

基于多任务卷积神经网络的舌象分类研究

汤一平 王丽冉 何 霞 陈 朋 袁公萍
(浙江工业大学信息工程学院 杭州 310023)

摘 要 针对现有技术难以并行实现舌象多标签的高效分类和识别,难以利用标签间的相关性进行综合分析等问题,提出了一种基于多任务卷积神经网络的舌象分类方法,构建了一种多任务联合学习模型,尝试实现传统中医舌诊中对舌色、苔色、裂纹和齿痕等多个标签的同时辨识。首先,在共享网络层对所有标签进行联合学习,从特征提取的角度自动挖掘和利用标签间的相关性;然后,在不同子网络层分别完成特定类别的学习任务,从而消除多标签分类中的歧义性;最后,训练多个 Softmax 分类器以实现对所有标签的并行预测。研究表明,所提方法能以端到端的方式同时提取舌象的多个特征并直接进行分类识别,在各分类评价指标上的最低值约为 0.96,多任务的总体识别时间为 34 ms,因此该方法在精度和速度上均具有明显优势。

关键词 舌象分类,多标签,多任务网络,相关性,迁移学习
中图法分类号 TP391.41 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.12.042

Classification of Tongue Image Based on Multi-task Deep Convolutional Neural Network

TANG Yi-ping WANG Li-ran HE Xia CHEN Peng YUAN Gong-ping
(School of Information Engineering,Zhejiang University of Technology,Hangzhou 310023,China)

Abstract It is difficult to exploit the existing methods to achieve efficient classification and identification of tongue image' labels in parallel,and it is also difficult to utilize the correlation between labels for comprehensive analysis. Aiming at the problems above,this paper proposed a classification method of tongue image based on multi-task deep convolutional neural network and constructed a multi-task joint learning model based on deep convolutional neural network to realize the simultaneous identification of tongue color,moss color,tongue crack and tooth marks in tongue diagnosis of Chinese medicine. First,the shared network layer is used to learn all labels,and the correlation between the tags is extracted and utilized automatically from the perspective of feature extraction. Then,the learning tasks of specific labels are completed in different sub-network layers to eliminate the ambiguity in the multi-label classification problem. Finally,multiple Softmax classifiers are trained to achieve parallel prediction of all labels. Experimental results suggest that the proposed method can simultaneous extract multiple features of tongue image and implement classification by means of end to end. The lowest value is about 0.96 in several evaluation indexes and the multi-task recognition rate is about 34 ms. Therefore,this algorithm has obvious advantages in accuracy and speed.

Keywords Tongue classification,Multi-label,Multi-task network,Correlation,Transfer learning

1 引言

在中医舌诊中,一张舌象可以同时反映出多种信息,这些信息大体上可分为颜色、纹理和形状三大类。其在舌体上表现为舌色、舌形等,在舌苔上表现为苔色、苔质等。因此,舌象属于多标签图像数据,且舌象的各个标签之间存在一定的相关性,明显的相关性如淡红舌多与薄白苔对应。

在多标签分类中,标签之间的关联性可以为多标签学习提供更多的有效信息,故现阶段多标签算法大都试图利用标

签之间的相关性作为先验知识来提升其分类性能^[1-4],但标签之间的相关性事先很难确定和获得,如舌象中舌色与苔色间的相关性较为明显,但舌色与裂纹间的相关性仍有待探索。

传统的舌象分类方法一般以单独的颜色、纹理或形状为主要研究内容,未考虑到多标签问题,不能对舌象进行综合分析。这与中医舌诊的整体观诊断思想不相符,同时也难以利用其中的相关性进行建模分析。文献[5]提取舌象的 RGB 像素值作为特征值,利用多种方法对 RBF 核函数进行参数寻优,训练 SVM 对舌色和苔色进行分类。文献[6]利用中医望

到稿日期:2017-11-08 返修日期:2018-02-06 本文受国家自然科学基金:基于物联网技术的生物式临震预测关键技术研究(61379078)资助。
汤一平(1958—),男,博士,教授,博士生导师,主要研究方向为全方位视觉传感器及应用、计算机视觉、机器学习等,E-mail:typ@zjut.edu.cn
(通信作者);王丽冉(1993—),女,硕士生,主要研究方向为计算机视觉、深度学习;何霞(1993—),女,硕士生,主要研究方向为计算机视觉、图像检索;陈朋(1992—),男,硕士生,主要研究方向为计算机视觉、机器学习;袁公萍(1992—),男,硕士生,主要研究方向为全方位视觉传感器及应用、机器学习。

诊的先验知识设计先验模板和标准色块作为辅助,再用 KNN 方法对颜色进行分类。文献[7]通过核假彩色变换增强裂纹与周围区域的对比度,并对裂纹进行提取。文献[8]利用道格拉斯-普克法提取齿痕舌的个数来分析齿痕舌的严重程度。

无论是对颜色的识别还是对纹理、形状的分析,现有关于特征提取的方法多为人工进行的。人工提取到的特征主要包含浅层的颜色信息,而颜色信息属于弱信息,容易受到拍摄环境的影响,且不同颜色的像素值可能较为相近,故人工提取的方法无法从根本上挖掘到具有明显区分度的特征,不能对舌象进行全面的表达。

近几年,深度学习在计算机视觉领域得到了广泛应用[9-11],这得益于卷积神经网络的快速发展[12-14]。卷积神经网络通过模拟人脑的阶层式结构逐层进行信息传递,自动挖掘数据的本质特征,这些特征在大量的任务中被证实比传统的手工提取的特征具有更强的表征能力,能够更完整地描述图像的整体结构。

目前,大多数卷积神经网络结构只支持单标签数据。利用现有的单标签网络实现舌象多标签分类的最直接方法是训练多个模型,从而将多分类转化为二分类,以消除多标签产生的歧义性问题。但将多个标签完全分开,分别进行特征提取和分类识别,也同样忽略了它们之间可能存在的关联对分类提供的有效信息。因此,解决多标签所产生的歧义性问题,自动挖掘标签之间的相关性,并有效地将其应用于分类模型中,是实现舌象多标签分类的关键问题。

为探究舌象多标签之间的关联性,实现舌象的高效、准确分类,本文从特征提取和相关性学习两个角度出发,利用深度卷积神经网络构建基于特征依赖性的多任务联合学习模型,以实现舌象的多标签分类。在不需要先验知识的条件下,以特征为探索点,自动挖掘数据间的相关性,从而提升其分类性能。本文在颜色上将舌色分为淡白舌、淡红舌、红舌、绛舌和青紫舌,将苔色分为白苔、黄苔和灰黑苔,并进行 5 种舌色和 3 种苔色的分类;在纹理和形状上,选取常见的裂纹和齿痕这两种属性进行判断。对舌象的主要属性同时进行识别分析,这也更符合中医舌诊的整体观诊断思想。

2 研究方法

2.1 整体方案

本文提出的舌象多标签识别方案以多任务联合学习模型为核心进行设计,具体操作流程如图 1 所示。其中舌体分割算法另有论文论述,不再赘述。

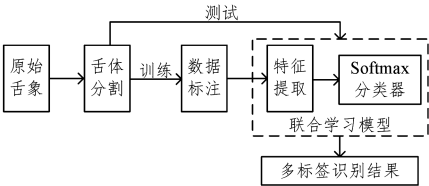


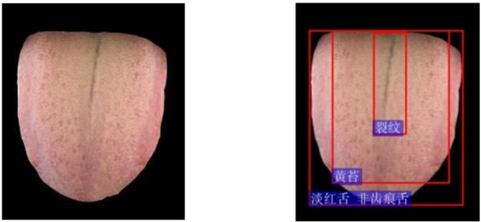
图 1 多标签识别的整体流程图

Fig. 1 Overall flow chart of multi-label recognition

2.2 数据标注

图 2 为训练数据的标记示意图。如图 2 所示,舌象同时

具有淡红舌、黄苔舌、裂纹舌和非齿痕舌 4 个标签。在进行标注时,不仅给定了类别标签,同时也给定了每个标签对应的位置坐标。该操作的优点是:1)简化多标签问题,如图 2(b)所示,将一个样本划分为 3 个样本块,减少了单个样本同时对应的标签数量;2)不需要进行苔质分离的预处理,也可以迫使网络集中关注舌象中每个标签所对应的区域,使得网络更高效、更准确地提取到该标签的特征向量,还可以避免可能的错误分离结果对颜色识别的影响;3)模型进行测试时,不仅可以预测舌象的类别标签,也可以预测属于该标签的舌面区域。这里主要针对裂纹舌,若该舌象属于裂纹舌,则可同时得到裂纹所在区域,便于对裂纹做进一步分析,如对裂纹进行分割提取等。



(a)待识别舌象

(b)标注舌象

图 2 待识别舌象及多标签舌象示意图

Fig. 2 Schematic diagram of unlabeled and multi-labeled tongue image

2.3 多任务网络结构设计

本节将从舌象分类问题本身出发,从多任务和多标签两个角度详细阐述构建网络的思路及方法。

虽然已经对多标签问题进行了简化,但如图 2(b)所示,舌色与齿痕的标注位置是一致的(所有数据标注情况与图 2(b)类似),都是对整个舌体进行类别及位置回归。该样本块同时具有两个类别标签,说明此时的舌象识别仍然属于多标签分类问题。在训练及测试过程中不可避免地会产生歧义,这也正是多标签分类中要解决的问题。一种直接、有效的方法是将多分类问题分解成多个独立的二分类问题,其具体操作作为训练两个模型,分别用于舌色的识别及齿痕舌的判别。然而,将两类标签完全分开,用两个网络分别进行训练和测试,会忽略它们之间可能存在的关联对分类提供的有效信息,也会增加时间成本、降低识别效率。

为兼顾标签的独一性与相关性,本文以增强各标签特征之间的依赖性为探索点,构建一种基于特征相关性的多任务联合学习网络对舌象进行识别,网络结构如图 3 所示。

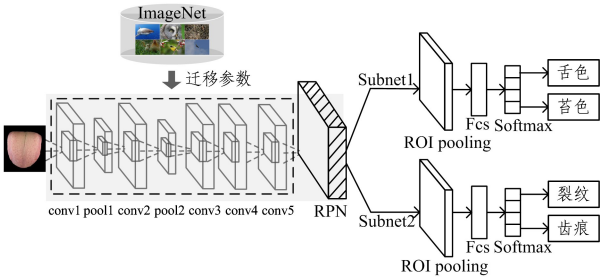


图 3 多任务联合学习网络的结构

Fig. 3 Structure of multi-task joint learning network

本文提出的多任务联合学习网络主要由两部分构成。第

一部分由图 3 中阴影框所示,网络层进行共享,对所有标签同时进行学习。共享层又分为浅层特征提取(图 3 中虚线框部分)和候选区域生成(图 3 中的 RPN^[14])两个模块。网络层参数的共享增强了各标签特征间的依赖性,隐性地地将各个标签联系在一起,从而挖掘和利用标签间的相关性。第二部分以网络分支的方式进行组建,根据图 2(b),为了消除歧义,需要将两个子网络分别用于舌色和苔色的识别训练。因此,构建两个子网络分别完成舌色和苔色、裂纹和齿痕的类别预测主任务,在每个主任务中又分别包含位置预测的子任务。最后对所有预测结果并行输出,实现对舌象多标签的同时识别。

2.4 多任务网络的训练

本节将对网络的训练策略、参数设置及训练过程等进行较为详细的阐述。

目前深度卷积神经网络的训练还依赖大量的样本,对于医学图像来说,数据的采集及标注都存在一定的困难,因此难以得到大规模的舌象数据集。利用小规模数据直接进行训练时,网络不易收敛,容易产生过拟合,难以达到预期效果。

为了解决以上问题,本文采用迁移学习^[15]策略进行网络的训练。网络的学习具有层级特性,浅层网络提取低层特征,这些特征一般具有共性,在不同的数据集上得到的低层特征相似度很大;高层网络提取抽象特征,这些特征在不同的数据集上往往具有很大的差异。因此低层网络通过迁移参数进行学习是可行的^[16-18],且有利于提高网络的收敛速度,防止出现过拟合。

图 3 中虚线框内的网络结构如下:第一个卷积层(conv1)的卷积核个数为 96,大小为 7×7 ,步长为 2;第一个池化层(pool1)的池化窗大小为 3×3 ,步长为 2;第二个卷积层(conv2)有 256 个卷积核,大小为 5×5 ,步长为 2;第二池化层(pool2)的池化窗为 3×3 ,步长为 2;第三个卷积层(conv3)有 384 个卷积核,大小为 3×3 ,步长为 1;第四个卷积层(conv4)有 384 个卷积核,大小为 3×3 ,步长为 1;第五个卷积层(conv5)有 256 个卷积核,大小为 3×3 ,步长为 1。

本文直接利用在大规模图像数据集 ImageNet^[19]上训练得到的模型对框内网络进行参数初始化,以提高收敛速度。对于后端的网络层均采用高斯函数进行参数初始化,进一步提取舌象的深层抽象特征。该部分网络的结构及参数设置将在后面做进一步说明。

舌象经过虚线框内的初步特征提取后,可得到 256 张特征图。如何在包含了所有标签信息的特征图中将不同的标签分开,从而送入不同的子网络进行训练,成为了解决同一样本块标签歧义性的关键点。在此,本文选择 RPN 网络进行候选区域的生成,RPN 网络相当于算法层,它通过区域建议的方式对目标进行定位,可分别得出图 3 中所有样本块候选区域的类别和位置信息。随后,可根据这些候选区域的类别标签,分别将其送入各自的子网络。

这两个网络模块共同构成了本文多任务联合学习模型的共享层。共享层训练过程为:网络接收任意尺寸的图像作为输入,利用卷积和池化两种操作进行初步特征提取,接着将得到的特征图作为 RPN 网络的输入。RPN 网络的具体操作为:首先在特征图上进行多尺度操作,以每个滑动窗口的中心

为中心点,生成 3 种尺度(128, 256, 512)和 3 种长宽比(1:1, 2:1, 1:2)的候选区域(Anchor),并将其映射到原舌象图中,即特征图上的每个点在原图中都对应 9 个候选区域。然后,以候选区域与真实标注框(Ground Truth, GT)的重叠率来进行候选区域的筛选。重叠率用式(1)计算:

$$IoU = \frac{Anchor \cap GT}{Anchor \cup GT} \quad (1)$$

根据经验设定阈值,将重叠率大于一定阈值(0.7)的候选区域视为正样本,将重叠率小于一定阈值(0.3)的候选区域视为负样本,并舍弃其余候选区域。

接下来,对所有正样本进行标记判断,将与其有最大重叠率的真实标注区域标签作为该正样本的标签,并为所有负样本添加背景标签。对正负样本进行随机采样,采样数目设为 128,采样比例设为 1:1。在后续的网络中正负样本一同参与训练,以增强标注样本与非标注样本的区分度。

共享层经过上述操作即可得到多个带有标签信息的正负样本,接下来根据每个样本的标签,并依据上述采样准则将其送入不同的子网络进行训练。在此,将标签为淡白舌等 5 种舌色和白苔等 3 种苔色的正样本送入子网络 1,将标签为齿痕舌、非齿痕舌和裂纹舌的正样本送入子网络 2。两个子网络结构相似,如图 4 所示,其中数字代表该层的输出神经元个数。

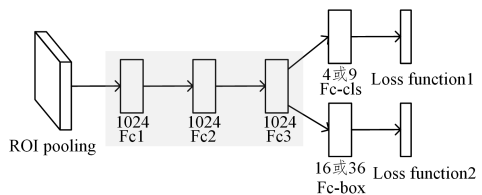


图 4 单个子网络的结构

Fig. 4 Structure of single subnet

如图 4 中阴影框区域所示,两个子网络主要利用全连接层对采样后的候选区域进行特征提取。正如前文所述,候选区域共有 9 种尺寸,而全连接层要求输入的尺寸一致,故在此先利用金字塔池化层进行尺度归一化,再将其送入 3 个全连接层进行深层特征提取。两个子网络中全连接层输出神经元的个数均设为 1024,可得到 1024 维的特征向量。接着,将该特征向量分别送入两个全连接层进行特征压缩,根据子网络所需完成的具体任务对这两个全连接层的输出神经元个数分别进行设置,在子网络 1 中设置为 9 和 36,在子网络 2 中设置为 4 和 16。最后,将输出值与真实标签值进行对比,对损失函数进行回归约束。

由于对类别及位置进行了预测,因此回归损失同时包含两个部分,以任一子网络为例,多任务损失函数可由式(2)表示:

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (2)$$

其中,类别损失函数由式(3)定义:

$$L_{cls}(p_i, p_i^*) = -\log[p_i^* p_i + (1 - p_i^*)(1 - p_i)] \quad (3)$$

位置回归损失函数由式(4)定义:

$$L_{reg}(t_i, t_i^*) = R(t_i - t_i^*) \quad (4)$$

其中, R 是鲁棒的损失函数 $smooth_{L1}$, 如式(5)所示:

$$smooth_{L1}(x)=\begin{cases}0.5x^2, & \text{if } |x|<1 \\ |x|-0.5, & \text{otherwise}\end{cases}\quad (5)$$

其中, N_{cls} 和 N_{reg} 是为避免过拟合的正则项; λ 是权重系数; i 是该候选区域的类别索引值; t_i 是该候选区域的预测坐标偏移量; t_i^* 是该候选区域的实际坐标偏移量; p_i 是预测候选区域属于第 i 类的概率; p_i^* 表示其真实类别, $p_i^* = 0$ 表示其不属于第 i 类, $p_i^* = 1$ 表示其属于第 i 类。

通过这两个损失函数分别计算预测值与给定真实值之间的误差, 利用反向传播算法将误差层层回传, 利用随机梯度下降法对每层的参数进行调整和更新, 使得网络的预测值更接近真实值, 即最后两个全连接层的输出更接近给定标注值中的类别和位置信息。

3 实验研究与分析

3.1 舌象数据集

本文从舌象库中选取了 1222 幅舌图像, 库中舌象由专业的仪器采集得到。选取的舌象要求舌体充分暴露, 舌面平展, 且均按比例缩放为最长边为 400 的图像。每种标签的舌象数目如表 1 所列。考虑到数据有限, 本文将数据集随机分为 5 组, 每次选择 4 组用于训练, 剩余 1 组用于测试。本文进行五折交叉实验, 以验证所提方法的可靠性。

表 1 实验数据
Table 1 Experimental data

标签	数目	标签	数目
淡白舌	233	白苔	279
淡红舌	330	黄苔	214
红舌	291	灰黑苔	62
绛舌	107	齿痕舌/非齿痕舌	76/1146
青紫舌	261	裂纹舌	1100

3.2 实验环境

具体的实验环境如表 2 所列。

表 2 实验环境
Table 2 Experimental environment

名称	参数
系统	LINUX64 Ubuntu14.04
框架	Caffe
语言	Python, C++
CPU	Intel®CoreTM i7-2600 3.4GHz×8
GPU	TITAN X (Pascal)
RAM	8.00GB

3.3 分类评价准则

多标签分类评价准则可分为两种类型: 1) 以样本为出发点, 计算样本的实际标记集合与预测标记集合之间的误差; 2) 以类别为出发点, 衡量分类器在单类别上二分类的结果。对于分类问题来说, 以样本为出发点进行评价更为综合。为全面评价分类算法的可靠性, 本文以基于图像(即基于样本)的方式为主进行评价, 将基于类别的评价方式作为辅助, 对每类的识别结果进行详细分析。每种方式都选择精确度 P 、召回率 R 、 F 值及杰卡德相似度 J 4 项指标。

如图 5 所示, 基于图像预测所含标签时, I 为所有标签, TP 为预测结果中属于该图像的标签, FP 为预测结果中不属

于该图像的标签, FN 为未被预测出来的标签; 基于类别预测所含图像时, I 为所有图像, TP 为预测结果中属于该类的图像, FP 为预测结果中不属于该类的图像, FN 为未被正确预测为该类的图像。

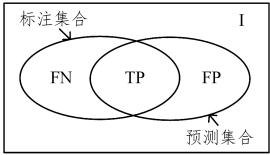


图 5 评价准则原理图

Fig. 5 Schematic diagram of evaluation criteria

精确度 P 、召回率 R 、 F 值及杰卡德相似度 J 的计算式分别如式(6)–式(9)所示:

$$P=\frac{TP}{TP+FP}\quad (6)$$

$$R=\frac{TP}{TP+FN}\quad (7)$$

$$F=\frac{2*P*R}{P+R}\quad (8)$$

$$J=\frac{TP}{TP+FP+FN}\quad (9)$$

P 衡量的是分类器的查准率, 与错分率相对; R 衡量的是分类器的查全率, 与漏分率相对; P 与 R 一般相互影响, F 则将 P 和 R 结合起来综合考虑, 是反映整体精度的评价指标; J 是更严格的评价指标。4 个指标的取值均在 0~1 之间, 且越接近 1 表明分类效果越好。

3.4 不同子网络结构的比较

网络学习效果取决于其不同的结构(表现为网络层数和输出神经元数目的差异), 故在此对子网络的结构进行分析。本文随机选择一组实验数据, 利用不同的子网络结构对其进行训练和测试。在评价分类效果时, 选择基于图像的方式, 并着重比较更为综合的评价指标 F 和 J 的值, 实验结果如表 3 所列。其中, 网络结构中包含该层用“√”表示, 其后的数值表示该层的输出个数, 用“×”表示网络中不包含该层。

表 3 不同子网络结构的分类结果对比(基于图像)

Table 3 Comparison of classification results of different subnet structures (based on images)

Fc1	Fc2	Fc3	P	R	F	J
√/4096	×	×	0.9850	0.9857	0.9844	0.9762
√/2048	×	×	0.9887	0.9846	0.9861	0.9794
√/1024	×	×	0.9896	0.9891	0.9885	0.9834
√/4096	√/4096	×	0.9765	0.9925	0.9831	0.9725
√/2048	√/2048	×	0.9891	0.9891	0.9884	0.9820
√/1024	√/1024	×	0.9876	0.9921	0.9892	0.9826
√/4096	√/4096	√/4096	0.9814	0.9896	0.9845	0.9757
√/2048	√/2048	√/2048	0.9846	0.9911	0.9872	0.9797
√/1024	√/1024	√/1024	0.9919	0.9928	0.9919	0.9871

由于舌象数据较自然图像数据而言更简单, 因此利用较少的输出参数就可获取到足够的信息, 参数过多反而会误学习到冗余信息, 不利于特征的提取。如表 3 所列, 在全连接层数一定的情况下, 在适当范围内参数越少, 识别精度(F 值和 J 值)越高。由表 3 中的信息可知, 本文利用 3 个全连接层提取深层特征, 每个全连接层的输出数目设为 1024。

3.5 实验结果与分析

为了说明本文方法的有效性,设计了两组对比实验:将本文方法与目前常用的舌象分类方法(由于分步进行,涉及方法较多,故在后文中统称为传统方法)和 Faster R-CNN 进行对比。

传统分类方法无法实现多特征的同时提取及多标签的同时识别,故需要分步进行。首先,综合考虑目前的颜色分类方法,提取舌质与苔质的 HSV 值作为特征向量,采用 KNN 法对舌色和苔色进行分类, K 分别取 5~10,选择分类精度最高的结果进行对比;其次,利用文献[20]中的方法分别提取裂纹

与齿痕特征值,设置合适的阈值(裂纹舌为 0.1 和 0.5,齿痕舌为 0.75)判断是否为裂纹舌或齿痕舌。Faster R-CNN 属于单标签网络,不能同时对舌色和齿痕进行判别,故需要训练两个模型来分别进行舌色和苔色、裂纹和齿痕的识别。

表 4 和表 5 分别为基于图像预测所含标签、基于类别预测所含图像时不同方法的 P, R, F, J 的均值及标准差。根据表 4 和表 5 中的数据可衡量不同分类方法在所有图像上的分类性能,也可对各方法在每个类别上的分类效果进行综合考量。图 6 为不同方法在每类上的 F 值和 J 值的均值和标准差。下面对这些方法在每个类别上的分类情况进行详细分析。

表 4 不同方法的分类结果(基于图像)

Table 4 Classification results of different methods (based on images)

方法	P	R	F	J
传统方法	0.8018 ± 0.0039	0.7311 ± 0.0041	0.7581 ± 0.0030	0.6651 ± 0.0030
Faster R-CNN	0.9731 ± 0.0076	0.9908 ± 0.0023	0.9803 ± 0.0054	0.9678 ± 0.0086
多标签 Faster R-CNN	0.9898 ± 0.0029	0.9888 ± 0.0014	0.9885 ± 0.0018	0.9820 ± 0.0026
本文方法	0.9894 ± 0.0034	0.9907 ± 0.0026	0.9892 ± 0.0030	0.9829 ± 0.0044

表 5 不同方法的分类结果(基于类别)

Table 5 Classification results of different methods (based on classes)

方法	P	R	F	J
传统方法	0.7380 ± 0.0226	0.7025 ± 0.0120	0.6971 ± 0.0189	0.5675 ± 0.0159
Faster R-CNN	0.9589 ± 0.0078	0.9789 ± 0.0091	0.9674 ± 0.0074	0.9395 ± 0.0132
多标签 Faster R-CNN	0.9815 ± 0.0054	0.9812 ± 0.0065	0.9810 ± 0.0050	0.9637 ± 0.0090
本文方法	0.9833 ± 0.0055	0.9845 ± 0.0079	0.9837 ± 0.0064	0.9688 ± 0.0116

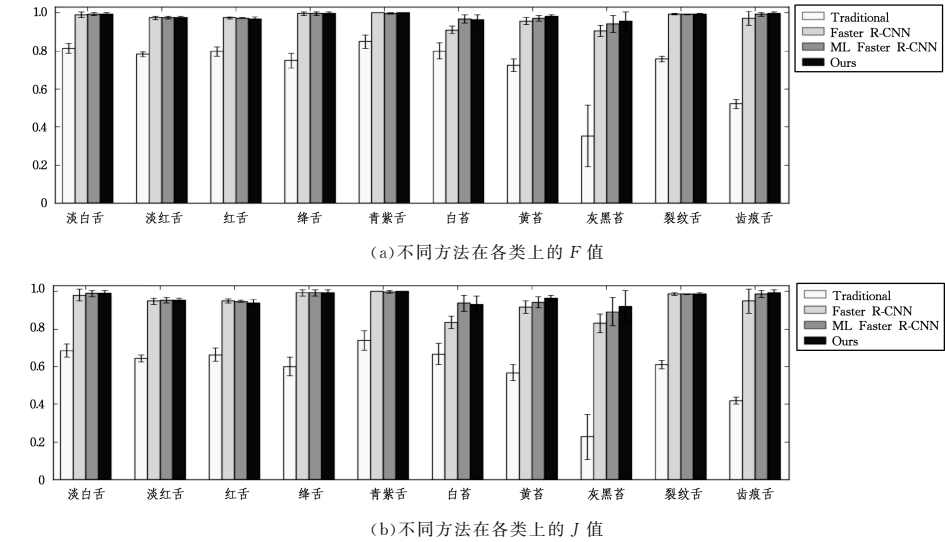


图 6 不同方法在各类上的 F 值和 J 值(基于类别)

Fig. 6 F and J values of different methods in all classes (based on classes)

由于人工提取特征主要利用了颜色信息,不能得到具有完整表达力的区分性特征,因此整体分类效果不佳。特别地,齿痕舌的判别主要利用了舌体的边缘信息,而人的舌体形状各异,故直接提取边缘信息、利用单一阈值进行判断有较大的片面性和局限性,不能很好地进行识别。从表 4 和表 5(第 2, 3, 5 行)可以直观地看出,基于深度卷积神经网络的分类方法明显优于人工提取特征再分类的传统方法,证明了深度学习方法应用在舌象分类中的可行性。从整体上看,本文提出的多任务联合学习方法在分类效果上明显优于 Faster R-CNN。

本文提出的多任务联合学习模型在利用深度卷积神经网

络自动提取特征的基础上,通过网络的共享增强了特征之间的依赖性,从而利用标签相关性进一步提升分类性能。为进一步验证本文思路的正确性,对单标签网络 Faster R-CNN 进行改进,得到与本文结构类似的网络(共享层与子网络层相结合的模式),以实现多个标签的同时判别,简称为多标签 Faster R-CNN。同样进行五折交叉实验,结果如表 4 和表 5(第 4 行)所列。可以看出,改进后的多标签 Faster R-CNN 较原 Faster R-CNN 分类性能有明显提升。特别地,图 6 展示了改进前后两种方法在每个类别上的分类结果,可以看出原本识别精度较低的白苔、黄苔等几类的分类效果都得到了明显提升。

该实验结果表明,本文对网络进行共享的方法可以实现对标签相关性的有效利用,对标签之间相关性的考虑有利于提高舌象分类性能;且只需训练一个模型即可实现对多个标签的识别,实现端到端的分类识别。

本文的舌象多标签分类方法在完成类别预测这一主任务的同时,还完成了位置预测的辅助任务。这里的位置预测主要对裂纹舌有较大帮助,可作为先验知识对裂纹舌做进一步分析。

下面对裂纹位置预测的效果做进一步分析。选用上文提及的分类评价准则,如图 5 所示, I 表示舌象图中的所有像素点集合, $TP+FP$ 表示预测框内的所有像素点集合, $TP+FN$ 表示标注框内的像素点集合。选择较为严格的评价指标 J ,计算 5 次实验中 J 的均值和标准差,以对裂纹舌的位置预测进行评价,并将其与原 Faster R-CNN、改进后的多标签 Faster R-CNN 进行对比,对比结果如表 6 所列。

表 6 舌裂纹定位评价结果

Table 6 Position evaluation results of tongue crack

方法	时间/(s/张)
Faster R-CNN	0.7731±0.0394
多标签 Faster R-CNN	0.8094±0.0174
本文方法	0.8160±0.0069

从表 6 可以看出,考虑到标签相关性的多标签 Faster R-CNN 及本文方法不仅使分类性能有所提升,对标注位置的预测也更准确,对裂纹能够进行较好的区域检测,有助于裂纹的后续处理。

经统计,在训练阶段,Faster R-CNN 用时 1.7~3.0 h,多标签 Faster R-CNN 用时 2.3~3.5 h,本文网络用时 1.9~2.1 h。对于深度学习方法而言,可直接利用训练好的模型进行特征提取与分类,故实际测试的耗时更值得关注。表 7 列出了各方法完成所有标签预测所需的时间。

表 7 分类所需时间的对比

Table 7 Comparison of classification time

方法	时间/(s/张)
传统方法	5.788
Faster R-CNN	0.078
多标签 Faster R-CNN	0.047
本文方法	0.034

从表 7 可以看出,基于深度学习的方法在效率上远远优于传统方法。由于进行了网络共享,减少了总的网络层数,使得改进后的多标签 Faster R-CNN 网络较原 Faster R-CNN 减少了约 40% 的时间消耗;且本文方法的效率最高,能够满足分类的实时性。

此外,本文在训练过程中采用了迁移学习策略,基于在大规模数据集 ImageNet 上训练得到的模型对低层网络进行微调和优化。训练过程中的损失情况是衡量模型是否表现良好的重要指标,损失值越小表明模型训练得越好。因此,在 5 组实验中随机选择一组做进一步分析,分别记录本文网络在采用迁移学习策略前后分别迭代 10 万次的损失值。如图 7 所示,横坐标为迭代次数(单位为万次),纵坐标为损失值。从图 7 可以直观地看出,采用了迁移学习策略后,网络损失值的下

降速度明显加快,波动幅度明显变小,最终趋于稳定时的损失值也更小。这说明采用迁移学习的训练策略有助于提高网络的收敛速度,改善网络对舌象分类的效果。

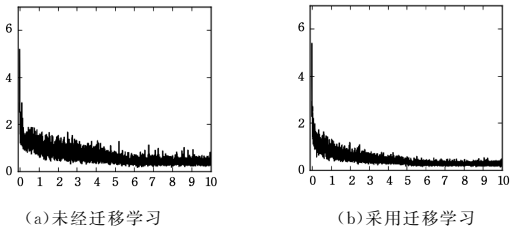


图 7 损失下降情况的对比

Fig. 7 Comparison of declining loss

综上所述,在利用卷积神经网络的基础上采用迁移学习策略进行深层特征的提取,并将标签相关性应用至分类系统中,可实现多标签的同时识别,使得本文方法在识别精度和速度上均具有明显优势。同时,对 Faster R-CNN 网络进行多标签化改进,进一步验证了本文方法的正确性。

结束语 本文以多标签之间的相关性为出发点,利用深度卷积神经网络构建多任务联合学习模型。在实现舌象多个特征自动提取的同时,增强了各特征之间的依赖性,利用数据间的相关性提升其分类性能,实现了端对端的舌象多标签分类,提高了分类的效率。且本文提出的网络结构简单,易于扩展和调整,可用于更多标签的分类任务。

然而,本文仍然存在一些不足之处。虽然本文进行了五折交叉实验来验证方法的可靠性,但更大的数据量才更能充分发挥深度学习方法的优点,更具有说服力。因此,今后还需要促进与医院的紧密合作,进一步采集数据。

苔质是舌象的重要特征之一,可细分为腐腻、润燥、厚薄等。接下来的工作首先是对多任务网络做进一步的改进,对多种苔质进行深入研究,得到更全面的舌象类型;其次是在舌象分析取得有效成果的前提下,加强与中医专家的合作,在临床实践的基础上研发一种基于深度学习的舌象自诊健康云服务平台,这也是本研究的最终目标。

参 考 文 献

[1] LEE J,KIM H,KIM N R,et al. An approach for multi-label classification by directed acyclic graph with label correlation maximization[J]. Information Sciences,2016,351(C):101-114.

[2] WU J,YE C,SHENG V S,et al. Active learning with label correlation exploration for multi-label image classification[J]. Iet Computer Vision,2017,11(7):577-584.

[3] LI S L,LIU R,LIU H. Multi-Label Learning for Improved RBF Neural Networks[J]. Computer Science,2015,42(4):316-320. (in Chinese)

李书玲,刘蓉,刘红. 改进型 RBF 神经网络的多标签算法研究[J]. 计算机科学,2015,42(4):316-320.

[4] XU X D,YAO M H,LIU H W,et al. Pre-processing Method of Multi-label Classification Based on kNN[J]. Computer Science, 2015,42(5):106-108,131. (in Chinese)

徐晓丹,姚明海,刘华文,等. 基于 kNN 的多标签分类预处理方法[J]. 计算机科学,2015,42(5):106-108,131.

[5] ZHANG Y F, HU G Q, ZHANG X F, et al. An algorithm study on tongue color recognition of patients[J]. Beijing Biomedical Engineering, 2016, 35(1): 7-11. (in Chinese)

张艺凡,胡广芹,张新峰,等. 基于支持向量机的痤疮患者舌苔色识别算法研究[J]. 北京生物医学工程, 2016, 35(1): 7-11.

[6] WANG Y G, YANG J, ZHOU Y, et al. Tongue Image Color Recognition in Traditional Chinese Medicine[J]. Journal of Biomedical Engineering, 2005, 22(6): 1116-1120. (in Chinese)

王永刚,杨杰,周越,等. 中医舌象颜色识别的研究[J]. 生物医学工程杂志, 2005, 22(6): 1116-1120.

[7] YANG Z H, ZHANG D P, LI N M. Kernel False-Colour Transformation and Line Extraction for Fissured Tongue Image[J]. Journal of Computer-Aided Design and Computer Graphics, 2010, 22(5): 771-776. (in Chinese)

杨朝辉,张大鹏,李乃民. 裂纹舌图像的核假彩色变换及其纹线提取[J]. 计算机辅助设计与图形学学报, 2010, 22(5): 771-776.

[8] ZHUMU L M, LU P, XIA C M, et al. Research on Douglas-Peucker Method in Feature Extraction from 55 Cases of Tooth-Marked Tongue Image[J]. Chinese Archives of Traditional Chinese Medicine, 2014, 32(9): 2138-2140. (in Chinese)

朱穆朗玛,陆萍,夏春明,等. 基于道格拉斯-普克法提取 55 例齿痕舌图像特征研究[J]. 中华中医药学刊, 2014, 32(9): 2138-2140.

[9] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C] // Proceedings of European Conference on Computer Vision. Amsterdam: Springer International Publishing, 2016: 21-37.

[10] JONATHAN L, EVAN S, TREVOR D. Fully convolutional networks for semantic segmentation[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 3431-3440.

[11] SUN Y, WANG X G, TANG X O. Deep Learning Face Representation from Predicting 10,000 Classes[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 1891-1898.

[12] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 580-587.

[13] GIRSHICK R. Fast R-CNN[C] // Proceedings of the IEEE Conference on International Conference on Computer Vision. Boston: IEEE, 2015: 1440-1448.

[14] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.

[15] HOOCHANG S, ROTH H R, GAO M, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning [J]. IEEE Transactions on Medical Imaging, 2016, 35(5): 1285-1298.

[16] DONAHUE J, JIA Y Q, VINYALS O, et al. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition [C] // Proceedings of the 31st International Conference on Machine Learning. 2014: I-647-I-655.

[17] RAZAVIAN A S, AZIZPOUR H, SULLIVAN J, et al. CNN features off-the-shelf: An Astounding Baseline for Recognition [C] // 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CV PRW). 2014: 512-519.

[18] YOSINSKI J, CLUNE J, BENGIO Y, et al. How transferable are features in deep neural networks? [C] // Proceedings of the IEEE Conference on Neural Information Processing Systems. Montreal: MIT Press, 2014: 3320-3328.

[19] DENG J, DONG W, SOCHER R, et al. ImageNet: A large-scale hierarchical image database[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Miami: IEEE, 2009: 248-255.

[20] KAN H X, ZHANG L Y, DONG C W. A Tongue Image Recognition Method Based on Type 2 Diabetes Traditional Chinese Medicine Syndrome Classification[J]. Chinese Journal of Biomedical Engineering, 2016, 35(6): 658-664. (in Chinese)

阙红星,张璐瑶,董昌武. 一种 2 型糖尿病中医证型的舌图像识别方法[J]. 中国生物医学工程学报, 2016, 35(6): 658-664.

(上接第 234 页)

[12] RATLIFF L J, BURDEN S A, SASTRY S S. Characterization and computation of local Nash equilibria in continuous games [C] // Proceedings of the 2013 51st Annual Allerton Conference on Communication, Control, and Computing. Allerton: IEEE Computer Society, 2013: 917-924.

[13] MIRZA M, OSINDERO S. Conditional Generative Adversarial Nets[J]. arXiv preprint arXiv:1411.1784, 2014.

[14] IOFFE S, SZEGEDY C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[C] // Proceedings of the 32nd International Conference on Machine Learning. Lille, France: International Machine Learning Society, 2015: 448-456.

[15] TREFNY J, MATAS J. Extended Set of Local Binary Patterns for Rapid Object Detection[C] // Proceedings of the 15th Computer Vision Winter Workshop. Nove Hradý, Czech Republic: CVWW, 2010: 1-7.

[16] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features[C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Kauai: IEEE Computer Society, 2001: 511-518.

[17] GULRAJANI I, AHMED F, ARJOVSKY M, et al. Improved Training of Wasserstein GANs[C] // Proceedings of Annual Conference on Neural Information Processing Systems. Long Beach, CA, USA, 2017: 5769-5779.

[18] KINGMA D P, BA J. Adam: A Method for Stochastic Optimization[J]. arXiv preprint arXiv:1412.6980, 2014.

[19] GANGNET M, BLAKE A. Poisson image editing [C] // Proceedings of the ACM SIGGRAPH. New York: Association for Computing Machinery, 2003: 313-318.