

# 基于深度学习的舌象特征分析



李渊彤<sup>1</sup> 罗裕升<sup>2</sup> 朱珍民<sup>2,3</sup>

1 湘潭大学信息工程学院 湖南 湘潭 411105

2 中国科学院计算机技术研究所 北京 100080

3 移动计算与新型终端北京重点实验室 北京 100080

(641985483@qq.com)

**摘要** 中医舌诊因其直观稳定且易于观察的特点,以及具有较高的临床应用价值和快捷方便的实用性,成为了一个重要的研究课题。目前,将医学图像处理技术、人工智能技术和中医专家的临床经验相结合,实现了对中医舌诊的客观化、量化和自动化,这是当前中医舌诊现代化研究的主流方向。文中研究了基于迁移学习和深度学习的舌体精确分割和舌象特征识别的关键技术,提出了一种基于区域关联的单像素损失函数的舌体分割方法,新的损失函数不仅考虑到了区域像素之间的关联关系,而且有效利用了像素标签语义的监督信息,能够更好地指导模型进行训练学习,在测试集上的 MIoU 指标达到了 96.32%。然后,针对舌象几何特征提出了一个融合空间转换网络和 VGG16 模型的舌象几何特征分析模型,使用了空间转换网络来显式地学习空间不变性,并复用了 VGG16 模型的卷积部分,使得可以用舌体分割任务学习到的知识来进行参数迁移学习。通过两组对比实验,验证了空间转换网络对提高模型空间不变性的有效性,以及舌体分割的知识迁移能使模型更快、更平稳地收敛。同时,提出了一个融合深度纹理编码网络和 VGG16 模型的舌象纹理特征分类模型,使用深度纹理编码网络能将卷积得到的有序特征图编码成无序的纹理语义表示,以更有效地表达纹理信息。通过实验对比分析验证了深度纹理编码网络的无序编码对舌象纹理语义表示的有效性。

**关键词:** 中医舌诊;舌象分析;舌体分割;深度学习;迁移学习

**中图法分类号** TP391

## Tongue Image Analysis in Traditional Chinese Medicine Based on Deep Learning

LI Yuan-tong<sup>1</sup>, LUO Yu-sheng<sup>2</sup> and ZHU Zhen-min<sup>2,3</sup>

1 College of Information Engineering, Xiangtan University, Xiangtan, Hunan 411105, China

2 Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China

3 Beijing Key Laboratory of Mobile Computing and Pervasive Device, Beijing 100080, China

**Abstract** The traditional Chinese medicine tongue diagnosis, because of its intuition and easy to be observed, as well as its high clinical value, convenience and practicability, has become one of the important research subjects. At present, the combination of medical image processing technology, artificial intelligence technology and clinical experience of Chinese medicine experts to achieve objectification, quantification and automation of TCM tongue diagnosis is the mainstream of modernization research of TCM tongue diagnosis. In this paper, the key techniques of tongue segmentation and tongue image feature recognition based on migration learning and deep learning are studied. A tongue segmentation method based on region-based single pixel loss function is proposed. It can instruct the training and learning of the model by combining the color correlation and the semantic correlation between neighboring pixels, and the semantic information of target pixel labels. The experiments show that it partly improves the segmentation effect of the model, the MIoU index on the test set reached 96.32%. Then, a classification model of the tongue image geometric features, which combines spatial transformation network and VGG16 model, is proposed to identify and extract the geometric features of tongue image, providing a basis for syndromic inference of tongue image. Considering the orderliness of the geometric features of the data on the two-dimensional plane, the spatial transformation network is used to explicitly learn the spatial invariance in the model. And the convolution part of the VGG16 model is reused, so that the knowledge learned from the

到稿日期: 2019-10-16 返修日期: 2020-03-07 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家重点研发计划项目(2018YFC2000605)

This work was supported by the National Key R&D Program of China (2018YFC2000605).

通信作者:朱珍民(zmzhu@ict.ac.cn)

tongue segmentation task can be used for parameter transfer learning. Through two sets of comparative experiments, the validity of the spatial transformation network is proved to improve the spatial invariance of the model, and the knowledge of transfer learning is proved to make the model converge faster and more smoothly. At the same time, a classification model of the tongue image texture features, based on the deep texture coding network and VGG16 model, is proposed to recognize and extract the texture features of tongue image, providing a basis for syndromic inference of tongue image. According to the disorder of texture features in two-dimensional plane, a deep texture coding network is used to encode the ordered feature map, obtained by convolution layers, into an orderless texture semantic representation, which can express texture information more effectively. And the deep texture encoding network can enable the whole model to input images of any size, which gets rid of the loss of texture information caused by scaling operations of fixed input size. The validity of the orderless encoding of the deep texture encoding network for texture semantic representation is verified by the comparative analysis of experiments.

**Keywords** Traditional chinese medicine tongue diagnosis, Tongue image analysis, Tongue segmentation, Deep learning, Transfer learning

## 1 引言

中医舌诊是中医通过辨识舌象来了解人体的生理、病理变化,进行健康状态和疾病诊断的一种方法<sup>[1]</sup>,是在数千年的临床实践和经验积累中形成的相对完整的诊疗理论,对疾病的认识、理论方药的指导、疗效的评价都发挥着重要的作用。与西医常用的诊断手段相比,舌诊存在很多优势:

- 1) 稳定,扰动小,不同于中医“切诊”或西医听诊容易受到个人情绪的影响;
- 2) 无痛无伤,诊断温和,不需要通过穿刺、采血等检测手段就能了解人体的健康状态;
- 3) 简单易行,成本低,不需要依靠复杂昂贵的医疗设备,诊断过程快捷方便;
- 4) 舌诊或者中医,不同于西医诊断需要获取很多具体的生理指标,而是从整体通过辩证分析来把握人体的健康状态,能够发现西医难以定诊的一些人体亚健康状态。

然而,中医舌诊仍然避免不了传统中医面临的问题:传统舌诊需要用肉眼观察,且诊断结果又重于形象描述而轻于客观量化,甚至使用模糊多样的描述性文字,使得整个诊断过程缺少客观的评价指标以及统一的现代科学语言描述,由此必导致诊断结果易受医生个人经验、知识水平、主观意识以及诊断环境的影响,从而产生偏差。因此,对中医舌诊的现代化和客观化研究有着极其重要的意义和应用前景:有望为临床舌诊提供标准化依据,摆脱主观因素的影响,提高准确率;有望数字化舌象信息,定量化诊断结果,促进知识传承与学术交流;有望留存诊断过程中的量化数据,完善医疗事故评价体系。

在大数字与信息化时代,结合医学图像处理技术、智能信息技术和中医专家的临床经验等来解决中医舌诊过程中的技术标准化和数据化等关键问题,深入挖掘中医舌诊模式的科学内涵,进一步提高其客观性和临床有效性,实现舌诊的自动化,是当前中医舌诊现代化研究的主流方向。目前,舌诊自动化主流的解决方案如图1所示。首先采集舌像并进行色彩校正等图像预处理工作,接着分割并获取舌体,识别并提取舌体的各类视觉特征,最后从中医舌诊的知识库中获取相应的知识规则来完成从舌象特征到中医证候的推断。

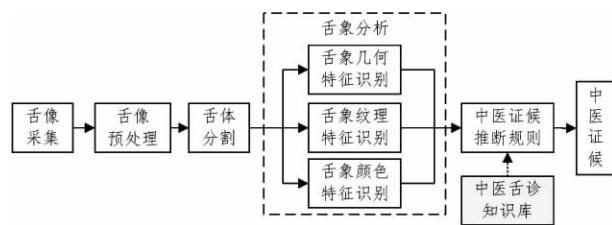


图1 舌诊自动化流程

Fig. 1 Tongue diagnosis automated process

目前,对于中医舌诊自动化的研究,已有不少学者在舌像采集、舌像预处理、舌体分割到舌象特征的定量化等多个方面做了一些相对成熟的工作。在舌体分割的研究中,Zhou 等<sup>[2]</sup>和 Zhang 等<sup>[3]</sup>使用了基于阈值的分割方法,Wu 等<sup>[4]</sup>、Zhang<sup>[5]</sup>和 Huang 等<sup>[6]</sup>使用了分水岭、区域生长等基于区域的分割手段,而 Liu 等<sup>[7]</sup>和 Sun 等<sup>[8]</sup>则基于边缘检测的方式来研究舌体分割。这些工作大多基于传统的数字图像处理技术,对标准舌像采集设备或仪器采集的图片进行分割。但由于舌体与周围组织颜色相似,轮廓相对模糊,这些传统的分割方法难以实现舌体的精准分割,且在非标准仪器采集的舌图(如手机照片等)上分割效果更不理想。另外,这类方法在分割过程中一般需要人工的直接或间接参与,属于半自动化手段。

在过去的研宄中,很多学者使用传统的机器学习方法来完成舌象特征的定量化工作,一般需要先进行图像的预处理,再人为设计或提取大量数据特征,如 Wei 等<sup>[9]</sup>、Xu 等<sup>[10]</sup>、Lu<sup>[11]</sup>在研究胖大舌、齿痕舌等的工作中就设计了很多的几何特征,Zhai 等<sup>[12]</sup>、Yang<sup>[13]</sup>研究舌苔时也利用小波、共生矩阵、分形维等手段设计了很多的纹理特征。这些方法无法做到端到端的学习,且设计好的特征通常需要后期的分析与筛选,不仅工作量极大,而且容易遗漏有效特征。Liu<sup>[14]</sup>和 Hu 等<sup>[15]</sup>使用了神经网络来完成舌象证候的分类,摆脱了特征设计的麻烦,但这些方法都跳过了对舌象特征的提取,直接通过舌图对中医的部分证候进行分类。而证候推断是相当复杂的,一般需要除舌诊外的其他诊断信息才能准确确认,因此这些工作的研宄价值相对不大。另外,目前的研究或者只是在单一舌象特征上进行分类或量化研究,或者基于多个模型分别提

取多个舌象特征,并未考虑到各舌象特征之间的依赖性或者关联性,如胖大舌同时也会伴随齿痕舌的舌象,舌尖红常伴随芒刺等。如果在提取多个舌象特征的过程中能够融入舌象特征之间的关联关系,理论上预测效果会更好。

针对舌诊自动化的两个关键技术点(舌体分割和舌象分析)做了相关分析与研究。本文基于深度学习和迁移学习技术,分别提出了一个舌体精确分割方案和一个舌象特征识别方案。首先,使用了 FCN-16s 分割模型并提出了一种新的基于区域关联性的单像素损失函数,以显式地指导模型学习局部区域像素间的相关性。在舌体分割的基础上,本文提出了一个融合空间转换网络和 VGG16 模型的多任务分类模型,以及一个融合深度纹理编码网络和 VGG16 模型的舌象纹理特征分类模型,来分别提取舌象的几何视觉特征和纹理视觉特征。

## 2 相关工作

国内外对传统中医舌诊的现代化工作始于 20 世纪 80 年代,并且随着计算机图像处理技术的发展,中国科技大学与安徽中医学院<sup>[16]</sup>率先开展了将图像处理识别技术应用于中医舌诊的研究,进行了舌色的量化分析实验,使得往后自动化舌诊的研究路线变得基本清晰,舌诊客观化的研究也由此进入了崭新的一页。结合图像处理技术与模式识别技术完成对中医舌诊的客观化、定量化和自动化是当前研究的主流方向,因此国内外的研究基本集中在舌像采集和舌像预处理、舌体分割和苔质分离、舌象特征(舌象的视觉特征)定量化,以及完整的舌诊分析系统这几方面,本节主要介绍舌体分割和舌象分析的相关工作。

### 2.1 舌体分割

舌像采集的过程中,一般都会拍摄进其他非舌体的部分,如嘴唇、牙齿和外围皮肤等杂乱的背景,有些采集装置甚至会把装置本身的一部分也采集进去。这些非舌体的存在,对医生或专家诊断舌象的影响并不大,但对于计算机自动化分析舌象却是很大的阻碍,特别是有关舌体形态的舌象特征的分析对舌体轮廓识别的精确度要求很高,因此,如何实现舌体的精准分割是中医舌诊自动化的一个至关重要的前提工作。目前,舌体分割大多是对传统的数字图像处理技术的部分改进,如基于阈值的方法,Zhou 等<sup>[2]</sup>针对在舌图像预处理中通过阈值分离方法来解决舌体分割问题存在的不足,提出了利用数学形态学法来进行二次处理的新方法;Zhang 等<sup>[3]</sup>针对舌图像的灰度和颜色特点,先对舌像的亮度灰度图像分别进行水平和垂直方向的灰度投影来确定舌体所在的区域,再采用自动选取阈值的 Otsu 法对该区域进行分割(如基于区域的分割);Wu 等<sup>[4]</sup>提出了一种基于分水岭变换和主动轮廓模型相结合的自动轮廓分割算法;Zhang<sup>[5]</sup>提出了基于初始区域和位置信息、基于种子点特征信息的两种改进的随机游走方法;Huang 等<sup>[6]</sup>提出了基于最大类间方差法和子块生长的舌体自动分割算法(如基于边缘检测的分割手段);Liu 等<sup>[7]</sup>通过将舌像从 RGB 空间转到 HSV 和 YCbCr 空间,利用数学形态学方法提取 Cb 和 V 分量上的舌体轮廓作为初始轮廓,再使用 Snake 模型实现分割;Sun 等<sup>[8]</sup>从舌体上方一层特殊的水

膜出发,在 HSV 颜色空间上预处理原始的舌像来得到舌体大致的边缘轮廓线,再应用改进的 Snake 最小能量函数进行边缘收敛来达到舌体提取的目的;还有许多与一些特定理论、方法相结合的舌像分割方法,如 Wang 等<sup>[17]</sup>提出了一种基于舌体的位置、颜色等先验信息来分割舌体的方法;Gao 等<sup>[18]</sup>提出了一种基于改进模糊算子和形态学的舌体分割方法;Li 等<sup>[19]</sup>借鉴光谱角度匹配算法,通过光谱角度匹配算法将超光谱舌图像立方体转换成光谱角度立方体,再采用一维脉冲波形检测出的边缘信息来实现基于超光谱图像的舌体分割;Liu 等<sup>[20]</sup>探讨了基于视觉注意机制和支持向量机相结合的舌体自动提取方法,为模式识别应用到舌体图像分割进行了初步探索;Wang 等<sup>[21]</sup>通过将现代三维点云处理技术与传统中医诊断经验进行有效融合,提出了基于扩展快速点特征颜色直方图特征值的欧氏聚类的舌体分割算法。

近年来,深度学习技术越来越成熟,在图像分割领域也出现了不少成熟的深度语义分割模型,但目前较少有研究将其应用于舌像分割,Wang 等<sup>[22]</sup>提出了一种基于两阶段卷积神经网络的舌体分割方法,该方法先在粗分割阶段使用卷积层和全连接层构建的网络来得到舌体所在的最小矩形区域,然后在精分割阶段使用卷积层与反卷积层构建的网络对粗分割得到的舌像进行进一步的语义分割,最后采用形态学相关算法对精分割后的舌体图像进行后续操作,即消除噪点和边缘粗糙点。

### 2.2 舌象分析

舌象是人体健康状态的外在表现,可以反映人体的气血、腑脏功能状态,是中医辩证诊断的主要依据之一,研究中医舌诊的客观化或者自动化就是研究如何通过计算机从舌图中自动分析出舌象特征及其病理变化。舌象一般包括舌质、舌苔以及舌下静脉的变化,主要特征可粗略划分为舌神、舌色、舌态、舌形、苔质、苔色和舌下脉络等几类。从计算机的角度看,舌象特征是指舌体的视觉特征,包括舌体的颜色特征、几何特征(形态特征)和纹理特征,因此舌象分析就是对舌体的 3 类视觉特征的分类识别。

Gong 等<sup>[23]</sup>探讨了 4 种常见病理苔质的定量方法及其与疾病的相关性,研究指出厚薄苔、润燥苔、腐腻苔、剥落苔等病理苔质是可以定量的,且其定量值有较明显的特异性;Sheng 等<sup>[24]</sup>基于舌象是低维流形结构在少数隐含变量的“控制”下所表现出来的多种形态的假设,采用流形学习方法对舌图像进行分析来考察舌象的内在变化规律;Yang<sup>[13]</sup>系统深入地研究了舌裂纹增强、舌裂纹区域提取、舌裂纹特征描述、面向疾病和中医证候的诊断分类等问题,并提出了一种基于 3 阶单位矩阵代数特征的彩色裂纹舌图像颜色变换算法、一种基于间隔差异度和先验知识的舌裂纹提取算法以及一种计算机化的舌裂纹诊病模型;Hu<sup>[25]</sup>结合传统的中医理论,利用主成分分析的统计方法,在 HSV 颜色空间中对舌象进行特征提取、降维,提取有效的舌象特征,并通过 AdaBoost 算法对舌象颜色进行了深入的分类研究;Lu<sup>[11]</sup>以齿痕舌象分析为切入点,提出了一种改进的齿痕个数提取算法,能有效解决曲线拟合精度无法界定的问题以及因边缘形态变化不明显而引起的齿痕个数难以提取的问题;Zhai 等<sup>[12]</sup>提出了一种基于 Gabor 小

波变换的舌苔腐腻识别方法,并针对 Gabor 小波变换的边缘检测效应使纹理描述弱化这一问题,提出了一种弱化边缘的方法;Zhang 等<sup>[26]</sup>以分形理论为基础,设计实现了计算数字化舌图分形维数的算法,并通过实验证明了分形维数对舌苔的厚薄、腐腻等数字化舌图图像特征敏感,指出其可作为数字化舌象纹理分析的有效参考参数;对于舌苔的润燥特征,Xie<sup>[27]</sup>利用二分光反射模型以及润燥系数,分别从识别和量化方面提出了新的识别研究方法;Wang 等<sup>[28]</sup>基于斑点检测、支持向量机和 K 均值聚类算法,提出了对舌诊图像中点刺和瘀点的识别及提取方法;Liu<sup>[24]</sup>探究了卷积神经网络在中医舌象分类模型中的可行性,设计了从舌图像直接到中医证候的分类实验;Liu 等<sup>[29]</sup>提出了一种结合灰度共生矩阵和小波纹理特征提取来自动描述舌苔厚度的方法,实现了舌苔厚度的定量化。

### 3 舌体分割方法的研究

舌象分析研究的主体对象是舌体,检测并分割出舌体,排除非目标对象的干扰,使舌象分析的结果更加精准,是中医舌诊自动化的一个至关重要的前提工作。其主要任务是去除诸如脸部、下巴以及嘴牙等外周皮肤组织部位,分割并提取舌体,且一般要求分离出的舌体边缘光滑且连续。由于舌体与周围组织颜色相似,轮廓相对模糊,不同人群的舌体形态差异较大,传统图像处理方法难以实现舌体的精准分割,且无法灵活应对开放的拍摄环境下的复杂背景问题。为此,本文提出了一种基于区域关联损失的舌体分割方法,提高了分割的准确度。

目前,深度语义分割模型都是对各个像素进行分类,输出的是单个像素的边缘概率分布,且使用的损失函数也是基于单个像素的负对数损失(Negative Logarithm Loss, NLLoss),这样就无法显式地保证模型学习到了像素与像素之间的关联性,忽略了在通常的基于像素分类的分割方法中使用空间规整(Spatial Regularization)的步骤,缺乏空间一致性。一个自然的想法是让模型能够输出局部区域的联合概率分布,基于联合概率分布设计损失函数,但这个想法不易实现,假设是简单的二分类问题,基于  $3 \times 3$  区域的联合概率分布将有  $2^9$  种情况,使模型计算变得复杂。基于区域联合概率分布输出的模型因计算复杂而难以设计,因此基于区域联合概率分布的损失函数也相应地不易设计。鉴于以上原因,本文设计了一种基于区域像素之间的相关性来设计单像素损失函数,这样既考虑到了局部区域像素之间的关联关系,又能将其应用于现有的单像素边缘概率分布的输出模型。

基于局部区域的相关性存在以下 4 个事实:

- 1) 在局部区域内,领域的真实语义与目标像素的真实语义越相近,目标像素被预测正确的概率就越大,相应地,错误的预测对损失函数的贡献也更大。
- 2) 在局部区域内,如果目标像素与邻域像素在颜色空间上的距离越接近,即目标像素与这些邻域像素存在越强的颜色关联性,则同样目标像素被预测正确的概率更大。
- 3) 在局部区域内,如果与目标像素真实语义相近(相同标签)的邻域像素已经被预测正确(分配正确的标签),则目标像

素被预测正确的概率更大。

4) 在局部区域内,与目标像素颜色相近的邻域像素应该尽量被预测成相近的语义(即分配相同的标签)。

基于前两个事实可以发现,在同一张图片中每个像素对损失的贡献程度是不一样的。因此,可以根据局部区域的关联性给每一个单像素的损失赋予不同的权重,权重的计算公式如下:

$$w_i = 1 + \frac{1}{k^2 - 1} \sum_{j=1}^{k^2-1} label_{sim_j} * (1 + color_{sim_j} + pred_{prob_j}) \quad (1)$$

其中,  $k$  表示大小为  $k \times k$  的矩形邻域,  $label_{sim_j}$  表示邻域像素  $j$  与目标像素  $i$  是否相等, 其值为 0 或 1,  $color_{sim_j}$  表示邻域像素  $j$  与目标像素  $i$  在 RGB 空间的余弦相似度,  $pred_{prob_j}$  表示邻域像素被预测为正确标签的概率。

由上述公式可以看出,  $1 + \frac{1}{k^2 - 1} \sum_{j=1}^{k^2-1} label_{sim_j}$  是对应于第一个事实的, 而  $1 + \frac{1}{k^2 - 1} \sum_{j=1}^{k^2-1} label_{sim_j} * color_{sim_j}$  对应于第二个事实,  $1 + \frac{1}{k^2 - 1} \sum_{j=1}^{k^2-1} label_{sim_j} * pred_{prob_j}$  则对应于第三个事实。整个公式与 Focal Loss<sup>[30]</sup> 类似, 给予单像素的损失不同的权重, 区别在于后者是在标签层面上分配权重, 旨在弱化目标预测中正负样本比例严重失衡的问题; 而前者是基于局部关联性分配权重, 旨在确保模型学习的指导方向正确。

基于第四个事实, 可对预测的语义相似度额外定义一个基于颜色的损失, 旨在体现颜色相近的像素之间应该有相近的语义分布, 即分配相同的标签。计算公式如下:

$$color\_loss_i = -\frac{1}{k^2 - 1} \sum_{j=1}^{k^2-1} \log(pred_{sim_j} + \epsilon) * (color_{sim_j} - \beta) * label_{sim_j} \quad (2)$$

其中,  $pred_{sim_j}$  表示邻域像素  $j$  与目标像素  $i$  的边缘概率分布的余弦相似度,  $\epsilon$  是一个极小正数(这里取  $10^{-6}$ ),  $\beta$  是一个可调节的参数, 这里默认取值为 0.5, 表示颜色相似度的一个阈值, 使用负对数是为了保持与原损失函数值域范围的一致性。

从式(2)中可以看出, 当  $color_{sim_j}$  小于阈值  $\beta$  时,  $pred_{sim_j}$  越小(语义距离越大), 损失就越小; 反之,  $pred_{sim_j}$  越大(语义距离越小), 损失就越小。这个损失只关心与目标像素真实语义相近且颜色也相近的邻域像素尽量分配与目标像素相同的标签,而不关心标签是否分配正确,因此该损失函数理论上可以在无真实标签参与的情况下指导模型学习到部分信息。

综上,改进的损失函数称为局部关联损失(Local Correlation Loss, LCLoss), 定义为:

$$Loss_i = \frac{1}{K} \sum_i^K w_i * \log(p_i) + \alpha * color_{loss_i} \quad (3)$$

其中,  $K$  是数据样本个数,  $\alpha$  是一个可调节的参数, 这里默认取值为 0.1。

该损失函数既考虑到了区域像素之间的关联关系, 又有效利用了像素标签语义的监督信息, 理论上能够更好地指导模型进行训练学习。

### 4 舌象分析方法的研究

舌象是人体健康状态的外在表现,可以反映人体的气血、腑脏功能状态,是中医辩证诊断的主要依据之一。在中医理

论中,舌象一般是指舌质、舌苔以及舌下静脉的病理变化,主要包括舌神、舌色、舌形、苔质、苔色、舌下脉络的颜色及纹路等视觉特征的变化。因此,舌象分析的任务是对舌体的各个视觉特征的分类识别,属于多任务分类问题。

从计算机的角度看,舌体的视觉特征广义上就是指舌体的颜色特征、几何特征(形态特征)和纹理特征,舌象分析的任务也可相应地划分为舌象颜色特征分析、舌象几何特征分析以及舌象纹理特征分析。自动化舌象分析的过程如图 2 所示,先对舌体进行分割与裁剪,得到目标舌体后,再通过舌象识别得到舌象的 3 类视觉特征。

本文在对舌象数据进行充分分析的基础上,分别对舌体(舌质和舌苔)的几何特征识别和纹理特征(包含颜色特征)识别提出了基于迁移学习的舌象分析方案。

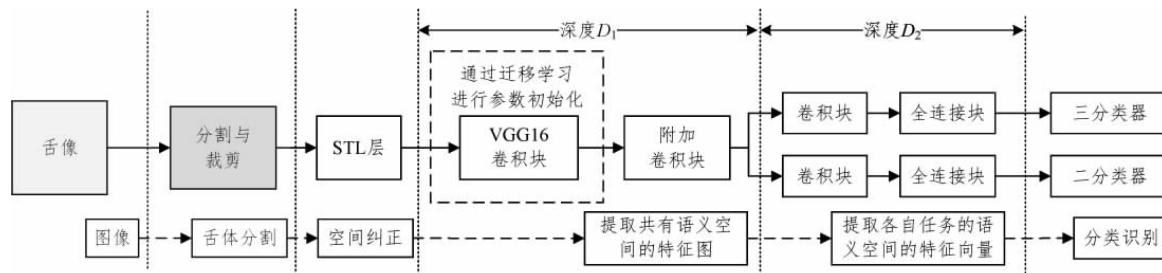


图 2 舌象自动化分析过程

Fig. 2 Tongue image automated analysis process

#### 4.1 融合 STN 和 VGG16 的几何特征分析方法

舌象的几何特征分析是指对舌质胖瘦和舌质齿痕的视觉特征的分类识别。虽然舌象特征识别 T\_1 与舌体分割 T\_2 属于不同的任务,但考虑到两者的数据 x\_1 与 x\_2 之间有着很强的相似性,本文设计了如图 3 所示的神经网络结构,并使用舌体分割任务学习到的知识进行参数迁移,初始化模型的部分参数。

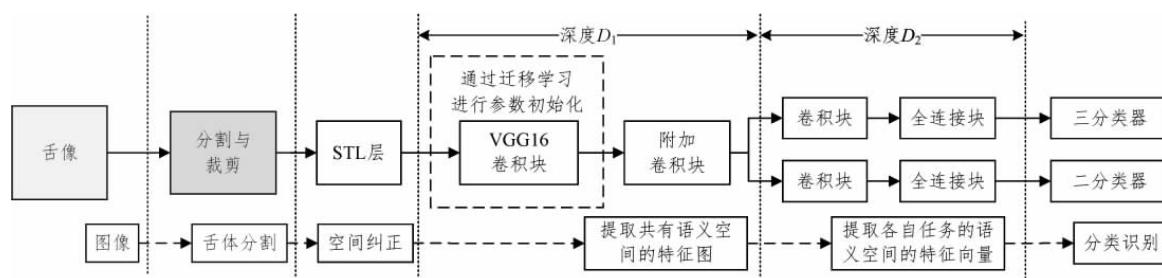


图 3 几何特征识别模型的基本架构

Fig. 3 Basic architecture of geometric feature recognition model

整个网络结构可分为 5 个模块:

第一个模块主要进行舌体分割处理,先使用改进的 FCN-16s 模型分割并提取舌体,再以分割的舌体的最小外包方形裁剪舌图,从而获得舌象分析的对象。

第二个模块是空间变换层 STL(来自空间转换网络<sup>[31]</sup>),用于纠正图像目标的空间位置,获得相应的空间不变性,处理开放的拍摄环境下引起的舌体姿势和位置不统一的问题。

第三个模块由一个 VGG16 模型的卷积部分和一个附加卷积块(多个卷积层和池化层的组合)组成,主要用于获得一个能够在多个分类任务之间共享的语义特征图,它在一定程度上体现了任务之间的关联性,即类标之间的依赖关系。理论上,该模块的网络层数 D1 越深,第三个模块的层数 D2 相对 D1 就越浅,因此更能捕捉多个分类任务在更深层的语义空间的共享信息和学习到类标之间的依赖关系。VGG16 的卷积块的参数可通过在 ImageNet 数据集上预训练的 VGG16 模型或者前一步工作的舌体分割模型的参数迁移学习来初始化。

第四个模块由一个卷积块(同样是多个卷积层和池化层的组合)和一个全连接块(多个全连接层的组合)组成,主要用于在共同的语义特征图上进一步抽取与各自任务相关的语义特征向量。

第五个模块是针对各个任务的分类器,基于提取的语义特征向量来完成分类预测,这里都使用 Softmax 分类器。

该模型具体的细节信息如图 4 所示,其中只表示出了每一层的输出信息,隐藏了相应的输入信息,另外也隐藏了池化层和 Dropout 层。由于包含形状信息的几何特征并不会随着图像的放缩操作而产生太大的信息损失,因此整个模型的输

入的图片大小是固定的,为  $3 \times 300 \times 300$ 。另外几何特征对于二维平面来说是有序的,其对平移、旋转和镜像等空间操作比较敏感,因此在模型的第一层使用了 STL 来学习相应的空间不变性。

Input(300 * 300 RGB image)
STL(Spatial Transformer Layer)
Conv3-64
Conv3-64
Maxpool
Conv3-128
Conv3-128
Maxpool
Conv3-256
Conv3-256
Conv1-256
Maxpool
Conv3-512
Conv3-512
Conv1-512
Maxpool
Conv3-1024
Conv3-2048
Conv3-1024
FC-1024
FC-512
FC-3
Softmax
Conv3-1024
FC-1024
FC-512
FC-2
Softmax

舌质胖瘦识别 舌质齿痕识别

图 4 几何特征识别模型

Fig. 4 Geometric feature recognition model

#### 4.2 融合 LREL 和 VGG16 的纹理特征分析方法

舌象的纹理特征分析是指对舌质裂纹、舌质芒刺、舌质瘀点、舌苔腐腻、舌苔厚薄以及舌苔苔色(准确地说苔色并不属于纹理,但本文认为它也可以通过纹理特征识别模型来预测识别)的视觉特征的分类识别。同样地,可以利用舌体分割任务学习到的知识进行参数迁移,因此本文设计了如图 5 所示

的神经网络结构。整个网络结构可分为 5 个模块:

第一个模块与几何特征识别模型的第一个模块是同一模块,通过舌体分割和裁剪来获得舌象分析的对象。

第二个模块与几何特征识别模型的第二个模块类似,同样用于获得一个能够在多个分类任务之间共享的语义特征图,一定程度上可以学到类标之间的依赖关系。

第三个模块是一个可学习的残差编码层 LREL(来自深度纹理编码网络<sup>[32]</sup>),主要用于把共享语义特征图转换成更适合纹理表达的语义编码特征。传统的卷积层是一个类似滑

动窗口的卷积操作,输出的特征图本质上是有序的(Ordered),使用 LREL 能够把任意大小的有序的特征图编码成固定大小的无序(Orderless)的特征图,从而获得一个空间不变性的语义表示。

第四个模块是一个全连接块(多个全连接层的组合),主要用于在共同的语义特征上进一步抽取与各自任务相关的语义特征向量。

第五个模块是针对各个任务的分类器,基于提取的语义特征向量完成分类预测,这里都使用 Softmax 分类器。

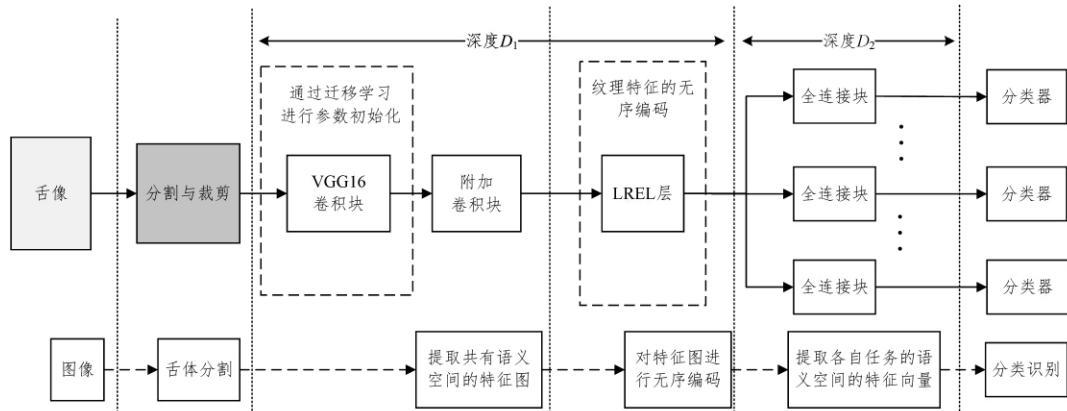


图 5 纹理特征识别模型的基本架构

Fig. 5 Basic architecture of texture feature recognition model

该模型具体的细节信息如图 6 所示,同样只表示了每一层的输出信息,隐藏了相应的输入信息、池化层和 Dropout 层。由于纹理信息一般对平移、旋转和镜像操作来说不敏感,因此没有必要使用 STL 或类似的操作来进行空间纠正处理。但纹理信息会因为放缩操作而产生一定程度的损失,因此在该模型中使用了深度纹理编码技术(即 LREL),将提取的共有的有序特征图进一步编码为一个无序的纹理语义表示,以更有效地表达纹理信息,使整个模型能够输入任意大小的图片。

Input(* RGB image)					
Conv3-64 Conv3-64 Maxpool					
Conv3-128 Conv3-128 Maxpool					
Conv3-256 Conv3-256 Conv1-256 Maxpool					
Conv3-512 Conv3-512 Conv1-512 Maxpool					
Conv3-1024	Conv3-1024	Conv3-1024	Conv3-1024	Conv3-1024	Conv3-1024
LREL(D512-K64)	LREL(D512-K64)	LREL(D512-K64)	LREL(D512-K64)	LREL(D512-K64)	LREL(D512-K64)
FC-1024 FC-512 FC-3 Softmax	FC-1024 FC-512 FC-3 Softmax	FC-1024 FC-512 FC-3 Softmax	FC-1024 FC-512 FC-3 Softmax	FC-1024 FC-512 FC-3 Softmax	FC-1024 FC-512 FC-3 Softmax
舌质裂纹识别	舌质芒刺识别	舌质瘀点识别	舌苔腐腻识别	舌苔厚薄识别	舌苔苔色识别

图 6 纹理特征识别模型

Fig. 6 Texture feature recognition model

## 5 实验设计与结果分析

### 5.1 改进的损失函数的舌像分割对比实验

#### 5.1.1 数据信息与评价指标

舌像分割实验采用来源于 Github 的开源数据集,是使用标准的舌像采集装置采集的 300 张舌图,分辨率为 768 \* 576,包含对应的掩模图片。为了增强模型的泛化能力,本文对舌像数据集进行旋转、镜像、缩放等随机增强操作,最后得到的训练集有 400 张图片,验证集和测试集各有 100 张图片。

对于分割模型的有效性,可以从执行时间、内存占用和分割效果等多个不同方面的度量标准来测试和衡量评估。评价图像分割的效果也有很多衡量标准,且各有优劣,目前通常使用的是像素精度和交并比以及它们的变种。本文采用的评价指标有以下 4 种:1) 表示标记正确的像素占总像素的比例的像素精度(PA);2) 通过计算每个类中被正确分类的像素数占该类中总像素的比例,来求得所有类的平均像素精度(MPA);3) 真实值与预测值两个集合的交集和并集之比的均交并比(MIoU);4) 对所有类求 IoU 的加权平均值的频权交并比(FWIoU)。其中,MIoU 由于简洁、代表性强而成为最常用的度量标准,本文也使用 MIoU 来作为模型选择的依据。

#### 5.1.2 实验设计

本实验主要对 U-Net,FCN-32s,改进的 FCN-16s 和改进的 FCN-8s 分别在 3 种不同的损失函数下进行对比分析。实验过程与上文实验类似,但损失函数依次替换为负对数损失 NNLoss, Focal Loss 以及本文提出的局部关联损失 LCloss。实验结果如图 7 所示,从图中可直观地看到,使用 Focal Loss

或是本文提出的局部关联损失 LCLoss,结果都比单纯地使用负对数损失 NLLoss 要好。其中,FCN 模型使用 Fcoal Loss 训练后其 MIoU 指标约提高了 0.2%~0.5%,而使用了 LCLoss 后其 MIoU 指标约提高了 0.8%~1%。U-Net 模型在使用 Fcoal Loss 和 LCLoss 后各项指标提升相对来看都比较显著,其中 MIoU 指标在两种损失下分别提高了约 4% 和

6%。FCN 模型由于有迁移学习的先验知识指导,损失函数对学习的指导作用相对不大,因此效果虽然有提升但不明显,而 U-Net 的效果提升较为明显。实验结果说明,改进损失函数确实能够更好地指导模型学习。再观察 U-Net 在使用了局部关联损失下的训练过程(见图 8)可以发现,虽然各项评价指标都有所上升,但仍然避免不了模型不收敛的问题。

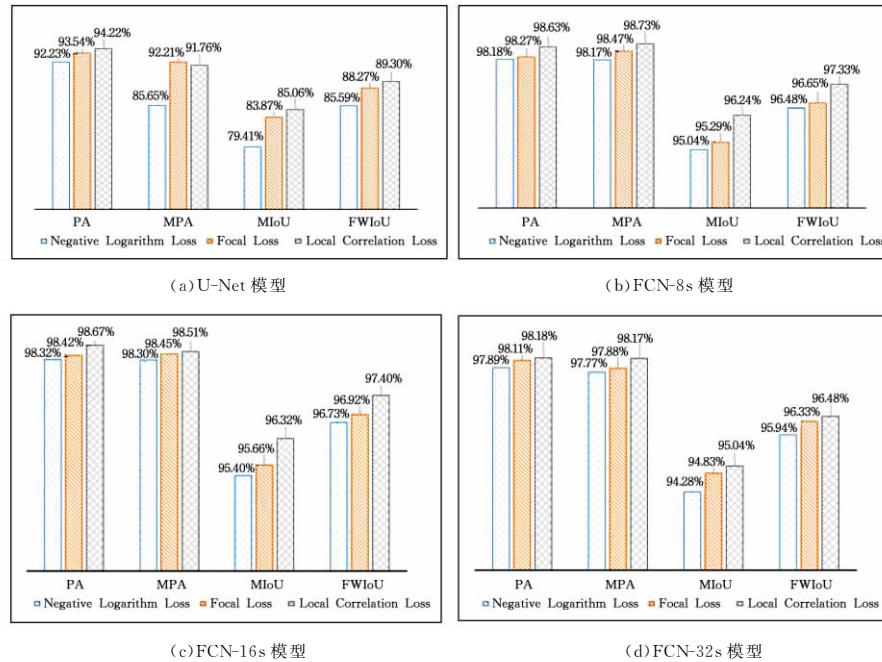


图 7 各模型在不同损失函数下的实验结果对比

Fig. 7 Comparison of experimental results of each model under different loss functions

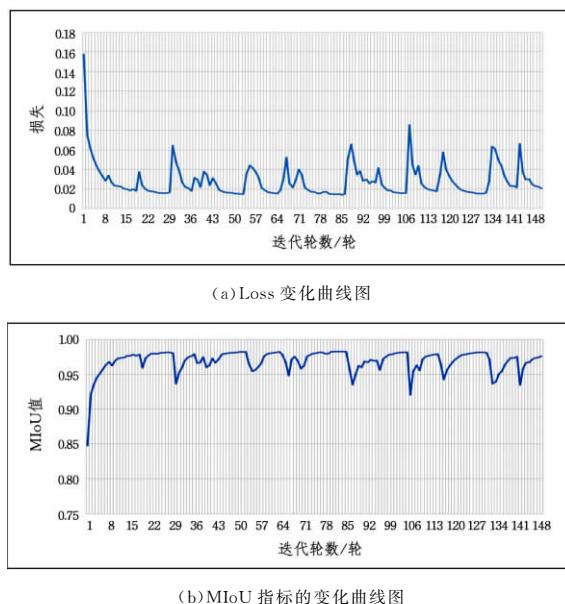


图 8 U-Net 模型在局域关联损失下的训练过程

Fig. 8 Training process of U-Net model under local correlation loss

### 5.1.3 结果分析

本实验主要对比分析了基于区域关联性的单像素损失函数对模型的指导学习能力,实验发现新的损失函数对 FCN 模型的评价指标略有提升,MIoU 指标大约提高了 0.8%~1%,

而对 U-Net 模型有较显著的提升,MIoU 指标约提高了 6%,但是仍未能解决 U-Net 模型不收敛的问题。最后,改进的 FCN-16s 模型加上 LCLoss 的指导训练,整个舌体分割模型在测试集上的 MIoU 指标可以达到 96.32%,如表 1 所列。

表 1 舌体分割模型的最终测试结果

Table 1 Final test results of tongue segmentation model

Model	PA/%	MPA/%	MIoU/%	FWIoU/%	Time/s
FCN-16s	98.00	97.71	94.54	96.13	261
FCN-16s+ LCLoss	98.67	98.51	96.32	97.40	262

从模型在测试集上的分割效果图(见图 9)可以看出,模型的直观效果较为平滑,而且十分完整。实验证明新的损失函数能在一定程度上提供 FCN 模型的各项分割指标,为舌像分割的进一步研究提供了基础。



注:奇数行为真实值,偶数行为预测值

图 9 分割效果图

Fig. 9 Segmentation effect map

## 5.2 舌象几何特征分析对比实验

### 5.2.1 数据信息和分布

舌象的几何分析的数据来源于上海中医药大学附属曙光医院整理出的部分诊断案例,共收集到了 528 张舌图(增强后为 1056 张),分辨率为  $300 \times 300$ 。每张舌图共标注有 8 个字段:舌质胖瘦、舌质齿痕、舌质裂纹、舌质芒刺、舌质瘀点、舌苔腐腻、舌苔厚薄以及舌苔苔色,其中前 2 个字段为舌体的几何特征,中间 5 个字段为纹理特征,最后一个字段是颜色特征,每个字段对应的取值如表 2 所列。

表 2 舌象标签

Table 2 Tongue Label

几何特征	胖	瘦	正常, 胖, 瘦
	齿痕		正常, 齿痕
纹理特征	裂纹		正常, 裂纹
	芒刺		正常, 芒刺
瘀点		瘀点	正常, 瘴点
		腐腻	正常, 腐, 腻
厚薄	厚		薄, 厚
	苔色		淡白, 白, 黄, 灰, 黑

另外,通过数据统计和分析发现,除了厚薄和齿痕外,其他特征都或多或少存在类别不平衡的问题,其中几何特征分布如图 10 所示。舌象数据集不平衡是不可避免的,在中医舌诊案例中,黑苔属于极为严重的病症,因此这类数据比较少见,同样地,在病理舌中,腻苔较腐苔更为常见,胖大舌较瘦舌也更为常见。大部分舌象特征都或多或少因类似的原因而存在数据不平衡的问题,因此舌象分析的数据一般都较难获取。数据难以获取,而标记和标准化也是大数据时代以来医疗数据采集面临的一大难题。本文为了充分利用数据集,并避免过拟合,因此不再划分验证集,而是直接在训练过程中使用 5 折交叉验证的方式进行验证。本文还使用舌体分割任务学习到的知识进行参数迁移,从一定程度上保证了模型的可靠性。

本文通过 FocalLoss 在一定程度上能同时解决样本类别不平衡的问题和简单与复杂样本的问题。平衡因子  $\alpha$  能根据类别样本的比例来平衡不同类别样本的重要性,一定程度上可以得到与采样方法相同的效果。另外,FocalLoss 在  $\gamma$  大于 0 时,可减少易分类样本的权重,从而使得模型在训练时更专注于难分类的样本。

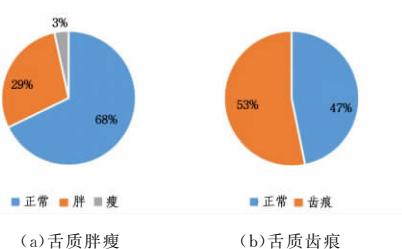


图 10 几何特征数据分布

Fig. 10 Geometric feature data distribution

### 5.2.2 实验设计

针对舌象几何特征分析的实验,本文选取 800 张舌图数据(包含增强后的数据)作为训练集,并使用 5 折交叉验证的方式训练并验证模型,选取剩下的 256 张舌图作为测试集,总

共完成了两组对比实验。由于在多分类任务下,“微平均”指标等于准确率,即  $Accuracy = P_{micro} = R_{micro} = F1_{micro}$ ,因此本文使用“宏平均”指标和准确率作为单个多分类任务分类效果的评价指标,又因为舌象分析属于多任务分类问题,所以在多个任务的评价指标上求平均,以此作为本实验的评价指标(本质上还是一种“宏平均”)。

#### 1) STL 对比实验

本实验就舌象几何特征分析模型是否使用 STL 进行空间变换的情况进行对比分析,实验过程的大致描述如下:使用 Focal Loss 作为损失函数,使用 Adam 算法作为优化方法,以每 800 张舌图作为一个 epoch 训练模型,使用 5 折交叉验证方法保存最好的验证结果与相应的模型参数,总共训练 40 个 epoch,训练结束后在测试集上计算保存的模型的平均准确率  $Average$ 、宏查准率  $P_{macro}$ 、宏查全率  $R_{macro}$  以及  $F1_{macro}$ 。

在测试集上的实验结果如表 3 所列,使用 STL 使得模型计算的时间开销多了 35ms 左右,但 4 个评价指标都有所提高,其中平均准确率提高了约 6%,宏精确度提高了约 1%,宏召回率提高了约 8%,而宏 F1 值提高了约 5%。可以认为,使用 STL 做空间变换虽然增加了计算量,但只是略微增加了时间开销,而分类识别的效果却得到了显著提升。因此,本文实验验证了空间转换层对提高模型空间不变性的有效性。

表 3 STL 在测试集上的对比实验结果

Table 3 Comparative experimental results of STL on test sets

Model	Average / %	$P_{macro}$ / %	$R_{macro}$ / %	$F1_{macro}$ / %	Time / ms
未用 STL	76.80	87.56	76.06	81.41	1 098
使用 STL	82.58	88.87	84.34	86.54	1 124

#### 2) 迁移学习对比实验

本实验主要针对舌象几何特征分析模型在没有迁移学习的知识参与、预训练 VGG16 的知识迁移和舌体分割模型的知识迁移这 3 种情况下的学习能力做对比分析。图 11 为模型在 3 种情况下的损失函数变化曲线图,其中横坐标是训练的 epoch,纵坐标是损失函数在一个 epoch 内的平均值。



图 11 舌象几何特征分析模型的损失函数变化曲线图

Fig. 11 Loss function curve of tongue image geometric feature analysis model

从图中可看出,借助于预训练模型的知识迁移,模型要比简单的训练学习收敛得更快且更稳定。另外,预训练 VGG16 的知识迁移和舌体分割模型的知识迁移的学习能力差别不大。如图 12 所示,三者在测试集上的评价指标相近,有迁移学习的则效果稍好一些。

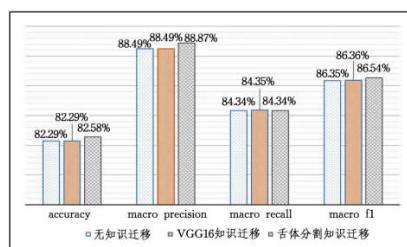


图 12 舌象几何特征分析模型在测试集上的评价指标对比  
Fig. 12 Comparison of evaluation indexes of tongue image geometric feature analysis model on test set

### 5.2.3 结果分析

舌象几何特征分析模型在测试集上对每个子任务的预测效果如表 4 所列,可以看到,对于舌质齿痕的识别,评价指标均高于 85%,预测效果还算理想;但是对于舌质胖瘦的识别,准确率仅为 79.17%,宏 F1 值仅为 83.38%,预测效果相对一般,原因可能是:

- 1) 数据集的胖瘦类别样本不平衡,相应的训练样本过少;
- 2) 舌质的胖瘦虽然可以通过舌体轮廓的形状、舌体的厚度和长度等信息加以识别<sup>[9-10]</sup>,但医生在诊断过程中一般还会参照嘴巴与舌体的大小关系(如胖大舌常常会出现“伸舌满口”或“肿大塞口”的状况),而经过分割步骤后,仅保留了舌体,缺少嘴巴的信息来作对比,因此可能会导致模型识别效果变差。最后,整个模型的平均准确率为 82.58%,宏 F1 值为 84.34%。

表 4 舌象几何特征分析模型的预测效果

Table 4 Prediction effect of geometrical feature analysis model of tongue image

Feature	(单位: %)			
	Aurracy	P <sub>macro</sub>	R <sub>macro</sub>	F1 <sub>macro</sub>
舌质胖瘦	79.17	86.52	80.47	83.38
舌质齿痕	85.98	91.22	88.22	89.69
平均值	82.58	88.87	84.34	86.54

### 5.3 舌象纹理特征分析对比实验

舌象纹理特征分析的实验过程与几何特征分析的实验过程几乎一样,损失函数和优化方法依旧使用 Focal Loss 和 Adam,评价指标使用平均准确 Aurracy、宏查准率 P<sub>macro</sub>、宏查全率 R<sub>macro</sub> 以及 R<sub>macro</sub>。

#### 5.3.1 数据信息与分布

舌象纹理特征分析的数据与几何特征分析的数据一致,其中纹理特征的数据分布如图 13 所示。

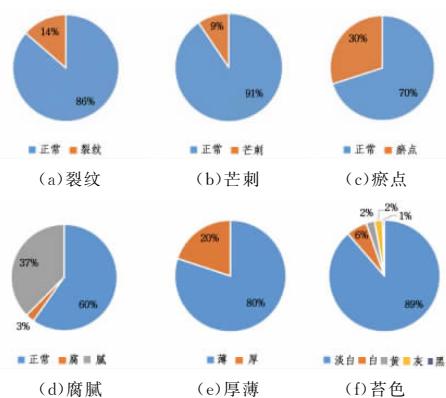


图 13 纹理特征数据分布

Fig. 13 Texture feature data distribution

### 5.3.2 实验设计

#### 1)LREL 对比实验

本实验同样就舌象纹理特征分析模型是否使用 LREL 进行纹理编码的情况进行了对比分析,由于删除 LREL 会导致原模型前后层的输出输入不一致(不像 STL 可以随意插入或者删除),因此这里不能简单地删除 LREL,而是替换成一个卷积层和多个“并联的卷积层”,以承前启后,如图 14 所示(虚线框部分为替换后的结构),这是一个与未使用 STL 的舌象几何特征分析模型类似的结构。

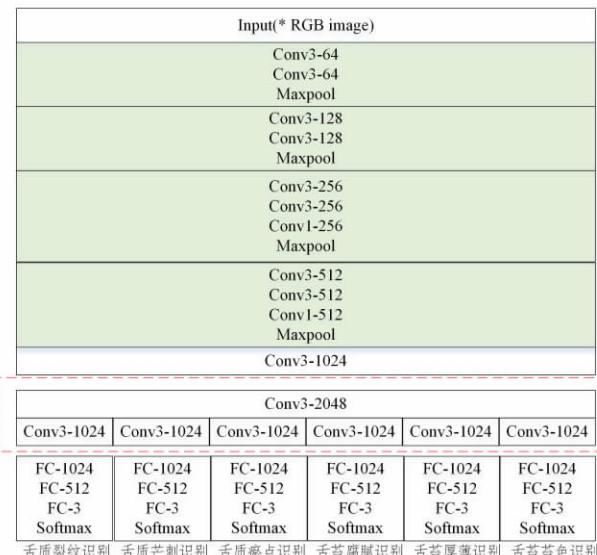


图 14 未使用 LREL 的纹理特征识别模型

Fig. 14 Texture feature recognition model without LREL

在测试集上的实验结果如表 5 所列,使用 LREL 使得时间开销增加了约 84ms,但评价指标都获得了显著的提升,可以看到,平均准确率和宏 F1 值都提高了近 10%,而宏精确度和宏召回率分别提高了约 8% 和 10%。因此,本实验验证了 LREL 的无序编码对舌象纹理语义表示的有效性。

表 5 LREL 在测试集上的对比实验结果

Table 5 Comparative experimental results of LREL on test set

Model	Aurracy / %	P <sub>macro</sub> / %	R <sub>macro</sub> / %	F1 <sub>macro</sub> / %	Time / ms
未用 LREL	70.52	79.82	73.70	76.08	1 243
使用 LREL	82.07	87.53	84.23	85.69	1 327

#### 2)迁移学习对比实验

本实验同样是针对舌象纹理特征分析模型在没有迁移学习的知识参与、预训练 VGG16 的知识迁移和舌体分割模型的知识迁移这 3 种情况下的学习能力的对比分析。图 15 为模型在 3 种情况下的损失函数变化曲线图,同样可看出,迁移学习比简单的训练学习收敛得更快且更稳定。另外,可以看到,在使用舌体分割模型的知识迁移的情况下,模型训练学习的速度比使用预训练 VGG16 的知识迁移稍快一些,可以理解为原舌体分割模型在处理舌体分割任务时捕捉到了更多的纹理信息。如图 16 所示,三者在测试集上的评价指标相近,基于舌体分割模型的迁移学习的预测效果比其他两个稍好一些。

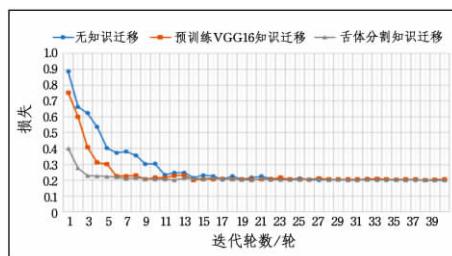


图 15 舌象纹理特征分析模型的损失函数变化曲线图

Fig. 15 Loss function curve of tongue image texture feature analysis model

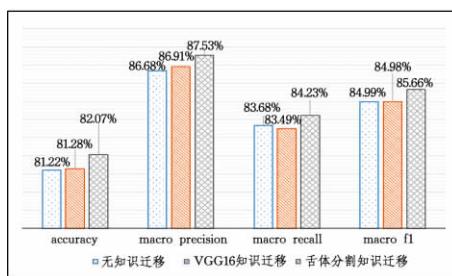


图 16 舌象纹理特征分析模型在测试集上的评级指标对比

Fig. 16 Comparison of rating indicators of tongue image texture feature analysis model on test set

### 5.3.3 结果分析

舌象纹理特征分析模型在测试集上对每个子任务的预测效果如表 6 所列,模型对于舌质裂纹和舌苔厚薄的识别效果相对较好,对于舌质芒刺的识别效果一般,而对于剩下的子任务识别效果相对较差,特别是舌苔苔色和舌苔腐腻。从表中可看出,模型对数据集中存在样本不均衡的类别识别能力都相对较差,说明样本不均衡问题确实对模型的学习产生了影响,降低了模型的识别能力。最后,整个模型的平均准确率为 82.07%,宏 F1 值为 85.66%。

表 6 舌象纹理特征分析模型的预测效果

Table 6 Prediction effect of tongue image texture feature analysis model

(单位:%)

Feature	Aurracy	$P_{macro}$	$R_{macro}$	$F1_{macro}$
舌质裂纹	91.29	97.03	88.82	92.74
舌质芒刺	81.63	79.94	90.33	84.82
舌质瘀点	78.79	86.21	80.17	83.08
舌苔腐腻	76.33	77.95	82.30	80.06
舌苔厚薄	87.12	95.40	88.18	91.65
舌苔苔色	77.27	88.67	75.57	81.60
平均值	82.07	87.53	84.23	85.66

本文针对舌象的几何特征和纹理特征的特点提出了相应的迁移学习解决方案,即一个融合空间转换网络和 VGG16 模型的几何特征分类模型以及一个融合深度纹理编码网络和 VGG16 模型的纹理分类识别模型,并使用前一任务中舌体分割模型学到的知识进行参数迁移学习。最后分别对两个模型进行了两组对比实验,实验验证了 STL 和 LREL 对各自的分类模型的有效性,以及迁移学习的知识能够使模型更快更平稳地收敛。

**结束语** 本文针对语义分割模型提出了一种新的基于区

域关联性的单像素损失函数,以显式地指导模型学习局部区域像素间的相关性,使得整个舌体分割模型在测试集上的 MIoU 指标可以达到 96.32%。

本文提出了一个融合空间转换网络和 VGG16 模型的几何特征分类模型来完成对舌象几何特征的识别提取,舌象几何特征分析模型的平均准确率为 82.58%,宏 F1 值为 84.34%。本文在充分分析舌象数据的基础上,根据数据的纹理特征的“无序”特性,提出了一个融合深度纹理编码网络和 VGG16 模型的纹理特征分类模型来完成对舌象纹理特征的识别提取,舌象纹理特征分析模型的平均准确率为 82.07%,宏 F1 值为 85.66%。虽然本文已在初步阶段取得了以上这些研究成果,但仍然存在很多问题,需要继续深入研究,在分类识别多个舌象特征的过程中如果能够融入舌象特征之间的关联关系,理论上预测效果会更好,如何强化这种关联关系并利用它来提升模型的分类识别能力,是后续研究的难点。中医舌象比较复杂,目前还有一些字段未被考虑,且有些舌象特征还可以分区域进行细分,如舌根腐腻、舌边齿痕、舌尖红等,而且目前的舌象信息不够细化,如果需要完成下一步的中医证候推断,从舌象数据采集到舌象分析都还有大量的工作要做。

中医舌诊客观化是一个意义重大并且极具挑战性的科研难题,本文将迁移学习和深度学习技术应用到舌象分析,尝试实现全自动化的舌象分析,希望能对中医舌诊的客观化和标准化工作有所贡献。相信随着计算机技术的发展以及中医现代化理论的成熟,中医诊疗技术的现代化必将推进到一个全新的阶段。

## 参 考 文 献

- [1] XU J T. Clinical map of traditional Chinese medicine tongue diagnosis [M]. BeiJing:Chemical Industry Press,2017.
- [2] ZHOU Z, HUANG F. Application of mathematical morphology method on the segmentationof tongue images [J]. Technological Development of Enterprise,2009,28(3):164-166.
- [3] ZHANG L, QIN J. Tongue-image segmentation based on gray projection and threshold-adaptive method [J]. Chinese Tissue Engineering Research and Clinical Rehabilitation, 2010, 14 (9): 1638-1641.
- [4] WU J, ZHANG Y H, BAI J, et al. Tongue contour image extraction using a watershed transform and an active contour model [J]. Journal of Tsinghua University (Science and Technology), 2008(6):1040-1043.
- [5] ZHANG X Y. Tongue image segmentation based on random walk algorithm [D]. Beijing: Beijing Institute of Technology, 2016.
- [6] HUANG Z P, HUANG Y S, YI F L, et al. An automatic tongue segmentation algorithm based on OTSU and region growing[J]. Lishizhen Medicine and Materia Medica Research, 2017, 28(12):3062-3064.
- [7] LIU Y B, YANG S. New algorithm on image automatic segmentation of body of tongue[J]. Journal of Shenyang Normal University (Natural Science Edition), 2011, 29(4):514-517.
- [8] SUN X L, PANG C Y. An improved snake model method on

- tongue segmentation [J]. Journal of Changchun University of Science and Technology (Natural Science Edition), 2013, 36(5): 154-156.
- [9] WEI B G, SHEN L S. Automatic analysis for plumpness and slenderness of tongue [J]. Computer Engineering, 2004(11): 25-26.
- [10] XU J T, ZHANG Z F, REN H F, et al. An imaging diagnostic method about analyzing for plumpness and slenderness of tongue [J]. Chinese Imaging Journal of Integrated Traditional and Western Medicine, 2009, 7(06): 407-410.
- [11] LU P. Study of teeth-marked tongue based on image processing and pattern recognition [D]. Shanghai: East China University of Science and Technology, 2014.
- [12] ZHAI T T, XIA C M, WANG Y Q. Recognition of Greasy or curdy tongue coating based on gabor wavelet transformation [J]. Computer Applications and Software, 2016, 33(10): 162-166.
- [13] YANG Z H. Dissertation for the doctoral degree in engineering [D]. Harbin: Harbin Institute of Technology, 2010.
- [14] LIU G Z. Research on application of traditional Chinese medicine tongue images classification based on CNN [D]. Jilin: Jilin University, 2018.
- [15] HU J L, KAN H X. Tongue classification based on convolutional neural network [J]. Journal of Anqing Normal University (Natural Science Edition), 2018, 24(4): 44-49.
- [16] SUN L Y, CHENG Z, GAO F S, et al. Discussion on Objective Research of Tongue Diagnosis by Computer Image Recognition Technology [J]. Journal of Anhui Traditional Chinese Medical College, 1986(4): 5-7.
- [17] WANG Y Q, WEI B G, CAI Y H, et al. A knowledge-based arithmetic for automatic tongue segmentation [J]. Acta Electronica Sinica, 2004(3): 489-491.
- [18] GAO L, LING X M. A tongue segmentation method based on improved fuzzy operator and morphology [J]. Journal of Lanzhou Jiaotong University, 2006(3): 89-91.
- [19] LI Q L, XUE Y Q, WANG J Y, et al. Automated tongue segmentation algorithm based on hyperspectral image [J]. Journal of Infrared and Millimeter Waves, 2007(1): 77-80.
- [20] LIU Z, CHEN J X, ZHAO Y M, et al. Automatic tongue image segmentation based on visual attention and support vector machine [J]. Journal of Beijing University of Traditional Chinese Medicine, 2013, 36(1): 18-20.
- [21] WANG P, YANG W C, SUN C K, et al. Tongue segmentation and tongue crack extraction of tongue 3D color point cloud [J]. Infrared and Laser Engineering, 2017, 46(S1): 88-95.
- [22] WANGLR, TANGYP, CHENP, et al. Two-phase convolutional neural network design for tongue segmentation [J]. Journal of Image and Graphics, 2018, 23(10): 1571-1581.
- [23] GONG Y P, CHEN S Z, LIAN Y S, et al. Quantity study on the pathological nature of the tongue fur [J]. Chinese Journal of Information on Traditional Chinese Medicine, 2006(11): 28-29.
- [24] SHENG S Y, LI B, YUE X Q, et al. Color feature extraction of tongue image based on manifold learning [J]. Space Medicine & Medical Engineering, 2008(5): 435-439.
- [25] HU S N. Research on color recognition for tongue image in traditional Chinese medicine [D]. Huangzhou: Zhejiang Sci-Tech University, 2010.
- [26] ZHANG J, QIAN J, DONG H Y, et al. Analysis of traditional Chinese medicine digital tongue texture based on fractal theory [J]. China Journal of Traditional Chinese Medicine and Pharmacy, 2016, 31(1): 104-106.
- [27] XIE T. A new approach to the tongue-image segmentation and moistening analysis based on image processing [D]. Shanghai: East China University of Science and Technology, 2017.
- [28] WANG S, LIU K H, WANG L T. Tongue spots and petechiae recognition and extraction in tongue diagnosis images [J]. Computer Engineering & Science, 2017, 39(6): 1126-1132.
- [29] LIU B, HU G Q, ZHANG X F, et al. An improved automatic description method of tongue coating thickness in Chinese medicine [J]. Beijing Biomedical Engineering, 2018, 37(2): 157-163.
- [30] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C] // IEEE transactions on pattern analysis and machine intelligence. 2018.
- [31] ZJADERBERG M, SIMONYAN K, ZISSERMAN A. Spatial transformer networks [C] // Advances in Neural Information Processing Systems. 2015.
- [32] ZHANG H, XUE J, DANA K. Deep TEN: Texture encoding Network [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017.



**LI Yuan-tong**, born in 1993, postgraduate, is a member of China Computer Federation. His main research interests include deep learning and pervasive computing.



**ZHU Zhen-min**, born in 1962, Ph. D, professor, is a member of China Computer Federation. His main research interests include embedded system technology and pervasive computing.