

改进 Res2Net 和注意力的中药饮片识别模型^{*}

谷 瑞^{1,2} 宋翠玲¹ 李元昊³

(1. 南京大学 南京 210003; 2. 苏州工业园区服务外包职业学院 苏州 215123;
3. 上海中医药大学 上海 200003)

摘要: 中药饮片是指药材经炮制后可直接用于中医临床或制剂生产的药品,针对中药饮片种类繁多、形状各异、鉴别困难的问题,提出一种改进 Res2Net 和注意力的中药饮片识别模型 BIM-Res2Net50-IECA。首先,在 Res2Net 的基础上引入双向融合策略,促进不同尺度特征之间的有效交互,获取更精细和丰富的特征信息;其次,使用最大池化改进 ECA 注意力机制,同时增强全局视角和显著性特征,突出中药饮片重要的特征区域;最后,结合 Softmax Loss 和 Center Loss 构造联合损失函数,有效地调节类内以及类间距离,提高分类的准确性。实验表明,基线模型能有效提取多尺度特征,BIM-Res2Net50-IECA 在构建的 16 类中药饮片数据集上的准确率、精确率、召回率和 F1-Score 分别为 94.74%、94.27%、94.83% 和 94.55%,与先进的 Transformer 分类模型相比,具有更低的计算复杂度和更高的准确率,能为中药饮片的智能识别提供有力支持。

关键词: 中药饮片; BIM-Res2Net50-IECA; 多尺度特征; 注意力机制

中图分类号: TP317.4 文献标识码:A 国家标准学科分类代码: 520

Improved Res2Net and attention for Chinese herbal piece recognition model

Gu Rui^{1,2} Song Cuiling¹ Li Yuanhao³

(1. Nanjing University, Nanjing 210003, China; 2. Suzhou Industrial Park Institute of Services Outsourcing, Suzhou 215123, China; 3. Shanghai University of Traditional Chinese Medicine, Shanghai 200003, China)

Abstract: Chinese herbal medicine decoction pieces refer to drugs that can be directly used in clinical practice or formulation production after processing. In response to the problems of the wide variety, different shapes, and difficult identification of Chinese herbal medicine decoction pieces, this article proposes an improved Res2Net and attention mechanism Chinese herbal medicine decoction piece recognition model BIM-Res2Net50—IECA. Firstly, a bidirectional fusion strategy is introduced on the basis of Res2Net to promote effective interaction between features of different scales and obtain more refined and rich feature information. Secondly, using max pooling to improve the ECA attention mechanism, while enhancing the global perspective and salient features, highlighting the important feature regions of traditional Chinese medicine decoction pieces. Finally, a joint loss function is constructed by combining Softmax Loss and Center Loss, which effectively adjusts the intra — and inter class distances and improves the accuracy of classification. Experiments have shown that the baseline model proposed in this paper can effectively extract multi-scale features. BIM-Res2Net50—IECA achieved accuracy, precision, recall, and F1-Score of 94.74%, 94.27%, 94.83%, and 94.55%, respectively. Compared with the advanced Transformer classification model, it has lower computational complexity and higher accuracy, providing strong support for intelligent recognition of traditional Chinese medicine decoction pieces.

Keywords: Chinese herbal medicine decoction pieces; BIM-Res2Net50—IECA; multi-scale features; attention mechanism

收稿日期:2024-06-17

* 基金项目:2023 年江苏省高职院校教师专业带头人高端研修项目(2023TDFX010)资助

0 引言

中药饮片是指药材经过炮制后可直接用于中医临床或制剂生产的药品,其上可承接中药材,向下可承接中成药,是中医临床处方的重要组成部分,也是中医药文化的重要载体^[1]。然而,由于中药饮片种类繁多、形态相似,导致其鉴定十分困难。早期依靠专业人员通过观察、触摸、品尝、鼻闻等方式鉴定^[2],其准确性依赖于技术人员的经验积累,易受主观因素影响,不利于大规模进行检测,很难适应信息化环境下的鉴别需求。

许多学者尝试借助特征描述算子,从形状^[3]、颜色^[4]、纹理^[5]等方面提取图像特征,然后使用支持向量机、朴素贝叶斯等算法对中药饮片识别。陈仕妍等^[6]构建尺度和旋转无关的颜色匹配模板,提取中药饮片前景颜色参数。王天舒等^[7]利用图像相邻像素点间的灰度关系提取纹理特征,通过 K-临近算法输出药材产地信息。周晴等^[8]基于支持向量机构建药性判别模型,实现对唇形科中药辛味药性的高效判别。然而,传统的机器学习方法需人工设计特征,针对不同的任务需设计不同特征描述子,不具备通用性,导致中药饮片识别的准确率不高。

近几年,基于人工智能的卷积神经网络因适应性好、特征提取能力强、准确率高受到了普遍关注,越来越多的研究人员将其应用到中药饮片识别领域。张宜等^[9]对比传统算法和深度学习的差异,证明深度学习在中药饮片识别上的有效性。王健庆等^[10]使用 GoogleNet 提取中药饮片语义特征,实现复杂背景下对人参与西洋饮片的检测与识别。李玥辰等^[11]针对中药饮片识别鲁棒性不足的问题,通过图像扩充、细节增强以及 AlexNet 优化,提升模型的泛化性。韩勇等^[12]基于 DenseNet 的稠密块和过渡层的交替连接,并加入随机丢弃来防止过拟合,实现对 43 种中药材的平均识别准确率达到 93.6%。胡晓东^[13]通过对 ResNet34 进行改进,将尺寸为 7×7 的卷积分解为 3 个并联的 3×3 的卷积,提高模型对中药饮片图像识别的精度。张琦等^[14]以 Xception 以基准网络,采用迁移学习实现对 60 种中药饮片图像的快速识别。Miao 等^[15]将 ConvNeXt 与 ACMix 网络融合,提升对中药饮片特征的提取性能。Sun 等^[16]使用 LetNet 和 ResNet 分别提取中药饮片特征,然后将两个特征向量拼接用于分类。虽然上述方法提高了中药饮片的识别精度,但在实际的分类任务中,可供辨别的特征集中在形状尺寸、色泽质地、截面纹理等部位,增加这些关键特征的表达有助于改善中药饮片的识别效果。

为突出重要的特征信息,一些研究者引入注意力机制以增强对中药饮片关键特征的提取。周丽媛等^[17]利用通道注意力,引导模型提取复杂背景下的纹理特征,实现对藏药材切片图像的识别。Mo 等^[18]使用卷积块注意力从空间和通道两个维度对中药饮片分类起作用的像素区域进行关注。Wang 等^[19]提出一种高效注意力机制 (effi-

cient channel attention, ECA),采用特征重标定的形式,强调有用区域并抑制用处不大的区域。实践表明,虽然注意力可增强中药饮片关键特征的表达,但通道注意力的降维不利于学习通道之间的依赖关系,卷积块注意力过度聚焦局部区域,影响模型特征学习能力,ECA 全局平均池化造成局部细节的丢失,降低模型对关键部位的敏感性,且中药饮片形状尺寸多样,单一尺寸的卷积核无法充分提取特征,导致有效信息丢失。

基于以上研究现状和问题,本文从增强特征多尺度适应性和突出重要特征区域两方面入手,以提升中药饮片的识别精度。

1 模型架构

本文提出一种改进 Res2Net 和注意力的中药饮片识别模型 BIM-Res2Net50-IECA,如图 1 所示,该模型在 Res2Net50 的基础上进行如下改进:1)鉴于中药饮片形状尺寸多样,引入双向融合策略对 Res2Net 模块进行改进,提升不同尺度特征融合的效率和有效性,获取更精细和丰富的特征信息;2)针对 ECA 全局平均池化造成的局部细节丢失,融入最大池化同时增强全局视角和显著性特征,提升中药饮片色泽质地,截面纹理等关键特征的权重,突出有效的特征区域;3)结合 Softmax Loss 和 Focal Loss 构建联合损失函数,有效缩小同类特征之间的差异,进一步提高分类的准确率。本文方法在构建的 16 类中药饮片数据集上进行一系列的实验,并与其它先进方法进行对比,证明了具有良好的识别效果。

1.1 双向多尺度 Res2Net 模块

不同尺度的图像含有不同层次的特征信息^[20],通过提取多尺度特征,有效地融合不同感受野的特征信息,可以提升图像分类的准确性。现有方法如图像金字塔及 Inception 等都是以分组方式提取多尺度特征。Gao 等^[21]基于 ResNet 提出一种新的多尺度特征提取模块 Res2Net,所不同的是,前者(图 2(a))通过在每个残差块构造恒等映射和跳跃连接确保特征的可重用性,但每个残差块内特征提取尺度单一,而 Res2Net 模块(图 2(b))则是将输入的特征图划分为不同的特征的子集,并使用类残差结构连接不同数量和不同感受野特征组合,以完成多尺度特征融合。然而在 Res2Net 模块内部不同尺度特征的单向融合,容易造成细节信息的丢失,导致特征信息很难从底层传递到最后一层。

受双向多尺度特征融合网络的启发,本文提取一种新的双向多尺度 Res2Net (bidirectional multi-scale Res2Net, BIM-Res2Net) 模块(图 1(c)),该模块通过在 Res2Net 内部构建从左到右和从右到左的双向融合策略,以促进不同尺度特征之间的有效交互和信息传递,提升特征融合的效率和有效性,获取更精细和丰富的特征信息,增强网络的特征表达能力。

在 BiM-Res2Net 模块中,输入的特征经过 1×1 卷积

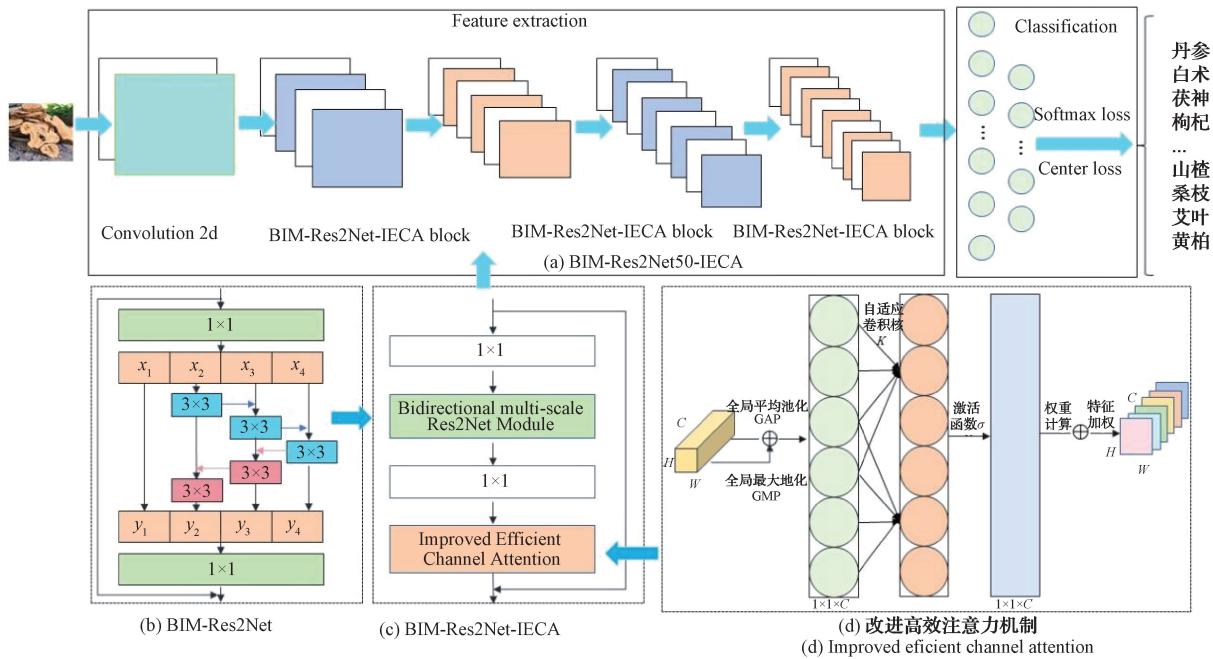


图 1 BIM-Res2Net50-IECA 网络模型架构
Fig. 1 BIM-Res2Net50-IECA network model architecture

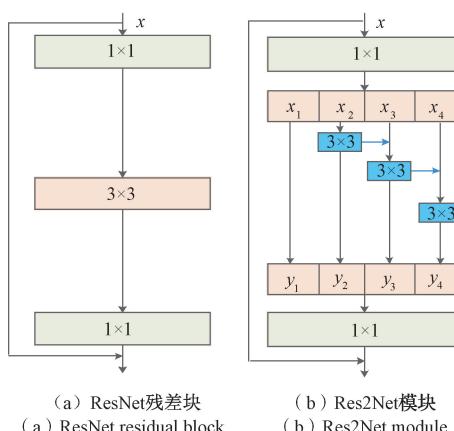


图 2 不同残差块对比
Fig. 2 Comparison residual blocks

进行维度变换,将得到的 n 个通道的特征均匀地分割为 s (本文 $s = 4$) 个特征子集,表示为 x_i ,其中 $i \in \{1, 2, \dots, s\}$,每个特征子集的通道数为 n/s 。除 x_1 直接传输给 y_1 外,每个 x_i ($i = 2, 3, \dots, s$) 都有一个对应的 3×3 小卷积层,从左到右的多尺度特征提取如式(1)所示。

$$P_i = \begin{cases} x_i, & i = 1 \\ k_i(x_i), & i = 2 \\ k_i(x_i + P_{i-1}), & 2 < i \leq s \end{cases} \quad (1)$$

其中,从 $i=2$ 开始,每组特征都会输入到对应的小卷积层,将卷积结果与下一组特征相加输入到下一个卷积层。从右到左的多尺度信息提取如下:

$$y_i = \begin{cases} x_i, & i = 1 \\ P_i, & i = s \\ k_i(P_i + y_{i+1}), & 2 < i < s \end{cases} \quad (2)$$

最后将所有特征图 y_i ($i = 1, 2, \dots, s$) 沿着通道方向进行拼接,通过 1×1 卷积实现多尺度特征融合,并与原始特征融合得到最终的特征图。由于双向多尺度特征融合,使用网络能更加有效地组合和利用不同尺度的信息,提升网络的特征表达能力,从而提高分类的精准度。

1.2 改进的高效通道注意力

ECA 作为一个高效的通道注意力模块,在提高模型特征表达能力的同时,显著减少参数量并降低计算复杂度^[22]。与 SE 类似,ECA 也是输出不同通道的特征权重,但采用了大小为 k 的一维卷积实现局部跨通道交互,避免 2 个全连接层带来的维度缩减。然而 ECA 使用全局平均池化整合特征图的空间信息,容易造成局部细节的丢失,降低了模型对特定区域的敏感性,而全局最大池化通过对特征图每个通道提取最大值,能突出最显著特征的信息,有助于模型更加精准地定位关键区域。基于此,本文通过将全局最大池化融入到 ECA 中,形成改进的高效通道注意力机制(improved efficient channel attention, IECA),提升模型对中药饮片色泽质地,截面纹理等关键信息的提取能力,其结构如图 1(d) 所示,IECA 首先使用全局平均池化和最大池化融合每个通道的局部信息和全局统计信息;然后使用一维卷积实现局部跨通道交互,并通过 Sigmoid 函数对卷积结果进行归一化处理,与原始特征图按元素相乘,最终获取加权的特征向量。

1)全局平均池化和最大池化

对输入的特征图 $U \in R^{H \times W \times C}$ 进行全局平均池化 (global average pooling, GAP) 和全局最大池化操作 (global max pooling, GMP)。GMP 对特征图每个通道进行平均值计算,能够提取特征图的整体信息,GAP 是对特征图每个通道进行最大值提取,够突出特征图中最显著的信息,然后将两者进行相加融合,同时增强全局视角和显著性特征,池化后特征图大小由 (H, W, C) 降到 $(1, 1, C)$, 即:

$$Z_{ca} = \frac{1}{H \times W} \sum_{i=1}^{i=H} \sum_{j=1}^{j=W} U_c(i, j) \quad (3)$$

$$Z_{cm} = \max(U_c(0, 0), U_c(0, 1), \dots, U_c(i, j)) \quad (4)$$

$$Z_c = \lambda Z_{ca} + (1 - \lambda) Z_{cm} \quad (5)$$

式中: Z_{ca} 表示第 c 通道的平均池化结果; Z_{cm} 表示第 c 通道的最大池化结果; Z_c 表示第 c 通道的最终输出; $U_c(i, j)$ 代表第 c 个通道 i 行、 j 列的元素值; λ 用于调节两者的比例,本文取值 0.4。

2)计算自适应卷积核尺寸

IECA 通道权重的生成是借助执行卷积核为 k 的一维卷积操作,而 k 则通过通道维度 C 的映射来自适应确定,可以表示为:

$$k = \left\lceil \frac{\log_2 C + b}{\gamma} \right\rceil_{odd} \quad (6)$$

式中: k 表示输出的卷积核尺寸; C 表示通道数; γ 用于调节通道数 C 和卷积核 k 之间的比例; $\lceil \cdot \rceil_{odd}$ 表示 k 的值只能是奇数。本文设置 $\gamma = 2$, 计算得到 $k = 5$ 。

3)权重计算

进行卷积核大小为 k 的一维卷积操作,生成每个通道的自适应权重,通过 Sigmoid 函数对权重进行归一化:

$$\omega = \sigma(Conv_k(y)) \quad (7)$$

式中: ω 表示归一化的权重; σ 表示 Sigmoid 激活函数; $Conv_k$ 表示大小为 k 的一维卷积; y 表示全局平均池化后的特征。

4)特征加权

将得到的归一化权重与对应通道特征进行逐通道相乘,获得加权后的特征图,从而增强模型对重要特征的学习和提取。

如图 1(c)所示,将 IECA 注意力机制嵌入到 Res2Net 残差中,形成 BIM-Res2Net-IECA 模块。该模块可以在多尺度特征中学习不同通道的重要性,利用跳跃连接,缓解神经网络的梯度消失。使网络更加集中地关注对当前任务更重要的特征,既能增强网络的特征学习能力,同时又保持了相对低的计算复杂度。

1.3 改进的损失函数

在神经网络训练中,提高分类效果可以从增大类别间距离和减少类内间距入手。Softmax 损失函数实现了不同类别清晰可分,却没有考虑类别内部的差异性。在中药饮片分类任务中,不同种类的中药饮片具有一定的相似

性,同一种类的饮片由于拍摄角度、光照等因素的影响呈现出较大的差别。为了获得更好的分类效果,在损失函数设计上需要考虑缩小同类别的距离的同时,扩大不同类别间的距离^[23],因此,本文引入能够区分类别间特征的 Softmax 损失函数和减小类内差异的 Center Loss,构建联合损失函数进行模型训练。

Softmax 损失函数可以增加类别间的距离,降低模型过拟合的风险,从而提高分类准确性和鲁棒性,函数定义如下:

$$L_s = - \sum_{i=1}^m \ln \frac{e^{w_{yi}^T x_i + b_{yi}}}{\sum_{j=1}^n e^{w_j^T x_i + b_j}} \quad (8)$$

式中: y_i 表示中药饮片的第 i 个类别; m 和 n 分别代表训练批次和类别数量。

Center Loss 缩小类内距离的能力较强,使得相同种类的数据表现得更加紧凑,有利于提高分类效果,函数定义为:

$$L_c = \frac{1}{2m} \sum_{i=1}^m \|x_i - c_{yi}\|_2^2 \quad (9)$$

式中: m 表示数据集中样本个数; x_i 表示输入的第 i 个表情特征; c_{yi} 表示该类表情所有样本的中心。每个样本特征距该类样本中心的距离的平方和越小,类内差距就越小。中药饮片分类模型联合损失函数可表示为:

$$L = L_s + L_c \quad (10)$$

2 实验数据与方法

2.1 数据集

1)数据集构建

由于目前还没有公开权威的中药饮片数据集,为满足实验场景的需求,本文通过实验室拍摄和网络爬取两种形式构建数据集。第 1 种方式是在室内自然光条件下,以不同底色为背景,使用 Huawei Mate Xs 2 手机采集了包含丹参、白术、山楂、桑枝等 16 种常见中药饮片样本共计 6 872 张,样本采集过程使用的焦距、曝光参数和拍摄角度略有差别。第 2 种方式使用分布式爬虫技术从网络中爬取中药饮片图像,剔除错误图像并进行筛选后每种保留 120~150 张。由于网络抓取的图像存在背景干扰、图像模糊以及图片水印等问题,因此采用两种方法相结合的方式构建数据集,不仅有利于保证同种类别数据的多样性,还有利于验证模型的识别性能。如图 3 所示,图 3(a)、(b) 为室内拍摄,图 3(c)、(d) 为从互联网上获取。

2)数据增强扩充

为进一步增强模型的泛化性和鲁棒性,避免因训练数据不足导致的过拟合,对数据进行增强扩充处理。(1)图像旋转,不改变原始样本的画质,随机以不同角度旋转图像,模拟不同的拍摄角度;(2)色彩抖动,随机调整图像的亮度、对比度和饱和度,模拟不同的光照条件;(3)高斯噪



图3 部分中药饮片图像

Fig. 3 Images of traditional Chinese medicine decoction pieces

声,添加不同程度的高斯噪声,模拟不同的图像采集设备。原始中药饮片经过数据增强操作后,样本数量扩充到29 043张,将增强后的图像按照8:2的比例划分训练集和验证集,分别用于模型的训练与验证。为加快训练速度,所有中药饮片图像统一调整为 $224 \times 224 \times 3$ 。数据增强前后不同样本分布如表1所示。

表1 数据集不同样本分布

Table 1 Different sample distributions in the dataset

种类	原始样本	增强样本	训练集	验证集
丹参	456	1 368	1 094	274
白术	578	1 734	1 387	347
白芍	689	2 067	1 654	413
醋龟甲	721	2 163	1 730	433
鹅管石	679	2 037	1 630	407
牡丹皮	568	1 704	1 363	341
党参	496	1 488	1 190	298
茯神	536	1 608	1 286	322
枸杞	499	1 497	1 198	299
桑枝	577	1 731	1 385	346
山楂	604	1 812	1 450	362
桂枝	651	1 953	1 562	391
猪苓	720	2 160	1 728	432
艾叶	669	2 007	1 606	401
金钱草	586	1 758	1 406	352
黄柏	652	1 956	1 565	391
总计	9 681	29 043	23 234	5 809

2.2 实验环境设置

本文实验所使用的软硬件平台如下:处理器采用Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40 GHz,内存型号256 GB DDR4-2400,运行平台是64位的Microsoft Windows 10,PyTorch2.0搭建和训练中药饮片识别模型,并使用英伟达高性能显卡RTX 4060 Ti(8 GB) GPU加速训练过程,实验环境参数设置如表2所示。

表2 实验环境参数配置

Table 2 Experimental environment parameters

名称	属性
操作系统	Microsoft Windows 10
CPU	Intel(R) Xeon(R) CPU E5-2680 v4
GPU	RTX 4060 Ti(8GB)
编程语言	Python 3.8
深度学习框架	PyTorch2.0
Cuda	11.4

为验证本文模型的识别精度,实验对模型超参数设置如下:训练批次设为80轮,32张图片为一个batch,初始学习率为0.001,Adam优化算法自适应地调整每个参数的学习率。

2.3 评价指标

为全面评估本文模型对中药饮片数据集的识别性能,实验选用准确率(accuracy)、精确度(precision)、召回率(recall)和F1-Score作为评估指标。准确率是正确分类的样本占所有样本数的比例,用于衡量模型避免分类错误的能力。

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (11)$$

精确度是被分类为正类的样本中,实际为正类的样本的比例,用于衡量模型预测为正类结果的可信度。

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \quad (12)$$

召回率是实际为正类的样本中,被模型正确预测为正类的样本的比例。用于衡量模型捕获正类样本的能力。

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \quad (13)$$

F1-score是精确度和召回率的调和平均值,同时考虑精确度和召回率,反映模型总体的性能指标。

$$\text{F1-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% \quad (14)$$

3 实验结果分析

3.1 基线模型有效性分析

为验证本文BIM-Res2Net50基线模型的特征提取能力,本文复现部分经典深度神经网络模型GoogLeNet^[24]、ResNet50^[25]、ResNeXt50^[26]、Res2Net50^[21],并在中药饮片

数据集上进行训练和验证测试,实验结果如表3所示。

表3 不同基线模型对比分析

Table 3 Comparison of different baseline models

方法	参数量/ ($\times 10^6$)	浮点数/ GFLOPS	准确率/%
GoogLeNet ^[24]	71.4	6.78	84.81
ResNet50 ^[25]	25.6	4.13	85.96
ResNeXt50 ^[26]	25.0	4.20	87.24
Res2Net50 ^[21]	24.3	4.20	88.31
BIM-Res2Net50	24.5	4.32	89.77

文献[24]并行使用不同尺度的卷积核提取多尺度特征,在网络末端使用 GAP 替代全连接层,然而网络参数庞大,计算复杂度高,容易导致模型过拟合;文献[25]使用瓶颈结构大幅减少参数量,通过残差连接防止梯度消失,但单一尺度的卷积核限制了模型的特征提取能力;文献[26]

是 ResNet 和 GoogLeNet 的组合体,它将输入的特征图分成若干个组,在组卷积中分组数据来改善模型的性能,相比文献[24-25],准确率分别提升 2.43% 和 1.28%;文献[21]以分层类残差方式连接不同的滤波器组,获取不同数量和不同感受野的特征组合,但在 Res2Net 模块内部不同尺度特征单向融合,容易造成细节信息的丢失;本文基线模型在文献[21]的基础上,建立双向融合策略获得更精细和丰富的特征信息。从参数量、计算复杂度和准确率维度综合考虑,本文基线模型特征提取能力更强,具有更明显的优势。

3.2 消融实验

为了评估各个模块的有效性,本文分别对 BIM-Res2Net 模块、IECA 注意力模块和 Center Loss 损失函数进行了消融实验。对不添加任何模块的 Res2Net50 网络模型标记为 Base,在验证集上将各个模块依次加入进行对比实验,实验结果如表4所示。

表4 消融实验

Table 4 Ablation experiment

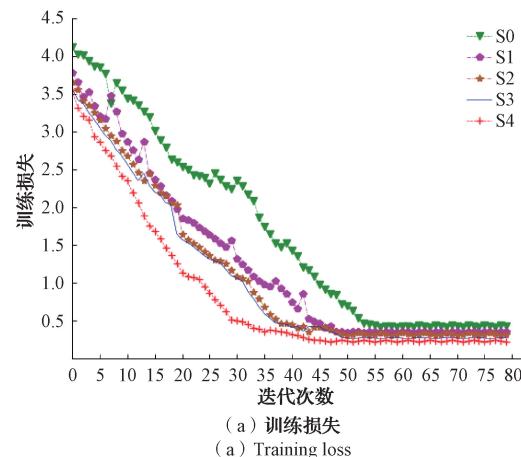
方案	Base	BIM-Res2Net	IECA	Center Loss	准确率/%	精确率/%	召回率/%	F1-Score/%
S0	✓	—	—	—	88.31	88.66	88.58	88.62
S1	—	✓	—	—	89.77	90.03	90.62	90.32
S2	—	✓	—	✓	92.82	92.67	92.79	92.73
S3	—	✓	✓	—	93.24	93.59	93.03	93.31
S4	—	✓	✓	✓	94.74	94.27	94.83	94.55

从表4可以看出,每个模块均在一定程度上提升了模型的性能。方案 S0 以 Res2Net50 为特征提取网络,通过类残差结构融合不同尺度的特征,在验证集上的准确率、精确率、召回率和 F1-Score 分别为 88.31%、88.66%、88.58% 和 88.62%。方案 S1 将特征提取网络替换成为 BIM-Res2Net50 后,由于采取双向融合策略提升多尺度特征融合效率,可以有效促进不同尺度特征之间的交互和信息传递,与 Base 相比,其准确率上升了 1.46%。方案 S2 是在 S1 的基础上添加 IECA 注意力,使模型能同时增强对全局信息和显著性特征的表达,提升有效特征区域的权重。原始图像经过随机旋转、色彩抖动和高斯噪声等数据增强处理后,相同类别的图像受到外界干扰因素的影响较大,表现出了较大的差异性,方案 S3 加入 Center Loss 损失函数后,能使同类中药饮片特征尽量向样本中心靠拢,使其相似性变大,从而提高了模型识别的精度。实验结果表明,将 3 个模块融合起来效果最佳,与 Base 相比模型在各个评价维度上都获得了较大幅度的提升,证明本文模型既能增强模型多尺度特征学习能力,又能强调有效的特征区域,并调节类内以及类间距离,提高分类的准确率,从而发挥模型的对中药饮片的最佳识别效果。

模型迭代次数与训练损失之间的关系如图4(a)所示。方案 S4 随着训练次数的增加,损失函数值逐渐减小,

未出现较大幅度的上下波动,前 20 轮损失函数的值快速下降,35~40 轮之间损失函数的值下降速度减缓,第 40 轮之后,损失函数的值趋于稳定,并最终保持在 0.2 左右,证明了预测结果与真实结果之间的差异较小,模型对中药饮片图像具有较好的拟合能力。

迭代次数与准确率之间的关系曲线如图4(b)所示,可以看到,本文模型能够以较高的准确率实现对中药饮片图像的识别。在整个训练期间,准确率保持了较高的增长速度,并在第 45 轮之后并保持识别精度在 94% 以上,证明



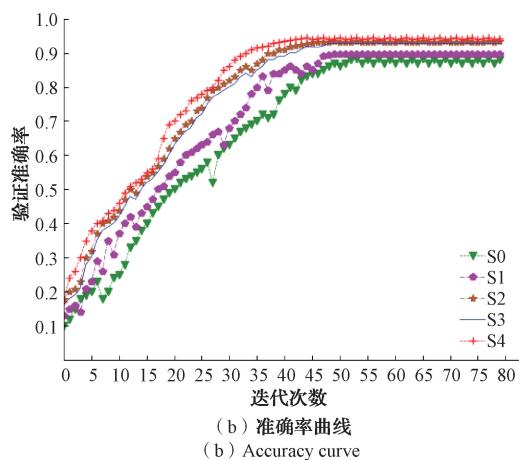


图 4 各方案训练损失与准确率曲线

Fig. 4 Train loss and accuracy curves for each scheme

本文模型具有较强的中药饮片图像识别能力。

3.3 模型有效性验证

模型验证可以帮助评估算法是否能够准确的预测未知的数据，并在实际的应用中表现出优越的性能^[27]。为验证本文模型对每种中药饮片的识别效果，将训练好的模型在验证集上进行测试，并对每种中药饮片最终的准确率、精确率、召回率和 F1-score 值进行统计，以评估本文所提模型的有效性，结果如表 5 所示。

由表 5 可知，除个别种类外，模型对大部分中药饮片的识别均取得了令人满意的效果，在 5 809 张测试集图片中，能正确识别其中的 5 504 张，占测试集样本总数的 94.74%。其中醋龟甲、鹅管石、枸杞、山楂、艾叶等特征区分明显的中药饮片样本，其识别准确率、精确率、召回率和 F1-Score 均在 97% 以上。牡丹皮、党参、金钱草、黄柏等含有较多从网络抓取的样本，虽然存在一定的背景干扰、

表 5 每种中药饮片识别效果

Table 5 Identification effect of each traditional Chinese medicine decoction piece (%)

类别	准确率	精确率	召回率	F1-Score	类别	准确率	精确率	召回率	F1-Score
丹参	88.76	89.93	90.01	89.97	枸杞	98.26	96.21	97.35	96.78
白术	91.15	91.86	91.78	91.82	桑枝	92.53	93.02	92.94	92.98
白芍	92.26	92.19	92.65	92.42	山楂	97.94	97.03	97.51	97.27
醋龟甲	98.91	97.82	98.37	98.09	桂枝	90.71	90.26	90.69	90.47
鹅管石	97.54	97.45	97.21	97.33	猪苓	91.39	91.56	91.27	91.41
牡丹皮	95.32	96.05	95.64	95.84	艾叶	98.08	98.34	98.43	98.38
党参	95.35	95.29	95.32	95.30	金钱草	96.62	96.56	97.04	96.80
茯神	89.94	89.78	90.28	90.03	黄柏	95.48	95.86	95.67	95.76

图像模糊问题，但也取得了 95% 以上的识别精度，远远超过了人类的分辨水平。模型的实际输出效果如图 5 所示，可以看出，本文算法能对不同背景、不同分辨率、不同角度

和不同干扰的中药饮片进行正确识别，并具有较高的置信度，说明了模型具有较强的泛化能力，能有效的对各种环境下的中药饮片进行正确识别。



图 5 不同类别样本测试结果

Fig. 5 Test results of different category samples

为进一步分析模型分类错误的中药饮片的分布情况，模型对各类中药饮片图像识别的混淆矩阵如图 6 所示，其中的每一列之和是验证集中每一类中药饮片的数量，主角线上的数值表示预测正确的中药饮片的数量，对角线以外的数值表示预测错误的中药饮片数量。

虽然本文方法取得了有意义的分类效果，但在实验中也暴露中所存在的一些缺陷，对形状尺寸、色泽质地，截面

纹理等特征相似的不同种类的中药饮片，容易造成分类错误(图 7)。如桂枝、丹参和茯神在颜色和纹理结构上具有一定的相似性，难以提取具有辨别性的特征，导致模型难以区分，三者的准确率都在 90% 左右；白术和猪苓具有相似的形状，但断面结构不同，白芍和桑枝形状不同，但断面结构均有光泽，因此虽然特征具有一定的区分度，但也容易引起分类混淆，四者的准确率在 91%~93%；另外少量

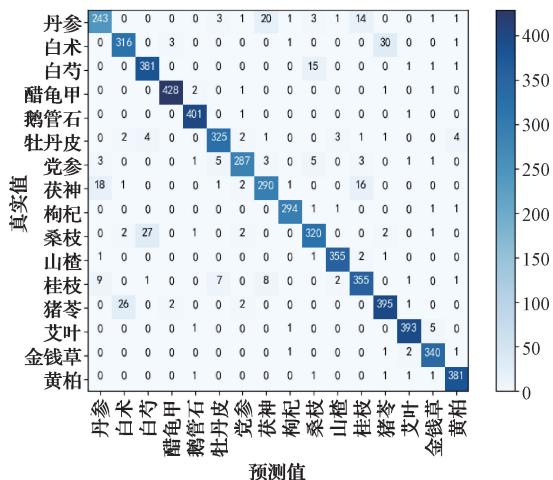


图 6 混淆矩阵

Fig. 6 Confusion matrix



图 7 容易误分的类别

Fig. 7 Categories that are prone to misclassification

表 6 不同算法对比分析

Table 6 Comparison of different algorithms

方法	参数量/ ($\times 10^6$)	浮点数/ GFLOPS	准确率/%
Swin Transformer ^[28]	88.0	15.4	91.89
Vision Transformer ^[29]	86.6	12.6	92.22
Pyramid Vision Transformer ^[30]	61.4	9.8	92.94
Data-Efficient Image Transformer ^[31]	22.01	5.97	93.58
本文	25.3	4.72	94.74

数据的高效处理,然而模型缺乏空间信息捕捉能力。文献[30]通过金字塔结构,在不同分辨率的特征图上进行高效的Transformer操作,并使用局部注意力机制关注局部

样本存在的背景干扰、数字水印等问题容易导致识别错误。

3.4 模型先进性分析

近年来,视觉Transformer将图像的每个像素区域视为序列中的一个位置,利用自注意力机制将特征映射成不同的向量空间,增强模型的感知能力,为图像分类任务带来了革命性的变化。为验证本文算法的先进性,与基于Transformer的分类模型进行对比实验,其结果如表6所示。

由表6可知,文献[28]基于滑动窗口的自注意力机制,将特征图划成多个不相交区域,在每个区域内部进行自注意力计算,并采用层次化的方法融合多尺度特征,但模型参数量较大,资源消耗高。文献[29]采用多头自注意力机制,将输入图像裁剪为固定尺寸的图像块,对其进行线性映射后加入位置编码和类别嵌入等信息,实现对图像

区域内的关键点,提高模型的表示能力。文献[31]在文献[30]的基础上引入知识蒸馏技术,减少模型中的冗余参数和过度依赖特定的隐藏单元,提升模型的泛化能力。本文模型在Res2Net的基础上,通过双向融合策略提升不同尺度特征的融合效率,并利用IECA注意力机制,使更加关注中药饮片重要的特征区域。与文献[28-31]相比参数量和计算复杂度显著降低,而准确率分别提升2.85%、2.52%、1.81%和1.16%,证明了本文算法的先进性。

4 结论

本文提出一种改进Res2Net和注意力机制的中药饮片识别模型,通过引入双向融合策略,促进不同尺度特征之间的有效交互,使用最大池化改进ECA注意力机制,增强全局视角和显著性特征,并结合联合损失函数进一步提

高分类的准确性。实验表明该模型识别精度高,能有效满足对中药饮片识别需求,具有一定的实用价值。

然而由于中药饮片种类繁多,且不同种类的中药饮片存在着形状、纹理和颜色等特征的相似性,下一步的研究方向如下:1)引入对比学习并结合图像分割技术,提取更精细的特征,减少相似中药饮片之间的混淆;2)不断完善模型,进一步评估模型对其他中药饮片的识别能力,并部署于实际的使用环境中,提升中药饮片识别的智能化水平。

参考文献

- [1] 李海洋,桂新景,侯富国.中药材和中药饮片快速鉴定技术研究进展[J].分析测试学报,2023,42(8):1026-1038.
- WANG H Y, GUI X J, HOU F G. Research progress on rapid identification technology of Chinese crude drugs and prepared slices of Chinese crude drugs[J]. Journal of Instrumental Analysis, 2023, 42(8): 1026-1038.
- [2] 徐雅静,俞捷,余远盼.人工智能在中药材及饮片鉴别领域的应用[J].中华中医药学刊,2022,40(8):47-50.
- XU Y J, YU J, YU Y F. Application of artificial in identification of Chinese medicinal materials and pieces[J]. Chinese Archives of Traditional Chinese, 2022, 40(8): 47-50.
- [3] 谭超群,解达帅,程小恩.基于大数据技术的中药饮片外观性状与内在成分数据的研究与应用[J].中国药房,2018,29(16):2287-2290.
- TAN CH Q, JIE D SH, CHENG X EN. Research and application of appearance and internal composition data of traditional Chinese medicine decoction pieces based on big data technology[J]. China Pharmacy, 2018, 29(16): 2287-2290.
- [4] 谢树莹,周明周,金海.基于OpenCV的中药饮片“辨色”识别研究[J].时珍国医国药,2018,29(2):510-512.
- XIE SH Y, ZHOU M ZH, JIN H. Research of OpenCV-based recognition system of herbal decoction pieces[J]. Lishizhen Medicine and Materia Medica Research, 2018, 29(2): 510-512.
- [5] 陶欧,张百霞,张燕玲.不同中药饮片横切面图像纹理特征参数的比较研究[J].世界科学技术—中医药现代化,2014,16(12):2550-2557.
- TAO O, ZHANG B X, ZHANG Y L. Comparative studies on parameters of texture feature of transverse section in different chinese herbal medicine [J].
- Modernization of Traditional Chinese Medicine and Materia Medica-World Science and Technology, 2014, 16(12):2550-2557.
- [6] 陈仕妍,卢文彪,王凤梅.基于颜色匹配模板的中药饮片图像识别[J].中国实验方剂学杂志,2020,26(6):158-162.
- CHEN SH Y, LU W B, WANG F M. Image recognition of Chinese herbal pieces based on color matching template [J]. Chinese Journal of Experimental Traditional Medical Formulae, 2020, 26(6): 158-162.
- [7] 王天舒,严辉,胡孔法.基于图像结构纹理信息的当归药材产地识别研究[J].中国中药杂志,2021,46(16):4096-4102.
- WANG T SH, YAN H, HU K F. Recognition of producing areas of angelicae sinensis radix based on structure+texture image decomposition [J]. China Journal of Chinese Materia Medica, 2021, 46 (16): 4096-4102.
- [8] 周晴,张传耀,王鹏.基于支持向量机的64种唇形科中药辛味药性模式识别研究[J].时珍国医国药,2023,34(9):2280-2283.
- ZHOU Q, ZHANG CH Y, WANG P. Pattern recognition for pungent of drug property of 64 kinds of labiate chinese herbs based on support vector machine[J]. Lishizhen Medicine and Materia Medica Research, 2023, 34(9): 2280-2283.
- [9] 张谊,万华,涂淑琴.基于计算机视觉的中药饮片分类技术综述与案例研究[J].计算机应用,2022,14(10):3224-3234.
- ZHANG Y, WANG H, TU SH Q. Technical review and case study on classification of Chinese herbal slices based on computer vision [J]. Journal of Computer Applications, 2022, 14(10): 3224-3234.
- [10] 王健庆,戴恺,李子柔.基于深度学习的中药饮片图像识别研究[J].时珍国医国药,2020,31(12):2930-2933.
- WANG J Q, DAI K, LI Z R. Research on image recognition of traditional chinese medicine pieces based on deep learning [J]. Lishizhen Medicine and Materia Medica Research, 2020, 31(12): 2930-2933.
- [11] 李玥辰,赵晓,王若男.基于改进卷积神经网络的中药饮片图像识别[J].科学技术与工程,2024,24(9):3596-3604.
- LI Y CH, ZHAO X, WANG R N. Image recognition of chinese herbal slices based on improved convolution neural network [J]. Science Technology and Engineering, 2024, 24(9): 3596-3604.

- [12] 韩勇, 兰杰, 郭瑞瑶. 基于深度学习的中药材饮片识别[J]. 西北农业学报, 2023, 32(11): 138-145.
HAN Y, LAN J, GUO R Y. Identification of chinese herbal medicine slices based on deep learning[J]. Acta Agriculturae Boreali-occidentalis Sinica, 2023, 32(11): 138-145.
- [13] 胡晓东. 基于改进深度学习算法的中药饮片图像识别研究[D]. 长春: 吉林农业大学, 2023: 1-5.
HU X D. Research on image recognition of traditional chinese medicine decoction pieces based on improved deep learning algorithm [D]. Changchun: Jilin Agricultural University, 2023: 1-5.
- [14] 张琦, 区锦锋, 周华英. 基于 Xception 与迁移学习的中药饮片图像识别研究[J]. 现代电子技术, 2024(3): 29-33.
ZHANG Q, QU J F, ZHOU H Y. Research on traditional Chinese medicine piece image recognition based on Xception and transfer learning[J]. Modern Electronics Technique, 2024(3): 29-33.
- [15] MIAO J F, HUANG Y A, WANG Z S. Image recognition of traditional Chinese medicine based on deep learning [J]. Frontiers in Bioengineering and Biotechnology, 2023, 23(11): 2031-2046.
- [16] SUN X, QIAN H N. Chinese herbal medicine image recognition and retrieval by neural network [J]. PlosOne, 2016, 11(6): e0156327.
- [17] 周丽媛, 赵启军, 高定国. 基于注意力引导深度纹理特征学习的复杂背景藏药材切片图像识别[J]. 世界科学技术—中医药现代化, 2022, 24(12): 4825-4832.
ZHOUL Y, ZHAO J Q, GAO D G. Recognition of tibetan medicinal slices in images with complex background via attention-guided deep texture feature learning [J]. Modernization of Traditional Chinese Medicine and Materia Medica-World Science and Technology, 2022, 24(12): 4825-4832.
- [18] MO W T, WANG J Q, WU Y. Combined channel attention and spatial attention module network for chinese herbal slices automated recognition [J]. Frontiers in Neuroscience, 2022, 13(16): 1017-1032.
- [19] WANG Q, WU B, ZHU P. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Recognition Pattern, 2020, 11534-11542.
- [20] 王肖, 张俊华, 王泽彤. 并行多尺度特征融合的肺炎 CT 分割方法[J]. 国外电子测量技术, 2023, 14(11): 15-23.
WANG X, ZHANG J H, WANG Z T. Parallel multi-scale feature fusion for pneumonia CT segmentation method [J]. Foreign Electronic Measurement Technology, 2023, 14(11): 15-23.
- [21] GAO S H, CHENG M M, ZHAO K, et al. Res2Net: A new multi-scale backbone architecture [J]. IEEE transactions on pattern Analysis and Machine Intelligence, 2019, 43(2): 652-662.
- [22] 金婷婷, 房建东, 赵于东. 注意力特征融合的番茄叶部早期病斑诊断算法[J]. 电子测量技术, 2024, 47(4): 156-164.
JIN T T, FANG J D, ZHAO Y D. Early lesion diagnosis algorithm of tomato leaf based on attention feature fusion [J]. Electronic Measurement Technology, 2024, 47(4): 156-164.
- [23] 陈万志, 唐浩博, 王天元. 融合轮廓增强和注意力机制的改进 GaitSet 步态识别方法[J]. 电子测量与仪器报, 2024, 38(1): 203-210.
CHEN W ZH, TANG H B, WANG T Y. Improved GaitSalt method for gait recognition via funsion of sillhouette enhancement and attention mechanism[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(1): 203-210.
- [24] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1-9.
- [25] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [26] XIE S, GIRSHICK R, DOLLAR P, et al. Aggregated residual transformations for deep neural networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1492-1500.
- [27] 陈大鹏, 陈庚, 刘佳. 基于纹理图像和用户动作信息的纹理触觉再现方法研究[J]. 仪器仪表学报, 2024, 45(4): 145-154.
CHEN D P, CHEN G, LIU J. Research on texture haptic display method based on texture image and user action information [J]. Chinese Journal of Scientific Instrument, 2024, 45(4): 145-154.
- [28] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10012-10022.
- [29] LIU Z, LIN Y T, CAO Y. Swin Transformer: Hierarchical vision transformer using shifted windows[C]. Proceedings of the IEEE/CVF

- International Conference on Computer Vision, 2021: 10012-10022.
- [30] WANG H, XIE E, FAN P, et al. Pyramid vision Transformer: A versatile backbone for dense prediction without convolutions [C]. International Conference on Computer Vision (ICCV). IEEE, 2021: 568-578.
- [31] YANG Z, ZENG A, LI Z, et al. From knowledge distillation to self-knowledge distillation: A unified approach with normalized loss and customized soft labels [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023: 17-32.

作者简介

谷瑞,副教授,主要研究方向为计算机视觉、数据挖掘等。

E-mail:gur@siso.edu.cn

宋翠玲(通信作者),博士,教授,主要研究方向为机器学习与数据挖掘。

E-mail:songcl@nju.edu.cn

李元昊,硕士,主要研究方向为中医诊断。

E-mail:lyh2021314@163.com