

DOI:10.3969/j.issn.1005-202X.2021.04.023

医学人工智能

# 基于Mask Scoring R-CNN的齿痕舌象识别

芮迎迎,孔祥勇,刘亚楠,董鑫,蔡健,卢严砖,况忠伶  
上海理工大学医疗器械与食品学院,上海 200093

**【摘要】目的:**提出一种基于Mask Scoring R-CNN和迁移学习的舌象特征识别方法。**方法:**首先使用CNN提取特征,使用ResNet-101和特征金字塔网络(FPN)的主干网络,可以从低层次和高层次的网络中提取特征,根据不同比例绘制金字塔特征的级别。接着使用区域生成网络将从主干网络中提取的特征生成候选感兴趣区域(ROI)。最后为每个ROI检测并分割齿痕。**结果:**在232例样本的测试集上进行测试, $F_1$ 分数为0.95,准确率为0.93,精确率为0.99,召回率为0.914。**结论:**该方法能够在小样本舌象数据集上有效识别齿痕特征、准确定位齿痕位置、标定齿痕大小、提取齿痕个数,该方法具有良好的有效性、通用性、泛化性,能够为后续齿痕严重程度分析提供依据。同时为疾病预防、移动医疗保健或从生物信息学角度跟踪疾病进展提供客观、方便的计算机辅助舌诊方法。

**【关键词】**Mask Scoring R-CNN;深度学习;迁移学习;齿痕舌;舌象分类

**【中图分类号】**R318                      **【文献标志码】**A                      **【文章编号】**1005-202X(2021)04-0523-06

## Tooth-marked tongue recognition using Mask Scoring R-CNN

RUI Yingying, KONG Xiangyong, LIU Ya'nan, DONG Xin, CAI Jian, LU Yanzhuan, KUANG Zhongling  
School of Medical Instrument and Food Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

**Abstract: Objective** To identify tongue features based on Mask Scoring R-CNN and transfer learning. **Methods** After the features were extracted by convolutional neural network, the backbone networks of ResNet-101 and feature pyramid network were used to extract features from low-level and high-level networks, and the levels of pyramid features were plotted according to different proportions. Region generation network was then used to generate candidate regions of interest from the features extracted from the backbone network. Finally, tooth marks in each region of interest were detected and segmented. **Results** The test on the test set with 232 samples showed that  $F_1$  score was 0.95, and that the accuracy rate, precision rate and recall rate of the proposed method were 0.93, 0.99 and 0.914, respectively. **Conclusion** Using the proposed method can effectively identify the features of tooth marks, accurately locate the position of tooth marks, calibrate the size of tooth marks, and extract the number of tooth marks on small-sample tongue image data set. The proposed method which has good effectiveness, generality and generality provides a basis for the severity analysis of tooth marks and serves as an objective and convenient computer-assisted tongue diagnosis method for monitoring disease progression from the perspective of disease prevention, mobile healthcare or bioinformatics.

**Keywords:** Mask Scoring R-CNN; deep learning; transfer learning; tooth-marked tongue; tongue classification

### 前言

移动健康技术在中医和远程医疗中的应用呈增长趋势,尤其是在新型冠状病毒肺炎(COVID-19)爆发期间。

舌诊是中医诊断中的重要环节,因其无接触、有效、无创、实时、便捷、成本低等西医检测无法比拟的优势,能在任何时间、任何地点进行辅助诊断,满足全球初级保健系统的需求。舌象能够反映内脏虚实、气血盛衰、病位深浅、预后好坏<sup>[1-2]</sup>。通过对舌象的客观化研究,能够观察COVID-19患者的舌象特征及变化,为防治COVID-19提供治疗依据<sup>[3-4]</sup>。由于传统舌诊主要依据医师个人经验,易受环境影响,存在主观性、评判标准不一致等问题。因此,越来越多的计算机研究人员开始将其结合计算机视觉领域,建立客观、定量的中医辅助舌诊系统。

**【收稿日期】**2020-12-21  
**【基金项目】**国家自然科学基金青年科学基金(61906121)  
**【作者简介】**芮迎迎,在读硕士,研究方向:医疗人工智能,E-mail: yinying\_rui@163.com  
**【通信作者】**孔祥勇,硕士,讲师,研究方向:医疗人工智能,E-mail: kxy-@usst.edu.cn

齿痕舌的客观化研究具有重要性和挑战性,齿痕舌是中医诊断的一个重要指标。临床实验结果表明在不同临床分型COVID-19患者的舌形分布研究中,齿痕舌易见于普通型患者。据研究,人群中的齿痕舌占比56%,其中严重者占比11%<sup>[5]</sup>。齿痕舌的识别更具挑战性,主要原因为:(1)齿痕形状多样,变化大,如不同颜色、不同形状、不同类型的齿痕;(2)由于个人隐私和图像获取的限制,舌部的图像数量有限;(3)用于金标准的图像数量有限。舌象的判读需要多个专家的参与,但专家在有限时间内只能将图片分类为有齿痕和无齿痕,并未在图片上标注出齿痕的位置,无法提供齿痕的位置分布、大小、深度等信息。

舌体分割<sup>[6-9]</sup>、舌象特征提取及分类取得一定进展,越来越多的学者研究舌象特征与疾病、处方之间的关系<sup>[10-12]</sup>。Li等<sup>[13]</sup>等针对当齿痕舌不是凹入区域时,利用凹入区域信息对齿痕舌进行分类性能不稳定这一问题,提出一种使用多示例学习和卷积神经网络(CNN)三阶段的方法。Sun等<sup>[14]</sup>提出一种视觉解释方法来解决神经网络学习到的特征可视化问题,该方法将整个舌头作为输入,使用卷积神经网络来提取特征,然后对舌头进行分类并生成粗略的定位图,最后使用梯度加权类激活映射齿痕标记区域。Kanawong等<sup>[15]</sup>使用机器学习方法中的AdaBoost和SVM以及多层感知网络建立舌象和相应的症之间的

关系,通过对舌头上的瘀斑点大小、尺寸、分布,来研究其与胃炎之间的联系。Xiao等<sup>[16]</sup>通过使用CNN构建舌诊的寒症、热症、正常3种状态的识别,实验结果表明该模型对二分类问题识别准确率高于三分类问题。Hu等<sup>[17]</sup>构建真实世界的舌部图像和相应处方对应的神经网络模型,为移动医疗系统中的医疗服务提供新思路。

深度学习的对象检测取得巨大成功<sup>[18]</sup>,为提高舌象特征识别准确率提供了契机<sup>[19]</sup>。本文针对齿痕形状大小多样难以识别、用于金标准的舌象数据有限等问题,提出一种基于Mask Scoring R-CNN(以下简称MS R-CNN)<sup>[20]</sup>和迁移学习的舌象特征识别方法,能够识别齿痕特征、准确定位齿痕位置、标定齿痕大小、提取齿痕个数,可以在一定程度上降低小样本数据上应用深度学习技术的难度系数,获得较高的分类准确度。

### 1 实验方法

本实验舌象分类流程如图1所示。在离线训练阶段,首先筛选符合条件的舌部图片,构建舌象数据集。接着对图片进行预处理,并由多个中医专家对舌象进行标注,最后使用迁移学习的方法训练模型。在线测试阶段,使用测试图片对模型进行测试,并评估模型性能。

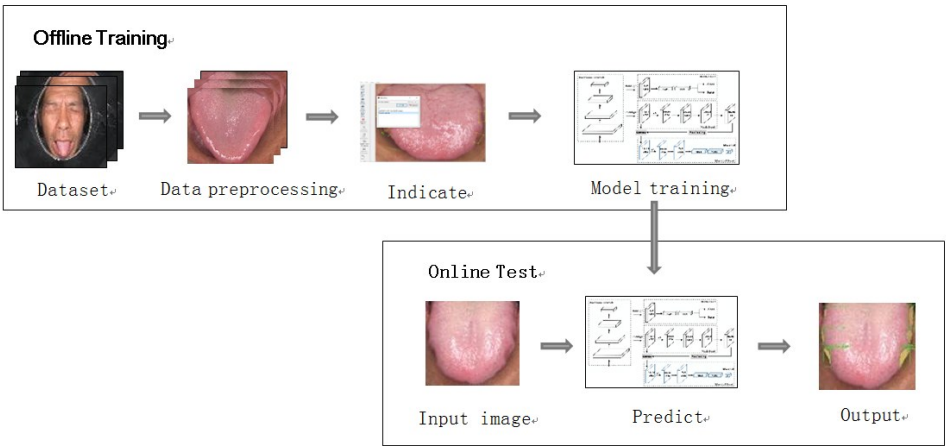


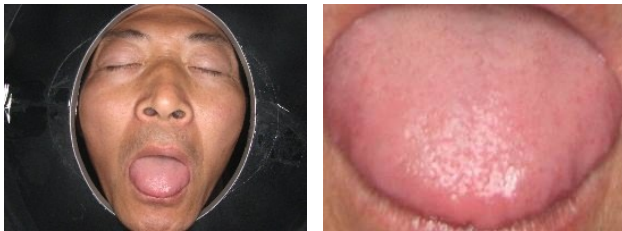
图1 舌象特征识别实验流程图

Fig.1 Flowchart of tongue feature recognition experiment

#### 1.1 数据筛选与预处理

本课题组与上海中医药大学附属曙光医院合作,根据采集标准使用专业采集设备,采集1 500张临床舌象照片。为了提高后续研究舌象特征的准确率,由多位专业中医从业人员帮助标注,完成一致性评价确认。本次实验共筛选出无齿痕舌400例,齿痕舌756例,分辨率为2 816×2 112。对筛选后的图片首

先进行统一命名,接着使用训练好的舌部分割网络模型对采集到的舌部图像进行裁剪,如图2所示。一是以减少图片中面部其他区域对后续工作的影响;二是原图像数据分辨率大小为2 816×2 112,分辨率较大会增大网络训练的难度和训练时长。最终得到的图片分辨率大小为563×397,可以直接输入神经网络。



a: 原始图像                      b: 预处理后的图像

图2 原始图像与处理后的图像

Fig.2 Original image and processed image

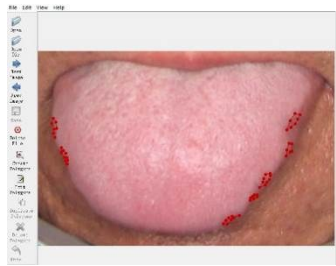


图3 使用标注软件Labelme对舌象特征进行标注

Fig.3 Marking tongue features by Labelme

1.2 舌象标注与数据集划分

MS R-CNN网络的训练不仅需要图像还要有与之相对应的掩膜信息。使用基于Python的Labelme开源图像标注软件,对舌部区域进行标注和制作掩膜,得到舌体图像中齿痕的分割信息,如图3所示。为了保证有充分的数据用于模型的训练,同时预留一定量的数据用于验证和测试,避免因数据划分引入其他偏差,参考数据集划分的常见方法:留出法。将所有标注完成的数据随机分为训练集、验证集、测试集,最后将训练集和验证集的注释信息按照COCO数据集格式整合到JSON格式文件中。

1.3 舌象特征提取模型构建

1.3.1 特征提取网络模型构建 MS R-CNN是首个解决实例分割假设打分问题的框架,简单、有效、通用。Huang等<sup>[20]</sup>在Mask R-CNN的基础上增加用于了解预测掩码质量的网络块,可以检测目标并同时能够

更准确地评估预测的质量<sup>[21-22]</sup>。该网络结合了用于目标检测的Faster R-CNN和用于语义分割的FCN,在Faster R-CNN检测到目标后,将FCN用于掩膜预测、边框回归、分类、掩膜评分。MS R-CNN在Mask R-CNN的基础上增加一个网络块,用来学习预测的实例掩膜的质量。掩膜评分策略校正掩膜质量与掩膜得分之间的偏差,提高实例分割性能,为实例分割的改进提供一个新的方向。

MS R-CNN方法分为4个阶段,其网络结构如图4所示。第一阶段为特征提取,将预处理后的舌图片输入ResNet+FPN的主干网络,从特征金字塔的不同层次提取特征。第二阶段为感兴趣区域(ROI)生成,通过RPN提取出候选ROI。第三阶段使用ROIAlign从每个候选ROI中提取特征,并进行边框回归、softmax分类和掩膜预测。最后一个阶段是MaskIoU Head,在预测的掩膜和groundtruth掩膜之间进行回归运算。

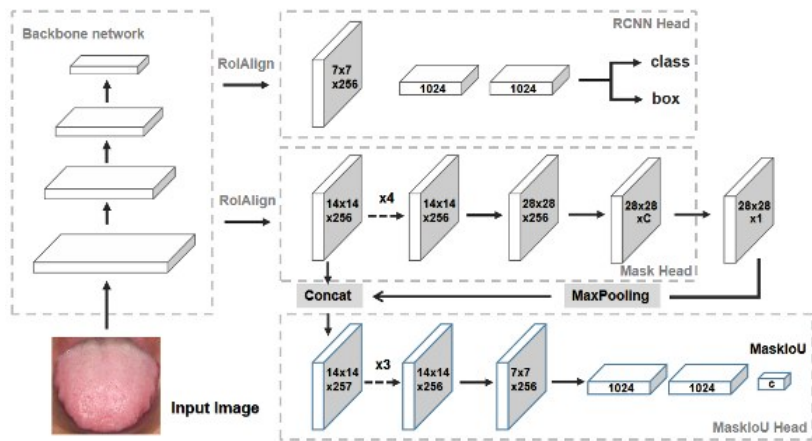


图4 基于Mask Scoring R-CNN的舌象特征提取网络结构图

Fig.4 Network architecture of tongue feature extraction based on Mask Scoring R-CNN

1.3.2 迁移学习的应用 本文应用迁移学习方法减少训练时间,防止神经网络的过拟合。尽管尽量减少神经网络模型参数可以避免过拟合,然而医学图像本身数据难以采集、可以作为金标准的数据量少,无法与十几万的开源数据集相比。迁移学习是将在一个任务上训练好的模型通过调整应用到另一个任务

上<sup>[23-26]</sup>。在本实验中,使用在COCO数据集上预训练的ResNet主干网络,删除预训练模型的最后一层。因为预训练的模型是为了不同的分类任务而训练的,去除全连接层和与分类得分、边界框预测、掩膜预测层相关的权重和偏差。使用未经训练的层来替换移除的层,将类别调整为3类,最后将数据集和相



关的掩码输入到 MS R-CNN 模型架构中,如图 5 所示。

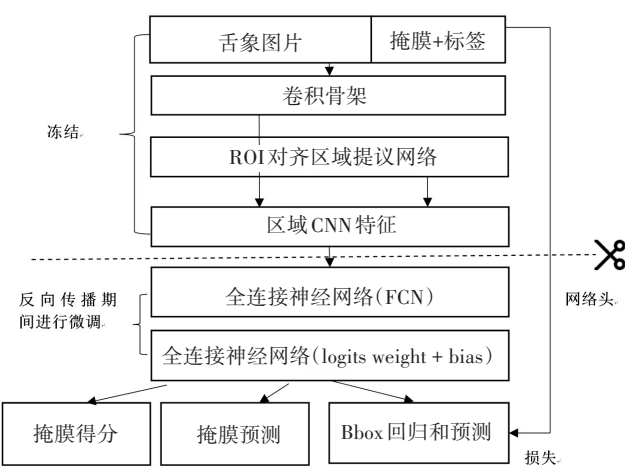


图5 用于舌象数据集上迁移学习的Mask Scoring R-CNN结构  
Fig.5 Mask Scoring R-CNN architecture adopted for transfer learning on the tongue image data set

1.4 模型效果评价标准

本实验采用深度学习中目标检测任务常用的评价指标:准确率、精确率、召回率(recall)、F1分数。将神经网络预测类别与真实类别的组合划分为真阳性(TN)、假阴性(FN)、真阴性(TN)、假阳性(FP),以判别非齿痕舌和齿痕舌为例,若真实类别为齿痕舌,预测类别相同为真阳性,预测类别为非齿痕舌则为假阴性;若真实类别为非齿痕舌,预测类别相同为真阴性,相反为假阳性,具体计算公式如下所示:

准确率 =  $\frac{TP + TN}{TP + TN + FP + FN}$  (1)

精确率 =  $\frac{TP}{TP + FP}$  (2)

召回率 =  $\frac{TP}{TP + FN}$  (3)

$F_1 = \frac{2 \times \text{精确率} \times \text{召回率}}{\text{精确率} + \text{召回率}}$  (4)

2 实验结果

2.1 训练平台

模型训练在 Ubuntu64 位操作系统中离线完成, CPU 为 Intel Core i7-9700F, 内存 32 G, GPU 为 GeForce GTX 2080Ti。本文选取 Pytorch 作为学习框架,与 Tensorflow 相比,PyTorch 的优点是动态图,能够在强大的 GPU 加速的基础上实现张量和动态神经网络。而 Tensorflow 等都是静态图,不利于扩展。

2.2 实验细节

实验中使用的主干网络是由 50 层或 101 层残差神经网络(ResNet-50、ResNet-101)衍生而来的特征金字塔网络,使用 ResNet-50 主干网络模型的计算量低于 ResNet-101,使用 ResNet-101 可以在不改变模型或者训练的情况下显著改善结果。

本实验使用最初在 COCO 数据集上训练的 Resnet-101 网络,并删除全连接层,因为预训练的模型识别 88 类样本。全连接层的去除删除了与 classscore、Bbox 预测、mask 预测层相关的权重,然后使用未经训练的层替换。接着将数据集和相应的掩膜输入到 MS R-CNN 模型架构中。

在迁移学习阶段,冻结预训练的 Resnet-101 主干网络的权重,使模型能够从舌象数据中提取通用的低级特征。然后训练新添加的全连接层,根据新数据集的分布调整权值。在模型训练中,网络进行了更新和微调。

2.3 测试结果与分析

使用 232 张舌部测试图片(包含无齿痕和不同齿痕严重程度的齿痕舌)对模型进行评估。TP=101, FN=15, FP=1, TN=115, 得到模型  $F_1$  分数为 0.95, 准确率为 0.93, 精确率为 0.99, 召回率为 0.914。从预测结果来看,该模型具有良好的泛化性,针对齿痕舌不同严重程度(图 6)(轻度、中度、重度),舌型(图 7)(圆形、三角形、椭圆形、方形等),不同伸舌方向(图 8)具有较好的识别效果。其中,共有 15 例无齿痕舌被误识别为齿痕舌,通过与中医专家共同分析得出,被模型识别错误的无齿痕舌中包含先天性和由牙齿畸形导致的非齿痕凹陷。因此,在下一步工作中将加入更多非齿痕凹陷信息。

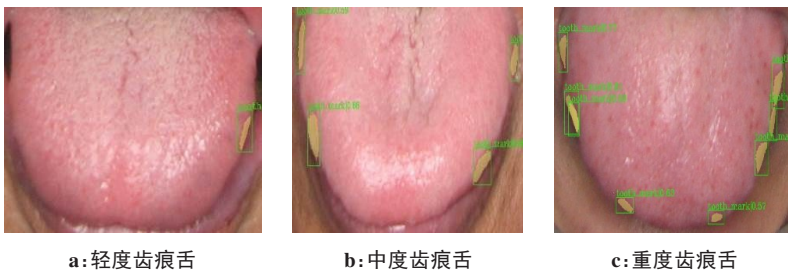


图6 对不同齿痕程度的齿痕舌识别效果

Fig.6 Recognition of different degrees of tooth marks

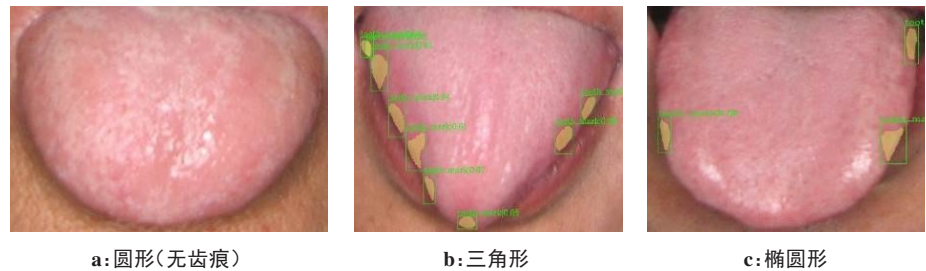


图7 对不同形状的齿痕舌识别效果

Fig.7 Recognition of different shapes of tooth marks

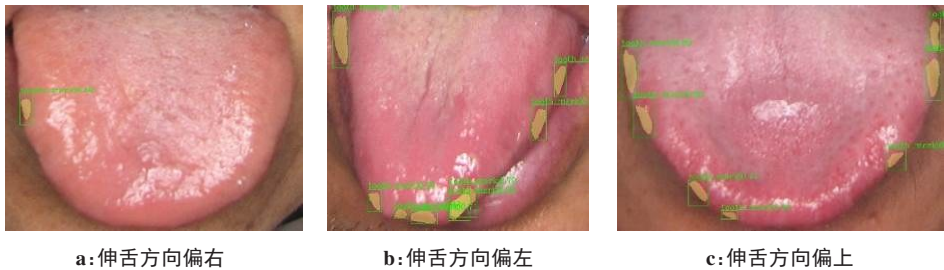


图8 对不同伸舌方向的齿痕舌识别效果

Fig.8 Recognition of tongue marks with different tongue extension directions

从可视化角度来看,相比于简单给出舌象的分类,该模型通过提供齿痕部分位置和大小的高分辨率可视化,结合置信度、齿痕的数量使研究者更好地理解网络,为齿痕严重程度分析提供依据,更加适用于移动医疗、远程医疗。

3 总结与展望

本实验将图像处理领域的最新研究成果应用到中医齿痕舌象特征提取及分类中,解决目前舌象识别方法未提供齿痕部分位置和大小的高分辨率可视化,深度学习技术难以应用在小样本医学数据的问题,提高齿痕舌象分类精度,对提高舌诊客观化分析准确度和医疗保健中的实际应用具有重要意义。该方法首先根据舌象特征识别任务的特点结合深度学习选择合适的神经网络;然后为提高算法的分类准确率和多种场景下齿痕舌象分类的鲁棒性,采用迁移学习方法。最后对模型进行测试和分析,得到较高的齿痕舌象分类准确率、准确的齿痕位置和齿痕数量。

未来,该研究将朝着以下3个方向继续深入。首先,提升模型的准确率。接下来将针对舌诊客观化改进模型,使模型更具轻量化,并提高准确率。其次,通过 MS R-CNN 的分析与实验,已了解 MS R-CNN 在处理中医舌象分类具有良好的表现。齿痕只是中医舌象中的一种,该研究继续将问题扩展到更多分类问题,研究更多舌象分类问题可以更好地帮助诊断和治疗。最后,该实验重点研究舌象特征提取与分类,然而,从中医的角度来看,要分析一个人

的身体健康,应收集望、闻、问、切这四诊的所有信息。因此在接下来的工作中,将结合先前的面诊、问诊等研究,建立更全面的深度神经网络。希望该工作可以促进医疗健康信息化进一步发展!

【参考文献】

[1] 何黎,曹悦,胡远樟,等. 基于中医“治未病”理论的中医四诊在中医体检中的应用及意义[J]. 科学技术创新, 2020(3): 4-6.  
HE L, CAO Y, HU Y Z, et al. Application and significance of four diagnoses of traditional Chinese medicine based on the theory of "curing disease" in traditional Chinese medicine physical examination [J]. Science and Technology Innovation, 2020(3): 4-6.

[2] 黄淑琼,张云龙,周静,等. 浅谈中医舌象客观化、量化、标准化研究[J]. 中华中医药杂志, 2017, 32(4): 1625-1627.  
HUANG S Q, ZHANG Y L, ZHOU J, et al. Study on objectification, quantification and standardization of TCM tongue image [J]. Chinese Journal of Traditional Chinese Medicine, 2017, 32(4): 1625-1627.

[3] 周耿标,黄东晖,蔡彦,等. 新型冠状病毒肺炎患者舌象特征与临床分型的关系[J]. 中医杂志, 2020, 61(19): 1657-1660.  
ZHOU G B, HUANG D H, CAI Y, et al. The relationship between tongue characteristics and clinical classification of COVID-19 patients [J]. Journal of Traditional Chinese Medicine, 2020, 61(19): 1657-1660.

[4] 董丽,杨思进,罗钢,等. 四川地区新型冠状病毒肺炎患者舌象变化规律研究[J]. 中药药理与临床, 2021, 37(1): 6-9.  
DONG L, YANG S J, LUO G, et al. Study on the change of tongue pattern of COVID-19 patients in Szechuan area [J]. Pharmacology and Clinics of Traditional Chinese Medicine, 2021, 37(1): 6-9.

[5] 任玉杰,梁嵘,张莹,等. 体检人群的齿痕舌与中医体质的相关性研究[J]. 世界科学技术(中医药现代化), 2012, 14(6): 2283-2289.  
REN Y J, LIANG R, ZHANG Y, et al. Study on the correlation between tongue marks and TCM constitution in physical examination population [J]. World Science and Technology (Modernization of Traditional Chinese Medicine), 2012, 14(6): 2283-2289.

[6] 周浩,胡广芹,张新峰. 基于舌图像深度特征融合的中医体质分类方法研究[J]. 北京生物医学工程, 2020, 39(3): 5-10.  
ZHOU H, HU G Q, ZHANG X F. Study on TCM constitution classification method based on tongue image depth feature fusion [J]. Beijing Biomedical Engineering, 2020, 39(3): 5-10.

- [7] ZHOU J, ZHANG Q, ZHANG B, et al. TongueNet: a precise and fast tongue segmentation system using U-Net with a morphological processing layer[J]. Appl Sci, 2019, 9(15): 3128-3147.
- [8] LI X L, YANG D, WANG Y, et al. Automatic tongue image segmentation for real-time remote diagnosis [C]//2019 IEEE International Conference on Bioinformatics and Biomedicine, 2019: 409-414.
- [9] QU P, HUI Z, LI Z, et al. Automatic tongue image segmentation for traditional chinese medicine using deep neural network [C]// International Conference on Intelligent Computing, 2017: 247-259.
- [10] 谢晟洁, 李福凤. 糖尿病患者舌、面诊参数与实验室检查指标的的相关性及其治疗前后的参数变化分析研究[J]. 中国全科医学, 2020, 23(21): 2631-2634.  
XIE S J, LI F F. Study on the correlation between tongue and facial parameters and laboratory examination indexes in patients with diabetes mellitus and the changes of parameters before and after treatment[J]. Chinese Journal of General Practice, 2020, 23(21): 2631-2634.
- [11] 冯宇, 周曼丽, 王健章, 等. 中医舌诊方法现代研究进展[J]. 陕西中医, 2020, 41(6): 838-840.  
FENG Y, ZHOU M L, WANG J Z, et al. Modern research progress of tongue diagnosis in traditional Chinese medicine [J]. Shaanxi Traditional Chinese Medicine, 2020, 41(6): 838-840.
- [12] DAI Y, WANG G. Analyzing tongue images using a conceptual alignment deep autoencoder[J]. IEEE Access, 2018, 6(99): 5962-5972.
- [13] LI X Q, ZHANG Y, QING C, et al. Tooth-marked tongue recognition using multiple instance learning and CNN features[J]. IEEE Trans Cybern, 2019, 49(2): 380-387.
- [14] SUN Y, DAI S M, LI J D, et al. Tooth-marked tongue recognition using gradient-weighted class activation maps[J]. Future Internet, 2019, 11(2): 45.
- [15] KANAWONG R, OBAFEMI-AJAYI T, LIU D H, et al. Tongue image analysis and its mobile app development for health diagnosis[M]. Singapore: Springer, 2017: 1005.
- [16] XIAO M, LIU G, XIA Y, et al. A deep learning approach for tongue diagnosis [C]//International Conference on Advanced Machine Learning Technologies and Applications. Cham: Springer, 2019: 3-12.
- [17] HU Y, WEN G, LIAO H, et al. Automatic construction of chinese herbal prescriptions from tongue images using CNNs and auxiliary latent therapy topics[J]. IEEE Trans Cybern, 2019, 51(2): 708-721.
- [18] WU X, SAHOO D, HOI S C H, et al. Recent advances in deep learning for object detection[J]. Neurocomputing, 2020, 396(1): 39-64.
- [19] TANIA M H, LWIN K, HOSSAIN M A. Advances in automated tongue diagnosis techniques[J]. Integr Med Res, 2019, 8(1): 42-56.
- [20] HUANG Z, HUANG L, GONG Y, et al. Mask scoring R-CNN[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 6409-6418.
- [21] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]// Proceedings of the IEEE International Conference on Computer Vision. 2017: 2961-2969.
- [22] JOHNSON J W. Automatic nucleus segmentation with mask-RCNN [J]. Adv Comput Vis, 2018, 944: 53709-53721.
- [23] PAN S J, TSANG I W, KWOK J T, et al. Domain adaptation *via* transfer component analysis[J]. IEEE Trans Neural Netw, 2010, 22(2): 199-210.
- [24] YOSINSKI J, CLUNE J, BENGIO Y, et al. How transferable are features in deep neural networks[C]//Advances in Neural Information Processing Systems. 2014: 3320-3328.
- [25] LIU T, YANG Q, TAO D. Understanding how feature structure transfers in transfer learning[C]. IJCAI, 2017: 2365-2371.
- [26] DONAHUE J, JIA Y, VINYALS O, et al. Decaf: a deep convolutional activation feature for generic visual recognition [C]//International Conference on Machine Learning, 2014: 647-655.

(编辑:陈丽霞)