

基于 Mask R-CNN 的舌图像分割研究^{*}

颜建军¹, 徐姿¹, 郭睿^{2**}, 燕海霞³, 王忆勤^{3**}

(1. 华东理工大学机械与动力工程学院 上海 200237; 2. 上海中医药大学交叉科学研究院 上海 201203;
3. 上海中医药大学四诊信息综合实验室 上海 201203)

摘要:目的 舌图像分割是舌诊客观化的关键之一,易受舌体附近嘴唇和皮肤等带来的影响,而增加分割的难度。针对该问题,为确保舌图像分割的准确性,本研究提出一种基于卷积神经网络 Mask R-CNN 的舌图像分割方法。方法 首先用标注工具 labelme 对舌图像进行标注,然后进行 Mask R-CNN 舌图像分割模型的训练和舌图像分割测试。结果 采用该方法进行舌图像分割,获得的舌体边缘比较准确,并且四个定量评价指标均像素准确度、平均准确度、均交并比、频权交并比均高于 84.6%。结论 本研究取得了较好的舌体分割效果,能够改善舌体周围的嘴唇和皮肤颜色与舌体颜色接近导致舌体分割轮廓不准确的问题,为舌图像分割提供了一种新的思路与方法,对舌诊客观化具有一定实用价值和借鉴意义。

关键词:舌图像分割 舌诊客观化 卷积神经网络 实例分割 深度学习

doi: 10.11842/wst.20190611006 中图分类号: R2-0 文献标识码: A

中医通过“望、闻、问、切”以获得体征和症状等相关信息,经过综合分析归纳后,提取病机,从而进行相应的论治。舌诊作为一种中医诊断方法,主要通过观察舌苔、舌质等的舌像特征来了解人体生理病理的变化,在中医诊断中具有重要的应用价值。舌诊存在的问题是主观性较强,无法精确量化,结果难以记录,使其发展受到了限制。因此,运用机器视觉技术进行舌图像的分析处理,实现中医舌诊客观化和定量化,是现代中医舌诊发展的重要工作。其中舌图像分割是中医舌诊客观化的基础,舌体分割准确程度会直接影响舌诊系统中舌色苔色识别以及形态分析诊断^{[1][2]}。

当前舌图像分割方法主要包括阈值分割算法、主动轮廓模型及图分割算法等。刘关松等^[3]提出基于阈值分割的舌体分割方法;Zhai X 等^[4]提出基于双 Snake 的舌像分割方法;Pang B、Zhang H 和 Fu Z 等提出主动轮廓模型与双椭圆形可变模板等检测器结合法^{[5][6][7]};吴佳等^[8]提出分水岭变换与主动轮廓模型组合法;Yan Z 等^[9]采用基于图分割的 GrabCut 算法进行舌图像分

割。这些方法在一定程度上改进了舌图像分割结果,但都存在不足:基于阈值分割方法对采集舌像的光照要求较高且分割轮廓不太精确;利用主动轮廓模型的分割方法容易受到嘴唇干扰而存在分割误差;采用 GrabCut 算法进行舌图像分割,计算量大且速度较慢。

在舌图像分割过程中,当舌体周围的嘴唇或皮肤颜色较接近舌体时,会导致分割的难度增大,所以容易造成分割效率低、准确度不高的结果,对后续的舌图像分析产生较大影响。但是现有的舌图像分割方法均不能很好地解决这一问题,因此,舌图像的准确分割仍然是舌诊客观化中的一个难点,这在一定程度上影响了舌诊客观化的进程。

近年来,越来越多的学者着手进行基于像素的图像分割研究。对舌图像分割而言,像素特征的提取虽然容易实现,但很难表达舌体的位置信息,且当舌体与背景之间颜色相近、边界模糊时,两者的像素特征可能会重合。故仅提取舌像的像素特征较难全面地表达舌体信息,还需要进一步提取图像的深层语义信

收稿日期:2019-06-11

修回日期:2019-08-06

* 国家自然科学基金委员会面上项目(81673880):基于中医四诊大数据的冠心病风险评估与预测模型研究,负责人:王忆勤。

** 通讯作者:王忆勤,教授,本刊编委,主要研究方向:中医诊断规范化、标准化研究;郭睿,副研究员,主要研究方向:中医四诊客观化。

息,得到更完整的特征以实现舌图像的精确分割。而基于深度学习的图像分割方法能够通过模拟人脑的阶层式结构进行信息传递,实现从像素级原始数据到抽象语义特征的转换,在医学图像处理等领域取得了较好效果^{[10][11]}。特别是在Faster R-CNN的基础上提出的Mask R-CNN目标实例分割框架^[12],能够解决以往网络不能被很好优化的退化问题,可以进行有效目标探测、准确分类和高质量分割,得到的准确率远高于HyperColumns、FCIS和SIS等经典实例分割方法^[13-15]。

本研究提出基于Mask R-CNN的舌图像分割方法,在较大样本舌图像标注的基础上,建立舌图像分割的深度学习模型,并对分割效果进行评估和分析。

1 基于 Mask R-CNN 的舌图像分割

近年来,实例分割已经逐渐发展成为一项支持各种基于内容的多媒体应用不可或缺的关键技术^[16]。目前,深度学习的发展尤为突出,已经在图像识别领域的应用取得了较好的成果。和传统的方法相比,深度学习有以下三大优点:①在计算机视觉领域识别率最高;②充分利用大数据的优势进行特征的自动学习;③可最大程度发挥特征表示和分类器联合优化的性能。深度学习在实例分割上的应用逐渐成为一个研究热点,特别是基于Faster R-CNN模型改进的Mask R-CNN实例分割网络模型在图像分割领域取得了不错的效果。因此,本研究采用Mask R-CNN实例分割网络模型进行舌图像的分割。

Mask R-CNN在Faster R-CNN的基础上主要进行三点改进:①在边界框识别分支上添加一个用于预测目标掩码的并行分支;②为每个类别分别进行二元掩码的预测;③提出简洁非量化层RoIAlign,该层不但可以保留大致的空间位置,还能将掩码精确度相对提高10%~50%。因此,Mask R-CNN可以在更严格的评价指标下得到更好的结果,具有框架简洁灵活、训练和分割速度较快以及分割精确度较高的优点。

Mask R-CNN采用RoIAlign层替换了Faster R-CNN中原有的RoIPool层这一策略是其一大亮点。RoIPool层基于的原理是从每个感兴趣区域(Region of Interest,RoI)中提取小的特征图(如 $7 \times 77 \times 7$),基于其选择的特征图区域,会与原图中的区域有轻微出入。相对应具体的做法是先对浮点数RoI进行量化,再提取分块的直方图,最后进行最大池化(Max pooling)

操作。Max pooling会先将整个图像不重叠地分割成若干个同样大小的小块(pooling size),然后每个小块内只取最大的数字,再舍弃其他节点后,最终保持原有的平面结构得出输出结果。由此可见,RoIPool层是基于stride间隔来取整从原图RoI得到特征图RoI,这会使得在把特征图RoI映射回原图RoI时,出现stride间隔造成的误差,导致原图RoI和提取特征RoI间的不重合,该不重合性在max pooling后特征图的RoI与原RoI间的空间不对齐时更加明显。这一点对于精确的像素级masks预测和分割具有较大的负面影响,因此需要像素级的空间对齐。

RoIAlign能够去除RoIPool引入的不重合性,准确地对齐输入的提取特征,以此来避免RoI边界或bins进行量化的发生,如采用 $x/16x/16$ 来替代rounding($x/16$)rounding($x/16$)[四舍五入处理];根据每个RoI bin的四个采样点,采用双线性插入来计算输入特征的精确值,并采用取最大值或平均值来组合结果。假设点(x,y)(x,y),取其周围最近的四个采样点,在Y方向进行两次插值,再在X方向进行两次插值,以得到新的插值。这种处理方式完全不会影响RoI的空间布局。假设一个 128×128 的图像, 25×25 的特征图,其原始图像的每一个像素与特征图上的 $25/128$ 个像素对应。想要得到与原始图像左上角 15×15 位置对应的特征区域,就要在特征图上需要选择 $15 \times 25/128 \approx 2.93$ 个像素。对于这种情形,RoIPool会舍去零头选择两个像素,从而导致排列问题。但是RoIAlign避免使用去掉小数点后取整的方式,而是使用双线性插值直接准确地获得2.93像素位置的信息,避免了排列错误。

因此,RoIAlign既确保mask分支的每一网络层均可保持 $m \times mm \times m$ 的object空间布局,而不采用会导致空间信息损失的压扁拉伸成向量表示形式,又保证RoI特征图的对齐性,从而不影响pixel-to-pixel操作,以保留per-pixel空间对应关系。

Mask R-CNN是基于Python 3,Keras以及TensorFlow平台实现的,是以ResNeXt101神经网络作主结构、特征金字塔网络(FPN)作辅结构的ResNeXt-101-FPN特征提取网络。用于实例分割的Mask R-CNN框架,如图1所示。

1.1 舌图像分割的过程

基于Mask R-CNN的舌图像分割包括舌体轮廓标注、模型训练、舌图像分割及分割结果评估这四个主

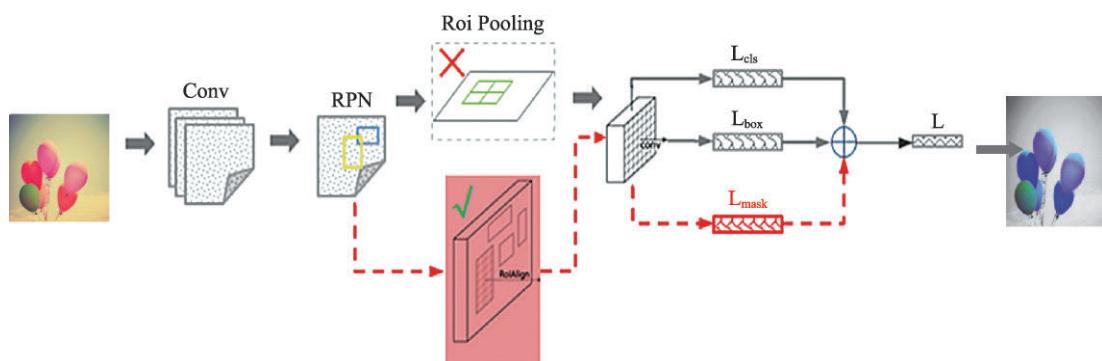


图1 Mask R-CNN 目标实例分割框架

要步骤,具体如图2所示。

1.1.1 舌体轮廓标注

由于深度学习需要用到大量标注样本进行训练,因此,需要先利用标注工具对样本的舌体轮廓标注。

本研究使用标注工具labelme进行舌图像样本标注,该工具基于多边形框对目标物体进行标注。标注时利用标注点和标注点之间的最短路径,得到能够较好贴合舌体轮廓的多边形,获得标注结果如图3所示。

1.1.2 舌图像分割模型的训练

舌图像分割的网络模型训练在tensorflow环境下使用Mask R-CNN目标实例分割框架进行的。

Mask R-CNN整个网络结构为Faster R-CNN + FCN+ROIAlign,不仅在Faster R-CNN的基础上增加FCN产生对应的Mask分支,而且提出采用ROIAlign策略替换ROI Pooling。Faster R-CNN可以既快又准地完成目标检测;FCN能够精准地进行语义分割;ROIAlign则解决了Faster R-CNN中存在的ROI Pooling像素偏差问题。除此之外,主网络ResNeXt-101网络很好地解决了传统网络构架中随着网络层数增加准确率先上升后饱和最后下降的问题。这些都有益于提高舌图像分割的分割速度和准确率。

训练的主要参数有TRAIN_ROIS_PER_IMAGE、STEPS_PER_EPOCH 和 VALIDATION_STEPS。其中,参数TRAIN_ROIS_PER_IMAGE代表每张样本建议的ROI数量,并将其作为后续分类和掩膜网络的输入,一般目标占整体比例越大,该值的取值就越大;STEPS_PER_EPOCH为keras中的fit_generator函数中的参数,fit_generator函数的作用是在训练集太大时可自动生成batch,该参数表示每轮迭代次数,因为tensorboard会自动更新并在每个epoch结束时保存更新,所以进行该参数设置时不需要考虑训练集大小。

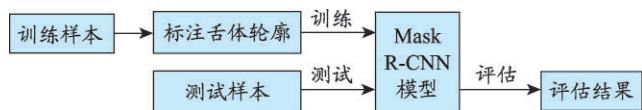


图2 基于 Mask R-CNN 模型的舌图像分割方法步骤

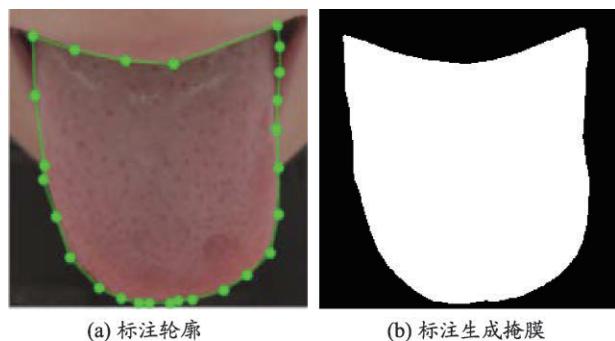


图3 labelme标注工具标注结果

同时又由于验证集也在epoch最后记录,比较花时间,因此该参数不能设置太小,以免花费大量时间进行验证;VALIDATION_STEPS也是fit_generator中的参数,表示验证的步数,但此参数越大验证就越精确,而训练速度越慢。因此,本研究将TRAIN_ROIS_PER_IMAGE、STEPS_PER_EPOCH 和 VALIDATION_STEPS分别设置为300、5000和5。

此外,实验采用的GPU显卡为NVIDIA Tesla V100,CPU为Intel Core i7-8700K @3.70GHz六核,内存大小32GB。

1.1.3 舌图像分割的评估指标

为了检验基于Mask R-CNN的方法进行舌图像分割的效果,本研究将中医专家评估和定量指标评估结合起来,进行图像分割结果评估,这样能够更加全面地反映舌图像分割的结果。用于定量评估的四个指标分别是均像素准确度(Mean Pixel Accuracy, MPA)、平均准确度(Mean Accuracy, MA)、均交并比(Mean

Intersection over Union, MIoU)、频权交并比 (Mean Frequency Weighted Intersection over Union, MFWIoU), 其定义如下:

$$\text{MPA} = \frac{\sum_{i=1}^n A_i}{P_n} \quad (1)$$

$$\text{MA} = \frac{\sum_{i=1}^n B_i}{n} \quad (2)$$

$$\text{MIoU} = \frac{\sum_{i=1}^n \text{IoU}_i}{n} \quad (3)$$

$$\text{MFWIoU} = \frac{\sum_{i=1}^n \text{IU}_i \times C_{pi}}{\sum_{i=1}^n E_i} \quad (4)$$

其中, A_i 为像素准确预测的像素点; P_n 为图像像素总数得到的准确率; i 为像素点的个数; B_i 为每个类别被正确分类像素数目的比例; i 为所有类别; $\text{IoU} = \frac{A_p}{C_p} + D_p - A_p$, A_p 为某类别被正确分类像素数目; C_p 为某类别真实的像素总数; D_p 为某类别预测像素总数; $\text{IU} = C_p + D_p - A_p$; C_p 为某类别真实的像素总数; D_p 为某类别预测像素总数; A_p 为某类别被正确分类像素数目; E_i 为图像像素总数。

2 结果与讨论

本研究舌图像样本由上海中医药大学四诊信息综合实验室提供,共计 2870 张,其中 1870 张作为训练样本,1000 张作为测试样本。舌图像中大多数区域为舌体,还有嘴唇、牙齿、部分面颊以及少量背景。

将标注的舌图像样本输入到 Mask R-CNN 中进行训练,建立舌图像分割模型,并利用测试样本进行舌体分割及结果评估。舌图像分割测试结果,如图 4 所示。

由图 4 可以看出,Mask R-CNN 舌图像分割模型得到的舌体轮廓较为清晰准确,能较好地处理舌体周围嘴唇以及皮肤的干扰,分割结果得到了中医认可。

为了进一步验证该模型能够较好地解决舌体周围嘴唇以及皮肤对舌图像分割的干扰问题,本研究针对整个舌图像样本,分别进行了靠近嘴唇的舌体部分同舌体附近的嘴唇部分及同舌体周边的皮肤部分的色差计算统计,并对舌图像进行了边缘检测,分析验证模型的鲁棒性。



(a) 原始舌图像 (b) 舌图像分割结果

图 4 Mask R-CNN 舌图像分割模型结果

LAB 空间是基于人的视觉系统对颜色的感觉来设计的,能较为贴切地反映人眼视觉系统对颜色的感知,所以此处采用 LAB 颜色空间来计算舌体同周围嘴唇及皮肤的色差。在该空间中存在 L、A、B 三个通道,其中 L 表示明度, A 表示红绿色差, B 表示蓝黄色差。两个颜色之间的相似程度,即色差,其计算公式如下:

$$\Delta E = [(\Delta L)^2 + (\Delta A)^2 + (\Delta B)^2]^{1/2} \quad (5)$$

其中, ΔE 表示色差, $\Delta L/\Delta a/\Delta b$ 分别表示两个颜色之间在不同分量的差值。

使用上述 LAB 颜色空间色差计算方法,针对有关靠近嘴唇的舌体部分同舌体附近的嘴唇部分,本研究的舌图像样本得到的色差范围是 0–337.29。其中,图 4(a) 的两张舌图像样本的两部分色差值分别为 0 和 1.41,由图 4(b) 的舌体分割结果图可以直观地看出:当舌体部分的颜色同周围嘴唇颜色极为接近甚至相同时,该舌体分割模型还是能够较好地将目标舌体分割出来,不受周围相似颜色嘴唇的干扰。同样地,针对有关靠近皮肤的舌体部分同靠近舌体的皮肤部分,本研究的舌图像样本得到的色差范围是 0–306.88。其中,图 4(a) 的两张舌图像样本的两部分色差值分别为 3.32 和 1.41,由图 4(b) 的舌体分割结果图可以直观地看出:当舌体区域的颜色同周围皮肤颜色比较接近时,该舌体分割模型依旧能够较好地将目标舌体分割出来,不受周围相似颜色皮肤的影响。

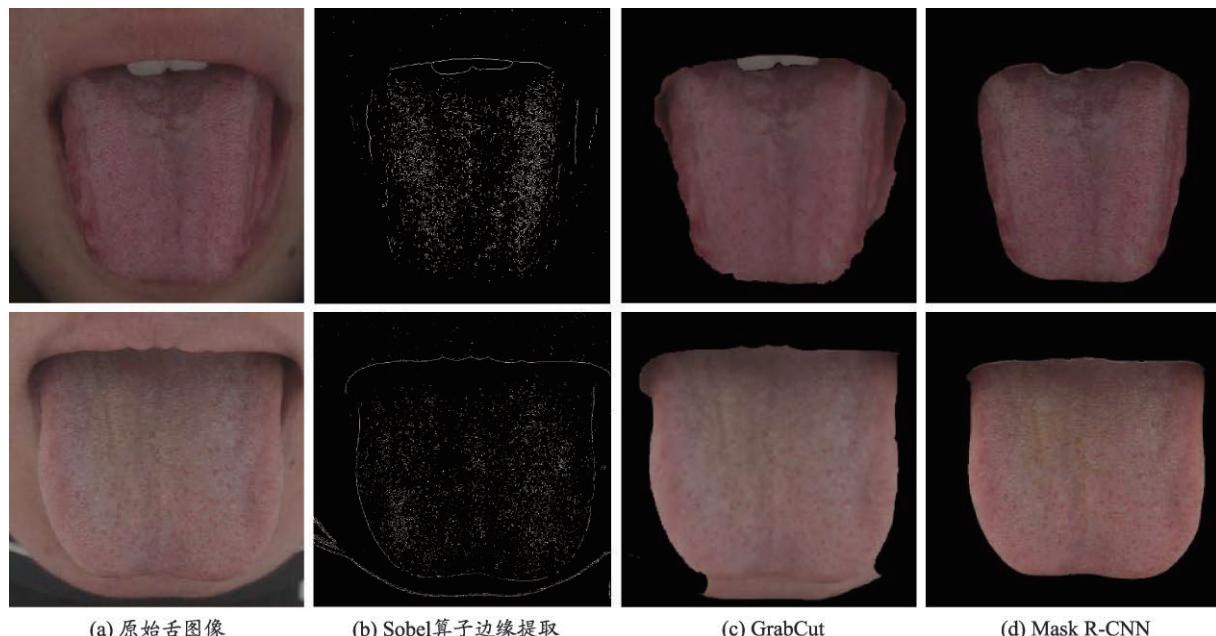


图5 GrabCut算法和Mask R-CNN分割结果对比

以上 LAB 颜色空间色差计算,主要考虑感兴趣区域同邻近背景区域之间的区域色差接近的情况。但是在舌图像样本中,有关兴趣区域和邻近背景区域的分割还会涉及到舌体区域边缘问题。在图像处理中使用较多的一个特征就是图像的边缘,即灰度值变化剧烈的地方。而边缘检测就是对一张图像的输入到输出边缘信息的整个处理过程。由于边缘检测中的 Sobel 算子具有平滑处理降低噪声的特性,故本研究采用该算子进行边缘检测。该算子先对输入图像的像素做卷积,即求像素的梯度值;然后,对生成的新像素灰度值做阈值运算,以此来确定边缘信息。由于邻域的像素对当前像素产生的影响不是等价的,距离不同的像素具有不同的卷积核,对算子结果产生的影响也不同。一般来说,距离越远,产生的影响越小。

当舌体区域的颜色同周围嘴唇或皮肤的颜色较为接近,如果舌体的边缘较为明显,舌体的分割效果较好且易于实现;当舌体区域的颜色同周围嘴唇或皮肤的颜色相差较大时,即使舌体区域的边缘不太明显甚至几乎没有,也可以较容易得到较好的舌体分割结果。但是当舌体区域与周围嘴唇或皮肤的颜色较为接近,并且舌体区域的边缘不太明显时,舌体分割就较难进行。但是本研究建立的舌图像分割模型能够较好地处理这种情况,得到较为满意的分割结果。此外,还与基于 GrabCut 的舌图像分割方法进行了结果

对比,舌图像分割结果如图 5 所示。

由图 5 可以看出,当嘴唇颜色同舌体较为接近,且舌体边缘很少甚至几乎没有时,利用 GrabCut 算法不能够完全避免嘴唇甚至牙齿的影响,得到包含牙齿和周围嘴唇部分的舌体分割结果,致使分割效果不佳;而 Mask R-CNN 能够有效地去除嘴唇及牙齿的影响,较好地从原始舌图像中分割出准确的舌体,获得较为满意的分割结果。当皮肤与舌体颜色相近,且舌体下半部分边缘不明显时,利用 GrabCut 算法无法很好地区分出舌体和同舌体颜色较为接近的皮肤,导致分割得到的舌体下半部分误将邻近的部分皮肤分割在内;而 Mask R-CNN 能够较好地减小皮肤与舌体颜色接近带来的不利影响,获得较为准确的舌体轮廓。

综上所述,Mask R-CNN 能够有效地解决舌体周围的嘴唇和皮肤颜色较接近舌体时导致舌体分割轮廓不准确的问题,获得较好的分割结果。

为了更加精确地度量 Mask R-CNN 的分割结果,根据公式(1)–(4),本研究对舌图像分割结果进行四个评估指标的计算,其结果如表 1 所示。

由表 1 可以看出,不同舌色、苔色舌图像分割的均像素准确度 MPA 均达到 92.7% 以上,平均为 93.03%,其中黄苔舌图像最高,为 93.87%;平均准确度 MA 均达到 93.1% 以上,平均为 93.44%,其中白苔舌图像最高,为 93.59%;均交并比 MIoU 均达到 84.6% 以上,平均为 86.69%,其中黄苔舌图像最高,为 87.66%;频权

交并比 MFWIoU 均达到 85.8% 以上,平均为 87.16%,其中黄苔舌图像最高,为 87.74%。综合来看,黄苔舌图像的 MPA、MIoU 及 MFWIoU 三项评估指标最高,而红舌和绛舌的 MPA、MA 两个评估指标均达到 93% 以上。因此,采用本文的舌图像分割方法得到的三类苔色、五类舌色的舌图像分割结果,将会给一些疾病诊断提供一定的舌诊客观化基础。

此外,基于 Mask R-CNN 的舌图像分割方法在分割时平均每张处理时间大约 0.21 秒,分割速度较快。这说明了该方法在舌图像分割上具有准确率高、速度快的优势。在 Mask R-CNN 整个网络结构中,Faster R-CNN 作为经典的目标检测算法,利用一个神经网络将特征提取网络、生成 ROI、ROI 的分类和 ROI 回归这 4 个模块结合起来,进行一个端到端网络的训练,实现了既快又准地完成目标检测的功能;FCN 能够生成精准的像素级掩膜,从而获得高分割准确率;ROIAlign 策略解决了 ROI Pooling 中所存在的像素偏差问题,克服了以往网络的退化难题。相比弱监督学习使用图像级别或物体级别的标注作为数据集,Mask R-CNN 作为有监督的目标实例分割框架使用像素级别的标注数据集,能够从含有完善标注信息的数据集中学习规律,进行分割模型建立,最后得到较好的分割效果。

因此,基于 Mask R-CNN 的舌图像分割方法不仅分割速度较快,而且在很大程度上减少了嘴唇或皮肤等对舌体分割结果的影响,能够得到了较好的舌体分割效果,证明 Mask R-CNN 目标实例分割框架在舌图像分割中具有较高的实用价值。

3 结束语

舌诊是中医望诊的重要内容之一,通过对舌苔、舌质的观察,从而了解病变所在,据以辨证论治。中医学认为“舌为心之候,苔乃胃之明徵”“有诸内者必形于外”。舌图像分割是舌诊客观化的重要基础,其

表 1 舌图像的定量指标评估结果

指标类型	MPA	MA	MIoU	MFWIoU
白苔	0.9330	0.9359	0.8735	0.8763
淡白舌	0.9278	0.9316	0.8469	0.8589
淡红舌	0.9344	0.9342	0.8636	0.8701
红舌	0.9312	0.9333	0.8585	0.8660
黄苔	0.9387	0.9354	0.8766	0.8774
灰黑苔	0.9379	0.9350	0.8746	0.8759
绛舌	0.9368	0.9350	0.8707	0.8741
紫舌	0.9367	0.9349	0.8714	0.8739
总均值	0.9303	0.9344	0.8669	0.8716

分割的准确性直接影响到后续舌诊分析的结果。

为了克服已有舌图像分割算法效率较低、准确率不高的缺点,本研究提出应用 Mask R-CNN 进行舌图像的分割。先利用标注工具 labelme 进行舌图像的样本标注,再基于 Mask R-CNN 建立了舌图像分割模型。对 1000 张不同舌色和苔色舌图像进行了自动分割和结果评估,四个定量评估指标均像素准确度、平均准确度、均交并比、频权交并比均高于 84.6%,达到了较好的舌图像分割效果,并与 GrabCut 算法进行了比较,体现了其良好的不受周围嘴唇及皮肤干扰的特性。因此,基于 Mask R-CNN 的舌图像分割方法速度快、准确率高,为舌图像分割研究提供了有益的参考,在舌诊客观化方面具有较好的实际应用价值。

目前,深度学习已经成为计算机视觉任务处理的研究热点,在舌图像处理领域也取得了较为满意的结果。但是深度学习在该领域的应用仍然存在一些挑战和限制,如高质量标注训练样本的不足,深度学习模型的可解释性等。然而,目前存在的挑战也为深度学习在机器视觉领域的应用提供了良好的机会和广阔的研究空间。随着深度学习算法的不断完善、计算能力的提高和大样本舌图像的采集,深度学习在舌图像处理领域的应用前景将会更加广阔。

参考文献

- 1 Jung C J, Jeon Y J, Kim J Y, et al. Review on the current trends in tongue diagnosis systems. *Integrative Medicine Research* (ISSN: 2213-4220), 2012, 1(1): 13–20.
- 2 Kass M, Witkin A, Terzopoulos D. Snakes: active contour models. *International Journal of Computer Vision* (ISSN: 0920-5691), 1988, 1 (4): 321–331.
- 3 刘关松,徐建国,高敦岳.舌图像自动分割方法.计算机工程,2003,
- 29(12): 63–64, 74.
- 4 Zhai X M, Lu H D, Zhang L Z. Application of image segmentation technique in tongue diagnosis// International Forum on Information Technology & Applications. Chengdu: IEEE Computer Society, 2009.
- 5 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* (ISSN: 0278-0062), 2005, 24

- (8): 946–956.
- 6 Zhang H Z, Zuo W M, Wang K Q, et al. A snake-based approach to automated segmentation of tongue image using polar edge detector. *International Journal of Imaging Systems and Technology* (ISSN: 0899–9457), 2006, 16(4): 103–112.
 - 7 Fu Z, Li W, Li X, et al. Automatic tongue location and segmentation// 2008 International Conference on Audio. *Shanghai: Shanghai University*, 2008.
 - 8 Wu J, Zhang Y H, Bai J. Tongue area extraction in tongue diagnosis of traditional Chinese medicine//2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. *Shanghai: IEEE Pulse*, 2006.
 - 9 Yan Z F, Ding H L, Li N M. Automatic segmentation of sublingual excrescences in color sublingual images// International Conference on Medical Biometrics. *Shenzhen: Chinese Association of Integrative Medicine*, 2014.
 - 10 Esteva A, Kuprel B, Novoa R A, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* (ISSN: 0028–0836), 2017, 542(7639): 115–118.
 - 11 刘飞, 张俊然, 杨豪. 基于深度学习的医学图像识别研究进展. *中国生物医学工程学报*, 2018, 37(1): 86–94.
 - 12 He K M, Gkioxari G, Dollar P, et al. Mask R-CNN. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2020, 42(2): 386–397.
 - 13 Hariharan B, Arbeláez, Pablo, et al. Hypercolumns for object segmentation and fine-grained localization. <http://www.oalib.com/>, (2014.11.21)[2019.1.4].
 - 14 Li Y, Qi H Z, Dai J F, et al. Fully convolutional instance-aware semantic segmentation//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). *Chengdu: IEEE Computer Society*, 2017.
 - 15 De Brabandere B, Neven D, Van Gool L. Semantic Instance Segmentation with a Discriminative Loss Function. <http://www.oalib.com/>, (2017.8.8)[2019.1.4].
 - 16 Ning J, Zhang D, Wu C, et al. Automatic tongue image segmentation based on gradient vector flow and region merging. *Neural Computing and Applications*(ISSN: 0941–0643), 2012, 21(8): 1819–1826.

Research on Tongue Image Segmentation Based on Mask R-CNN

Yan Jianjun¹, Xu Zi¹, Guo Rui², Yan Haixia³, Wang Yiqin³

(1. School of Mechanical and Power Engineering, East China University of Science and Technology, Shanghai 200237, China; 2. Institute of Interdisciplinary Research Complex, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China; 3. Laboratory of Information Access and Synthesis of Traditional Chinese Medicine Four Diagnosis, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China)

Abstract: Tongue image segmentation is one of the keys to the objectification of tongue diagnosis. When the color of the tongue is close to the lips or the skin, it will have a greater influence on the segmentation of tongue. This increases the difficulty of segmentation. Therefore, this paper proposed a tongue image segmentation method based on Mask R-CNN of convolutional neural network. In order to improve the accuracy of tongue image segmentation, firstly, the tongue image is labeled with the labeling tool labelme, and then the tongue image segmentation model of Mask R-CNN is trained and the test of tongue image segmentation is carried out. The tongue segmentation edge obtained by the tongue image segmentation method is more accurate. And the four quantitative evaluation indexes of the method are higher than 84.6%, which are Mean Pixel Accuracy, Mean Accuracy, Mean Intersection over Union and Mean Frequency Weighted Intersection over Union. This method achieves better tongue segmentation effect. It can solve the problem of inaccurate segmentation of the tongue image caused by the color of the lips and skin being closer to that of the tongue. This research provides a new idea and method for tongue image segmentation, which has certain practical value and reference significance for the objectification of tongue diagnosis.

Keywords: Tongue Image Segmentation, Tongue Diagnosis Objectification, Convolutional Neural Network, Image Annotation, Deep Learning

(责任编辑:闫群, 责任译审:钱灵姝)