A Tongue Image Segmentation Method Based on Enhanced HSV Convolutional Neural Network

Jiang Li, Baochuan Xu(✉), Xiaojuan Ban, Ping Tai, and Boyuan Ma

University of Science and Technology Beijing, Beijing, China
790966919@qq.com, xubc_2005@163.com,
banxj@ustb.edu.cn, pingkflod@sina.com, 793960328@qq.com

Abstract. In the procedure of the Chinese medical tongue diagnosis, it's necessary to carry out the original tongue image segmentation to reduce interference to the tongue feature extraction caused by the non-tongue part of the face. In this paper, we propose a new method based on enhanced HSV color model convolutional neural network for tongue image segmentation. This method can get a better in tongue image segmentation results compared with others. This method also has a great advantage over other methods in the processing speed.

Keywords: Tongue image segmentation · Convolutional neural network · Snake model

1 Introduction

Tongue diagnosis is a unique diagnosis method of the Chinese traditional medical science. Doctors can get the information of patients' illness conditions by observing their tongue features [1]. It's important to diagnose the patient's health condition and make a prescription. Modern research shows that tongue diagnosis, as a unique diagnosis method to identify the functional status of human body, has great value in diagnostics, so it's necessary to inherit and carry forward tongue diagnosis.

The automation process of tongue diagnosis based on tongue image analysis is to analyse tongue features by extracting them from tongue images captured by digital image device. So, the tongue body segmentation is the first step of tongue image analysis and accurate tongue body segmentation can reduce interference caused by non-tongue body of the face to tongue feature extraction [2, 3].

Image segmentation is the hot research field all the time, and there are several methods which can be used in tongue image segmentation. Currently the method applied to tongue body segmentation are mainly traditional image processing algorithm such as threshold segmentation method, region growing method, cluster partition method, watershed transform method, edge detection segmentation method and snake model method [4–7]. The snake model method has better performance than the others. But all of these methods of tongue segmentation still have their imperfections because it's hard to precisely divide tongue and lips as their colors are similar.

Y. Luo (Ed.): CDVE 2017, LNCS 10451, pp. 252–260, 2017.
DOI: 10.1007/978-3-319-66805-5_32

In some of the collected images, the boundary of the human body's facial region and the tongue body region is not obvious, and the color of the face area and the tongue area is relatively similar. In this case, the threshold segmentation method can't obtain accurate segmentation results, and the effect of segmentation method based on snake model is bad because the energy function of snake model can't converge. Besides the snake model method can't use RGB image directly and need to convert RGB image to gray image in advance.

In recent years, deep learning has developed rapidly and deep neural network in image understanding has achieved good results. The convolutional neural network is an application of the deep learning method in image processing field [8, 9]. In this paper the enhanced HSV convolutional neural network based method for tongue body segmentation was proposed for the first time. Set the enhanced tongue image as the input of convolutional neural network and calibrate the image of tongue body as the tag of tongue, then use convolutional neural network to conduct training. The convolutional neural network finally output a binary image of the tongue body which has same size with the original tongue image. The white area of the binary image is the tongue body and the black area is the non-tongue body of the face. Merge the output binary image with the original image, the resulting image is the image of tongue body extracted by the convolutional neural network.

Specially, we did some preprocessing for the image in order to obtain more exact features. Firstly, we convert image from RGB color model to HSV color model because that the HSV color model can show the tongue edge more clearly. Then we use image enhancement to strengthen the edge. Our method which used enhanced HSV color model based on convolutional neural network obtain a good effect in the tongue image segmentation.

2 Enhanced HSV COLOR Model

2.1 HSV Color Model

Generally, the extracted image data was RGB color model. The RGB color model is an additive color model in which red, green and blue light are added together in various way to reproduce a broad array of colors. And the HSV color model is another representation of the image attribute value. HSV stands for hue, saturation, and value, and is also often called HSB (B for brightness). In our experiments, we convert image from RGB color model to HSV color model because we found that HSV color model was easy to show the tongue rough edge.

As we can see in the Fig. 1, the hue channel of the tongue image shows a distinct edge between the tongue bottom edge with the face and the saturation channel of the tongue image shows a distinct edge between the tongue top edge with the mouth.

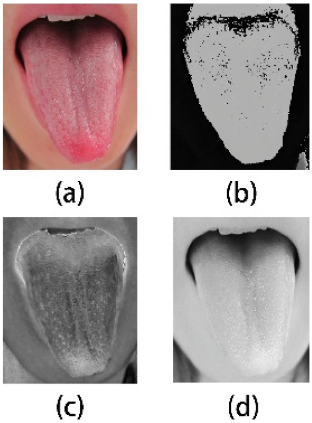

Fig. 1. Split RGB image into HSV channel (Color figure online)

2.2 Image Enhancement

After convert RGB color model to HSV color model, we Contrast Limited Adaptive histogram equalization (CLAHE) [10, 11] to enhance the tongue image in order to obtain a clearer edge.

AHE algorithm is a histogram method in order to improve image contrast locally by the cumulative distribution function. But image is often distored by AHE because the image local contrast was improved too much. In order to solve this problem, we limit the local contrast. In the histogram equalization, the relation of the mapping curve T and the cumulative distribution function (CDF) is shown as in (1).

$$T(i) = \frac{\mathrm{M}}{\mathrm{N}} CDF(i) \tag{1}$$

The cumulative distribution function (CDF) is the integral of the histogram of the gray scale so we can limit the slope of the CDF to limit the contrast.

$$\frac{d}{di} CDF(i) = Hist(i) \tag{2}$$

We cut the histogram obtained in the subblock so that the amplitude is lower than a certain upper limit, and the cut-off portion can't be discarded. We also distribute the cut value evenly over the whole gray interval to ensure that the total area of the histogram unchanged (Fig. 2).

Contrast Limited Adaptive histogram equalization (CLAHE) can't be applied to multi-channel image directly so we apply CLAHE to each channel of HSV image and then merge these three channels. The tongue image was enhanced obviously by this method and our result was shown in Fig. 3.

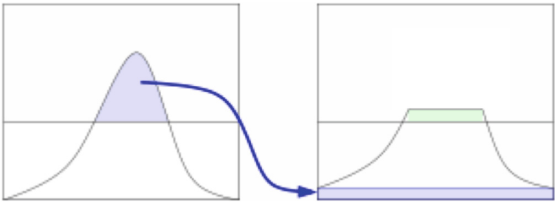

Fig. 2. CLAHE

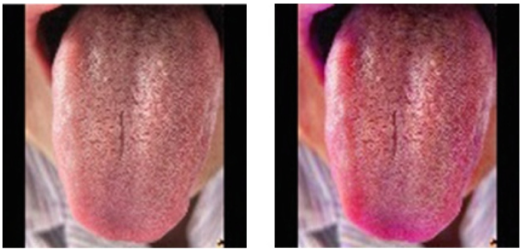

Fig. 3. Enhance image by CLAHE

3 Convolutional Neural Network

The convolutional neural network is a multi-layer artificial neural network which can extract features from input data automatically and it was always used to classify images. In convolutional neural network, each layer is composed of many 2-dimension planes and each plane is composed of many independent neurons. Neurons in two adjacent layers were interconnected but neurons in a same layer were not connected [12]. In the convolutional layer, neurons are locally connected with the next layer's neurons and this structure can reduce the number of the weight params. But in the full-connected layer, all the features are connected to the full-connected layer neurons.

3.1 Convolutions

Natural images have the property of being stationary, meaning that the statistics of one part of the image are the same as any other part. Therefore, the features that we learn at one part of the image can also be applied to other parts of the image, and we can use the same features at all locations. Having learned features over small patches sampled randomly from the larger image, apply this learned feature detector anywhere in the image. Specifically take the learned features and convolve them with the larger image, thus obtain a different feature activation value at each location in the image [13].

3.2 Max Pooling

Features that are useful in one region are also likely to be useful for other regions because of the stationarity property of image. Pooling is an operation to aggregate statistics of these features at various locations. Divide the convolutional features of image into pooling area by defining pooling size, then obtain the pooled convolutional features by meaning pooling or max pooling. Pooling is effective to reduce dimension of the features and improve results (less over-fitting).

3.3 Rectified Linear Unit

Instead of sigmod and tanh, we choose ReLU [14] as activation to model a neurons output. These saturating nonlinearities are much slower than the non-saturating nonlinearities. CNNs with ReLU are trained several times faster than with tanh on CIFAR-10 dataset. Moreover, when executing back propagation, sigmod and tanh will cause gradient vanishing because of saturating, which leads to information loss. ReLU makes some output of neurons zero that causes sparse of network, which prevents overfitting.

3.4 Dropout

Dropout [15] is a recently-introduced technique preventing overfitting when train data is not that much. Dropout sets part of the output of each hidden neuron to zero. These neurons do not contribute to the forward pass and do not participate in back propagation. Every time input is presented, parts of the hidden neurons are "deleted" randomly. The architecture trains lots of networks with some neurons and learns more robust features that are useful. Most correct subsets of networks impact on final result, but other subsets are abandoned. A 50% dropout rate is employed to discourage the co-adaption of feature detectors, which is proved better. Figure 4 shows the operation mode of dropout.

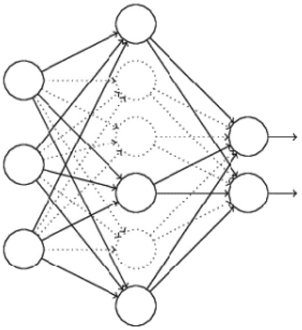

Fig. 4. Half of hidden neurons are dropped out

3.5 Full-Connection

The last layer before output layer in the convolutional neural network is full-connected layer. Each neuron in the layer connected to all the features which was generated by the convolutional layers.

4 Enhanced HSV Convolutional Neural Network

4.1 Convolutional Neural Network Model

In the experiment, we used a ten-layer convolutional neural network to train the tongue image. The convolutional neural network structure was shown as Fig. 5.

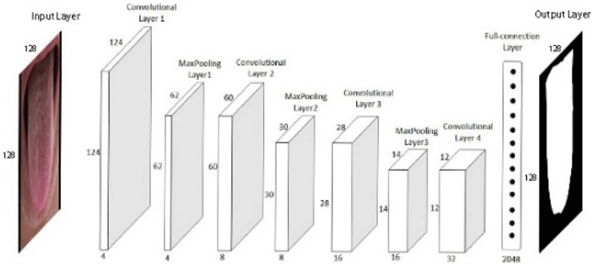

Fig. 5. Convolutional neural network structure (Color figure online)

From Fig. 5, we can see that the input layer is a marked tongue image which is a color picture. The 2, 4, 6 and 8 layer are the convolutional layers and the second lay uses 5 * 5 convolutional kernel and others use 3 * 3 convolutional kernel. The 3, 5 and 7 layer are the max pooling layers and their pooling size is 2 * 2. The 9 layer is the full-connection layer with 2048 neurons. The 10 layer is the output layer with 128 * 128 neurons and we can transform the outputs to a binary image as the tongue outline.

4.2 Dataset

The original tongue images were captured from a hospital by a digital camera. There are 264 tongue images in total and we used 211 tongue images for training and 53 images for test. We cut the training tongue images randomly (reserve at least 70% tongue body) to generate more training data. Standardize the original tongue images and make the tongue images in same size. Draw the outline of the tongue artificially in the standardized image, then blacken the tongue body. Executing the XOR operation to the marked tongue image and the original image, the resulting image is outline of the

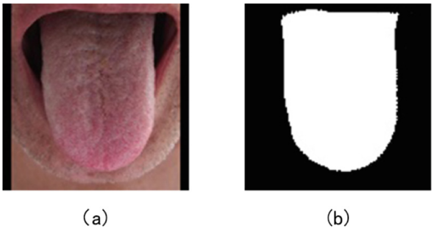

Fig. 6. Image preprocessing result (a) standardized tongue 128 image (b) marked tongue image

tongue and the tag of the original tongue image which is the training target of the convolutional neural network. The standardized tongue image and the marked tongue image were shown in Fig. 6.

5 Experiment Result

After training and testing the convolutional neural network, the tongue body segmentation result was shown in Fig. 3, we can see that the tongue body was accurately segmented by the convolutional neural method.

To validate the availability and practicability of the proposed algorithm, take a tongue image for example, we do the contrast segmentation experiment with traditional snake model algorithm and the proposed improved snake algorithm. The result image of these two methods are shown as Fig. 7.

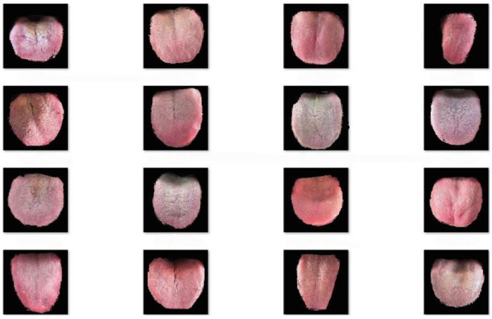

Fig. 7. Experiment result

In Fig. 8, we can see that the proposed tongue segmentation method based on convolutional neural network obtained a better result than the method based on snake model. The former tongue segmentation method kept more details of the tongue than the latter method especially in the part of the tongue edge where the color of tongue was similar to the color of face. Result by the snake model method get a poor

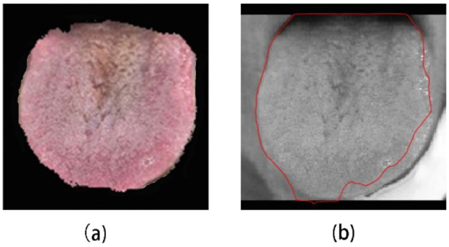

Fig. 8. Comparison of EH-CNN method and snake model method (a) result by EH-CNN (b) result by snake model (Color figure online)

performance when the edge was indistinct and in this case the tongue contour line often moves in the wrong direction. In our experiments, there are several tongue images which didn't get a convergent contour lines by using snake model method. We use the mean error pixel rate to evaluate these two segmentation algorithms, result was shown in Table 1.

Table 1. Mean error pixels rate comparison.

Method	Mean error pixel rate
EH-CNN	5.3%
Snake model method	8.7%

Our enhanced HSV convolutional neural network also has a great advantage over the snake model method in the processing speed. When apply these methods to large amounts of data, faster processing speed has more import application value. The processing speed comparison of these two method was shown in Table 2.

Table 2. Processing speed comparison.

Method	Time (second) per image
EH-CNN	0.0275
Snake model method	3.1355

6 Conclusion

The convolutional neural network is used widely in image recognition. The proposed tongue segmentation method based on enhanced HSV color model convolutional neural network obtains good results particularly performs better in this condition. That the tongue image has an indistinct edge between the tongue body and the face. Tongue image segmentation by snake model method can't get a convergent tongue contour line occasionally. However, we get lower mean error pixel rate by using our enhanced HSV convolutional neural network. Besides, the method we proposed in this paper has a great advantage over the snake model method in tongue image segmentation.

This method also has potential to get better results by increasing the number of training samples and making the artificial marks of tongue body more accurate. Our next stage is to optimize the network structure by adding more layers and fuse multi-layer data features.

References

1. Yuan, L., Liw, E., Yao, J., et al.: Research progress of information processing technology on tongue diagnosis of traditional Chinese medicine. Acta Univ. Tradit. Med. Sin. Pharmacol. Shanghai **25**(02), 80–86 (2011)

2. Guo, R., Wang, Y.-Q., Yan, J.-J., et al.: Study on the objectivity of traditional Chinese medicinal tongue inspection. Chin. J. Integr. Tradit. West. Med. **29**(07), 642–645 (2009)
3. Chiu, C.-C.: A novel approach based on computerized image analysis for traditional Chinese medical diagnosis of the tongue. Comput. Methods Programs Biomed. **61**(2), 77–89 (2000)
4. Sun, X., Pang, C.: An improved snake model method on tongue segmentation. J. Chang. Univ. Sci. Technol. **36**(5), 154–156 (2013)
5. Wang, K., Guo, Q., Zhuang, D.: An image segmentation method based on the improved snake model. In: IEEE International Conference on Mechatronics and Automation, pp. 532–536 (2006)
6. Xu, C.Y., Prince, J.L.: Snakes and gradient vector flow. IEEE Trans. Image Process. **7**(3), 359–369 (1998)
7. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: active contour models. Int. J. Comput. Vis. **1** (4), 321–331 (1988)
8. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, vol. 25, no. 2 (2012)
9. Sun, Y., Wang, X., Tang, X.: Deep learning face representation from predicting 10,000 classes. In: Computer Vision and Pattern Recognition, pp. 1891–1898. IEEE (2014)
10. Zuiderveld, K.: Contrast Limited Adaptive Histogram Equalization. Graphics Gems, pp. 474–485. Academic Press, San Diego (1994)
11. Zhang, L.: Contrast limited adaptive histogram equalization. Comput. Knowl. Technol. (2010)
12. Lecun, Y., Bengio, Y.: Convolutional Networks for Images, Speech, and Time Series. The Handbook of Brain Theory and Neural. MIT Press, Cambridge (1997)
13. Unsupervised Feature Learning and Deep Learning. http://ufldl.stanford.edu/
14. Glorot, X., Bordes, A., Bengio, Y.: Deep sparse rectifier neural networks. In: International Conference on Artificial Intelligence and Statistics (2011)
15. Hinton, G.E., Srivastava, N., Krizhevsky, A., et al.: Improving neural networks by preventing co-adaptation of feature detectors. Comput. Sci. **3**(4), 212–223 (2012)