

Intelligent quality control of traditional chinese medical tongue diagnosis images based on deep learning

Xuezhong Zhou^{a,1}, Chenxi Li^{a,1}, Hai Su^a, Zhixian Tang^d, Ping Li^e, Yijie Li^f, Jiawei Hou^{b,*} and Xun Sun^{c,*}

^aCommunity Health Center of Tianping Subdistrict, Xuhui District, Shanghai, China

^bLonghua Hospital Shanghai University of Traditional Chinese Medicine, Shanghai, China

ORCID: <https://orcid.org/0009-0009-2132-5342>

^cLonghua Hospital Shanghai University of Traditional Chinese Medicine, Shanghai, China

ORCID: <https://orcid.org/0000-0003-3586-7934>

^dCollege of Medical Imaging, Shanghai University of Medicine and Health Sciences, Shanghai, China

^eCollege of Medical Instrumentation, Shanghai University of Medicine and Health Sciences, Shanghai, China

^fCollege of Information Technology, Shanghai Ocean University, Shanghai, China

Abstract.

BACKGROUND: Computer-aided tongue and face diagnosis technology can make Traditional Chinese Medicine (TCM) more standardized, objective and quantified. However, many tongue images collected by the instrument may not meet the standard in clinical applications, which affects the subsequent quantitative analysis. The common tongue diagnosis instrument cannot determine whether the patient has fully extended the tongue or collected the face.

OBJECTIVE: This paper proposes an image quality control algorithm based on deep learning to verify the eligibility of TCM tongue diagnosis images.

METHODS: We firstly gathered enough images and categorized them into five states. Secondly, we preprocessed the training images. Thirdly, we built a ResNet34 model and trained it by the transfer learning method. Finally, we input the test images into the trained model and automatically filter out unqualified images and point out the reasons.

RESULTS: Experimental results show that the model's quality control accuracy rate of the test dataset is as high as 97.06%. Our methods have the strong discriminative power of the learned representation. Compared with previous studies, it can guarantee subsequent tongue image processing.

CONCLUSIONS: Our methods can guarantee the subsequent quantitative analysis of tongue shape, tongue state, tongue spirit, and facial complexion.

Keywords: Image quality control, tongue image classification, deep learning

1. Introduction

The four diagnostic methods in Traditional Chinese Medicine (TCM) are divided into “inspection,

¹Xuezhong Zhou and Chenxi Li are co-first authors of the article.

*Corresponding authors: Jiawei Hou and Xun Sun, Longhua Hospital Shanghai University of Traditional Chinese Medicine, No. 725 Wanping South Road, Xuhui District, Shanghai, China. E-mail: houjw@vip.qq.com and shsunxun@sina.cn.

auscultation and olfaction, inquiry, pulse-taking and palpation” [1]. Among them, “Inspection” is listed in the first of the four diagnoses of TCM. It is crucial for TCM doctors to preliminarily know the patient’s physiological state, diagnose the patient’s disease types and pathological changes. This is because the functions of the internal organs often have their external manifestations.

Tongue and facial examination are essential parts of “Inspection” [2]. In TCM theory, doctors must observe the shape, color, and texture of the patient’s tongue. High-quality tongue photography requires patients to open their mouth in a “non-smiling state”, extend the tongue naturally and expose the root of the tongue. In addition, asking patients to open their eyes can provide more information for follow-up facial examination. However, traditional facial and tongue examinations are directly observed by human eyes in clinical practice. With the popularity of tongue diagnosis instruments and the emergence of mobile applications related to TCM, the analysis of tongue and facial images in TCM has moved towards digitization and quantification. Therefore, how to obtain a qualified image so that the computer can conduct subsequent quantitative analysis and avoid producing incorrect results and ambiguities becomes particularly important.

The acquisition of tongue image is the most basic and critical part in the objectification of TCM tongue diagnosis. The acquisition of high-quality tongue image is conducive to the subsequent detailed analysis of tongue characteristics and high-accuracy pathologic research. In the past 10 years, many tongue image acquisition systems have emerged in the market [3,4]. However, these systems have different light sources and color temperatures, resulting in considerable differences in the obtained tongue images. After the tongue image is acquired, some image processing algorithms need to be used, among which the most essential work is tongue detection and segmentation. These methods can be roughly divided into the following types.

The first type is based on the classical mathematical model, such as threshold method [5], since the color of the tongue is usually red. For instance, Gu et al. [6] utilized an automatic tongue segmentation algorithm based on the threshold, whose experiment results show that the segmented area is complete and close to the actual target. Fachrurrozi et al. [7] used hybrid multilevel thresholding and harmony search algorithm to segment the tongue. Snakes or other active contour models [8] can dynamically change the deformation model to find the boundary of the region of interest, to achieve segmentation based on prior knowledge. For example, Sapaico et al. [9] combined the mouth appearance model with edge detection algorithm to fulfill tongue detection. However, the threshold segmentation method depends on the quality of the image, while the statistical model relies on the initial curve.

The second type is based on the traditional machine learning. In the early 21st century, machine learning has been developed rapidly. Machine learning has also made remarkable achievements in tongue image detection and segmentation. Zhou et al. [10] used a reconstruction enhanced probabilistic model for semi-supervised tongue segmentation. Hu et al. [11] put forward an unsupervised tongue segmentation method validated in comparisons with the snake method. Liu et al. [12] proposed a patch-driven segmentation method based on sparse representation. Sapaico et al. [13] utilized a 3-layer cascade of SVM learning classifiers to detect the tongue. Li et al. [14] used histogram projection and learning-based digital matting to extract a tongue from images. Liu et al. [15] put forward a tongue recognition and localization method based on the LBP feature and cascade classifier. However, traditional machine learning algorithms require manual design of the tongue’s features, with limited feature expression capabilities.

The third type is based on the deep learning. In contrast to traditional machine learning algorithms, deep learning enables the automatic extraction of image features, eliminating the requirement for manual design. For instance, Welikala et al. [16] assessed two deep learning structures based computer vision approaches for the automated detection and classification of oral lesions. Shamim et al. [17] have applied

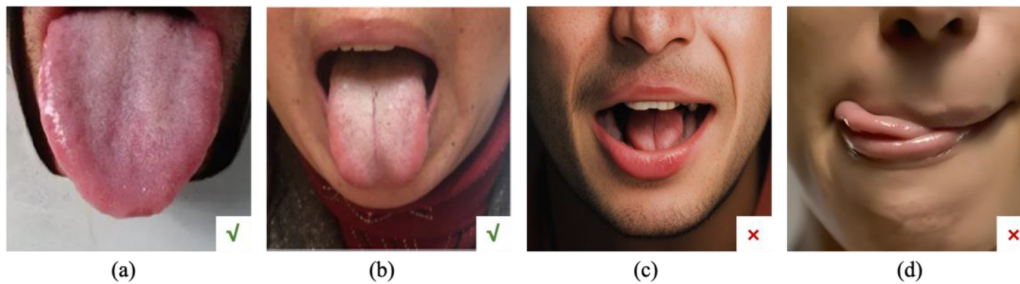


Fig. 1. Common tongue images, where (a) and (b) are qualified tongue images, and the tongues in (c) and (d) are not extended, rendering them unqualified.

and evaluated the efficacy of six deep convolutional neural networks models using transfer learning for identifying pre-cancerous tongue lesions directly. A Multi-Task Joint learning method for segmenting and classifying tongue images is proposed by Qiang Xu et al. [18]. Tang's [19] and Kong's [20] team both used CNN to complete tongue detection. Cai et al. [21] applied a robust interclass and intraclass loss function to improve the performance of deep learning frameworks. Li et al. [22] used the U-Net model to achieve tongue image segmentation, which obtained 98.26% accuracy. Zhou et al. [23] added a morphological layer to the last layer of traditional U-Net, accomplishing segmentation of small-scale tongue image. Huang et al. [24] proposed a Tongue U-Net (TU-Net), which combined the U-Net structure with Squeeze-and-Excitation block, Dense Atrous Convolution block and Residual Multi-kernel Pooling block to segment the tongue region. More recently, Tang et al. [25] and Jiang et al. [26] both utilized deep learning technology to realize tongue image detection and quality control. Deep learning has achieved some good results in the specific field of tongue image analysis, but its clinical application scenarios in traditional Chinese medicine still need to be further expanded.

In summary, there are still many shortcomings in the design of the tongue image analysis system. First, there is no universal standard for tongue image. Different acquisition devices have different light source color temperatures and different display indices. The imaging camera models are changeable. Moreover, the acquisition environment can be open, semi-open, or closed. Second, many images are unqualified to even be of no clinical value in the practical application. As depicted in Fig. 1, the correct and qualified image should reveal the tip, middle, and root of the tongue, while images where the tongue is not extended or straightened are considered unqualified. However, there are few acquisition devices to judge the eligibility of the image. Third, the common tongue diagnosis instrument only considers the tongue and does not collect a complete facial image. In the process of TCM inspection, the collection of facial features is vital. Also, eye-opening or closing will significantly affect the results of the facial diagnosis.

The basis of computerized tongue diagnosis is to collect a complete tongue image. This paper proposes an intelligent quality control algorithm for TCM tongue diagnosis images based on deep learning to judge the qualification of images. The method uses the ResNet framework to evaluate the qualification of collected image data to improve the quality of data sets and guarantee subsequent tongue image detection accuracy. The contributions of this paper mainly include: Firstly, it presents a fast and precise quality control method for traditional Chinese medicine images based on deep learning, which combines artificial intelligence with traditional Chinese medicine. Secondly, it integrates the joint quality control of facial and tongue images, and can indicate the reasons for unqualified images. In this paper, we are going to discuss our problem as follows: Materials and Methods which propose a fast and accurate quality control method for TCM images based on Resnet34. Results which show the effectiveness of our method; Discussion which analyzes the advantages and limitations of this method, and puts forward future research directions.

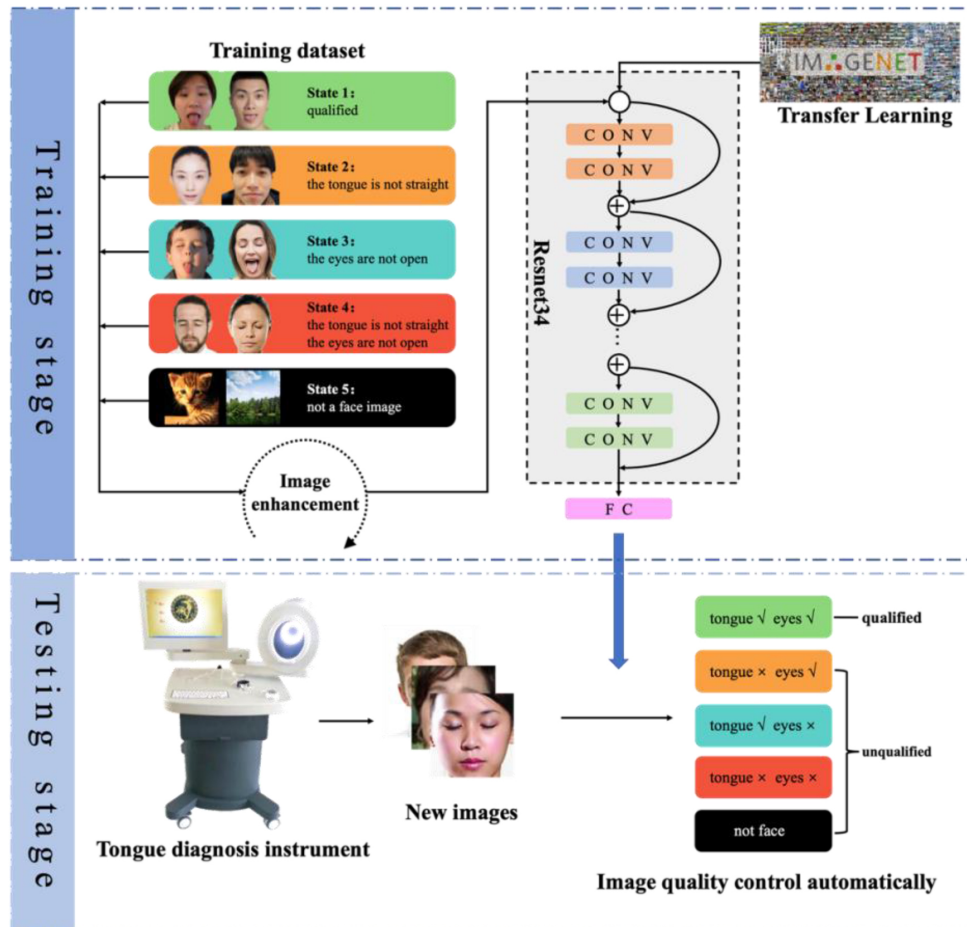


Fig. 2. The workflow of our method. The workflow comprises two stages: the training stage depicted above utilizes a Resnet34-based transfer learning model with training data including five types of images; the testing stage uses the well trained model to identify new images and indicates the reasons for unqualified images.

2. Method

2.1. Algorithm description

The essence of judging whether an image is qualified is a multi-classification problem. This paper proposes an image quality control algorithm based on deep learning to verify the eligibility of TCM tongue diagnosis images. The algorithm workflow is shown in Fig. 2.

The algorithm includes the training stage and testing stage. In the training stage, we first collected enough images and divided them into five states according to whether it was a human face, whether the tongue was straightened and whether the eyes were open. We do these two steps to form a training data set. Secondly, perform image preprocessing on all training images. Then build a convolutional neural network model and use the normalized training image as input to train the network model. In the testing stage, firstly take the patient's images and normalize them. Then, input the test images into the trained convolutional neural network model to achieve automatic image classification. Our method uses the neural network model to analyze and judge the eligibility of the image. It improves the quality of the TCM

tongue diagnosis image and enhances the accuracy of tongue recognition by digging into the high-level features of the tongue.

2.2. Image preprocessing

To use the ImageNet dataset to perform transfer learning, the image must be preprocessed. For each image, the size will be scaled to 224×224 , and the RGB three-channel color will be normalized according to the ImageNet standard. The processing formula is as follows:

$$x' = \frac{x - \text{mean}(x)}{\text{std}(x)} \quad (1)$$

Where the x' represents the RGB channel of the image after preprocessing, x represents the RGB channel of the image before preprocessing, $\text{mean}(x)$ represents the mean value of the image data, and $\text{std}(x)$ is the variance of the image data. After image preprocessing, the values in the three RGB channels will all be in the range of $[-1, 1]$. To solve the problem of unbalanced or insufficient data, our method also enhanced the training data set, including translation, rotation, cropping, flipping, etc.

2.3. Resnet residual network

The depth of deep learning is critical to the performance of the model. When the number of network layers increases, the network can extract more complex feature patterns. But if the network depth is too deep, it will lead to degradation problems. ResNet significantly reduces the attenuation of this gradient correlation, making it possible for the network to have hundreds or even thousands of layers.

The residual network is composed of a stack of residual units. The residual unit is shown in Fig. 3a.

The residual unit can be expressed as:

$$y_l = x_l + G(x_l, W_l) \quad (2)$$

$$x_{l+1} = g(y_l) \quad (3)$$

Where x_l and x_{l+1} represent the input and output of the l -th residual unit, and each residual unit contains a multi-layer structure. G is the residual function, meaning the learned residual and g is the ReLU activation function. W_l is the weight of residual unit. Based on the above formula, the L -layer features can be obtained by learning at the l -layer ($L > 1$):

$$x_L = x_l + \sum_{i=l}^{L-1} G(x_i, W_i) \quad (4)$$

Using the chain rule, the gradient of back propagation can be obtained:

$$\frac{\partial \text{loss}}{\partial x_l} = \frac{\partial \text{loss}}{\partial x_L} \cdot \frac{\partial x_L}{\partial x_l} = \frac{\partial \text{loss}}{\partial x_L} \cdot \left(1 + \frac{\partial}{\partial x_l} \sum_{i=1}^{L-1} G(x_i, W_i) \right) \quad (5)$$

Where loss represents the loss function. $\frac{\partial \text{loss}}{\partial x_L}$ can ensure that the information can be transmitted to any shallow unit during back propagation. $\frac{\partial}{\partial x_l} \sum_{i=1}^{L-1} G(x_i, W_i)$ cannot always be -1 , so the problem of gradient disappearance can be avoided in the residual network.

The path that directly maps the upper layer to the lower layer in the residual network is called a shortcut. Due to the existence of the shortcut, the upper layer information can be passed to the next

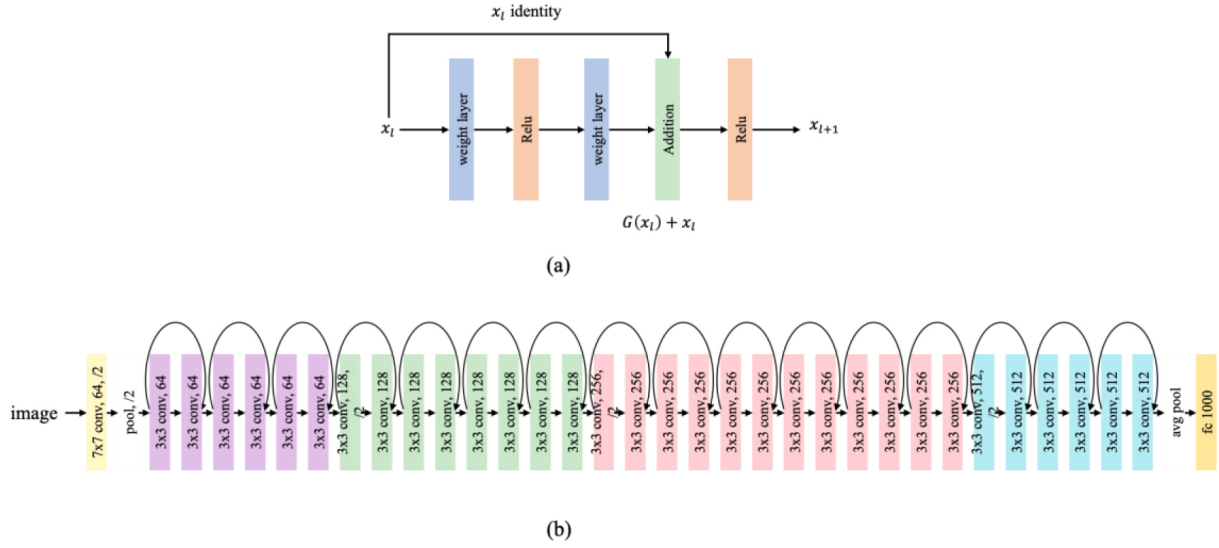


Fig. 3. (a) Residual units. The core of residual units is to introduce a skip connection in each layer of the network, passing the output of the previous layer directly to the next and adding it. This helps retain information and alleviate the gradient disappearance problem. (b) The structure of Resnet34. Resnet34 consists of 34 convolutional layers and adds skip connections in the network, enabling information to be transferred more efficiently between layers.

layer without loss. It will not affect the entire training. The residual network can be regarded as another network composed of adaptive algorithms. No matter how many layers are stacked, the actual layer with parameters is determined by the training itself. Many residual units and a fully connected classification layer can form a ResNet34 network. The specific structure is shown in Fig. 3b.

2.4. Image quality control of tongue diagnosis based on Resnet34 and transfer learning

We use Resnet34 as the fundamental backbone of the network and add a fully connected layer at the end for classification. Since it only needs to distinguish whether the tongue is straight, whether the eyes are open, and whether it is a face photo, it can be boiled down to a 5-classification problem. That is, the final linear layer is 5-dimensional. In practical applications, we first train the Resnet34 network in the ImageNet dataset and then transfer learning to the image classification problem. The method adopts Adam optimizer to train the model and take cross loss entropy as the loss function, shown as Eq. (6).

$$loss = - \sum_{i=1}^5 y_i \log(p_i) \quad (6)$$

Where y_i means the label of state i , if the image is classified correctly, then $y_i = 1$, otherwise it is equal to 0. p_i is the neural network's output and represents the probability that the image is classified into state i .

3. Computer simulation and results

3.1. Data sets

We recruited 300 volunteers, including 147 males and 153 females, and the volunteers consisted of

children, young adults, and the elderly. Subsequently, A total number of 310 tongue and face images were collected using diverse camera devices, including 120 qualified images, 290 unqualified images. To train a robust model, we collected 270 natural images to facilitate the model to learn which is the correct input of our model. It needs to be emphasized that natural images are essential to be added to the experiment. Due to one of the future trends in the development of tongue diagnosis, it can be packaged in mobile apps for the public to use independently. Concretely, a total of 680 images were taken in the experiment. 476 images were used for training, 136 images for validation, and 68 images for testing.

3.2. Training environment and hyperparameters

The experiment used Pytorch to build the deep learning network, and completed the network training on ubuntu 20.04, and RTX Nvidia 2080Ti. The network used the Adam optimizer, the weight decay was $1e-6$, the learning rate was $1e-4$, batch size was 4, and trained 50 Epoch.

3.3. Evaluation metrics

The evaluation metrics in this paper are as follows: accuracy rate, precision rate, recall rate and F1 value. The evaluation metrics are defined as follows.

Accuracy: The ratio of accurate predictions to the total samples.

Macro-Precision: The percentage of accurate predictions that are positive to all positive predictions.

$$mP = \frac{1}{n} \sum_i \frac{TP_i}{TP_i + FP_i} \quad (7)$$

Macro-Recall: The number of positive cases divided into positive cases.

$$mRecall = \frac{1}{n} \sum_i \frac{TP_i}{TP_i + FN_i} \quad (8)$$

Macro-F1 value: The arithmetic mean divided by the geometric mean.

$$mF1 = \frac{1}{n} \sum_i \frac{2TP_i}{2TP_i + FP_i + FN_i} \quad (9)$$

Where w_i represents the proportion of each category. TP_i means true positive, FP_i means false positive, TN_i means true negative, FN_i means false negative for category i . n is the number of categories. For the above four metrics, higher scores indicate better performance.

3.4. Algorithm performance

We reproduced several classic deep learning classification models, including Alexnet, VGG16, and GoogLeNet [27]. All models used the same hyperparameters. The experimental results are shown in Table 1. For the test dataset, the classification accuracy of our algorithm can reach 97.06%. Moreover, the classification speed of each picture is in milliseconds.

From Fig. 4, we can see that our trained model achieved highly promising classification performance. In the category, the first two letters represent the state of the eyes, where ‘‘EY’’ stands for qualified eyes and ‘‘EN’’ stands for unqualified eyes. The last two letters indicate the state of the tongue, where ‘‘TY’’ represents a qualified tongue and ‘‘TN’’ represents an unqualified tongue. Only ‘ENTN’ and ‘EYTN’ produced one wrong classification result among the five classes.

Table 1
Classification results in test dataset compare with other models

Model	Accuracy	mP	mRecall	mF1
Alexnet	94.11%	91.81%	93.24%	92.52%
VGG16	94.11%	93.33%	93.57%	93.45%
GoogLeNet	95.59%	95.14%	93.95%	94.54%
Ours	97.06%	96.64%	95.95%	96.29%

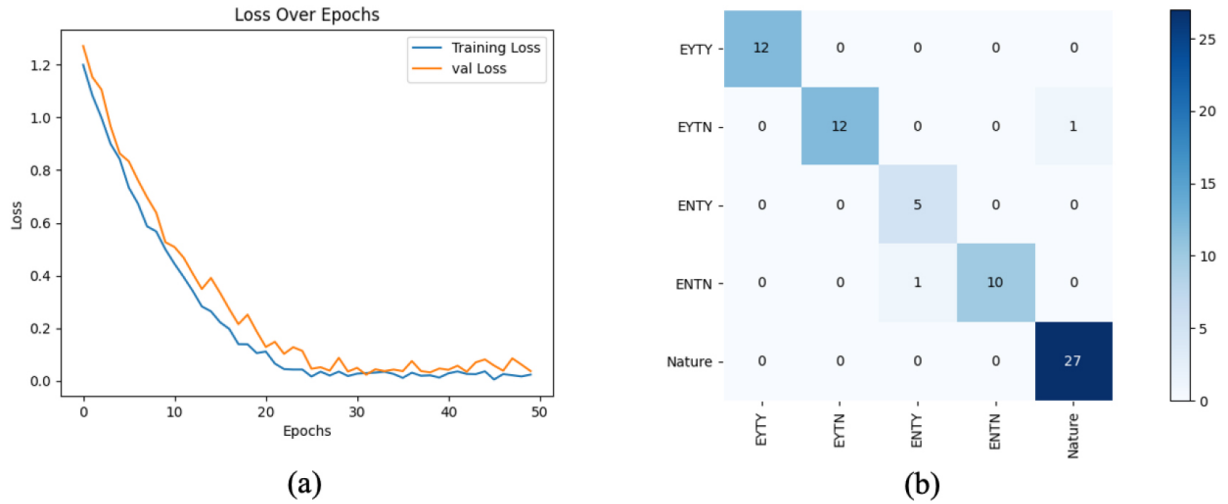


Fig. 4. Loss curve and confusion matrix of the classification results. (a) represents the loss curve during training, and it can be found that during the training process, the convergence of the loss function is very fast. (b) represents confusion matrix of testing data. The row represents the predicted results, and the column represents the ground truth.

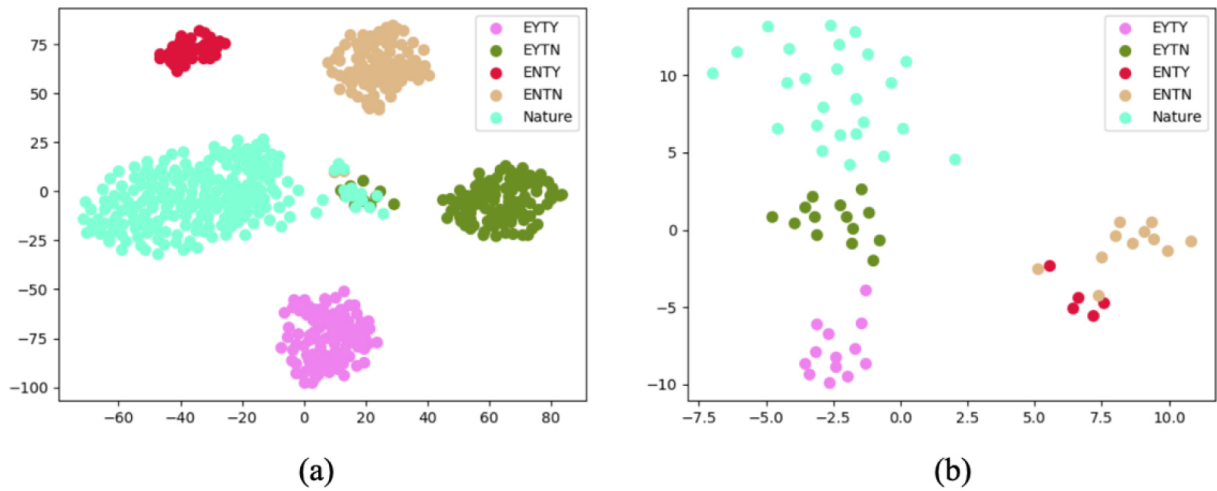


Fig. 5. Visualization of learned representations using t-SNE. (a) represents the train set (b) represents the test set. Different colored dots indicate the distribution of features in various categories via deep learning. Features within the same category are relatively proximate, while those in different categories have a certain distance.

To have an intuitive understanding of the classification performance of our models, we visualize the learned features on the train and test set in Fig. 5, respectively. We map the high-level features to 2D space by applying t-SNE [28]. The t-SNE technology can reduce the dimensions of data by constructing the probability distribution between samples, so as to better understand the structure and relationship of the data. From Fig. 5, we can clearly see that features from different categories are separated well, reflecting our proposed method's strong discriminative power of the learned representation. Additionally, we can learn that the classes 'EYTN' and 'Nature' are somewhat inseparable, and they are still separated from their original class, which indicates some noise data. These noise data can encourage the model to be more robust.

4. Conclusion

With the progress of image processing, researchers use computer vision to create a new approach for objectifying and standardizing TCM tongue and face examination. However, current research has the following shortcomings: non-standardized image collection lacks qualification testing and standardization, making these images less suitable for tongue detection.

To achieve the quantification of data and optimize the process, this paper proposes an intelligent quality control algorithm for Chinese tongue diagnosis images based on ResNet34 and transfer learning. First, the tongue image is preprocessed, the data is uniformly scaled to 224×224 size, and regularization is used for image preprocessing. Secondly, using ResNet34 as the backbone of the neural network and adding a linear layer at the end for classification, the images are divided into five categories: non-face images, tongue not straightened, eyes closed, tongue not straightened and eyes closed, and qualified.

The method proposed in this paper has the following advantages: (1) The use of deep learning methods for image detection has fast speed, better robustness, and higher accuracy. (2) Compared with previous studies, the image standardization research ensures subsequent tongue image segmentation and feature extraction. Due to the limitation of one's ability and the influence of external objective factors, the method proposed in this paper still has some shortcomings. For instance, incorrect judgments by the algorithm still occur. Two examples of errors in the experiment are as follows: a child's image with a straight tongue is identified as having a non-straight tongue, and an image of an animal is recognized as a human face. Furthermore, when collecting images, subjects are all required to remove their glasses. However, in real life, it is highly probable that people will be photographed wearing glasses. In the following study, we will continuously improve the analysis accuracy and take more factors, such as adopting more robust classification models, or introducing attention mechanisms, into account to achieve higher efficiency and precision.

Acknowledgments

This research was supported by the Shanghai Municipal Health Commission, General Program (20215003).

Conflict of interest

None to report.

References

- [1] Velik R. An objective review of the technological developments for radial pulse diagnosis in Traditional Chinese Medicine. *European Journal of Integrative Medicine*. 2015; 7: 321-331.
- [2] Tian Z, Wang D, Sun X, et al. Current status and trends of artificial intelligence research on the four traditional Chinese medicine diagnostic methods: A scientometric study. *Annals of Translational Medicine*. 2023; 11(3): 145.
- [3] Zhang B, Wang X, You JJ, et al. Tongue color analysis for medical application. *Evidence-based Complementary and Alternative Medicine: eCAM*. 2013; 4: 264742.
- [4] Kawanabe T, Kamarudin ND, Ooi CY, et al. Quantification of tongue colour using machine learning in Kampo medicine. *European Journal of Integrative Medicine*. 2016; 8: 932-941.
- [5] Li Z, Yu Z, Liu W, et al. Tongue image segmentation via thresholding and clustering. 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). 2017. pp. 1-5.
- [6] Gu H, Yang Z, Chen H. Automatic Tongue Image Segmentation Based on Thresholding and an Improved Level Set Model. 2020 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS). 2020. pp. 149-152.
- [7] Fachrurrozi M, Erwin, Saparudin, et al. Tongue image segmentation using hybrid multilevel otsu thresholding and harmony search algorithm. *Journal of Physics: Conference Series*. Vol. 1196; 2019. p. 012072.
- [8] Shi MJ, Li GZ, Li FF. C2G2FSnake: Automatic tongue image segmentation utilizing prior knowledge. *Sci China Inf Sci*. 2013; 561-14.
- [9] Sapaico LR, Laga H, Nakajima M. Detection of tongue protrusion gestures from video. *IEICE Transactions on Information and Systems*. 2011; E94-D(8): 1671-1682.
- [10] Zhou C, Fan H, Zhao W, et al. Reconstruction enhanced probabilistic model for semisupervised tongue image segmentation. *Concurrency and Computation: Practice and Experience*. 2020; 32: e5844.
- [11] Hu X, Liu Z, Yang X, et al. An Unsupervised Tongue Segmentation Algorithm Based on Improved gPb-owt-ucm. *Journal of Medical Imaging and Health Informatics*. 2021; 11: 688-696.
- [12] Liu W, Zhou C, Li Z, et al. Patch-driven tongue image segmentation using sparse representation. *IEEE Access*. 2020; 8: 41372-41383.
- [13] Sapaico LR, Nakajima M. Toward a tongue-based task triggering interface for computer interaction. *Conference on Applications of Digital Image Processing*. 2007.
- [14] Li X, Li J, Wang D. Automatic tongue image segmentation based on histogram projection and matting. *IEEE International Conference on Bioinformatics & Biomedicine*. 2014. pp. 76-81.
- [15] Liu Y, Chen D, Yu F, et al. A tongue segmentation algorithm based on lbp feature and cascade classifier. *International Conference on Artificial Intelligence and Advanced Manufacture*. 2020. pp. 109-112.
- [16] Welikala RA, Remagnino P, Lim JH, et al. Automated detection and classification of oral lesions using deep learning for early detection of oral cancer. *IEEE Access*. 2020; 8: 132677-132693.
- [17] Shamim MZM, Syed S, Shiblee M, et al. Automated detection of oral pre-cancerous tongue lesions using deep learning for early diagnosis of oral cavity cancer. *The Computer Journal*. 2022; 65: 91-104.
- [18] Guo J, Xu Q, Zeng Y, et al. Multi-task joint learning model for segmenting and classifying tongue images using a deep neural network. *IEEE Journal of Biomedical and Health Informatics*. 2020; 24(9): 2481-2489.
- [19] Tang Q, Yang T, Yoshimura Y, et al. Learning-based tongue detection for automatic tongue color diagnosis system. *Artificial Life and Robotics*. 2020; 25: 363-369.
- [20] Kong X, Rui Y, Dong X, et al. Tooth-marked tongue recognition based on mask scoring R-CNN. *Journal of Physics: Conference Series*. 2020; 1651: 012185.
- [21] Cai Y, Wang T, Liu W, et al. A robust interclass and intraclass loss function for deep learning based tongue segmentation. *Concurrency and Computation: Practice and Experience*. 2020; 32: e5849.
- [22] Li L, Luo Z, Zhang M, et al. An iterative transfer learning framework for cross-domain tongue segmentation. *Concurrency and Computation: Practice and Experience*. 2020; 32: e5714.
- [23] Zhou J, Zhang Q, Zhang B, et al. TongueNet: A Precise and fast tongue segmentation system using U-net with a morphological processing layer. *Applied Sciences*. 2019; 9: 3128.
- [24] Huang Y, Lai Z, Wang W. TU-Net: A precise network for tongue segmentation. 9th International Conference on Computing and Pattern Recognition. 2020.
- [25] Tang W, Gao Y, Liu L, et al. An automatic recognition of tooth-marked tongue based on tongue region detection and tongue landmark detection via deep learning. *IEEE Access*. 2020; 8: 153470-153478.
- [26] Jiang T, Hu X, Yao X, et al. Tongue image quality assessment based on a deep convolutional neural network. *BMC Medical Informatics and Decision Making*. 2021; 21: 147.
- [27] Jena B, Nayak GK, Saxena S. Convolutional neural network and its pretrained models for image classification and object detection: A survey. *Concurrency and Computation: Practice and Experience*. 2021; 34(6): e6767.
- [28] Van DML, Hinton G. Visualizing data using t-SNE. *Journal of Machine Learning Research*. 2008; 9: 2579-2625.