

基于 ConvNeXt 的相似中药饮片图像分类

兰邵琨¹ 朱彦陈¹ 李会丽^{1*} 郑文华¹ 张志凯¹

LAN Shaokun ZHU Yanchen LI Huili ZHENG Wenhua ZHANG Zhikai

摘要

针对药材市场上出现越来越多以假乱真、以次充好的现象,根本原因在于人工识别效率低、准确性易受主观因素影响,而精准区分外观相似的中药饮片,对保障临床用药安全与疗效至关重要。基于计算机视觉的分类技术具备智能、高效的独特优势,可为中药饮片筛选提供可靠解决方案。为此,文章研究并提出一种基于 ConvNeXt 的同类中药饮片识别算法:通过调整全连接层节点数量、冻结该层参数,结合迁移学习方法对模型进行训练优化,最终构建出新的识别模型。实验以 26 类易混淆中药饮片为研究对象,构建包含 1 124 幅图片的数据集,结果表明,该模型对中药饮片影像的分类准确率达 99% 以上,显著优于传统 CNN 模型。该研究为中药饮片从业人员提供了强有力的技术支持,能有效辅助中药饮片的鉴定工作,为规范中药市场、保障用药安全提供了新路径。

关键词

卷积神经网络;迁移学习;中药饮片图像分类;ConvNeXt 网络

doi: 10.3969/j.issn.1672-9528.2025.12.002

0 引言

中药饮片包括原形药材和经过切制、炮制后的饮片^[1]。国内中药材资源十分丰富,中药饮片中存在外观相似但功效不同的饮片,许多药品饮片的炮制方式和储存方式的不同,饮片本身的纹理、颜色会受到不同程度的影响,通过主观的人工鉴别,对中药饮片的外形、颜色、气味、味道等特征判断相似药品的类别的方式是效率和准确率都较低,且个人主观性以及经验等方面的不同会影响判断结果。相似饮片的鉴定工作或将是加快中药饮片产业化发展和传统中医学用药安全的首要任务。

如今,国内的药材市场也有许多商家为追求利益,以次充好、以假乱真^[2]。在计算机视觉不断发展的今天,卷积神经网络的图像分类能力被认为是超过人工检测的,这就提供了一种利用网络模型解决中药饮片规模识别的方法,目前,在医药分类领域应用深度学习的技术已经有一定的进步,主要是通过构建经典的 CNN 模型来实现对饮片品种的鉴别。传统的 CNN 模型的劣势也很明显,其存在混淆率高,特征提取不充分等缺点。李志成等^[3]采用改进全连接层节进行中药识别,达到了 97.68% 的高准确率;周法律^[4]则通过基于图像的叶片分类方法实现中药植物自动检测分类,达到了 90% 以上的准确率;吴冲等^[5]的研究基于多特征融合策略,整合了方向梯度直方图(HOG)的轮廓描述能力和局部二值

模式(LBP)的纹理表征优势,在中药饮片图像分类任务中取得了突破性进展。实验结果表明,该方法的分类准确率较传统单一特征提取方法提升,同时有效解决了形态相似饮片的误判问题。通过优化特征权重分配和采用自适应融合算法,系统在保持较高运算效率的前提下,将平均识别错误率控制在 10% 以下。这一成果为中医药材的智能化鉴别提供了新的技术路径,特别是在处理外观特征相近的饮片样本时展现出显著优势,平均准确率超过 91%。

与传统的 CNN 模型相比,基于 Transformer 的跨模态识别和多特征融合算法有着更高的准确率等优点,但也具有局限性,比如小样本场景下的泛化能力不足,ConvNeXt 通过优化网络结构和训练策略,在标准视觉任务中展现出了优秀的泛化能力,尤其是在数据分布与训练集相似的情况下。这是第一次在饮片分类研究中应用 ConvNeXt 模型,旨在为识别可视特征相似的饮片提供参考。该实验模型将迁移学习与 ConvNeXt 相结合,对中药饮片图像进行分类,其准确率达到 99.6%,比原模型有较大提高。

1 研究方法

1.1 卷积神经网络

卷积神经网络(CNN)是应用最广泛的深度学习模型之一,通常由卷积层、池化层(也称为下采样层)、全连接层和输出层构成^[6]。

1.1.1 卷积层

每个卷积层在卷积神经网络中包含多个用于从输入中提取特征的卷积核。卷积核的工作原理是扫描输入特征,在感

1. 江西中医药大学 江西南昌 330004

[基金项目] 纳米/分子电子器件设计及输运性质的研究(GJJ211225)

受野内对输入特征进行矩阵元素乘法,然后求和并加上偏差,最后通过激活函数生成新的特征图。

1.1.2 池化层

池化层采用池化操作可以减小特征图的尺寸。池化操作的特点是尽量减少信息丢失,不会对计算和内存开销产生显著影响。在卷积神经网络中,池化层是不可或缺的,降低了特征图的尺寸,减少了网络计算的参数量,实现了空间不变性,并扩展了后续卷积的感受野^[7]。常见的池化方法包括最大池化和平均池化^[8]如图1所示。

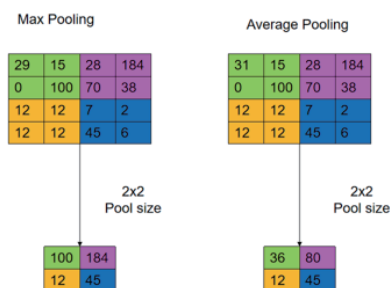


图1 池化层

1.1.3 全连接层

全连接层 (fully connected layer, FCLayer) 广泛应用于各种深度学习模型,特别是在卷积神经网络 (CNN) 和多层感知机 (MLP) 中,是神经网络中的一个基本组成部分。全连接层的主要功能是为了产生最终的输出,整合前一层的特性。在全连接层中,每一个输入节点都连接着每一个输出节点,所以叫“AllConnection”。具体而言,设输入特性向量为 $\mathbf{x}$;全连接层的权重矩阵为 $\mathbf{W}$;偏置向量为 $\mathbf{B}$;全连接层的输出为 $\mathbf{Y}$ 。则全连接层的输出 $\mathbf{Y}$ 可以表示为:

$$\mathbf{Y} = \mathbf{W}\mathbf{x} + \mathbf{B} \quad (1)$$

1.2 迁移学习

迁移学习的目的在于利用源域中获得的知识来解决目标域的任务^[9]。在样本稀缺的应用场景中,跨域迁移学习通过参数共享机制显著优化模型泛化能力。迁移学习主要有两种方式:(1)将源域训练好的模型作为特征提取器^[10]。首先剥离源模型末端的全连接分类器,保留其预训练的特征提取主干网络;随后针对目标域特性,构建可调分类器模块并重新初始化权重参数。在这种方法中,仅更新新加的全连接层的参数。(2)微调:在这种方法中,将源域训练好的模型参数迁至目标域,并对整个网络进行微调。这样的做法让模型在目标区内更好地适应任务。迁移学习可以通过这两种方式对不同领域的任务进行有效帮助。

1.3 GELU 激活函数

高斯误差线性单元 (gaussian error linear unit, GELU) 是一种广泛应用于深度学习模型的激活函数,于2016年提出。其核心思想是通过结合输入值的高斯分布特性,以更符合自然数据特性的方式实现非线性变换。其模拟了神经元在输入

受随机噪声影响时的激活状态,使其在接近零的区间呈现平滑过渡,而非传统ReLU的硬阈值特性。实验表明,GELU在图像分类、机器翻译等任务中常优于传统激活函数,成为当前深度学习中激活函数的重要选择之一,尤其适用于需要精细梯度调节和高层特征抽象的复杂模型。

2 ConvNeXt

ConvNeXt网络是一种新型的卷积神经网络 (CNN) 模型。相比于传统模型,ConvNeXt在ImageNet-1K数据集上达到了87.8%的Top-1准确率,这一成绩显著优于许多传统CNN模型,如ResNet系列和EfficientNet等^[11]。如今的先进模型如Transformer,ConvNeXt基于纯卷积结构设计,相比Transformer架构中的自注意力机制,其计算复杂度更低。结合实际,中药饮片图像包含丰富的纹理、形状等局部特征,ConvNeXt通过继承卷积神经网络 (CNN) 的局部感受野特性,能够更高效地捕捉这些局部细节,这对于饮片分类中的形态学特征提取至关重要。中药饮片图像数据集通常规模较小,ConvNeXt引入大核卷积增强模型表达能力,在小规模数据集上能取得良好表现,但在大规模数据集上的表现可能不如Transformer。为了适应这一限制,通常需要对输入图像进行裁剪或缩放,这可能会影响模型的性能。而SPP则解决了这一问题。对传统CNN中的池化层进行了改进。通过在不同的尺度上对输入进行金字塔池化SPP可以捕捉到多尺度的特征,从而使CNN的精确性得到提高。ConvNeXt Block输出公式为:

$$\mathbf{Y} = \text{ConvNeXtBlock}(\mathbf{X}) \\ = \text{Conv} \left(\text{Relu} \left(\text{BatchNorm}(\text{Con}(\mathbf{X}, \mathbf{W})) \right) \right) \quad (2)$$

式中: $\mathbf{W}$ 表示卷积核;Con表示卷积操作。

ConvNeXt及ConvNeXt Block结构如图2所示。

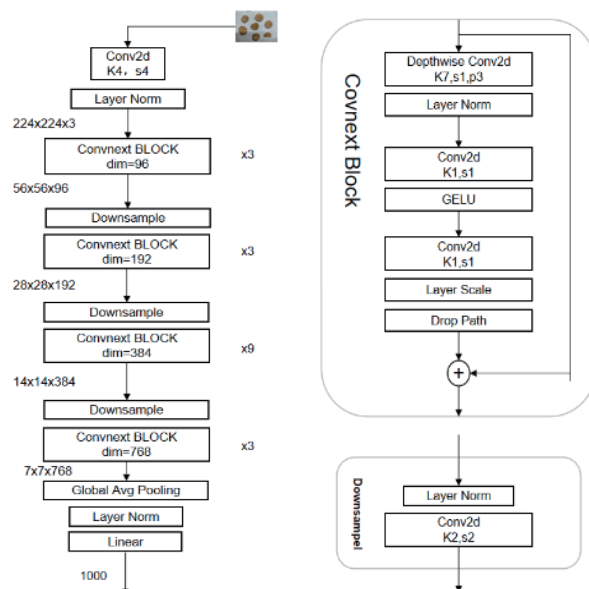


图2 ConvNeXt及ConvNeXt Block结构

3 仿真分析

目前, 中药饮片识别领域缺乏公开的标准数据集, 因此本研究决定使用自建的饮片数据集进行实验, 本文所使用数据集中数据均拍摄于江西中医药大学岐黄国医书院, 拍摄设备为华为 mate20 等手机。其中含有两种易混淆类, 类似白芍类和类似乌梅类, 其中类似白芍类有白芍、泽泻、浙贝母、人参、牛膝、桑枝、炮附片; 类似乌梅类有乌梅、神曲、熟地黄、五味子。数据集共含有 53 类饮片, 易混淆类中每种饮片包含 50 张图片, 数据集共含有 1 174 张图片。其中几种易混淆图像如图 3 所示。

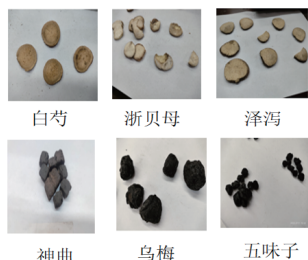


图 3 中药饮片图像

3.1 实验环境

实验操作系统环境为 Windows 10 (64 位), CPU 为 Intel(R) Xeon(R)。采用 Gold 5218 CPU 处理器, GPU 为 Tesla K80, 应用 Python3.7, 深度学习框架及版本为 PyTorch1.8.9, 默认参数设置: epoch 大小为 100, batch_size 大小为 8, 学习率为 1×10^{-3} , 使用 AdamW 优化器, 模型的输入图像尺寸由训练模型要求的图片大小决定。

3.2 数据增强及划分

由于数据总量较少, 为了增强模型的泛化能力, 采用了多种数据增强的方法对训练集进行扩充。这些方法包括增加高斯噪音, 图像翻转和转动等。

该研究采用了实验样本较少的 53 种饮片 1 124 张图像, 并采用了进一步扩充数据集的方法, 如增加高斯噪音、旋转翻转等增强方式, 使扩展后的图像总量达到 5 310 张。

为设置好更优的训练集与测试集的划分比例, 不同训练样本占比 (10%~90%) 下的准确率变化如图 4 所示。

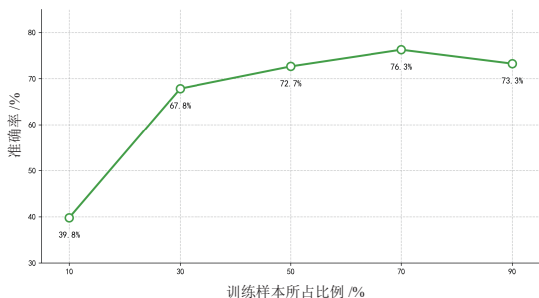


图 4 不同比例训练样本下的准确率

当仅使用 10% 的样本作为训练数据时, 准确率约为 39.8%; 当训练样本占比提升至 50% 时, 准确率上升至 67.8%; 训练样本占比为 70% 时, 准确率达到了 76.3%; 当训练样本占比增至 90% 时, 准确率约为 73.3%。实验结果表明,

当训练集占比达到 70% 时, 模型准确率趋于稳定。本研究采用 7:3 的比例划分训练集 (3 717 张中药饮片图像) 和测试集 (1 593 张中药饮片图像), 该划分方式可在样本总量较少时有效检验模型的泛化能力。

3.3 实验结果及分析

在 ImageNet 数据集上对 ConvNeXt 进行训练, 在训练结束后获取参数和权重。将获取到的参数载入模型中, 修改全连接层的节点数, 冻结全连接层的参数, 继续使用自建相似饮片数据集进行训练。原模型相似饮片训练识别效果如图 5 所示, ConvNeXt 相似饮片识别效果如图 6 所示。

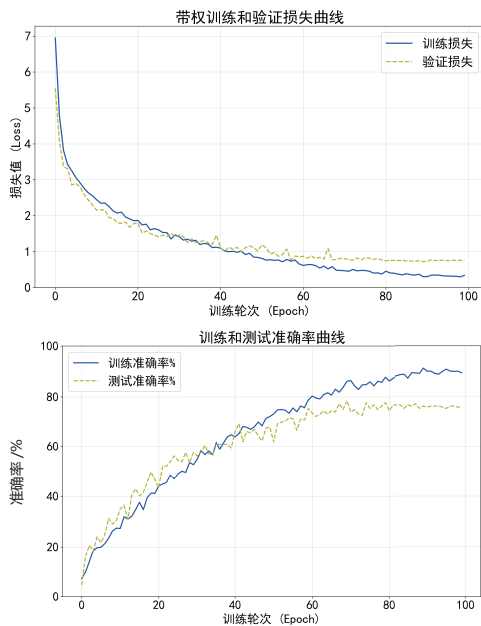


图 5 原始模型训练效果

3.4 迁移学习对实验结果的影响

在模型训练过程中, 首先将训练好的参数导入模型, 并调整全连接层的节点数量。接着利用自建的中药饮片资料集, 对除全连接层以外的全部参数进行冻结, 以备训练之用。从图中能看出, 原模型训练准确率和测试准确率不稳定, 波动幅度较大。此后, 损失值曲线几乎在第 80 轮才逐渐收敛, 训练损失持续下降。说明该模型在训练集中表现不错, 而在测试集中成绩不佳, 两者差距逐步拉大, 过拟合现象明显。

将两个实验数据进行比较, 采用 ConvNeXt 并结合迁移学习的方法, 改进后的模型能够更好地适应中药饮片数据集。实验结果如图 6 所示。相对于原模型准确率曲线波动, 改进后的模型从第 20 轮开始, 测试集与训练集的准确率曲线就趋于平稳, 且差距不大, 改进后损失值曲线相对于原模型损失值, 在第 20 轮左右就趋于平稳, 与原模型相比测试集的精确度有明显提高。模型在中药饮片资料集中的拟合效果, 通过迁移学习, 明显提高, 避免了过拟合的问题。

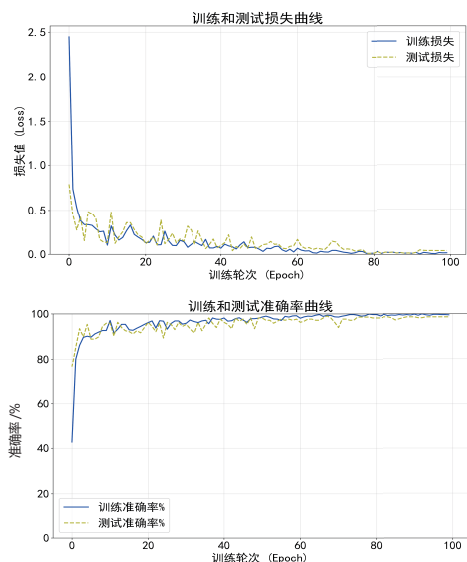


图6 ConvNeXt 结合迁移学习后的效果

4 算法对比分析

通过实验，将相同的数据集放入模型中进行对比，发现经过改进后的模型相比于原模型准确率提升了6.5%，对比其他经典的CNN模型准确率平均提升了20%，表1为比较结果。实验显示，基于迁移学习的ConvNeXt准确率对相似饮片的识别率高于其他热门模型。

表1 各模型对比结果

模型	准确率 / %
本研究算法	99.6
ResNet50	97.8
Vgg16	69.6
GoogLeNet	69.8
原 ConvNeXt	93.1

5 结论

当前，中药饮片图像的识别准确率普遍偏低，成为制约中药饮片智能化发展的关键瓶颈。为攻克这一难题，本文将迁移学习技术与前沿的ConvNeXt模型深度融合，并创新性地将此组合模型应用于中药饮片图像分类任务。

通过实验验证，在自建的中药饮片数据集上，该模型展现出了卓越的分类性能，分类准确率达99.6%，这一成果为中药饮片图像分类领域带来了新的突破。然而，我们也清醒地认识到，当前数据集存在数据分布不均衡的问题，这可能会对模型的泛化能力产生潜在影响。此外，光照变化作为实际场景中不可忽视的因素，对中药饮片图像的识别精度有着重要影响。未来，有必要深入研究光照变化对该模型的影响机制，进而检测在复杂背景下该模型对中药饮片图像的识别精度。这一研究方向的深入探索，有望为研究人员提供更为强大、可靠的工具，助力中药饮片种类实现精准识别，进而有力推动中医药现代化进程迈向新的高度。

参考文献:

- [1] 王传娟. 常见中药饮片伪品和易混淆品的鉴别[J]. 心理月刊, 2019, 14(8): 182.
- [2] 李明利, 高文雅, 王宏洁, 等. 基于“经验-成分-活性-电子感官”的中药饮片质量综合评价体系的探讨[J]. 中国中药杂志, 2022, 47(8): 1995-2007.
- [3] 李志成, 朱彦陈, 杜建强, 等. 基于迁移学习的中药饮片图像分类[J]. 计算机仿真, 2024, 41(8): 221-227.
- [4] 周法律. 基于数字图像的中药植物叶片分类方法研究[D]. 合肥: 安徽中医药大学, 2016.
- [5] 吴冲, 谭超群, 黄永亮, 等. 基于深度学习算法的川贝母、山楂及半夏饮片的智能鉴别[J]. 中国实验方剂学杂志, 2020, 26(21): 195-201.
- [6] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [7] CHEN Y P, LI J N, XIAO H X, et al. Dual Path networks[EB/OL]. (2017-08-01)[2025-11-22]. <https://doi.org/10.48550/arXiv.1707.01629>.
- [8] BOUREAU Y L, PONCE J, LECUN Y. A theoretical analysis of feature pooling in visual recognition[C]// Proceedings of the 27th international conference on machine learning (ICML-10). New York: ACM, 2010: 111-118.
- [9] 刘飞, 陈仁文, 邢凯玲, 等. 基于迁移学习与深度残差网络的滚动轴承快速故障诊断算法[J]. 振动与冲击, 2022, 41(3): 154-164.
- [10] 谭代庆, 肖志鸿, 吴浩忠, 等. 融合改进残差网络和Transformer的易混淆中药饮片识别研究[J]. 中国数字医学, 2023, 18(6): 42-50.
- [11] LIU Z, MAO H Z, WU C Y, et al. A ConvNet for the 2020s[C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Piscataway: IEEE, 2022: 11976-11986.

【作者简介】

兰邵琨(1999—), 男, 湖北黄冈人, 硕士研究生, 研究方向: 图像处理。

朱彦陈(1980—), 男, 山东临沂人, 硕士研究生, 副教授、硕士研究生导师, 研究方向: 数字图像处理、深度学习。

李会丽(1979—), 通信作者(email: 20101030@jxutcm.edu.cn), 女, 河南扶沟人, 博士, 教授、硕士研究生导师, 研究方向: 凝聚态计算物理。

郑文华(1999—), 男, 广东普宁人, 硕士研究生, 研究方向: 图像处理。

张志凯(1998—), 男, 江西抚州人, 硕士研究生, 研究方向: 图像处理。

(收稿日期: 2025-07-02 修回日期: 2025-12-12)