JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Tobacco Sales Bill Recognition Based on Multi-Branch Residual Network

Yuxiang Shan[1], Cheng Wang[1], Qin Ren[1], and Xiuhui Wang[2,*]

Abstract

Tobacco sales enterprises often need to summarize and verify the daily sales bills, which may consume substantial manpower, and manual verification is prone to occasional errors. The use of artificial intelligence technology to realize the automatic identification and verification of such bills offers important practical significance. This study presents a novel multi-branch residual network for tobacco sales bills to improve the efficiency and accuracy of tobacco sales. First, geometric correction and edge alignment were performed on the input sales bill image. Second, the multi-branch residual network recognition model is established and trained using the preprocessed data. The comparative experimental results demonstrated that the correct recognition rate of the proposed method reached 98.84% on the China Tobacco Bill Image dataset, which is superior to that of most existing recognition methods.

Keywords

Artificial Intelligence, Image Recognition, Residual Network

1. Introduction

As important aspects of artificial intelligence technology, image classification and recognition [1,2] have been used extensively in industry, agriculture and daily life. For example, a fine-grained recognition method [3] was proposed to improve the real-time performance and accuracy of vehicle recognition in an intelligent transportation system. Chen et al. [4] presented a load forecasting framework based on self-adaptive dropout, which improves the load supply robustness of power systems. However, as the input data and specific requirements vary significantly in different application scenarios, it is difficult to use a general model to solve all problems. For example, in the identification of tobacco sales bills, different regions have varying management departments, and different customers adopt varying management clients, which results in many differences in the formats and contents of the sales bills.

In view of the above problems, this paper presents a novel multi-branch residual network for tobacco sales bills to improve the efficiency and accuracy of tobacco sales reconciliation. Our work can be summarized as follows:

(1) A novel multi-branch residual network is proposed for the recognition of tobacco sales bills. The proposed network integrates a multi-branch conv-block module, and spatial and channel squeeze

and excitation (SCSE) module.

(2) Score advancements are achieved on large-scale datasets. We conducted comparative experiments using the China Tobacco Bill Image (CT-BI) dataset, which consists of more than 125,000 images.

The rest of this paper is structured as follows: Section 2 discusses related work on image classification and recognition for different tasks. Then, the methodology of the proposed framework is described in detail in Section 3. In Section 4, several comparative experiments that were conducted on CT-BI dataset are outline, and conclusions are drawn in Section 5.

2. Related Work

The application of image recognition technology in various fields has resulted in a series of phased achievements. A weakly supervised fine-grained image recognition method that can accurately locate objects and parts without annotation was proposed in [5]. The experimental results demonstrated that the target mask module and salient point detection module in the method could suppress the background interference and improve the correct recognition rate. Zhao et al. [6] proposed a nonlinear loosely coupled nonnegative matrix decomposition method for low-resolution image recognition, in which the target images were regarded as being composed of different local features.

Furthermore, to evaluate the reliability of image recognition applications that are driven by deep learning technology to change the image background area, Zhang et al. [7] introduced a deformation test method for image recognition systems. To alleviate the difficulty of understanding the network system owing to the unpredictability of the network structure, a coordination method based on deep learning technology was proposed [8] to solve the network structure reasoning problem by integrating a residual network and fully-connected network. Yan et al. [9] proposed a method for predicting future traffic, and constructed a prediction interval based on the combination of a residual neural network as well as upper- and lower-bound estimation. This method generated a combined residual network by optimizing the objective function and adjusting the type of the remaining blocks, which effectively improved the accuracy of quantifying the uncertainty prediction of future traffic. To solve the problem of obtaining subtle clues in the process of fine-grained image recognition more effectively, Kim et al. [10] proposed a method for generating the characteristics of hard negative samples that reduced the dependence on the number of tuples of hard negative samples.

3. Methodology

The proposed tobacco sales bill recognition model consists of three multi-branch conv-blocks and an SCSE module, as shown in Fig. 1.

3.1 Data Preprocessing

In this study, the data preprocessing consisted of two sequential processes: image calibration and image segmentation, as shown in Fig. 2.
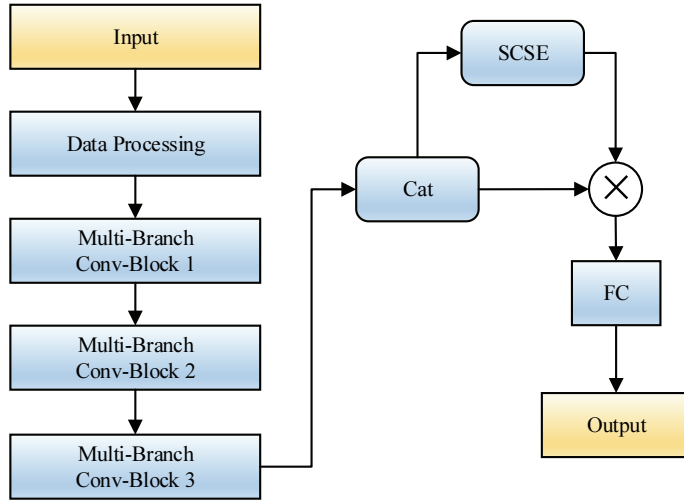
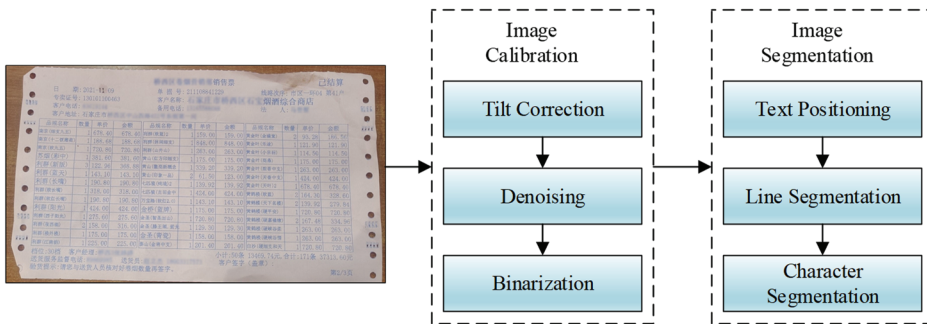**Fig. 1.** Architecture of the proposed recognition model.

Fig. 2. Flow of data preprocessing.

In the image correction process, the input image was tilt-corrected; the four top corners of the document were extracted to calculate the length and width of the image, and rotation alignment was performed. Thereafter, in order to reduce the impact of image noise on the recognition rate, a Gaussian difference filter was applied for noise reduction and the image was binarized. The transfer and impact response functions are presented in Eqs. (1) and (2), respectively.

$$G(s) = Me^{-s^2/2\alpha_1^2} - Ne^{-s^2/2\alpha_2^2} \tag{1}$$

$$g(t) = \frac{M}{\sqrt{2\pi\sigma_1^2}} e^{-t^2/2\sigma_1^2} - \frac{N}{\sqrt{2\pi\sigma_2^2}} e^{-t^2/2\sigma_2^2} \tag{2}$$

where $M \geq N$, $\alpha_1 > \alpha_2$, $\sigma_i = 1/2\pi\alpha_i$.

Based on the image calibration, the Sobel operator was used to determine the gradient in the X-direction of the input image to realize the text positioning operation. Each pixel in the input image was convoluted using two convolution kernels of the Sobel operator. One of the two convolution kernels had the largest response to the vertical edge, whereas the other had the largest response to the horizontal edge. The

maximum value of the two convolution results was used as the output of the pixel. Then, the processed image was further expanded and corroded to detect the text area, to realize line segmentation of the entire image. Finally, a vertical projection operation was used to divide each line into a series of characters for processing.

3.2 Recognition Model

The preprocessed image was input into the proposed recognition model that consists of a multi-branch conv-block module and an SCSE module, as shown in Fig. 1.

The design of the former integrates the branch idea of the Inception series model and residual mechanism of the ResNET network. To reduce the total number of parameters, we referred to the Inception network and used two 3×3 convolutions instead of large 5×5 convolutions. This improvement could reduce the number of model parameters and establish more nonlinear transformations, which increased the capability of the conv-block for learning features. Moreover, using this structure, the sparse matrix could be clustered into dense submatrices to improve the computational performance. As illustrated in Fig. 3, the input of the multi-branch conv-block module was composed of four branches: branch 0 was composed of one BasicConv2d; branch 1 was composed of two BasicConv2d; branch 2 was composed of three BasicConv2d; and branch 3 was composed of an average pool and one BasicConv2d. The input tensors passed through the above four branches, respectively, following which the results were spliced together. The advantage of this procedure was that visual information could be processed on different scales and subsequently aggregated, and features could be extracted from different scales simultaneously.
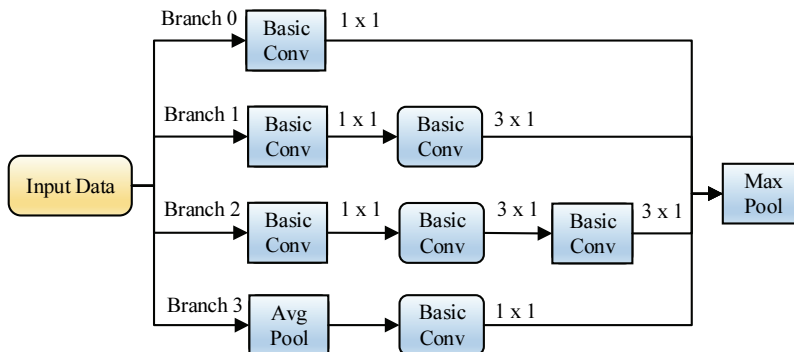
Fig. 3. Multi-branch conv-block module.

The SCSE module consists of the sSE and cSE. The cSE is a channel attention module. The specific process is as follows: the global average pooling method is used to convert the feature map from [C, H, W] to [C, 1, 1], and two 1×1 convolutions are used for information processing to obtain the C-dimensional vector. Thereafter, the sigmoid function is used for normalization to obtain the corresponding mask. Finally, the feature map that is calibrated by the information is obtained by channel-wise multiplication. The sSE module is a spatial attention module, whose implementation process is as follows: a 1×1 convolution operation is conducted directly on the feature maps and their dimensions are converted from [C, H, W] to [1, H, W]. Subsequently, the feature maps are activated with a sigmoid function to obtain a spatial attention

map, which is applied to the original feature map to complete the spatial information calibration. The SCSE is a parallel connection between the two modules. Input features from the sSE and cSE modules are added to obtain a more accurately calibrated feature map. Finally, the result is added to the input tensor as the output of the block. The structure of the SCSE is presented in Fig. 4. By introducing an attention mechanism, the network can focus on more critical information in the current task to solve the problem of information overload, and the efficiency and accuracy of the task processing can be improved.
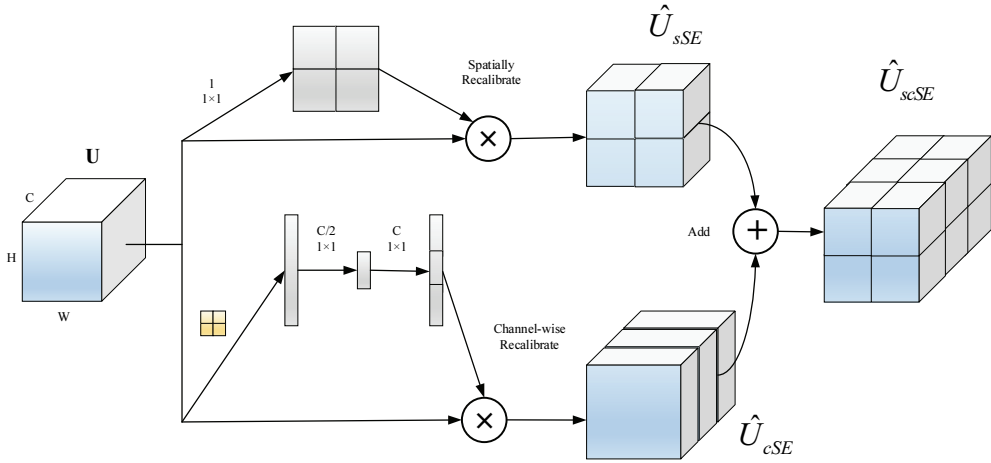
Fig. 4. SCSE module.

4. Experiments

We conducted a series of comparative experiments to demonstrate the effectiveness of the proposed recognition approach through comparisons with existing methods.

4.1 Experimental Configuration

The hardware of the experimental environment consisted of an NVIDIA Titan X graphics card, 128 G of running memory, and an Intel E5-2678V3 CPU. The software environment comprised an Ubuntu 16 system, Python 3.6 and the PyTorch 1.0 development environment. The experiments were conducted using the integrated development environment Python 3.6+PyTorch 0.4.0.

The experimental data were obtained from the CT-BI dataset. This dataset contained more than 1.2 million tobacco sales bills and corresponding statistical data, collected from different regions and dealers. These sales bills were saved in the JPG image format according to different tobacco types and statistical data were saved in a sheet in the .xlsx format. Each sample image contained the store name, monopoly license number, and tobacco commodity sales data and amounts. Fig. 5 presents several samples from the CT-BI dataset. In the experiments, the correct recognition rate of the algorithm was tested by identifying the sales and amounts of specific types of cigarettes in the image bill and verifying these with the data in the statistical table. The experimental results were analyzed quantitatively using the Top-1 to Top-5 error rate indexes.

Fig. 5. Image examples from CT-BI dataset.

4.2 Experiment I: Different Tobacco Types

In this experiment, we selected seven types of tobacco sales—namely, SUYAN, NANJING, ZHONGHUA, TAISHAN, HUANJINYE, WANBAOLU and LIQUN—to verify the accuracy of the proposed recognition method. The experimental results are listed in Table 1.

It can be observed from Table 1 that the recognition accuracies of the different tobacco were slightly different. This is mainly because the number and complexity of the Chinese characters in the names of these types of tobacco products differed, resulting in changes in accuracy when positioning the commodity names from the sample images with interference information. For example, the Chinese name for HUANJINYE contains three complex Chinese characters, whereas the Chinese name for ZHONGHUA contains only two simple Chinese characters. This resulted in a difference of 0.88% in the Top-1 index and 0.36% in the Top-5 index. However, Table 1 also indicates that the maximum standard deviation corresponding to each index from Top-1 to Top-5 was only 0.62, which reflects the stability of the proposed method.

Table 1. Experimental results for different tobacco (unit: %)

Item	Top-1	Top-2	Top-3	Top-4	Top-5
SUYAN	6.48	5.52	3.90	2.62	1.47
NANJING	5.94	4.32	3.05	2.29	1.27
ZHONGHUA	5.62	3.96	2.78	2.04	1.13
TAISHAN	5.34	3.78	2.67	1.85	1.09
HUANJINYE	6.50	4.95	3.31	2.56	1.49
WANBAOLU	6.42	4.97	3.40	2.31	1.26
LIQUN	5.46	4.39	3.14	2.01	1.16
Standard deviation	0.50	0.62	0.41	0.29	0.16

4.3 Experiment II: Different Recognition Methods

We compared the proposed method with several existing methods. The experimental data were the sales data of LIQUN tobacco. The experimental results are presented in Table 2.

It can be observed from Table 2 that the proposed method was superior to the other four methods in terms of the Top-1 to Top-5 indicators. For example, compared to the second-ranked Inception V3 method, the Top-1 of our method increased by 1.14%. Moreover, the Top-5 increased by almost 5% compared to the closest methods of Inception V3 and NasNet. The direct reason for this performance improvement was that our method integrated the multi-branch concept of Inception network, and provided the conv-block with a stronger learning feature ability by establishing more nonlinear transformations. Another possible reason was that the proposed method could focus on more critical information in the current task by introducing an attention mechanism, to improve the efficiency and accuracy of the task processing.

Table 2. Experimental results for different methods (unit: %)

Method	Top-1	Top-2	Top-3	Top-4	Top-5
Inception V3	6.60	5.29	3.87	2.44	1.58
NasNet	6.64	5.57	4.04	2.68	1.58
MobileNet	7.08	5.02	3.75	3.45	1.99
ResNet-18	8.91	8.49	6.54	5.98	3.12
Our method	5.46	4.39	3.14	2.01	1.16

5. Conclusion

In this study, as one of the typical applications of artificial intelligence technology in traditional industries, a new multi-branch residual framework was developed for the recognition of tobacco sales bills. A multi-branch residual network recognition model was designed and trained based on the geometric correction and edge alignment of input images. Finally, the effectiveness of the proposed approach was verified through comparative experiments on a large-scale tobacco sales bill dataset.

Acknowledgement

References

[1] K. Ohri and M. Kumar, "Review on self-supervised image recognition using deep neural networks," *Knowledge-Based Systems*, vol. 224, article no. 107090, 2021. https://doi.org/10.1016/j.knosys.2021.107090

[2] W. Ma, X. Tu, B. Luo, and G. Wang, "Semantic clustering based deduction learning for image recognition and classification," *Pattern Recognition*, vol. 124, article no. 108440, 2022. https://doi.org/10.1016/j.patcog.2021.108440

[3] Y. Zhou, "Vehicle image recognition using deep convolution neural network and compressed dictionary learning," *Journal of Information Processing Systems*, vol. 17, no. 2, pp. 411-425, 2021.

[4] Q. Chen, W. Zhang, K. Zhu, D. Zhou, H. Dai, and Q. Wu, "A novel trilinear deep residual network with self-adaptive Dropout method for short-term load forecasting," *Expert Systems with Applications*, vol. 182, article no. 115272, 2021. https://doi.org/10.1016/j.eswa.2021.115272

[5] J. Chen, J. Hu, and S. Li, "Learning to locate for fine-grained image recognition," *Computer Vision and Image Understanding*, vol. 206, article no. 103184, 2021. https://doi.org/10.1016/j.cviu.2021.103184

[6] Y. Zhao, C. Wang, J. Pei, and X. Yang, "Nonlinear loose coupled non-negative matrix factorization for low-resolution image recognition," *Neurocomputing*, vol. 443, pp. 183-198, 2021.

[7] Z. Zhang, P. Wang, H. Guo, Z. Wang, Y. Zhou, and Z. Huang, "DeepBackground: metamorphic testing for deep-learning-driven image recognition systems accompanied by background-relevance," *Information and Software Technology*, vol. 140, article no. 106701, 2021. https://doi.org/10.1016/j.infsof.2021.106701

[8] K. Huang, S. Li, W. Deng, Z. Yu, and L. Ma, "Structure inference of networked system with the synergy of deep residual network and fully connected layer network," *Neural Networks*, vol. 145, pp. 288-299, 2022.

[9] L. Yan, J. Feng, T. Hang, and Y. Zhu, "Flow interval prediction based on deep residual network and lower and upper boundary estimation method," *Applied Soft Computing*, vol. 104, article no. 107228, 2021. https://doi.org/10.1016/j.asoc.2021.107228

[10] T. Kim, K. Hong, and H. Byun, "The feature generator of hard negative samples for fine-grained image recognition," *Neurocomputing*, vol. 439, pp. 374-382, 2021.

Yuxiang Shan https://orcid.org/0000-0002-4703-924X

He received his M.S. degree in School of Computer Science and Technology from Zhejiang University in 2013. He is now an engineer of Information Center of China Tobacco Zhejiang Industrial Co. Ltd. His current research interests include image recognition and artificial intelligence.

Cheng Wang https://orcid.org/0000-0002-8620-6717

He received his B.S. degree in School of Human Resources Management from Nanjing Audit University in 2010. Since then, he joined Zhejiang Tobacco Industry Company as custom manager. In 2020, he joined the brand operation department, engaged in data operation and customer operation.

Qin Ren https://orcid.org/0000-0002-2911-0647

She received a bachelor's degree in marketing from Zhejiang Normal University in 2010. Since August 2011, she has worked in China Tobacco Zhejiang Industrial Co. Ltd. engaged in Tobacco Marketing and Internet Marketing Research, respectively.

Xiuhui Wang https://orcid.org/0000-0003-1773-9760

He received his master's degree and doctor's degree from Zhejiang University in 2003 and 2007, respectively. He is now a professor in the Computer Department of China Jiliang University. His current research interests include computer graphics, pattern recognition and artificial intelligence.