Classification of Tongue Color Based on CNN

Jun Hou, Hong-Yi Su, Bo Yan, Hong Zheng, Zhao-Liang Sun, Xiao-Cong Cai
Key Lab of Intelligent Information Technology
Beijing Institute of Technology
Beijing, 100081, China
e-mail: {yijifanhua, henrysu, yanbo, hongzheng }@bit.edu.cn, sunzl8417@163.com, 2120160979@bit.edu.cn

***Abstract*—Tongue manifestation is one of the most significant basic criteria for the diagnosis of Traditional Chinese Medicine (TCM). And tongue color recognition with high accuracy will contribute to the efficiency of TCM diagnosis. The drawbacks of traditional tongue diagnosis methods are that the features need to be designed artificially. While the feature acquisition from the deep learning is a process of simulating the brain activities and learning behaviors of human beings, and it has achieved fruitful results in many aspects, including image classification, face recognition, objects detection and so on. Therefore, the method combining deep learning with tongue color classification is proposed. First, the preprocessed and enhanced images are created as a tongue image database. Then, the parameters of the traditional network are modified for tongue color classification. Finally, it is more targeted to use our own model to fine-tune our neural networks. The experimental results show that this method is more practical and accurate than the traditional one.**

Keywords-classification of tongue color; convolutional neural network; CaffeNet

I. INTRODUCTION

Traditional Chinese Medicine (TCM) has a long history in the treatment of various diseases in East Asian countries and it is also a complementary and alternative medical system in western countries. TCM diagnosis is based on the information obtained from four diagnostic processes, which include observation, auscultation, question and pulse-taking. Among which the most common task is inspecting the tongue. Doctors can judge the health condition of patients by examining the color, shape, textures and other features of their tongues. Together with other health information, the possible causes of the pathogen can be terminated and then treated.

However, diagnosis of the tongue usually depends on the subjective judgment of the doctor. The environmental factors, such as different light sources and brightness, have great influence on doctors to obtain good diagnostic results through the tongue. Besides, the examination outcome in traditional tongue diagnosis could not be described scientifically and quantitatively. Hence, it is urgent for TCM tongue diagnosis to develop a modern medical system, which ought to be in the direction of the leading modern standardization of science and technology, objectivity, quantification, automation and exhibition. In order to address the issues that the details and characteristics of the tongue can easily be overlooked, a method of automatic tongue color classification based on neural network is presented. This neural network called Convolutional Neural Network (CNN) can accurately examine all the features in pictures and send the right one to computers. So we use CNN to study the details and characteristics of tongues. What's more, computer can effectively avoid the errors of visual acuity or light.

In this paper, we mainly make two contributions: 1) we constructed tongue images dataset of our own, which contain about 1500 photos, and then horizontal flip technique is used to double the dataset. 2) we modify the traditional CaffeNet [1] model and propose a new training method.

The rest of this paper is organized as follows. Section 2 describes the related preparatory work; Section 3 introduces our methods; Section 4 reports a series of experiments to evaluate the performances of tongue color classification; and finally Section 5 makes a conclusion based on the researches and experiments.

II. RELATED WORK

In recent years, studies on the combination of traditional Chinese medicine and computer science about tongue color classification have been widely emphasized and implemented. Before the advent of neural networks, the most common method is that images are categorized according to the color space. With the emergence of BP neural network, more and more people began to classify tongue color by it.

A. Based on Color Space

The traditional tongue color recognition mainly uses the color space feature as classification algorithm. Chen Songhe et. al, (2007) used the CIELAB and the LCH color space, and there were the fusiform 3D shape distribution of different tongue color datum, and there is also the satisfied distinguish in the a*-b* plane, the L*-a* plane, the C*-H* plane and the L*-C* plane different tongue color spatial distribution[2]. Ni Hao et. al,(2011) used HSV color space and reasonable quantitative color, and then calculated the color histogram as the tongue image feature to retrieval[3], which is the most common method. The other is based on the statistical analysis of the local pixel point RGB diagnostic results for quantitative analysis of the tongue color by Beijing Institute of traditional Chinese medicine (Beijing Chinese medicine hospital). The limitation of this approach is that its effect is highly dependent on the design of the

978-1-5090-3619-6/17/$31.00 ©2017 IEEE

feature, and the good features can accurately extract the favorable information in the picture. While the relatively poor characteristics are not available to extract useful information for training.

B. Based on Traditional Neural Networks

The second general direction is to use BP neural network. Wu Xia et. al, (2007) applied the SVM and BP(Back Propagation) neural network to achieve TIR(Tongue image recognition)[4]. The other is based on BP-ANN with Self-Adaptive Network Structure algorithm for the purpose of use classifying the different tongue colors based on spectral data of tongue[5]. Although many research institutions have made deep studies in this area and obtained some achievements, they have not proposed an easy way to collect data set and obtained a high rate of accuracy.

In our course of the study, many details of the characteristics in images are ignored and a large number of network parameters can easily lead to overfitting, which are the biggest vulnerability of the traditional methods.

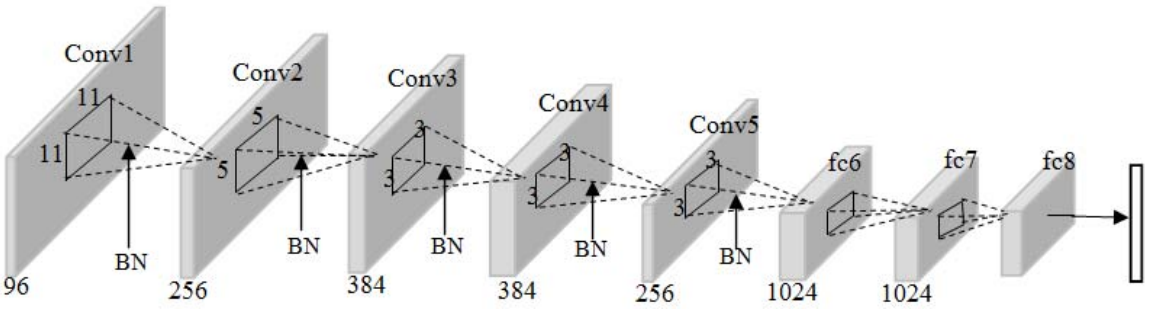

Figure 1. Our model. Based on CaffeNet, we modified the network structure of the model, and add the BN layer between convolutional layer and scale layer, to prevent overfitting. Because of the limited number of tongue images, some characteristics of the tongue will be gradually weakened until disappeared in the learning process. So we reduced the output numbers of the fc6 layer and fc7 layer

III. OUR METHODS

The entire process of developing a tongue color classification model using CNN can be described as follows: 1) the pretreatment and augmentation of tongue images. 2) modification of the structure of CNN. 3) using a special method to train CNN. The whole process is divided into three stages in the following subsections, starting with the collection of images for the classification process using the convolutional neural network.

A. Dataset

Pictures of the tongue is different from the ordinary images, it is very difficult to photograph and arrange. We took several months to collect about 2000 photos, removing some unintelligible and inappropriate images, with the remaining of 1,500. The biggest advantage of our dataset is the universality of our photos, which is obtained without any manual intervention and taken randomly.

Images in the dataset are grouped into six different classes, but it distributes unevenly. There are many classic datasets that are derived from non-uniform ones. (e.g: Oxford flowers datasets, Caltech-UCSD Birds-200-2011). But there is no practical significance for us to obtain results of experiments from the non-uniform datasets. That is to say, the pictures that we can choose to use among 1,500 in total are even less.

In order to distinguish healthy tongue color from diseased ones, one more class is added in the dataset. It only contains images of healthy tongues. Next, we enrich the dataset with augmented tongue images. The main goal of this study is to train the network to learn the features that distinguish one class from the others. Therefore, the more augmented images we use, the more chance for the network to learn the appropriate features will be increased [6]. In addition, it also helps to reduce overfitting during the training period. The image augmentation contained one of several transformation techniques including horizontal reflections, affine transformation and perspective transformation. In this paper, we use the first method to augment datasets.

Table 1-3 shows the accuracy of different categories with the number of original images and number of augmented images for every class used as training and testing dataset for the tongue color classification model. According to the theory of traditional Chinese medicine and actual collection situation, the six categories of tongue color, ranging from pale white, light red, red, sharp red, side red and dark red.

B. Image Preprocessing

To obtain better feature extraction, the final image of the dataset intended for use as a convolutional neural network classifier is preprocessed for consistency [7]. The first step is to extract contours from the tongue image using a mature contour extraction algorithm. This operation can reduce the interference from background information. Besides, it is difficult to classify the tongue photos into a specific category because many of them are in the critical value, so we need to enhance the features with some pretreatments. The RGB values of the same classification are in a certain range, so we can calculate the ranges by the standard tongue colors to normalize the critical value to a specific value. Because of that some tongue colors are not in the scope of our tongue color calibration. In this case, we can retain the unidentified sites to deal with the tongue images. This paper has adopted the sample template shown in Fig. 2. Picture (b) and (d)

shows the original pictures of tongue. Picture (a) and (c) tells the tongue image keeps the original color after dealt with the pretreatment.

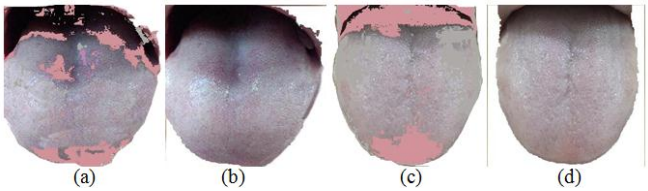

Figure 2. (a) Keep the original color (b) Original picture (c) keep the original color (d) Original picture

C. Research on Traditional CNN

It is known that the traditional tongue color classification feature extraction mainly depends on the individual thoughts of experimenters, leading to that there is no quantitative, qualitative criteria. So, we use CNN to classify tongue color. The CNN has a hierarchical structure: input layer, convolutional layer (conv layer), relu layer, pooling layer, fully connected layer (fc layer). The convolutional layer is the indispensable layer of the convolutional neural network. A number of learnable kernels that run through the whole input volume [5] are the core components of these layer. Every convolutional layer has N maps, N_x and N_y, and a kernel of size X_x which is shifted over the certain region of the input image [7]. The primary role of the convolutional layer is that extract features. Rectified Linear Units (ReLU) adaptively learns the parameters of rectifiers and improves the accuracy of negligible extra computational cost [7]. It is defined as

$$f(z_i) = \max(0, z_i) \tag{1}$$

Another essential layer is pooling layer, which belongs to nonlinear downsampling. Pooling operation gives us the form of translation invariance [8]. The pooling layer can be used to calculate the local sensitivity and extract feature of the computing layer. In this way, the structure of the two feature extraction reduces the feature resolution and diminish the number of parameters that need to be optimized.

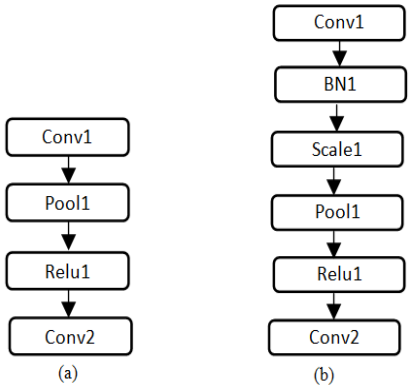

Figure 3. (a)Original module of CaffeNet between the two convolutional layers (b) Modified module of CaffeNet between the two convolutional layers

CNN can be used to identify two or three-dimensional images of displacement, scaling, and other forms of distortion-invariance. The parameters of CNN's feature extraction layer are learned through training data, so it avoids artificial feature extraction, and learns from training data. Besides, neurons of the same feature map share the same weight, which reduces the network parameters and it is an advantage, compared with the fully connected network.

D. Our CNN

In this paper, we use the CaffeNet [1] to classify tongue color. CaffeNet is a deep CNN which has multiple layers that progressively compute features from input images [9]. Specifically, the network contains eight learning layers and five convolutional and three fully connected layers [10], this framework was used as the basic network, along with the set of weights learned from our own model. Due to the limitation of the pictures, it is difficult to grasp the direction of feature learning in the course of training, and it is prone to overfitting. So, the modified model is showed as Fig. 1.

Our first improvement is that we add BN layer and reduce network parameters to avoid overfitting. As we can see from Fig. 3, the left is original module, the right is modified module which adds the BN (Batch Normalization) layer between the two convolutional layers. This layer draws its strength from making normalization a part of the model architecture and performing the normalization for each training mini-batch [13] to prevent the gradient disappear or explosion, speed up the training of CaffeNet. In spite of the recently-introduced technique, called "dropout" [11], consists of setting to zero the output of each hidden neuron with probability 0.5[12], there is no obvious effect. BN layer is added to take a step towards reducing internal covariate shift, and in doing so dramatically accelerates the training of deep neural nets [13]. Besides, reducing the output numbers of fc (InnerProduct layer) is useful for the training of our network. The original output numbers of InnerProduct layer (fc6 and fc7) is reduced from 4096 to 1024 to avoid overfitting. After preprocessing images and improving model, the convergence speed becomes faster and the overfitting is basically solved.

The next is that we use a new method of high learning efficiency to train our networks. When the training loss value minus the test loss value is greater than a certain threshold, we stop training, modify the parameters and use the generated model to fine-tune the modified network. This method has dual advantages: 1) compared with the lack of pertinence in the traditional model, our new model is pertinent to this network. 2) the learning rate is dynamically changed to avoid overfitting as well as speed up the convergence. The results of all these improvements are shown in the Fig. 4 and 5.

TABLE I. THE TABLE SHOWS THE SIX CATEGORIES AND DIFFERENT PREPROCESS BASED ON THE SAME DATASET.

Number of categories	*Original pictures(per class)*	*Preprocess (per class)*
6- classification	Training:128 Testing:24 Accuacy:0.77	Training:128 Testing:24 Accuacy:0.83

TABLE II. THE TABLE SHOWS THE FIVE CATEGORIES AND DIFFERENT PREPROCESS BASED ON THE SAME DATASET.

Number of categories	*Original pictures(per class)*	*Preprocess (per class)*
5- classification	Training:128 Testing:24 Accuacy:0.80	Training:128 Testing:24 Accuacy:0.87

TABLE III. THE TABLE SHOWS THE DIFFERENT NUMBER OF DATASET BASED ON THE SAME CATEGORIES.

Number of categories	*Original pictures(per class)*	*Preprocess (per class)*
4- classification(1)	Training:128 Testing:24 Accuacy:0.79	Training:128 Testing:24 Accuacy:0.88
4- classification(2)	Training:256 Testing:48 Accuacy:0.89	Training:256 Testing:48 Accuacy:0.93
4- classification(3)	Training:440 Testing:88 Accuacy:0.82	Training:440 Testing:88 Accuacy:0.91

To quicken the whole experimental progress, we use Caffe [1] platform and GPU training networks, which can greatly shorten the train and test time and have a very important application significance.

IV. EXPERIMRNT AND RESULTS

This experiment was performed using Tesla K20C. We built a convolutional neural network using Caffe [1], a deep learning framework made with expression, speed, and modularity in mind.

We can see that many methods are mentioned in section 2. But there is no condition to do some comparative experiments because of different classification standard and inconsistent datasets.

According to the suggestion of the TCM experts, three group experiments are conducted included 6-classification (light white, light red, red, sharp red, dark and edge sharp), 5-classification (light white, light red, red, sharp red and dark) and 4-classification (light white, light red, red and sharp red). The number of labeled training data and the test data are shown at Table I-III.

In Fig. 4, there are three sets figures, the first set represents the right classification rate of 6-classification testing samples, the second is the 5-classification, and the last is the 4-classification. As learn from the overall trend of the Fig. 4, with the different categories, we can see that along the increase of the number of categories, the accuracy is lower. The purpose of this group of experiments is to find out which color is difficult to recognize. From the results, we can see that it is different for computer to distinguish between light white, dark and light red. The solution to this problem is to increase the datasets.

Fig. 5 shows the accuracy of four categories with the number of original images, augmented images and more images for every class used as training and testing dataset for the tongue color classification. We can see that the number of dataset is the decisive factor for accuracy. But, by comparing the second and third columns, we can see the problem that accuracy is reduced with the increase of dataset. We compared labels and images one by one, finding that the gap within the class is growing when the datasets increased. So when the data set increases, the accuracy of tongue color classification has declined. So it is necessary to establish a large tongue image database.

From the Fig. 4 and 5, the effect of data enhancement is very obvious. It turns out that the modification of the network is still valid for tongue color classification.

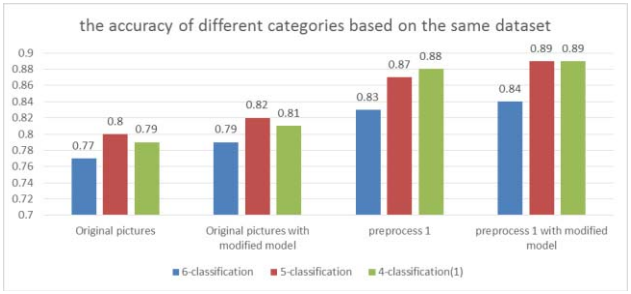

Figure 4. The accuracy of different categories based on the same dataset

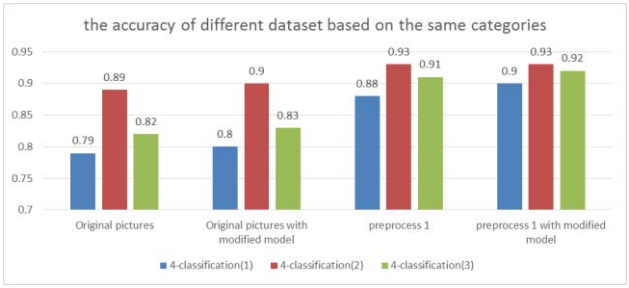

Figure 5. The accuracy of different dataset based on the same categories

V. CONCLUSION

Many methods are used to study the classification of tongue color, but still, this research field is lacking. In this paper, a new approach of using convolutional neural network method was explored in order to classify tongue color from tongue images. The complete process was described, respectively, from collecting the tongue images used for training and testing to image preprocessing and augmentation and finally the procedure of training our own CNN and fine-tune by our model. The experimental results show that as the dataset increases the accuracy becomes higher on the same categories and the number of datasets is critical to different categories. In the future, we will continue to collect photos to improve our datasets. By extending this research, the authors hope to achieve a valuable impact on TCM development.

REFERENCES

[1] Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional Architecture for Fast Feature Embedding[C]. acm multimedia, 2014: 675-678.

[2] CHEN Song-he. Study on tongue color analysis of digital tongue [D]. Beijing University of Chinese Medicine, 2007.

[3] NI Hao. Study on Color - and Texture - Based Retrieval Technology of Chinese [D]. Guangdong University of Technology, 2011.

[4] WU Xia. Pattern Recognition of Tongue Colors [D]. Nankai University, 2007.

[5] G.Montavon,M. L. Braun, and K.-R.M ¨ uller, "Kernel analysis of deep networks," The Journal of Machine Learning Research, vol.12, pp. 2563–2581, 2011.

[6] Sladojevic S, Arsenovic M, Anderla A, et al. Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification.[J]. Computational Intelligence and Neuroscience, 2016.

[7] C. Ciresan Dan, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber, "Flexible, high performance convolutional neural networks for image classification," in Proceedings of the International JointConference onArtificial Intelligence (IJCAI'11), vol.22, no. 1, pp. 1237–1242, 2011.

[8] A. Romero, Assisting the training of deep neural networks with applications to computer vision [Ph.D. thesis], Universitat de Barcelona, Barcelona, Spain, 2015.

[9] A. K. Reyes, J. C. Caicedo, and J. E. Camargo, "Fine-tuning deep convolutional networks for plant recognition," in Proceedings of the orking Notes of CLEF 2015 Conference, 2015, http://ceur-ws.org/Vol-1391/121-CR.pdf.

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, Imagenet Classification with Deep Convolutional Neural Networks, Advances in Neural Information Processing Systems, 2012.

[11] G.E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R.R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580, 2012.

[12] Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks[J]. Advances in Neural Information Processing Systems, 2012, 25(2):2012.

[13] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[C]. international conference on machine learning, 2015: 448-456.