

DOI 10.19656/j.cnki.1002-2406.20221106

一种基于深度学习的二阶段舌象分割网络模型

刘东华, 张伟, 顾旋, 梁富娥, 吕珊珊

(甘肃中医药大学信息工程学院, 甘肃 兰州 730000)

【摘要】目的:为了改善传统的图像分割算法需要人为干预且分割精度低,探讨基于深度学习的二阶段舌象分割网络模型提高分割精度。方法:将基于深度学习的二阶段舌象分割网络模型分为粗分割和精分割两个部分,粗分割定位舌体去除无关干扰信息,精分割实现舌象的像素级精细分类分割。对分割后的舌象图片进行形态学优化,进一步优化分割结果。结果:本分割网络模型平均交并比为95.25%,比主流卷积神经网络模型高出3.10%。结论:基于深度学习的二阶段舌象分割网络模型优于传统图像分割算法和主流卷积神经网络模型,在不同类型的数据集上能准确分割舌体,有较高的精度和鲁棒性。

【关键词】深度学习;卷积神经网络;舌象分割;形态学优化

【引用格式】

刘东华, 张伟, 顾旋, 等. 一种基于深度学习的二阶段舌象分割网络模型[J]. 中医药信息, 2022, 39(11): 35-39, 46.

LIU D H, ZHANG W, GU X, et al. A two-stage tongue demonstration segmentation network model based on in-depth learning[J]. Information on TCM, 2022, 39(11): 35-39, 46.

舌诊是指医生观察病人的舌质和舌苔来进行诊断的一种方法,属于中医“望、闻、问、切”四诊中望诊的重要方法,也是中医特色诊法^[1]。应用现代分析手段分析舌象,首先要从图像中将舌体分割出来才能继续对舌象进行分类识别和诊断,所以舌象分割是后续研究的前提和关键。由于舌象分割的复杂性和特殊性,人们提出了各种各样的分割方法。阈值法^[2]分割舌体,利用舌体颜色像素值的差别来进行划分,分割速度快但精度不高;边缘检测算法^[3],根据图像边缘像素值变化剧烈这一特点来进行定位,但舌体背景环境复杂,出现错误分割率高;以主动轮廓模型^[4]为主的算法,模型算法过度依赖轮廓线的好坏,很容易进行错误的分割。因为这些算法都是利用的舌体的颜色,纹理等浅层的特征,以上传统图像处理方法在精度上很难有较大的提升,想要提高舌体的分割精度就需要利用舌象更深层次的特征信息。

近几年深度学习在计算机视觉领域得到了广泛的

应用,这是因为卷积神经网络可以提取图像深层次抽象的语义特征。卷积神经网络从R-CNN^[5]发展到Faster R-CNN^[6],从CNN到FCN^[7],在图像目标检测、分类、分割领域都得到了广泛的应用。卷积神经网络在医学图像领域已经取得了不错的成果,但应用在舌象领域还不是很成熟。本文在卷积神经网络的基础上改进了一个二阶段舌象分割网络模型,网络分为粗分割(Rsnet)和精分割(Psnet)两部分级联组成,两部分网络可以单独训练,减轻整体网络的复杂度,降低对计算机硬件的要求。

1 舌象分割网络及方法

本算法的整体网络结构见图1,网络结构主要可以分为粗分割和精分割。粗分割阶段,将图像进行卷积后提取不同尺度的特征图融合,传入全连接层从脸部图像中提取出感兴趣区域,定位舌体,去除无关干扰信息。精分割阶段,利用深度卷积神经网络处理上一阶段提取出的感兴趣区域,传入特征增强网络强化特

基金项目:甘肃中医药大学科技创新基金资助项目(KCYB2018-6)

第一作者简介:刘东华(1998-),男,硕士研究生,主要研究方向:深度学习,计算机视觉。

征信息,然后分类器对图像每一个像素点进行分割,实现舌体精分割。形态学优化,利用腐蚀膨胀操作对精

分割后的图像进行优化处理,进一步优化分割的结果。

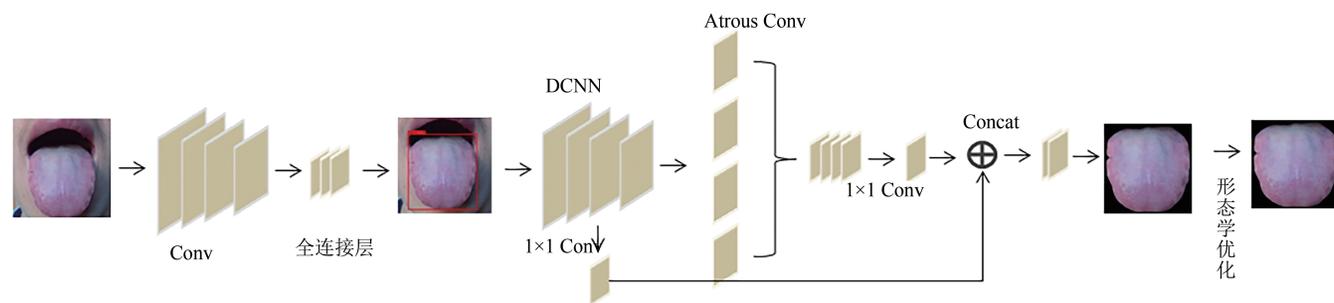


图1 舌象分割算法网络结构

1.1 粗分割阶段

舌体分割的本质就是对舌象图像中的每一像素点进行分割,舌象图像中无关因素越多就会影响分类器的效率,影响分割的精度。所以减少舌象中无关因素的影响,准确地提取出感兴趣区域,定位舌体区域是关键。

粗分割阶段网络的主干部分由带深度可分离卷积^[8]的 Mobilenet_v1^[9]组成,对舌象图片进行卷积和池化操作,提取图像深层次的抽象特征。舌象经过卷积池化后,笔者取出网络的最后3个有效特征层来构建特征金字塔,不同尺度的特征图包含了不同层次的特征信息,将这些不同尺度的特征图进行有效融合,融合舌象不同层次的特征信息,进行特征增强处理。特征增强后的图像传入全连接层,全连接层利用这3个增强后的特征进一步分类回归预测结果。

深度可分离卷积可以实现普通卷积一样的结果,但其所需要的参数和运算量大大减少。深度可分离卷积模块可以分为逐深度卷积和逐点卷积,逐深度卷积对输入特征图的每个通道分别使用一个卷积核,然后将所有卷积核的输出再进行拼接得到它的最终输出。逐点卷积实际为 1×1 卷积,在深度可分离卷积中主要起两个方面的作用,一个作用是能够自由改变输出通道的数量,另一个作用是对逐深度卷积输出的特征图进行通道融合。

假设输入特征图的尺寸为 $D_k \times D_k \times M$,卷积核的尺寸为 $D_f \times D_f \times M$,数量为 N 。假设对应特征图空间位置中的每一个点都会进行一次卷积操作,那么可知单个卷积共需要进行 $D_k \times D_k \times D_f \times D_f \times M$ 次计算。这是因为特征图空间维度共包含 $D_k \times D_k$ 个点,而对每个点进行卷积操作的计算量与卷积核的尺寸一致,即 $D_f \times D_f \times M$,所以对单个卷积而言,其总计算量为 $D_k \times D_k \times D_f \times D_f \times M$,对 N 个卷积的总计算量为 $D_k \times$

$D_k \times D_f \times D_f \times M \times N$ 。对逐深度卷积计算量为 $D_k \times D_k \times D_f \times D_f \times M$,逐点卷积计算量为 $D_k \times D_k \times M \times N$,所以深度可分离总计算量为 $D_k \times D_k \times D_f \times D_f \times M + D_k \times D_k \times M \times N$ 。对比可知深度可分离卷积的计算量与普通卷积的比为 $\frac{1}{N} + \frac{1}{D_f^2}$,可知深度可分离卷积计算效率远优于普通卷积。

1.2 精分割阶段

本文的精分割网络模型是编码-解码结构,编码部分网络由带空洞卷积的深度卷积神经网络(DCNN)模块^[10]和带空洞卷积的空间金字塔模块^[11]组成。深度卷积神经网络模块对粗分割网络传入的图像进行下采样特征提取,然后特征金字塔模块对特征层利用不同rate的空洞卷积进行特征提取,特征金字塔融合后再进行 1×1 卷积压缩特征。解码部分主要作用就是将编码模块中的低维特征与空间特征金字塔融合的高维特征进行有效合并,然后进行上采样恢复图像大小并分割预测。

在图像分割领域,DCNN通常会多次使用卷积和池化操作,这样可以提取图像深层次的高维特征,但是经过卷积池化后特征图会不断变小,低层的边界信息和语义信息会丢失,这样就会降低分割的精度。为了解决类似问题,人们提出了空洞卷积模块,它能代替下采样和上采样操作,且不会造成精度损失。

空洞卷积可以不丢弃任何信息并增大特征图的感受野,让每个输出特征图都包含大范围信息。空洞卷积与正常的卷积相比,空洞卷积加入了一个叫作扩张率(dilation rate)的参数,该参数定义了卷积核处理数据时各值的间距。

图2中三幅 3×3 卷积图像中(a)为普通卷积,(b)(c)为不同扩张率的空洞卷积,图像最外层的大框表示输入图像,黑色的圆点表示 3×3 的卷积核,灰色框格表示的是卷积后的感受野。图中(a)是普通的卷积过

程(dilation rate = 1),卷积后的感受野为3,(b)是扩张率(dilation rate = 2)的空洞卷积,卷积后的感受野为5,(c)是扩张率(dilation rate = 3)的空洞卷积,卷积后的感受野为8。从图中可以看出,普通卷积可以看作空

洞卷积的一种特殊情况,同样一个 3×3 的卷积,空洞卷积却可以起到 5×5 、 7×7 等卷积的效果,以上可以看出空洞卷积能够不增加网络的参数量,却可以增大特征图的感受野。

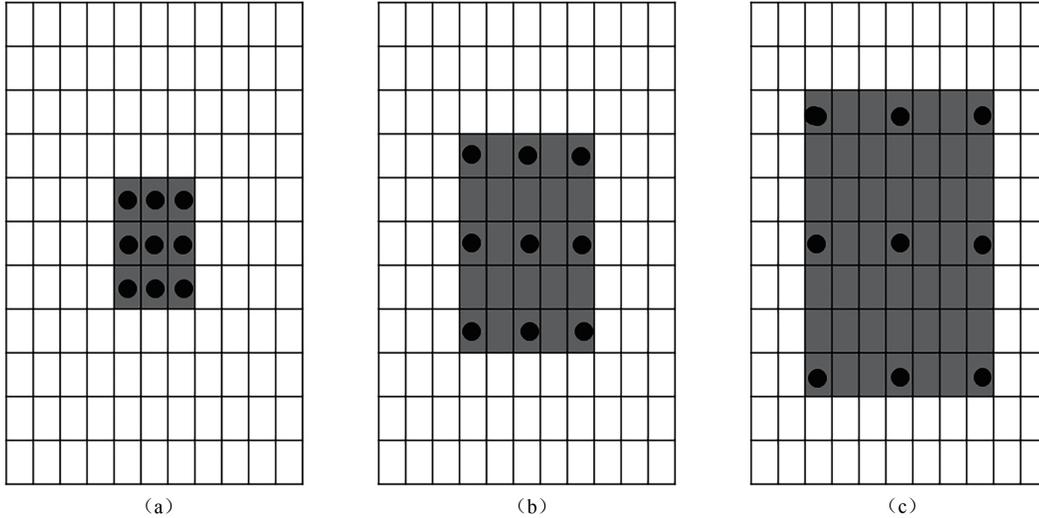


图2 普通卷积和空洞卷积对比图

空洞卷积感受野的计算公式如下,假设空洞卷积的卷积核大小为 k ,空洞数为 d ,则其等效卷积核大小 k' :

$$k' = k + (k - 1) \times (d - 1)$$

当前层网络感受野计算,其中, RF_{i+1} 表示当前层的感受野, RF_i 表示上一层的感受野, k' 表示卷积核的大小:

$$RF_{i+1} = RF_i + (k' - 1) \times S_i$$

S_i 表示之前所有层的步长的乘积(不包含本层):

$$S_i = \prod_{i=1}^i stride_i$$

1.3 形态学优化

形态学优化的主要作用是利用形态学图像处理算法来进行图像优化处理,分割后的图像背景中会出现一些细小噪声和小的连通域,边缘会产生毛刺和不平整,经过形态学算法优化处理后会消除噪声并平滑边缘。形态学图像优化算法常被用于图像的预处理和后处理中,是一种不错的图像增强技术。

形态学优化算法过程:

(1)首先将分割后的图像做二值化处理,然后选取适合舌体形状大小的结构元素对二值化处理后的图像进行多次腐蚀膨胀操作。

(2)将上一步腐蚀膨胀处理后的图像转换为布尔类型数组,以舌体为主要部分,消除舌体周围小的连通区域即可以消除被误分割的背景噪声并平滑边缘。

(3)将舌体分割输入的原图与布尔类型数组相乘

就可以得到形态学优化后的分割图像。

2 实验研究与分析

2.1 数据集

为了提高算法模型的泛化能力,本文的数据主要来源于两种方式:(1)专业舌象仪在标准光源环境下对受试者进行舌象数据采集;(2)智能手机在非标准环境下对受试者进行舌象数据采集。最终通过筛选从数据库中选取了1025张舌象图片,然后随机将这1025张图片分成5等份,205张图片作为测试集,820张图片作为训练集。

2.2 实验环境

本文为了训练分割网络模型,应用tensorflow框架搭建分割网络,代码实验全在如下计算机上完成:Windows 10系统,AMD R7 5800H Cpu,16 G Ram,NVIDIA GeForce GTX1660 4G Gpu,使用的编程语言为python。

2.3 算法评判标准

图像分割算法效果评价,一般是把算法网络自动分割的图像与实际分割图像对比分析,计算出分割算法的好坏。本文采用像素精度(PA)、平均像素精度(MPA)和平均交并比(MIOU)3个标准来对算法进行评判。计算公式如下:

$$PA = \frac{\sum_{i=0}^m P_{ii}}{\sum_{i=0}^m \sum_{j=0}^m P_{ij}}$$

$$MPA = \frac{1}{m+1} \sum_{i=0}^m \frac{P_{ii}}{\sum_{j=0}^m P_{ij}}$$

$$MIOU = \frac{1}{m+1} \frac{\sum_{i=0}^m P_{ij}}{\sum_{j=0}^m P_{ij} + \sum_{j=0}^m P_{ji} - P_{ii}}$$

式中： P_{ii} 为将第*i*类正确地分割为第*i*类的像素点数量， P_{ij} 为将第*i*类错误地分割为第*j*类的像素点数量， P_{ji} 为将第*j*类错误地分割为第*i*类的像素点数量， $m+1$ 是类别个数。这三个评判指标取值都在0~1之间，值越大说明分割精度越好。

2.4 结果分析

将本文方法与Otsu算法、Grabcut算法^[12]等传统图像分割方法以及基于深度学习思想的Segnet算法^[13]进行对比分析，见图3。第一列5张舌象图片中的前3张为标准环境下舌象仪采集的数据，后2张为智能手机在非标准环境下采集的数据，通过不同类型的数据集来验证算法的鲁棒性。

由图3可以看出，Otsu算法根据自适应阈值来分割图像，选取一个适当的阈值来对图像进行像素级分类，

此方法在标准环境下采集的图像中分割效果不错，但在非标准环境下采集的图像中效果较差，由此可知Otsu算法在不同类型的数据集上鲁棒性较差。Grabcut算法需要手动标记出分割目标区域，然后进行多次算法迭代运算分割出结果，算法不能自动进行，需要人为干预，分割效率也较低。Otsu算法和Grabcut算法都是传统的图像分割算法，利用的是舌象最底层的像素特征，颜色、形状、纹理等。根据底层特征进行分割速度快，但精度不高，像舌体的颜色有时会和嘴唇的颜色相近，所以分割时会把嘴唇误分割为舌体，并且算法在不同类型的舌象数据集上鲁棒性较差。Segnet算法是基于深度学习的网络模型，由图3和表1数据可以看出其分割效率比传统图像分割算法要高出许多，而且可以在不同类型的舌象数据集上运行，并保持分割精度基本不变，所以其分割精度和鲁棒性都较传统的图像分割算法精度更优。本文分割算法平均交并比达到了95.25%，比Segnet算法高3.10%，在不同类型的数据集上可以有效地避免多数像素的错误分类和漏分类，在大部分图像上可以实现理想的分割效果，有较好的分割精度和鲁棒性，可以满足后续分类识别的精度要求。

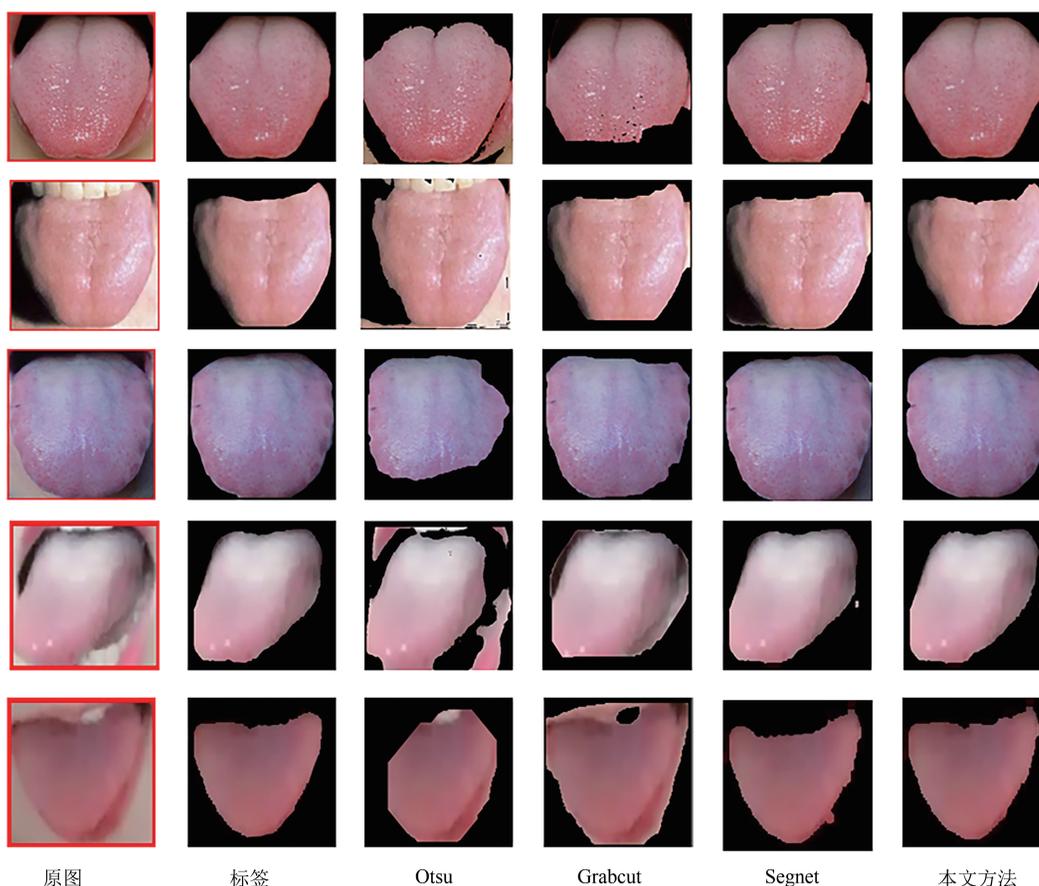


图3 不同分割算法结果对比

表1 不同分割算法的结果(%)

算法	PA	MPA	MIOU
Otsu	68.56	68.12	65.62
Grabcut	86.66	85.46	85.19
Segnet	94.78	93.56	92.15
本文算法	97.86	96.42	95.25

本文提出的分割算法网络是二阶段的,其中粗分割(Rsnet)提取感兴趣区域减少背景因素干扰,精分割(Psnet)通过对舌体像素级的分类来实现精准分割,后接形态学优化算法来消除细小噪声和平滑舌体边缘信息,并通过实验来验证每部分在整体网络中的作用大小,将各部分拆装后再进行有效拼接,进行实验对比分析。MIOU越高表明像素分类正确率高,舌体分割精度高。从表2可以看出直接使用精分割网络,不经过粗分割网络提取感兴趣区域的分割精度有所下降,直接进入精分割网络分割舌体,图像中的背景因素会产生干扰,增大像素错误分类的概率。经过形态学优化算法处理后,舌体边缘因背景因素引起的错误分割像素点被消除,减少了被错误分类的像素点数量,整体分割精度又得到了提升。从表2中也可以看出经过粗分割后再精分割的算法精度比直接进行精分割的平均交并比高0.42%,说明经过粗分割提取目标区域,去除无关因素干扰对网络整体精度的提升有很大帮助。形态学优化二阶段分割出来的舌象平均交并比提升了0.05%,可以看出二阶段分割网络的整体效率高,基本不会出现较大区域的错分像素点。

表2 不同拼接算法的结果(%)

拼接算法	PA	MPA	MIOU
Psnet	96.68	95.82	94.78
Psnet + 优化	97.36	96.55	95.07
Rsnet + Psnet	97.58	96.36	95.20
本文算法	97.86	96.42	95.25

3 结论

本文针对传统图像分割算法需要预处理,人为干预、分割精度低等问题,提出了基于深度学习的二阶段舌象分割网络模型。首先对图像目标进行感兴趣区域提取,减少背景区域的干扰,然后利用空洞卷积在不损失信息的情况下增大特征图感受野的方法来对图像进行精准分割,提高分割精度,最后采用形态学优化算法对分割后的舌体图像进行优化处理,消除细小噪声和平滑边缘。本文网络采用级联的方法来对舌象进行自

动分割,分割精度高而且模型灵活并对计算机硬件要求低。与传统图像分割算法相比,本文网络分割精度高、鲁棒性好且自动化程度高。与其他主流分割算法相比,本文网络在分割精度上比其有所提升,能够满足后续分类识别的条件要求。在接下来的研究工作中,我们可以在高精度分割网络的基础上,进一步研究舌象的分类识别,根据舌象识别相应病症。

【参考文献】

- [1] 陆冠龙,黄益栓,张琦,等.融合舌象和形体特征的辅助中医体质辨识模型研究[J].时珍国医国药,2019,30(1):244-246.
- [2] 王旭阳,刘世健.基于多方向阈值的超分辨率图像噪声识别仿真[J].计算机仿真,2021,38(12):132-135,81.
- [3] 刘建思,尹丽菊,潘金凤,等.基于参数化对数图像处理模型的光照不均匀图像的边缘检测算法[J].激光与光电子学进展,2021,58(22):140-149.
- [4] 刘冬梅,常发亮.结合Retinex校正和显著性的主动轮廓图像分割[J].光学精密工程,2019,27(7):1593-1600.
- [5] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]// Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA:IEEE,2014:580-587.
- [6] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real time object detection with region proposal networks [C]// Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Canada:MIT Press,2015:91-99.
- [7] SHELHAMER E, LONG J, DARRELL T. Fully Convolutional networks for semantic segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2017,39(4):640-651.
- [8] 孔令军,刘伟光,周耀威,等.一种基于深度可分离卷积的轻量级人体关键点检测算法[J].无线电工程,2022,52(1):76-82.
- [9] 张骏,朱标,吉涛.基于MobileNet-SSD的红外人脸检测算法[J].激光与红外,2021,51(1):107-113.
- [10] 张新峰,郭宇桐,蔡轶珩,等.基于DCNN和全连接CRF的舌图像分割算法[J].北京航空航天大学学报,2019,45(12):2364-2374.
- [11] 金一凡,余雷,费树岷.基于空洞卷积神经网络的噪声水平可调的高斯去噪方法[J].模式识别与人工智能,2021,34(11):979-989.
- [12] ROTHER C, KOLMOGOROV V, BLAKE A. GrabCut: interactive foreground extraction using iterated graph cuts [J]. ACM Transactions on Graphics,2004,23(3):309-314.
- [13] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2017,39(12):2481-2495.

(收稿日期:2022-03-11)

(下转第46页)