, no. 1, pp. 1–8, 2019.

[11] Z. Ma, H. Yu, Z. Tan, and J. Guo, “Text-Independent Speaker Identification Using the Histogram Transform Model,” IEEE Access, vol. 4, pp. 9733–9739, Jan. 2017.