



Research Article

Deep Learning Based Tongue Prickles Detection in Traditional Chinese Medicine

Xinzhou Wang,^{1,2} Siyan Luo,^{3,4} Guihua Tian,⁵ Xiangrong Rao ,³ Bin He,¹ and Fuchun Sun ²

¹College of Electronic and Information Engineering, Tongji University, Shanghai 200092, China

²Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

³Guang'Anmen Hospital, China Academy of Chinese Medical Sciences, Beijing 100053, China

⁴Beijing University of Chinese Medicine, Beijing 100029, China

⁵Dongzhimen Hospital, Beijing University of Chinese Medicine, Beijing 100700, China

Correspondence should be addressed to Fuchun Sun; fcsun@tsinghua.edu.cn

Received 21 April 2022; Revised 8 August 2022; Accepted 26 August 2022; Published 22 September 2022

Academic Editor: Jianan Xia

Copyright © 2022 Xinzhou Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Tongue diagnosis is a convenient and noninvasive clinical practice of traditional Chinese medicine (TCM), having existed for thousands of years. Prickle, as an essential indicator in TCM, appears as a large number of red thorns protruding from the tongue. The term “prickly tongue” has been used to describe the flow of qi and blood in TCM and assess the conditions of disease as well as the health status of subhealthy people. Different location and density of prickles indicate different symptoms. As proved by modern medical research, the prickles originate in the fungiform papillae, which are enlarged and protrude to form spikes like awn. Prickle recognition, however, is subjective, burdensome, and susceptible to external factors. To solve this issue, an end-to-end prickle detection workflow based on deep learning is proposed. First, raw tongue images are fed into the Swin Transformer to remove interference information. Then, segmented tongues are partitioned into four areas: root, center, tip, and margin. We manually labeled the prickles on 224 tongue images with the assistance of an OpenCV spot detector. After training on the labeled dataset, the super-resolutionfaster-RCNN extracts advanced tongue features and predicts the bounding box of each single prickle. We show the synergy of deep learning and TCM by achieving a 92.42% recall, which is 2.52% higher than the previous work. This work provides a quantitative perspective for symptoms and disease diagnosis according to tongue characteristics. Furthermore, it is convenient to transfer this portable model to detect petechiae or tooth-marks on tongue images.

1. Introduction

Based on the clinical practice of doctors, traditional Chinese medicine has been developed for thousands of years and has achieved brilliant results both in the past and in modern times. Tongue diagnosis, as a role of vital importance in TCM clinical diagnosis, is a convenient and noninvasive method based on the health status information carried by the appearance of the tongue [1]. The chromatic features and morphological characteristics of the tongue, the number of prickles and the form of tongue coating reveal the pathological changes of internal organs, as shown in Figure 1 [2, 3]. There have been reports that prickles are associated with tumors, kidney disease, gastric

disease, and new crowns [4]. However, traditional tongue diagnosis is an empirical procedure that relies heavily on the personal experience and subjective judgment of TCM doctors. With the assistance of artificial intelligence (AI), tongue diagnosis will be objective and people without medical knowledge can give themselves a preliminary diagnosis of a health condition. In recent years, much effort has been spent on AI-based tongue diagnosis, especially in the field of tongue color recognition [5, 6], tongue shape analysis [7], cracks segmentation [8], thickness, and moisture of tongue coating classification [9, 10].

Prickle, also called red-pointe, appears as a large number of red thorns protruding from the tongue. It indicates blood heat or excess heat in the internal organs. The color and

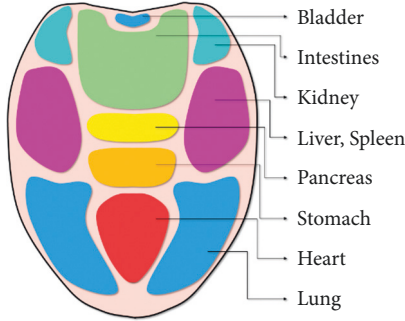


FIGURE 1: Tongue reflexology chart. Different areas on the tongue reflect the state of different organs.

number of the prickles can help estimate the flow of qi and blood. As proved by the study, the prickles originate in the fungiform papillae, which are enlarged and protrude to form spikes like awn [11]. Shang et al. further analyzed the association of prickles with the gastric sinus [12]. On the one hand, prickles mean increased blood flow and thus congestion. On the other hand, prickles represent thermal burns to blood vessels, resulting in blood spillage and mucosal erythema. The automatic detection of prickles can not only release the burden of doctors but also enable patients without medical knowledge to give themselves a brief examination.

Though AI-based tongue diagnosis has attracted a lot of attention, there is little literature on prickle detection due to its difficulty. In most tongue images, each prickle only occupies a few pixels and has little difference in tongue color under natural light. Moreover, the similarity between prickles and petechiae in both morphological and chromatic characteristics makes it a challenging task to distinguish them. Xu et al. were the first to introduce template feature matching to detect the prickles and petechiae, and then distinguished them based on RGB value range, gray average, and the position of the detected object [13]. Zhang employed the fuzzy C -means color cluster and noise reduction methods to detect prickles in the tongue edge image. Wang et al. used multistep threshold spot detection to detect prickles and petechiae. After extracting the features of spots (including prickles and petechiae), support vector machines and k -means were introduced to distinguish prickles and petechiae [11]. Wang et al. proposed a prickle detection method based on an auxiliary light source and a LOG operator edge detection method [14]. The last two works are most similar to our work, for they provided a quantitative description of prickles.

By eliminating background areas such as the face and lips, the tongue region segmentation can enhance the performance of downstream tasks, including prickle detection. Practice has proved that the neural network is very effective in the task of tongue segmentation. Zhang et al. introduced a DCNN-based tongue segmentation algorithm [15]. Wang et al. designed a coarse-to-fine segmentor based on RsNet and FsNet [16]. Jiang proposed an HSV enhanced CNN to segment the tongue region [17]. Zhang et al. combined superpixel with CNN to increase decoding performance [18].

Though previous researchers put much efforts into prickles and petechiae detection, the existing methods all

rely heavily on manual parameter tuning. This not only adds to the burden of researchers but also causes the model to overfit to specific circumstances and equipment. Moreover, there is no end-to-end prickle detection method, which could provide a quantitative description of prickles without manually segmenting the tongue raw images. Finally, most methods only took gray values of the exact pixels into consideration and lose the color information and the context information around the prickles.

The method proposed in this paper solved the question mentioned above from a completely new perspective: Deep Learning. We designed an end-to-end workflow to detect prickle automatically. The entire workflow and intermediate results are depicted in Figure 2.

2. Dataset and Methods

2.1. Dataset Collection. In this paper, the tongue images and segmentation annotations come from the bio-HIT tongue image dataset [19] (<https://github.com/BioHit/TongeImageDataset>). The tongue images dataset contains 300 RGB images with 576×768 pixels, and the images are obtained by the tongue image acquisition device shown in Figure 3. The device is designed as a semiclosed black box with a camera and illuminated on each side of the camera. The daylight illuminant D50, recommended by CIE (Commission Internationale de l'Éclairage) [20], was utilized as daylight illumination. According to the guidelines provided by CIE, the angle between the incident and outgoing rays is 45° . We elaborately screened out images with poor quality and got 224 images to train the model. The device has a closed image acquisition environment with an independent stable light source and a head restraint to ensure all the images are sampled to one standard. In addition, the image registration and calibration is not necessary either. Four volunteers in HIT elaborately annotated the image segmentation labels and the best one was chosen [19]. All the images were standardized to meet the standard normal distribution.

2.2. Tongue Segmentation. The AI-assisted tongue diagnosis is based on the information obtained from the tongue image. When concentrated on the tongue, irrelevant elements, including the lips, cheek, and chin, distract I neural network. With interference eliminated and the tongue matted, the contour line of the tongue becomes apparent and the performance of feature extraction is guaranteed. Therefore, it is necessary to segment the tongue before the next step, and we introduced the Swin Transformer [21] as the segmentor. The core concept of the Swin Transformer is self-attention, as shown in equations:

$$\begin{aligned} Q &= XW^Q, \\ K &= XW^K, \\ V &= XW^V, \end{aligned} \quad (1)$$

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T V}{\sqrt{d_k}}\right), \quad (2)$$

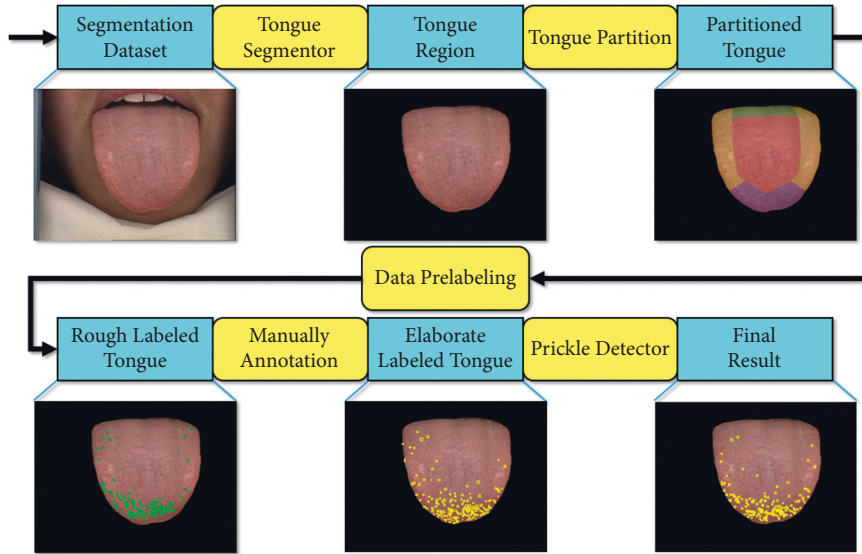


FIGURE 2: Prickle detection workflow. Blue rectangles represent images, and yellow rectangles represent image processing. First, in tongue segmentor, we introduced Swin Transformer, a state-of-the-art computer vision segmentation neural network, to mat the tongue region out of the raw picture. Second, in tongue partition, the tongue is partitioned into four areas: root, margin, tip, and center. Third, in data prelabeling, a spot detector is applied based on tongue areas. Fourth, elaborate manual annotation is conducted with the help of TCM doctors. Fifth, to fully embody the advantages of the neural network, a super-resolutionfaster-RCNN based detector is deployed to detect the prickles from a matted tongue image.

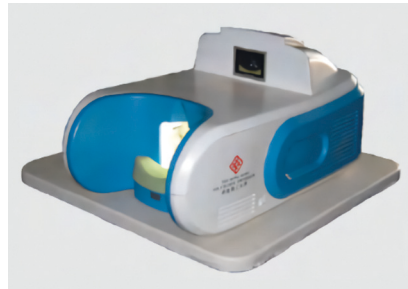


FIGURE 3: The tongue image acquisition device. The device is designed to obtain tongue images in uniform illumination and facial poses.

where the Q is the query vector, the K is the key vector, and the V is the value vector. W^Q, W^K, W^V are all weight matrices. With the self-attention mechanism, the network can perceive global semantic information. The entire architecture of Swin Transformer is shown in Figure 4, where $Z_0, Z_2, Z_4, Z_6, Z_8,$ and Z_{10} used multihead self-attention with regular windowing (W-MSA) and the others used multihead self-attention with shifted windowing (SW-MSA). Considering the fact that the dataset only contains 224 images, which is insufficient for training a network from scratch, we adopted a paradigmatic strategy in computer vision: pretraining and fine-tuning with data augmentation. Microsoft has trained the model with over 20,000 images on the ADE20K dataset [22], and we fine-tuned the model on the tongue segmentation dataset. The data augmentation pipeline includes flipping, cropping, and photometric distortion. Photometric distortion applies the following transformations with a probability of 0.5: random brightness, random contrast, color space converting, random saturation, random hue, and randomly swapping channels. With data augmentation, the

model will be more robust when the illumination or sampling device varies. Though the neural network is able to classify each pixel as tongue or background, there is no guarantee that the segmented tongue region has structural integrity. To address this issue, we analyzed the connected components of each mask, filled the blank areas in the tongue and eliminated the outliers using two-pass connected component analysis [23]. The algorithm is shown in Algorithm 1 and the result is shown in Figure 5.

2.3. Prickles Annotation. Usually, hundreds of prickles appear on the tongue in groups. Therefore, it will be a challenging task if we manually annotate the whole dataset. In this paper, we employed partition spots detection [11] with LAB chromatic aberration filtering to give a primitive annotation of the prickles.

The LAB color space is based on the human eye's perception of color and can represent all colors the human eye can perceive. "L" represents lightness, "A" represents red-

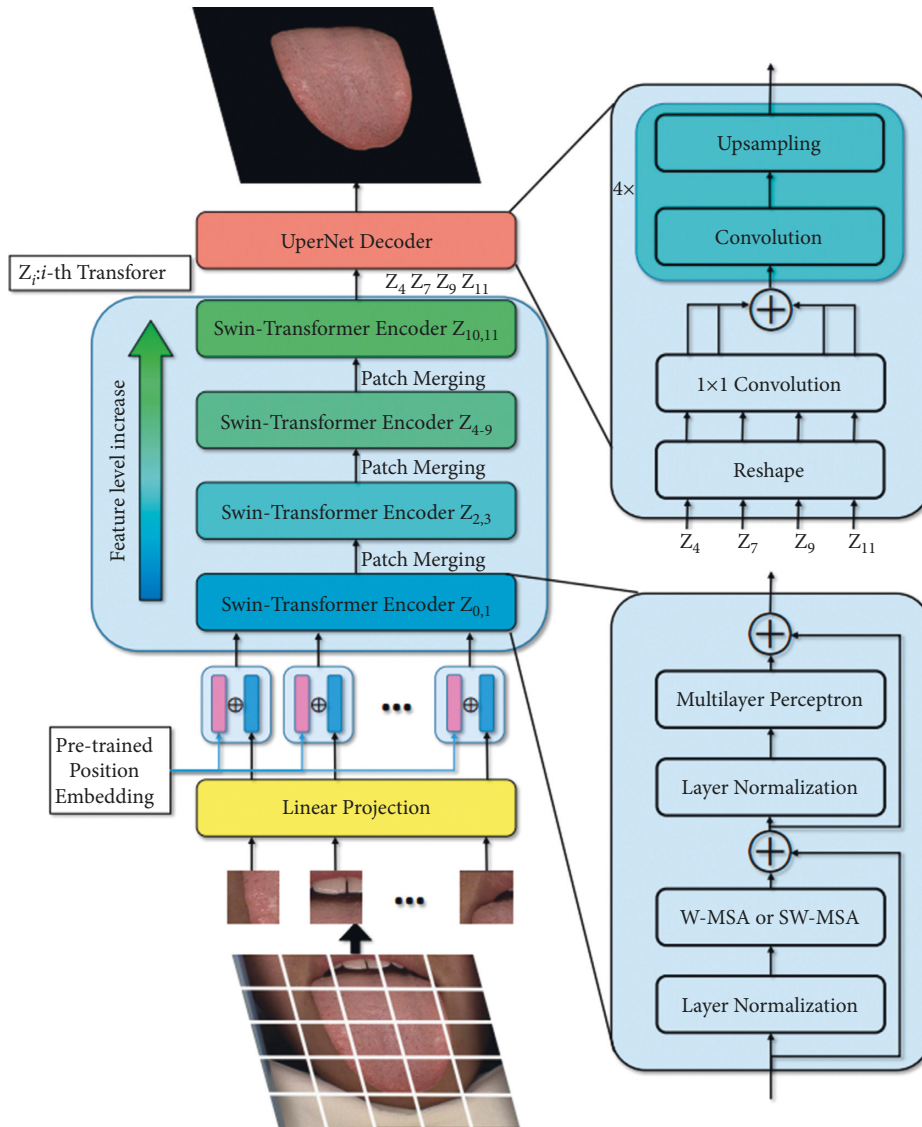


FIGURE 4: Architecture of tongue segmentor. The tongue image is divided into several patches and then added with position embeddings to retain spatial information. The encoder consists of cascaded Swin Transformers, while UperNet decoder is applied to aggregate multilevel features from encoder.



FIGURE 5: Mask morphology processing. The connected component with the largest area in black or white will be marked as “1” and other connected components will be eliminated.

green difference, and “B” represents blue-yellow difference. The spot detection algorithm works on gray images and the color information is lost. Therefore, we filtered the spots by

the chromatic aberration between the spots and the manually picked prickles. Given an RGB value, an approximate LAB chromatic aberration can be calculated as follows:

```

Input: Binary tongue mask  $I$  width  $W$  and height  $H$ 
Output: Tongue mask with structural integrity
 $num\_label = 0$ 
for  $y = 0$  to  $H - 1$  do
  for  $x = 0$  to  $W - 1$  do
     $N = \text{neighbours}(I[y, x])$ 
    if  $I[y, x] = 1$  then
      if  $0 \text{ not in } N$  then
         $I[y, x] = num\_label$  // Give each mask pixel a label
         $num\_label += 1$ 
      else if  $\text{has\_mask}(N)$  then
         $label = \min(N)$ 
         $I[y, x] = label$  // If neighbours have label, use the minimum one.
         $labelSet[label].append(N.all\_labels())$ 
  for  $y = 0$  to  $H - 1$  do
    for  $x = 0$  to  $W - 1$  do
       $I(y, x) = \min(labelSet[I(y, x)])$  // Unify the label of each component
  Select the mask component with largest area as tongue and discard others
  Select the background component with largest area and discard others // Fill holes in mask
Return  $I$ 

```

ALGORITHM 1: Mask morphology processing.

$$\tilde{r} = C_{1,R} + \frac{C_{2,R}}{2},$$

$$\Delta R = C_{1,R} - C_{2,R},$$

$$\Delta G = C_{1,G} - C_{2,G},$$

$$\Delta B = C_{1,B} - C_{2,B},$$

(3)

$$\text{Chromatic Aberration} = \sqrt{\left(2 + \frac{\tilde{r}}{256}\right) \times \Delta R^2 + 4 \times \Delta G^2 + \left(2 + \frac{(255 - \tilde{r})}{256}\right) \times \Delta B^2}.$$

The tongue is partitioned before detection so that we can elaborately set different detection parameters for different areas. The tongue coating is distributed on the tongue surface, which is usually slightly thicker in the center or root of the tongue, and the prickles covered by the coating have different characteristics from the prickles on the margin and tip. In addition, the cracks on the root and center of the tongue tend to be detected mistakenly by the spot detection algorithm. To solve this problem, we divided the tongue into four areas: root, margin, tip, and center before preliminary annotation. Then we set the threshold of chromatic aberration, area, circularity, and convexity tighter in the root and center than in other areas. This setting avoids the mis-detection of cracks while maintaining the detection rate.

The algorithm of annotation is shown in Figure 6. First, the parallel-line method is introduced to build a reference line for tongue partition [14]. Second, the tongue is divided into four areas by the relative thickness compared to the overall scale of the tongue. The result is depicted in Figure 7. Third, a simple blob detector based on OpenCV (Open Source Computer Vision Library) is deployed with different parameters in different tongue regions. Fourth, the detected

spots are filtered by LAB chromatic aberration. Finally, a professional TCM doctor revised the roughly annotated bounding boxes with the MIT Labelme annotation tool (<https://labelme.csail.mit.edu>) and two other TCM doctors checked the result under the same diagnostic criteria on the same monitor [24, 25].

2.4. Prickle Detection. In this paper, we take the Faster-RCNN as the prickle detector. In 2016, Ross B. Girshick proposed a new object detection neural network called Faster-RCNN, which is depicted in Figure 8. The Faster-RCNN first uses a set of basic convolution layers, ReLU function, and pooling layers to extract an image feature map, which is subsequently shared by region proposal networks (RPN) and fully connected layers. The RPN network is designed to generate region proposals with the softmax layer to determine whether the anchors are positive or negative, and then it employs bounding box regression to correct the anchors to obtain accurate proposals. The roi-pooling layer collects the input feature maps and proposals, extracts proposal feature maps after synthesizing the information,

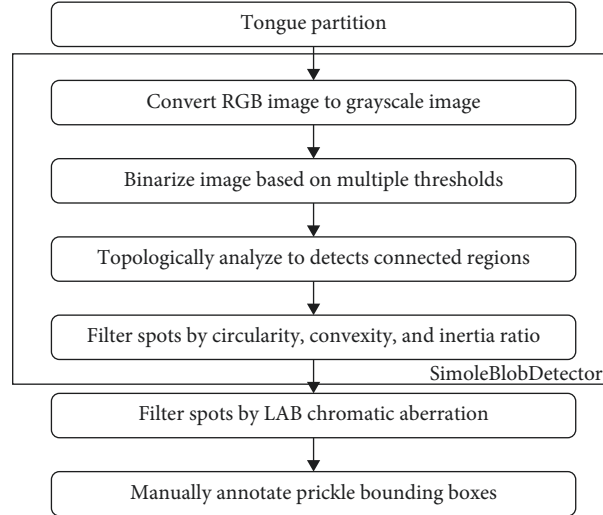


FIGURE 6: Prickle bounding boxes annotation workflow. Prickle bounding boxes are labeled automatically and then manually adjusted by TCM doctors.

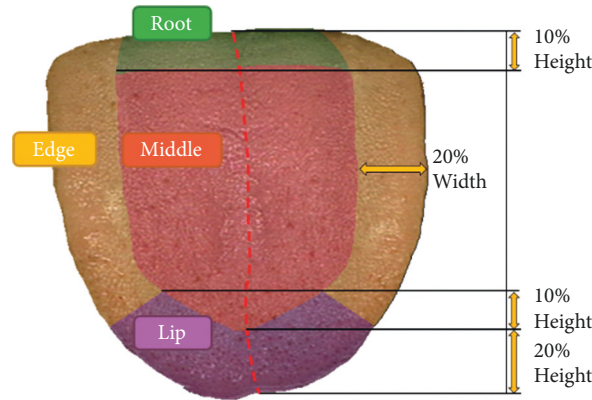


FIGURE 7: Tongue partition. The tongue is divided into four areas automatically based on the midline and the detection parameters vary with areas adaptively.

and sends them to the subsequent fully connected layer to determine the target category [26].

There are two obstacles that need to be overcome before training the Faster-RCNN. First, insufficient data makes the model hard to train. Therefore, we introduced data augmentation, including flip, crop, scale, translation, and rotation, as shown in Figure 9. In addition, we followed the paradigm of computer vision workflow: we pretrained the model on a general detection dataset and used transfer learning to fine-tune the model on the prickle detection dataset. Second, the Faster-RCNN is designed for the detection of the target on a normal scale. When the target is smaller than 32×32 pixels, the performance of the model will decrease sharply. To address this issue, we used $4 \times$ bilinear interpolation for upsampling to improve the model performance. Since the neural network aims to build an end-to-end prickle detection model and it is proven that the color calibration and irrelevant noise filtering may cause a degradation in model performance [27], image registration and filtering are not employed.

2.5. Evaluation Metrics. The segmentation tasks in computer vision field could be considered as a pixel-wise classification, and there are four types of the results: true positive (TP), false positive (FP), true negative (TN), and false negative (FN), as shown in Figure 10. The standard evaluation metrics of segmentation task is intersection over union (IoU), defined as follows:

$$\text{IoU} = \frac{\text{TP}}{(\text{FP} + \text{FN} + \text{TP})}, \quad (4)$$

where TP, FP, and FN are about pixel-wise classification results. We also employed precision and accuracy as metrics to fully demonstrate the performance of the proposed method, which are defined as follows:

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})}, \quad (5)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}.$$

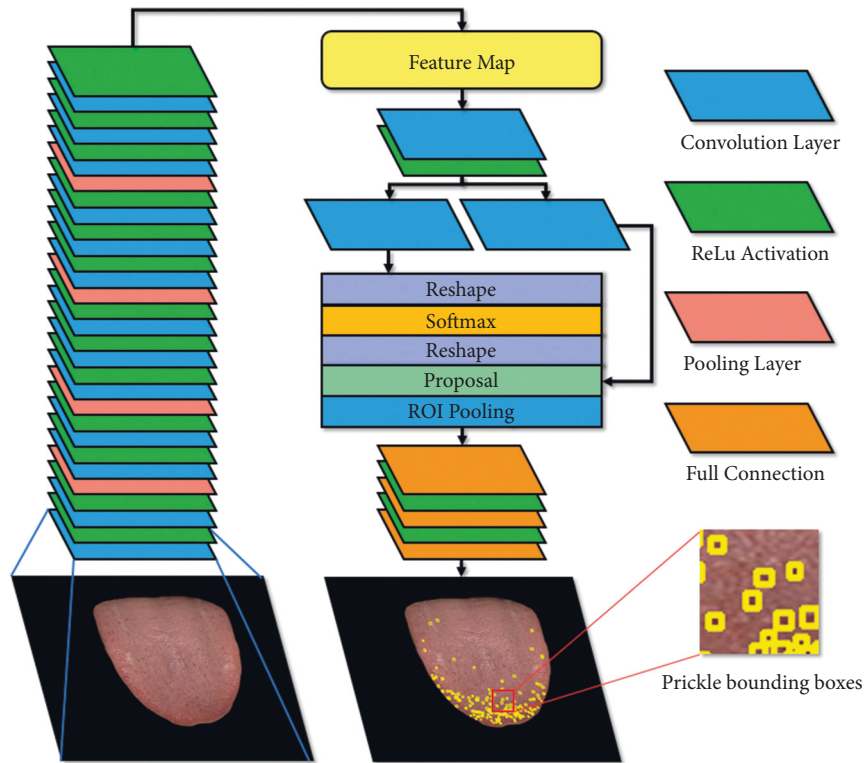


FIGURE 8: Architecture of prickle detector. CNN encoder extract features from images. Region proposal networks generate region proposals and bounding box regressor modify anchors to predict precise prickle bounding boxes.

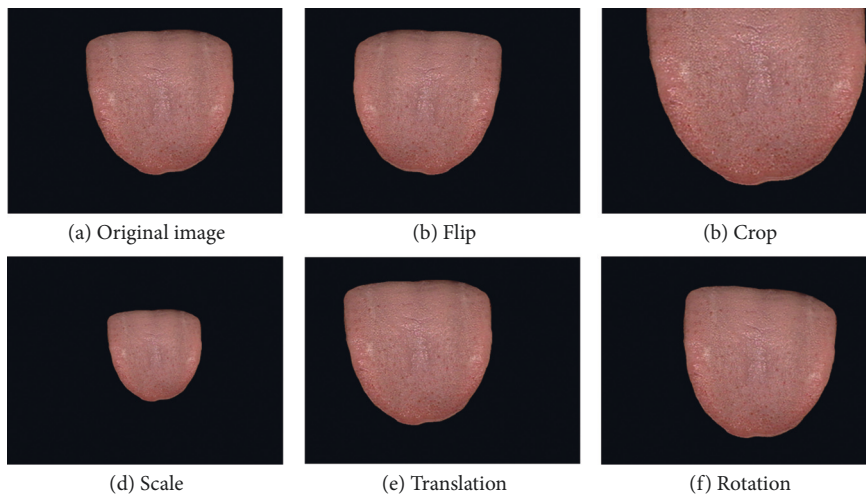


FIGURE 9: Data augmentation. Super-resolution, flip, crop, scale, translation, and rotation are conducted to increase dataset size and improve network robustness without collecting new data. (a) Original image (b) flip (c) crop (d) scale (e) translation and (f) rotation.

A prickle is classified as detected correctly when the bounding box has IoU over the threshold.

The metrics for prickles detection are precision defined above and recall defined as follows:

$$\text{Recall} = \frac{TP}{(TP + FN)}. \tag{6}$$

Where FP is the number of misdetections, TP and FN is the number of manually annotated prickles that were detected and undetected, respectively.

3. Experiments and Results

3.1. Experiment Setting. In our work, the models were trained and tested based on the Python deep-learning

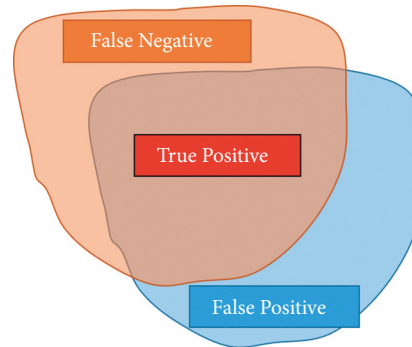


FIGURE 10: Types of the predicted segmentation. The orange area represents the ground truth, and the blue represents the predicted segmentation.

framework PyTorch (<https://pytorch.org>) and the computing platform is a Linux server with Intel Xeon (R) E5-2620 CPU, 4 NVIDIA RTX2080Ti GPUs, and 128 GB memory. The total training time is about 2 hours. The parameters of segmentor and detector in the training stage are shown in Table 1.

3.2. Result of Tongue Segmentation. The Swin Transformer is a flexible neural network, which means it requires more training compared to the conventional CNN. To address this issue, we introduced the pretraining model provided by Microsoft (<https://github.com/SwinTransformer/Swin-Transformer-Semantic-Segmentation>). Microsoft has trained the model with over 20,000 images on ADE20K dataset [16] and we fine-tuned the model on our tongue segmentation dataset. The training set and test set contained 178 and 46 annotated images, respectively. The splitting of the dataset took a cross-validation strategy. After we got the predicted segmentation, we filled the blank areas in the tongue and eliminated the outliers. The result is shown in Table 2 and Figure 11(b).

The excellent performance of the Swin Transformer makes the segmentation results basically the same as manual segmentation. Compared to the conventional machine vision segmentation methods, neural network is an end-to-end progress without the requirement of manual tuning parameters. It is also more robust when the illumination or sampling device varies and the result proves the superiority of the transformer architecture in tongue segmentation.

3.3. Prickle Labeling. We employed the simple blob detector with LAB chromatic aberration filtering to detect the prickles coarsely. The simple blob detector is a multistep threshold spot detection method for processing gray images. We took the parameters of the simple blob detector in [11] as the initial value and introduced the grid search to find the optimal parameters for each area. The searching step length is set to 10% of the initial value and the searching range is set from 50% to 150% of the initial value. The parameters of the partitioned simple blob detector are shown in Table 3.

Petechiae are usually found on the center and roots of the tongue. To reduce the probability of misdetection, the

TABLE 1: Parameters for training Swin Transformer.

Hyper-parameter	Segmentor	Detector
Epoch	20	100
Batch size	4	4
Optimizer	AdamW	SGD
Learning rate	6e-05	5e-2
Learning rate policy	Polynomial	Step
Weight decay	1e-2	1e-4
Loss function	Cross entropy	Cross entropy

constraints of spots in those areas are more stringent. In addition, to take full advantage of the color information, we sampled the RGB value of prickles in different tongues and areas and filtered the detected spots by the LAB chromatic aberration. The automatic annotation result with and without partitioning is shown in Figure 12 to give a clear demonstration of how the partitioning helps the prickle labeling. After automatic annotation, three well-trained TCM doctors manually revised the labels. The annotation is shown in the Figure 11(d).

3.4. Result of Prickle Detection. Object detection is a challenging task requiring a lot of training. We downloaded the pretrained model provided by CUHK and SenseTime (<https://github.com/open-mmlab>) and fine-tuned it on the prickle detection dataset. The training set and test set contained 178 and 46 annotated images, respectively. The splitting of the dataset took a cross-validation strategy. Compared to the original Faster-RCNN, we employed a 4× bilinear interpolation for upsampling and modified the anchor size to match the prickle detection. To demonstrate the superiority of our modified Faster-RCNN, we took the vanilla Faster-RCNN and other detection algorithms for comparison. The predicted result of the entire dataset is shown in Table 4.

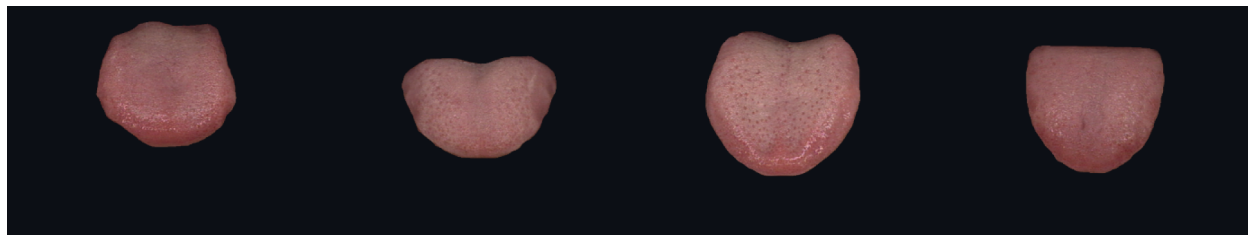
The recall of our method outperformed the existing methods without fine-tuning the parameters manually. We did not choose YOLO because it is a one-stage detection algorithm and the resolution is fixed, which limits the performance of detecting tiny targets. It is worth mentioning that we tried other learning-based algorithms, including DCNv2 [29] and SSD [30], but they were unable to predict

TABLE 2: The performance of segmentor and previous studies.

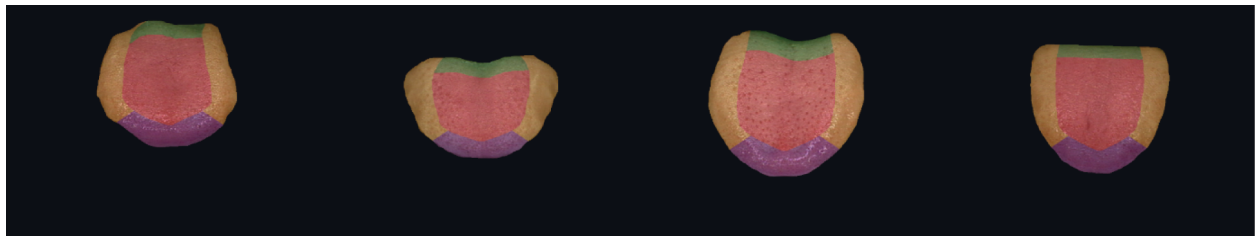
Method	IoU (%)	Precision (%)	Accuracy (%)
DCNN [8]	—	97.94	99.41
RsNet and FsNet [9]	—	97.85	99.04
HSV enhanced CNN [10]	—	94.70	97.88
Ours	99.08	99.47	99.79



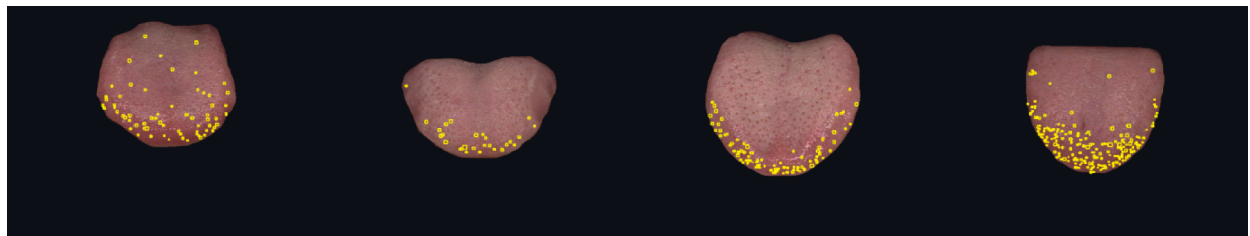
(a)



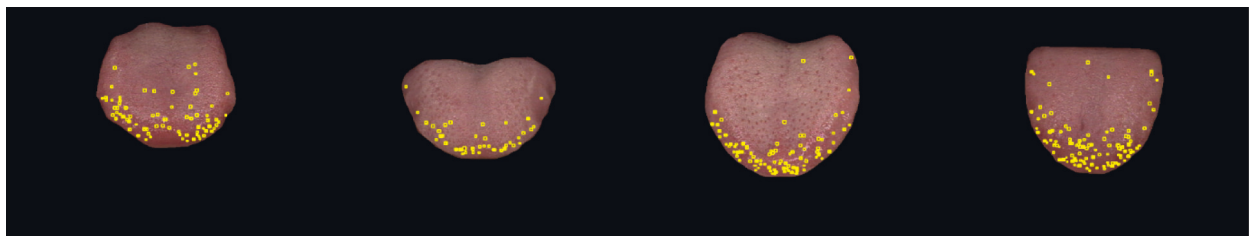
(b)



(c)

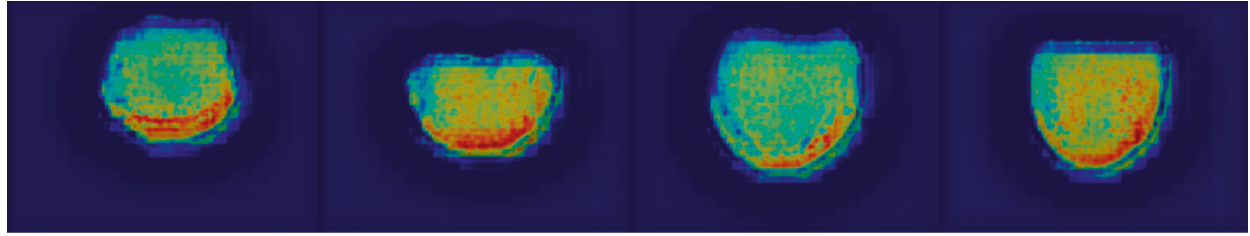


(d)



(e)

FIGURE 11: Continued.

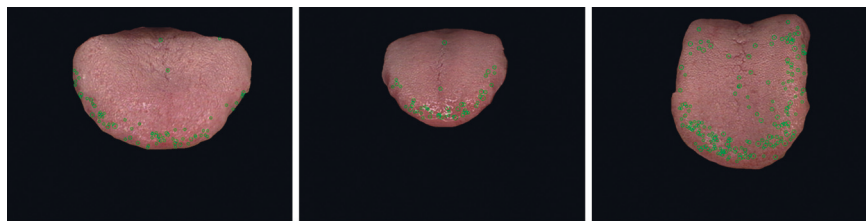


(f)

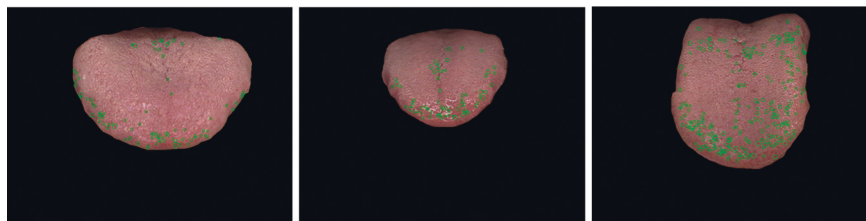
FIGURE 11: Prickle detection result. (a) Original images (b) segmentation result (c) partitioned tongues (d) prickle labels (e) prickle detection result and (f) neural network attention map. The intermediate result of the whole workflow is shown (zoom in for better viewing).

TABLE 3: Parameters of the simple blob detector in different areas.

Parameter	Root and center	Tip	Margin
Min threshold	60	60	60
Max threshold	100	100	100
Threshold step	2	2	2
Min repeatability	8	4	4
Min area	4	2	2
Max area	25	40	40
Min circularity	0.8	0.4	0.4
Min convexity	0.8	0.4	0.4
Min inertia ratio	0.5	0.4	0.4
Max aberration	85	100	100



(a)



(b)

FIGURE 12: Automatic annotation results with and without partitioning. (a) Annotation with partitioning. (b) Annotation without partitioning. With annotation parameters differing between areas, the misdetections of center cracks are reduced.

TABLE 4: Prickle detection results and comparison with previous studies.

Method	Recall			Total (%)	Accuracy Total (%)
	Root & center (%)	Tip (%)	Margin (%)		
LOG [7]	84.97	67.37	84.23	86.95	—
SVM [5]	—	—	—	89.90	—
YOLO [28]	73.28	67.68	65.48	67.95	60.44
Vanilla faster-RCNN [26]	69.43	66.48	61.22	65.58	72.13
Ours w/o segmentation	88.89	89.45	83.34	89.77	86.69
Ours	98.62	89.92	94.59	92.42	88.46

any bounding box. As shown in Figure 11(e), the detected prickle patterns are similar to the annotation while it tends to detect prickles on the tip. To illustrate the principle of our neural network, we provided an attention heat map of the detector, as shown in Figure 11(f). The neural network had more attention on the edge of the tongue, where the prickles are most likely to appear. This attention is learned by the neural network automatically and has the potential to provide TCM doctors an intuition to distinguish a tongue with prickles. In addition, we fed the raw images into the detector as an ablation experiment to validate the necessity of segmentation, and the result is shown in the 3rd line of Table 4.

4. Discussion

Prickle, as an essential syndrome feature of TCM, can be used to assist in the diagnosis and treatment of subhealthy people. But the ambiguity of prickle recognition and the subjective preferences of doctors limit its further application. Combined with modern AI technology, we proposed an end-to-end computer vision-based prickle detection workflow, which makes prickle detection more precise and objective. This workflow is divided into three main steps. Firstly, Swin Transformer, a state-of-the-art semantic segmentation neural network, was employed to segment the tongue region out of a raw image. The segmented tongues were partitioned into four areas: root, center, tip, and margin to help the parameters setting of follow-up spot detection and prickle detection. Secondly, we manually labeled the prickles on 224 tongue images with the help of a spot detector and fed the result into the Faster-RCNN. Finally, the neural network extracted the features of images at both a texture level and morphological level to detect the prickles on the tongues. The precision of our segmentation is 99.47%, and the recall of our detection is 92.42%, which outperformed the existing methods. The result illustrates that the utilization of transfer learning made it possible to train neural networks on a limited number of images. Compared to previously published studies, it gets rid of the trivial parameters tuning procedure and releases the burden on researchers. Meanwhile, the workflow proposed is more portable, which means you can transfer the model to arbitrary tongue characteristics or image acquisition equipment.

In the context of artificial intelligence and big data, the informatization of TCM diagnosis is an area that urgently needs in-depth research. Our work took full advantage of the deep learning algorithm to implement an intelligent recognition of prickles and provided the possibility of establishing the quantitative association between the tongue image and clinical symptoms. In addition, incorporating a prickle detection model into the smartphone will allow people without medical knowledge to give themselves a simple health status assessment, and a quantitative and objective tongue diagnosis will also benefit the integration of TCM and modern Western medicine. The precise segmentation and feature recognition of the tongue can also be used for throat swab robot perception.

Though our method is state-of-the-art, there are some aspects for further research. Firstly, the model was trained on the images sampled by a standard acquisition device, which limits its generalization and robustness. A larger dataset consisting of images from different devices would help the model to establish a greater degree of accuracy in this matter. Secondly, our model provides a paradigm for tongue feature detection. With petechiae, cracks, toothmarks, and other TCM features of the tongues labeled, the model has the potential to achieve an acceptable result. Thirdly, most existing learning-based tongue feature detection methods aim to find bounding boxes. It is possible to use a segmentor to classify every single pixel in the image into a kind of TCM tongue feature.

Data Availability

The experimental data and code used to support the findings of this study will be available on <https://github.com/zz7379/PrickleDetection> or on contacting the authors with reasonable request (wangxinzhou@tongji.edu.cn).

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Authors' Contributions

Xinzhou Wang and Siyan Luo contributed equally to this work.

Acknowledgments

The research in the paper is jointly funded by the Sino-German Collaborative Research Project Crossmodal Learning (NSFC 62061136001/DFG TRR169), Key Technologies on a Pharyngeal Swab Sampling Robot using Human-Machine Fusion Based on the Prey Mechanism of the Tip of the Chameleon Tongue and its Demonstration Applications (No. 2020GQG0006), and National Science Foundation No. 81973683.

References

- [1] L. Zhi, D. Zhang, J. Q. Yan, Q. L. Li, and Q. L. Tang, "Classification of hyperspectral medical tongue images for tongue diagnosis," *Computerized Medical Imaging and Graphics*, vol. 31, no. 8, pp. 672–678, 2007.
- [2] B. Zhang, H. Zhang, and G. Li, "Significant geometry features in tongue image analysis," *Evidence-based Complementary and Alternative Medicine*, vol. 2015, Article ID 897580, 8 pages, 2015.
- [3] E. Vocaturo and E. Zumpano, "Machine learning opportunities for automatic tongue diagnosis systems," in *Proceedings of the 2020 IEEE International Conference On Bioinformatics and Biomedicine (BIBM)*, pp. 1498–1502, IEEE, Seoul, South Korea, 2020.
- [4] S. Han, X. Yang, Q. Qi et al., "Potential screening and early diagnosis method for cancer: tongue diagnosis," *International Journal of Oncology*, vol. 48, no. 6, pp. 2257–2264, 2016.

- [5] B. Zhang, X. Wang, J. You, and D. Zhang, "Tongue color analysis for medical application," *Evidence-Based Complementary and Alternative Medicine*, vol. 2013, Article ID 264742, 11 pages, 2013.
- [6] J. Hou, H. Su, B. Yan, H. Zheng, Z. Sun, and X.-C. Cai, "Classification of tongue color based on cnn," in *Proceedings of the 2017 IEEE 2nd International Conference On Big Data Analysis*, IEEE, Beijing, China, 2017.
- [7] B. Huang, J. Wu, D. Zhang, and N. Li, "Tongue shape classification by geometric features," *Information Sciences*, vol. 180, no. 2, pp. 312–324, 2010.
- [8] Z. Shi and C. Zhou, "Fissure extraction and analysis of image of tongue," *Computer Technology and Development*, vol. 1, no. 5, pp. 245–248, 2007.
- [9] J. Kim, G. J. Han, B. H. Choi et al., "Development of differential criteria on tongue coating thickness in tongue diagnosis," *Complementary Therapies in Medicine*, vol. 20, no. 5, pp. 316–322, 2012.
- [10] D. Meng, G. Cao, Y. Duan et al., "Tongue images classification based on constrained high dispersal network," *Evidence-Based Complementary and Alternative Medicine*, vol. 2017, Article ID 7452427, 12 pages, 2017.
- [11] S. Wang, K. Liu, and L. Wang, "Tongue spots and petechiae recognition and extraction in tongue diagnosis images," *Computer Engineering and Science*, vol. 39, no. 6, pp. 1126–1132, 2017.
- [12] Z. Shang, Z.-G. Du, B. Guan et al., "Correlation analysis between characteristics under gastroscop and image information of tongue in patients with chronic gastritis," *Journal of Traditional Chinese Medicine*, vol. 42, no. 1, pp. 102–107, 2022.
- [13] J. Xu, Y. Sun, Z. Zhang, Y. Bao, and W. Li, "Recognition of acantha and ecchymosis in tongue pattern," *Journal of Shanghai University of Traditional Chinese Medicine*, vol. 3, no. 4, pp. 38–40, 2004.
- [14] X. Wang, R. Wang, D. Guo, X. Lu, and P. Zhou, "A research about tongue-prickled recognition method based on auxiliary light source," *Chinese Journal of Sensors and Actuators*, vol. 29, no. 10, pp. 1553–1559, 2016.
- [15] X. Zhang, Y. Guo, Y. Cai, and M. Sun, "Tongue image segmentation algorithm based on deep convolutional neural network and fully conditional random fields," *Journal of Beijing University of Aeronautics and Astronautics*, vol. 45, no. 12, pp. 2364–2374, 2019.
- [16] L. Wang, Y. Tang, P. Chen, X. He, and G. Yuan, "Segmentation, two-phase convolutional neural network," *Journal of Image and Graphics*, vol. 23, no. 10, pp. 1571–1581, 2018.
- [17] J. Li, B. Xu, X. Ban, P. Tai, and B. Ma, "A tongue image segmentation method based on enhanced hsv convolutional neural network," *International Conference on Cooperative Design, Visualization and Engineering*, vol. 1, pp. 252–260, 2017.
- [18] H. Zhang, R. Jiang, T. Yang, J. Gao, Y. Wang, and J. Zhang, "Study on TCM tongue image segmentation model based on convolutional neural network fused with superpixel," *Evidence-Based Complementary and Alternative Medicine*, vol. 2022, Article ID 3943920, 12 pages, 2022.
- [19] L. Wu, X. Luo, and Y. Xu, "Using convolutional neural network for diabetes mellitus diagnosis based on tongue images," *Journal of Engineering*, vol. 13, pp. 635–638, 2020.
- [20] G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods Quantitative Data and Formulae*, Wiley, New York, NY, USA, 2000.
- [21] Z. Liu, Y. Lin, Y. Cao et al., "Swin transformer: hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10012–10022, IEEE, Seoul, Korea, 2021.
- [22] B. Zhou, H. Zhao, X. Puig et al., "Semantic understanding of scenes through the ade20k dataset," *International Journal of Computer Vision*, vol. 127, no. 3, pp. 302–321, 2019.
- [23] D. G. Bailey and M. J. Klaiber, "Zig-zag based single-pass connected components analysis," *Journal of imaging*, vol. 5, no. 4, p. 45, 2019.
- [24] X. Wang, J. Liu, C. Wu et al., "Artificial intelligence in tongue diagnosis: using deep convolutional neural network for recognizing unhealthy tongue with tooth-mark," *Computational and Structural Biotechnology Journal*, vol. 18, pp. 973–980, 2020.
- [25] Z. Qi, L. Tu, Z. Luo et al., "Tongue image database construction based on the expert opinions: assessment for individual agreement and methods for expert selection," *Evidence-Based Complementary and Alternative Medicine*, vol. 2018, Article ID 8491057, 9 pages, 2018.
- [26] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 1, no. 28, pp. 91–99, 2015.
- [27] S. Hosseinzadeh Kassani and P. Hosseinzadeh Kassani, "A comparative study of deep learning architectures on melanoma detection," *Tissue and Cell*, vol. 58, pp. 76–83, 2019.
- [28] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788, Las Vegas, NV, USA, 2016.
- [29] X. Zhu, H. Hu, S. Lin, and J. Dai, "Deformable convnets v2: more deformable, better results," in *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Long Beach, CA, USA, 2019.
- [30] W. Liu, D. Anguelov, D. Erhan et al., "SSD: single shot multibox detector," in *European Conference on Computer Vision*, pp. 21–37, Springer, Berlin, Germany, 2016.
- [31] L.-C. Lo, T.-L. Cheng, J. Y. Chiang, and N. Damdinsuren, "Breast cancer index: a perspective on tongue diagnosis in traditional Chinese medicine," *Journal of Traditional and Complementary Medicine*, vol. 3, no. 3, pp. 194–203, 2013.
- [32] P.-C. Hsu, H.-K. Wu, H.-H. Chang, J.-M. Chen, J. Y. Chiang, and L.-C. Lo, "A perspective on tongue diagnosis in patients with breast cancer," *Evidence-Based Complementary and Alternative Medicine*, vol. 2021, Article ID 4441192, 9 pages, 2021.
- [33] C.-J. Chung, C.-H. Wu, W.-L. Hu, C. H. Shih, Y. N. Liao, and Y. C. Hung, "Tongue diagnosis index of chronic kidney disease," *Biomedical Journal*, 2022.
- [34] W. Pang, D. Zhang, J. Zhang et al., "Tongue features of patients with coronavirus disease 2019: a retrospective cross-sectional study," *Integrative medicine research*, vol. 9, no. 3, Article ID 100493, 2020.