

# Artificial Intelligence in Tongue Image Recognition

Hongli Chu, South China Normal University, China

Yanhong Ji, South China Normal University, China

Dingju Zhu, South China Normal University, China\*

Zhanhao Ye, South China Normal University, China

Jianbin Tan, South China Normal University, China

Xianping Hou, Sun Yat Sen Memorial Hospital, Sun Yat Sen University, China

Yujie Lin, Sun Yat Sen Memorial Hospital, Sun Yat Sen University, China

## ABSTRACT

Tongue image recognition is a traditional Chinese medicine diagnosis method, which uses the shape, color, and texture of the tongue to judge the health of the human body. With the rapid development of artificial intelligence technology, the application of artificial intelligence in the field of tongue recognition has been widely considered. Based on the intelligent analysis of tongue diagnosis in traditional Chinese medicine, this paper reviews the application progress of artificial intelligence in tongue image recognition in recent years and analyzes its potential and challenges in this field. Firstly, this paper introduces three steps of tongue image recognition, including tongue image acquisition, tongue image preprocessing, and tongue image feature analysis. The application of traditional methods and artificial intelligence methods in the whole process of tongue image recognition is reviewed, especially the tongue body segmentation, and the advantages and disadvantages of convolutional neural networks are analyzed and compared. Artificial intelligence can use technologies such as deep learning and computer vision to automatically analyze and extract features from tongue images. By constructing a tongue image recognition model, tongue shape, color, texture, and other features can be accurately recognized and quantitatively analyzed. Finally, this paper summarizes the problems existing in artificial intelligence in tongue image recognition and looks forward to the future developmental direction of this field. It can promote the modernization of TCM diagnostic methods, achieve early disease screening and prevention, personalized medicine and treatment optimization, and support medical research and knowledge accumulation. However, there is still a need for further validation and practice, with a focus on patient privacy and data security.

## KEYWORDS

AI, Deep Learning, Image Segmentation, Machine Learning, Tongue Diagnosis

DOI: 10.4018/IJSSCI.328771

\*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

## 1. INTRODUCTION

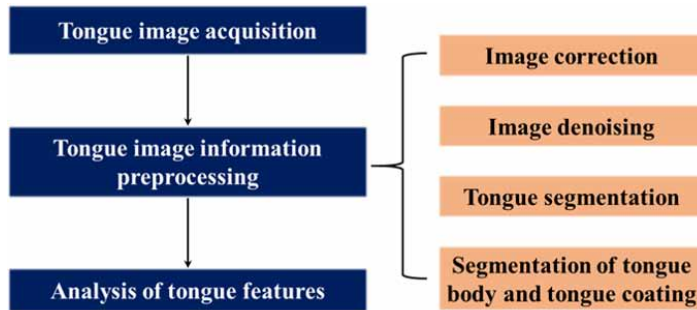
In the thousands of years of Chinese history, traditional Chinese medicine has formed its systematic branch of culture. Traditional Chinese medicine believes that the changes in tongue color, coating, and quality can reflect the rise and fall of Qi and blood, the deficiency and reality of viscera, the nature of the disease, and the depth of disease. Tongue diagnosis, as an important dialectical diagnosis and treatment method of traditional Chinese medicine, has been well inherited in China. It is convenient, fast, and low cost, and does not require expensive medical equipment to assist in diagnosis. It is not affected by personal emotions. In the past, the doctor macroscopic observation and the analysis of individual experiences considered patients with illness, but its shortcomings are obvious. Every doctor among individuals vary in medical level and clinical experience. The lack of quantification of tongue diagnosis standard is subject to the influence of environmental factors, such as different light conditions, different viewing angle, etc. To the naked eye, disease recognition accuracy is not high. It also increases the fatigue of individual doctors' vision and other problems, which very easily causes the wrong diagnosis of patients. The consequences of the wrong diagnosis will undoubtedly lead to loss of life and property.

Artificial intelligence (AI) (Xu, 2013) is a discipline that records, accumulates, reproduces, and uses knowledge by simulating human beings. The application of artificial intelligence in tongue diagnosis can overcome the limitation of doctors' observation of tongue images with the naked eye, and realize the precision and digitalization of traditional Chinese medicine (TCM) diagnosis. AI has continuously improved the accuracy of tongue image analysis instruments, and the changes in tongue image are related to diseases. Some tongue images are difficult to distinguish by the naked eye. The comprehensive use of an artificial intelligence system combined with TCM tongue diagnosis can also analyze complex and hidden diseases, reducing the rate of misdiagnosis and miss diagnosis. Therefore, it will be an important modernization development of TCM diagnosis and treatment to make good use of artificial intelligence and establish an intelligent diagnosis and decision support system with TCM characteristics, and to provide a powerful objective basis for the diagnosis, treatment, and prognosis of diseases (Bhatnagar & Bansod, 2022; Wan & Chin, 2021).

Tongue diagnosis based on a machine learning algorithm has been applied to many diseases, such as diabetes (Fan et al., 2021), stomach trouble (Yuan et al., 2023), appendicitis (Pang et al., 2005), etc. (Chung et al., 2022; Lee et al., 2016a; Lo et al., 2015; Park et al., 2022; Wu et al., 2018). Jiang et al. (2012) demonstrated that visual signatures of the microbiome in tongue coatings reflect health status, specifically GI disorders. Han et al. (2016) suggested that tongue diagnosis based on images analyzing tongue features, tongue color, and tongue coating could provide a potential screening and early diagnosis method for cancer. Lo et al. (2013) reported significant differences between breast cancer patients and healthy subjects in tongue characteristics such as the amount of tongue hair in the spleen and stomach region, the largest area covered by tongue coating, thin tongue coating, number of tooth marks, red points, and red points in the spleen and stomach region. In addition, associations of tongue color, tongue coating, and sublingual vessels with risk factors and clinical characteristics in patients with ischemic stroke have been reported (He et al., 2016).

Tongue image processing is a key step in digital tongue diagnosis. It includes image correction, image noise reduction, tongue body segmentation, and tongue coating segmentation (Li et al., 2022). The intellectualized process analysis diagram of tongue diagnosis is shown in Figure 1. Firstly, to determine the acquisition standard in the standard acquisition environment, the use of a digital camera and other shooting equipment to collect tongue images will be transferred to the computer. Secondly, the tongue is labeled. Then the tongue is segmented from the image and the segmented tongue is separated from the moss. Finally, the separated tongue body and tongue coating are classified respectively and an intelligent tongue diagnosis system is formed. To meet the high computational and low latency requirements of edge computing for remote smart tongue diagnostic modeling, Zhang et al. (2021) introduced a similar data transfer strategy to effectively transfer the necessary knowledge

Figure 1. Intelligent analysis model of tongue diagnosis in traditional Chinese medicine



and overcome the shortage of clinical tongue images. The network is then pruned while preserving mobility in domain adaptation to generate a simplified structure. Finally, a compact model combining two sparse networks is designed to match limited edge devices.

Tongue image acquisition is mainly affected by the acquisition environment and acquisition equipment. Among them, light condition is an important factor along with the existing tongue image acquisition analyzer, such as Four Diagnostic Instruments of Traditional Chinese Medicine (DS01-series) (Duan et al., 2018), TDA-1 tongue imager (Zhang et al., 2017), etc. These devices are large in size, high in cost, not convenient, and not suitable for complex acquisition environments, which restricts the development of remote diagnosis and treatment cloud services. Therefore, tongue image acquisition in the future should be developed in the direction of convenience, speed, and efficiency.

Tongue image processing is the key step in digital tongue diagnosis, which includes image correction, image noise reduction, tongue body segmentation, tongue coating segmentation, and other preprocessing (Mukai et al., 2022). Artificial intelligence plays a major role in this step. Tongue body segmentation is essentially image segmentation. Many traditional image segmentation methods have been applied to tongue image segmentation, such as the threshold method and region-based segmentation method. With the continuous application of deep learning in medicine, a variety of new segmentation methods and representative semantic segmentation networks have been developed, such as convolutional neural network (CNN) (Krizhevsky et al., 2017), fully convolutional neural network (FCN) (Long et al., 2015), VGGNet (Simonyan & Zisserman, 2014), GoogleNet (Szegedy et al., 2015), ResNet (He et al., 2016), DenseNet (Huang et al., 2017), U-Net (Ronneberger et al., 2015), etc. The above-segmented networks can also be combined by optimization. For example, Zhuang et al. (2022), based on the artificial intelligence framework of deep learning convolutional neural network, applied ResNet34 to the dataset to automatically extract image features and realize tongue image classification. In addition, they also applied the VGG16 network framework to the dataset to compare the classification model and classification effect. The method they used improved the accuracy of the tooth mark recognition model by more than 10%. In the latest report, Li et al. (2022) constructed an “end-to-end” deep learning network for intelligent analysis of tongue images based on TCM tongue diagnosis images. The tongue target region in the original image is segmented by the U-Net tongue segmentation model at the front of the network. After segmentation, the feature vectors of the tongue target region were extracted by the ResNet network, and then the blood pressure on the day of shooting was fused with the feature vectors extracted by the ResNet network by the method of convolution operation to complete the data sets of two tongue features and fused features. U-Net tongue segmentation model combined with ResNet network can realize automatic extraction of tongue image features. The extracted features combined with machine learning modeling can be used to explore the complex hierarchical mathematical associations between tongue images and clinical data. The experimental results show that the multi-modal data fusion method is an important way to explore the clinical value of tongue images in TCM. In the latest study by Zhang et al. (2022), U-Net

was used as the model skeleton, while ResNet18 introduced residual network as the feature extraction layer of the U-Net encoding path to improve the sensitivity of feature mapping to output changes and improve tongue segmentation accuracy. A tongue image segmentation model UrNet (U-Net-Resnet18), including encoding and decoding, was constructed to achieve rough segmentation of the tongue. On this basis, superpixel features are added to optimize the process of coarse segmentation results. Subsequent experiments confirm that SpurNet (UrNet+ superpixel) (Figure 2) can effectively solve the problems of poor margin segmentation, over-segmentation, and under-segmentation of background.

This paper will take the intelligent analysis steps of TCM tongue diagnosis as the clue, analyze the research status of the steps one by one, and the research progress of the design to the application of artificial intelligence. Its structure is as follows: The first part introduces the research progress of tongue image acquisition in official selection and equipment. The second part focuses on analyzing the theory and research progress of tongue image processing and comparing the advantages and disadvantages of various methods. The third part mainly explains the methods of different tongue characteristics, and the characteristics of tongue diagnosis combined with traditional Chinese medicine are explained. Finally, we summarize the success and problems of artificial intelligence in tongue image recognition and make a prospect about the future.

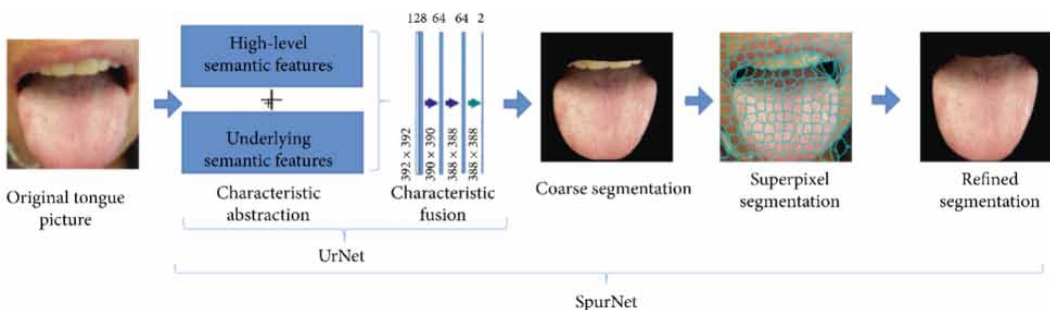
## 2. TONGUE IMAGE ACQUISITION

In the process of tongue image acquisition, the following is considered: the tongue extension posture, the internal factors of different subjects' diet and exercise, as well as the external factors of different acquisition equipment (digital camera, lens), and various light sources. How to effectively avoid the interference of the above factors to capture clear and real tongue images is the focus of this kind of research.

### 2.1 Choices of the Light Source

Tongue image acquisition involves the design of light source and optical path, the selection of color card and algorithm for color correction, and the verification of tongue image color consistency. Tongue image acquisition devices in the 1990s were mainly used for tongue image color measurement, which involved using physical methods, or the use of some material characteristics, tongue image color measurement, such as tongue plate color measurement method, fluorescence spectrum analysis color measurement method, spectrophotometric color measurement method, spectral spectrophotometry. Standardized tongue image acquisition is more conducive to reducing unnecessary interference in subsequent processing and improving the accuracy of tongue image recognition. The acquisition light source can be divided into two kinds. One is natural light, which has the advantages of a continuous spectrum, uniform illumination, simplicity, and practicality. In this case, the color temperature and brightness are greatly influenced by the environment, which brings uncertain factors to tongue image

Figure 2. SpurNet framework (H. Zhang et al., 2022)



acquisition. The second is an artificial light source. The artificial light source is less affected by the external environment, but it also has some shortcomings, such as single spectral discontinuity, uneven illumination, and complicated equipment. For instance, Yamamoto et al. (2010) proposed a hyperspectral imaging system consisting of an integrating sphere, an artificial solar lamp, and a hyperspectral camera for regional image analysis of tongue color spectra. The system automatically examines the uncoated tongue and quantifies the spectral factors of the uncoated tongue, coated tongue, lips, and perioral region. In the study by Qi et al. (2016), a LED illuminator at 6447 K with a color rendering index CIE  $R_a$  of 98 and 2413 lx was placed in the middle of a telescopic cylinder to obtain the color features of the tongue images.

## 2.2 Development of Acquisition Equipment

As the media of tongue image acquisition, tongue imager needs to collect the real information of tongue diagnosis objectively. At present, the tongue imager mainly combines the RGB color model with digital technology to realize the quantitative analysis and recognition of the color and shape of tongue and tongue coating. Representative products include the DS01 series Daosheng Four diagnostic instrument jointly developed by Shanghai Daosheng Tang and Shanghai University of Traditional Chinese Medicine, TDA-1 tongue image instrument jointly developed by Shanghai University of Traditional Chinese Medicine, and YM-iii series tongue image data measuring instrument jointly developed by Tianjin Medical University and Tianjin University. The characteristics of several tongue image instruments are compared in Table 1.

With the development of computer technology, tongue image acquisition devices have been further studied, the functions of tongue image processing and analysis have been increased, the acquisition and storage of tongue image have been realized, and the color correction and subsequent processing and analysis can be carried out. With the development of mobile devices, tongue image acquisition devices are also developing in the direction of miniaturization, portability, and intelligence.

## 3. TONGUE IMAGE PROCESSING

Tongue diagnosis in traditional Chinese medicine is combined with artificial intelligence. With artificial intelligence as the medium, tongue images with real information are collected by equipment and processed. To make the subsequent tongue image recognition more complex and practical, the collected images after tongue image processing should be able to objectively reflect the real

Table 1. Several types of tongue image instruments and their characteristics

| Types                                    |               | Characteristics                                                                                                                                                                                                                                                        |
|------------------------------------------|---------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>DS01 Series</b><br>(Luo et al., 2020) | <b>DS01-A</b> | Mainly evaluates the changes of focal inflammatory response in the stomach indirectly by objectively detecting parameters of tongue image.                                                                                                                             |
|                                          | <b>DS01-B</b> | Performs computer-assisted tongue diagnosis to analyze characteristic parameters and preliminarily screen high-risk groups.                                                                                                                                            |
|                                          | <b>DS01-C</b> | Paid attention to 12 characteristics of the tongue – such as tooth marks, red tongue edge and tip, and tongue coating color – and combined with the research results to determine the characteristics of the lesions and severity of the disease.                      |
| <b>TDA1</b> (J. Huang et al., 2018)      |               | Mainly used to quantify the characteristics of tongue quality, tongue coating color, shape, and texture, and to explore the parameters of tongue diagnosis in specific diseases.                                                                                       |
| <b>YM-III</b> (B. Wang et al., 2019)     |               | Focuses on tongue color, moss color, moss quality, tongue shape, and other parameters, which are quantified according to the scoring system to assist the diagnosis. This is mainly used for coronary heart disease and hyperuricemia.                                 |
| <b>XM-SX-III</b>                         |               | A combination of data mining technology and machine vision technology. It is suitable for the auxiliary analysis of the doctor's tongue diagnosis and has a wide range of applications. This is one of the more advanced tongue instrument representatives at present. |

information of the real tongue, such as color, shape, state, tooth marks, cracks, and so on, to make a correct and reasonable diagnosis of patients combined with the traditional Chinese medicine tongue diagnosis. Therefore, tongue image processing is the most critical step in the intellectualization of tongue diagnosis in traditional Chinese medicine, and it is also the key content discussed in this paper.

### 3.1 Image Correction

When using a digital camera to collect tongue information, will not only be affected by the external acquisition environment, but its image generated by the RGB color space information also will be affected by the camera device inside the sensor type and chip rendering algorithm. The influence of these collections will lead to the tongue pictures on some color distortion, and negative influence on the back of the research. To eliminate these effects, researchers designed tongue image color correction algorithms to correct the collected color information. Zhuo et al. (2014) proposed a tongue image color correction algorithm based on simulated annealing (SA) -Genetic algorithm (GA) -back propagation (BP) neural network.

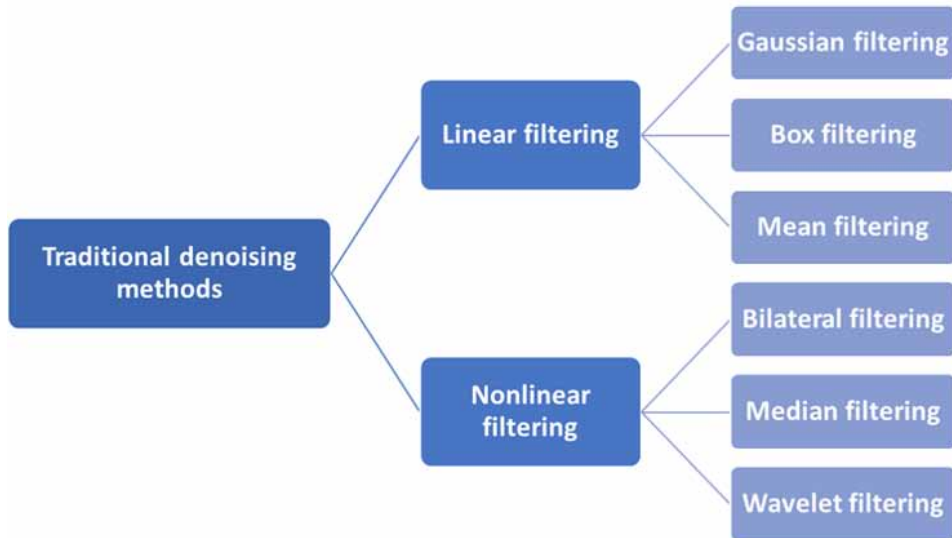
The methods used in the objective acquisition of clinical tongue images have evolved from digital cameras to self-developed devices. In the early stage, most of the tongue image information was captured by high imaging digital cameras. However, the disadvantage is that the acquisition environment has high requirements, such as acquisition angle and lighting conditions, which makes the collected images lack unified standards (Han et al., 2018). Therefore, researchers acquire tongue images through self-developed acquisition equipment. Liu et al. used a hyperspectral tongue imager instead of a digital camera to capture tongue images (Li et al., 2010; Li & Liu, 2009); Liu et al. (2011) also proposed a hyperspectral imaging system, which uses spectral acousto-optic tunable filters and spectral adapters to collect reflectance data and measure and analyze the reflection spectrum of the human tongue with a high spatial resolution for the detection of tongue tumors. Lu et al. (2020) proposed a two-stage deep color correction network method. In the first stage, a deep color correction network was designed to simulate the mapping between images captured under standard lighting conditions and objective chromaticity values of the target. The second stage provides flexible color adjustment options to accommodate different work environments and physicians' subjective preferences for visually perceived color appearance. The results obtained in both stages have satisfactory perceptual adaptation.

### 3.2 Image Noise Reduction

At present, more and more medical universities and pharmaceutical companies begin to explore the development route of TCM tongue diagnosis combined with computer science and technology, and a series of scientific research achievements have been achieved (Guangyu et al., 2021; Li et al., 2017a; Selvaraju et al., 2017; Tang et al., 2018; Wang & Zhao, 2000). Medical image is inevitably affected by noise in the process of storage or signal transmission, resulting in the loss of image details, which then affects the subsequent image processing effect and clinical application. To eliminate noise in images, researchers have proposed many denoising methods, such as linear algorithms such as Gaussian, Wiener filter, and mean filter, and nonlinear algorithms such as bilateral filter, median filter, and wavelet filter. In the process of image storage or transmission, due to the influence of noise, the original details of the image will be lost, affecting the subsequent analysis of image features, so it is necessary to carry out denoising processing to obtain a clear image with obvious features (Thakur et al., 2021).

Some traditional denoising methods are shown in Figure 3. Linear filtering can suppress Gaussian noise well, but it has inherent defects, that is, it cannot protect image details well, blurring the edges and structure of the image, and destroying the original image information. Nonlinear filtering is better than linear filtering in the treatment of salt and pepper noise, but it is not good in the treatment of Gaussian noise. Therefore, linear filtering or nonlinear filtering cannot be used to denoise images. To this end, many researchers have proposed improved methods.

Figure 3. Traditional denoising methods



### 3.3 Tongue Body Segmentation

The tongue image collected by the device includes not only the tongue body but also some non-tongue parts such as teeth, lips, and even the face, which will cause interference to the tongue image analysis by the computer. For better and more accurate tongue image analysis, an accurate and efficient segmentation method is needed to segment the tongue body. In other words, tongue segmentation refers to the extraction of the tongue part from the collected image. Its purpose is to simplify or change the representation form of the tongue image and make it easier to analyze the tongue image. The essence of tongue image segmentation is an image segmentation task, which assigns a specific label to each pixel in the image. The pixel set with the same label in the image has the same visual features, such as brightness, texture, and color.

The tongue body is similar to the background color, which affects the later computer judgment, and the workload of manual matting is huge, so the research of tongue image segmentation technology is also the key technology of tongue image information preprocessing. Traditional tongue segmentation methods include region-based segmentation, threshold-based segmentation, and so on. With the development and utilization of deep learning, a variety of image segmentation methods have been applied to tongue segmentation, such as convolutional neural networks, fully convolutional neural network, VGGNet, GoogleNet, etc. The use of a semantic segmentation network in tongue image segmentation greatly promotes the accuracy of tongue image recognition. The tongue image segmentation method based on deep learning has more accurate results compared with traditional methods. However, the generalization performance of the above methods is weak. When the test data and training data come from different data distributions, additional data annotation is usually required for the new data to reduce the loss of performance.

#### 3.3.1 Traditional Segmentation Method

##### 3.3.1.1 Region-Based Segmentation Method

The region-based segmentation method is to segment along the boundary of different regions, mainly including watershed and boundary detection. The segmentation method based on edge detection is

also easily disturbed by noise. Because of this, it is difficult to accurately segment the image with complex boundaries, yet it is easy to divide different small areas into the same parts.

The active contour model (ACM) (Zhang et al., 2006), also known as the Snakes model, is the most widely used method in tongue image segmentation. The algorithm initializes a contour and puts forward an energy function that expresses both internal energy and external energy to constrain the contour. When the energy function reaches the minimum value, the position of the curve is the contour of the target. The active contour model is based on the initial target contour and can optimize the contour position by calculating the energy function. However, this kind of method is sensitive to the initial contour, and the energy function is a non-convex function, which is easy to achieve the local optimum, and it cannot fit the very depressed part of the contour.

The image is divided into different parts according to the similarity criteria (such as color, texture, etc.). According to the starting point and growing direction of the algorithm, it can be divided into two methods: region-based growing and region-based splitting and merging. This method is relatively simple, but it is sensitive to noise and the segmentation speed is slow when a target region is too large. Wu et al. (Li et al., 2009) reported that the tongue image was segmented by brightness enhancement processing, which determined the difference between the tongue area and the surrounding area. The difference was combined with the image roughness to realize the tongue image segmentation.

There is also a tongue segmentation method based on an adaptive threshold in region recognition. This method divides the tongue image data into several image subblocks and obtains the optimal threshold of each word block by iterative calculation. Then, it forms a matrix according to the optimal threshold of each subblock, which is used to segment the tongue image (Li & Wei, 2011). The region recognition method uses the common characteristics of the pixel data in the tongue image and combines with some machine learning algorithms to realize the segmentation of the tongue body, which provides support for the early research on the objectification of tongue diagnosis. However, due to its poor segmentation accuracy and automation, the above methods have rarely been used in the related work of tongue diagnosis objectification.

The variable image segmentation model of two-dimensional images was first proposed by Kass et al. (1988), whose principle is to use energy pan-function minimization to gradually approach the true boundary of the target image, which is called the Snake model according to its dynamic characteristics. Guo et al. (2017) used the K-means clustering algorithm to set the initial edge in the Snake model, thereby realizing the automatic recognition of the tongue target region based on Snake. Compared with the above traditional image recognition methods, the Snake model has improved in both accuracy and stability and is still one of the important tongue segmentation methods in the objectification research of tongue diagnosis (Zhang et al., 2006). The shortcoming of this model is that the initial edge needs to be set in advance or automatically combined with other algorithms, which leads to the decrease of the automation degree and the segmentation speed of the model to a certain extent. It is sensitive to a local minimum and easy to fall into local extremum. During the minimization of contour energy, small features are easily ignored, and the accuracy depends on the convergence strategy.

In the process of segmentation, traditional image recognition methods are easily affected by environmental light factors, which leads to the deviation of the tongue image feature analysis from the actual situation. In addition, the tongue color is often too similar to the facial tissue color, resulting in insufficient segmentation boundary accuracy.

### *3.3.1.2 Thresholding-Based Segmentation Method*

The threshold method is the most basic method of tongue image segmentation. According to the difference between tongue color and background color, the appropriate threshold is selected to extract the tongue region from the image. The key to threshold method is the selection of threshold, simple calculation, high operation efficiency, and can quickly get the segmentation result of the tongue body. However, this kind of method is sensitive to image noise and is affected by too many external factors,



so it is difficult to select an appropriate threshold (Jian-Qiang et al., 2008; Zhang & Qin, 2010). It is difficult to effectively segment the tongue body from the original image using the threshold method because the color of the region outside the tongue body in the tongue image is very close to the tongue color and the threshold method cannot distinguish the foreground and background of the tongue body.

Zhao et al. (1999) introduced the HIS model into mathematical morphology and transformed the color RGB tongue image into HIS space. Since the H component could better present bimodal property, the H component was binarized according to the threshold value of the gray histogram. Then, the maximum target area is preserved by clustering in the target area. Finally, the edges of the tongue body are detected according to the corrosion or growth operation in mathematical morphology, and the regions within the edges are filled to obtain the mentioned tongue body.

### 3.3.2 Deep Learning

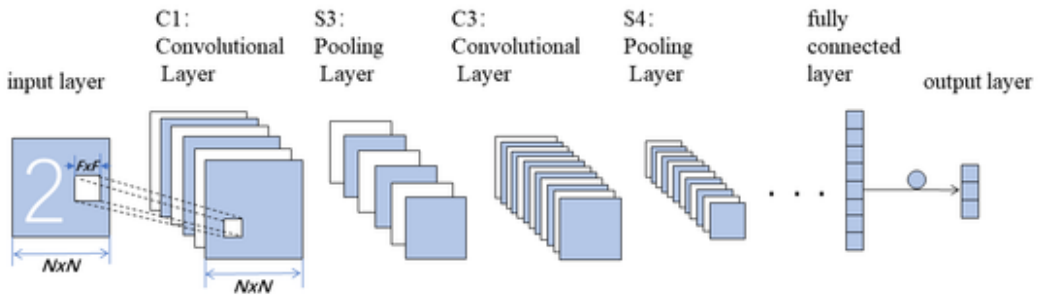
Deep learning is a branch of machine learning that emphasizes the use of multiple levels of data abstraction. Deep learning is not a new technology; its concept originated from artificial neural networks. In essence, it refers to an effective training method for neural networks with deep structure. Deep learning combines low-level features to form more abstract high-level representation attribute categories or features to find distributed feature representations of data (Stoecklein et al., 2017). The motivation for deep learning is to build a neural network that simulates the human brain for analytical learning. It mimics the way the human brain interprets data, such as images, sounds, and text. In addition, it can automatically abstract and extract low, medium, and high-level features directly from the original tongue image for end-to-end composition (Kajikawa et al., 2019; Stoecklein et al., 2017; Yoo & Oh, 2018). Combining traditional Chinese medicine theory with deep learning technology and constructing a neural network model to analyze tongue images not only provides a new idea for computer-aided tongue diagnosis but also improves the modernization and automation level of disease diagnosis. Semantic segmentation is a kind of image recognition task. By establishing category labels with real meanings, pixels are accurately classified into the labels to which their meanings belong, allowing them to realize image recognition (Wang et al., 2020).

In the early stage, the task of semantic segmentation mainly relies on machine learning algorithms, among which support vector machine (Yang et al., 2012), conditional random field (Kumar & Hebert, 2003), Random Forest (Zhang, 2017), and other models are widely used. After the emergence of deep learning technology, FCNN (Long et al., 2015), UNet (Ronneberger et al., 2015), ResNet (He et al., 2016), and other algorithm models have been widely used for image recognition applications in a variety of complex environments. In recent years, with the development of computer hardware technology, model architectures based on deep learning are constantly updated. The emergence of new network architectures, such as Deeplab V3 (Chen et al., 2017), makes deep learning more prominent in semantic segmentation and expands the application scope of deep learning. In addition, there are many new network architectures, but due to the lack of space, we will focus on the common neural network architectures.

#### 3.3.2.1 CNN

A convolutional neural network is one of the most representative models in deep learning, and its structure can be regarded as a branch of an artificial neural network. Based on the artificial neural network, the basic architecture adds two parts: the convolutional layer and the pooling layer. A convolutional neural network is widely used in the field of machine vision, and its performance is outstanding in the learning of image, video, audio, and other data forms. The basic principle involves using matrix convolution operation to filter the image data and reduce the dimensions to extract data features, and then use neural networks for data classification and recognition. The main architecture of the convolutional neural network is composed of five parts, which are successively divided into the input layer, convolutional layer, pooling layer, fully connected layer, and output layer (see Figure 4).

Figure 4. Convolutional neural network structure diagram (Gu et al., 2018)



Convolutional neural networks are widely used in image processing problems, such as image classification and object segmentation, and have recently also been applied to natural language processing and speech processing. It can effectively reduce the dimensionality of large data images into small data images. At that time, the bottleneck of image processing technology was that the amount of data needing to be processed was too large, which not only required a lot of cost but also had low efficiency. Can effectively retain the image features, in line with the principle of image processing, so that the original image features can be retained in the process of image digitalization and improve the accuracy of image processing.

### 3.3.2.2 VGGNet

Proposed by the Visual Geometry Group at Oxford University, VGGNet (Simonyan & Zisserman, 2014) improves AlexNet by replacing large, kernel-size filters with multiple 3\*3 kernel-size filters (Hu et al., 2021). For a given receptive field, multiple nonlinear layers increase the depth of the network, so multiple stacked kernels with smaller sizes are better than kernels with larger sizes. More complex features can be learned at a lower cost (Simonyan & Zisserman, 2014).

### 3.3.2.3 FCN

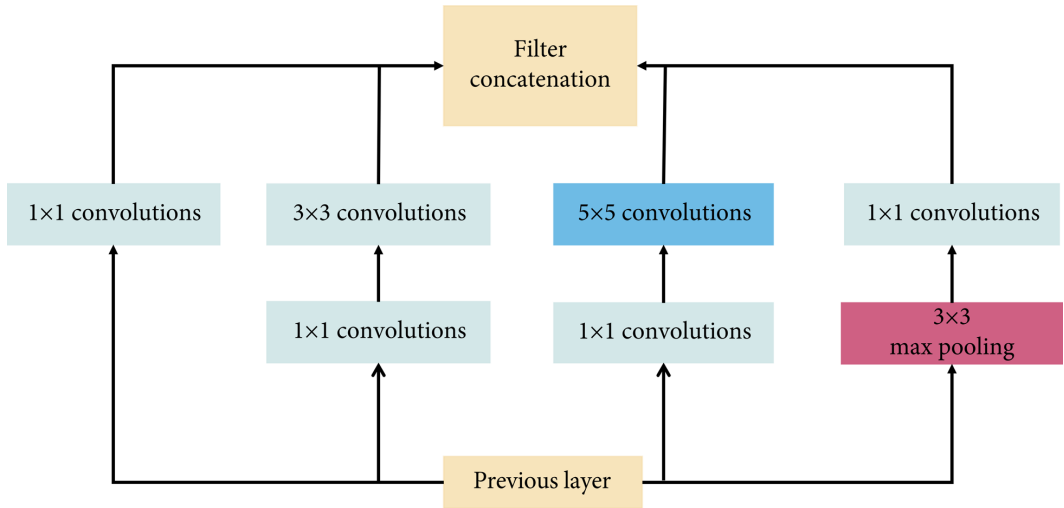
FCN (Long et al., 2015) is another milestone for CNN in the field of image classification. By marking each pixel of the target object with the correct semantic type, the target object can be accurately segmented. FCN has become a framework model for image semantic segmentation. Devconv Net and Seg Net are developed based on this framework. The difference between FCN and CNN is that FCN replaces the last series of convolutional layers with fully connected layers, but the size of the convolution kernel is  $N * 1 * 1$ . In this way, compared with CNN, only the structure of network parameters is changed, but the weight parameters in the network are not changed. At the same time, FCN generally adopts transfer learning to reuse and transform the pre-trained image classification model.

### 3.3.2.4 GoogleNet

GoogleNet (Szegedy et al., 2015) focused on how to build deeper network structures and introduced a new basic structure, the Inception module (see Figure 5), to increase the width of the network. Google Network V1 is deeper than Alex Net or VGGNet, but it has less computational power than Alex Net and far better accuracy than Alex Net, which is a very practical model. GoogleNet V1 reduces fewer parameters but works well for the following reasons: First, it removes the final fully connected layer and replaces it with a global average pooling layer, which makes the model training faster and reduces overfitting. In addition, the initial module improves the utilization of parameters.

Deep networks are often more difficult to train as the number of layers increases. As some networks begin to converge, they may also have degradation problems that cause accuracy to saturate rapidly. The deeper the level, the higher the error rate. More surprisingly, the higher error rate caused

Figure 5. Inception module with dimensionality reduction from the googlenet architecture (Li et al., 2021)

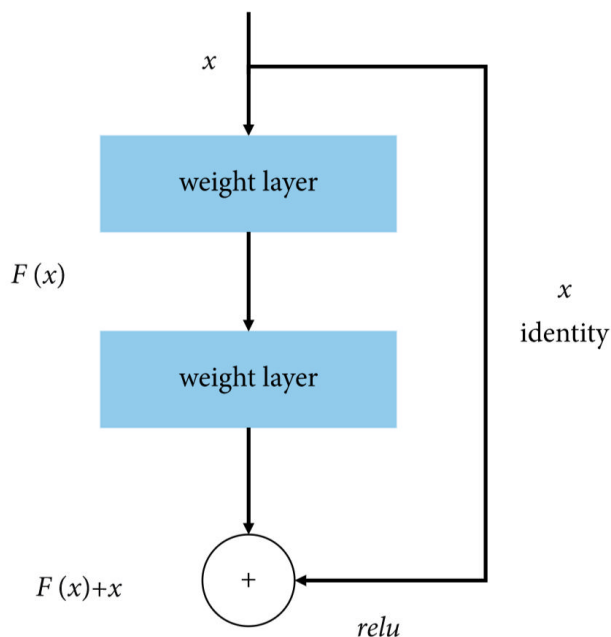


by this degradation is not overfitting, but the addition of more layers. To solve the degradation problem, a deep residual learning framework is proposed, which can successfully train hundreds of residual networks. In contrast to ordinary neural networks, residual networks introduce cross-layer connections or shortcut connections to construct residual modules (see Figure 6).

### 3.3.2.5 ResNet

ResNet, proposed by He et al. (2016), Microsoft Research, is the champion of ImageNet2015. It is more powerful than the traditional CNN in depth and accuracy by letting the network learn the residuals.

Figure 6. Residual block from the ResNet architecture (Li et al., 2021)



The ResNet structure is designed with 8 modules, which are successively connected from the input layer to a convolution layer, and then to the Contlock1-4 modules. At the end of the Contlock4 output connection network, the modules are pooled at an average rate of 11 to retain more coding information. Traditional convolutional neural networks always encounter the problem of gradient disappearance or degradation when the network is deep. The biggest improvement of residual networks is that the identity mapping layer is added to the network structure, which makes the network not degenerate with the increase of depth, not easy to overfit, and has a good convergence effect. Figure 7 shows the ResNet-34 network structure.

### 3.3.2.6 DenseNet

DenseNet (Huang et al., 2017) is a convolutional neural network with dense links, as shown in Figure 8. In DenseNet, each layer is connected to every other layer in a feedforward manner. There is a direct link between any two layers, that is to say the input of each layer of the network is the union of the output of all previous layers. Compared with traditional convolutional networks, it can effectively alleviate the problem of gradient disappearance, strengthen feature propagation, and enhance feature reuse through such dense links. In addition, the number of parameters can be greatly reduced because there is no need to relearn redundant feature maps.

### 3.3.2.7 U-Net

The U-Net (Ronneberger et al., 2015) tongue segmentation network model is a deep learning network model composed of 11 “convolutional pool” layers (Figure 9(a)). Semantic segmentation

Figure 7. ResNet-34 structure (Li et al., 2022)

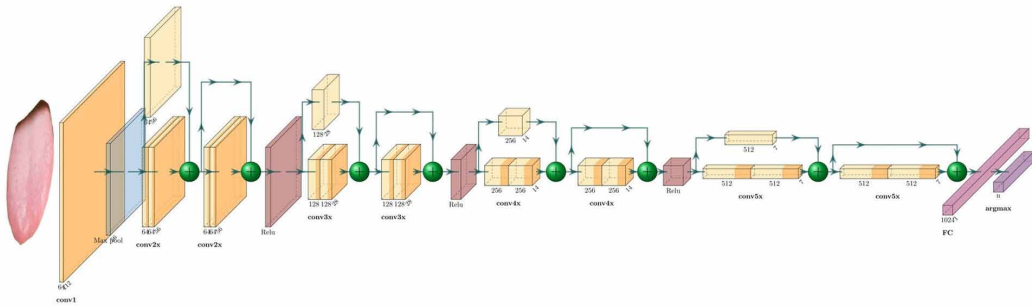


Figure 8. A 5-layer dense block with a growth rate of  $k=4$ . Each layer takes all preceding feature maps as input (Huang et al., 2017).

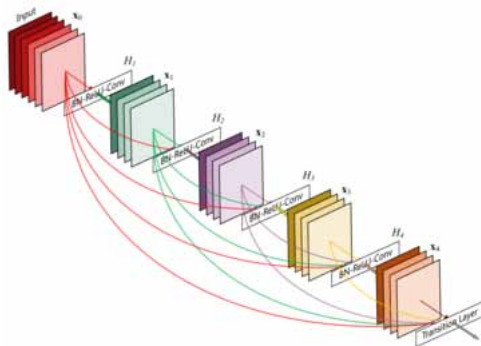
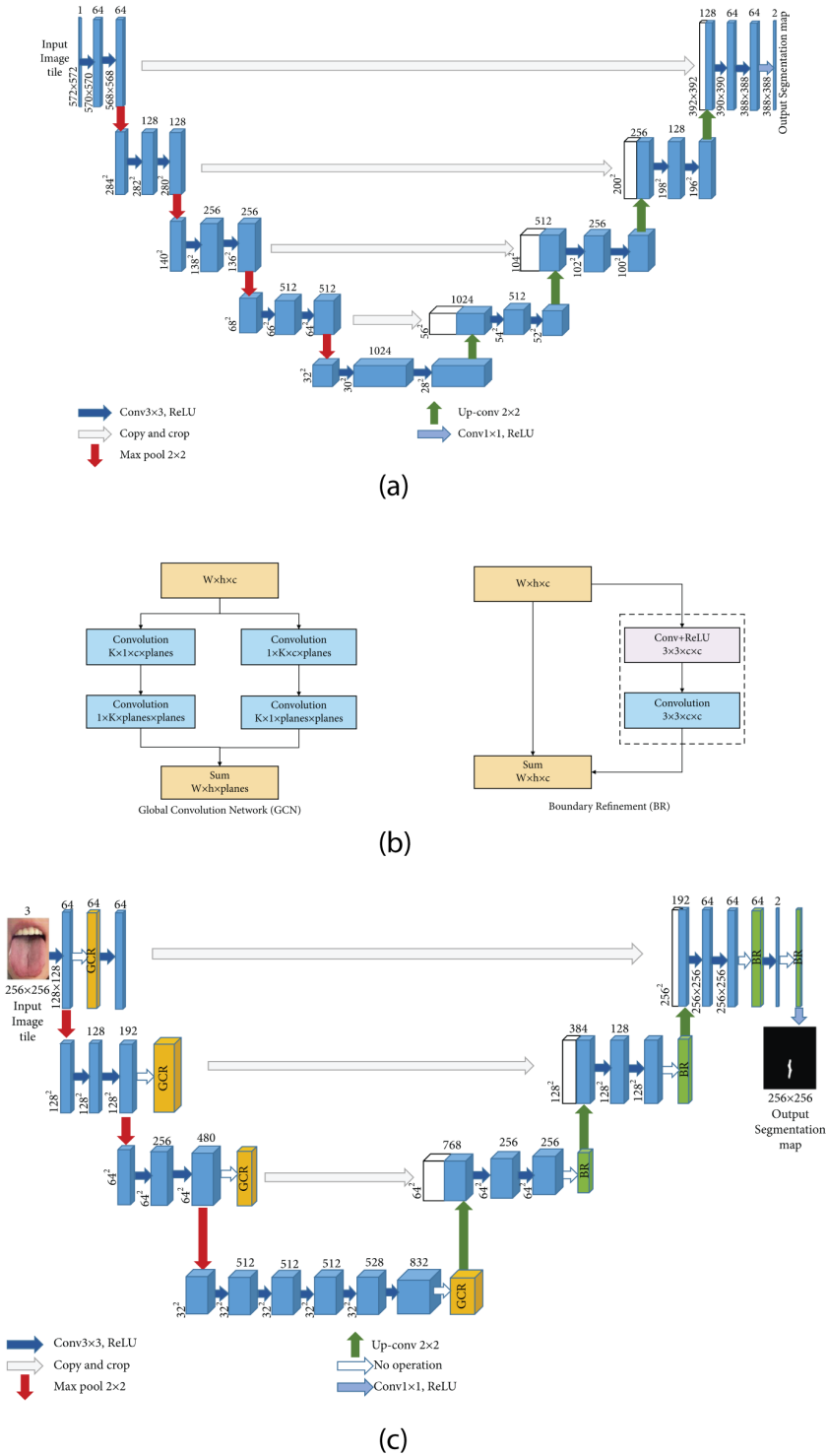


Figure 9. The 3D U-Net architecture. Blue boxes represent feature maps. The number of channels is denoted above each feature map. (a) U-Net network structure, (b) GCR module and BR module, and (c) Improved U-Net network structure (Li et al., 2021).



test results show that the average intersection ratio (MIoU) of the proposed model is 91% and the pixel accuracy (PA) is 93%.

U-Net has good performance in medical image segmentation and is superior to other codec structure networks in small target segmentation performance. Therefore, the U-Net network is selected as the main segmentation model for Chinese medicine split-tongue image. Due to the influence of light intensity, diet, and drugs, tongue images are characterized by a large amount of information and many features. FCN (Long et al., 2015) and SegNet (Badrinarayanan et al., 2017) are not fine enough for crack segmentation of TCM tongue images, and it is easy to lose detailed information. Compared with them, U-Net can obtain a better segmentation effect. Therefore, we propose an improved U-Net network structure (Li et al., 2021), boundary refinement block (BR) (Peng et al., 2017) (see Figure 9(b)), to replace the post-processing common report format module to solve the problem that small targets are difficult to accurately segment. The improved U-Net model uses pre-trained GoogleNet as the feature extraction network for image feature extraction. After feature extraction, feature information is added through the Global Convolutional (GCN) module and BR module. Through this operation, the decoder can better recover the image details and spatial dimensions through the upsampling operation. The improved U-Net network can effectively increase the size of the receiving field by increasing the size of the convolution kernel and improving the segmentation accuracy of small targets. The improved U-Net model still retains the encoder-decoder structure, as shown in Figure 9(c).

Deep learning technology is feasible and has advantages in the research of tongue diagnosis in traditional Chinese medicine. However, due to its short application time and rapid technological innovation, a large number of experiments are still needed to further explore the application value of its different models in the objectification of tongue diagnosis in traditional Chinese medicine. On the other hand, deep learning technology also has some shortcomings. For example, compared with traditional image recognition methods, deep learning models need the training set data model for early training. Here, the sample size and labeling cost required by the training set limit the application of deep learning technology to a certain extent.

### 3.3.3 Deep Learning for Tongue Segmentation

Since 2012, deep learning has been successfully applied to various computer vision tasks. For the tongue image segmentation, researchers are either using existing networks directly (Cai et al., 2020; Li et al., 2019; Qu et al., 2017; Zhou et al., 2019) or modifying them to design specific networks (Huang et al., 2022; Li et al., 2017b; Lin et al., 2018). The application of deep learning in tongue segmentation has many advantages, such as high accuracy, automation, strong adaptability, and scalability, but it also has some disadvantages, such as large data demand, large consumption of computing resources, poor interpretability, and overfitting risk. In practical application, it is necessary to take these factors into consideration comprehensively and combine professional knowledge and auxiliary technology to obtain better tongue segmentation results.

## 3.4 Tongue/Coating Segmentation

Tongue diagnosis can be used to identify constitution types using tongue images. This is because the color and texture features of tongue coating images reflect the health status of patients (Kamarudin et al., 2016). The purpose of tongue texture and coating segmentation is to remove the interference environment, such as face and lips, from the acquired tongue image, segment the effective color details of tongue parts that can be used for analysis, fully retain the patient's tongue information, and eliminate the interference of irrelevant background for further classification and recognition of the tongue body.

Zhang et al. (2019) achieved the separation of tongue lines and coating on the tongue image by using computer technology. The tongue pattern and tongue coating template are identified according to the location information, and the color information of the tongue pattern and tongue coating area is obtained, respectively. The color information is used as the input of the Back-Propagation neural

network, and the tongue pattern color and tongue coating color are used as the output. Finally, the automatic classification of tongue pattern color and tongue coating color is realized by a neural network. The tongue pattern color and tongue coating color can be automatically classified with high accuracy. However, there are not enough image samples to detect the color of various tongue colors and coatings. Therefore, Ma et al. (2019) applied the deep learning method to propose a novel complexity sensing classification method to automatically realize the system framework of body mass recognition. The system framework consists of tongue image acquisition, tongue coating detection, tongue coating calibration, tongue feature extraction, and body mass classification. The method considers the complexity of individual-level instances to reduce the influence of uneven image distribution caused by various environmental conditions, such as lighting and resolution. The proposed complexity-aware method is universal and can be easily extended to other application scenarios.

#### 4. ANALYSIS OF TONGUE IMAGE FEATURES

In traditional Chinese medicine, tongue color can reflect the health status of the human body, which is one of the important dialectical means in tongue diagnosis and can also be used as an indicator for daily health self-monitoring (Yuan-tong et al., 2020). Therefore, it is of great practical significance to classify and identify tongue color and moss color. The color of the tongue is generally divided into five kinds: light red, light white, red crimson, green, and purple (Figure 10) (Chen et al., 2022; Oji et al., 2014). The color of tongue coating is generally divided into three kinds: white, yellow, and gray-black (Figure 10). Tongue coating color classification is based on neural convolutional network architecture, while tongue image classification is based on a small sample fully connected neural network model. At present, most convolutional neural network structures are suitable for the classification task of a single label. By constructing multiple network models, the classification tasks, such as tongue color and morphology, are mostly carried out separately, ignoring the possible relevance of different features of the tongue.

Tongue image classification and recognition include tongue color recognition, tongue shape recognition, tooth-marked, and so on. Different tongue states are shown in Figure 11. Before deep

Figure 10. Lingual coating information (Chen et al., 2022; Lundgren et al., 2007; Tang et al., 2021)

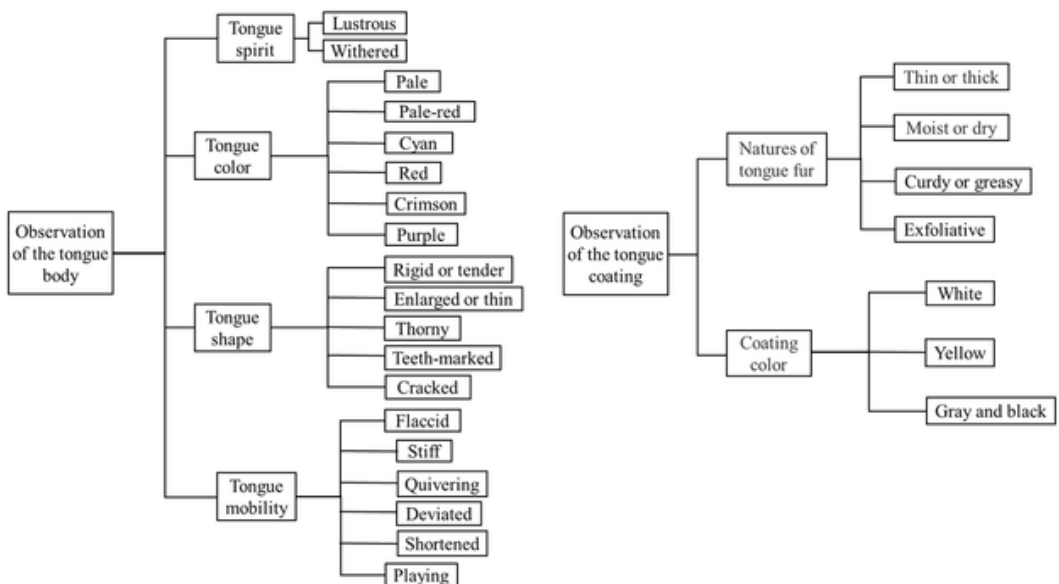
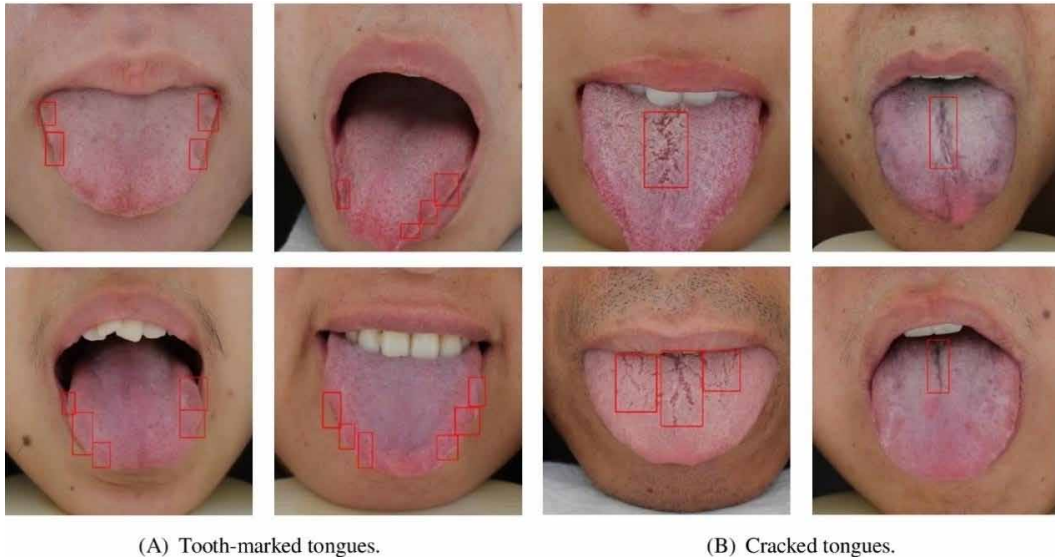


Figure 11. Sample images of (a) tooth-marked tongue and (b) cracked tongue (Gunal & Edizkan, 2008)



learning is mature, clustering algorithms, support vector machines, Bayesian network, and other methods are the mainstream methods for tongue image regime recognition.

#### 4.1 Tongue Color

Tongue color is one of the important indicators in tongue diagnosis in traditional Chinese medicine. Usually, tongue color shows the description of common color names on the tongue body. Chinese medicine is a very convenient way to express the tongue color of the patient. For example, a normal tongue is identified as light red, while a pale tongue indicates a “cold” condition. However, it is not an exact commutation method. Therefore, it is necessary to use more scientific methods to assist naked eye judgment in the process of contemporary TCM diagnosis. Chen et al. (2022) succeeded in scientific quantification and computational simulation analysis of tongue color in TCM calculation of tongue diagnosis. The hue-lightness-chroma data of the chosen Munsell color charts were transformed to CIE xyY using the look-up table computation, then further converted to CIELAB values and sRGB data. On the other hand, the establishment of visual assessment data based on visual assessment experiments has been proven to be an objective, reliable method when applied to color science and color engineering (Kawanabe et al., 2016).

#### 4.2 Tongue Shape

The tongue consists of a body, a tip, and a root. These parts give the tongue its shape, which is not permanent but may change. In traditional Chinese medicine, these changes are used to indicate specific pathologies (Gunal & Edizkan, 2008). The normal tongue is considered an oval tongue, but traditional Chinese medicine also has six other types of tongue shapes: square, rectangle, round, acute triangle, blunt triangle, and hammer. Changes in the shape of the tongue, which do not begin and end within 24 hours, but may occur and persist for months, reflect the course of the disease. This constant change creates diagnostic uncertainty, which is the ambiguity that falls into one or more categories, only up to a point. This uncertainty makes it difficult to analyze the tongue shape. Therefore, Huang et al. (2010) proposed to correct tongue deflection by using automatic contour extraction and a combination of length, area, and angle criteria. The features of tongue shape were then defined using seven sub-features defined from length, area, and angle information. To translate



human judgments into classification decisions, a decision support tool called Analytical Hierarchy Process (AHP) is applied, where the relationship of each sub-feature to shape has been characterized and given weights on a standardized numerical scale. The fuzzy fusion framework is used to combine the seven modules of AHP, classify each tongue image into seven classes of tongue shape, and model the relationship between diagnostic uncertainty and definition class. The proposed shape correction method reduces the deflection of tongue shape, and the shape classification method has been tested on a total of 362 tongue samples with an accuracy of 90.3%.

### 4.3 Tongue Crack

In the research field of objectification of tongue diagnosis in traditional Chinese medicine, there are few specific studies on analyzing the texture of tongue coating in tongue images (Shenhua & Jiang, 2016). Correlational research including Dapeng Zhang (Tingting et al., 2016) put forward based on Gabor wavelet transforming tongue picture of tongue coating texture analysis of digital tongue coating texture analysis is focused on the management of digital tongue image analysis and identification of thin/thick and greasy tongue coating characteristics (namely the tongue coating is thin or thick, whether greasy tongue coating) (Lee et al., 2016b). Yan et al. (2022) proposed a tongue image texture classification method based on image interpolation and convolutional neural network, and completed the classification of tough tongue, soft tongue, and normal tongue texture types. As described in 2.3.7, we proposed an improved split-tongue image segmentation model based on the U-Net model and constructed a database of split-tongue images (Li et al., 2021). The improved U-Net structure has good performance, achieves crack extraction, and achieves a good trade-off between effective acceptance and the number of parameters.

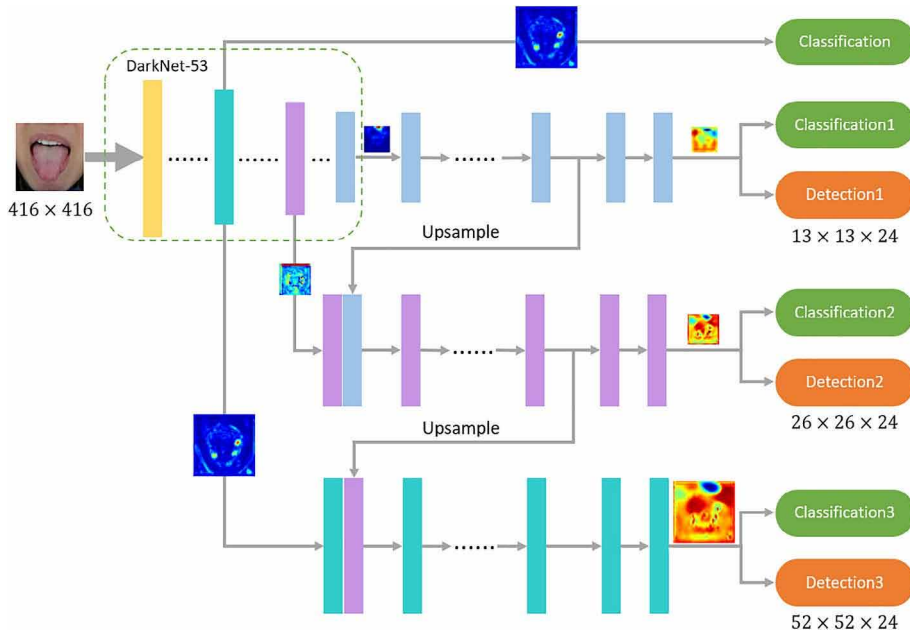
Although the improved U-Net model has been greatly improved to a certain extent, the experiment also has some limitations. As can be seen from the performance of the model on the test dataset, the model still needs to be improved. Meanwhile, the mapping from input to output in the process of neural network learning is discontinuous (Szegedy et al., 2013). This discontinuity means the picture may deceive the model and produce wrong judgment after proper modification (Mani et al., 2020). In subsequent work, we need to conduct adversarial example attack experiments on the model and modify the training samples (Goodfellow et al., 2015). By adding more adversarial samples to the training set, we can effectively avoid some attacks. We can test the model based on clean data by adding a small amount of noise that the human eye cannot detect. In the encoder design part, we can add additional networks on top of the GoogleNet network to keep the original network unchanged.

### 4.4 Tooth-Marked

In tongue diagnosis, identifying the tongue with dental markers plays an important role in assessing the patient's status. Teeth marks appear mainly along the posterior edge of the tongue, which is mainly caused by the compression of adjacent teeth. According to the theory of traditional Chinese medicine, the dentate tongue usually has an enlarged shape, which is related to spleen deficiency, qi weakness, Yang deficiency, and so on. Identification of tongues with dental markers can provide meaningful guidance for the differentiation of clinical syndromes.

Weng et al. (2021) identified dentate tongue and tongue crack simultaneously as an object detection task, identifying the dentate tongue and dehiscent tongue at the same time, and locating the location of tooth marks and cracks in tongue images. Based on DarkNet, a weakly supervised YOLO model is proposed, which consists of a multi-scale feature coding module, classification module, and detection module (as shown in Figure 12). The localization information can be further used to assess tongue severity and calculate more detailed information. The method can be trained with full bounding box level annotations and coarse image level annotations. Experimental results on challenging tongue images demonstrate the effectiveness of the proposed method, which can significantly reduce the annotation workload of tooth marks and crack detection tasks.

Figure 12. The network architecture of the proposed weakly supervised YOLO (Weng et al., 2021)



## 5. SUMMARY

Tongue diagnosis is one of the important diagnostic methods of traditional Chinese medicine, and the objective realization of tongue diagnosis is also the development direction of tongue diagnosis in traditional Chinese medicine. The machine learning algorithm is widely used in the process of image segmentation, tongue image classification, and feature recognition, which has good recognition effect and generalization ability, and greatly promotes the objectification process of tongue image diagnosis in traditional Chinese medicine. There are still some problems in the intelligent process of tongue diagnosis:

1. The performance of AI-based tongue image recognition methods is highly dependent on the data set used. However, most current objectification studies lack balanced large sample data. Therefore, further efforts should be placed on building larger, diverse, and standardized tongue image datasets. Such a dataset could include samples of different ages, genders, disease types, and severity to better reflect the real clinical situation.
2. Tongue image recognition may have some limitations in the diagnosis of tongue image only. The lack of classified storage, efficient utilization, and comprehensive analysis of a large number of tongue image data fails to effectively establish the corresponding relationship between tongue image information and TCM syndrome types, which is disconnected from clinical practice. Association analysis of tongue image features with other medical data (such as clinical symptoms, signs, and medical history) can improve the comprehensive diagnostic ability of tongue image recognition. For example, by combining tongue features with clinical data from patients, models can be built to predict the risk of a specific disease or the progression of the disease, providing more accurate guidance for personalized treatment.
3. The establishment of multi-disciplinary and multi-expert systems is not perfect. The determination of tongue image needs to involve the comprehensive diagnosis of tongue god, tongue color, tongue shape, tongue form, tongue coating, and sublingual choroid veins to lay the foundation. The tongue

image diagnosis system developed at present is usually a single system. The problem-solving method is relatively single, the problem-solving field is relatively narrow, and the lack of unified quantitative standards mean that the tongue diagnosis information cannot be comprehensively integrated with recognition, judgment, analysis, and processing.

4. Tongue recognition models based on deep learning are often considered to be black box models whose decision-making process lacks interpretation and interpretability. This could raise some concerns in the medical field, where doctors and patients need to understand how models make judgments based on tongue features. Therefore, future research should focus on improving the interpretability of tongue image recognition models. For example, by visualizing important features, the model's attention mechanism, or by generating explanatory reports to increase the credibility and acceptability of the models.
5. Tongue image recognition methods involve personal privacy and ethical issues of data use. When using tongue image data, it is necessary to ensure the security and confidentiality of the data and to comply with applicable privacy and ethical regulations. In addition, appropriate ethical review and informed consent should be obtained for research involving human participation.

To sum up, tongue image recognition methods based on artificial intelligence have potential in tongue image diagnosis, but still face some challenges. By improving the quality and diversity of datasets, correlating tongue features with other medical data, improving the interpretability and interpretability of models, conducting clinical validation, and addressing privacy and ethical issues, the field can be further advanced and more accurate and reliable AIDS for TCM diagnosis and treatment can be provided. Therefore, we put forward the development direction of intelligent tongue diagnosis in TCM:

1. **Improve the quality and quantity of tongue images:** Further improve the quality and accuracy of tongue image. Construct larger, diversified, and standardized tongue image data sets.
2. **Feature extraction:** Defining and extracting more features of tongue images is the most important task for computerized tongue diagnosis in the future. In the future computer tongue diagnosis system, TCM tongue diagnosis can be assisted by using only the mapping relationship between a certain feature and clinical disease, thus further simplifying the steps of computer tongue diagnosis.
3. **Feature fusion:** Computer tongue diagnosis combined with other diagnostic methods to promote the objective study of the four TCM diagnoses.
4. **System integration and testing:** Integrating the research results of computer tongue diagnosis into the system and conducting large-scale clinical trials in some hospitals is the key step for computer tongue diagnosis technology to go to the market.

## REFERENCES

- Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), 2481–2495. doi:10.1109/TPAMI.2016.2644615 PMID:28060704
- Bhatnagar, V., & Bansod, P. P. (2022). Challenges and solutions in automated tongue diagnosis techniques: A review. *Critical Reviews in Biomedical Engineering*, 50(1), 47–63. doi:10.1615/CritRevBiomedEng.2022044392 PMID:35997110
- Cai, Y., Wang, T., Liu, W., & Luo, Z. (2020). A robust interclass and intraclass loss function for deep learning based tongue segmentation. *Concurrency and Computation*, 32(22). Advance online publication. doi:10.1002/cpe.5849
- Chen, H. S., Chen, S. M., Jiang, C. Y., Zhang, Y. C., Lin, C. Y., Lin, C. E., & Lee, J. A. (2022). Computational tongue color simulation in tongue diagnosis. *Color Research and Application*, 47(1), 121–134. doi:10.1002/col.22717
- Chen, L.-C., Papandreou, G., Schroff, F., & Adam, H. (2017). *Rethinking atrous convolution for semantic image segmentation*. arXiv preprint arXiv:1706.05587. /arxiv.1706.05587<ALIGNMENT.qj></ALIGNMENT>10.48550
- Chung, C. J., Wu, C. H., Hu, W. L., Shih, C. H., Liao, Y. N., & Hung, Y. C. (2022). Tongue diagnosis index of chronic kidney disease. *Biomedical Journal*. Advance online publication. doi:10.1016/j.bj.2022.02.001 PMID:35158075
- da Xu, L. (2013). Introduction: Systems science in industrial sectors. *Systems Research and Behavioral Science*, 30(3), 211–213. doi:10.1002/sres.2186
- Du, J. Q., Lu, Y. S., Zhu, M. F., Zhang, K., & Ding, C. H. (2008, May). A novel algorithm of color tongue image segmentation based on HSI. *2008 International Conference on BioMedical Engineering and Informatics*, 1, 733–737. doi:10.1109/BMEI.2008.271
- Duan, J. L., Deng, B., Song, G. H., Chen, Z. F., Gong, Y. W., He, Y. H., & Jia, L. Q. (2018). Application of computer-aided tongue inspection for preliminary screening of esophageal cancer. *Chinese Journal of Integrative Medicine*, 24(10), 746–751. doi:10.1007/s11655-018-2840-6 PMID:29671118
- Fan, S., Chen, B., Zhang, X., Hu, X., Bao, L., Yang, X., Liu, Z., & Yu, Y. (2021). Machine learning algorithms in classifying TCM tongue features in diabetes mellitus and symptoms of gastric disease. *European Journal of Integrative Medicine*, 43, 101288. doi:10.1016/j.eujim.2021.101288
- Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. *Computer Science*. <https://arxiv.org/abs/1412.6572>
- Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., & Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, 77, 354–377. doi:10.1016/j.patcog.2017.10.013
- Guangyu, N., Caiqun, W., Bo, Y., & Yong, P. (2021). Tongue color classification based on convolutional neural network. *Advances in Intelligent Systems and Computing*, 1364, 649–662. <ALIGNMENT.qj></ALIGNMENT>10.1007/978-3-030-73103-8\_46/COVER
- Gunal, S., & Edizkan, R. (2008). Subspace based feature selection for pattern recognition. *Information Sciences*, 178(19), 3716–3726. doi:10.1016/j.ins.2008.06.001
- Guo, J., Yang, Y., Wu, Q., Su, J., & Ma, F. (2016, October). Adaptive active contour model based automatic tongue image segmentation. *2016 9th International Congress on image and signal processing, BioMedical engineering and informatics (CISP-BMEI)*, 1386–1390. doi:10.1109/CISP-BMEI.2016.7852933
- Han, K., Qian, J., Zhang, W., Li, F., & Li, X. (2018). Research of tongue diagnosis characteristic of colorectal cancer. *Chinese Archives of Traditional Chinese Medicine*, 36(3), 668–671. .issn. 1673-7717. 2018.03.036<ALIGNMENT.qj></ALIGNMENT>10. 13193 /j

- Han, S., Yang, X., Qi, Q., Pan, Y., Chen, Y., Shen, J., Liao, H., & Ji, Z. (2016). Potential screening and early diagnosis method for cancer: Tongue diagnosis. *International Journal of Oncology*, 48(6), 2257–2264. doi:10.3892/ijo.2016.3466 PMID:27035407
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE computer society conference on computer vision and pattern recognition*, 770–778. doi:10.1109/CVPR.2016.90
- He, M. Q., Liu, L., Ma, J., Li, J., Zhen, X. M., & Zhang, Y. Y. (2016). Risk factors and clinical features in ischemic stroke patients with different tongue conditions. *Chinese Journal of Integrative Medicine*, 1–9. doi:10.1007/s11655-015-2442-5 PMID:26956463
- Hu, Y., Wen, G., Luo, M., Yang, P., Dai, D., Yu, Z., Wang, C., & Hall, W. (2021). Fully-channel regional attention network for disease-location recognition with tongue images. *Artificial Intelligence in Medicine*, 118, 102110. doi:10.1016/j.artmed.2021.102110 PMID:34412836
- Huang, B., Wu, J., Zhang, D., & Li, N. (2010). Tongue shape classification by geometric features. *Information Sciences*, 180(2), 312–324. doi:10.1016/j.ins.2009.09.016
- Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE Conference on computer vision and pattern recognition*, 2261–2269. doi:10.1109/CVPR.2017.243
- Huang, J., Xu, J., Zhang, Z., Tu, L., Hu, X., Cui, L., & Jiang, T. (2018). Influence of ‘daily rhythm’ factors on tongue image characteristics of healthy people. *CJT CMP*, 33(8), 3462–3465.
- Huang, X., Zhuo, L., Zhang, H., Li, X., & Zhang, J. (2022). Lw-TISNet: Light-weight convolutional neural network incorporating attention mechanism and multiple supervision strategy for tongue image segmentation. *Sensing and Imaging*, 23(1), 6. Advance online publication. doi:10.1007/s11220-021-00375-x
- Jiang, B., Liang, X., Chen, Y., Ma, T., Liu, L., Li, J., Jiang, R., Chen, T., Zhang, X., & Li, S. (2012). Integrating next-generation sequencing and traditional tongue diagnosis to determine tongue coating microbiome. *Scientific Reports*, 2(1), 936. Advance online publication. doi:10.1038/srep00936 PMID:23226834
- Kajikawa, T., Kadoya, N., Ito, K., Takayama, Y., Chiba, T., Tomori, S., Nemoto, H., Dobashi, S., Takeda, K., & Jingu, K. (2019). A convolutional neural network approach for IMRT dose distribution prediction in prostate cancer patients. *Journal of Radiation Research*, 60(5), 685–693. doi:10.1093/jrr/rrz051 PMID:31322704
- Kamarudin, N. D., Ooi, C. Y., Kawanabe, T., & Mi, X. (2016). Tongue’s substance and coating recognition analysis using HSV color threshold in tongue diagnosis. *First International Workshop on Pattern Recognition, 10011*, 109–113. doi:10.1117/12.2242404
- Kass, M., Witkin, A., & Terzopoulos, D. (1988). Snakes: Active contour models. *International Journal of Computer Vision*, 1(4), 321–331. doi:10.1007/BF00133570
- Kawanabe, T., Kamarudin, N. D., Ooi, C. Y., Kobayashi, F., Mi, X., Sekine, M., Wakasugi, A., Odaguchi, H., & Hanawa, T. (2016). Quantification of tongue colour using machine learning in Kampo medicine. *European Journal of Integrative Medicine*, 8(6), 932–941. doi:10.1016/j.eujim.2016.04.002
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90. doi:10.1145/3065386
- Kumar, S., & Hebert, M. (2003). Discriminative random fields: A discriminative framework for contextual interaction in classification. *Proceedings - IEEE International Conference on Computer Vision*, 2, 1150–1157. doi:10.1109/ICCV.2003.1238478
- Lee, T. C., Lo, L. C., & Wu, F. C. (2016). Traditional Chinese medicine for metabolic syndrome via TCM pattern differentiation: Tongue diagnosis for predictor. *Evidence-Based Complementary and Alternative Medicine*, 2016, 1–8. Advance online publication. doi:10.1155/2016/1971295 PMID:27313640
- Li, D., & Wei, Y. (2011). Tongue image segmentation method based on adaptive thresholds. *Computer Technology and Development*, 21(9), 64–65.
- Li, J., Pan, B., & Wei, Y. (2009). Tongue image segmentation based on fuzzy rough sets. *2009 International Conference on Environmental Science and Information Application Technology*, 3, 367–369. doi:10.1109/ESIAT.2009.211

- Li, J., Xu, B., Ban, X., Tai, P., & Ma, B. (2017a). A tongue image segmentation method based on enhanced HSV convolutional neural network. *Lecture Notes in Computer Science*, 10451, 252–260. doi:<ALIGNMENT.qj></ALIGNMENT>10.1007/978-3-319-66805-5\_32/COVER
- Li, J., Xu, B., Ban, X., Tai, P., & Ma, B. (2017b). A tongue image segmentation method based on enhanced HSV convolutional neural network. *Cooperative Design, Visualization, and Engineering: 14th International Conference*, 252–260. doi:10.1007/978-3-319-66805-5\_32
- Li, J., Zhang, Z., Zhu, X., Zhao, Y., Ma, Y., Zang, J., Li, B., Cao, X., & Xue, C. (2022). Automatic classification framework of tongue feature based on convolutional neural networks. *Micromachines*, 13(4), 501. doi:10.3390/mi13040501 PMID:35457806
- Li, M. Y., Zhu, D. J., Xu, W., Lin, Y. J., Yung, K. L., & Ip, A. W. H. (2021). Application of U-Net with global convolution network module in computer-aided tongue diagnosis. *Journal of Healthcare Engineering*, 2021(22), 1948–1952. doi:10.1155/2021/5853128 PMID:34840700
- Li, Q., & Liu, Z. (2009). Tongue color analysis and discrimination based on hyperspectral images. *Computerized Medical Imaging and Graphics*, 33(3), 217–221. doi:10.1016/j.compmedimag.2008.12.004 PMID:19157779
- Li, Q., Wang, Y., Liu, H., Sun, Z., & Liu, Z. (2010). Tongue fissure extraction and classification using hyperspectral imaging technology. *Applied Optics*, 49(11), 2006–2013. doi:10.1364/AO.49.002006 PMID:20389998
- Li, X., Yang, D., Wang, Y., Yang, S., Qi, L., Li, F., Gan, Z., & Zhang, W. (2019). Automatic tongue image segmentation for real-time remote diagnosis. *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 409–414. doi:10.1109/BIBM47256.2019.8982947
- Li, Z., Ren, X., Xiao, L., Qi, J., Fu, T., & Li, W. (2022). Research on data analysis network of TCM tongue diagnosis based on deep learning technology. *Journal of Healthcare Engineering*, 2022, 1–9. Advance online publication. doi:10.1155/2022/9372807 PMID:35392154
- Lin, B., Xle, J., Li, C., & Qu, Y. (2018). Deeptongue: Tongue segmentation via resnet. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1035–1039. doi:10.1109/ICASSP.2018.8462650
- Liu, Z., Wang, H., & Li, Q. (2011). Tongue tumor detection in medical hyperspectral images. *Sensors (Basel)*, 12(1), 162–174. doi:10.3390/s120100162 PMID:22368462
- Lo, L. C., Cheng, T. L., Chen, Y. J., Natsagdorj, S., & Chiang, J. Y. (2015). TCM tongue diagnosis index of early-stage breast cancer. *Complementary Therapies in Medicine*, 23(5), 705–713. doi:10.1016/j.ctim.2015.07.001 PMID:26365451
- Lo, L. C., Cheng, T. L., Chiang, J. Y., & Damdinsuren, N. (2013). Breast cancer index: A perspective on tongue diagnosis in traditional Chinese medicine. *Journal of Traditional and Complementary Medicine*, 3(3), 194–203. doi:10.4103/2225-4110.114901 PMID:24716178
- Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 431–440. doi:10.1109/CVPR.2015.7298965
- Lu, Y., Li, X., Gong, Z., Zhuo, L., & Zhang, H. (2020). TDCCN: A two-phase deep color correction network for traditional Chinese medicine tongue images. *Applied Sciences (Basel, Switzerland)*, 10(5), 1784. doi:10.3390/app10051784
- Lundgren, T., Mobilia, A., Hallström, H., & Egelberg, J. (2007). Evaluation of tongue coating indices. *Oral Diseases*, 13(2), 177–180. doi:10.1111/j.1601-0825.2006.01261.x PMID:17305619
- Luo, J., Wang, J., Shu, C., Tu, L., Ma, X., Li, J., Shi, Q., Xu, J., & Wang, Y. (2020). Tongue parameter features of the cases with primary osteoporosis. *Geriatr Health Care*, 26(1), 25–28.
- Ma, J., Wen, G., Wang, C., & Jiang, L. (2019). Complexity perception classification method for tongue constitution recognition. *Artificial Intelligence in Medicine*, 96, 123–133. doi:10.1016/j.artmed.2019.03.008 PMID:31164206
- Mani, N., Moh, M., & Moh, T. S. (2020). Defending deep learning models against adversarial attacks. *International Journal of Software Science and Computational Intelligence*, 13(1), 72–89. doi:10.4018/IJSSCI.2021010105

- Mukai, N., Mori, K., & Takei, Y. (2022). Tongue model construction based on ultrasound images with image processing and deep learning method. *Journal of Medical Ultrasonics*, 49(2), 153–161. doi:10.1007/s10396-022-01193-8 PMID:35181818
- Oji, T., Namiki, T., Nakaguchi, T., Ueda, K., Takeda, K., Nakamura, M., Okamoto, H., & Hirasaki, Y. (2014). Study of factors involved in tongue color diagnosis by Kampo medical practitioners using the Farnsworth-Munsell 100 hue test and tongue color images. *Evidence-Based Complementary and Alternative Medicine*, 2014, 1–9. Advance online publication. doi:10.1155/2014/783102 PMID:24808919
- Pang, B., Zhang, D., & Wang, K. (2005). Tongue image analysis for appendicitis diagnosis. *Information Sciences*, 175(3), 160–176. doi:10.1016/j.ins.2005.01.010
- Park, S. H., Shin, N. R., Yang, M., Bose, S., Kwon, O., Nam, D. H., Lee, J. H., Song, E. J., do Nam, Y., & Kim, H. (2022). A clinical study on the relationship among insomnia, tongue diagnosis, and oral microbiome. *The American Journal of Chinese Medicine*, 50(3), 773–797. doi:10.1142/S0192415X2250032X PMID:35380093
- Peng, C., Zhang, X., Yu, G., Luo, G., & Sun, J. (2017). Large kernel matters - Improve semantic segmentation by global convolutional network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1743–1751. doi:10.1109/CVPR.2017.189
- Qi, Z., Tu, L. P., Chen, J. B., Hu, X. J., Xu, J. T., & Zhang, Z. F. (2016). The classification of tongue colors with standardized acquisition and ICC profile correction in traditional Chinese medicine. *BioMed Research International*, 2016, 1–9. Advance online publication. doi:10.1155/2016/3510807 PMID:28050555
- Qu, P., Zhang, H., Zhuo, L., Zhang, J., & Chen, G. (2017). Automatic Tongue Image Segmentation for Traditional Chinese Medicine Using Deep Neural Network. *Intelligent Computing Theories and Application: 13th International Conference*, 247–259. doi:10.1007/978-3-319-63309-1\_23
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science*, 9351, 234–241. doi:<ALIGNMENT.qj></ALIGNMENT>10.1007/978-3-319-24574-4\_28/COVER
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision*, 618–626. doi:10.1109/ICCV.2017.74
- Shenhua, J., & Jiang, L. (2016). Review on objective and essential research of tongue image. *Shanghai J Traditional Chinese Med*, 50(7), 94–97.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations*. doi:10.48550/arxiv.1409.1556
- Stoecklein, D., Lore, K. G., Davies, M., Sarkar, S., & Ganapathysubramanian, B. (2017). Deep learning for flow sculpting: Insights into efficient learning using scientific simulation data. *Scientific Reports*, 7(1), 1–11. doi:10.1038/srep46368 PMID:28402332
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1–9. doi:10.1109/CVPR.2015.7298594
- Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (2013). Intriguing properties of neural networks. *Computer Science*. Advance online publication. doi:10.48550/arxiv.1312.6199
- Tang, Y., Sun, Y., Chiang, J. Y., & Li, X. (2021). Research on multiple-instance learning for tongue coating classification. *IEEE Access : Practical Innovations, Open Solutions*, 9, 66361–66370. doi:10.1109/ACCESS.2021.3076604
- Tang, Y., Wang, L., He, X., Chen, P., & Yuan, G. (2018). Classification of tongue image based on multi-task deep convolutional neural network. *Computer Science*, 45(12), 255–261. doi:10.11896/J.ISSN.1002-137X.2018.12.042
- Thakur, R. S., Chatterjee, S., Yadav, R. N., & Gupta, L. (2021). Image de-noising with machine learning: A review. *IEEE Access : Practical Innovations, Open Solutions*, 9, 93338–93363. doi:10.1109/ACCESS.2021.3092425

Tingting, Q., Chunming, X., Yiqin, W., & Mulangma, Z. (2016). Recognition of greasy or curdy tongue coating based on gabor wavelet transformation. *Com Appl Softw*, 33(10), 162–166. [https://jglobal.jst.go.jp/en/detail?JGLOBAL\\_ID=201702291769801369](https://jglobal.jst.go.jp/en/detail?JGLOBAL_ID=201702291769801369)

Wan, H. C., & Chin, K. S. (2021). Exploring internet of healthcare things for establishing an integrated care link system in the healthcare industry. *International Journal of Engineering Business Management*, 13. Advance online publication. doi:10.1177/18479790211019526

Wang, A., Zhao, Z., & Shen, L. (2000). Research on the tongue color classification in automatic tongue analysis of traditional Chinese medicine. *Beijing Biomedical Engineering*, 3, 136–142.

Wang, B., Fang, G., Zhou, X., & Yu, C. (2019). A comparative study on tongue diagnostic parameters between phlegmstasis binding pattern and non-phlegm-stasis binding pattern of coronary heart disease at different time points. *Journal of Beijing University of Traditional Chinese Medicine*, 42(11), 963–968. . issn. 1006-2157. 2019. 11. 013<ALIGNMENT.qj></ALIGNMENT>10. 3969 /j

Wang, L., Chen, X., Hu, L., & Li, H. (2020). Overview of image semantic segmentation technology. *2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference*, 9, 19–26. doi:10.1109/ITAIC49862.2020.9338770

Weng, H., Li, L., Lei, H., Luo, Z., Li, C., & Li, S. (2021). A weakly supervised tooth-mark and crack detection method in tongue image. *Concurrency and Computation*, 33(16), e6262. doi:10.1002/cpe.6262

Wu, T. C., Wu, K. L., Hu, W. L., Sheen, J. M., Lu, C. N., Chiang, J. Y., & Hung, Y. C. (2018). Tongue diagnosis indices for upper gastrointestinal disorders: Protocol for a cross-sectional, case-controlled observational study. *Medicine (United States)*, 97(2), e9607. Advance online publication. doi:10.1097/MD.00000000000009607 PMID:29480863

Yamamoto, S., Tsumura, N., Nakaguchi, T., Namiki, T., Kasahara, Y., Terasawa, K., & Miyake, Y. (2010). Regional image analysis of the tongue color spectrum. *International Journal of Computer Assisted Radiology and Surgery*, 6(1), 143–152. doi:10.1007/s11548-010-0492-x PMID:20532648

Yan, J., Chen, B., Guo, R., Zeng, M., Yan, H., Xu, Z., & Wang, Y. (2022). Tongue image texture classification based on image inpainting and convolutional neural network. *Computational and Mathematical Methods in Medicine*, 2022, 1–11. doi:10.1155/2022/6066640 PMID:36570335

Yang, Y., Hallman, S., Ramanan, D., & Fowlkes, C. C. (2012). Layered object models for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(9), 1731–1743. doi:10.1109/TPAMI.2011.208 PMID:22813957

Yoo, S. G., & Oh, D. (2018). An artificial neural network–based fall detection. *International Journal of Engineering Business Management*, 10. Advance online publication. doi:10.1177/1847979018787905

Yuan, L., Yang, L., Zhang, S., Xu, Z., Qin, J., Shi, Y., Yu, P., Wang, Y., Bao, Z., Xia, Y., Sun, J., He, W., Chen, T., Chen, X., Hu, C., Zhang, Y., Dong, C., Zhao, P., Wang, Y., & Cheng, X. (2023). Development of a tongue image-based machine learning tool for the diagnosis of gastric cancer: A prospective multicentre clinical cohort study. *EClinicalMedicine*, 57, 101834. Advance online publication. doi:10.1016/j.eclinm.2023.101834 PMID:36825238

Yuan-tong, L., Yusheng, L., & Zhen-min, Z. (2020). Tongue image analysis in traditional Chinese medicine based on deep learning. *Computer Science*, 47(11), 148–158. doi:10.11896/JSJKX.191000104

Zhang, H., Jiang, R., Yang, T., Gao, J., Wang, Y., & Zhang, J. (2022). Study on TCM tongue image segmentation model based on convolutional neural network fused with superpixel. *Evidence-Based Complementary and Alternative Medicine*, 2022, 1–12. Advance online publication. doi:10.1155/2022/3943920 PMID:35300068

Zhang, H., Zuo, W., Wang, K., & Zhang, D. (2006). A snake-based approach to automated segmentation of tongue image using polar edge detector. *International Journal of Imaging Systems and Technology*, 16(4), 103–112. doi:10.1002/ima.20075

Zhang, J., Xu, J., Hu, X., Chen, Q., Tu, L., Huang, J., & Cui, J. (2017). Diagnostic method of diabetes based on support vector machine and tongue images. *BioMed Research International*, 2017, 1–9. Advance online publication. doi:10.1155/2017/7961494 PMID:28133611



- Zhang, K., Zhang, H., & Zhu, M. (2019). Automatic classification of tongue texture color and tongue coating color based on BP neural network. *Journal of Physics: Conference Series*, 1423(1), 012056. doi:10.1088/1742-6596/1423/1/012056
- Zhang, L., & Qin, J. (2010). Tongue-image segmentation based on gray projection and threshold-adaptive method. *Journal of Clinical Rehabilitative Tissue Engineering Research*, 14(9), 1638–1641. doi:10.3969/J. ISSN.1673-8225.2010.09.027
- Zhang, W. (2017). Semantic segmentation algorithm based on random forests. *Electronic Sci. & Tech*, 30(2), 72–75. . cnki. issn1007 - 7820. 2017. 02. 01910. 16180 /j
- Zhang, X., Huang, W., Gao, J., Wang, D., Bai, C., Chen, Z., Zhang, X., Huang, W., Gao, J., Wang, D., Bai, C., & Chen, Z. (2021). Deep sparse transfer learning for remote smart tongue diagnosis. *Mathematical Biosciences and Engineering*, 18(2), 1169–1186. doi:10.3934/mbe.2021063 PMID:33757181
- Zhao, Z., Wang, A., & Shen, L. (1999). The color tongue image segmentation based on mathematical morphology and HIS model. *Journal of Beijing Polytechnic University*, 25(2), 67–71.
- Zhou, J., Zhang, Q., Zhang, B., & Chen, X. (2019). TongueNet: A precise and fast tongue segmentation system using U-Net with a morphological processing layer. *Applied Sciences (Basel, Switzerland)*, 9(15), 3128. doi:10.3390/app9153128
- Zhuang, Q., Gan, S., & Zhang, L. (2022). Human-computer interaction based health diagnostics using ResNet34 for tongue image classification. *Computer Methods and Programs in Biomedicine*, 107096. Advance online publication. doi:10.1016/j.cmpb.2022.107096 PMID:36191350
- Zhuo, L., Zhang, J., Dong, P., Zhao, Y., & Peng, B. (2014). An SA–GA–BP neural network-based color correction algorithm for TCM tongue images. *Neurocomputing*, 134, 111–116. doi:10.1016/j.neucom.2012.12.080