DOI: 10.13382/j.jemi.B2407273

改进 RetinaNet 的舌面齿痕和裂纹检测模型*

曹溪源 张德龙 朱枭龙 张志东 薛晨阳
(中北大学仪器科学与动态测试教育部重点实验室 太原 030051)

摘 要:中医舌诊通过观察舌特征能够进行脏腑虚实和功能盛衰的判断,具有无创便捷等优势。伴随着计算机视觉技术的飞速发展与广泛应用,开发一种能够进行自动检测、提取和识别舌象特征的模型至关重要。面向中医临床及健康监测对舌诊数字化的需求,提出了一种基于改进 RetinaNet 的舌面齿痕和裂纹特征自动检测模型。该模型首先在 RetinaNet 基准模型的骨干网络中引入 SimPSA-ResNet 模块和 SimSPPF 模块,用以增强网络的特征提取能力和模型的鲁棒性;同时,改进多级特征金字塔网络结构,提高模型的特征融合能力,进一步聚焦舌面特征的关键信息;最后,去除冗余输出特征层,并结合 ASFF 结构,保留重要的特征信息,提高信息利用率。将改进后的 RetinaNet 模型在自制的舌象数据集中进行训练和预测,得到的平均检测精度(mAP)为 94.37%,相较原算法提升了 2.77%。实验结果表明改进 RetinaNet 模型能够有效提高舌面齿痕和裂纹特征的检测精度,有助于用户的日常自检、健康管理以及辅助医生进行诊断。

关键词: RetinaNet;深度学习;目标检测;异常舌;舌诊
中图分类号: TP391.41;TN911.73 **文献标识码**: A **国家标准学科分类代码**: 520.20

Improved tongue tooth mark and fissure detection model of RetinaNet

Cao Xiyuan Zhang Delong Zhu Xiaolong Zhang Zhidong Xue Chenyang
(Key Laboratory of Instrumentation Science & Dynamic Measurement of Ministy of Education, North University of China, Taiyuan 030051, China)

Abstract: Tongue diagnosis in traditional Chinese medicine (TCM) judges the deficiency and strength of internal organs as well as the vitality of functions by observing tongue features. It has the advantages of being non-invasive and convenient. Accompanied by the rapid development and wide application of computer vision technology, it is crucial to develop a model that can perform automatic detection, extraction and recognition of tongue features. Toward demands for digital tongue diagnosis in traditional Chinese medicine clinic and health monitoring, an automatic detection model for tongue tooth mark and fissure features was proposed based on improved RetinaNet. The SimPSA-ResNet and SimSPPF module were introduced into the backbone of RetinaNet to enhance the feature extraction capability and robustness of the network. Meanwhile, the multi-level feature pyramid network structure was improved to ensure that the model can better integrate information from different scales, thereby focusing more accurately on the key information pertinent to tongue features. Finally, to further streamline the model's output, redundant output feature layers were eliminated and integrated with the Attention-guided Spatial Feature Fusion structure. This step helps retain important features while improving the utilization of information within the network. The improved RetinaNet model was trained and predicted by using the self-built tongue image dataset, and the mean average precision(mAP) reaches 94.37%, which is 2.77% higher than that of the original algorithm. Experimental results conclusively demonstrate that the improved RetinaNet model can effectively elevate the detection accuracy of tongue tooth mark and fissure features. This advancement holds tremendous potential for facilitating daily self-examination, health management and assisting doctors in diagnosis.

Keywords: RetinaNet; deep learning; object detection; abnormal tongue; tongue diagnosis

收稿日期: 2024-02-15 Received Date: 2024-02-15
*基金项目:山西省重点研发计划(202102130501011)、国家自然科学基金(62204231)项目资助

0　引　言

计算机视觉研究从生物视觉中获取灵感和启发,其核心目的是通过算法识别图片或视频中所需的信息[1]。随着信息技术的发展,计算机视觉所适用的领域不断拓宽,例如视频监控、行为检测、工业安防等应用[2-3]。舌诊作为中医望诊的重点内容之一,通过观察舌质和舌苔的颜色、形状进而做出病理判断[4-6],具有无接触、无创、实时、便捷等优势[7]。将计算机视觉技术与传统舌诊相结合,可以实现舌象特征的自动提取,为舌象诊断提供新的方法和思路。

齿痕和裂纹是常见的两种异常舌特征,在舌诊的研究中引起众多学者的关注。齿痕即舌体的边缘存在牙齿的痕迹,一般出现在舌体的两侧,呈现锯齿状。舌裂纹是指舌面存在数量、形态和分布不同的纹理[8]。在检测异常舌特征时,舌体容易和嘴唇混淆,其伸展姿势的差异也会影响判断,且不同患者的舌头颜色深浅纹理不同,因此,舌面齿痕和裂纹的自动检测是一个具有挑战性的任务。

近年来,研究人员利用传统图像处理的方法进行异常舌的识别。张璐瑶等[9]基于舌体的红绿蓝颜色空间(red-green-blue,RGB)图像,选择蓝色(B)颜色分量作为裂纹的灰度图像,采用局部灰度法分离裂纹区域和背景区域,并且为了节省处理时间利用区域一致性原理对舌裂纹特征进行预判断。基于 RGB 颜色分量的舌图像处理方法操作简单,且计算量小,但是存在一定的噪声导致提取不准确。钟少丹等[10]提出利用色相-饱和度-明度颜色空间(hue-saturation-value,HSV)彩色空间来分析齿痕舌,用阈值分割的方法来反映舌体的基本轮廓,具有处理速度较快的优势,然而这个方法提取的舌象边缘较为粗糙,同样面临噪声的问题。Shao 等[11]提出了一种结合舌边缘凹面特征和亮度变化特征的分类决策算法,基于阈值分析来进行齿痕检测,但是这种方法容易被图片中的其他因素干扰,且依赖于亮度的阈值。

随着医学科学和人工智能的融合发展,深度学习逐渐在舌象特征检测中崭露头角。Li 等[12]提出一种识别舌齿痕的三阶段方法,首先生成所有可能的齿痕区域,然后使用卷积神经网络(convolutional neural network,CNN)提取这些区域的特征向量,最后通过多实例支持向量机(support vector machine,SVM)实现舌头图像的分类。Hu 等[13]提出了基于 CNN 的 TongueNet 模型,用以提取舌裂纹,并根据特征进行三种证候的分类,实现了综合症的定量和客观诊断。王一丁等[14]对 U-Net 分割算法进行改进,增加了压缩激励(squeeze and excitation,SE)通道注意力机制,并应用到舌裂纹分割任务中。Weng 等[15]以 YOLO 模型为基础提出一种弱监督方法,可以利用粗糙标注的数据集来训练齿痕和裂纹检测模型,但是检测效果较差。综上可知,相较于传统图像处理方法,基于深度学习的检测模型拥有较好的泛化能力,不容易被图片中的单一因素所干扰。然而现阶段的研究大多只对齿痕或裂纹进行单一特征识别,没有将两个可检测的特征进行同步处理,或者对于两种特征检测的准确率较低。

本文针对舌面齿痕和裂纹两种特征提出了基于改进 RetinaNet 算法的自动检测模型。首先,在骨干网络中引入了改进后的 SimPSA-ResNet 模块和 SimSPPF 以增强网络的特征提取能力和鲁棒性;其次,优化了 RetinaNet 的多级金字塔特征融合结构,增加对图像细节的感知能力;另外,针对齿痕和裂纹特征的特点,去除了冗余的输出特征层,在保证检测精度不变的同时减少了模型参数;最后将输出特征层与自适应特征融合结构(adaptive structure feature fusion,ASFF)结合,进一步提高信息利用率。最终训练结果 mAP 为 94.37%,能够较为精确检测齿痕和裂纹的位置。改进 RetinaNet 优于其他主流目标检测算法,且较初始模型提高了 2.77%,验证了优化措施的有效性,并且可以同时检测齿痕和裂纹两种特征。因此,基于深度学习的目标检测算法为异常舌检测提供了新的可选方案。

1　改进 RetinaNet 模型

随着深度学习与卷积神经网络研究的逐渐成熟,目标检测已经成为了计算机视觉领域中的一个重要热点。RetinaNet 是 Lin 等[16]在 2017 年提出的单阶段目标检测模型,结构如图 1 所示。其中采用 ResNet[17]作为特征提取网络,经过一系列的卷积和池化等操作后生成包含深层语义信息的特征图;利用特征金字塔网络(feature pyramid networks,FPN)融合不同级别的特征信息,实现多尺度目标检测;最后采用两个全卷积子网络分别进行目标类别的预测和位置回归。

RetinaNet 在传统的二分类交叉熵损失函数的基础上,提出了 Focal loss 损失函数,可以自动判断正负样本并决定该样本对总损失的贡献,解决了由于类别不平衡造成检测精度低的问题,这也是 RetinaNet 的优势所在。Focal loss 的计算式如式(1)所示。

$$FL(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t) \tag{1}$$

式中:p_t 表示训练样本类别预测的置信度;γ 为调制因子。

通过对 RetinaNet 模型进行改进,提出一种可用于舌面齿痕和裂纹检测的模型,主要改进工作如下:1)利用 SimAM 注意力机制和 PSA 模块,构造 SimPSA-ResNet 模块,增强网络的特征提取能力;2)引入 SimSPPF 模块,提

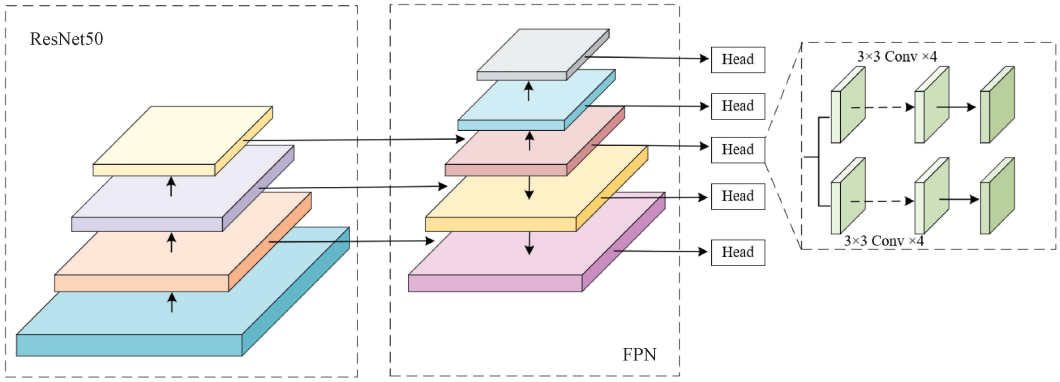


图 1 RetinaNet 框架

Fig. 1 Framework of RetinaNet

高多尺度目标的检测精度;3)改进 FPN 结构,提高特征信息的融合能力;4)去除冗余的输出层,并且添加 ASFF 结构,进一步增强不同尺度的特征融合。

1.1 SimPSA-ResNet 模块

RetinaNet 的骨干网络采用了经典模型 ResNet-50,并由后 3 个特征层的输出进行特征金字塔结构的特征融合。针对 ResNet 网络中缺乏注意力机制的问题,通过在基础的 ResNet Block 中添加金字塔压缩注意力(pyramid squeeze attention, PSA)和简单无参注意力(simple parameter-free attention module,SimAM)模块,能够更加有效的提取特征信息,增强骨干网络的特征提取能力。如图 2 所示,将 ResNet Block 中的 3×3 卷积层替换为了 PSA 模块,并在后面添加了能够更加有效的提取特征信息,增强骨干网络的 SimAM 模块。PSA 模块能够得到多尺度信息表现能力更丰富的特征,而 SimAM 模块能够在不引入额外的训练参数的情况下增强网络的特征提取能力,使模型更加关注重要信息。为了避免过多的进行注意力操作,将骨干网络中第一个 ResNet Block 模块替换成了 SimPSA-ResNet 模块。

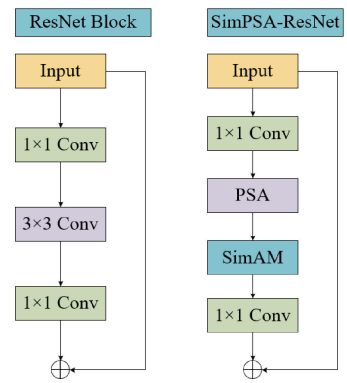


图 2 SimPSA-ResNet 模块结构

Fig. 2 Strtcture of SimPSA-ResNet module

1)SimAM 模块

SimAM 模块是 Yang 等[18]提出的一种 3D 注意力模块,结构如图 3 所示,其通过一种能量函数来发掘每一个神经元的重要性并分配权重。

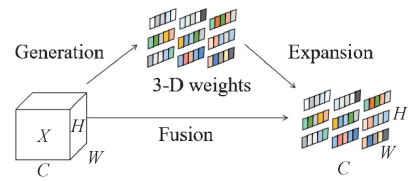


图 3 SimAM 模块结构

Fig. 3 Structure of SimAM module

与传统的通道和空间注意力机制不同,SimAM 模块更加关注局部相似性而并非全局上下文信息,没有额外的参数,可以随意的嵌入到卷积神经网络中。其为神经元定义的能量函数如式(2)所示。

$$e_t(w_t,b_t,y,x_i)=(y_t-\hat{t})^2+\frac{1}{M-1}\sum_{i=1}^{M-1}(y_0-\hat{x}_i)^2 \tag{2}$$

式中:$\hat{t}=w_t t+b_t$;$\hat{x}_i=w_t x_i+b_t$。t 和 x_i 则是目标神经元和输出特征。i 是空间维度上的索引,$M=H\times W$ 则为该通道上的神经元数量。

对 y_t 和 y_o 采用二值标记,并且添加正则项后的能量函数如式(3)所示。

$$e_t(w_t,b_t,y,x_i)=\frac{1}{M-1}\sum_{i=1}^{M-1}(-1-(w_t x_i+b_t))^2+(1-(w_t t+b_t))^2+\lambda w_t^2 \tag{3}$$

式中:λ 为正则化系数。其解析解为:

$$w_t=-\frac{2(t-\mu_t)}{(t-\mu_t)^2+2\sigma_t^2+2\lambda} \tag{4}$$

$$b_t=-\frac{1}{2}(t+\mu_t)w_t \tag{5}$$

因此,可以得出最小能量为:

$$e_t^* = \frac{4(\hat{\sigma}2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}2 + 2\lambda} \tag{6}$$

由此可得,能量越低,神经元与周围神经元的区别越大,对视觉处理也越重要。每个神经元的重要性可由最小能量的倒数表示。能量函数最终被细化为:

$$\tilde{X} = sigmoid(\frac{1}{E}) \otimes X \tag{7}$$

式中:X 为输入的特征张量;$\otimes$为 Hadamard 乘积;E 将所有通道和空间维度中的能量函数 e_t^* 进行分组。加入 *sigmoid* 函数则是为了限制 E 过大,并不会影响每个神经元的相对重要性。

2)PSA 模块

PSA 模块[19]的结构如图 4 所示,首先会对输入特征的通道进行分组,提取不同通道的特征信息后再拼接在一起,得到各通道的多尺度特征图;其次,利用 SEWeight 模块处理不同尺度的特征图,获取通道方向的注意力向量,再经过 Softmax 进行归一化处理,实现特征的重新校准,得到新的通道注意力权重;最后,将权重与相应的特征图进行元素乘积操作,从而得到一个多尺度特征注意力加权之后的特征图。

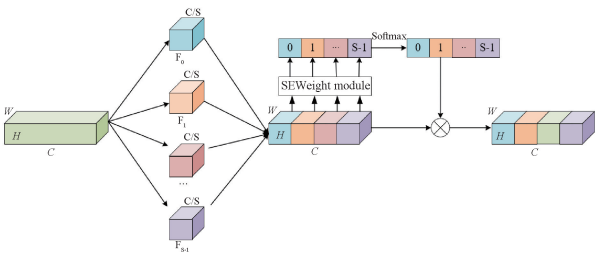


图 4　PSA 模块结构

Fig. 4　Structure of PSA module

1.2　SimSPPF 模块

快速简化空间金字塔池化(simplified spatial pyramid pooling-fast,SimSPPF)模块[20]如图 5 所示,将 1 个卷积层与 3 个相同的最大池化层依次相连,并且将计算结果拼接在一起,避免了对图像区域裁剪、缩放操作导致的图像失真问题,提高了检测精度,并且解决了卷积神经网络对图特征重复提取的问题,有效减少了冗余计算量。SimSPPF 模块被添加在 Backbone 的最后,具备进一步融合特征的优势。

1.3　多级特征金字塔网络(MFPN)

不同尺度特征的融合可以有效提高网络的目标检测性能。在 RetinaNet 中,FPN 通过对不同层级的特征信息进行融合,能够得到具有更加丰富语义信息的深层特征,增强网络对不同尺度特征的表达能力。然而,FPN 仅采用了自下而上的单向特征融合路径,并不能充分利用每一层的特征。

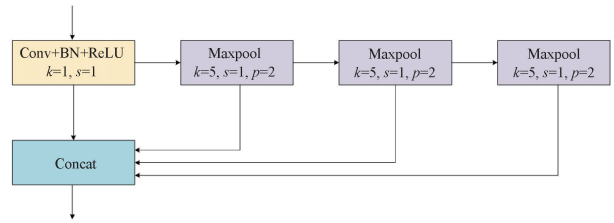


图 5　SimSPPF 模块结构

Fig. 5　Structure of SimSPPF module

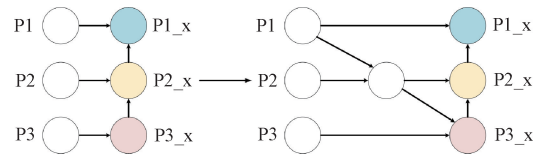


图 6　多级特征金字塔网络结构

Fig. 6　Multi-level feature pyramid network structure

对此,针对 FPN 结构进行了改进,添加了 2 个自上而下的连接路径,加强不同尺度特征之间的融合。具体地,如图 6 所示,在 P2 特征层的后面添加了一个中间特征层,首先融合 P1 特征层,然后作为输入与 P3 特征层融合。由于中间特征层处于特征融合的中间位置,包含了较为丰富的语义信息,同时不会引入过于庞大的计算量,因此改进后的多级特征金字塔网络可以增强模型的特征融合能力,提升网络检测性能。完整的改进后的 RetinaNet 模型结构如图 7 所示。

1.4　Head 的改进

1)输出层的改进

有效的特征输出是提高检测性能的关键。在 RetinaNet 中,原特征融合模块有 5 个输出特征层,随着特征的提取,输出特征层的尺度随之缩小。由于齿痕和裂纹多为中小型目标,主要由前面的特征层进行检测。考虑到后续多级特征金字塔网络和 ASFF 结构的改进,改进后的模型去除了 RetinaNet 特征融合网络的最后 2 个输出层,提高了模型的回归效率。如表 1 所示,对不同个数的输出特征层在统一环境下进行训练,结果显示,当去除后面的输出特征层时训练结果相差不大,但是节省了与 MFPN 和 ASFF 结构相连的冗余计算量。

表 1　不同个数输出特征层的训练结果

Table 1　Training results of output feature layers with different numbers

输出层个数	*mAP*/%	*Parameters*/M	*FLOPs*/G
3	91.69	31.415	83.966
4	91.57	36.134	84.96
5	91.60	36.724	85.105

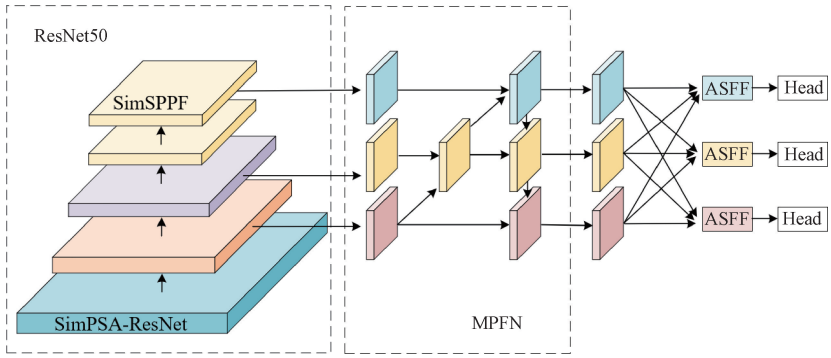


图 7 改进后的 RetinaNet 框架

Fig. 7 Framework of improved RetinaNet

2) ASFF 结构

在采用 FPN 特征融合结构时存在一个缺陷,即当某个目标在某一输出层中作为正类时,此目标可能在其他层会被当作背景甚至负类,导致不同层的特征之间不一致且存在差异。改进后的模型在去除冗余输出特征层后加入了 ASFF 结构,使模型自适应地融合多尺度特征。

ASFF 结构[21]提出了一种类似于空间注意力机制的算法来学习每个尺度的特征融合权重,使得检测头可以同时接受不同尺度的输出特征层,并自适应地学习每个尺度上特征映射的融合空间权重,更好地保留重要的特征信息,提高信息利用率。

如图 8 所示,ASFF 结构首先接收来自 3 个不同尺度的特征层,并调整到相同尺寸和维度,然后将调整后的 3 个特征层通过 1×1 的卷积得到权重参数,再将特征层分别乘以各自的权重参数并相加得到加权融合后的新特征。

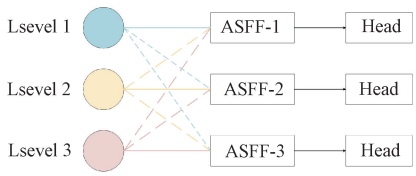


图 8 ASFF 结构示意图

Fig. 8 Schematic diagram of the ASFF structure

此结构的具体算法如式(8)所示。

$$Y_{ij}^{\ l} = \alpha_{ij}^{\ l} \times x_{ij}^{\ 1\to l} + \beta_{ij}^{\ l} \times x_{ij}^{\ 2\to l} + \gamma_{ij}^{\ l} \times x_{ij}^{\ 3\to l} \tag{8}$$

式中: $\alpha_{ij}^{\ l}$,$\beta_{ij}^{\ l}$,$\gamma_{ij}^{\ l}$ 表示不同特征层的权重参数,满足 $\alpha_{ij}^{\ l} + \beta_{ij}^{\ l} + \gamma_{ij}^{\ l} = 1$ 和 $\{\alpha_{ij}^{\ l},\beta_{ij}^{\ l},\gamma_{ij}^{\ l}\} \in [0,1]$;$x_{ij}^{\ n\to l}$ 表示调整后的 n 层到 l 层特征图的特征向量。

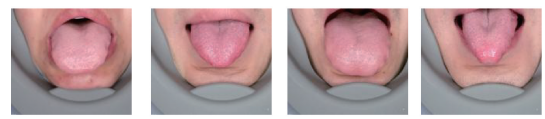
2 实验结果与分析

2.1 数据集

为了保证分类的客观性和有效性,如图 9(a)所示,本课题组使用统一的仪器采集患者的舌面图像。在采集过程中,保持拍摄环境标准且稳定,严格保证舌头图像的质量,并且在每次使用后对采集设备进行消毒通风,确保采集环境的卫生和安全。该数据集包含了 600 名以上患者的舌头图片,年龄分布在 18~70 岁之间,图片像素大小为 3 264×2 448,拍摄的部分舌图像如图 9(b)所示。

(a) 数据采集仪器

(a) Data acquisition instruments



(b) 舌图像示例

(b) Examples of tongue images

图 9 数据集来源

Fig. 9 Source of dataset

这些图片经过旋转、镜像、增加噪声等增强处理,最终形成了含有 1 290 张图片的数据集,其中包含了 1 091 个齿痕病灶和 1 248 个裂纹病灶。同时使用 Make Sense 在线标注网站对采集到的舌头数据集进行标注,并如表 2 所示按照 7∶2∶1 的比例分为了训练集、验证集和测试集。

表 2 数据集的图像分布情况

Table 2 Distribution of images in the dataset

数据集	总数据集	训练集	验证集	测试集
数量	1 290	928	233	129

2.2　实验环境

此模型的训练是基于 Ubuntu 20.04 操作系统和 Pytorch 深度学习框架完成的,具体的显卡、CPU 等实验环境如表 3 所示。迭代次数设置为 200 个 epoch,其中前 50 个 epoch 为冻结训练;使用的优化器是随机梯度下降法(stochastic gradient descent,SGD),冻结训练和非冻结训练的 batch-size 均设置为 16。为了避免学习率过大错过最优解,设置了学习率自动衰减机制。

表 3　实验环境

Table 3　Experimental environment

条目	版本
显卡	Nvidia RTX4090
CPU	Intel Xeon Silver 4316
系统	Ubuntu 20.04 操作系统
CUDA	11.6
Python	3.7
框架	Pytorch

2.3　评价指标

本文通过均值平均精度(mAP)、精确率(Precision)和召回率(Recall)作为评价指标来评估舌齿痕和裂纹检测模型的有效性。AP 是平均精度,在数值上等于 P-R 曲线下积分后得到的面积。mAP 表示平均精度的平均值,即各类的均值平均精度,用以衡量训练模型在各个目标中的检测能力。mAP 计算方法如式(9)所示。

$$mAP = \frac{1}{Q_R}\sum_{q=Q_R} AP(q) \tag{9}$$

其中,Q_R 表示类别的个数。

Precision 和 Recall 分别表示模型判断正确类别的准确率和分辨所有类别中正类的能力。二者的关系一般是负相关的,即 Precision 提高时,Recall 往往会降低。二者的计算方式如式(10)和(11)所示。

$$P = \frac{TP}{TP + FP} \tag{10}$$

$$R = \frac{TP}{TP + FN} \tag{11}$$

式中:TP(true positives)代表真实特征中成功检测的个数;FP(false positives)代表真实特征中错误检测的个数;FN(false negatives)代表漏检个数。

2.4　实验结果

实验分为 2 个部分,第 1 部分是将初始模型和改进后的模型在同一个数据集上进行训练和对比,分析改进前后的客观指标;第 2 部分,为了进一步验证改进后模型所具有的优势,使用其他主流的目标检测网络,在统一环境下进行训练,与改进后的 RetinaNet 模型进行对比。

表 4　消融实验的结果对比

Table 4　Comparison of ablation experimental results

Group	SimPSA-ResNet	SimSPPF	Output	MFPN	ASFF	*mAP*/%	*Precision*/%	*Recal*/%l
第 1 组						91.60	89.62	88.23
第 2 组	√					92.00(↑0.40)	90.45	87.36
第 3 组		√				92.46(↑0.86)	88.81	87.10
第 4 组			√			91.69(↑0.09)	88.42	87.93
第 5 组			√	√		92.07(↑0.47)	91.12	84.64
第 6 组					√	91.29(↓0.31)	88.76	88.24
第 7 组			√	√	√	92.43(↑0.83)	90.32	85.91
第 8 组	√	√				92.64(↑1.04)	90.33	85.97
第 9 组	√		√	√		92.73(↑1.13)	87.26	88.85
第 10 组	√	√	√	√		93.84(↑2.24)	90.39	90.19
第 11 组	√	√	√	√	√	94.37(↑2.77)	90.38	89.77

第 1 部分,首先使用包括了 129 张图片的测试集对改进前后的模型进行性能测试,图 10 为模型改进前后的 mAP 曲线对比,从图中可以看出,改进后的 RetinaNet 的 mAP 较初始模型有所提升,表示在舌齿痕和裂纹检测数据集中改进后的模型性能更加有效。

为了验证各个改进方法对于 RetinaNet 算法的影响,对所添加的优化措施进行消融实验,使用 mAP、Precision 和 Recall 指标来衡量模型的优劣。优化前后 RetinaNet 模型的对比结果如表 4 所示,指标均取分类平均值。其中初始 RetinaNet 模型的 mAP 为 91.6%,而优化后的为 94.37%,综合提高了 2.77%。具体而言,如表 4 中的第 2~6 组所示,SimPSA-ResNet 模块的改进使得 mAP 提高了 0.4%,SimSPPF 模块的添加则使得 mAP 提高了 0.86%,冗余输出特征层的去除使得 mAP 提高了 0.09%。由于多级特征金字塔融合结构是在去除冗余输出特征层的基础上改进的,所以对二者的结合进行了实验,mAP 提高了 0.47%。ASFF 结构的加入使得 mAP 下降了 0.31%,但是与其他优化措施结合后则可以提高 mAP。另外,改进后 RetinaNet 算法的 Precision 和 Recall 指标相较初始模型也有所提升。总体而言,优化措施有

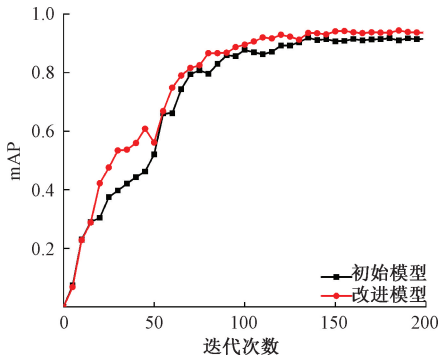


图 10 改进前后的 mAP 曲线

Fig. 10 Curve of mAP before and after improvement

效地提高了 RetinaNet 算法的性能。

模型的消融实验结果明确地表明了每一次单独优化机制和添加不同机制的效果。相较于单独的优化机制，不同的机制加在一起的效果也会不同。如表 4 中的第 7~11 组所示，改进后的 SimPSA-ResNet 模块结合 SimSPPF 模块使得 mAP 提升了 1.04%，而结合冗余输出特征层的去除则可以提升 1.13%。随着 SimSPPF 的添加和 ASFF 结构的加入，模型训练的效果随之逐步提升。另外，根据实验结果可以得出，模型的优化效果并不能随着优化机制的增加而线性叠加。

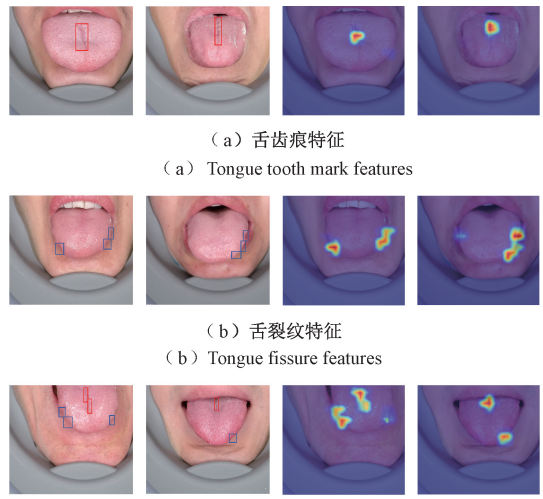
第 2 部分，为了进一步验证改进后的模型所具有的优势，如表 5 所示，分别使用其他主流目标检测模型 Faster RCNN[22]、SSD[23]、YOLOv3[24]、YOLOv5、YOLO-X[25]、YOLOv7[26]、CenterNet[27] 和 EfficientDet[28] 在同一个数据集中进行训练，并且和优化前后的 RetinaNet 模型进行各项指标的对比。可以看出，本文优化后的模型在舌齿痕和裂纹两种特征的平均检测精度都要优于其他几种算法，mAP 比以上 8 种算法分别高了 22.45%、22%、39.09%、19.62%、5.89%、9.4%、4.35% 和 1.89%。训练数据表明，由于数据集中裂纹的病灶数量略高于齿痕，因此对于裂纹特征的检测精度也要略高于齿痕特征，综上所述，改进后的 RetinaNet 更加适合齿痕和裂纹异常舌特征的检测。

最后，使用改进后的模型对测试集中的数据进行预测。对于舌齿痕和裂纹的识别效果如图 11 所示，其中，分别对只有舌齿痕特征（图 11(a)）、只有舌裂纹特征（图 11(b)）以及二者都有（图 11(c)）的舌数据进行检测。两种舌特征均可以被准确无误地框选出来，表明了改进后模型在舌特征识别中的可靠性。由此可得，本文的模型可以准确识别出舌头上存在的齿痕和裂纹，检测不同大小的特征，表现出了良好的泛化能力。另外，为了更好地对模型进行评估和解释，采用热力图可视化（Grad-CAM）[29] 技术对改进后的模型进行了处理，如图 11 所示，Grad-CAM 将模型关注的地方进行了突出显示，其关注区域与相应特征的实际位置高度吻合，为模型检测结果提供了更加直观有力的解释，也进一步表明改进后的模型可以对齿痕和裂纹部分进行精确检测。

表 5 与其他算法进行的对比

Table 5 Comparison with other algorithms (%)

模型	类别	*AP*	*Precision*	*Recall*	*mAP*
Faster RCNN	齿痕	61.37	81.52	45.83	71.92
	裂纹	82.95	83.64	57.90	
SSD	齿痕	68.75	86.24	30.91	72.37
	裂纹	75.93	86.85	64.26	
YOLOv3	齿痕	47.69	75.42	14.84	55.28
	裂纹	62.73	89.14	33.46	
YOLOv5	齿痕	71.75	77.78	57.73	74.75
	裂纹	77.75	82.90	65.31	
YOLO-X	齿痕	85.65	88.00	79.38	88.48
	裂纹	91.31	90.73	91.84	
YOLOv7	齿痕	81.24	78.11	80.93	84.97
	裂纹	88.69	85.84	81.63	
CenterNet	齿痕	88.39	93.25	78.35	90.02
	裂纹	91.66	94.62	86.12	
EfficientDet	齿痕	91.52	86.49	82.47	92.48
	裂纹	93.44	87.64	92.65	
RetinaNet	齿痕	88.51	89.44	82.99	91.6
	裂纹	94.68	89.80	93.47	
本文	齿痕	92.55	90.06	81.02	94.37
	裂纹	96.18	90.70	95.51	



（a）舌齿痕特征

（a） Tongue tooth mark features

（b）舌裂纹特征

（b） Tongue fissure features

（c）二者都有

（c） Both of them

图 11 改进后模型的预测结果及热力图可视化

Fig. 11 Prediction results and heat map visualization of the improved model

3 结 论

基于改进 RetinaNet 的舌面齿痕和裂纹检测模型为

日常自检、健康管理和辅助医生诊断提供了更有效的途径。通过数据增强处理扩充了数据集,并且对初始的 RetinaNet 算法进行了优化。首先,提出了改进后的 SimPSA-ResNet 模块,增强了网络的特征提取能力;其次,在骨干网络中添加了 SimSPPF 模块,进一步提高检测精度;然后优化了多级特征金字塔融合结构,更好地捕捉图像数据的语义信息;另外去除了冗余输出特征层,结合 ASFF 结构,使模型自适应地融合多尺度特征。实验结果表明,优化的模型 mAP 达到 94.37%,比初始 RetinaNet 的 mAP 高 2.77%,并且综合优于其他算法,可以较为准确的检测舌齿痕和裂纹的位置。

然而,验证结果表明该研究仍然存在需要改进的地方。目前,只能检测齿痕和裂纹两种类别,检测异常舌特征的种类较少,泛化能力有待提高,因此下一步将继续研究其他异常舌形态,拓宽模型的检测范围,以实现更加可靠有效的日常自检、健康管理及辅助医生诊断。

参考文献

[1] ZHANG H, MA T. Acne detection by ensemble neural networks[J]. Sensors, 2022, 22(18): 6828.

[2] 尹宏鹏,陈波,柴毅, 等. 基于视觉的目标检测与跟踪综述[J]. 自动化学报,2016,42(10):1466-1489.
YIN H P, CHEN B, CHAI Y, et al. Vision-based object detection and tracking: A review [J]. Journal of Automation, 2016, 42(10): 1466-1489.

[3] 杜宇宁,刘其文. 深度学习框架在计算机视觉领域的应用[J]. 中国安防,2022, 194(5):34-40.
DU Y N, LIU Q W. Deep learning framework in computer vision[J]. China Security, 2022, 194(5): 34-40.

[4] 黄诗茜, 王峰, 王晓洒, 等. 卷积神经网络在中医舌诊中的应用综述[J]. 电脑知识与技术, 2020, 16(26): 20-22.
HUANG SH Q, WANG F, WANG X S, et al. Review on the application of convolutional neural networks in Chinese medicine tongue diagnosis [J]. Computer Knowledge and Technology, 2020, 16(26): 20-22.

[5] 王玲, 林依凡, 李璐. 智能诊疗在舌象研究中的应用进展[J]. 中华中医药杂志, 2021, 36(1): 342-346.
WANG L, LIN Y F, LI L. Application progress of intelligent diagnosis and treatment in tongue manifestation research[J]. Chinese Journal of Traditional Chinese Medicine, 2021, 36(1): 342-346.

[6] WANG H, ZHANG X, CAI Y. Research on teeth marks recognition in tongue image[C]. Diagnostic Committee of the Chinese Society of Integrative Medicine. Proceedings of the Eighth National Diagnostic Conference of the Chinese Society of Integrative Medicine. 2014: 291-295.

[7] 芮迎迎, 孔祥勇, 刘亚楠,等. 基于 Mask Scoring R-CNN 的齿痕舌象识别[J]. 中国医学物理学杂志, 2021, 38(4): 523-528.
RUI Y Y, KONG X Y, LIU Y N, et al. Tooth-marked tongue recognition using Mask Scoring R-CNN [J]. Chinese Journal of Medical Physics, 2021, 38(4): 523-528.

[8] 赵颖,李玉双,武小荣. 基于舌图像多特征融合与机器学习的裂纹舌识别算法[J]. 燕山大学学报,2022, 46(6):522-528.
ZHAO Y, LI Y SH, WU X R. Crack tongue recognition algorithm based on tongue image multi-features fusion and machine learning[J]. Journal of Yanshan University, 2022, 46(6): 522-528.

[9] 张璐瑶, 汪莉, 包璇,等. 基于局部灰度阈值的舌象裂纹检测方法[J]. 电脑知识与技术, 2017, 13(29): 163-165.
ZHANG L Y, WANG L, BAO X, et al. A local gray-scale threshold-based lingual crack detection method[J]. Computer Knowledge and Technology, 2017, 13(29): 163-165.

[10] 钟少丹, 韦玉科, 谢铮桂. 齿痕舌像自动分割的方法[J]. 计算机技术与发展, 2009, 19(1): 245-247.
ZHONG SH D, WEI Y K, XIE ZH G. Method of automatic tongue area extraction in tooth-marked tongue images[J]. Computer Technology and Development, 2009, 19(1): 245-247.

[11] SHAO Q, LI X Q, FU ZH CH. Recognition of teeth-marked tongue based on gradient of concave region[C]. 2014 International Conference on Audio, Language and Image Processing. IEEE, 2014: 968-972.

[12] LI X Q, ZHANG Y, CUI Q, et al. Tooth-marked tongue recognition using multiple instance learning and CNN features[J]. IEEE Transactions on Cybernetics, 2018, 49(2): 380-387.

[13] HU J W, YAN ZH ZH, JIANG J H. Classification of fissured tongue images using deep neural networks[J]. Technology and Health Care, 2022,30(S1):271-283.

[14] 王一丁, 孙常浩, 崔家礼, 等. 基于深度学习的舌裂分割算法研究[J]. 世界科学技术-中医药现代化, 2021, 23(9): 3065-3073.
WANG Y D, SUN CH H, CUI J L, et al. Research on segmentation of tongue cleft based on deep learning[J]. World Science and Technology-Modernization of Chinese Medicine, 2021, 23(9): 3065-3073.

[15] WENG H, LI L, LEI H, et al. A weakly supervised tooth-mark and crack detection method in tongue image[J]. Concurrency and Computation: Practice and Experience,

2021, 33(16): e6262.
[16] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]. Proceedings of the IEEE International Conference on Computer Vision. 2017: 2980-2988.
[17] HE K M, ZHANG X, REN S Q, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.
[18] YANG L, ZHANG R Y, LI L, et al. Simam: A simple, parameter-free attention module for convolutional neural networks[C]. International Conference on Machine Learning. PMLR, 2021: 11863-11874.
[19] ZHANG H, ZU K, LU J, et al. EPSANet: An efficient pyramid squeeze attention block on convolutional neural network[C]. Proceedings of the Asian Conference on Computer Vision. 2022: 1161-1177.
[20] LI C, LI L, JIANG H, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. ArXiv preprint arXiv:2209.02976, 2022.
[21] LIU S T, HUANG D, WANG Y H. Learning spatial fusion for single-shot object detection[J]. ArXiv preprint arXiv: 1911.09516, 2019.
[22] 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. 2014: 580-587.
[23] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]. Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.
[24] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 779-788.
[25] GE Z, LIU S, WANG F, et al. YOLOX: Exceeding YOLO series in 2021[J]. ArXiv preprint arXiv:2107.08430, 2021.
[26] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 7464-7475.
[27] DUAN K, BAI S, XIE L, et al. Centernet: Keypoint triplets for object detection[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 6569-6578.
[28] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 10781-10790.
[29] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization[C]. Proceedings of the IEEE International Conference on Computer Vision. ICCV, 2017: 618-626.

作者简介

曹溪源,2018 年于华东师范大学获得博士学位,现为中北大学仪器与电子学院副教授,主要研究方向为微纳集成器件、神经网络等。
E-mail: caoxiyuan@nuc.edu.cn

Cao Xiyuan received her Ph. D. from East China Normal University in 2018. She is now an associate professor at the School of Instrumentation and Electronics, North University of China. Her main research interests include micro and nano integrated devices, neural networks.

张志东(通信作者),2014 年于西南交通大学获得博士学位,现为中北大学仪器与电子学院教授,主要研究方向为微纳测试技术、智慧医疗等。
E-mail: zdzhang@nuc.edu.cn

Zhang Zhidong (Corresponding author) received his Ph. D. from Southwest Jiaotong University in 2014. He is now a professor at the School of Instrumentation and Electronics, North University of China. His research interests include micro and nano testing technology and smart medical treatment.