IJAMML 16:1 (2022) 97-112 September 2022
ISSN: 2394-2258
Available at http://scientificadvances.co.in
DOI: http://dx.doi.org/10.18642/ijamml_7100122259

AN IMPROVED METHOD OF CONVOLUTIONAL NEURAL NETWORK BASED ON IMAGE RECOGNITION OF CHINESE HERBAL MEDICINE

Zhendi Qin, Linxuan Li, Xuan Chen, Lie Chen, Jianqiang Gao and Fanbin Meng

School of Medical Information Engineering, Jining Medical University, Rizhao Shandong 276826, P. R. China

Abstract

Chinese herbal medicine (CHM) have a long history, a wide variety, and similar shapes, which lead to the difficulty and low efficiency of traditional CHM identification methods (mainly including original identification, morphological identification, microscopic identification, and physical and chemical identification). In view of the above problems, this paper proposes an improved method based on the traditional convolutional neural network VGG. First of all, this paper is no longer limited to the Conv-Batch Normalization (BN)-Max Pooling network structure model, but uses three (Conv-BN, Conv-Max Pooling and Conv-BN-Max Pooling) interlaced network structures. Second, we use GAP to connect

*Corresponding author.

E-mail address: fanbinmeng@qq.com (Fanbin Meng).

2020 Mathematics Subject Classification: 68T10, 68U10.
Submitted by Balan Sethuramalingam.
Received August 2, 2022; Revised August 8, 2022

the convolutional and fully connected layers (FC). The experimental results show that compared with the Conv-BN-Max Pooling network structure, using the above improved network structure can improve the recognition accuracy rate by 3.3% (Epoch 100), and finally achieve a correct recognition rate of 97.86% for 12 traditional Chinese medicine images, which proves that the proposed improved method has certain advantages compared with other network structures, and explores new ideas for the modern identification of Chinese medicinal materials.

Keywords: Chinese herbal medicine image recognition, convolutional neural network, neural network improvement method, VGG, computer vision.

---

1. Introduction

CHM is the native medicine in China. At the same time, the modernization of CHM is an important step to promote the development of CHM industry [1]. In recent years, with the rapid development of computer vision applications and promoting the implementation of CHM revitalization and development major projects not only makes people's demand for CHM show a blowout growth, but also promotes CHM development to become a national strategy.

Due to the serious impact of counterfeiting and deterioration of CHM in the market, the identification of CHM has become extremely important [2]. Of course, the correct use of CHM, including the identification and classification of CHM, is vital to the life safety of the patients [3]. However, traditional Chinese medicinal materials identification methods rely on long-term experience, the identification process requires professional tools and techniques, the identification steps are complex and cumbersome, requires expensive machinery and equipment and is difficult to widely use [4], mostly used by professionals. It is difficult to use traditional identification methods for many mixed and counterfeit CHM on the market. Therefore, it is urgent to develop a new method that can quickly and accurately identify CHM. The intelligent identification of CHM based on Convolutional Neural Network (CNN) has become the inheritance sword of CHM, which makes the identification of CHM more general.

Convolutional neural networks (CNNs) have been widely used to solve image recognition problems with great success, and researchers have explored many CNNs with different depth structures [5]. At present, the coupling fields of CHM identification and deep learning mainly include the following researches: Liu et al. [5] classified 50 kinds of CHM with complex natural background and made its accuracy reach 89.4%. Sun et al. [6] established a public database of CHM images and used it for research and application. Through experiments, an average recognition accuracy of 71% can be achieved for 95 kinds of CHM. Juei et al. [7] have come up with a real-time smartphone app that can not only perform image recognition based on CNNs, but also display details of the herb. Xu et al. [8] published the Multi-Attention Pyramid Network for Chinese Medicine Recognition, proposing a new Chinese Medicine Recognition Attention Pyramid Network (APN), which can adaptively model images of Chinese medicines with different characteristic scales [9]. With the advent of deep CNN models such as VGGNet, AlexNet [10], and many others have made many contributions to the identification of modern CHM based on CNNs, which will not be listed in this article. The above research content has realized the identification of modern Chinese medicinal materials, and has a good identification effect. In this paper, 12 kinds of CHM and 3161 CHM images are studied. Through the standardization and normalization of images, the preprocessed images are input into the improved network structure for multiple iteration training, so as to achieve a better classification effect.

In recent years, the development of deep learning has become more and more prosperous. This is not only the impetus of a large amount of data and powerful hardware, but also the collision of new ideas and new algorithms [11]. The collision of ideas is the most valuable. Based on this, this paper proposes an improved method of CNN based on image recognition of CHM, which provides a new idea for modern CHM identification methods. The experimental results show that the model has a good classification effect.

2. Materials

2.1. Dataset

2.1.1. Acquisition of datasets

The datasets used in the experiments are mainly from network queries and downloaded locally for preliminary data screening.

Figure 1. This figure shows an example of obtaining pictures through network query, and 3161 pictures of CHM were obtained in this experiment.

2.1.2. Preprocessing of datasets

The network obtains images for operations such as deduplication and deletion of invalid images, and it is noted that the picture background as a dataset should be as clean as possible, and as the recognized object should be centered as much as possible, so as to avoid invalid images after cropping. After obtaining the valid images, they are standardized and normalized, the dataset is divided into training sets and test sets at 8:2, and finally the images are fed into the first layer of the improved CNN at (256, 256, 3) using RGB channels.

2.1.3. Amplification of datasets

In order to increase the noise and enrich the feature space of the model in the later stage, the data set should be amplified (Table 1), and care should be taken to ensure that the data sets of various categories are balanced, and the order of magnitude gap should not exceed 102. Both undersampling and oversampling affect the correct recognition rate, which can be solved by using random noise, adding or subtracting images, setting weights and so on.

Table 1. The number of detailed classifications of the CHM dataset used for training. There are 12 categories of CHM, with an average of 264 pictures per category

Serial number	Dataset name	Number
1	Dangshen	251
2	Cordyceps sinensis	244
3	Wolfberry	274
4	Forsythia	291
5	Momordica grosvenori	254
6	Rose	266
7	Dandelion	300
8	Sunflower	250
9	Daisy	257
10	Tulips	272
11	Houttuynia cordata	250
12	Plantain	252

2.2. CNNs

The model in this article is an improvement on the traditional neural network VGG (Table 2), VGGNet is often used for transfer learning, and the model is highly portable [12]. The convolutional layers start at 64 in the first layer, and each convolutional layer sequentially increases the convolution(Conv) by a factor of 2 until it reaches 512 [13], sliding feature extraction using small convolutional nuclei of 3×3, set the match size to 32 and the learning rate is 0.0001. In 2011, the ReLU activation function was proposed, and experiments proved that it has good nonlinearity and effectively inhibits the disappearance of gradients [6]. In this paper, the ReLU activation function is used to increase the nonlinearity of the CNN, thereby transforming the low-order features into high-order features, and iteratively train the network structure and the introduction of GAP with three interleaved uses as its innovation points, so as to improve the generalization and robustness of the model, so that the model has better global expression ability and learning ability.

3. Methods

3.1. Network model structure

It should be noted that the tiniest and most easily ignored, the information obtained by Max Pooling after Conv is not complete, although the discarded information is not key information in our view, but the complete neural network is interconnected, as long as the CNN undergoes small changes, it will be cumulatively amplified in later training. Similarly, BN is performed after Conv, and although such whitening operations force data to be normalized, the more Convs there are, the more unexplained the features extracted by CNN become, and the whitening operations taken on each layer of Conv become unreasonable. Therefore, the balance of convergence speed and information integrity can be achieved through interleaved networks, and the generalization and robustness of the model can be improved, so as to achieve a higher correct recognition rate.

Use GAP to connect convolutional and FC. The introduction of GAP is equivalent to adding a layer of full connection to the network model in this paper, and improving the nonlinear expression ability and learning ability of the model. At the same time, GAP can greatly reduce the network parameters, solve the problem of slowing down the convergence speed caused by the introduction of too much noise in the amplification dataset, and improve the global expression ability of the network model.

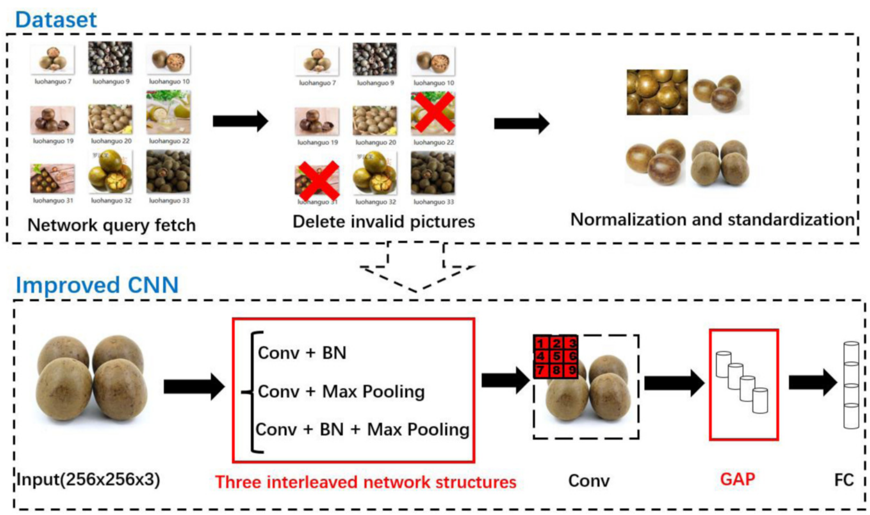

Figure 2. The diagram shows the overall experimental process, which is mainly divided into early data processing and later network model construction.

Table 2. The figure shows the model structure based on VGG network improved in this paper, using three interleaved network structures and GAP to connect convolutional and FC

Input (256,256,3)	Convolutional nuclei (3×3)	ReLU activate the function
Conv-64	Batch Normalization	–
Conv-64	Batch Normalization	Max Pooling
Conv-128	–	Max Pooling
Conv-128	Batch Normalization	Max Pooling
Conv-256	Batch Normalization	–
Conv-256	–	–
Conv-512	Batch Normalization	–
Conv-512	Batch Normalization	–
Conv-512	Batch Normalization	–
Conv-512	Batch Normalization	–
Global Average Pooling	–	–
Dense-1024	Batch Normalization	–
Dense-12	SoftMax	–

"–" is data missing symbol.

Table 3. The network structure for the comparison experiment is shown, using the Conv - BN - Max Pooling mode, which is compared before Conv-512 for ease of experimentation

Input (256,256,3)	Convolutional nuclei (3×3)	ReLU activate the function
Conv-64	Batch Normalization	Max Pooling
Conv-64	Batch Normalization	Max Pooling
Conv-128	Batch Normalization	Max Pooling
Conv-128	Batch Normalization	Max Pooling
Conv-256	Batch Normalization	Max Pooling
Conv-256	Batch Normalization	Max Pooling
Conv-512	Batch Normalization	–
Conv-512	Batch Normalization	–
Conv-512	Batch Normalization	–
Conv-512	Batch Normalization	–
Global Average Pooling	–	–
Dense-1024	Batch Normalization	–
Dense-12	SoftMax	–

"–" is data missing symbol.

3.2. Network structures interleaved

Both BN and Max Pooling in CNNs can accelerate the convergence speed of neural networks, where BN performs whitening operations (data normalization), that is, the data distribution of each layer of neural networks is consistent to improve convergence speed, while Max Pooling uses the dimensionality reduction of network parameters, that is, reduces network parameters to improve convergence speed.

From a specific point of view, a certain layer of CNNs can be divided into two types, one is a direct CNN, equivalent to a network with normal vision. The other is to use improved interleaved neural networks, which discard part of the data to a certain extent and the degree of discard of the data is different, equivalent to a myopic network, but the degree of myopia is different. When the degree of myopia is low, for a certain close-range neural network, the information that can be obtained by the two is the same, while for a relatively long-distance neural network, the direct CNN obtains more information, the sensory field is larger, and the noise is greater. But improved interleaved neural networks have a faster convergence rate due to the abandonment of some data.

We expect the layer neural network to get a faster convergence speed by discarding some of the data, but the actual operation cannot discard too much data, so we thought of building a convolutional layer with different convergence speeds, that is, using three interleaved network structure (Table 2) to achieve a complete balance of convergence speed and information.

3.3. GAP connects convolutional and FC

We know that although the neural network seems to be learning, it is actually cheating, no matter where to start training, the descent gradient will always choose the steepest, CNN will also choose the simplest texture map to solve the problem, so the later model training needs to increase the noise, of which the faster way to see the effect is to increase the data set. But in late training experiments, we found that too much noise caused the convergence rate to slow down. So we proposed using GAP to connect convolutional and FC. In this paper, the last layer of Conv-512 of the improved model outputs 512 feature maps of 32×32, and the 512 feature maps after the global average are obtained after GAP, and the feature maps output correspond to the spatial positions of the previous layer neural network. The use of GAP is equivalent to adding a layer of full connection to the network model, which can increase the complexity and nonlinear expression ability of the model, improve the learning ability of the model, and improve the correct recognition rate.

GAP has the following advantages over FC:

(1) Greatly reduce network parameters, speed up convergence, reduce overfitting, prevent parameter explosion, and eliminate fully connected black box operations.

(2) The spatial position information of the feature map will not be damaged, and the characteristic map through the GAP output is a one-to-one correspondence (Figure 3).

(3) The network structure enables higher expression ability. Although some data information is discarded, each texture map output has a global sensing field, and even the bottom layer of the network can feel the global information.

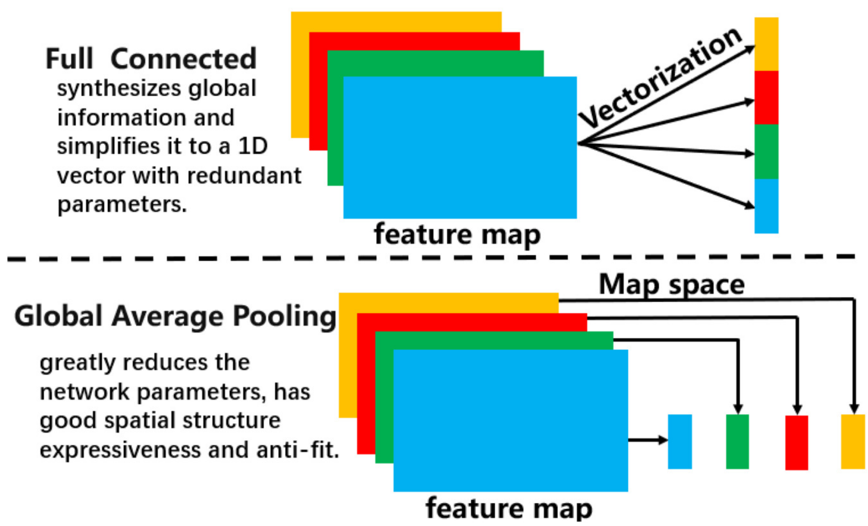

Figure 3. The figure shows the difference between full connection and GAP, which is more advantageous in comparison.

4. Results

4.1. Network structures interleaved

In order to facilitate the next experiment, the staggered comparison model is only compared before the Conv-256 layer of the improved model in this paper. The two network models (Table 2 and Table 3) are iteratively trained respectively, and the correct recognition rate of the staggered comparison model after Epoch 100 is compared. Reaching 83.05%, the correct recognition rate of the improved model in this paper increased by 3.3% to 86.35% (Figure 4). In addition, it can be observed that after Epoch 25, the training set and test set of the improved model are constantly oscillating and have not fully converged, while the staggered comparison model has converged on the training set and test set after Epoch 25, and the correct recognition rate is relatively high. This experiment shows that the improved model in this paper has a high correct recognition rate and has a large room for improvement in later training.

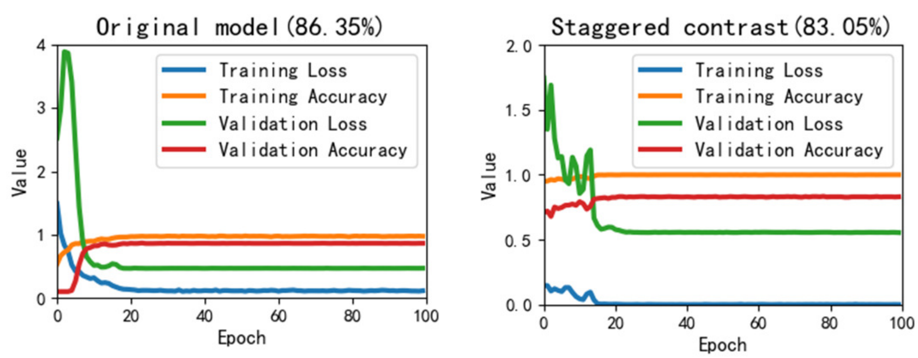

Figure 4. The left picture shows the training process of the original model Epoch100, and the right picture shows the training process of the staggered comparison model Epoch. Comparing the two, the correct recognition rate of the improved model is higher.

4.2. GAP connects convolutional and FC

In this paper, after multiple iterations and expansion of the data set using the network model in Figure 1, the correct recognition rate of the images of the 12 kinds of CHM studied finally reached 97.86%. When the data set is not amplified, the correct recognition rate of 50 Epoch times reaches 93.93%; after the first amplification of the data set, the correct recognition rate reaches 95.57% after 50 Epoch times; later, it is found that the imbalance of different types of CHM data sets will affect the correctness recognition rate, the second time to expand the data set, to ensure that the average number of pictures available for each CHM is 263. After continuing the Epoch for many times, the test set converges, and the correct recognition rate is 97.86% (Figure 5).

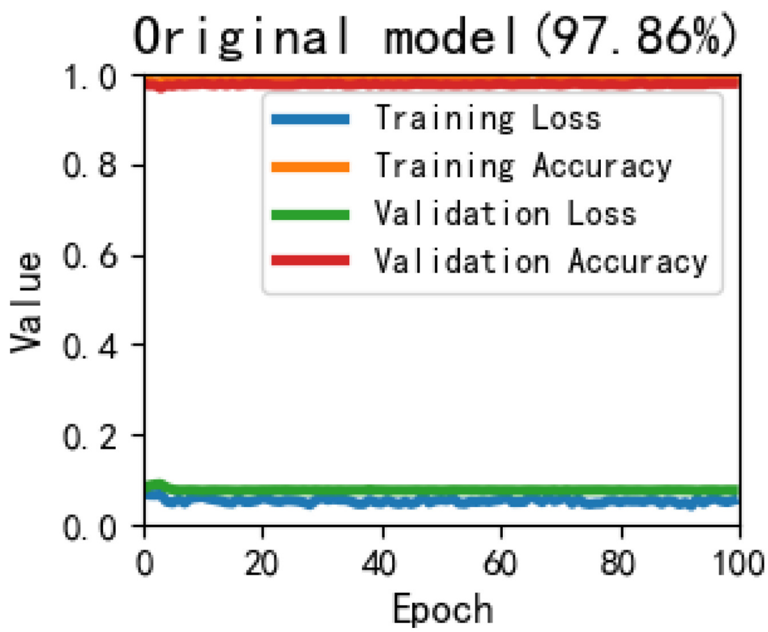

Figure 5. The figure shows the final result obtained after multiple training and amplification dataset experiments, which has a recognition accuracy rate of 97.86%.

5. Discussion

In this paper, the CNN improvement method proposed for the problem of difficulty and low efficiency of identification of CHM is proposed, that is, the improved interleaved network structure is used to achieve the balance of model information integrity and convergence speed, use GAP to connect convolutional and FC to achieve a balance of nonlinear expression capabilities and convergence speeds. Experimental results show that the interleaved network structure in this paper can be improved by 3.3% compared with Conv-BN-Max Pooling network structure, and the initial training has a large improvement space. The correct recognition rate of the late introduction of noise after multiple iterations was trained to reach 97.86%, which is suitable for the dataset with moderate sample size and the correct recognition rate in the later stage.

6. Conclusion

With the rapid development of the digital society, rich data information also provides more attempts and innovations for this article. Experiments also show that the improved method of network structure proposed in this paper has better recognition performance, but the model still has a lot of room for improvement, considering whether the identification of CHM is correct is crucial for life safety. How to further improve more the correct recognition rate of CHM classification, shortening the training time and trying more network models are our next research directions.

Acknowledgements

This work is partly supported by the Doctoral Research Foundation of Jining Medical University under Grant No. 600529001 and the Teachers' research of Jining Medical University under Grant No. JYFC2019KJ014, and the 2022 undergraduate innovation training program of Jining Medical University under Grant Nos. cx2022002z, cx2022009z, and cx2022022z.

References

[1] Qing Liu, Xiaolong Zhao, Hongping Yang, Dongfeng Li, Weijun Ling, Feiping Lu and Yuxiang Zhao, Image feature extraction and retrieval of the Euler number to Chinese herbal medicine based on PCNN, Journal of Physics: Conference Series 1335 (2019); Article 012016.

DOI: https://doi.org/10.1088/1742-6596/1335/1/012016

[2] Tianhao Li, Fengyang Sun, R. Sun, Lin Wang, Meihui Li and Huawei Yang, Chinese herbal medicine classification using convolutional neural network with multiscale images and data augmentation, 2018 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), IEEE, 2018.

DOI: https://doi.org/10.1109/SPAC46244.2018.8965566

[3] Rong-Rong Chen and Ying-Jun Chen, Intelligent Identification of Traditional Chinese Medicine Materials Based on Multi-feature Extraction and Pattern Recognition, Proceedings of 2019 International Conference on Modeling, Analysis, Simulation Technologies and Applications (MASTA 2019), Ed. Atlantis Press, (2019), 399-404.

DOI: https://doi.org/10.2991/masta-19.2019.66

[4] Changwei Cai, Shuangrong Liu, Lin Wang, Bo Yang, Mengfan Zhi, Rui Wang and Weikai He, Classification of Chinese herbal medicine using combination of broad learning system and convolutional neural network, 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), 2019.

DOI: https://doi.org/10.1109/SMC.2019.8914437

[5] Shupeng Liu, Weiyang Chen and Xiangjun Dong, Automatic classification of Chinese herbal based on deep learning method, 2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), 2018.

DOI: https://doi.org/10.1109/FSKD.2018.8687165

[6] Xin Sun and Huinan Qian, Chinese herbal medicine image recognition and retrieval by convolutional neural network, PloS One 11(6) (2016); Article e0156327.

DOI: https://doi.org/10.1371/journal.pone.0156327

[7] Juei-Chun Weng, Min-Chun Hu and Kun-Chan Lan, Recognition of easily-confused TCM herbs using deep learning, MMSys'17: Proceedings of the 8th ACM on Multimedia Systems Conference (2017), 233-234.

DOI: https://doi.org/10.1145/3083187.3083226

[8] Yingxue Xu, Guihua Wen, Yang Hu, Mingnan Luo, Dan Dai, Yishan Zhuang and Wendy Hall, Multiple attentional pyramid networks for Chinese herbal recognition, Pattern Recognition 110 (2020).

DOI: https://doi.org/10.1016/j.patcog.2020.107558

[9] Haijian Ye, Hang Han, Linna Zhu and Qingling Duan, Vegetable pest image recognition method based on improved VGG convolution neural network, Journal of Physics: Conference Series 1237(3) (2019); Article 032018.

DOI: https://doi.org/10.1088/1742-6596/1237/3/032018

[10] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke and Andrew Rabinovich, Going deeper with convolutions, 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015).

[11] Peng Jiang, Yuehan Chen, Bin Liu, Dongjian He and Chunquan Liang, Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks, Advanced Optical Imaging for Extreme Environments 7 (2019).

DOI: https://doi.org/10.1109/ACCESS.2019.2914929

[12] Karen Simonyan and Andrew Zisserman, Very deep convolutional networks for large-scale image recognition, The 3rd International Conference on Learning Representations, CoRR abs/1409.1556 (2014).

[13] Anh H. Vo, Hoa T. Dang, Bao T. Nguyen and Van-Huy Pham, Vietnamese herbal plant recognition using deep convolutional features, International Journal of Machine Learning and Computing 9(3) (2019), 363-367.

DOI: https://doi.org/10.18178/ijmlc.2019.9.3.811

■