



Article Title: Automated Tongue Diagnosis: A Deep Autoencoder Neural Network and Clustering-Based Image Segmentation Approach

Automated Tongue Diagnosis: A Deep Autoencoder Neural Network and Clustering-Based Image Segmentation Approach

Abisha L, K. Sindhu

PG Scholar, Department of ECE (CS), Bethlahem Institute of Engineering
Assistant Professor, Department of ECE, Bethlahem Institute of Engineering

ABSTRACT

Automatic tongue image segmentation and tongue image classification are two crucial tongue characterization tasks in traditional Chinese medicine (TCM). Due to the complexity of tongue segmentation and fine-grained traits of tongue image classification, both tasks are challenging. Fortunately, from the perspective of computer vision, these two Tasks are highly interrelated, making them compatible with the idea of Multi-Task Joint learning (MTL). By sharing the underlying parameters and adding two different task loss functions, an MTL method for segmenting and classifying tongue images is proposed in this paper. Moreover, two state-of-the-art deep neural network variants (UNET and Discriminative Filter Learning (DFL)) are focused into the MTL to perform these two tasks. To the best of our knowledge, our method is the first attempt to manage both tasks simultaneously with MTL. We conducted extensive experiments with the proposed method. The experimental results show that our joint method outperforms the existing tongue characterization methods. The process of tongue diagnosis by extracting meaningful features from tongue images and segmenting the relevant regions for analysis. The deep auto encoder neural network is employed to learn a compact representation of tongue images by encoding and decoding the input data.

Index Terms – Tongue characterization, tongue classification, tongue segmentation, multi-task joint learning, deep learning

1 Introduction

Tongue diagnosis is one of the most important diagnostic methods in the TCM context, which could provide an effective, noninvasive way to perform an auxiliary evaluation on the patient's body conditions. However, the popularization of traditional tongue diagnosis has been limited because the diagnosis process involves a great deal of subjective information and experiential knowledge. Specifically, automatic tongue characterization consists of the following two main tasks: automatic tongue segmentation and automatic tongue classification. Because digitally acquired tongue images include not only the tongue region but also parts of the non-tongue regions such as lips, teeth, etc., tongue segmentation from the whole image is the prerequisite for tongue characterization. The tongue classification task refers to identifying the type of tongue image to help TCM physicians make the further diagnostic decision. TCM theory, tongue image can be classified under the following two schemas:



Article Title: Automated Tongue Diagnosis: A Deep Autoencoder Neural Network and Clustering-Based Image Segmentation Approach

tongue body classification and tongue coating classification. In the TCM context, tongue coating refers to the thin layer that appears on the surface of the tongue body, which is a sensitive indicator for revealing disorders and pathological changes in the human body. MTL is a promising technique in machine learning that aims to leverage useful information contained in multiple learning tasks to help train a more accurate model for each task. MTL has been widely used in various computer vision scenarios by joint training on sets of heterogeneous and subtly correlated tasks, such as facial landmark detection, facial detection, facial expression recognition, head pose estimation. Tongue image segmentation and tongue image classification are interrelated tasks, which makes them compatible with the idea of MTL. The experimental results show that our method outperforms both the existing tongue segmentation methods and the existing tongue classification methods by sharing the underlying parameters and adding two different task loss functions.

2 Existing System

2.1 Tongue segmentation

Tongue segmentation methods can be classified into three main categories. The first segmentation subcategory is based on using color information in an image to divide it into feature areas and edges. Proposed an integrated method for tongue detection based on B-spline representation by merging prior knowledge and computer graphics technologies into the tongue detection procedure.

Introduced a color enhancement method based on $L^*a^*b^*$ color space for tongue segmentation. In a method that used color decomposition and thresholding (CDT) was employed for tongue segmentation. Different from the previous studies of which involved a series of features, the method required only a few features such as color distribution and region size information, thereby offering better robustness. The second segmentation category is based on the Snake model (also called active contour) and its variants. Proposed an automated tongue segmentation method by combining a polar edge detector and an active contour model (ACM) technique. Developed a deformable contour approach, the BEDC, for automated tongue edge extraction for a tongue diagnosis system. A color gradient was introduced into the gradient vector flow snake. A double snake model proposed a method to automatically extract the tongue contour, in which the initial tongue boundary was first extracted based on a simple initial-boundary procedure.

Dis-advantages

Mainly in three aspects:

- 1) Sensitive to illumination changes and cluttered backgrounds
- 2) Cannot segment the tongue from the lips accurately due to their similar colors, especially in the Snake-based methods



Article Title: Automated Tongue Diagnosis: A Deep Autoencoder Neural Network and Clustering-Based Image Segmentation Approach

3) Require pre-processing, such as tongue-body detection or require the initial region to be specified before segmentation begins. Recently, deep learning has achieved significant breakthroughs in computer vision fields. The third tongue segmentation category is based on a deep convolutional neural network.

The latest methods applying a Res Net variant and a DeepLabV3 variant for tongue segmentation outperformed some traditional tongue segmentation methods.

2.2 Tongue classification

TCM practitioners diagnose diseases based on abnormal changes in the tongue, including color, wet-ness, thickness, cracks, and tooth marks. From the perspective of digital imaging, these abnormalities are the composition of miscellaneous features such as color distribution, geometric feature, texture features, etc. In a statistical distribution feature descriptor of tongue color using a one-class SVM was proposed for tongue classification. In thirteen geometric features, including measurements, distances, areas, and their ratios, were extracted from tongue images and then classified by an SVM. Presented a classification approach to recognize and analyse tongue shapes based on geometric features automatically. In chromatic and textural metrics were extracted from tongue images. Then, Bayesian networks were employed for tongue classification. Proposed a set of features based on shape geometry and polynomial equations to realize automated recognition and classification of tongue images.

In general, the aforementioned classification methods have been primarily based on specific features, which is inconsistent with the holistic methods of tongue diagnosis in TCM. Thus feature extraction required skilled experts and is extremely time-consuming and tedious. Moreover, handcrafted features must be redesigned for different types of classifications and different datasets. Recently, due to the outstanding ability of feature learning and representation, it is mainstream for automatic feature ex-traction by the CNN and its variants for the invariant to image translation, scaling, skewing and other distortions. Meanwhile, it is a tendency to generalize a pipeline-like approach to handle feature extraction with the tongue classification task simultaneously. However, because it is necessary to remove the non-tongue area before tongue classification to reduce noise, there are few preferable methods for the tongue classification due to the current way for tongue classification is almost based on manually segmented tongue image, which reduces the practicability and automation degree of tongue classification.

3 Proposed System

We propose a new multi-task joint model to manage the tongue segmentation task and tongue coating classification task simultaneously. The tongue seg-mentation task and tongue coating classification task are performed by the semantic segmentation module and fine- grained classification module, respectively.



Article Title: Automated Tongue Diagnosis: A Deep Autoencoder Neural Network and Clustering-Based Image Segmentation Approach

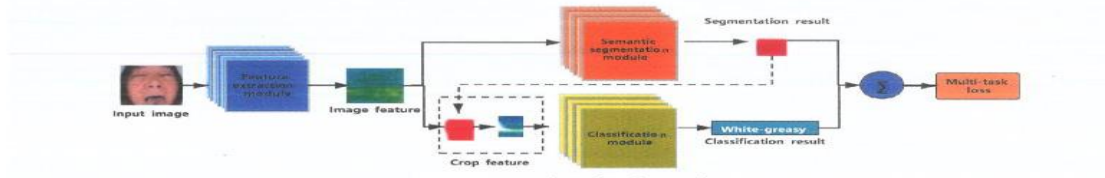


Figure 1: An overview of our framework

3.1 Feature extraction module

Feature extraction from an original image is a prerequisite for all CNN-based image processing. Joint learning model includes semantic segmentation tasks, the feature extraction module should have excellent pixel-level feature extraction capability and produce an image feature with the same size as the original image. With good pixel-level feature extraction ability and few training images, the UNET network and its variants have achieved excellent results for a series of semantic segmentation problems, especially in medical image processing such as liver tumor segmentation, brain tumor segmentation, retinal layer segmentation, etc. By adding the up sampling phase and many feature channels, plentiful information from the original image texture is propagated to the high-resolution layer. The UNET network generates feature maps of the same size as the original image with no contextual features missing. Therefore, we use a UNET variant as the feature extraction module for our joint learning model. The conventional UNET network is based on a convolutional neural network (CNN) and includes both down sampling and up sampling structures. The down sampling path (left side), also called the contracting path, repeatedly applies two (padded) 3×3 convolutional layers with a stride of 1, each of which is followed by a rectified linear unit (ReLU) and a 2×2 max-pooling operation with a stride of 2 through 4 levels. The up sampling path (right side), also called the expansion path, recovers the original dimensions of the input images by up sampling the feature maps, performing a concatenation with the corresponding feature channels from the contractive path and two 3×3 convolutional layers, the first followed by ReLU and a dropout layer and the second by ReLU. In addition to the encoding and decoding parts, a 1×1 convolutional layer with n channels is used for the tongue semantic segmentation result, where n represents the number of categories.

3.2 Semantic segmentation module

In the semantic segmentation module, three dilated convolutional layers with strides of 1, 1, and 1 and padding of 3, 4, and 5 are used repeatedly to receive image features from the feature extraction module to improve the receptive field of the convolutional layer while



Article Title: Automated Tongue Diagnosis: A Deep Autoencoder Neural Network and Clustering-Based Image Segmentation Approach

simultaneously reducing the computational complexity. In dilated convolutions, the alignment of the kernel weights is expanded by the dilation factor. By increasing this factor, the weights are placed far away at given intervals, and the kernel sizes increase accordingly. Therefore, by monotonically increasing the dilation factors through layers, the receptive field can be effectively expanded without a loss of resolution. Following the dilated convolutions, a 1×1 convolutional layer with a soft max activation function is used to obtain the semantic segmentation results.

3.3 Fine-grained classification module

With small inner-class variance and the large intra-class variance, tongue coating classification is a natural fine-grained classification task. The key to successful fine-grained classification lies in developing an automated method that can accurately identify the informative regions in an anomalous tongue image. However, labelling the discriminative regions requires significant manual annotation efforts and is therefore difficult to scale effectively. To avoid this issue, a DFL variant is used for our fine-grained classification and avoid the need for explicit additional object annotation. By introducing a bank of discriminative filters, DFL enhanced the mid-level learning capability of classical CNNs and achieved state-of-the-art performance on different datasets. DFL has recently emerged as one of the mainstream approaches for fine-grained classification tasks.

3.4 Joint loss function

The loss function of the tongue segmentation task is defined as Eq1, in which $f(x)$ represents the semantic segmentation method in this paper, j represents the j training sample, and k represents the k^{th} pixel of the j^{th} training sample. The loss function of the tongue classification task is defined as Eq2, in which $g(x)$, $p(x)$, and $s(x)$ represent the G-Stream, P-Stream and Side-Stream methods of the fine-grained classification module in this paper, respectively.

4 Experiments

4.1 Datasets

All the tongue images are reliable, and their high quality is ensured. The images were captured by a specialized device with a high-end industrial CCD camera and commonly used standard illumination (D50).

4.2 Training details

We use the model trained by Image Net for transfer learning. We split the datasets into training (80%) and validation (20%) datasets randomly. Random rotation and random horizontal flip are applied for data augmentation.



Article Title: Automated Tongue Diagnosis: A Deep Autoencoder Neural Network and Clustering-Based Image Segmentation Approach

4.3 Tongue segmentation experiments

In this experiment, we compare the tongue segmentation results of the JSC model with those of existing methods. These experiments are conducted on two datasets: normal and abnormal tongue datasets. The normal tongue datasets consist of thin-white coating images, which are usually seen on a healthy person, and the abnormal tongue datasets contain the remaining images.

4.4 Tongue coating classification experiments

In this experiment, we compare the classification results of the JSC model with those of existing classification methods for tongue coating classification.

Classification evaluation

We use the F1-score as them criterion to evaluate the performance of tongue classification.

Baselines for classification:

In JSC model, all the existing methods for the tongue classification are two-stage methods (also called non-joint methods). Their first stage involves separating the tongue region using a tongue semantic segmentation method, and their second stage involves classifying the tongue using a specific classification method. To ensure a fair comparison, for the non-joint methods, we use the UNET model for the first stage of tongue semantic segmentation. Then, the image containing only the tongue region is input to the second stage—the tongue coating classification

5 Results

JSC model achieves a balanced improvement on each class with regard to both precision and recall, which suggests that our method better differentiates among similar types and thereby reaches a dramatic reduction in false positives. The joint model only performs feature extraction once, the Inference time of our JSC model (32 milliseconds) is faster than those of non-joint models, which need to carry out feature extraction twice (i.e. once corresponds to the segmentation task and another corresponds to the classification task).

Visualization

We decouple the two tasks (i.e. first perform segmentation and then use the results of this step to perform tongue classification) as a none-joint approach. Intuitively, either the loss of segmentation task or classification task in our joint approach drop more rapidly than those in the none-joint approaches, which further demonstrates the effectiveness of our joint approach. Besides, We can visualize captures the blackened parts of the tongue area, which is highly consistent with human perception. Furthermore, we can visualize the output of the GMP pool, i.e. conv2's subsequent layer, to see how discriminative representations work for



Article Title: Automated Tongue Diagnosis: A Deep Autoencoder Neural Network and Clustering-Based Image Segmentation Approach

the different tongue coating classes. We have obtained many class-specific discriminative filters that are helpful for tongue coating classification tasks.

6 Conclusions

Tongue segmentation and tongue classification are interrelated tasks, which makes them compatible with the idea of MTL. By sharing the underlying parameters and adding two different task loss functions, a method of MTL for segmenting and classifying tongue image is proposed in this paper. Moreover, two state-of-the-art deep neural network variants (UNET and DFL) are fused into the MTL to perform the tongue segmentation and tongue classification tasks. The experimental results show that our method outperforms both the existing tongue segmentation methods and the existing tongue classification methods.

References

1. C. H. Li; P. C. Yuen, Year: 2002, "Tongue image matching using colour content," Pattern Recognition, Vol: 35, pp. 407-419.
2. Hao Zhou; Guangqin Hu; Xinfeng Zhang, Year: 2018, "Constitution identification of tongue image based on cnn", In 2018, 11th International Congress on Image and Signal Processing, Bio Medical Engineering and Informatics (CISP-BMEI), pp. 1–5. IEEE.
3. Qiang Xu; Wenjun Tang; Fei Teng; Wei Peng; Yifan Zhang; Weihong Li; Chuanbiao Wen; Jinhong Guo, Year: 2019, "Intelligent syndrome differentiation of traditional Chinese medicine by ann: A case study of chronic obstructive pulmonary disease", IEEE Access, Vol: 7, no: 1, pp. 76167–76175.
4. R. Pohle; K. D. Toennies, Year: 2001, "Segmentation of medical images using adaptive region growing," in Proceedings of SPIE, Vol: 4322, pp. 1337-1346.
5. S. Xie; S. Shan; X. Chen; J. Chen, Year: 2010, "Fusing Local Patterns of Gabor Magnitude and Phase for Face Recognition," in IEEE Transactions On Image Processing, Vol: 19, no. 5.
6. W. Zuo; K. Wang; D. Zhang; H. Zhang, Year: 2004, "Combination of Polar Edge Detection and Active Contour Model for Automated Tongue Segmentation," in proceedings of the Third International Conference on Image and Graphics, pp. 270-273.
7. W. Li; S. Hu; H. Li; S. Wang, Year: 2009, "A novel segmentation of tongue image," International Journal of Functional Informatics and Personalized Medicine, Vol: 2, no: 3, pp. 315 - 324.