

Research and application of tongue and face diagnosis based on deep learning

Digital Health
Volume 8: 1–18
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076221124436
journals.sagepub.com/home/dhj



Li Feng¹, Zong Hai Huang¹, Yan Mei Zhong¹, WenKe Xiao¹, Chuan Biao Wen¹,
Hai Bei Song¹  and Jin Hong Guo² 

Abstract

Objective: To explore the technical research and application characteristics of deep learning in tongue-facial diagnosis.

Methods: Through summarizing the merits and demerits of current image processing techniques used in the traditional medical tongue and face diagnosis, the research status of deep learning in tongue image preprocessing, segmentation, and classification was analyzed and reviewed, and the algorithm was compared and verified with the real tongue and face image. Images of the face and tongue used for diagnosis in conventional medicine were systematically reviewed, from acquisition and pre-processing to segmentation, classification, algorithm comparison, result from analysis, and application.

Results: Deep learning improved the speed and accuracy of tongue and face diagnostic image data processing. Among them, the average intersection ratio of U-net and Seg-net models exceeded 0.98, and the segmentation speed ranged from 54 to 58 ms.

Conclusion: There is no unified standard for lingual-facial diagnosis objectification in terms of image acquisition conditions and image processing methods, thus further research is indispensable. It is feasible to use the images acquired by mobile in the field of medical image analysis by reducing the influence of environmental and other factors on the quality of lingual-facial diagnosis images and improving the efficiency of image processing.

Keywords

Deep learning, traditional medicine, tongue diagnosis, face diagnosis, image processing

Submission date: 29 July 2022; Acceptance date: 18 August 2022

Introduction

Traditional medicine employs a variety of diagnostic methods, among which traditional Chinese medicine (TCM) diagnosis involves a doctor's visual observation of all external manifestations and discharges of the human body to determine the health status of the body.^{1,2} According to TCM, the manifestations of the face and tongue are important indicators reflecting the qi and blood status and overall functions of the body. However, the authenticity and accuracy of diagnosis and treatment information are easily affected by the clinical experience of the physician, the patient's expression, and the difference in the patient's constitution, so it is challenging to form a unified standard of TCM diagnosis. Presently, intelligent tongue

and facial diagnosis systems primarily apply modern instruments to acquire tongue and facial images³; it generally

¹Chengdu University of Traditional Chinese Medicine College of Intelligent Medicine, Chengdu, Sichuan, China

²University of Electronic Science and Technology of China, Chengdu, Sichuan, China

Corresponding authors:

Jin Hong Guo, University of Electronic Science and Technology of China, No. 2006, Xiyuan Avenue, West Hi-tech Zone, Chengdu, Sichuan 611731, China.
Email: guojinhong@uestc.edu.cn

Hai Bei Song, Chengdu University of Traditional Chinese Medicine College of Intelligent Medicine, Chengdu, Sichuan, China.
Email: songhaibei@cudtcm.edu.cn



involves data pre-processing, image positioning and segmentation, feature extraction, pattern recognition, and other steps to achieve the objective, quantified and informationized tongue and face diagnosis,⁴ so as to achieve the goal of high efficiency and high accuracy processing of tongue and facial diagnosis data.

In recent years, scholars have studied deep-learning-assisted medical diagnosis and its various explorations in the medical field. Gautam and Sharma⁵ focused on deep learning in diagnosing nervous system diseases, such as cerebrovascular diseases, Alzheimer's disease, Parkinson's disease, etc., described the performance of the convolutional neural network (CNN), recurrent neural network (RNN), restricted Boltzmann machine (RBM), deep belief network (DBN), deep Boltzmann machine (DBM), deep neural network (DNN) and deep learning technology, listed the optimal deep learning method to diagnose various diseases, and emphasized the relationship between deep learning method and those diseases. Ahsan and Siddique⁶ reviewed that when there is an imbalance in patient data, the utilization of deep learning methods can improve the performance of the models, but cannot prove the reliability of the models. Therefore, subsequent work is directed at extending deep learning experiments to increase the explanatory power of model predictions. Liu et al.⁷ proposed to use of an active matrix sensing array assisted with a machine learning approach to measure plantar pressure during walking, so as to improve the diagnostic process of LDD (lumbar degenerative disease). The system can classify common human movements such as squatting, squatting, jumping, walking, and jogging, and the accuracy is as high as 99.2%. It can accurately identify personal activities, the diagnostic accuracy of LDD patients is 100%, and give post-operative recovery evaluation. This research shows that deep learning can be applied to almost every aspect of medicine and poses certain far-reaching impacts. At the same time, deep learning can be combined with other technologies to play a substantial role in smart medicine, telemedicine, and other aspects in the future, for example, it can combine with the Internet of things technology to monitor and manage patients with psychological and neurological disorders⁸ and COVID-19 patients,⁹ which is conducive to the recovery of patients' physical and mental state. Therefore, deep learning technology can exert a great impact on the medical field; it has unique advantages in diagnosis, prognosis management, health monitoring, risk assessment, and so on. Combining deep learning technology with traditional medicine will improve the diagnosis and treatment efficiency of traditional medicine and contribute to the development of traditional medicine.

This work not only analyzed the development of tongue image processing technology and studied face image processing, but also summarized the hardware conditions of image acquisition. It is a systematic overview of the face and tongue images used in traditional medicine for

diagnosis, covering acquisition, preprocessing, segmentation, classification, algorithm comparison, result from analysis, and application. Following the evolution of mobile technology, real-time and universality are the direction of future medical development. The use of images collected by mobile terminals for disease diagnosis may provide new ideas for mobile medicine.

Related works

Equipment for tongue and face diagnostic image acquisition

Image acquisition under unrestricted scenarios. There would be no convenience due to the fixed location and time for using specialized image acquisition equipment such as tongue diagnostic devices. In contrast, mobile acquisition of tongue and face images with a greater range of applications is less demanding,¹⁰ so utilizing images taken by mobile devices for tongue diagnosis image analysis is the future trend of intelligent tongue and face diagnosis. The difference between the images captured by mobile and the images captured by standard tongue diagnostic devices lies in noise and picture resolution, and the images captured by mobile in the tongue image processing method require algorithms to exclude the influence of light and background. Song et al.¹¹ applied the algorithm of LBP (Local Binary Patterns) combined with Adaboost in a cell phone terminal to achieve near real-time tongue image detection, and the accuracy of image analysis was achieved by deep learning algorithm through a residual network without setting the scene. In order to construct a color correction function model for tongue images based on cell phone terminals, Wang et al.¹² compared tongue images taken by different models of cell phones with those taken by the TDA-1 micro-tongue diagnostic instrument, and analyzed more than 300 cases of tongue images for standard color card L*a*b color value acquisition; with the optimized back propagation (BP) neural network as the model, the constructed network can correct tongue color and moss color. Tongue images captured by mobile devices often feature in extremely complex lighting and background environments, therefore they need to be algorithmically pre-processed prior to analysis.

Image acquisition equipment under certain scenario situations

TCM physicians tend to determine the clinical significance of tongue manifestation and facial manifestation based on their own experience and treatment conceptions, which is highly subjective. By setting the parameters and using mathematical models to describe the established targets, the machine can objectively reflect the content of the pictures qualitatively, quantitatively, and locally.¹³ Since the

1980s, Yan et al.¹⁴ focused on TCM tongue-face diagnosis and conducted qualitative research on the objectification of the four diagnostic methods of TCM. In 2002, Wei et al.¹⁵ developed for the first time a computer-based digital TCM tongue analysis instrument, which broke through the subjective limitations of traditional TCM tongue diagnosis and was of great significance to the objectification of TCM clinical diagnosis and the development of TCM scientific research. Since then, the objectification of TCM tongue-facial diagnosis has entered a rapid development stage.

At present, the research on the objectification of tongue-face diagnosis centers on the following two aspects: (1) developing accurate instruments; (2) exploring specific indicators that can directly reflect the images of tongue and face diagnosis. In the process of tongue-facial diagnosis image acquisition, environmental factors usually directly affect the image quality and even the analysis results. In order to reduce the influence of these factors, Ding and Ding¹⁶ designed a tongue diagnostic instrument based on the theory of TCM inspection diagnosis, which maintains the authenticity of the tongue signs and excludes the environmental influence such as the background, while fixing the light source parameters under the set scene.

Research progress of the facial diagnostic instrument. The facial diagnostic instrument is designed based on the theory of TCM facial diagnosis, integrating facial diagnosis image acquisition, image segmentation, and information processing.^{17,18} The light source in the facial diagnostic instrument can affect the realism of facial color,¹⁹ so it is crucial to keep the stability of the light source. Li et al.²⁰ employed light-emitting diodes with color temperatures 4000 K–11,000 K as the light source of the dark box system for facial diagnosis information collection, which enhanced the standardization of facial diagnosis information collection; Shi et al.²¹ concluded that D50 can be treated as an illumination light source for objectified acquisition in comparison with natural light; Zheng et al.²² used a xenon lamp with a color temperature of 5500 K to simulate a daylight light source and used a digital camera with a resolution of 4752×3168 to acquire images. Their stability of the light source, color rendering, and uniformity of the light source were greater than 95%, thus the collected data could meet the needs of TCM color diagnosis and contribute to the objectified study of color diagnosis information.

Research progress of tongue diagnostic instrument. The tongue diagnostic instrument²³ mainly consists of two parts: the digital tongue image acquisition system and the tongue feature processing system.²⁴ Among them, the hardware of the digital tongue image acquisition system is mainly the combination of image acquisition equipment, light source, illumination environment, and image closure

platform. The tongue feature processing system²⁵ is to extract and identify tongue features such as tooth marks, fissures, sublingual vessels, and tongue coating, which mainly applies tongue classification and tongue segmentation (see Figures 1 and 2).

Jiang et al.,²⁶ who was inspired by the small handheld tongue analyzer, distributed tongue analyzer, and integrated sphere tongue diagnostic instrument, designed an all-in-one closed tongue acquisition system and investigated a method to manually adjust the camera white balance parameters. Their work laid the foundation for later image color correction. Zhang et al.²⁷ proposed a non-invasive method instrument based on three groups of tongue features that can capture the tongue image and complete image correction simultaneously and improve the quality of the tongue image. In order to solve the problem of the inconsistent color of the captured images during tongue objectification, some scholars compared and evaluated the imaging camera and illumination environment in the image acquisition equipment by three indexes: peak signal-to-noise ratio, blur coefficient, and quality factor for the imaging quality. The effect of the illumination environment on image color was explored using LED light sources of different wavelengths to produce a new tongue image acquisition device, and experiments proved that the color consistency of images acquired by the new device was significantly improved. The vast majority of tongue image acquisition devices are fixed posterior, and due to the different sizes and shapes of the patient's tongue, it cannot be guaranteed that the captured tongue images are complete and clear. Zhang et al.²⁸ proposed the concept of dynamic acquisition for the first time and realized vertical photography of the tongue body by controlling the camera with an embedded

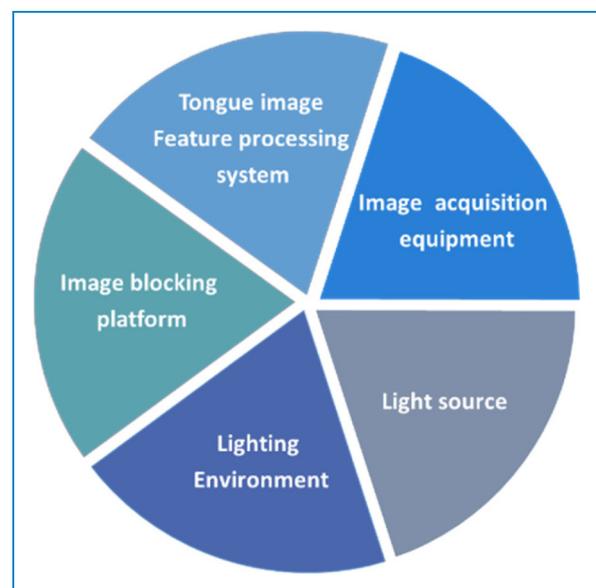


Figure 1. Composition of tongue diagnostic instrument.

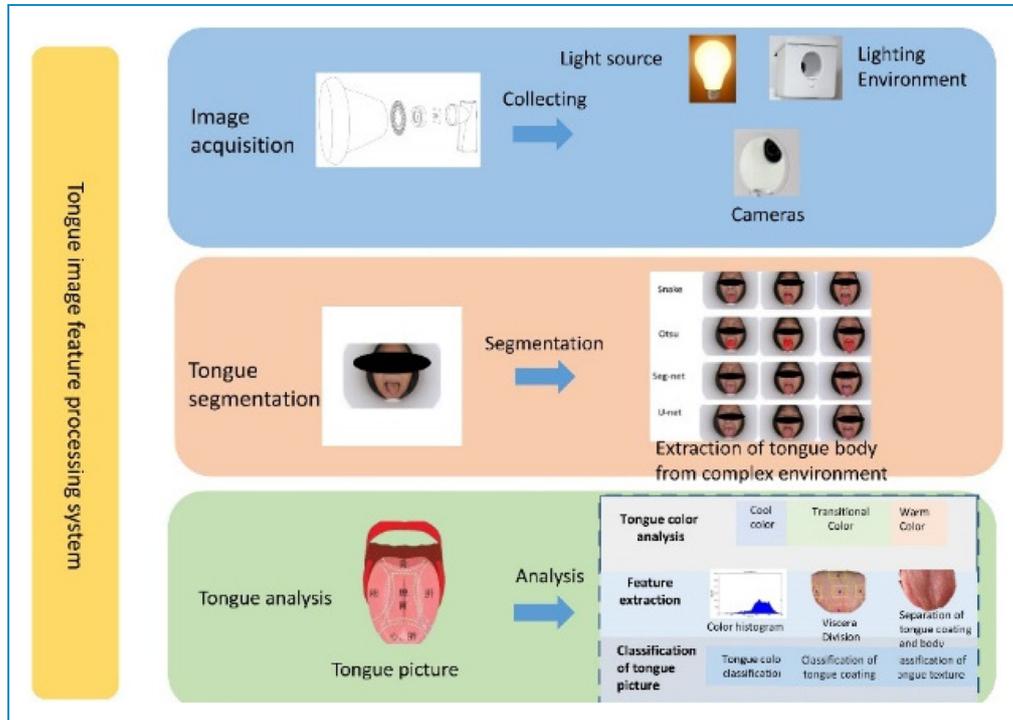


Figure 2. The constitution of tongue image feature processing system.

system. They achieved the clarity, completeness, fidelity, and approximation of tongue color in natural light, and obtained tongue pictures closer to those under natural light sources, overcoming the incomplete reflection of tongue information caused by the original static acquisition. After the tongue objectification instrument was put into clinical practice,²⁹ it was found that different subjects with different tongue exposure can cause inconsistent measurement results. Considering the ultrasonic ranging principle, the development of the specification of the tongue exposure of the subjects was helpful to improve the reliability of tongue diagnosis image data.

Separation of the image of tongue texture from coating³⁰ is difficult for characteristic processing of tongue image. The traditional K-means clustering algorithm unstably separates the image of tongue texture from the coating. Thus, Li and Li³¹ proposed an optimized K-means clustering model for texture and coating image separation, which integrated three color space properties of RGB, HSV, and $L^*a^*b^*$. After comparing the effect of single-channel, dual-channel, and three-channel tongue image clustering, it was confirmed that the dual-channel tongue image clustering was better than the other two segmentation schemes. And the single-channel tongue image classification histogram peak as the initial clustering centroid to separate the tongue coating and tongue texture are more accurate and stable than the traditional K-means clustering algorithm. Tongue fissures are relatively special features, in order to extract tongue fissures accurately, Zhang et al.³² proposed

a fissured tongue detection method based on a local grayscale threshold. It can separate the cracked region from the background region by local grayscale squared difference after operations such as Laplace enhancement, median filtering to remove noise, and regional consistency principle to determine whether cracks exist. This method can effectively and easily separate tongue cracks and background, which provides powerful technical support for tongue diagnosis objectification. The sublingual vein condition can provide an auxiliary diagnostic basis, but the size of the sublingual vein is much smaller than that of the tongue, and the difficulty of sublingual vein segmentation training is also much higher than that of tongue segmentation training. To improve the accuracy of sublingual vein segmentation, Yang et al.³³ proposed a collaborative attention network that decomposes the entire coding and decoding framework and collaboratively updates the parameters, which improves the training speed and maintains the training stability.

Research and progress of image processing algorithms

After the lingual diagnosis images are captured and transferred to the database, the image information needs to be identified, read and output completely and accurately. Algorithms of lingual diagnosis image processing can improve the speed and accuracy of information processing.

Early lingual diagnosis objectification was to convert the collected light of different wavelengths into digital signals by means of a photoelectric sensor/device. In recent years, lingual-facial diagnosis objectification is done by capturing images with a camera, pre-processing the images, and using machine learning to segment, and classify the lingual-facial images^{34,35} (see Figure 3).

Facial diagnosis image segmentation algorithm

There are few algorithms for facial diagnosis image segmentation and classification. Traditional image segmentation methods do not take into account the characteristics of the strong stereo nature of facial images, which brings about problems such as insufficient processing of facial details and distortion of chromaticity information in color space conversion. Liu et al.³⁶ designed an automatic facial image segmentation algorithm for TCM facial diagnostic instruments, using a grayscale adaptive enhancement method for pre-processing. On this basis, it combined with the adaptive nonlinear conversion method to reduce the risk of color conversion distortion, while the details were processed using clustering methods and mathematical morphological operations, accurately analyzing the facial diagnosis map. Lin et al.³⁷ combined elements of color space theory, statistical features of facial texture, and lip color features, and used machine learning methods such

as KNN, support vector machine (SVM), and BP neural network to classify and recognize the extracted facial features to achieve automatic segmentation of facial regions, whose best recognition rate reached 91.03%. To solve the visual inconsistency caused by different angles in the face diagnosis image acquisition process, Ning and Chen³⁸ adapted a columnar projection method based on facial features and combined the SIFT (Scale-invariant feature transform) feature matching algorithm and the RANSAC (Random Sample Consensus) matching optimization algorithm to extract image feature vectors. Thus it can eliminate matching errors and enable efficient, accurate, and smooth matching of images, which can generate face images quickly and effectively, contributing to the development of objectification in TCM face diagnosis.

Facial expressions can convey a variety of emotions; those emotions are directly related to the patient's disease condition, so the recognition of facial expressions is also part of intelligent inspection diagnosis. Huang et al.³⁹ used a residual deep neural network of internal evolutionary mechanisms and feature fusion algorithms to recognize image facial expressions. It achieved high recognition rates in both standard and low-quality face datasets. This method eliminates the effect of lighting conditions, occlusions, and age changes on the image quality. The facial expression changes are very subtle and the limitations of the technique lead to a low facial expression recognition rate. Some scholars have applied CNN

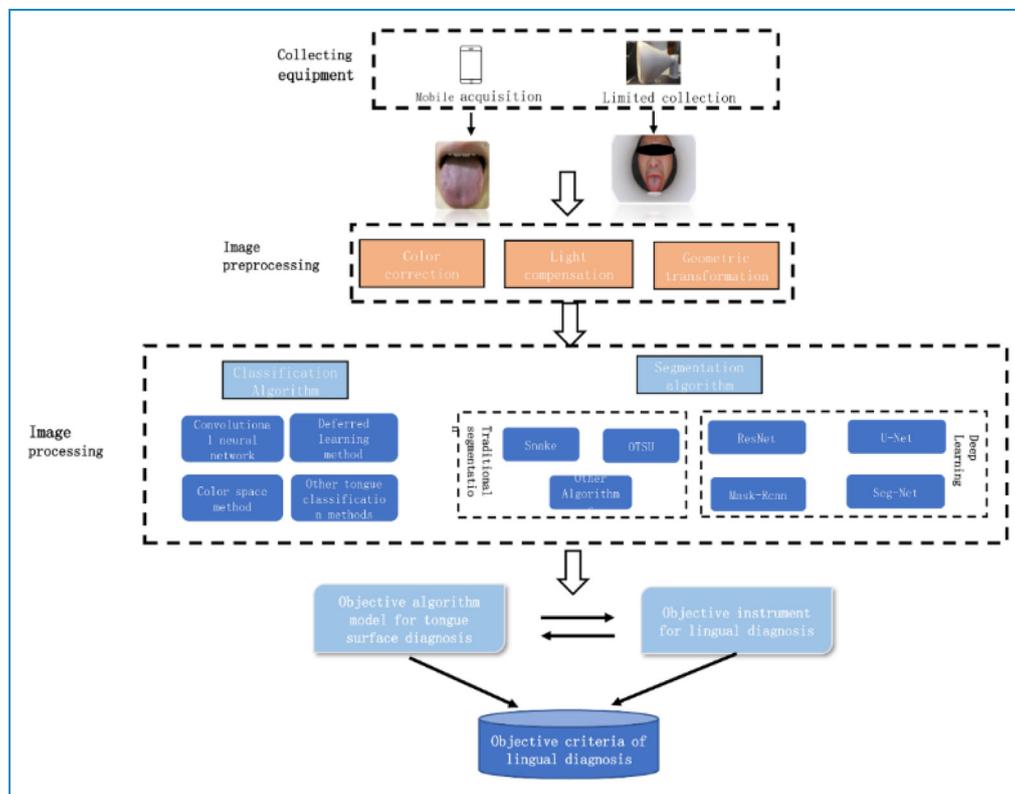


Figure 3. Flowchart of objectification of lingual diagnosis.

(convolutional network) models to expression recognition, and Jin et al.⁴⁰ used a model based on the VGG network structure for image training, which was optimized to continuously reduce the loss rate and improve the accuracy of recognition. Wu et al.⁴¹ designed a three-dimensional convolutional neural network (3D-CNN) micro-expression recognition algorithm, adding a batch normalization algorithm as well as a discard method on top of 3D-VGG-Block to improve the network depth and training speed and prevent data overfitting. At present, few studies incorporate facial expressions into the objectification index of facial diagnosis, but facial expressions possess special significance in TCM diagnosis, and the study of facial expression objectification is going deeper and in a high development period.

Face classification algorithms can be used not only for analyzing facial features but also for gender classification. In the process of social software access, identity authentication and login permission are achieved through facial recognition. Accurate gender classification can improve the accuracy of facial recognition⁴² and reduce the difficulty of recognition. Fekri-Ershad⁴³ designed a rotation invariant method for gender classification of face images based on an improved version of the local binary pattern (iLBP), which reduces the computational complexity of smartphone applications, reduces memory and CPU usage, and improves the performance of synchronous applications in smartphones. Zhang et al.⁴⁴ proposed a new method based on a multi-scale facial fusion feature (ms3f) to classify gender from facial images. The fusion features are extracted by the combination of a local binary pattern (LBP) and local phase quantization (LPQ) descriptors, and multi-scale features are generated by multi-block (MB) and multi-level (ML) methods. SVM is used as a classifier for gender classification. Those results indicate that the application of multi-scale fusion features greatly improves the performance of gender classification. In the future, gender classification can be applied to many scenes such as smartphones and intelligent medical treatment. The progress of facial image processing technology can promote the accuracy of gender classification.

Progress of tongue image feature processing algorithm

Before the tongue facial diagnosis image is input into the model, image preprocessing is required.⁴⁵ Image preprocessing includes light compensation, color correction,⁴⁶ geometric transformations, etc. Light compensation is widely utilized in image processing, and common light compensation methods are the Gray World color equalization algorithm and the white reference image-based algorithm. However, Yu et al.⁴⁷ proposed a method to remove the light imbalance without a white reference image, where the background is obtained by segmenting the image, estimating the light difference from different background

images, and then performing light compensation. This method can be used to segment microscopic medical images and also other medical images. The tongue images acquired by the camera produce noise due to uneven illumination, which seriously affects the quality of the images.⁴⁸ This can be done in terms of low-level image processing, point processing of the image, median filtering of the image, etc., by thresholding, expanding the contrast, reducing the gray value, and filtering the noise. This method can reduce the deviations generated during preprocessing while improving the image processing speed.

Color correction mainly consists of ambient lighting conditions correction, color space correction, color card color correction, and algorithm correction. Ambient lighting conditions correction specifies the illuminant and light source to complete color correction while capturing images. However, there are still obvious differences between standard light sources and natural light sources, and factors such as distance can also affect the illumination effect, so further correction by algorithms is still required subsequently. Shang et al.⁴⁹ superimposed the positive image of the healthy tongue image and the color negative image of the diseased tongue image, selected the information-sensitive area of the tongue image, and made use of the conversion relation between RGB and CIE Lab color model, the discrete chromatographic distribution is obtained, and the Lab color model is proved to be suitable for color discrimination. Yang et al.⁵⁰ explored the role of chromaticity for tongue color quantification and emphasized the significance of establishing color correction specifications for building a unified tongue image library. Due to physical characteristics, images acquired by different standardized devices still have color differences. He and Du⁵¹ converted color images to grayscale images and then converted grayscale images to binary images that can be entered into algorithmic models. Image preprocessing enhances the detectability of tongue diagnosis information in terms of light and color and lays the foundation for subsequent image processing.

In the process of image collection, due to different acquisition instruments, the size of the image is different. In order to improve the training efficiency and save memory space, the deep learning network has restrictions on the input pictures, so it is necessary to unify the picture size during preprocessing. The processing method is generally to reshape the input image size to facilitate model training. The commonly used reshaping methods are "interpolation," "cropping," "inclusion," "tiling," and "mirroring."⁵²

Progress of tongue image classification algorithm research

Tongue classification⁵³ covers tongue color classification and tongue coating classification, which are mainly designed to classify the features of the tongue, such as

tongue color, tongue coating, coating color, etc. The improvement of the tongue color classification algorithm is mainly to improve the speed and accuracy of tongue color and coating color discrimination, while the improvement of the tongue coating classification algorithm improves the classification accuracy of certain features of the tongue, such as cracks and teeth marks, which together contributes to the development of the tongue classification algorithm. In recent years, lingual classification algorithms based on deep CNN lingual classification model^{54,55} is a hot research topic, and to solve the problems of high training equipment requirements and long training time in deep learning in lingual classification training, and migration-learning-based lingual research methods are proposed. The current research on tongue image classification algorithms is shown in Table 1.

Advances in tongue segmentation methods

Traditional tongue segmentation methods. Compared with lingual image classification methods, lingual image segmentation methods have developed more rapidly. Traditional tongue image segmentation algorithms can be divided into five categories: region-based segmentation algorithms, active contour model-based algorithms,^{63–65} threshold-based segmentation methods,⁶⁶ edge-based segmentation methods, and graph theory-based segmentation methods. In recent years, the rise of deep learning-based lingual segmentation methods has brought breakthroughs in lingual image processing.

The early tongue image segmentation algorithms are mainly based on template matching,^{67–69} especially the Snake algorithm. The Snake algorithm needs to be given

Table 1. Research status of tongue image classification algorithm.

References	Methods	Classification object	Indicator	Effect
Tang et al. ⁵⁶	A joint multi-task learning model based on ImageNet and RPN network	Tongue color, coating color, fissure lines, tooth marks	Speed Accuracy	Speed compared to the traditional method 40% faster and more accurate 33% higher
Xiao et al. ⁵⁷	Tongue color classification method based on AlexNet network	Tongue color	Accuracy	94.85%
Yang and Zhang ⁵⁸	Based on Inception_v3 + 2NN and Inception_v3 + 3NN tongue classification methods	tongue features	Accuracy	Inception_v3 + 2NN Classification accuracy 90.30% Inception_v3 + 3NN Classification accuracy 93.98%
Song et al. ⁵⁹	Based on GoogLeNet, ResNet and ImageNet for deep migratory learning method for tongue study	Tongue features	Accuracy	Inception-v highest accuracy 94.88%
Kanawong et al. ⁶⁰	Feature extraction based on HSV and RGB extraction methods to build a linguistic classification model	Tongue color tongue coating	Comprehensive evaluation index (F-Measure)	F-measure mean 0.837
Yan et al. ⁶¹	YoloV5 deep learning algorithm based and random forest for dentate tongue classification method	Tongue features (tooth marks)	Accuracy	93.7%
Jiao et al. ⁶²	Weighted SVM method	Unbalanced tongue samples	Accuracy	Weighted SVM method accuracy rate is higher than the standard SVM method by more than 20% on average

an initial contour curve, and then the initial contour curve is adjusted to evolving to the real contour curve, so the main point of the Snake algorithm research is the acquisition of the initial contour and curve evolution. The initial contour curve is a set of control points that can control the shape of the curve, and all the points meet at the beginning and end to form a closed contour line representation. Assuming that $x(s)$ and $y(s)$ denote the coordinate position of each control point in the image, respectively, and s is the independent variable describing the boundary in the form of Fourier transform which is the arc length, then the energy function of the Snake curve is expressed as

$$\begin{aligned} E_{snake}^* &= \int_0^1 E_{snake}(v(s)) ds \\ &= \int_0^1 E_{int}(v(s)) ds + E_{image}(v(s)) + E_{con}(v(s)) ds \quad (1) \\ &= \int_0^1 E_{int}(v(s)) ds + E_{ext}(v(s)) ds \end{aligned}$$

where E_{int} is the internal energy. E_{image} is the image energy; E_{con} is the external constraint energy, and the image energy and external constraint energy are collectively referred to as the external energy.

The internal energy consists of the elastic energy (modulus of the first-order derivative) and the bending energy (modulus of the second-order derivative).

$$E_{int} = \frac{1}{2}(\alpha(s)|v_s(s)|^2 + \beta(s)|v_{ss}(s)|^2) \quad (2)$$

$$E_{image} = \omega_{line}E_{line} + \omega_{edge}E_{edge} + \omega_{term}E_{term} \quad (3)$$

Linear energy:

$$E_{line} = I(x, y) \quad (4)$$

Edge energy:

$$E_{edge} = -|\nabla I(x, y)|^2 \quad (5)$$

End energy:

$$\begin{aligned} E_{term} &= \frac{\partial \theta}{\partial n_{\perp}} = \frac{\partial C^2 / \partial n_{\perp}^2}{\partial C / \partial n} \\ &= \frac{C_{yy}C_x^2 - 2C_{xy}C_xC_y + C_{xx}C_y^2}{(C_x^2 + C_y^2)^{3/2}} \quad (6) \end{aligned}$$

Solving for: derivative

$$\begin{aligned} \alpha v_{ss}(s) - \beta v_{ssss}(s) - \nabla E_{image}(v(s)) - \nabla E_{con}(v(s)) \\ = 0 \quad (7) \end{aligned}$$

The optimization objective is to locally minimize the total energy function. The optimization objective is that the total energy function is locally minimum. The termination of the iteration is controlled by the minimal energy function or the number of iterations, and the minimized energy

function is solved by the Euler equation, based on $v(s) = [x(s), y(s)]$, rewriting the above equation as:

$$\begin{cases} \alpha x_{ss}(s) - \beta x_{ssss}(s) - \frac{\partial E_{ext}}{\partial x} = 0 \\ \alpha y_{ss}(s) - \beta y_{ssss}(s) - \frac{\partial E_{ext}}{\partial y} = 0 \end{cases} \quad (8)$$

Introducing external energy:

$$\begin{aligned} \alpha_i(v_i - v_{i-1}) - \alpha_{i+1}(v_{i+1} - v_i) + \beta_{i-1}[v_{i-2} - 2v_{i-1} + v_i] \\ - 2\beta_i[v_{i-1} - 2v_i + v_{i+1}] + \beta_{i+1}[v_i - 2v_{i+1} + v_{i+2}] \\ + (f_x(i), f_y(i)) = 0 \quad (9) \end{aligned}$$

Into an iterative evolutionary process with each step:

$$\begin{cases} x_t = (A + \gamma I)^{-1}(x_{t-1} - f_x(x_{t-1}, y_{t-1})) \\ y_t = (A + \gamma I)^{-1}(y_{t-1} - f_y(x_{t-1}, y_{t-1})) \end{cases} \quad (10)$$

Pang et al.^{70,71} proposed a double elliptic deformation contour method based on the Snake algorithm, which combines a double elliptic deformation template (BEDT) and an active contour model for obtaining the approximate shape features of the target and fitting the local details. The BEDT deformation template is used to roughly describe the tongue body. The initial contour curve is obtained by converting the BEDT energy function and replacing the traditional internal energy of the active contour model to evolve the curve, which leads to the tongue segmentation results. Ma et al.⁷² proposed a dual Snake algorithm based on the Snake model. After the median filter removes the noise image and transforms the color space, the complete tongue image is obtained using the double Snake model, which has better tongue segmentation performance. However, the Snake algorithm requires multiple interaction work, and if the initial contour of the algorithm is not accurately delineated, it will directly lead to subsequent segmentation underfitting, so the limitations of the Snake algorithm are relatively obvious. It is gradually replaced by other image segmentation algorithms at a later stage.

Thresholding is the classical image segmentation method, of which the OTSU algorithm has a wide range of applications. The OTSU algorithm uses one or more thresholds to calculate the grayscale image on the original image, compare the grayscale value of each pixel on the image with the corresponding threshold, and delineate the tongue image and the background, respectively according to the comparison results. Zhang et al. proposed to combine the threshold segmentation method with the grayscale projection method,^{73,74} Snake active contour method,⁷⁵ and HSV spatial thresholding method⁷⁶ for tongue image segmentation, however, further improvements can be realized in terms of algorithm performance. Jiang et al.⁷⁷ designed a tongue image segmentation algorithm by combining the Otsu algorithm and mathematical morphological algorithm. After the

Otsu threshold segmentation image, the morphological opening operation is added to the correct the tongue image. However, this method involves more operations in the correction step, resulting in a longer segmentation time and lower segmentation efficiency. Huang et al.⁷⁸ proposed a tongue image segmentation method based on Otsu and regional sub-block growth, incorporating classification based on image color features and a fast merging algorithm of sub-blocks after classification to achieve automatic tongue segmentation. Zhang and Liu⁷⁹ proposed a tongue extraction algorithm based on dynamic thresholding and correction model. The HSI color model^{80,81} was used to extract information such as lips and face from the tongue image. A dynamic threshold segmentation algorithm was used to extract the initial outline of the tongue body and a tongue body correction model was used to obtain the final tongue body. The data proved that the images obtained by this extraction method are superior in terms of noise immunity and accuracy and also have a good segmentation effect in the concave and convex areas of the tongue body.

Otsu algorithm is also called the adaptive threshold method. Referring to the reverse derivation of the adaptive threshold method, the following equation can be obtained:

Parameters:

TH: assumed threshold

C1: pixels less than the threshold TH

C2: pixels greater than threshold TH

p_1 : probability that the pixel is divided into C1

p_2 : probability that pixels are divided into C2

m_1 : average value of all pixels less than TH

m_2 : average value of all pixels greater than TH

mG : average value of all pixels in the image

$$p_1 * m_1 + p_2 * m_2 = mG \quad (11)$$

$$p_1 + p_2 = 1 \quad (12)$$

According to the concept of variance, the expression for the variance between classes is

$$\sigma^2 = p_1(m_1 - mG)^2 + p_2(m_2 - mG)^2 \quad (13)$$

Substituting (1) into (3), we get

$$\sigma^2 = p_1 p_2 (m_1 - m_2)^2 \quad (14)$$

The gray level k that maximizes the above equation is the OTSU threshold

$$p_1 = \sum_{i=0}^k P_i \quad (15)$$

$$m_1 = \frac{1}{p_1 * \sum_{i=0}^k i P_i} \quad (16)$$

$$m_2 = \frac{1}{p_2 * \sum_{i=k+1}^{L-1} i P_i} \quad (17)$$

The cumulative mean of the gray levels

$$m = \sum_{i=0}^k i P_i \quad (18)$$

Global mean of the image

$$mG = \sum_{i=0}^{L-1} i P_i \quad (19)$$

The final inter-class variance equation

$$\sigma^2 = \frac{(mG * p_1 - m)^2}{p_1(1 - p_1)} \quad (20)$$

Deep-learning-based tongue segmentation methods

Deep-learning-based tongue image segmentation models are mainly divided into two categories,⁸² one is U-Net^{83,84} and Seg-Net^{85,86} evolved from the fully convolutional network (FCN), and the other is Mask R-CNN improved from CNN,⁸⁷ which are widely used in various types of medical image segmentation. FCN, the pioneer of semantic segmentation, achieves pixel-level classification of images. The FCN model can accept image input of any size and upsample the feature map of the last convolutional layer using the deconvolution layer to recover it to the same size as the input image. The spatial information in the original input image is preserved, and then the upsampled image is classified pixel by pixel, and the softmax classification loss is calculated pixel by pixel, that is, each pixel corresponds to one training sample. So the features of FCN include full convolution, upsampling and skip-level structure (see Figure 4).

The U-Net model, an image processing method based on the FCN architecture,⁸⁸ is one of the three most classical F-CNNs (fully CNNs) models. It can handle a wide range of medical image segmentation tasks independent of different organs and imaging modalities. Li et al.⁸⁹ developed a U-Net model for accurate retinal vascular segmentation, which improved the existing image segmentation methods. Since organ image segmentation^{90,91} is error-prone due to the inhomogeneous and irregular shape of organs, Li et al.⁹² designed an attention-based nested segmentation network, ANU-Net.

They introduced an attention mechanism between nested convolutional blocks, allowing features extracted at different levels to be selected with task-relevant fusion. The model is also designed with a hybrid loss function to make full use of the resolution feature information.

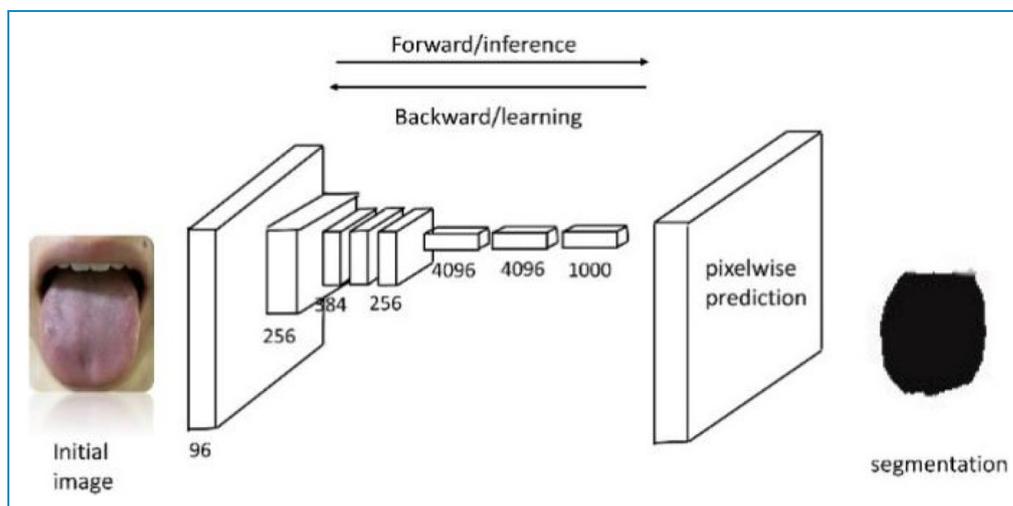


Figure 4. Fully convolutional network (FCN) model diagram.

ANU-Net is experimentally shown to be competitive among various medical image segmentation methods. Due to the increased demand for various medical image segmentation, the U-Net model plays an important role because of its excellent segmentation accuracy and robustness,⁹³ and the use of the U-Net model for tongue segmentation is a future trend for tongue objectification (see Figures 5 and 6).

CNN is a convolutional layer followed by some fully connected layers to obtain a fixed-length feature vector for classification, so the CNN structure is suitable for image classification and regression tasks. CNN architectures can be divided into two types, the first category includes Le-Net,⁹⁴ Alex-Net,⁹⁵ Res-Net,⁹⁶ etc., which are mainly used for image classification. LeNet is the modern convolutional neural network foundation and provides ideas for the later development of deep learning network structures. LeNet is mainly used for the recognition and classification of handwritten characters with an accuracy of 98%. AlexNet extends based on the LeNet network to reduce the model training time and enhance the generalization ability of the model. The emergence of ResNet (residual neural network) deepens the neural network training depth to avoid near saturation of accuracy.

The second class of CNN architectures is mainly used for target detection, including R CNN,⁹⁷ Fast R CNN,⁹⁸ Faster R CNN,⁹⁹ Mask-R CNN,^{100,101} etc., among which medical image segmentation mostly uses Mask-R CNN structure. Yan et al.¹⁰² came up with a tongue image segmentation method based on the convolutional neural network Mask-R CNN, and the obtained tongue edges were more accurate. The average pixel accuracy of four quantitative evaluation indexes was 93.03%, the average cross-merge ratio was 86.69%, and the average frequency-weight cross-merge ratio was 87.16%, which better improved the problem of unclear tongue segmentation

contours. Zhang et al.¹⁰³ also designed an end-to-end tongue image segmentation method for the unclear tongue segmentation problem,¹⁰⁴ which combined DCNN and fully connected CRF to refine the segmented image edges. This algorithm outperforms traditional tongue image segmentation algorithms and mainstream deep learning methods, with an average intersection ratio of 95.41% for segmentation accuracy.

There are some problems in the traditional semantic segmentation task: successive post-pooling downsampling in CNN leads to spatial resolution degradation; scale detection requires rescaling and aggregating feature maps, which leads to the excessive computational effort. The image classification task needs to ensure that the spatial transformation is invariant, so the Deeplab structure is introduced. Deeplab v1 is modified from VGG16 by converting the fully connected layers of VGG16 to convolution, and the last two pooling layers are removed and use null convolution. Deeplab v2 uses the ASSP base layer on top of the Deeplab v1 structure, and Deeplab v3¹⁰⁵ improves the ASSP module by proposing a more general framework that can be applied to all networks and improves the efficiency of image segmentation.

In addition to mainstream deep learning algorithms for tongue image segmentation, scholars also proposed to compare the chan vese method¹⁰⁶ with the canny algorithm. The results proved that the canny algorithm generates a large number of pseudo-edges after edge cutting, while the chan vese method can automatically select the best edge information and has a greater clinical value than the canny algorithm. Li et al.¹⁰⁷ proposed to use of calibrated neural networks mixed with other deep learning methods for improving the accuracy of tongue detection and changing the accuracy of tongue recognition taken under natural conditions. Li et al.¹⁰⁸ proposed to capture tongue images using hyperspectral tongues and then use hidden

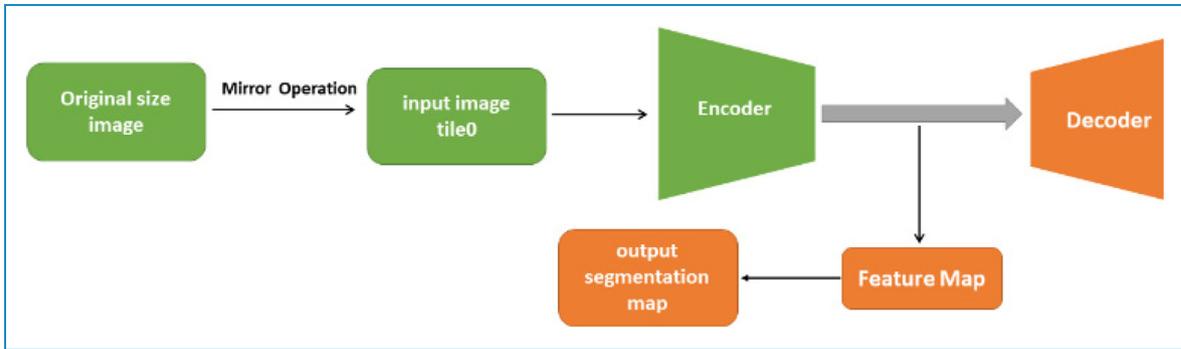


Figure 5. U-Net algorithm composition.

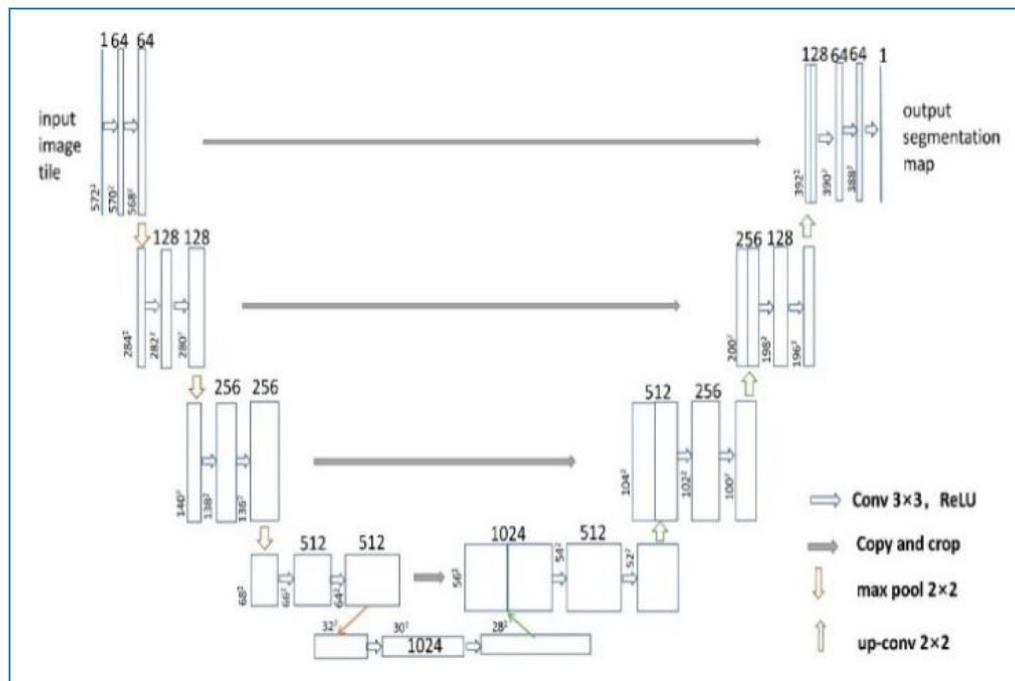


Figure 6. U-Net algorithm diagram.

Markov models to classify tongue fissures into 12 classes, and the method showed good performance in tongue classification. Shi et al.¹⁰⁹ proposed a double geo-vector flow (DGF) based tongue edge detection method, which is able to detect tongue edges and segment tongue regions for geological gradient vector flow evaluation of the tongue. Cui et al.¹¹⁰ proposed a new tongue segmentation method, GaborFM. This method uses a fast-marching algorithm to connect discontinuous contour segments to form a closed continuous tongue contour for ACM-based tongue segmentation. Qualitative and quantitative results show that the GaborFM method outperforms other methods; Zhou et al.¹¹¹ proposed a semi-supervised probabilistic model for tongue segmentation based on a combination of image reconstruction constraints and adversarial learning and an

inference model with a segmentation decoder and a reconstruction decoder. The effectiveness of this method was verified by generating tongue segmentation images and reconstructing tongue images, and supervising the images using a discriminator. All these models facilitate the process of tongue diagnosis objectification and provide new ideas for tongue diagnosis objectification. At the same time, these methods can only provide a reference for tongue diagnosis objectification and cannot form a standardized and effective set of tongue diagnosis objectification criteria, so further research and development are still needed for tongue diagnosis objectification in the future. In the last two years, tongue image segmentation algorithms have been developed based on traditional tongue image segmentation methods and deep learning algorithms. To

solve the problems of edge blurring and detail interference during tongue extraction, Huang et al.¹¹² designed an automatic tongue image segmentation method using an enhanced full convolutional network with an encoder–decoder structure with an average sensitivity of 98.97%, which is better than the four algorithms SegNet, FCN, PSPNet, and DeepLab v3+. Gao et al.¹¹³ proposed LSM-SEC based on convolutional neural networks as a model combining symmetric and edge-constrained level sets of tongue geometric features for tongue segmentation, which is suitable for tongue image segmentation under most conditions and can also improve the accuracy of subsequent model evolution.

These models promote the objectification process of tongue diagnosis and provide new ideas for the objectification of tongue diagnosis. However, these methods can only provide a reference for the objectification of tongue diagnosis, and cannot form a set of standardized and effective objectification standards of tongue diagnosis. Therefore, the objectification of tongue diagnosis in the future still needs further research and development. As the multi-headed attention mechanism has been proposed and mostly applied to human-computer interaction, such as speech emotion recognition, there have been attempts to apply the multi-headed attention mechanism to image segmentation. Lin et al.¹¹⁴ proposed a module for enhancing and fusing convolutional feature maps using a multi-headed attention approach, optimizing the scene recognition feature transformation algorithm and obtaining a framework that outperforms many state-of-the-art techniques. This demonstrates that the rapid development of deep learning algorithms offers the possibility of applying more novel algorithmic models to lingual diagnosis image segmentation (see Figure 7).

Performance comparison of tongue image algorithm segmentation

In this study, four models, Snake, Otsu, Seg-Net, and U-Net, were used for tongue image segmentation, and the algorithm performance was assessed by three parameters: time, memory usage, and accuracy. The Snake algorithm model is highly dependent on the initial contour correctness, and its iterative result accuracy is proportional to the initial contour accuracy. The Otsu algorithm is an algorithm that determines the image binarization segmentation threshold algorithm and maximizes the inter-class variance between foreground and background images after the binarization segmentation of images based on grayscale feature pairs. It is computationally simple and not affected by image brightness and contrast, so it is widely used in the field of digital image processing. While, it is sensitive to image noise, and the segmentation effect is poor when the target and the background-size ratio is disparate, so it is not suitable for images captured too far or too close. Seg-Net and U-Net both belong to the FCN semantic segmentation algorithm, and the semantic segmentation algorithm is to label each pixel point of the image with a category that is related to both the neighboring pixel category and the overall category to which this pixel point belongs. Using the image classification network structure, the determination needs can be satisfied by different levels of feature vectors. U-Net is one of the early algorithms using multi-scale feature semantic analysis; the unique U-shaped structure also provides ideas for the development of subsequent algorithms. The disadvantage is that the effective convolution increases the difficulty and generalization of the model design. Seg-Net uses a fully

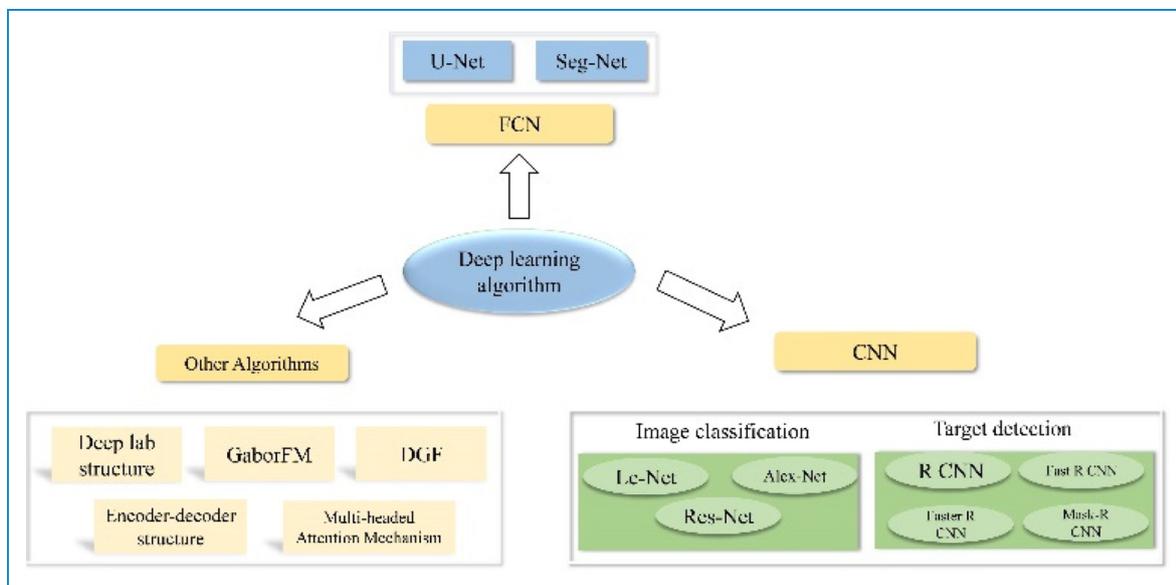


Figure 7. Types of deep learning algorithms for tongue and face image segmentation.

symmetric structure, convolution, and deconvolution symmetry, pooling and up pooling symmetry, and the overall composition of the encoder-decoder structure, which makes up for the disadvantage of the lower resolution of FCNN segmented images.

Methods

The evaluation metrics of image segmentation generally include inference time, memory occupation, and accuracy. Inference time is a very valuable metric to judge the segmentation efficiency, and the training speed is affected by hardware devices and background implementations, etc., which cannot fully represent the segmentation efficiency. But inference time can evaluate the effectiveness of segmentation methods and is the fastest method running in the same environment. Memory is another important factor in evaluating segmentation methods. The occupancy of an algorithm in image processing can affect the processing efficiency, and the CPU occupancy reflects the efficiency of the algorithm, so the memory occupancy is directly related to the processing efficiency when the running time is the same. There are many metrics to measure algorithm accuracy in image segmentation, usually variations of pixel accuracy and IoU, including PA (pixel accuracy), MPA (mean pixel accuracy), MIoU (mean intersection ratio), and FWIoU (frequency-weighted intersection ratio), with MIoU being the most commonly used metric due to its simplicity and representativeness.

To compare the segmentation performance of the algorithms, real tongue images captured by the tongue diagnostic instrument were used to compare the performance of Snake, Otsu, Seg-Net, and U-Net algorithms were validated. The implementation of this experiment is based on

the windows server2016 operating system. The CPU is Intel (R) Xeon (R) e5-2678 V3, the GPU is Tesla V100, and the CUDA version is v 10.1. In the environment of python3.7.4, Tensorflow 2.0.2, and Keras 2.3.1 are used as the main model framework to build the model.

First, the image is preprocessed. Because there may be pixel differences in the image, it is necessary to normalize the image to the same size first. For different pictures, we first fill their edges, take the long edge of the image as the target filling length, and fill the white edge to make it a square with the long edge of the image as the side length. The filled square resizes a standard image of $512 \times 512 \times 3$.

After completing the construction of the four algorithm models, 1233 preprocessed training set images are input into the algorithm for algorithm training. After the training, 98 random tongue images were input into the model to test, and the accuracy and processing speed of the algorithm was verified from the aspects of PA, runtime, and miou3 indicators. PA refers to the pixel accuracy, which is the proportion of the number of pixels with correct prediction categories to the total number of pixels. In deep learning, runtime refers to the running time of the algorithm. Miou is the ratio of the intersection and union of the predicted results of each category and the real values of the model. The sum and re-average results are used to measure the accuracy of image segmentation. The validation results are shown in Figure 8 and Table 2.

Results

With the same sample size, the Snake algorithm was more primitive in terms of operation speed, segmentation accuracy, and accuracy among the four algorithms. The Otsu

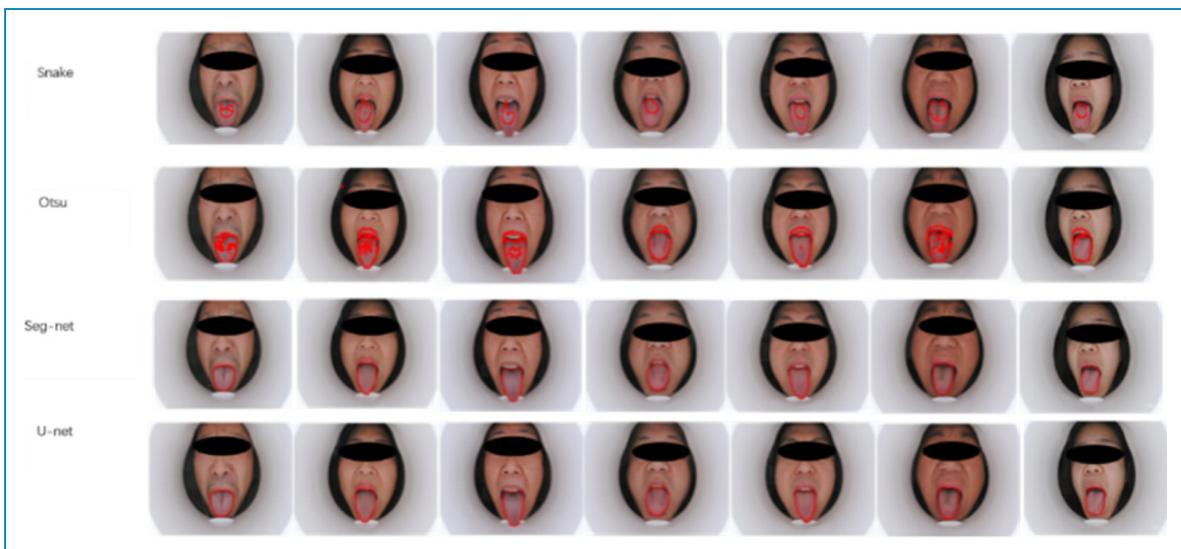


Figure 8. Four kinds of tongue image segmentation contrast map.

Table 2. Comparison results of four tongue segmentation algorithms.

Algorithm	Snake	Otsu	Seg-Net	U-Net
Sample size of training set	1233	1233	1233	1233
Sample size of test set	98	98	98	98
PA	0.988	0.983	0.999	0.999
Runtime	14,645 ms	6070 ms	54 ms	58 ms
MIOU	0.536	0.594	0.945	0.957

algorithm outperformed the Snake algorithm in terms of operation speed and segmentation accuracy, but the accuracy was still not significantly improved. The Seg-Net algorithm and the U-Net algorithm, compared with the first two, have substantially improved in terms of operation speed, segmentation accuracy, and accuracy improvement, and the performance difference between these two algorithms is small. It indicates that the depth learning method is superior to the traditional image processing methods in terms of operation speed and segmentation accuracy.

Discussion

These four algorithms performed differently and each with pros and cons. The advantage of the snake algorithm is that it can effectively find the contour of the target, and has a good ability to extract and tracking ability for the edge of the target in a specific region. The limitation of the snake algorithm is that it needs many interactions. If the initial contour of the algorithm is not accurately described, it will directly lead to subsequent segmentation underfitting. Therefore, the limitations of the snake algorithm are obvious.

Otsu algorithm excelled in its calculation simplicity and is not affected by image brightness and contrast. The limitation of the Otsu algorithm is that when the area of the target and the background in the image is very different, there is no obvious double peak in the histogram, or the size of the two peaks is very different, and the segmentation effect is poor, or there is a large overlap between the gray level of the target and the background, it cannot accurately separate the target and the background.

SEG net algorithm performed excellently in voice, semantic, visual, and various game tasks. The algorithm can be adjusted quickly to adapt to new problems. The limitation is that a large amount of data is needed for training, which requires high hardware configuration. The model is

in a black box state, and it is difficult to understand the internal mechanism.

The advantage of the u-net algorithm is that the u-net supports a small number of training models and introduces image mirroring operation to better train data; Each pixel can be segmented to obtain higher segmentation accuracy. The limitation is that effective convolution increases the difficulty and universality of model design.

Conclusion

As an important diagnosis and treatment method in TCM, the inspection diagnosis is primarily a medical activity performed through the physician's subjective judgment of the patient's signs and symptoms. A majority of doctors vary greatly in the efficiency of diagnosis and treatment, and it is highly demanding to form a unified treatment standard due to the variation in treatment ideas, diagnosis, and therapeutic methods. Therefore, using machines to capture images and then analyze images with deep learning algorithms can form an objective specification for tongue-facial diagnosis, which is also one of the research hotspots in recent years. Besides, It is an essential method to improve the efficiency of TCM diagnosis and treatment, and the research on the objectification of tongue-facial diagnosis is progressing faster.¹¹⁵

Deep learning algorithms can analyze the images captured by tongue-facial diagnosis instruments and process far more images than that manual processing. The future direction of lingual-facial diagnosis objectification research is to analyze images captured by mobile using deep learning algorithms. Image processing algorithms are mainly divided into image segmentation and image classification, with FCN and CNN and their extension algorithms in dominating position. The lingual images need to go through pre-processing steps such as light compensation and color correction before processing. Since the face has strong three-dimensionality and many detailed features, and the expression changes can reflect the patient's disease condition to some extent, the processing of facial angle and detail features are very important in the image acquisition and analysis process. The processing of facial images is more likely to combine machine learning methods with other optimization algorithms, such as combining residual neural networks with feature fusion algorithms to eliminate the influence of other factors on image quality and improve image recognition accuracy. Tongue image classification is the judgment of certain feature attributes of the tongue, such as tongue color, tongue coating, etc. Tongue image segmentation segments and labels the target, both of which enable accurate processing of tongue facial diagnosis images. Traditional tongue diagnosis image processing is represented by active contour model-based algorithms, etc., and the Snake algorithm and the OTSU algorithm are applied and developed, but the segmentation accuracy and speed are not ideal. With the development of

deep learning, the application of convolutional neural networks for look-alike image segmentation became popular. The U-Net network model formed based on the development of FCN can handle a variety of medical image segmentation tasks, the segmentation accuracy of the Mask-R CNN network model is high, and many scholars have also formed algorithms with high image processing efficiency based on the improvement of traditional image processing algorithms. However, there are many issues in the process of look-alike objectification: scholars currently pour attention to the objectification of inspection diagnosis without unified standards, specifications, the current acquisition environment, light source type, and intensity lack of national or industry unified standards, and the acquired images have no universality. Traditional deep learning algorithms for image classification focus on finding similar features, and the classification effect is not good when applied to research objects with small differences between samples like tongue image classification.¹¹⁶ At present, the research on the objectification of TCM diagnosis still stays at the stage of “color” inspection, ignoring the fact that tongue and facial morphology also have diagnostic significance, and cannot fully convey the concepts of TCM diagnosis. At the same time, the current research is still at the stage of small samples, the advantages of machine learning algorithms are reflected in large data, so a standard large database for the objective study of tongue and face diagnosis is should be constructed. The formation of standards for the objectification of inspection diagnosis will facilitate the traditional medicine inspection diagnosis to play a more important role in modern medicine and provide auxiliary functions in subsequent processes. Combining machine learning algorithms with intelligent inspection diagnosis is an inevitable trend.

Acknowledgements: I would like to express my gratitude to all those who helped me during the writing of this thesis. My deepest gratitude goes first and foremost to Professor Jin Hong Guo, my supervisor, for his constant encouragement and guidance.

Contributorship:

Li Feng, Data curation, Writing—original draft, Validation.
Hai Bei Song, Funding acquisition, Writing—review & editing.

Zong Hai Huang, Data curation, Writing—original draft.
Yan Mei Zhong, Conceptualization, Validation.
WenKe Xiao, Formal Analysis, Visualization.
Chuan Biao Wen, Writing—review & editing.
Jin Hong Guo, Writing—review & editing.

Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval: Written informed consent for publication of this paper was obtained from the Chengdu University of traditional Chinese and all authors.

Funding: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Key Research and Development Program of the Ministry of Science and Technology of China, National Nature Foundation of China (grant nos. 2018YFC1707606, 82004504) and Chinese Medicine Administration of Sichuan Province (grant nos. 2021MS199).

Guarantor: Hai Bei Song.

ORCID iDs: Hai Bei Song  <https://orcid.org/0000-0001-9793-8668>

Jin Hong Guo  <https://orcid.org/0000-0003-2659-3150>

Supplemental material: Supplemental material for this article is available online.

References

- Zheng N, Lu D and Ding J. Research progress on clinical application of modern tongue diagnosis information collection in traditional Chinese medicine. *Chin J Urban Rural Enterp Hyg* 2019; 34: 65–69.
- Zhang YY, Zhou H and Zhan SH. Application and research progress of new technologies for four diagnosis of traditional Chinese medicine. *Chin Comput Med Imaging* 2021; 27: 83–86.
- Wei XY, Zhan SH, Zhou H, et al. Research progress on new technology of tongue diagnosis in traditional Chinese medicine. *J Basic Chin Med* 2021; 27: 3.
- Xu X, Song HB, Wen CB, et al. Research on the objectification of tongue diagnosis based on intelligent information processing. *Comput Knowl Technol* 2020; 16: 182–184 + 197.
- Gautam R and Sharma M. Prevalence and diagnosis of neurological disorders using different deep learning techniques: a meta-analysis. *J Med Syst* 2020; 44: 49.
- Ahsan M M and Siddique Z. Machine learning-based heart disease diagnosis: a systematic literature review. 2021. eprint arXiv:2112.06459.
- Liu D, Zhang. DL, Sun ZR, et al. Active-matrix sensing array assisted with machine-learning approach for lumbar degenerative disease diagnosis and postoperative assessment. *Adv Funct Mater* 2022; 32: 13008.
- Sharma M and Samriti, Singh G. Need and design of smart and secure energy-efficient IoT-based healthcare framework. 2019; 206: 259–281.
- Sharma M, Sharma S and Singh G. Remote monitoring of physical and mental state of 2019-nCoV victims using social internet of things, fog and soft computing techniques. *Comput Methods Programs Biomed* 2020; 196: 105609.
- Sun X, Tian CY, Zhuang SH, et al. Bibliometric research on the collection technology of tongue diagnostic instrument. *J Tianjin Univ Traditional Chin Med* 2020; 39: 186–192.

11. Song CM, Lu JK, Xue FR, et al. Application of mobile information technology in clinical research of traditional Chinese medicine. *Chin J Clin Pharmacol* 2020; 36: 3862–3864.
12. Wang WY, Wang L, Xu HJ, et al. Research on color difference correction function model of tongue diagnosis based on mobile phone terminal. *China J Traditional Chin Med Pharm* 2021; 36: 1020–1024.
13. Yan XJ, Zhou ML, Qian P, et al. Research progress on Objectification of diagnosis in contemporary clinical background. *China J Traditional Chin Med Pharm* 2021; 36: 2199–2201.
14. Yan ZQ, Wang YZ, Zhou AF, et al. Physical analysis of tongue diagnosis - development and clinical observation of tongue diagnosis instrument. *Guizhou Medical J* 1984; 1984: 1–3.
15. Wei BG, Shen LS, Wang YQ, et al. Digital TCM tongue image analyzer. *Chin J Med Instrum* 2002; 03: 164–166 + 169.
16. Ding HJ and Ding JC. Modern techniques and methods of tongue diagnosis. *Lishizhen Med Mater Med Rese* 2010; 21: 1230–1232.
17. Zhu L, Liu F and Jin F. Research progress on Objectification of facial observation. *J GuiZhou Univ Traditional Chin Med* 2020; 42: 76–81.
18. Li Q, Yang YW, Zhang D, et al. Development and thinking of face-to-face diagnosis of traditional Chinese medicine. *Modernization Traditional Chin Med Mater Med-World Sci Technol* 2021; 23: 271–275.
19. Han PP, Wang TF, Liao JY, et al. Lu, Progress of objective quantitative research on facial color inspection in TCM. *Global Traditional Chin Med* 2021; 14: 749–755.
20. Li FF, Di D, Wang YQ, et al. Research on information collection and recognition of TCM facial color diagnosis based on computer technology. *Modernization Traditional Chin Med Mater Med-World Sci and Technol* 2008; 10: 71–76.
21. Shi Q, Tang WC, Li FF, et al. Preliminary study on light source selection in the objective study of tongue image information. *Acad J Shanghai Univ Traditional Chin Med* 2004; 02: 39–41.
22. Zheng DM, Guo DJ, Dai ZD, et al. Design, implementation and experimental research of color diagnosis image acquisition system in traditional Chinese medicine. *Chin J Biomed Eng* 2011; 30: 731–737.
23. Li DX, Guan J and Li F. The development of tongue diagnostic apparatus and its application in the objective research of tongue diagnosis. *World Chin Med* 2017; 12: 456–460.
24. Wang L, Lin YF and Li L. Application progress of intelligent diagnosis and treatment in tongue image research. *China J Traditional Chin Med Pharm* 2021; 36: 342–346.
25. Di D, Zhou M, Zhou HL, et al. Development of handheld tongue imaging instrument. *Shanghai J Traditional Chin Med* 2016; 50: 12–14.
26. Jiang YC, Fan CL, Ming X, et al. Design of an integrated TCM tongue image acquisition and analysis system. *Comput Meas Control* 2018; 26: 222–225.
27. Zhang B, Bhagavatula V, Zhang D. Detecting diabetes mellitus and nonproliferative diabetic retinopathy using tongue color, texture, and geometry features. *IEEE Trans Biomed Eng* 2014; 61: 491–501.
28. Zhang MM, Zhai TT, Yang CY, et al. Research on the design of a new object acquisition device for tongue images. *Prog Biomed Eng* 2015; 36: 65–68.
29. Liu M, Xie YM, Zhao J, et al. Study on the consistency of tongue exposure degree of TCM tongue image collection instrument. *Western J Traditional Chin Med* 2019; 32: 131–133.
30. Zhang LQ, Li MH, Gao SS, et al. A review of solutions to the key problems of computer-aided tongue diagnosis. *Comput Sci* 2021; 48: 256–269.
31. Li WB and Li RH. A method for separating tongue coating and tongue quality by optimizing K-means clustering. *Hebei J Ind Sci Technol* 2020; 37: 300–308.
32. Zhang LY, Wang L, Bao X, et al. Tongue crack detection method based on local gray threshold. *Comput Knowl Technol* 2017; 13: 163–165.
33. Yang TX, Yoshimura Y, Morita A, et al. Synergistic attention U-Net for sublingual vein segmentation. *Artif Life Robot* 2019; 24.
34. Chen R, Liu L, Wang YQ, et al. Research progress of artificial neural network in TCM tongue diagnosis. *China J Traditional Chin Med Pharm* 2020; 35: 1924–1926.
35. Liu HL, Feng Y, Xu H, et al. A review of tongue segmentation based on deep learning. *J Front Comput Sci Technol* 2021; 15: 2276–2291.
36. Liu Y, Zhao PC and Lu XZ. A segmentation algorithm for face diagnosis image. *Comput Knowl Technol* 2017; 13: 183–185.
37. Lin Y, Wang B, Xu JT, et al. Research on the color classification of traditional Chinese medical observation based on facial image feature fusion. *J Clin Med Pract* 2020; 24: 1–5.
38. Ning XL and Chen ZX. Face panoramic image mosaic algorithm for face diagnosis. *Chin J Med Phys* 2020; 37: 456–462. DOI:10.3969/j.issn.1005-202X.2020. 04.011.
39. Huang AY, Cheng BY and Sun JW. Research on face recognition algorithm based on convolution internal evolution mechanism and feature fusion. *Mod Comput* 2019; 10: 75–80.
40. Jin XZ, Lin F and Wang Y. Facial expression recognition algorithm based on CNN. *J Qilu Univ Technol* 2021; 35: 64–69.
41. Wu J, Min Y, Li C, et al. A micro expression recognition algorithm based on 3D-CNN. *Telecommun Eng* 2019; 59: 1115–1120.
42. Dhomne A, Kumar R and Bhan V. Gender recognition through face using deep learning. *Procedia Comput Sci* 2018; 132: 2–10.
43. Fekri-Ershad S. Gender classification in human face images for smart phone applications based on local texture information and evaluated Kullback-Leibler divergence. *Trait Signal* 2019; 36: 507–514.
44. Zhang C, Ding H, Shang Y, et al. Gender classification based on multiscale facial fusion feature. *Math Probl Eng* 2018; 2018: 1–6.
45. Wang XS, Wang F, Wang HW, et al. Application of image processing in traditional Chinese medicine inspection. *Comput Knowl Technol* 2019; 15: 212–213 + 221.
46. Yuan SM, Qian P and Li FF. Research progress of color correction methods for tongue and face diagnosis in traditional Chinese medicine. *China J Traditional Chin Med Pharm* 2019; 34: 4183–4185.

47. Yu J, Zhao HI, Shao Y, et al. Illumination compensation for microscope images based on illumination difference estimation. *Vis Comput*, 2021; 35: 1775–1786.
48. Tang JQ, Song WF and Chen LY. Study on tongue coating image preprocessing based on multi processor environment. *J Jiujiang Univ (Nat Sci Ed)* 2020; 35: 70–73.
49. Shang WW, Wang YW, Xue SS, et al. Tongue diagnosis method based on comparative analysis of tongue chromatography. *Laser Optoelectron Prog* 2020; 57: 194–202.
50. Yang XY, Liang R, Wang ZP, et al. Research status and analysis of tongue color classification based on colorimetry. *J Beijing Univ Traditional Chin Med* 2012; 35: 539–542 + 577.
51. He YM and Du JQ. Segment of tongue and tongue's venation based on OpenCV. In: *Proceedings of 2013 3rd International Conference on Advanced Measurement and Test (AMT 2013)*. Information Engineering Research Institute, 2013; 5: 2302–2306.
52. Ghosh S, Das N and Nasipuri M. Reshaping inputs for convolutional neural network: some common and uncommon methods. *Pattern Recognit* 2019; 93: 79–94.
53. Li YT, Luo YS and Zhu ZM. Tongue image feature analysis based on deep learning. *Comput Sci* 2020; 47: 148–158.
54. Zhai PB, Yang H, Song TT, et al. A multi-stage tongue image classification algorithm based on attention mechanism. *Comput Eng Des* 2021; 42: 1606–1613.
55. Huang SQ, Wang F, Wang XS, et al. Application of convolutional neural network in tongue diagnosis of TCM. *Comput Knowl Technol* 2020; 16: 20–22.
56. Tang YP, Wang LR, He X, et al. Study on tongue image classification based on multi task convolutional neural network. *Comput Sci* 2018; 45: 255–261.
57. Xiao QX, Zhang J, Zhang H, et al. Tongue coating color classification method based on light convolution neural network. *Meas Control Technol* 2019; 38: 26–31.
58. Yang JD and Zhang P. Tongue image classification method based on transfer learning and fully connected neural network. *Acad J Second Mil Med Univ* 2018; 39: 897–902.
59. Song C, Wang B and Xu JT. Research on tongue feature classification based on deep transfer learning. *Comput Eng Sci* 2021; 43: 1488–1496.
60. Kanawong R, Obafemi-Ajayi T, Ma T, et al. Automated tongue feature extraction for Zheng classification in traditional Chinese medicine. *Evidence-Based Complementary Altern Med* 2012; 2012: 14.
61. Yan JJ, Li DX, Guo R, et al. Study on classification of tooth mark tongue based on deep learning and random forest. *Chin Arch Traditional Chin Med* 2022; 40: 19–22 + 259–261.
62. Jiao Y, Zhang XF and Zhuo L. Study on the application of weighted SVM in TCM tongue image sample classification. *Meas Control Technol* 2010; 29: 1–4 + 13.
63. Liu CX, Zhang HY and Yang H. Application of GVF Snake model based on Perona-Malik method in tongue image segmentation. *Inf Technol Network Sec* 2017; 36: 94–96 + 102.
64. Soo JM and Beom LW. Improved snakes algorithm for tongue image segmentation in oriental tongue diagnosis. *J Inst InterNet, Broadcast Commun* 2016; 16: 125–131.
65. Zhang HZ, Zuo WM, Wang KQ, et al. A snake-based approach to automated segmentation of tongue image using polar edge detector. *Int J Imaging Syst Technol* 2006; 16: 103–112.
66. Liang YT, Meng YM, Zhu LY, et al. Two dimensional Otsu fitting line threshold image segmentation method. *Sci Technol Eng* 2021; 21: 3689–3697.
67. Wu K and Zhang D. Robust tongue segmentation by fusing region-based and edge-based approaches. *Expert Syst Appl* 2015; 42: 8027–8038.
68. Ning JF, Zhang D, Wu CK, et al. Automatic tongue image segmentation based on gradient vector flow and region merging. *Neural Comput Appl* 2012; 21: 1819–1826.
69. Shi MJ, Li GZ and Li FF. C~2G~2FSnake: automatic tongue image segmentation utilizing prior knowledge. *Sci China (Inf Sci)* 2013; 56: 154–167.
70. Pang B, Zhang D, Li NM, et al. Computerized tongue diagnosis based on Bayesian networks. *IEEE Trans Biomed Eng* 2004; 51: 1803–1810.
71. Pang B, Zhang D and Wang KQ. The bi-elliptical deformable contour and its application to automated tongue segmentation in Chinese medicine. *IEEE Trans Med Imaging* 2005; 24: 946–956.
72. Ma C, Tang ZD and Tang L. Application of image segmentation technology in tongue diagnosis of TCM. *Comput Simul* 2008; 02: 215–218.
73. Zhang L and Qin J. Tongue image segmentation method based on gray projection and automatic threshold selection. *Chin J Tissue Eng Res* 2010; 14: 1638–1641.
74. Liu W, Hu J, Li Z, et al. Tongue image segmentation via thresholding and gray projection. *KSII Trans. Internet Inf Syst* 2018; 13: 945–961.
75. Chen CM, Lu HH and Lin YC. An early vision-based snake model for ultrasound image segmentation. *Ultrasound Med Biol* 2000; 26: 273–285.
76. Zheng F, Huang XY, Wang BL, et al. Tongue image detection method based on image segmentation. *J Xiamen Univ (Nat Sci)* 2016; 55: 895–900.
77. Jiang S, Hu J, Xia CM, et al. Tongue image segmentation method based on Otsu threshold method and morphological adaptive modification. *Chin High Technol Lett* 2017; 27: 150–155.
78. Huang ZP, Huang YS, Yi FL, et al. Automatic tongue segmentation based on maximum inter class variance method and region growth. *Lishizhen Med Mater Med Res* 2017; 28: 3062–3064.
79. Zhang ZS and Liu Y. Tongue extraction algorithm based on dynamic threshold and modified model. *Comput Modernization* 2014; 11: 49–52.
80. Yu ZH, Zhang ZC, Li ZY, et al. Research on tongue image threshold segmentation algorithm based on multi color component fusion. *Comput Appl Softw* 2019; 36: 199–203 + 248.
81. Guo Z, Yang XZ, Si YC, et al. A tongue coating segmentation algorithm based on K-means clustering in CIELAB and HSI color space. *China J Traditional Chin Med Pharm* 2010; 25: 663–665.
82. Ma LX, Yang H, Song TT, et al. Research on tongue image segmentation algorithm based on high resolution features. *Comput Eng* 2020; 46: 248–252.
83. Li L, Luo ZM, Zhang MT, et al. An iterative transfer learning framework for cross-domain tongue segmentation. *Concurrency Comput: Pract Exp* 2020; 32: 5.

84. Zhou JH, Zhang Q, Zhang B, et al. TongueNet: A precise and fast tongue segmentation system using U-Net with a morphological processing layer. *Appl Sci* 2019; 9: 3128.
85. Kolhar S and Jagtap J. Convolutional neural network based encoder-decoder architectures for semantic segmentation of plants. *Ecol Inform* 2021; 64: 101373.
86. Lai C, Wang H, Wang F, et al. Autosegmentation of Prostate Zones and Cancer Regions from Biparametric Magnetic Resonance Images by Using Deep-Learning-Based Neural Networks. *Sensors (Basel, Switzerland)* 2021; 21: 8.
87. Shao K, Zhang YF, Bao FX, et al. A method of coronary artery segmentation based on mask RCNN and fused geometric features. *J Taiyuan Univ Technol* 2021; 52: 83–90.
88. Lee S, Negishi M, Urakubo H, et al. Mu-Net: Multi-scale U-Net for two-photon microscopy image denoising and restoration. *Neural Netw* 2020; 125: 92–103.
89. Li RR, Peng XT, Xiao GX, et al. Research on retinal vascular segmentation based on vascular connectivity. *China Digital Med* 2020; 15: 125–129.
90. Liu Z, Song YQ, Sheng VS, et al. Liver CT sequence segmentation based with improved U-Net and graph cut. *Expert Syst Appl* 2019; 126: 54–63.
91. Xu HW, Yan PX, Wu M, et al. Automatic segmentation of cyst kidney in CT images based on residual double attention u-net model. *Appl Res Comput* 2020; 37: 2237–2240.
92. Li C, Tan YS, Chen W, et al. ANU-Net: Attention-based nested U-Net to exploit full resolution features for medical image segmentation. *Comput Graph* 2020; 90: 11–20.
93. Wang Q, Qiang Y, Yang XT, et al. Lung nodule segmentation network model based on dual attention 3D-U-Net. *Comput Eng* 2021; 47: 307–313.
94. Akira H, Yongbum L, Yu T, et al. Automated classification of calcification and stent on computed tomography coronary angiography using deep learning. *Nihon Hoshasen Gijyutsu Gakkai zasshi* 2018; 74: 1138–1143.
95. Chen J., Wan Z.C., Zhang J.C., W.H., et al. Medical image segmentation and reconstruction of prostate tumor based on 3D AlexNet. *Comput Methods Programs Biomed*, 2020; 200: 105878.
96. Han SS, Kim MS, Lim W, et al. Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *J Invest Dermatol* 2018; 138: 1529–1538.
97. Lei Y, He XX, Yao JC, et al. Breast tumor segmentation in 3D automatic breast ultrasound using Mask scoring R-CNN. *Med Phys* 2021; 48: 204–214.
98. Ren SQ, He KM, Ross G, et al. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell* 2017; 39: 1137–1149.
99. Ma SL, Huang Y, Che XJ, et al. Faster RCNN-based detection of cervical spinal cord injury and disc degeneration. *J Appl Clin Med Phys* 2020; 21: 235–243.
100. Nazir T, Irtaza A, Starovoitov V, et al. Optic disc and optic cup segmentation for glaucoma detection from blur retinal images using improved mask-RCNN. *Int J Opt* 2021; 2021: 12.
101. Sara M, Chiara FM and Emanuele F. Mask-R-CNN: a distance-field regression version of Mask-RCNN for fetal-head delineation in ultrasound images. *Int J Comput Assist Radiol Surg* 2021; 16: 1711–1718.
102. Yan JJ, Guo Z, Guo R, et al. Tongue image segmentation based on Mask R-CNN. *Modernization Traditional Chin Med Mater Med-World Sci Technol* 2020; 22: 1532–1538.
103. Zhang XF, Guo YT, Cai YH, et al. Tongue image segmentation algorithm based on DCNN and fully connected CRF. *J Beijing Univ Aeronaut Astronaut* 2019; 45: 2364–2374.
104. Zhou C, Fan H and Li Z. SAR Target recognition via joint sparse representation of monogenic components with 2D canonical correlation analysis. *IEEE Access* 2019; 7: 1–1.
105. Wang J. and Liu X.P. Medical image recognition and segmentation of pathological slices of gastric cancer based on Deeplab v3+ neural network. *Comput Methods Programs Biomed*, 2021; 207: 106210–106210.
106. Chen YS, Chang YH and Lin J. Comparing intuitionistic fuzzy set theory method and canny algorithm for edge detection to tongue diagnosis in traditional Chinese medicine. *Adv Mater Res* 2013; 78: 3771–3774.
107. Li HH, Wen GH and Zeng HB. Natural tongue physique identification using hybrid deep learning methods. *Multimedia Tools And Applications* 2019; 78: 6847–6868.
108. Li QL, Wang YT, Liu HY, et al. Tongue fissure extraction and classification using hyperspectral imaging technology. *Appl Opt* 2010; 49: 2006–2013.
109. Shi MJ, Li GZ, Li FF, et al. Computerized tongue image segmentation via the double geo-vector flow. *Chin Med* 2014; 9: 7.
110. Cui ZC, Zhang HZ, Zhang D, et al. Fast marching over the 2D Gabor magnitude domain for tongue body segmentation. *Inventi Impact Signal Process* 2013; 2013: 1–13.
111. Zhou CG, Fan HY, Zhao W, et al. Reconstruction enhanced probabilistic model for semisupervised tongue image segmentation. *Concurrency Comput: Pract Exp* 2020; 32: e5844.
112. Huang X, Zhang H and Zhuo L. TISNet-enhanced fully convolutional network with encoder-decoder structure for tongue image segmentation in traditional Chinese medicine. *Comput Math Methods Med* 2020; 2020: 1–13.
113. Gao S, Guo N and Mao D. LSM-SEC: Tongue segmentation by the level set model with symmetry and edge constraints. *Comput Intell Neurosci* 2021; 2021: 14.
114. Lin C, Lee F, Cai J, et al. Global and graph encoded local discriminative region representation for scene recognition. *Comput Model Eng Sci* 2021; 128: 985–1006.
115. Wen H, Huang L, Liu J, et al. Research Progress on application of machine learning technology in clinical diagnosis and treatment of traditional Chinese medicine. *China Medical Herald* 2021; 18: 42–45.
116. Sun M and Zhang XF. Study on tongue image classification based on TripletLoss loss function. *Beijing Biomed Eng* 2020; 39: 131–137.