

中医舌象分割技术研究进展: 方法、性能与展望

卢运西^{1,2} 李晓光^{1,2} 张辉^{1,2} 张菁^{1,2} 卓力^{1,2,3}

摘要 中医舌诊的客观化、量化研究是中医现代化发展中的重要课题。数字化采集到的舌图像包括舌体及部分面部区域, 为了便于后续舌象自动分析, 需要首先将舌体部分从图像中分割出来, 分割效果将直接影响后续舌象特征分析的准确性。基于传统方法的舌象分割技术虽然取得了很大进展, 但其性能仅能达到半自动分割, 对较难分割的图像往往需要借助人机交互来完成。近年来, 深度学习技术在图像处理及计算机视觉等多个领域取得了突破, 其在图像语义分割任务中也取得了远超传统方法的进展。基于深度学习的舌象分割技术已经基本实现了全自动的鲁棒分割。本文首先从传统分割方法和基于深度学习的分割方法两方面对中医舌象分割技术发展中的主要方法进行综述; 其次, 采用我们收集的舌象数据库对典型的方法进行性能评估, 并对不同舌象分割方法的特点进行分析与讨论。最后, 对中医舌图像分割方法潜在的发展方向进行了展望。

关键词 中医舌图像, 语义分割, 迁移学习, 性能评估

引用格式 卢运西, 李晓光, 张辉, 张菁, 卓力. 中医舌象分割技术研究进展: 方法、性能与展望. 自动化学报, 2021, 47(5): 1005–1016

DOI 10.16383/j.aas.c180807

Review on Tongue Image Segmentation Technologies for Traditional Chinese Medicine: Methodologies, Performances and Prospects

LU Yun-Xi^{1,2} LI Xiao-Guang^{1,2} ZHANG Hui^{1,2} ZHANG Jing^{1,2} ZHUO Li^{1,2,3}

Abstract The objectification and quantitative analysis of tongue diagnosis is an important topic in the development of traditional Chinese medicine (TCM) modernization. The digitally acquired tongue images include the tongue and part of the face region. In order to facilitate the automatic analysis of the tongue image, the tongue needs to be segmented from the whole image, and the segmentation results will directly affect the accuracy of the tongue image feature analysis. Although traditional methods of tongue-image segmentation has made great progress, their performance can only achieve semi-automatic segmentation. There are images that difficult to segment perfectly without human-computer interaction. In recent years, with the breakthrough of deep learning technology in the field of image processing and computer vision, it has achieved far more performance than traditional methods in the semantic segmentation tasks. The deep-learning based tongue-image segmentation technologies have achieved fully automatic robust segmentation. This survey gives a detailed overview of the history, state of the art, and typical methods in this domain. Firstly, the typical segmentation methods are presented. Then, they are used for migration learning and network testing based on our self-built tongue image database. In addition, this paper analyzes the characteristics of these segmentation methods and obtains the advantages and disadvantages of them. Finally, this paper summarizes the methods of Chinese medicine tongue image segmentation, and discussed to the development direction.

Key words Chinese medicine tongue image, semantic segmentation, transfer learning, performance evaluation

Citation Lu Yun-Xi, Li Xiao-Guang, Zhang Hui, Zhang Jing, Zhuo Li. Review on tongue image segmentation technologies for traditional Chinese medicine: methodologies, performances and prospects. *Acta Automatica Sinica*, 2021, 47(5): 1005–1016

收稿日期 2018-12-05 录用日期 2019-02-13
Manuscript received December 5, 2018; accepted February 13, 2019

国家自然科学基金面上项目 (61871006) 资助
Supported by National Natural Science Foundation of China (61871006)

本文责任编辑 刘成林
Recommended by Associate Editor LIU Cheng-Lin

1. 北京工业大学信号与信息处理研究室 北京 100124 2. 北京工业大学信息学部微电子学院 北京 100124 3. 北京电动车辆协同创新中心 北京 100081

1. Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing University of Technology, Beijing 100124 2. College of Microelectronics, Faculty of Information

舌诊是中医学望诊的重要内容, 医生可通过观察舌象的变化, 了解人体生理功能和病理变化^[1]. 20 世纪 80 年代以来, 随着计算机技术的不断更新发展, 科研工作者开始将数字图像处理技术应用于舌诊客观化研究当中。采集到的中医舌图像除了舌体以外, 往往还包含脸部等背景区域, 而这些区域对舌诊起不到太大作用。为了避免这些区域对后续分析

Technology, Beijing University of Technology, Beijing 100124
3. Beijing Collaborative Innovation Center of Electric Vehicles, Beijing 100081

处理造成干扰,需要将舌体区域分割出来.同时,舌体自动分割的精确性会直接影响舌象特征分析相关算法的性能,精准的舌图像分割方法是舌诊客观化研究中的一项重要技术.

在舌体分割中,舌体大小和形状的差异以及舌体颜色与嘴唇颜色相近等客观因素大大增加了舌体分割的难度,使得传统的舌图像舌体分割方法在分割自动化程度和分割精度上有待进一步提升.此外,封闭环境具有稳定的光照环境,对舌体分割算法更有利,现有的舌图像自动分割方法主要研究应用于封闭式采集环境;开放式采集环境的舌图像分割问题受到光照和图像质量的影响,则更加难以解决.因此,利用图像处理、机器学习等领域的最新研究算法提升舌图像的分割精度具有重要意义.现有的舌体自动分割方法大致可以分为两类:基于传统技术的分割方法和基于深度学习的分割方法.

本文对基于传统技术的中医舌图像分割方法和基于深度学习的中医舌图像分割方法进行总结归纳,并采用我们临床采集和网络收集的舌象数据集对典型算法进行网络训练和性能评估.同时,我们对这些算法的特点进行分析和讨论.本文结构安排如下:第1节简要回顾基于传统技术的中医舌象分割方法;第2节具体介绍基于深度学习的中医舌象分割技术,着重对目前基于深度学习的中医舌象分割技术及典型方法进行梳理;第3节对深度学习舌象分割方法的关键问题进行分析与讨论;最后,给出结论与展望.

1 基于传统技术的中医舌象分割方法

1.1 基于图像特征的舌体分割方法

基于图像特征的舌体分割方法是利用图像的像素值特征实现的,即目标区域内的相邻像素值具有相似性,而目标区域边界处的相邻像素值则存在间断性或不连续性.根据侧重点的不同,又进一步将这类舌体分割方法划分为区域分割技术、边缘检测技术以及结合特定理论工具的分割技术^[2].

1.1.1 区域和边缘检测分割技术

区域分割是根据区域数值的相似性和空间的接近性,按照特定的评判标准将像素分配到某个区域,进而把这些像素与图像背景区域分开.可分为基于阈值的区域分割技术、特征空间聚类和基于种子点的区域自增长技术等方法. Kim 等^[3]提出了一种区域增长的方法,首先对舌图像进行下采样、直方图均衡化及边缘增强等预处理,再进行过分割操作获得大于目标区域的部分,之后利用区域增长技术实现区域融合,根据检测到的局部最小值确定舌体边缘,最后采用边缘平滑操作实现舌体分割.一种结合亮度和粗糙度信息的舌象分割方法^[4]采用自定义

的亮度信息,利用最大类间方差法自动选取舌图像亮度的阈值,初步提取出舌体较亮的部分,然后利用粗糙度剔除非舌体部分,最后采用数学形态学方法实现舌体分割.基于动态阈值和修正模型的舌体提取算法^[5]利用色调-亮度-饱和度(Hue-intensity-saturation, HIS)色彩模型去除嘴唇和脸部区域,并利用动态阈值分割方法提取舌体初始轮廓,最后运用舌体修正模型得到分割结果.赵忠旭等^[6]将 HIS 色度空间的 H 分量进行图像二值化,利用聚类算法去除非舌的背景区域,然后运用形态学方法去噪,最后得到舌体分割结果. Chen 等^[7]利用基于 Lab 颜色空间的颜色增强算法实现了一种新的舌图像分割方法,该方法将分割速度大大提高,但是存在分割边缘过于粗糙的问题,分割效果图如图 1(a)所示.李丹霞等^[8]提出的基于自适应阈值的舌象分割方法,首先把舌图像分割成多个图像子块,不断迭代计算出每个子块的最优阈值,然后根据局部最优阈值构成的阈值矩阵进行分割,最终实现舌象分割.

蒋依吾等^[9]提出的方法先检测包含舌体的矩形区域,之后进行对比度增强、二值化操作,最后通过边界检测方法得到舌体边缘. Zhi 等^[10]通过人工植入边缘种子点,并利用 B 样条(B-spline)方法拟合函数以得到目标区域的边缘信息提高分割可靠性.文献[11]假设舌根与唇之间总是存在一块黑色阴影区域,利用先验知识搜索局部直方图最优阈值可以准确提取阴影边缘.文献[7]基于 HSV 空间中的色调通道,得到阈值控制函数,然后提取感兴趣的区域,并在区域内使用颜色增强方法.最后,使用 Lab 颜色空间中的亮度特征来获得完整的轮廓. Li 等^[12]使用直方图投影和基于学习的数字模板从医学图像中提取舌体.

基于区域的分割技术和边缘检测的方法大都是利用图像的颜色信息进行特征区域和边缘的划分,但是舌图像较为复杂,脸部、嘴唇和舌头的颜色较为接近,该类方法存在不稳定性,分割准确性有待提升.

1.1.2 特定理论工具分割技术

文献[13]在实现舌体的定位时采用了 Adaboost 方法,并在获取轮廓时结合极坐标变换法和颜色对消法,然后实现舌图像分割.一种基于核模糊聚类的舌图像分割方法^[14],为了实现舌体与背景分离,通过提取舌图像的颜色、位置和纹理等特征并进行聚类. Liu 等^[15]将特征提取的方法应用到舌体区域提取,并利用支持向量机(Support vector machine, SVM)算法在多光谱图像中提取舌体,结果如图 1(b)所示.文献[16]利用直方图投影和基于学习的数字抠图方法对舌体进行提取,但是该方

法在光照不均匀时难以获得准确的舌图像分割结果, 分割效果不理想. 文献 [17] 根据颜色和空间的相似性采用均值漂移聚类的方法对图像内容进行分类, 然后将主成分分析算法应用到舌体检测算法, 利用投票策略实现分割. 文献 [18] 针对舌体的特点设计了双椭圆形变轮廓方法, 该方法针对舌体形状特意设计, 通过在参数空间中的能量函数来捕获粗体形状特征, 能够适应局部细节, 提高了舌图像分割的准确率. Shi 等^[19] 通过显著性窗口来细化临床舌象, 将舌区域初始化为上部分二值化模板和下部分水平集矩阵. 应用双重矢量流的方法来检测舌体边缘并在图像中分割舌体区域.

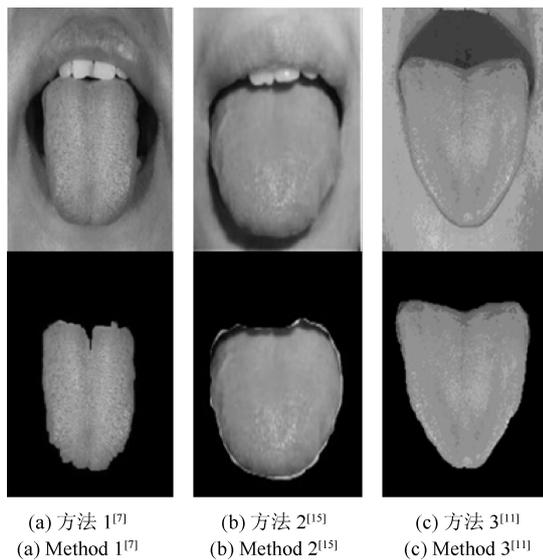


图1 几种基于图像特征的舌体分割方法分割效果
Fig.1 The results of several traditional algorithms

1.2 基于可变模型的分割方法

20世纪80年代中期, Kass等^[20]提出了二维的可变模型的概念, 又称为Snakes或动态轮廓模型. 该方法通过对能量函数的动态优化来逼近目标的真实轮廓, 其特点是结合了图像的高层和底层特征, 在生物医学图像分割中得到广泛使用^[21-22]. 在用于图像分割时, 其稳定性、精确度等方面均优于图像特征的分割方法^[23].

王爱民和沈兰荪^[24-25]提出了Catmull-Rom样条Snakes模型, 并将其应用于舌体自动分割. McInerney等^[26]提出了拓扑自适应的Snakes模型, 能够得到具有分支或由若干闭合曲线构成的复杂轮廓, 实现图像分割. 一种基于先验知识的自动舌体分割算法^[27]利用舌体的位置、颜色等先验信息, 通过Snakes模型得到舌体区域. 文献 [28] 提出了一种舌体边缘检测和梯度矢量流 (Gradient vector flow Snakes, GVF-Snakes) 动态轮廓线相结合的全

自动舌体分割方法. 文献 [29] 提出一种基于分水岭变换和主动轮廓模型的舌体分割方法, 该方法通过降采样的方法来提高分割速度. Yu等^[11]根据舌图像先验知识检测舌体的初始边界, 然后将彩色梯度信息引入GVF-Snakes方法, 将舌体区域分割出来, 结果如图1(c)所示. Zhai^[30]采用中值滤波去除图像中的噪声, 然后将图像映射到HIS色度空间, 利用对偶Snakes算法获取舌体轮廓, 实现舌图像分割. 一种基于初始化Snake轮廓线的混合舌图像分割算法^[31]为采用双层极坐标边缘检测算法获取舌体的粗略边缘, 之后采用Snakes模型修正得到准确的舌体边缘, 将舌体与背景信息分离. Pang等^[32]的分割方法中采用了一个双椭圆变形模板, 结合主动轮廓模型将舌体分割出来. Guo等^[33]提出了基于自适应Snakes模型的舌图像分割方法, 但是该方法在舌体边界初始化时容易受到唇部的干扰, 需要进行人工干预. 孙晓琳等^[34]使用转换颜色空间和优化的Snakes模型实现舌体分割, 提高了算法稳定性和准确性. 王明英等^[35]采用两次Snakes动态轮廓模型, 提高了舌体分割方法的分割精度. Zhang等^[36]提出了一种结合极边检测器和主动轮廓模型技术的新型自动舌分割方法, 使用一种极边检测器, 以有效地提取舌体边缘. 然后, 设计了一种边缘滤波方案, 并引入局部自适应边缘双阈值算法以执行边缘二值化. 最后, 使用启发式初始化和主动轮廓模型来从图像中分割舌体.

基于可变模型的分割方法大多采用Snakes算法, 需要指定初始区域, 然后使用Snakes进行精细分割. 但是初始区域的选取方法有时候会存在较大的误差, 或者较为复杂, 使得算法的实用性大大降低.

基于传统技术的中医舌象分割方法利用图像像素值特性和一些特定的算法实现了舌图像分割, 但这些算法的稳定性和性能多难以达到实际应用需求, 这些算法需要人为辅助, 自动化效果较差, 算法运行速度较慢. 绝大多数现有的舌图像分割方法是针对封闭式采集环境提出的, 当面对复杂的开放式环境时, 算法的鲁棒性较差. 因此, 利用新的图像处理技术提升分割方法的鲁棒性具有重要的理论意义和实际应用价值.

2 基于深度学习的分割方法

近年来, 深度学习在语义分割等计算机视觉领域取得了显著进步. 其中, 卷积神经网络 (Convolutional neural network, CNN)^[37] 凭借其强大的特征学习与表达能力广泛应用于图像语义分割. 这些方法大部分不是专门为中医舌图像分割问题而设计的, 同时基于深度学习舌图像分割方法非常少.

舌象分割问题与图像语义分割问题具有相似性. 语义分割也称为全像素语义分割, 对图像中的每个像素标注类别标签, 以此来识别图像中存在的内容以及位置. 舌象分割与自然图像语义分割相比具有以下特点: 1) 舌体表面反映很多病理细节, 对边缘提取有较大的影响, 如舌边缘可能存在齿痕, 舌体可能存在裂痕等. 2) 不同疾病、不同人采集到的舌体的形态差异大, 难以采用预先定义的变形模板解决舌象鲁棒分割问题; 3) 舌体颜色与唇色、肤色颜色相近, 灰度值也相近, 较难区分; 4) 舌体根部边缘不清晰, 单纯依靠边缘信息难以准确确定根部边缘. 因此, 我们可以借鉴深度学习在语义分割中取得突破性进展的先进技术解决舌象分割问题, 但又不能直接生搬硬套, 需要采用舌象分割数据集对现有网络进行训练微调, 以适应舌象分割的具体应用.

本节首先介绍典型的基于深度网络的语义分割方法, 然后阐述我们对其进行针对舌图像分割问题的迁移学习训练方法.

2.1 基于编解码网络框架的分割方法

编解码网络通常采用“沙漏型”网络结构, 网络分为编码部分和解码部分. 用于语义分割的网络通常由这种典型的编解码网络结构演化而来, 图像输入到传统 CNN 中, 经过卷积和池化后, 特征图的分辨率通常会不断降低, 同时其图像的感受野会不断增加, 但是由于图像分割预测是逐像素输出的, 所以将 CNN 输出的较小的特征图进行上采样到原始图像尺寸进行预测. 在编码过程中, 全卷积网络通常会逐层降低图像空间分辨率, 产生较为粗糙的分割结

果; 在解码过程中, 将低分辨率的结果上采样到原图大小. 现有大部分先进的分割网络都有相同或相似的编码结构, 之所以会产生不同精度的分割结果, 关键在于解码网络结构的不同.

全卷积分割网络 (Fully convolutional networks for semantic segmentation, FCN)^[38] 将传统 CNN 中的全连接层转化为多个并行的卷积层. 如图 2 CNN 卷积化所示, 在传统的 CNN 结构中, 前 5 层为卷积层, 第 6 层和第 7 层分别是一个长度为 4096 的一维向量, 第 8 层是长度为 C 的一维向量, C 为分类的类别数量, 此处 C 为 2, 对应 2 个类别即舌体区域和背景区域的概率. FCN 将这 3 层替换为卷积层, 卷积核的大小 (通道数, 宽, 高) 分别为 (4096, 1, 1)、(4096, 1, 1)、(2, 1, 1). 网络中解码部分所有的层都是卷积层, 之后对特征图进行上采样, 得到与输入图像大小相同的特征图, 因此称其为全卷积网络. 同时, 网络结构中加入深层和浅层的跳层连接, 用来促进梯度传递, 提高训练的有效性. 如图 3 所示, 其中, 不加跳层连接、并且输出图特征直接进行上采样和预测的网络结构称为 FCN32S; 在第 4 层特征和网络输出结果加入跳层连接, 进行特征融合, 然后进行上采样和分割预测的网络结构称为 FCN16S; 在第 3 层特征和 FCN16S 输出结果之间加入跳层连接, 特征融合后进行上采样和分割预测的网络结构为 FCN8S. Mask R-CNN (Region CNN)^[39] 在 Faster R-CNN^[40] 的基础上使用 FCN 增加了分割的分支实现了目标检测、目标分类识别和像素级目标分割.

SegNet 为用于图像分割的深度卷积编码器-解

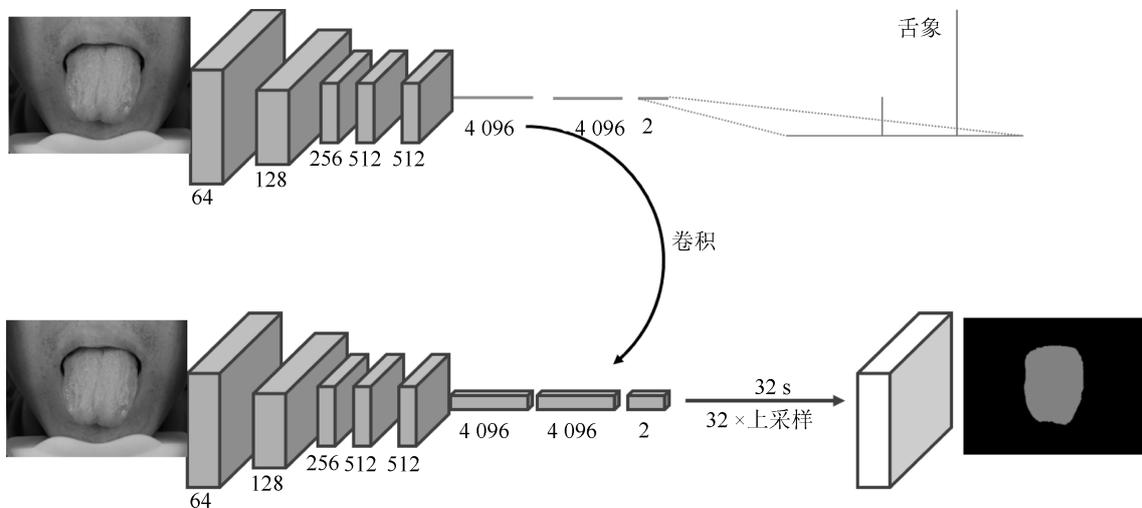


图 2 卷积神经网络卷积化

Fig. 2 Convolutionalization of CNN

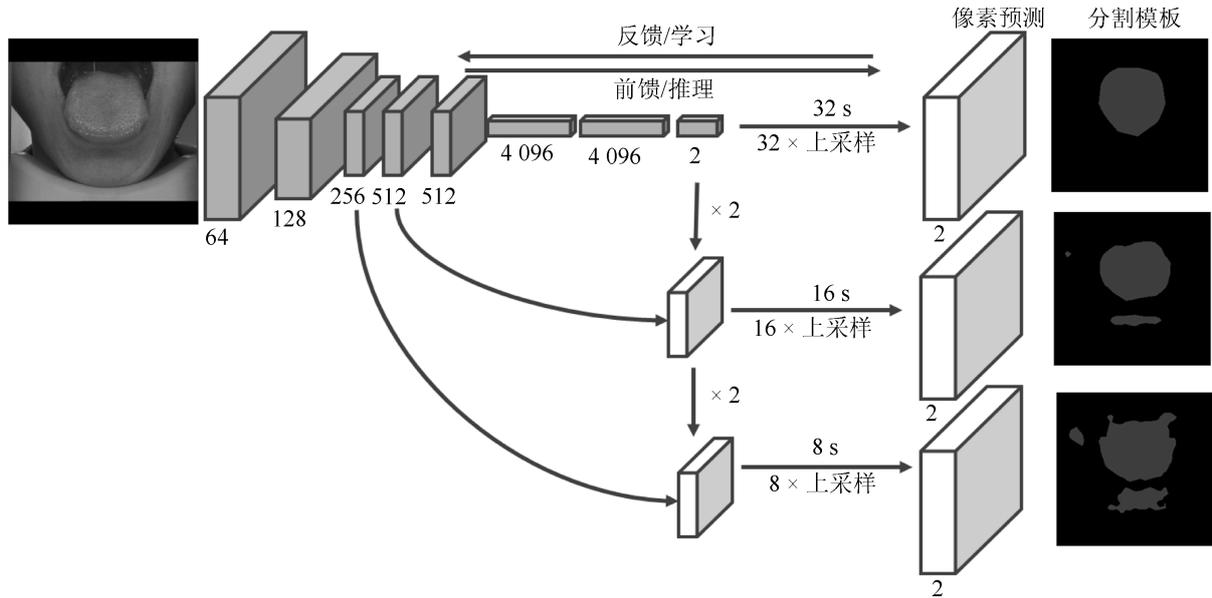


图 3 全卷积网络结构图
Fig. 3 The architecture of FCN

码器结构^[41], 如图 4 所示, 同样采用编解码的网络结构, 编码网络用于提取特征, 解码网络采用上采样的方法丰富输出的图像信息. Qu^[42] 等将 SegNet 成功应用到了舌图像分割当中, 并利用小样本实现网络的训练. 类似的分割网络 U-net^[43], 其编码网络结构采用卷积和最大池化来实现的特征提取, 并且在下采样过程中逐次增加特征通道; 解码网络部分对编码网络的输出特征图进行拷贝和剪裁, 并通过反卷积获得丰富的分割细节, 这样的结构使得在上采样过程中网络可以传播到更高分辨率的层. 同样采用编解码网结构的还有 2018 年初 Google 公司提出的 DeepLab V3+^[44], 该网络为了融合多尺度信息引入语义分割常用的编解码网络, 引入可任意控制编码器提取特征的分辨率, 通过 Atrous 卷积平衡精度和耗时.

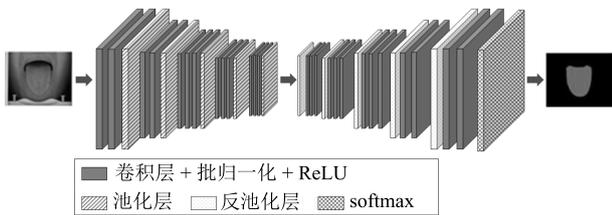


图 4 SegNet 网络结构图
Fig. 4 The architecture of SegNet

编解码网络框架结构中, 通过上采样到原始的图像尺寸进行预测. 编码-解码网络框架结构网络框架通过对图像的下采样获得高层语义特征, 将获得图像特征上采样到原始图像大小特征进行图像预

测.

2.2 基于 Atrous 卷积的图像分割方法

像素级别的图像分割问题是 Low-level 的问题, 从直觉上讲可能不需要深度网络提取抽象的高层语义信息. 但是, 需要考虑不同大小目标的多尺度分割问题. 为了避免卷积网络特征提取获得高层语义信息, 同时兼顾不同的尺度分割问题, Chen 等^[45] 在 Deeplab 中引入了空洞滤波的概念. 他们在卷积核里增加空洞, 也就是 Atrous 卷积. 其基本思想通过在卷积滤波器中插入零值来调整单层卷积层滤波器感受野的大小, 在不增加权值参数数量的情况下适应不同尺度目标的分割问题. 多分辨率的感受野可以更为有效地学习不同尺度目标的有效信息, 从而进一步提高目标的分割精度. 由于 Atrous 卷积良好特性, PSPNet^[46] 网络同样在卷积网络中使用了 Atrous 卷积, 提高网络性能.

Deeplab V2^[45] 将 Atrous 卷积应用到空间金字塔中, 增加更多的多视野区域, 如图 5 所示, 该方法分别使用 Atrous 卷积的孔径系数参数 Rate 为 6、12、18 和 24 的 4 个 3×3 卷积, 构成基于 Atrous 卷积的空间金字塔池化 (Atrous spatial pyramid pooling, ASPP). 同样 Deeplab V3^[47] 将 Atrous 卷积应用到基础网络和空间金字塔中, 其中基础网络中, 取 ResNet^[48] 最后一个 Block, 将其复制多个, 级联起来, 并将 Atrous 卷积应用到级联模块中. 这一结构在 DeeplabV3+ 中同样被采用.

图像分割网络的基础网络通过连续的下采样整合多尺度的内容信息, 降低图像分辨率. 卷积能够整

合多尺度内容信息,且不损失分辨率.在 Deeplab 引入 Atrous 卷积后,Deeplab V2、V3、V3+ 网络均采用了 Atrous 卷积,在不增加参数量的前提下保护了感受野及特征分辨率.引入 Atrous 卷积,图像语义分割效果有很大的提高,具体实验结果在本文后续部分进行详细讨论.

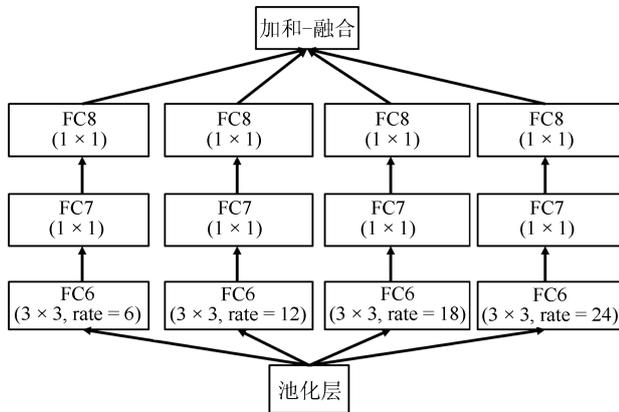


图 5 基于 Atrous 卷积的空间金字塔池化结构图
Fig. 5 The architecture of ASPP

2.3 空间金字塔池化的网络框架

深度神经网络一般都需要固定的输入图像尺寸,潜在性的弊端会降低识别精度.在该过程中,为了使图像尺寸相同,一定会涉及到图像的比例、非比例缩放,这就引入了尺度误差和形变误差.在解决分割物体多尺度的问题中,空间金字塔池化^[49] (Spatial pyramid pooling, SPP) 是一种非常有效的多分辨率策略,对目标形变等问题具有很好的鲁棒性. PSP-Net^[46] 使用金字塔池化模块能够聚合不同区域特征的上下文信息,从而提高获取全局信息的能力,有效提高了分割的精度.

Chen 等^[45] 提出的基于孔洞的空间金字塔模块,能够获得多尺度的特征信息,使得分割结果更具鲁棒性.基于孔洞的空间金字塔并行的采用多个采样率的 Atrous 卷积层获得不同特征,以多个比例获取对象以及图像上下文信息.但是随着采样率的增加,滤波器的有效权重逐渐变小.为了克服这个问题,并将全局上下文信息纳入模型,Deeplab V3 在模型的最后一个特征图采用全局平均池化,将重新生成的图像级别的特征提供给带 256 个滤波器和批归一化 (Batch normalization)^[50] 的 1×1 卷积,然后通过双线性插值将特征提升到所需的空间维度.改进的结构如图 6 所示.

空间金字塔池化的网络框架,使得任意大小的特征图像能够转换成固定大小的特征向量,之后送入后面的分割预测.同时,该结构能够有效解决输入图像大小不一的情况,提高了分割网络的精度和速

度.针对不同问题,优化的 ASPP 和改进的 ASPP 使得图像分割的精度不断提高.

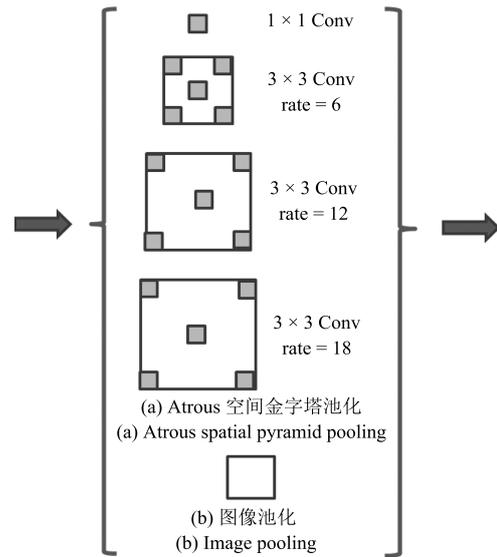


图 6 改进的基于 Atrous 卷积的空间金字塔池化结构图
Fig. 6 The architecture of the improved ASPP

2.4 面向舌象分割的迁移学习

经典的基于深度学习的语义分割网络没有针对特定种类的分割任务设计,算法训练和测试均是在公开数据集上进行.中医舌象分割有自身特点和要求.将经典算法应用到舌图像分割当中应充分考虑舌图分割的任务特点,其对精度和边缘信息要求更加严格.

2.4.1 训练样本的数据扩充

在深度学习的模型训练中,训练样本数据量大小对深度模型的性能有着直接的影响,如果样本数少,会造成模型的过拟合.由于本文构建的舌图像样本数据集 (Tongue dataset) 的样本数据有限,为了防止深度模型过拟合,本文采用了常用的方法来扩充样本数据量,即旋转、平移和水平镜像.其中,旋转采用在 $-70^\circ \sim +70^\circ$ 范围内随机产生不同角度的变化,平移采用了上、下、左、右四个方向的移动.采用该方式我们得到扩充后的舌图像和标签图像,两者作为图像对被送入深度网络中进行微调训练.

2.4.2 基于迁移学习的网络训练

深度学习需要大量的数据,经过上述的数据扩充在一定程度上改善了数据库数据量较少的问题,为了进一步提高训练网络的精度,迁移学习的策略可以让网络从 ImageNet 的训练模型中,简单地提取网络权重的初级特征,应用在中医舌图像分割任务中.同时,经过迁移学习可以避免网络过拟合,加快网络训练.迁移学习有两种策略:微调 (Fine tun-

ing) 和冻结与训练 (Freeze and train). 微调包括使用基础数据集上的预训练网络以及在目标数据集中训练所有层; 冻结与训练包括冻结除用于分类的全连层的所有层 (权重不更新) 并训练最后一层. 也可以冻结前面几层, 然后微调其余层. 我们使用微调的训练策略, 使用 ImageNet 数据集的预训练模型对网络初权值进行初始化, 在舌图像分割数据集中训练所有层, 从而使各个网络更好地应用到中医舌图像语义分割任务中来. 同时, 我们将网络最后一层的输出类别的改为 2, 即舌体区域和背景区域. 使用舌象数据对各个分割网络进行训练, 当网络损失曲线和分割精度曲线趋于平稳后停止训练.

3 分割性能评估与分析

经历了近 20 年的舌象分析研究, 对传统的舌象分割方法, 文献 [1] 的结论为“鉴于舌体分割问题的复杂性, 至今尚无哪种自动分割算法能对各类舌图像都取得理想的分割结果, 在有些场合还必须为用户提供人机交互的方法来提取舌体, 以保证系统的可靠运行”. 文献 [51] 集中阐述了香港理工大学张大鹏 (David Zhang) 教授课题组的舌象分割成果, 从文献中可得出结论: 通常, 除了主要的舌体, 捕获的舌头图像包含很多其他无关的信息. 因此, 为了提高舌象分析的准确性, 研究人员针对不同的任务对现有分割算法进行改进, 使其适合中医舌图像分割任务. 经过改进算法可以达到可接受的性能.

虽然个别的传统舌象分割方法取得了不错的结果, 但基本上都是在特定封闭光学环境和规定舌象姿态下完成的, 对开放环境下任意舌体的分割极少涉及. 因此, 本节将重点介绍基于深度语义分割网络迁移学习的舌象分割技术的结果测评与分析. 采用深度学习方法的舌图像语义分割算法, 无需人为选取舌图像特征, 通过网络训练获得分割模型, 测试阶段采用模型进行预测, 实现图像的精确分割. 为了更好地了解各个模型的分割特点, 我们对其中几种经典算法进行迁移学习, 测试评估, 分析各自分割性能.

3.1 数据集

目前, 国内外没有标准的用于中医舌图像自动分割的数据库. 我们根据不同环境下中医舌象仪研究的实际需求, 在开放式采集环境下构建了舌图像数据集 Tongue dataset¹. 该数据集由 3000 幅舌图像构成, 由互联网图像和利用手机、相机等不同图像采集设备在不同的环境和时间段拍摄得到照片组成. 因此, Tongue dataset 中的舌图像具有尺寸大小不一, 光照环境复杂多变, 舌体的大小、形状各异和位

置不固定等特点. 舌图像数据集 Tongue dataset 部分图像样例如图 7 所示.



图 7 Tongue dataset 中的部分舌图像

Fig. 7 Some pictures of the tongue dataset

我们采用这些数据训练模型, 以便提高模型对光照条件、图像大小、背景信息等各种因素变化的泛化能力. 同时, 我们对数据集中的舌图像进行像素级的标注, 分为背景区域和舌体区域. 标注后得到与原图大小相同的标签, 得到用于深度学习微调训练的数据.

3.2 评价指标

对于分割质量评价研究, 大多偏重于客观的评价, 借助于求分割后的图像与参考图像之间的差别来评判分割算法的优劣, 文献 [52] 提出了 3 种经典的评价方法: 基于轮廓的评价、基于区域的评价和基于参数的评价. 基于轮廓的评价方法使用距离分布函数和轮廓加权误分率测量精度; 基于区域的评价方法通过区域数量、位置和大小来评价分割的准确性; 基于参数的评价方法通过理想分割结果的连通域参数和分割结果的参数之间的差值来表示分割的质量. 文献 [24] 将分割质量评定为 3 个级别: 分割结果非常准确、比较准确和偏差较大, 通过各级图像所占的百分比来衡量分割算法的优劣. 该方法主观性较强, 误差较大.

对舌体分割正确率客观评价, 本文采用语义分割任务中常用的指标平均交叠率 (Mean intersection over Union, mIoU). 该方法相对文献 [52] 计算量较小, 测量精度相对文献 [24] 更加客观, 同时也符合主客观评价的一致性. 该指标的计算分别如式 (1) 和式 (2) 所示.

$$mIoU = \frac{1}{n_{cl}} \sum_i \frac{n_{ii}}{t_i + \sum_j n_{ji} - n_{ii}} \quad (1)$$

¹自建舌象数据库经过整理后将适时对外公开, 具体详情可邮件联系通信作者.

$$t_i = \sum_j n_{ij} \quad (2)$$

其中, n_{ji} 表示被分为 j 类的第 i 类像素的个数, n_{cl} 表示总的像素类别个数, 本文的像素类别个数为 2, t_i 表示第 i 类的像素总数.

在研究中, 有时候增加网络的复杂度会提高模型的精度, 但是在算法落地过程中, 模型预测速度同样十分重要. 因此, 我们除了计算各个模型的精度指标 mIoU 之外, 还对模型预测速度进行测量, 从模型的效率方面对方法进行性能评测.

实验中, 我们使用搭载有 TITAN X Pascal 显卡, 内存为 16 GB, 处理器为 Intel i7 处理器, 64 位 Linux 系统操作平台进行分割网络训练和性能测试.

3.3 实验结果与分析

在中医自动化分析中, 由于舌图像颜色呈现与光照关系很大, 对中医舌图像分析的算法基本上是用封闭环境下的中医舌图像. 在测试过程中, 为了测试网络的在实际的应用的效果, 选取了封闭环境下采集到的舌图像进行测试. 我们选取了 4 幅典型的舌图像进行展示. 图 8 中所示的舌图像为未分割的舌图像, 第 1 幅图为较正常的舌图像, 其他的舌图像存在常见的齿痕、裂纹、红点舌等问题. 本文主要对比了 FCN8S, FCN16S, FCN32S, SegNet, Mask R-CNN, Deeplab V2, Deeplab V3 和 Deeplab V3+ 等方法, 接下来将从方法的主观和客观两方面进行分析.



图 8 测试舌图像 (彩色图像见网络版)

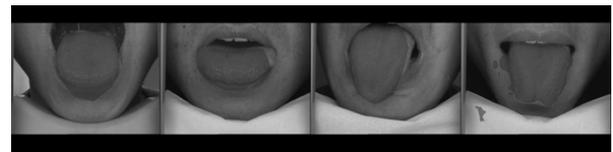
Fig. 8 Pictures of test dataset (Refer to the internet version for color images)

3.3.1 客观分割结果分析

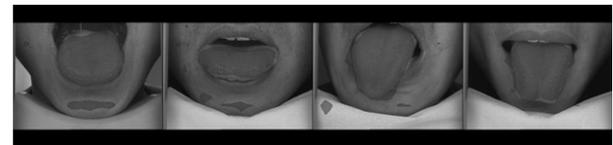
本文对 FCN8S, FCN16S, FCN32S, SegNet, Mask R-CNN, Deeplab V2, Deeplab V3 和 Deeplab V3+ 分割结果进行可视化. 为了更好地展示分割的效果, 本文采用分割结果图和原图进行叠加的方式对分割效果进行可视化, 结果图 9 所示 (彩色图参见本文网络版).

在 FCN 系列中, 如图 9 (a)~(c) 中所示, 从可视化结果图可以看出, FCN32 效果最好, FCN16S 次之, FCN8S 效果最差. 在 3 种方法分割结果中, FCN32S 分割结果中含有最少的分割噪声, FCN8S 的噪声最多. 网络结构上, FCN32S 加入跳层连接,

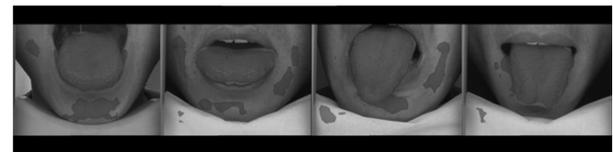
得到 FCN16S 和 FCN8S. 在文献 [38] 中, FCN16S 和 FCN8S 在 PASCAL VOC 等数据集上的分割效果要好于 FCN32S, 但是在舌图像语义分割任务中,



(a) FCN32S



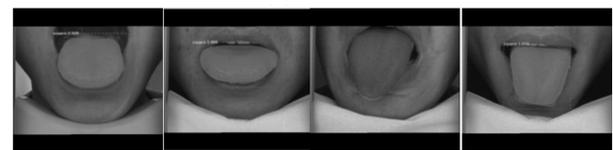
(b) FCN16S



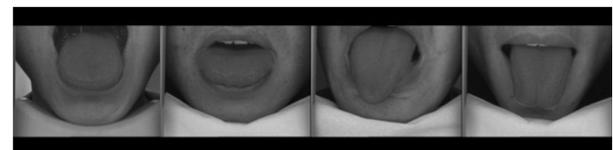
(c) FCN8S



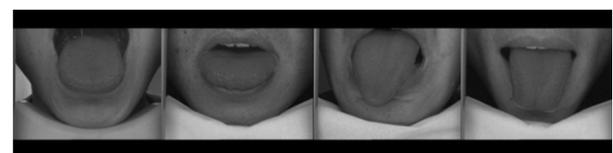
(d) SegNet



(e) Mask R-CNN



(f) Deeplab V2



(g) Deeplab V3



(h) Deeplab V3+

图 9 不同分割算法的分割效果 (彩色图像见网络版)

Fig. 9 The results of different segmentation algorithms (Refer to the internet version for color images)

却未能提升舌体分割的精度. 随着网络结构中的跳层连接增加, 分割结果图中的噪声不断增加, 跳层连接最多的 FCN8S 分割噪声最多. 这可能与图像语义分割类别数量有关, 中医舌图像语义分割只有两类, 即背景和舌体区域, 相对 PASCAL VOC 目标类别较少, 而且舌体在图像中占有的比例较大, 不需要更多的底层信息来丰富特征信息. 除此之外, 也可以观察到随着跳层连接的加入, 除去噪声的 Mask 部分舌体的预测确实在提升, FCN32S 预测结果中, 一些舌体区域没有被预测为舌体, FCN16S 和 FCN32S 的结果不断提升. 从另一个角度来说, 跳层连接使得预测结果融合浅层网络的信息能够提升模型预测性能.

SegNet 网络与 FCN 系列网络相比, 预测结果的噪声大幅度减少, 但是舌体边缘部分会存在一些粘连的噪声, 同时呈现出毛糙的边缘部分. 这对舌体分割来说影响较大, 因为毛糙的边缘会造成舌体存在齿痕的假象, 这对舌体分析造成较大的负面影响.

观察图 9 可以看出, Mask R-CNN 分割结果没有分割噪声和粗糙的分割边缘, 客观分割效果相对 FCN 和 SegNet 有提升. 仔细观察, 会发现分割结果多会存在预测舌体区域小于实际舌体区域的欠分割效果. 因此, 该方法丢失了舌体的边缘信息, 与实际存在一定的偏差.

最后, 在 Deeplab 系列网络的分割结果中, Deeplab V2 分割效果能够减少 Mask R-CNN 分割算法中出现的欠分割问题, Deeplab V3 的分割结果如图 9 所示, Deeplab V3 的主观结果将 Deeplab V2 中的欠分割的问题大大改善. Deeplab V3+ 在此基础上进一步优化, 在 Deeplab V3 的分割结果图中会有微小的欠分割的情况, Deeplab V3+ 则将该问题几乎完全解决, 而且单幅图像的预测时间减少了近一半.

8 种主要的深度学习方法在中医舌图像分割主观效果中, FCN 系列的跳层连接方法未能更好地提高分割性能. SegNet 编码解码网络结构分割处理分割边缘的性能有待提升. 不同于其他的深度学习方法, Deeplab 系列网络采用 Atrous 卷积和基于空间金字塔网络结构, 舌体分割结果没有出现噪声问题. Deeplab V3 对基于空间金字塔模块的改进, Deeplab V3+ 引入编解码的网络结构等方法, 该系列方法不断对网络结构进行调整, 使得分割效果不断提升.

3.3.2 客观分割结果分析

为了从客观上对上述方法进行比较, 我们对各个方法的客观分割精度进行测量. 同时, 对各个方法的模型测试时间进行统计. 表 1 给出了 8 种比较方

法的客观分割精度和单幅图像预测时间.

表 1 不同舌象分割方法性能比较
Table 1 Comparison on performances of different algorithms

网络名称	mIoU	预测时间 (s)
FCN8S	0.8322	7.4227
FCN16S	0.8718	7.4403
FCN32S	0.9272	7.5273
SegNet	0.9277	0.0014
Mask R-CNN	0.9361	0.1412
Deeplab V2	0.9678	0.0021
DeeplabV3	0.9810	4.9999×10^{-6}
DeeplabV3+	0.9818	2.9999×10^{-6}

FCN 系列在加入了跳层连接后, 单幅图像的预测时间变化不大, 但是客观分割精度 mIoU 和主观效果图一致, FCN32S 的 mIoU 值最高. SegNet 与 FCN 系列不同, 在解码模块进行改进, 精度和 FCN32S 相当, 但预测时间有明显的降低. Mask R-CNN 分割精度相对 FCN 系列和 SegNet 有所提高, 但预测时间相对 SegNet 有所增加. 可看到 Deeplab 系列的分割方法在分割精度和预测时间上都取得了不错的结果. Deeplab 系列分割精度较前几种方法来说有大幅度提高, 并且单幅图像的预测时间有大幅度减少.

在中医舌图像分割的客观评价中, 主观评价分割效果的结论在 mIoU 指标上进一步得到验证. 网络单幅图像测试速度方面可以发现, FCN 系列网络的跳层连接对网络速度影响不大, 与其他算法相比运行时间较长. 同样采用 FCN 结构的 Mask R-CNN 的单幅图像测试速度也比较长. 主观分割效果较好的 Deeplab 系列网络, 每一代的改进在提升分割效果的同时, 提高了网络的运行速度, 这与网络使用新的网络结构和训练策略有关.

4 总结与展望

通过本文综述, 我们梳理了中医舌象分割问题的研究最新进展. 我们得出的结论是: 目前基于深度学习的舌象分割方法克服了舌形态、光照、颜色多种因素的舌象分割挑战, 能够实现开放环境静态舌象的精准快速分割, 基本解决了困扰舌象分割多年的全自动鲁棒舌象分割的瓶颈问题. 这对中医舌象仪的自动化分析提供了传统方法难以达到的进展. 本论文以综述的形式汇报该领域的最新进展, 为相关研究人员提供参考. 在此基础上, 我们对中医舌象分割技术的发展提出了新的展望.

1) 舌象动态分割. 舌象的动态特征, 如患者伸舌过程中呈现的颤抖、舌体伸出的幅度等特征也为舌诊提供了重要的信息. 目前的中医舌象分析系统很少对动态特征的采集和分析开展研究. 精准鲁棒的深度网络静态舌象分割方法, 为舌体动态视频的分割提供了基础. 其中关键问题是如何在舌象视频中如何准确捕获舌体颤抖等动态信息.

2) 开放环境下的舌象分析研究. 基于深度网络的舌象分割方法不仅实现了封闭环境下的舌象精准鲁棒分割, 在开放环境下的舌象分割任务中, 也取得了令人满意的结果. 这为进一步开展开放环境下的舌象分析研究提供了基础.

3) 针对舌象分割任务, 设计并优化专门的深度网络结构. 本文所涉及的语义分割网络并非专门为舌象分割任务设计, 因此可能存在网络参数冗余等问题; 另外, 网络的训练和测试过程都需要借助图形计算卡等计算资源. 这在一定程度上限制了舌象分析仪的灵活性. 因此, 针对舌象分割任务, 设计专门的网络结构, 并进行面向普通计算机或嵌入式系统的算法优化, 也是中医舌象仪发展需要开展的重要工作.

总之, 深度神经网络在中医舌图像分割中已初步展现了巨大的潜力和很好的效果. 但是, 仍有很多尚未完善的工作, 需要更多的研究者开展富有创新性的工作.

References

- Shen Lan-Sun, Cai Yi-Heng, Zhang Xin-Feng. *Capture and Analysis of Chinese Medicine Tongue Images*. Beijing: Beijing University of Technology Press, 2007. (沈兰荪, 蔡轶珩, 张新峰. 中医舌象的采集与分析. 北京: 北京工业大学出版社, 2007.)
- Li Guo-Zheng, Shi Miao-Jing, Li Fu-Feng, Wang Yi-Qin. An empirical study on tongue image detection. *Journal of Shandong University (Engineering Science)*, 2010, **40**(5): 87–95 (李国正, 史淼晶, 李福凤, 王忆勤. 舌体图像分割技术的实验分析与改进. 山东大学学报(工学版), 2010, **40**(5): 87–95)
- Kim K H, Do J H, Ryu H, Kim J Y. Tongue diagnosis method for extraction of effective region and classification of tongue coating. In: Proceedings of the 1st Workshops on Image Processing Theory, Tools and Applications. Sousse, Tunisia: IEEE, 2008. 1–7
- Wu Wen-Jun, Ma Li-Zhuang, Xiao Xue-Zhong. Method of tongue image segmentation based on luminance and roughness information. *Journal of System Simulation*, 2006, **18**(S1): 374–376, 379 (邬文隽, 马利庄, 肖雪中. 一种结合亮度和粗糙度信息的舌像分割方法. 系统仿真学报, 2006, **18**(S1): 374–376, 379)
- Zhang Zhi-Shun, Liu Yong. Tongue extraction algorithm based on dynamic thresholds and correction model. *Computer and Modernization*, 2014, (11): 49–52 (张志顺, 刘勇. 基于动态阈值和修正模型的舌体提取算法. 计算机与现代化, 2014, (11): 49–52)
- Zhao Zhong-Xu, Wang Ai-Min, Shen Lan-Sun. The color tongue image segmentation based on mathematical morphology and HIS model. *Journal of Beijing Polytechnic University*, 1999, **25**(2): 67–71 (赵忠旭, 王爱民, 沈兰荪. 基于数学形态学和 HIS 模型的彩色舌图像分割. 北京工业大学学报, 1999, **25**(2): 67–71)
- Chen L, Wang D Y, Liu Y Q, Gao X H, Shang H L. A novel automatic tongue image segmentation algorithm: Color enhancement method based on $L^*a^*b^*$ color space. In: Proceedings of the 2015 IEEE International Conference on Bioinformatics and Biomedicine. Washington, DC, USA: IEEE, 2015. 990–993
- Li Dan-Xia, Wei Yu-Ke. Tongue image segmentation method based on adaptive thresholds. *Computer Technology and Development*, 2011, **21**(9): 63–65 (李丹霞, 韦玉科. 基于自适应阈值的舌像分割方法. 计算机技术与发展, 2011, **21**(9): 63–65)
- Jiang Yi-Wu, Chen Jian-Zhong, Zhang Heng-Hong. Computerized Chinese medicine tongue diagnosis system. *Chinese Journal of Integrated Traditional and Western Medicine*, 2000, **20**(2): 145–147 (蒋依吾, 陈建仲, 张恒鸿. 电脑化中医舌诊系统. 中国中西医结合杂志, 2000, **20**(2): 145–147)
- Zhi L, Yan J Q, Zhou T, Tang Q L. Tongue shape detection based on b-spline. In: Proceedings of the 2006 International Conference on Machine Learning and Cybernetics. Dalian, China: IEEE, 2006. 3829–3832
- Yu S Y, Yang J, Wang Y G, Zhang Y. Color active contour models based tongue segmentation in traditional Chinese medicine. In: Proceedings of the 1st International Conference on Bioinformatics and Biomedical Engineering. Wuhan, China: IEEE, 2007. 1065–1068
- Li X, Li J, Wang D. Automatic tongue image segmentation based on histogram projection and matting. In: Proceedings of the 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2014: 76–81
- Fu Z C, Li W, Li X Q, Li F F, Wang Y Q. Automatic tongue location and segmentation. In: Proceedings of the 2008 International Conference on Audio, Language, and Image Processing. Shanghai, China: IEEE, 2008. 1050–1055
- Xie T, Xia C M, Chen F F, Zhang S L, Zhang Y. A method of tongue image segmentation based on kernel FCM. In: Proceedings of the 9th International Congress on Image and Signal Processing, BioMedical Engineering, and Informatics. Datong, China: IEEE, 2016. 319–324
- Liu Z, Yan J Q, Zhang D, Li Q L. Automated tongue segmentation in hyperspectral images for medicine. *Applied Optics*, 2007, **46**(34): 8328–8334
- Li X Q, Li J D, Wang D. Automatic tongue image segmentation based on histogram projection and matting. In: Proceedings of the 2014 IEEE International Conference on Bioinformatics and Biomedicine. Belfast, UK: IEEE, 2014. 76–81
- Xu W T, Kanawong R, Xu D, Li S, Ma T, Zhang G X, et al. An automatic tongue detection and segmentation framework for computer-aided tongue image analysis. In: Proceedings of the 13th IEEE International Conference on e-Health Networking, Applications and Services. Columbia, MO, USA: IEEE, 2011. 189–192
- Pang B, Zhang D, Wang K Q. The bi-elliptical deformable contour and its application to automated tongue segmentation in Chinese medicine. *IEEE Transactions on Medical Imaging*, 2005, **24**(8): 946–956

- 19 Shi M J, Li G Z, Li F F, Xu C. Computerized tongue image segmentation via the double geo-vector flow. *Chinese Medicine*, 2014, **9**: Article No. 7
- 20 Kass M, Witkin A, Terzopoulos D. Snakes: Active contour models. *International Journal of Computer Vision*, 1988, **1**(4): 321–331
- 21 Kang D J. A fast and stable snake algorithm for medical images. *Pattern Recognition Letters*, 1999, **20**(5): 507–512
- 22 Luo H, Gaborski R, Acharya R. Automatic segmentation of lung regions in chest radiographs: A model guided approach. In: Proceedings of the 2000 International Conference on Image Processing. Vancouver, BC, Canada: IEEE, 2000. 483–486
- 23 McInerney T, Terzopoulos D. Deformable models in medical image analysis: A survey. *Medical Image Analysis*, 1996, **1**(2): 91–108
- 24 Wang Ai-Min, Shen Lan-Sun. Study surveys on image segmentation. *Measurement and Control Technology*, 2000, **19**(5): 1–6, 16
(王爱民, 沈兰荪. 图像分割研究综述. 测控技术, 2000, **19**(5): 1–6, 16)
- 25 Shen Lan-Sun, Wang Ai-Min. Hierarchy of image segmentation. *Journal of Beijing Polytechnic University*, 2000, **26**(3): 33–38
(沈兰荪, 王爱民. 图像分割的分级性研究. 北京工业大学学报, 2000, **26**(3): 33–38)
- 26 McInerney T, Terzopoulos D. Topologically adaptable snakes. In: Proceedings of the 1995 IEEE International Conference on Computer Vision. Cambridge, MA, USA: IEEE, 1995. 840–845
- 27 Wang Yan-Qing, Wei Bao-Guo, Cai Yi-Heng, Shen Lan-Sun. A knowledge-based arithmetic for automatic tongue segmentation. *Acta Electronica Sinica*, 2004, **32**(3): 489–491
(王艳清, 卫保国, 蔡轶珩, 沈兰荪. 一种基于先验知识的自动舌体分割算法. 电子学报, 2004, **32**(3): 489–491)
- 28 Wang Yu-Zhong, Yang Jie, Zhou Yue, Zheng Yuan-Jie, Wang Yi-Qin. Image segmentation in tongue characterization. *Journal of Biomedical Engineering*, 2005, **22**(6): 1128–1133
(王郁中, 杨杰, 周越, 郑元杰, 王忆勤. 图像分割技术在中医舌诊客观化研究中的应用. 生物医学工程杂志, 2005, **22**(6): 1128–1133)
- 29 Wu Jia, Zhang Yong-Hong, Bai Jing, Weng Wei-Liang, Wu Yu, Han Yao, et al. Tongue contour image extraction using a watershed transform and an active contour model. *Journal of Tsinghua University (Science and Technology)*, 2008, **48**(6): 1040–1043
(吴佳, 张永红, 白净, 翁维良, 吴煜, 韩垚, 等. 基于分水岭变换和主动轮廓模型的舌体轮廓图像分割. 清华大学学报(自然科学版), 2008, **48**(6): 1040–1043)
- 30 Zhai X M, Lu H D, Zhang L Z. Application of image segmentation technique in tongue diagnosis. In: Proceedings of the 2009 International Forum on Information Technology and Applications. Chengdu, China: IEEE, 2009. 768–771
- 31 Qin Wu-Xing, Li Bin, Yue Xiao-Qiang. A hybrid tongue image segmentation algorithm based on initialization of snake contours. *Journal of University of Science and Technology of China*, 2010, **40**(8): 807–811
(覃武星, 李斌, 岳小强. 一种基于初始化 Snake 轮廓线的混合舌图像分割算法. 中国科学技术大学学报, 2010, **40**(8): 807–811)
- 32 Pang B, Wang K Q, Zhang S, Zhang F M. On automated tongue image segmentation in Chinese medicine. In: Proceedings of the 2002 Object Recognition Supported by User Interaction for Service Robots. Quebec City, Quebec, Canada: IEEE, 2002. 616–619
- 33 Guo J W, Yang Y K, Wu Q W, Su J L, Ma F. Adaptive active contour model based automatic tongue image segmentation. In: Proceedings of the 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics. Datong, China: IEEE, 2016. 1386–1390
- 34 Sun Xiao-Lin, Pang Chun-Ying. An improved snake model method on tongue segmentation. *Journal of Changchun University of Science and Technology (Natural Science Edition)*, 2013, **36**(5): 154–156
(孙晓琳, 庞春颖. 一种改进 snake 模型的中医舌体分割方法. 长春理工大学学报(自然科学版), 2013, **36**(5): 154–156)
- 35 Wang Ming-Ying, Zhang Xin-Feng, Zhuo Li. An improved snakes model for tongue image segmentation. *Measurement and Control Technology*, 2011, **30**(5): 32–35
(王明英, 张新峰, 卓力. 一种基于 Snakes 模型的中医舌象分割改进算法. 测控技术, 2011, **30**(5): 32–35)
- 36 Zhang H Z, Zuo W M, Wang K Q, Zhang D. A snake-based approach to automated segmentation of tongue image using polar edge detector. *International Journal of Imaging Systems and Technology*, 2006, **16**(4): 103–112
- 37 Salakhutdinov R, Mnih A, Hinton G. Restricted Boltzmann machines for collaborative filtering. In: Proceedings of the 24th International Conference on Machine Learning. Corvallis, Oregon, USA: ACM, 2007. 791–798
- 38 Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, **39**(4): 640–651
- 39 He K M, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, **42**(2): 386–397
- 40 Ren S Q, He K M, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, **39**(6): 1137–1149
- 41 Badrinarayanan V, Kendall A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**(12): 2481–2495
- 42 Qu P L, Zhang H, Zhuo L, Zhang J, Chen G Y. Automatic tongue image segmentation for traditional Chinese medicine using deep neural network. In: Proceedings of the 13th International Conference on Intelligent Computing. Liverpool, UK: Springer, 2017. 247–259
- 43 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer, 2015. 234–241
- 44 Chen L C, Zhu Y K, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer, 2018. 833–851

- 45 Chen L C, Papandreou G, Kokkinos I, Murphy K, Yuille A L. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, **40**(4): 834–848
- 46 Zhao H S, Shi J P, Qi X J, Wang X G, Jia J Y. Pyramid scene parsing network. In: *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA: IEEE, 2017. 6230–6239
- 47 Chen L C, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation. arXiv: 1706.05587, 2017.
- 48 He K M, Zhang X Y, Ren S Q, Sun J. Deep residual learning for image recognition. In: *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA: IEEE, 2016. 770–778
- 49 He K M, Zhang X Y, Ren S Q, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. In: *Proceedings of the 13th European Conference on Computer Vision*. Zurich, Switzerland: Springer, 2014. 346–361
- 50 Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: *Proceedings of the 32nd International Conference on Machine Learning*. Lille, France: JMLR.org, 2015. 448–456
- 51 Zhang D, Zhang H Z, Zhang B. *Tongue Image Analysis*. Singapore: Springer, 2017.
- 52 Huang Q, Dom B. Quantitative methods of evaluating image segmentation. In: *Proceedings of the 1995 International Conference on Image Processing*. Washington, DC, USA: IEEE, 1995. 53–56



卢运西 北京工业大学计算机技术专业硕士研究生。2015 年获得北京工商大学信息工程系学士学位。主要研究方向为图像处理和模式识别。

E-mail: yunxilu@emails.bjut.edu.cn

(**LU Yun-Xi** Master student in computer science and technology at Beijing University of Technology. He received

his bachelor degree from the Department of information engineering at Beijing Technology and Business University in 2015. His research interest covers image processing and pattern recognition.)



李晓光 北京工业大学副教授。2003 年于北京工业大学电子与信息工程专业获得学士学位，2008 年获得北京工业大学博士学位。主要研究方向为计算机视觉/图像增强，图像复原。本文通信作者。E-mail: lxxg@bjut.edu.cn

(**LI Xiao-Guang** Associate professor at Beijing University of Technology. He

received his bachelor degree in electronic and information

engineering from Beijing University of Technology in 2003. He received his Ph.D. degree from Beijing University of Technology in 2008. His research interest covers computer vision, image enhancement, and image restoration. Corresponding author of this paper.)

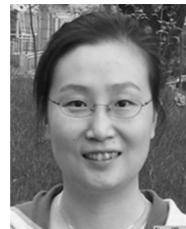


张辉 北京工业大学信息学部讲师。2010 年获得北京理工大学信号与信息处理专业博士学位。主要研究方向为计算机视觉，机器学习在多媒体内容分析、视觉追踪、目标检测中的应用。

E-mail: huizhang@bjut.edu.cn

(**ZHANG Hui** Lecturer at the Faculty of Information, Beijing University

of Technology. He received his Ph.D. degree in signal and information processing from Beijing Institute of Technology in 2010. His research interest covers computer vision, and machine learning techniques applied to multimedia content analysis, visual tracking and object detection.)



张菁 北京工业大学教授。2008 年获得北京工业大学博士学位。美国德州大学圣安东尼奥分校计算机科学系访问学者。主要研究方向为图像处理，图像识别，图像检索。E-mail: zhj@bjut.edu.cn

(**ZHANG Jing** Professor at Beijing University of Technology, visiting scholar in the Department of Computer

Science, University of Texas at San Antonio, USA. She received her Ph.D. degree from Beijing University of Technology in 2008. Her research interest covers image processing, image recognition, and image retrieval.)



卓力 北京工业大学教授。1992 年获得电子科技大学无线电技术系工学学士学位，1998 年和 2004 年分别获得东南大学信号与信息处理专业硕士学位和北京工业大学模式识别与智能系统专业博士学位。主要研究方向为图像/视频编码和传输，多媒体内容分析，多媒体信息安全。E-mail: zhuoli@bjut.edu.cn

(**ZHUO Li** Professor at Beijing University of Technology. She received her bachelor degree in radio technology from University of Electronic Science and Technology in 1992, master degree in signal and information processing from Southeast University in 1998, and Ph.D. degree in pattern recognition and intellectual system from Beijing University of Technology in 2004. Her research interest covers image/video coding and transmission, multimedia content analysis, and multimedia information security.)