Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 11, No. 3, 2023

Tongue image datasets and performance evaluation

Xiangyu Deng [a], Tianchao Ma [b]

School of Physics and Electronic Engineering, Northwest Normal University, Lanzhou 730070, China
[a]dengxy000@126.com, [b]1454140052@qq.com

Abstract: With the development of computer medicine, data has become an extremely important resource, and the construction and research of dataset is an important means to promote the development and innovation of medicine. Because there are few tongue image datasets and most tongue body image datasets are in a non-public state, the research of medical related algorithms is hindered to some extent, especially in computer vision and medical artificial intelligence. The constructed tongue image data set provides a large number of tongue images and related clinical information, which can be used to train computer medical models such as tongue image detection and tongue segmentation. A Tongue dataset based on deep learning was established by selecting and labeling the collected tongue images manually. The datasets contain 2,021 tongue images, which are randomly divided into training sets, verification sets and test sets according to the proportion, and the dataset performance evaluation experiments are carried out on the existing algorithms. Testing on convolutional neural network models such as YOLO detection and tongue segmentation, the average accuracy of detection and segmentation Miou value can reach 97.0% and 99.95%, respectively. The experimental results show that the data set can obtain high accuracy in detection and segmentation algorithms, and can be further applied to related research in the medical field. Tongue dataset data set can be downloaded at the following url: https://github.com/xiaohuomiao12/datasets.git.

Keywords: Deep learning; Computer medicine; Tongue image datasets; Tongue surface detection; Tongue segmentation.

1. Introduction

Tongue diagnosis is an important part of traditional Chinese medicine diagnosis and one of the traditional Chinese medicine diagnosis methods. Traditional Chinese medicine tongue diagnosis is mainly through observing the shape of the tongue, color, texture, moss quality and so on to judge the patient's physical condition. However, there are some problems in traditional Chinese tongue diagnosis, such as strong subjectivity and lack of standardization, which limit the popularization and application of traditional Chinese tongue diagnosis. The automation, standardization and scientific research of TCM tongue diagnosis methods have become a hot spot in the field of TCM tongue diagnosis. In recent years, with the development of computer vision, artificial intelligence and other technologies, the automated research of TCM tongue diagnosis has also made certain progress. However, the diagnostic accuracy of computer vision technology is not high, and the detailed information is ignored, resulting in certain errors in tongue surface detection, tongue surface segmentation [1] and physical classification. The establishment of this data set can help doctors to perform tongue diagnosis more quickly and accurately, and also help to improve the reliability and scientificity of TCM tongue diagnosis. Modern medicine attaches more and more importance to personalized diagnosis and treatment, and the traditional Chinese medicine diagnosis technology is gradually concerned by the world [2].

The establishment of tongue image datasets follows a standardized process, including tongue shooting, image processing, feature extraction and annotation. Among them, tongue shooting is a key part, which needs to follow certain shooting standards and conditions, such as illumination, Angle, distance, etc. Image processing and feature extraction are the processes of converting raw data into numerical features that can be used for analysis. Annotation is to associate tongue image data with TCM diagnosis results, so that the data set has availability and reference value. The application of tongue image datasets covers many fields, such as traditional Chinese medicine diagnosis, computer vision, artificial intelligence, biomedical engineering, health management and so on. In traditional Chinese medicine (TCM) diagnosis, tongue image datasets can help doctors make diagnosis and assist in judging the type and degree of disease [3].

2. Tongue image data set introduction

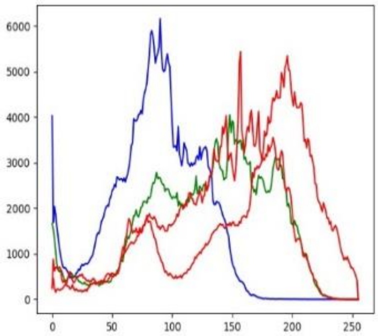
Sources of tongue image data sets the sources of tongue image datasets mainly include two aspects: one is to collect tongue images from hospitals, TCM clinics, TCM doctors, visiting personnel, remote disease diagnosis professionals, and institutions; the other is to collect tongue images of patients through the network. Through these methods, tongue image datasets can cover tongue images of different regions, different ages and different genders, with good representativeness and reliability. For the collection of good tongue image pictures also need to do the following aspects:

(1) Image preprocessing: Through image preprocessing, the noise in the image can be removed and the contrast and brightness of the image can be enhanced. The methods of image preprocessing include image denoising, histogram equalization, filtering and so on.

(2) Feature extraction: Feature extraction is the process of extracting and describing features in tongue images. Feature extraction methods include color feature, shape feature, texture feature, etc., as shown in Figure 4

(3) Data annotation: Data annotation is the process of marking the corresponding tongue quality, tongue coating and other information in the tongue image. Data annotation methods include manual annotation and automatic annotation. Tongue detection segmentation [4] datasets. In the selection of labeling tools, LableMe is selected as the labeling tool. After opening LabelMe software, select the create polygons

label function, mark the tongue label, save the label information and generate the corresponding json file. The json file mainly records the coordinates of each pixel point that makes up the polygon boundary label. By modifying the Python script labelme_draw_json that comes with LabelMe, the operation of converting json files to images in png format in batches is realized, so as to obtain the final label images,



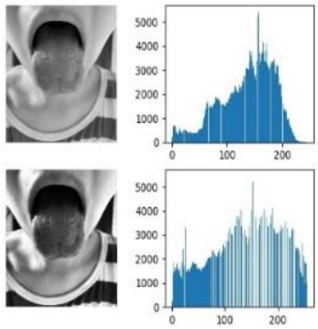


Fig.1 Original drawing **Fig.2** Color histogram **Fig.3** Histogram equalization

Fig.4 Feature map of tongue surface image

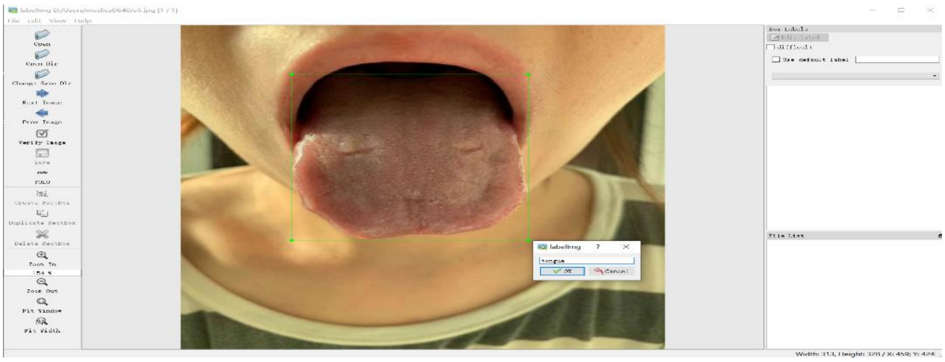


Fig .5 Example of detecting labels

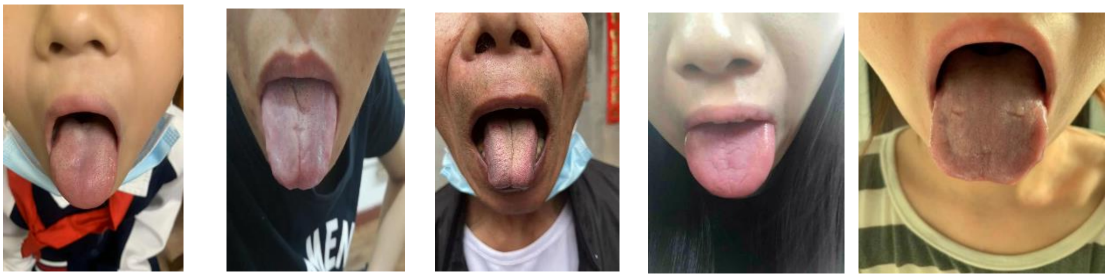
Fig.6 Example of segmentation label

The format of tongue image datasets mainly includes the following aspects: (1) Image format: The image format of tongue image data set can be jpg, png, bmp and other formats, usually the image resolution is more than 300dpi. (2) Data format: The data format of tongue image data set can be txt, xml, csv and other formats. (3) Folder structure: Tongue image data sets are usually stored in the form of folders, each folder corresponds to a patient's tongue image and related clinical information. The folder is usually named with the patient's number, name and other information.

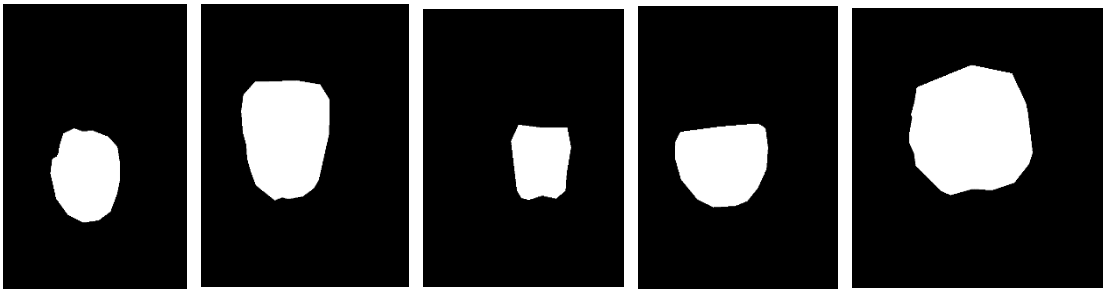
3. Application of data sets

3.1. Tongue surface detection

YOLO network model is a target recognition and location algorithm based on DCNN. The most important feature is fast detection speed, which can be used to detect targets quickly. The YOLO network model extracts the depth features of input images through the backbone feature extraction network, and can improve the accuracy of target detection by using feature fusion, thus further improving the effectiveness of features. The YOLO V4 Tiny Network model is a lightweight YOLO V4 network model that combines FPN and FCN (Full Convolutional networks) to ensure the accuracy of a particular model. It has the advantages of simple structure and lightweight model, and is suitable for detecting small targets. The model used in this paper is the improved YOLOv5 model,



(a) The original image



(b) label image

Fig.7 The original image and its label image

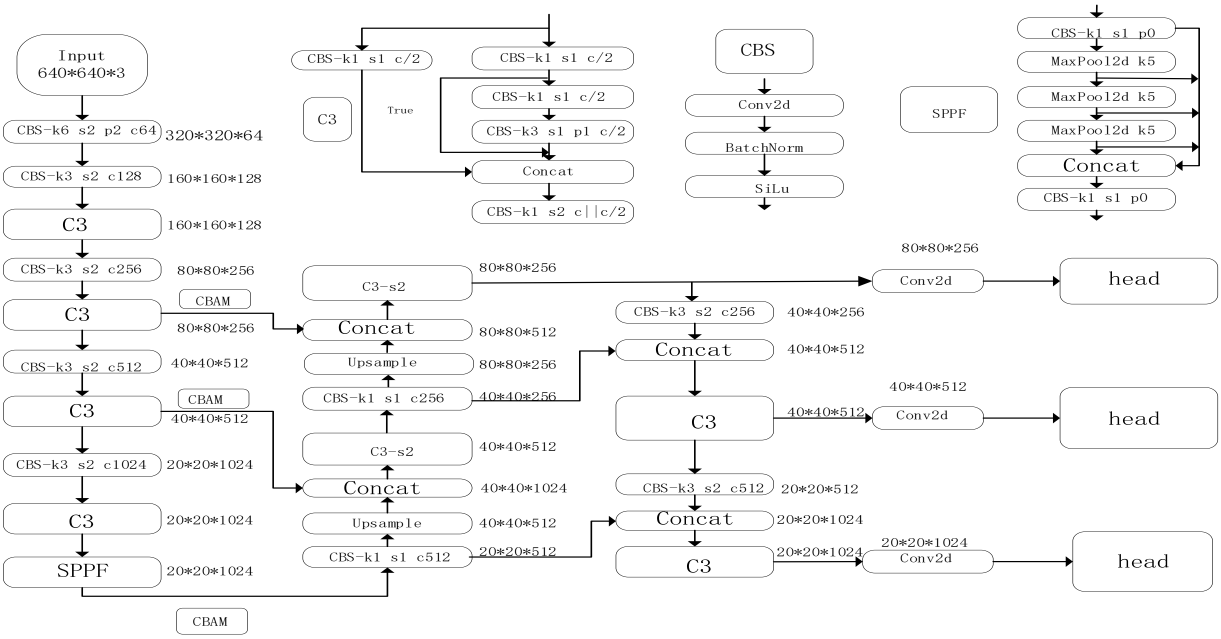


Fig. 8 Structure of tongue surface detection

3.2. Tongue and surface segmentation

YOLOv5 is an object detection algorithm based on deep learning, which can simultaneously detect the location and category of the target in the image. YOLOv5 is used to segment the tongue instance to achieve good segmentation effect. First, we need to prepare some data sets. We used the self-made Tongue dataset, which contains a large number of tongue surface labeled and segmented raw images and labeled images. Second, we train our model. The structure of the segmentation model is shown in the figure 9.

Or use some third-party tool to train our model. During the training process, we need to set some parameters, such as learning rate, batch size, number of iterations, etc. We also need to choose the appropriate loss function and optimizer to optimize our model.Once our model is trained, we can use it for instance segmentation. We can pass the input image to the model and get a prediction for each pixel. These predictions contain information about which category each pixel belongs to and its position. We can use these predictions to segment the image to get the tongue region to which each pixel belongs.Finally, we need to post-process the segmentation results to get more accurate results. There are techniques we can use to eliminate noise and errors, such as non-maximum suppression (NMS) and threshold splitting.We can also use some techniques to enhance the quality of segmentation results, such as semantic segmentation and instance segmentation fusion.In conclusion, using YOLOv5 for tongue instance segmentation is a very meaningful work. With this approach, we can more accurately identify the tongue region in oral medical images, thus providing better support for the diagnosis and treatment of oral diseases.

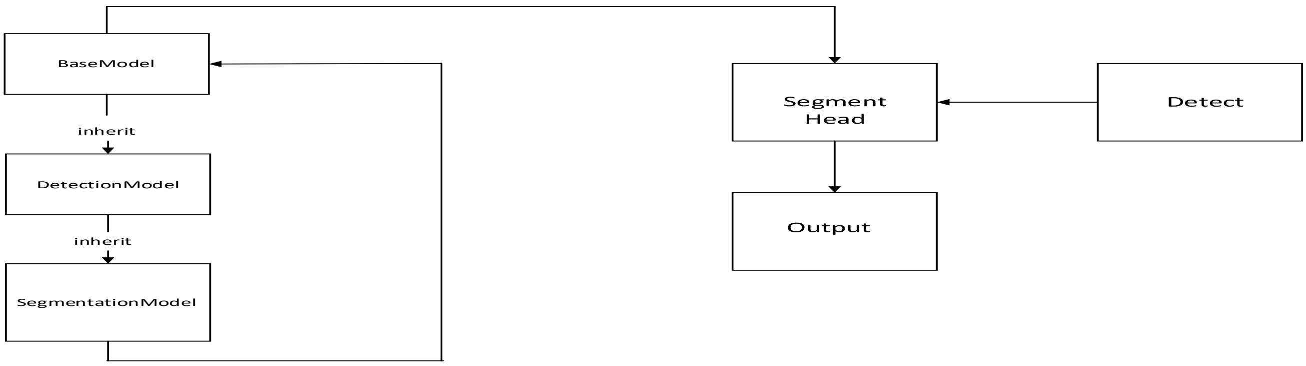


Fig.9 Structure of tongue and surface segmentation

4. Evaluation of experimental environment and results

4.1. Experimental environment

The scene classification model in this paper was implemented by PyTorch framework. Before model training, all images were scaled to 640 × 640 pixels. The number of training cycles of each model was 30epoch, and the models were optimized by stochastic gradient descent until convergence. It should be noted that data augmentation was not used in the experiments. The experimental environment is shown in Table 1

Table 1. The experimental environment

Configuration name	Configuration result
Graphics card	GeForce RTX™ 3050
Video memory size	4GB
CPU	Intel® Core™ i5-11400H
Memory size	16GB
Operating system	Windows10

4.2. Evaluation index

In order to quantitatively analyze the effectiveness of the proposed algorithm and evaluate its segmentation effect, we introduce four evaluation indicators:

Iou: Also known as the intersection ratio, is a measure of the overlap between the predicted bounding box and the real bounding box. When the two bounding boxes are completely overlapped, its mIou value is 1, and when they do not overlap, the mIou value is 0. The mIou is calculated as follows

$$MIoU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ii}} \quad (1)$$

Where, represents the true value, represents the predicted value, and represents the number of pixels to be predicted.

Precision: Also known as precision, it represents the percentage of positive samples detected by the classifier to all predicted positive samples, that is, the proportion of correct target prediction boxes detected in the current traversed prediction box. Its calculation method is formula 1, where is the positive sample predicted by the model as a positive class and the negative sample predicted by the model as a positive class. The training results are shown in Figure 10.

$$\Pr ecision = \frac{TP}{TP+FP} \quad (2)$$

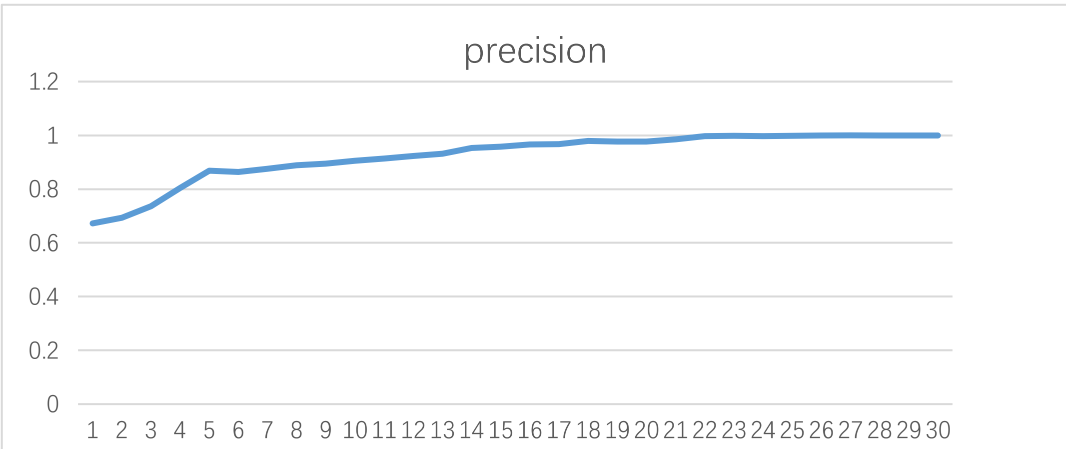


Fig.10 precision curve

Cross Entropy Loss: Called the cross-entropy loss function, it is one of the common loss functions used in machine learning to measure the difference between the model's predictions and the real label. The cross-entropy loss function is calculated as follows

$$L_{CE} = -\sum_{i=1}^{n} p(x_i)\log(q(x_i)) \quad (3)$$

Where, represents the true distribution of the sample and the distribution predicted by the model. The smaller the cross-entropy loss, the better the prediction effect of the model.

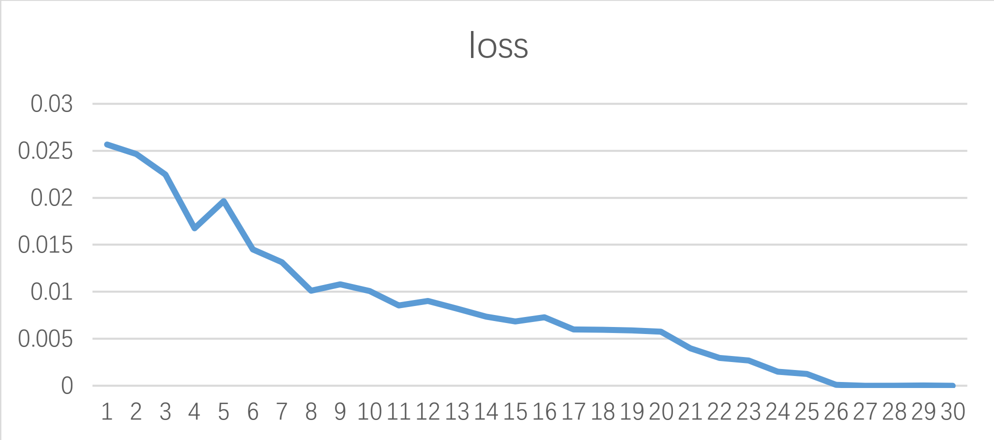


Fig. 11 Cross Entropy Loss curve

4.3. Experimental Results

We used the YoLov5 model to train the labeled Tongue dataset. The experiment found that this dataset could better obtain the tongue surface detection and segmentation model. The trained model could predict any image, and the model could efficiently and accurately detect and segment the tongue surface, which was better than the existing detection and segmentation results. It has provided great help to relieve the pressure of primary medical staff. The test and segmentation results are shown in Figure 11.

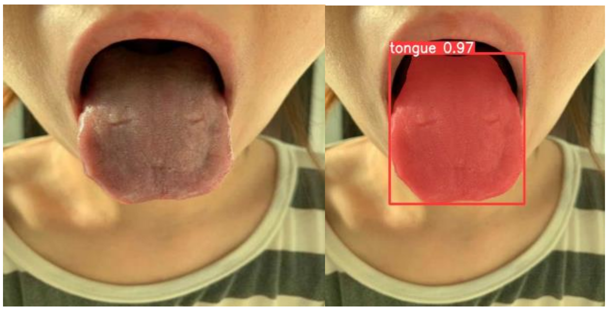


Fig.11 Experimental results

Table 1. Results of comparison with other models

Algorithms	Precision (%)	Recall (%)	mAP (%)	F1 (%)	AP (%)	Detection speed (f/s)	Model size (MB)	
Faster R-CNN	81.09	88.25	85.10	84.51	85.10	13.74	108.16	
YOLO V3	95.91	93.11	97.81	94.00	97.81	6.84	236.32	
YOLO V4	99.93	91.23	99.85	99.90	99.85	7.56	244.39	
SSD 300	67.78	87.12	76.24	84.32	84.12	12.04	90.47	
EfficientDet-D0	99.93	99.78	99.80	99.90	99.80	1.08	14.94	
YOLO V4-tiny	95.71	74.53	83.43	84.00	83.43	7.03	70.82	
Ours (100 epoch)	99.09	96.24	97.43	97.00	97.43	25.97	22.63	
Ours(1000epoch)	99.95	96.87	98.89	98.00	98.89	27.09	22.63	

Table 2. Results of comparison with other models

Model	MIou(%)	MPA (%)	mAP(%)	F1-Score (%)	G-Score (%)	AUC (%)	Parameterss(%)
PSPNet	81.49	93.30	92.71	79.35	81.19	98.33	46.81
SegNet	91.80	95.71	96.43	91.07	91.66	99.18	28.29
UNet[3]	88.70	95.37	95.70	88.37	90.06	99.02	25.83
DeepLabV3+	95.58	97.15	99.07	95.88	96.03	99.86	12.06
IAUNet	96.30	97.86	99.18	96.71	96.82	99.71	5.70
YoLov5	97.01	97.91	99.49	97.23	96.89	99.95	14.32

Table 1 Results of tongue surface detection compared with other models.

Comparing the results of this experiment with the published ones, it is found that this dataset has higher clarity than the published ones, and the detection and segmentation models obtained are better, which greatly accelerates the development of computer medicine. The experimental comparison results are shown in Table 2.

5. Significance of tongue image dataset

Tongue diagnosis is an important method in traditional Chinese medicine diagnosis. By observing the shape, color, texture and other characteristics of the tongue, we can infer the patient's physical condition and the severity of the disease, and provide support and reference for clinical diagnosis. The establishment, publication and sharing of tongue image data set plays an important role in promoting the standardization and automation of TCM tongue diagnosis.

Tongue image datasets can be used to train and test machine learning models for automated tongue image diagnosis. By analyzing a large number of tongue image and related medical data, the correlation model between tongue image and illness can be established, which provides new ideas and methods for TCM diagnosis. At the same time, the intelligent diagnosis system based on tongue image can improve the accuracy and efficiency of diagnosis, shorten the diagnosis time, and provide better medical services for patients.

Aiming at the research and application of tongue image based on deep learning, a tongue image detection and segmentation dataset was established, including 2021 images and label data. In this paper, the identification accuracy of Tongue dataset is tested under the classical deep learning framework. The experimental results show that the Tongue dataset is successfully run on the deep learning network, and the maximum value of MPA and MIoU can reach 97.0% and 99.95% respectively. However, Tongue dataset has the problem of relatively small scale and few label categories. In the next step, according to the work content and shooting images, multi-objective segmentation objects will be selected to further improve and expand Tongue datasets.

References

[1] Li X, Liu Z, et al. A novel tongue image segmentation method based on an improved fuzzy C-means clustering algorithm[J]. Computer Methods and Programs in Biomedicine, 2020, 12(3): 12-21.

[2] Zhang Y, Liu W, Wu L. Computer-aided tongue diagnosis using machine learning techniques: a review[J]. Evidence-Based Complementary and Alternative Medicine, 2020, 22(3): 7-13.

[3] Yang Y, Wang X, et al. A novel tongue image segmentation method based on the improved U-Net convolutional neural network[J]. Journal of Medical Imaging and Health Informatics, 10(10): 10-16.

[4] Swan R, Atha D, Leopold H, et al. AI4MARS: A Dataset for Terrain-Aware Autonomous Driving on Mars[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, USA, 2021:1982-1991.

[5] Li X, Li S, Li Y, et al.Tongue image analysis for computer-aided diagnosis of traditional Chinese medicine[J]. Evidence-Based Complementary and Alternative Medicine, 2017, 25(10):1-11.

[6] Du L,Zhang H, Wang S, et al. A novel tongue image segmentation method based on the improved U-Net convolutional neural network[J]. Journal of Medical Systems,2020, 44(4):1-9.

[7] Liao X, Li Y, Xu M, et al. Tongue image analysis for computer-aided diagnosis of disease: a review[J]. Quantitative Imaging in Medicine and Surgery,2021, 9(6), 1096-1114.