2018 Conference on Technologies and Applications of Artificial Intelligence (TAAI)

Tongue Fissure Visualization with Deep Learning

Wen-Hsien Chang[1], Hsueh-Ting Chu[2], Hen-Hong Chang[3,4]

[1]School of Post-Baccalaureate Chinese Medicine, China Medical University, Taichung, Taiwan

[2]Department of Computer Science and Information Engineering, Asia University, Taichung, Taiwan

[3]School of Post-Baccalaureate Chinese Medicine, Graduate Institute of Integrated Medicine, China Medical University, Taichung, Taiwan

[4]Department of Chinese Medicine, China Medical University Hospital, Taichung, Taiwan

changwsh@gmail.com; htchu.taiwan@gmail.com; tcmchh55@gmail.com

Abstract*—Tongue diagnosis is a unique practice in traditional Chinese medicine(TCM), which can be used to infer the health condition of a person. However, different TCM doctors may give different interpretations on the same tongue. If an artificial intelligence model can be developed based on a large number of doctor-interpreted tongue images, a more objective judgment will be obtained. Deep learning in artificial intelligence has excellent performance in image recognition, and feature extraction can be done automatically by deep learning without image processing experts. This study attempts to develop a deep learning model through a large number of tongue images, especially for tongue fissures. We also visualize the fissure regions with Gradient-weighted Class Activation Mapping(Grad-cam). Therefore, the model not only try to detect tongue fissures but also localize tongue fissure regions.

Keywords*—Chinese medicine, tongue diagnosis, artificial intelligence, deep learning, class activation mapping

I. Introduction

TCM doctors use four methods of diagnosis (i.e. observation, auscultation and olfaction, interrogation, and pulse feeling and palpation to learn about the status of organs, meridians and circulation of qi and blood, and then infer physiological and pathological changes and determines appropriate treatments. Tongue diagnosis in observation is a unique diagnosis method of TCM doctors [1]. Many disease can be reflected on the tongue. The degree and progression of disease are also highly correlated with the change of tongue. Besides, doctors can directly observe the tongue to avoid the deviation of the patient's private prosecution. Thus, tongue diagnosis is one of the clinically important objective evidence.

However, different doctors may have different interpretations on the same tongue. Even the same doctor cannot interpret the same tongue in the same way every time. Thus beginners cannot obtain a more consistent learning style. If an artificial intelligence model can be built according to tongue images diagnosed in the past, new tongue images can be interpreted in a more objective manner in clinical practice and errors of human interpretation can be reduced.

In the past, the automatic interpretation of the tongue is through conventional feature extraction algorithms combined with statistical method. In recent years, deep learning in artificial intelligence has significant progress in image processing [2] and does not require image processing experts to extract image features manually. Besides, pretrained deep learning models for large datasets sometimes can be applied to different datasets directly. We tried to use the pre-trained model to interpret and localize tongue features, especially for tongue fissures.

II. Related Work

Dan Meng et al. design CHDNet, which combined deep learning and support vector machine method to extract features and perform classification[3]. The extracted features are not features of tongue diagnosis and the classification category is “gastritis” and “non-gastritis.” Jun Hou et al. use deep learning to analyze tongue color and outperform the conventional method [4]. Hsu, Y.-C. et. al. use traditional image processing techniques to detect features of tongue and localize the corresponding regions, but do not provide detailed evaluation methods and results [5]. At present, no one has applied the deep learning visualization method on tongue diagnosis.

III. Methods

Deep learning models can determine the classification of an image, but they are unable to explain the reason to the users. They lack of intuition and comprehensibility. Consequently, when the model gives a wrong result, it is difficult for users to correct the errors. Thus, some people have tried to make the model more interpretable. For example, Zhou et al. propose Class Activation Mapping(CAM) which can localize the class-specific image regions. But CAM requires modifications when using in some deep learning models. Selvaraju et al. propose Gradient-weighted Class Activation Mapping(Grad-CAM) which is a generalization version of CAM. Grad-CAM does not need any modification in the network architecture [6]. We apply Grad-CAM on ResNet50 and try to localize tongue fissure regions [7].

Totally 489 tongue images that had been interpreted by a single TCM doctor are used as training and test data. All images are divided into two groups, namely A and B. All images in group A contain tongue fissures and all images in group B do not contain tongue fissures. There are 312 images in group A and 177 images in group B. All images are augmented with random horizontal flipping. Then we randomly split all images into training (80%) and test(20%) set. An ImageNet-pretrained ResNet50 model is used to perform a transfer learning. That is, the last one layer is replaced with a binary classifier and retrained. After the training being completed, Grad-CAM is combined with the model to localize the tongue fissures (Fig.4).

2376-6824/18/$31.00 ©2018 IEEE
DOI 10.1109/TAAI.2018.00013

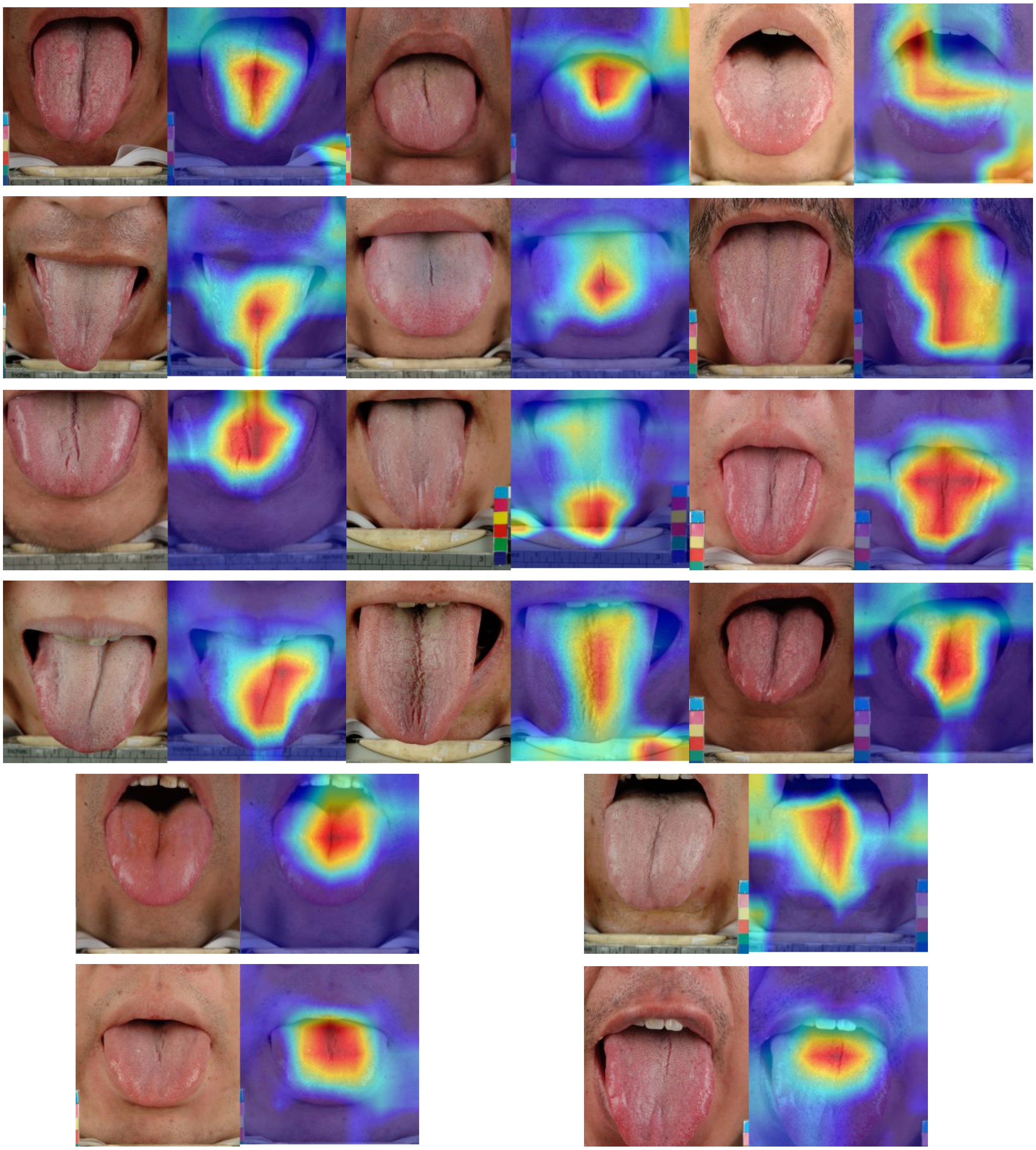

Figure 1. Tongue fissure localization by Grad-CAM on ResNet50. Here are some better results.

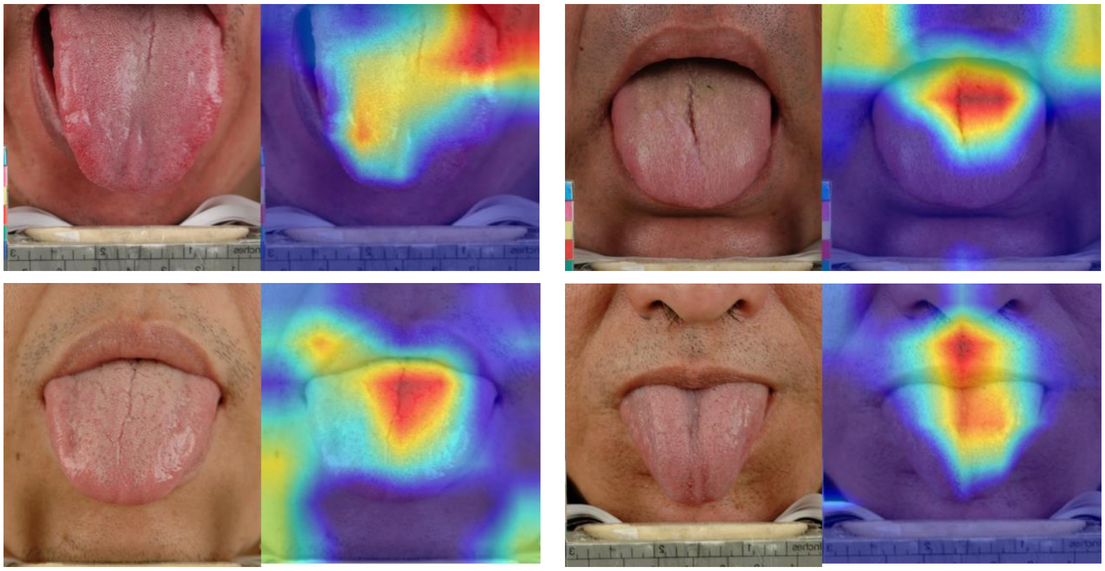

Figure 2. Misunderstood fissures. Some fissures on the face or other parts are also localized.

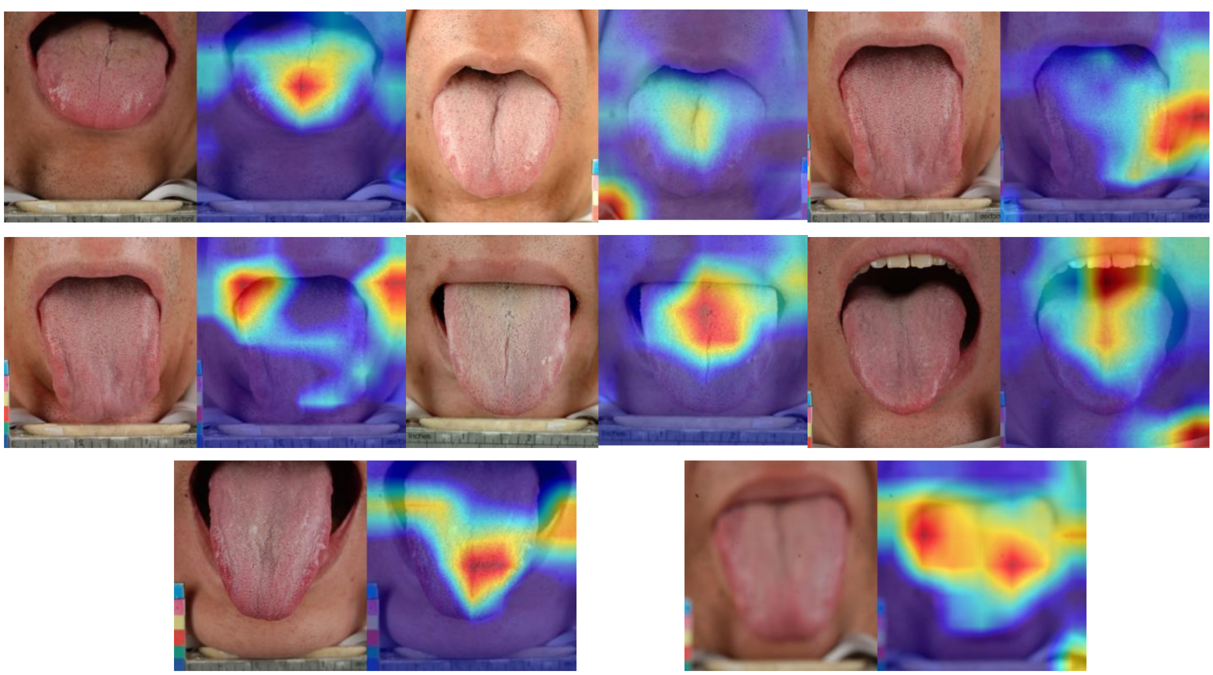

Figure 3. Wrong localization of tongue fissures. Tongue fissures are not localized but other fissures are localized.

IV. RESULT

We train our model without GPU for 6 epochs and each epoch takes less than 4 minutes. The accuracy on the test set is about 70%. Then we apply Grad-CAM to visualize tongue fissures on all images in group A regardless of the model's prediction. Surprisingly, even when an image in Group A is predicted as no fissures, Grad-CAM still be able to localize the fissure regions. Some better results are shown in Fig. 1. Besides, some fissures on the face or other parts are also localized by Grad-CAM. Although fissures on other parts are not our target features, it shows that the model has *"understood"* what a fissure looks like. Some misunderstood fissures are shown in Fig.2. In some images with tongue fissures, other fissures instead of real fissures are localized.

V. CONCLUSION AND DISCUSSION

In this study, we demonstrate how to quickly use the ImageNet-pretrained ResNet50 model to recognize and localize tongue fissure regions through a large number of images containing fissures or not. These fissure regions are not pre-labeled by TCM doctors. At present, the regions localized by the model is not accurate enough. It is still necessary to adjust the parameters to achieve better results. Applying other deep neural networks and fine tuning could also be taken into consideration to improve the performance.

REFERENCES

[1] Jung, C.J., et al., Review on the current trends in tongue diagnosis systems. Integrative Medicine Research, 2012. 1(1): p. 13-20.

[2] Schmidhuber, J., Deep learning in neural networks: an overview. Neural Netw, 2015. 61: p. 85-117.

[3] Meng, D., et al., Tongue Images Classification Based on Constrained High Dispersal Network. Evidence-Based Complementary and Alternative Medicine, 2017. 2017: p. 12.

[4] Hou, J., et al. Classification of tongue color based on CNN. in 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)(2017.

[5] Hsu, Y., et al. Automatic tongue feature extraction. in 2010 International Computer Symposium (ICS2010). 2010.

[6] Selvaraju, R.R., et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. in 2017 IEEE International Conference on Computer Vision (ICCV). 2017.

[7] He, K., et al. Deep Residual Learning for Image Recognition. ArXiv e-prints, 2015.

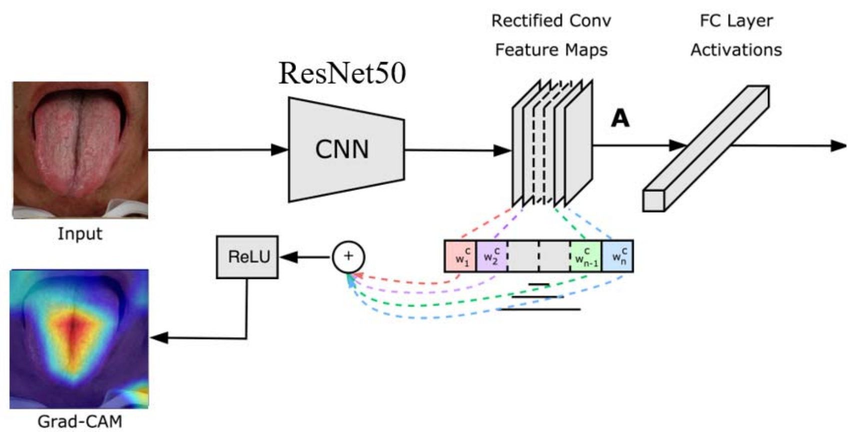

Figure 4. Gradient-weighted Class Activation Mapping(Grad-CAM) for tongue fissure visualization