基于多尺度特征深度神经网络的不同产地山楂细粒度图像识别

谭超群[1]，秦中翰[2]，黄欣然[2]，陈　虎[2]，黄永亮[3]，吴纯洁[4]，游志胜[2]
(1. 成都中医药大学 智能医学学院，成都，611137；
2. 四川大学 视觉合成图形图像技术国防重点学科实验室，成都 610065；
3. 成都中医药大学附属医院，成都 610032；4. 成都中医药大学药学院，成都 611137)

摘　要：中药是中医治疗疾病的主要途径，也是我国中医药事业传承与创新发展的物质基础，其真伪优劣也会直接影响中医临床的疗效，因此研究科学合理且高效的中药材质量检测方法符合当前行业热点. 山楂作为中国著名的药食两用类药材，在烹饪和治疗中具有保护心血管、降低血压的作用，被广泛应用；但由于自然环境与栽培条件的不同，不同产地的山楂易被混淆从而对品质产生影响. 尽管化学、生物鉴定的方法广泛而重要，但专业门槛高，耗时较长；且传统图像处理方法容易受外在环境因素干扰，可靠性差. 因此亟待研究快速准确的方法以实现山楂产地的精准鉴别；受 CoAtNet 与 Swin-Transformer 网络启发，本文结合 MBConv 模块中深度可分离卷积网络对局部信息建模的特点与 Swin Transformer 模块多层次结构可弥补网络非局部性损失的特性，提出一种多尺度特征的混合神经网络模型，通过获取图像不同层级特征，将获取的形状、颜色与纹理等浅层特征作为先验知识与高层级语义信息进行特征融合，研究了一种快速有效的识别方法以实现对不同产地山楂的有效鉴别；此外，本文提出一种新的局部空间注意力机制，通过形成通道注意力模块联合空间注意力模块的新结构，实现对图像细粒度特征的聚焦与学习. 实验结果表明，本文所提出的方法有最高的鉴别准确率为 89.306%，优于其他基线模型. 实践证明，本文的研究提高中药材鉴别的科技水平，拓宽传统中医药的研究思路.

关键词：多尺度特征；神经网络；山楂；细粒度识别
中图分类号：TP389.1　　**文献标志码**：A　　DOI：10.19907/j.0490-6756.2024.013003

Fine-grained image recognition of Cratargi Fructus from different origin based on multi-scale feature deep neural network

TAN Chao-Qun[1], *QIN Zhong-Han*[2], *HUANG Xin-Ran*[2],
CHEN Hu[2], *HUANG Yong-Liang*[3], *WU Chun-Jie*[4], *YOU Zhi-Sheng*[2]
(1. College of Intelligent Medicine, Chengdu University of Traditional Chinese Mdeicine, Chengdu 611137, China;
2. State Key Laboratory of Fundamental Science on Synthetic Vision, Sichuan University, Chengdu 610065, China;
3. Hospital of Chengdu University of Traditional Chinese Medicine, chengdu 610032, China;
4. College of Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China)

收稿日期：2022-12-01
基金项目：四川省科技厅应用基础研究课题(2018JY0435)；四川省中医药管理局科学技术研究专项课题(2021MS012)；成都中医药大学"杏林学者"学科人才科研提升计划人才项目(QNXZ2019018)
作者简介：谭超群(1995－)，女，四川南充人，博士研究生，研究方向为图像处理与机器智能. E-mail：18380161804@163.com
通讯作者：黄永亮. E-mail：hyl@cdutcm.edu.cn

Abstract: Traditional Chinese medicine (TCM) is the primary approach for treating diseases and is also the he foundation for the development and innovation of TCM, the authenticity of TCM directly impacts the clinical efficacy. Therefore, scientific, reasonable, and efficient quality detection of TCM is a pressing research topic. *Cratargi Fructus* (CF) as a well-known edible food in China, which has been widely used for the ability of protecting cardiovascular and lowering blood pressure in cooking and treatment. However, it is reported that the difference in natural environment and cultivation conditions affects the CF's quality and CF from different origins are easily confused, thus, the species authentication is necessary. Although physicochemical, biological, and manual identification methods are widely used, they have a high professional threshold and are inefficient. Image processing methods are easily affected by environmental factors, which reduces their reliability. Thus, there is an urgent need to study fast and accurate methods for the identification of CF. Inspired by CoAtNet and Swin-Transformer networks, we have proposed a hybrid neural network model with multi-scale features, combining the local information of the deep separable convolution network in MBConv and the non-local loss of the multi-level structure in Swin Transformer. By acquiring different features, the superficial features including shape, color and texture as prior knowledge have fused the high-level semantic information. A fast and effective recognition method is developed to realize the effective identification of CF from different origin. Furthermore, a new global spatial attention mechanism is introduced, which can focus and learn the fine-grained features of images by forming a new structure of channel attention module and spatial attention module. Our experimental results demonstrate that our proposed method has the highest identification accuracy of 89.306%, which outperforms other baseline models. This study highlights the potential for improving the scientific and technological level of TCM identification and broadening research on TCM.

Keywords: Multi-scale features; Neural network; *Cratargi Fructus*; Fine-grained recognition

1　引　言

中药是中医治疗疾病的主要途径,其真伪优劣直接影响中医的临床疗效. 中药材的生长与分布,因其自然成长环境以及栽培加工方式的不同具有一定的地域性;所谓"药材好,药才好",中药材的产地对药物质量与其疗效有直接关联,因此对不同产地药材的质量评价对行业发展有至关重要的作用.

山楂是蔷薇科植物山里红 *Crataegus pimiatifida* Bge. var *major* N. E. Br. (NEB)或山楂 *Crataegus pimiatifida* Bge. (CPB)的干燥成熟果实[1]. 山楂作为食用和药用材料,具有广泛的用途和市场价值. 山楂在传统上具有保护心血管、抗氧化、抗癌、抗炎、降低血压、胆固醇等功效[2,3],且广泛分布于多个国家,主要在亚洲、欧洲和北美等地区. 它的保健功效和极高的营养价值在中国已经有超过 500 年的历史,同时在河南、山东、安徽等 8 个省份广泛栽培. NEB 经过种植培育多作为水果,已被广泛加工成各种食品和药品;其中河北承德、山东平邑、费县等地的山楂质量优良,需求量大,经济价值高. CPB 大部分为野生,它的果实通常为暗红色,更多用作药物[4-6]. 据报道,不同产地山楂容易被混淆,从而对使用过程中质量产生影响. 因此,不同产地山楂的物种认定在实际应用中是迫切需要的.

随着中医药行业上千年的发展,形成了一套行之有效的传统中药质量评价技术. 传统中药质量评价技术以老药工的经验判别为主,主要依据药材的外观性状,以看、摸、闻、尝等感官方法来判断药材的真伪优劣. 当前的中药质量评价主要是通过检测《中华人民共和国药典》中有机酸和类黄酮等[1]有效成分来鉴别不同产地的中药材;然而理化鉴定和生物评价[7-10]很难通对一种或多种指标成分、特征物质含量的检测来充分反映其整体内在质量,且该方法依赖于特定的设备器械,耗时较长[11,12]. 此外,人工经验鉴别的方法过于依赖人的主观感受,缺乏客观科学性[13];基于传统图像处理方法虽然在一定程度上取得了较好的效果,但通过算法捕获的浅层特征是直接来自图像像素,而不具有高层语义的特征信息,容易受环境因素影响,可靠性差[14,15]. 因此,如何对不同产地山楂实现无损、快速高效的准确识别成为重要研究课题.

随着人工智能[16,17]与深度学习技术[18]的发

展，为中药材的经验鉴别与人工智能深度学习算法的交叉融合奠定了坚实理论基础，人工智能技术通过整合老药工经验知识，解决人工鉴别擅长领域的局限性问题. CNN 以对特征高效强大的学习能力和表达能力在食品、植物、农业、医疗等多个领域[19-25]的识别中得到广泛认可获得了广泛的关注. 尽管 ConvNets 的性能有了显著的提高，但它只能专注于模型的局部关系，而且具有较强的先验性，受到归纳偏差的影响[26,27]. 此外，基于 Transformer 的方法具有强大的全局信息提取能力，但计算成本高，对数据量要求较高，泛化能力差[28,29]；因此，ConvNets 与 Transformer 相结合解决图像分类问题引起了人们的关注. 本文针对不同产地的山楂细粒度图像进行视觉鉴别，受 CoAtNet 与 Swin-Transformer 网络启发，结合 MBConv 模块中深度可分离卷积网络对局部信息建模的特点与 Swin-Transformer 模块多层次结构可弥补网络非局部性损失的特性，研究了一种快速有效的识别方法以实现对细粒度图像的中药材不同产地的鉴别.

综上所述，本文有以下贡献：(1) 提出一种多尺度深度神经网络混合模型，通过注入浅层先验特征引导高层语义信息，捕获数据局部性与全局性特征，实现细粒度图像的识别. (2) 提出一种新的空间注意力模块，通过对线性变换的特征向量进行 SoftMax 操作，对获取的注意力权重平滑近似，将注意力进行分散，突出对图像低频区域的关注，增强网络对局部细粒度特征的感知. (3) 通过实验数据验证，本文提出的方法识别准确度高于其他主流模型，能更好地聚焦图像局部区域的细节特征；该方法相较于传统基于机器学习方法的中药材产地分析，更加快速便捷且高效，具有优良的实用价值.

2　相关工作

2.1　基于图像处理与模式识别的方法

随着计算机技术的飞速发展，图像处理与模式识别技术在中药质量鉴定中得到一定研究. 该技术主要利用计算机技术与数学方法对图像信息表示，先对获取的图像进行包括分割和平滑等预处理，联合机器视觉图像算法对图像进行质量评价与识别等研究. 该方法在中药质量中主要通过提取人工设计的形状、颜色和纹理等底层图像特征，随后采用 SVM 等机器学习方法进行识别鉴定，其具体流程如图 1 所示.

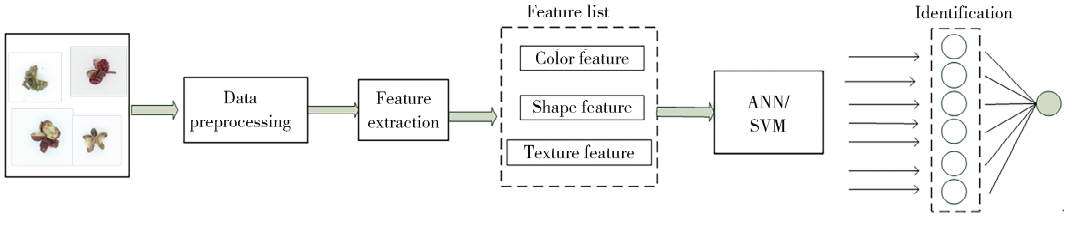

图 1　基于图像处理与模式识别的方法

Fig. 1　The method based on image processing and pattern recognition

基于此，不少学者也进行了相关研究，陶欧等[30]研究基于灰度共生矩阵提取中药材的纹理特征，Fataniya 等[31]应用形状、颜色和纹理特征研究了中药材显微图像的分类识别问题，李震[32]通过加权提取的中药材纹理片段，模糊直方图，Hu 不变矩形态特征，实现中药材的分类. 王耐等[33]利用颜色矩，灰度共生矩阵以及 Hu 不变矩来提取中药材视觉特征，实现对川牛膝的识别. 朱黎辉[34]采用梯度方向直方图，LBP 算子联合 SVM 分类器完成对中药材的分类鉴别.

传统图像处理方法针对视觉上的形状，颜色，纹理分别建立特征提取算法，再进行融合，算法复杂且未能考虑形状，颜色，纹理不同特征之间的关联性，效果较差，降低了识别效率. 同时，研究主要集中于视觉特征区别较大的分类问题，较少涉及视觉特征区别不明显的同类中药材的品质鉴定问题. 因此，基于图像处理与模式识别的中药质量检测在实际中受到很大限制.

2.2　基于深度学习技术的方法

随着人工智能技术发展的突飞猛进，学者提出结合深度学习技术来丰富传统数字图像处理问题中的算法空缺. 通过建立中药材质量与外在性状信息间的耦合关系，可实现中药质量的人工智能鉴别. 该方法具体训练流程如图 2 所示.

谭新宁等[35]利用 EasyDL 构建的深度学习模型，针对青葙子及其混伪品图像进行分类研究，其分类准确率可以达到 93.7%～94.8%，拓宽了中药品质评价的研究思路. 吴冲等[36]提出先用 YoLo

对图像进行检测分析,再用 ResNet50 等网络实现对多种不同中药材进行鉴别分类. 徐雅静等[37]基于人工智能深度学习技术及图像处理技术对中药材及中药饮片的识别方法和机制进行概述;史婷婷等[38]提出利用 GoogLeNet 网络模型对大量训练样本的学习对金银花整体分类精度可达 97.5%,面积总精度为 94.6%,完成对分散、不规则、细碎化的金银花精细分类分析.

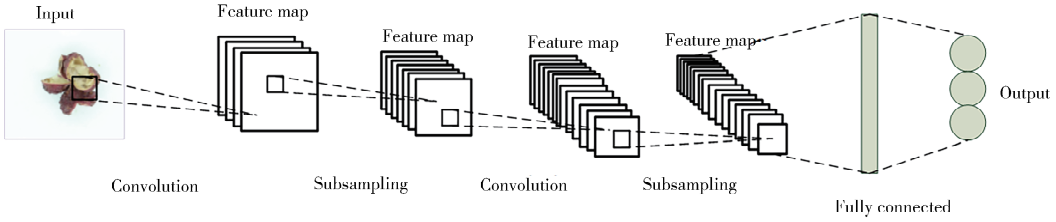

图 2　基于深度学习技术的方法

Fig. 2　The methods based on deep learning

中药饮片的性状评价结合人工智能技术作为"交叉突破口"发展突飞猛进,通过深度卷积神经网络对数据由浅层表象特征到深层语义特征进行提取,自适应的实现饮片数据的特征学习与特征表达,实现对中药材的智能鉴别与质量评价. 因此,基于人工智能算法的中药饮片鉴别分类与质量评价还有很大研究空间.

3　数据集介绍

山楂来自安徽、河南、河北、山东、山西、江苏等 10 个产地,来自不同的药材市场,均经成都中医药大学附属医院专家鉴定确认. 所有照片都是在一致的环境条件下拍摄的. 对所有图像进行注释和过滤,同时去除不完整、模糊和不合适的图像. 共采集图像 3960 张,山楂数据集如图 3 所示.

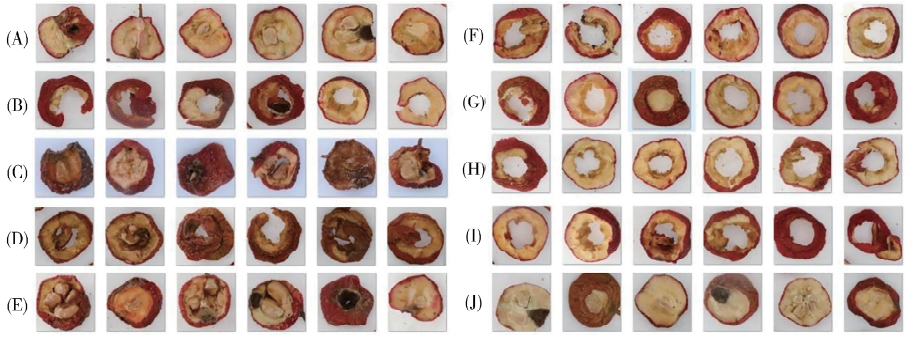

图 3　山楂数据集由 10 个不同产地组成

(A)安徽;(B)河北承德;(C)河南安阳;(D)河南野生;(E)山西;(F)河南辉县;(G)山东潍坊;(H)山东平邑;(I)山东费县;(J)江苏

Fig. 3　The CF dataset has consisted of 10 origins

(A) Anhui; (B) Chengde, Hebei; (C) Anyang, Henan; (D) Wild from Henan; (E) Shanxi; (F) Huixian, Henan; (G) Weifang, Shandong; (H) Pingyi, Shandong; (I) Linyi, Shandong; (J) Jiangsu

不同产地来源的图像具有相似的视觉特征,对于人工鉴别则要求更丰富的经验知识. 随着视觉分析技术的发展与进步,基于深度学习方法可以为中药材的快速识别提供帮助. 图 3 中不同产地山楂的数量如表 1 所示.

表 1　不同产地山楂的具体数量

Tab. 1　The number of images for CF of different origins

Name	Number	Name	Number
(A)	440	(F)	413
(B)	369	(G)	284
(C)	286	(H)	335
(D)	340	(I)	219
(E)	436	(J)	414

4　方法描述

4.1　网络结构

虽然 CNNs 模型性能有了显著的提高,但它只能专注于数据图像间的局部关系,而且具有较强的先验性,受到归纳偏差的影响. 此外,Transformer 模型方法具有强大的全局信息提取能力,但计算成本高,但泛化能力差. 因此,ConvNets 与 Transformer 相结合解决图像分类问题引起了人们的关注. 本文提出一种新的混合迭代网络,MBConv 模块采用了 Depthwise Convlution,相较于传统卷积,MBConv 结合深度可分离卷积与注意力

机制，在降低参数量的同时，提升模型特征学习能力[39]；Swin-Transformer 网络的多层次结构可弥补对图像非局部性区域损失[40]. 本文旨在利用 MBConv 模块可对局部信息建模的特点联合 Swin Transformer 可以产生不同层级特征图的属性特点，研究了一种快速有效的识别方法以实现对细粒度图像的中药材不同产地的鉴别，具体模型结构如图 4 所示. 图 4 中，SLA 为本文提出的空间注意力机制，后续内容会对其进行详细介绍；Swin 为 Swin Transformer Block.

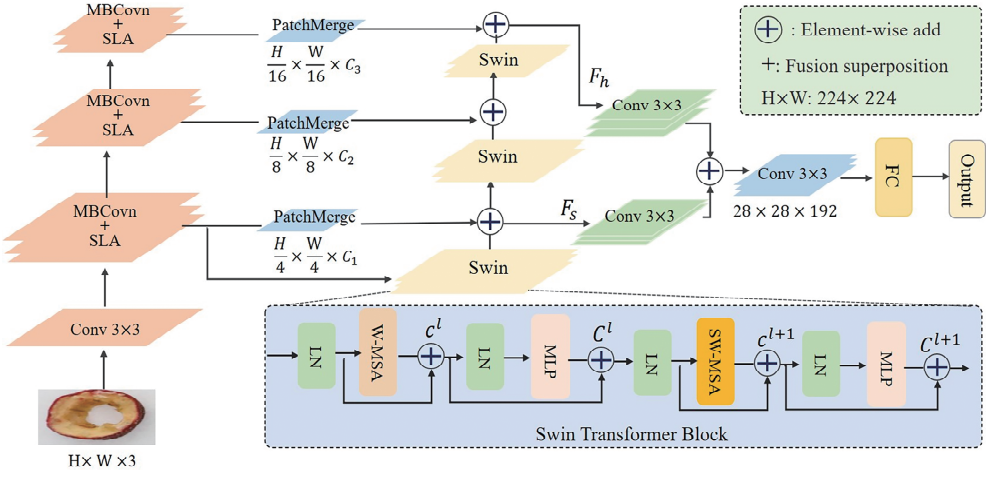

图 4　本文网络结构图

Fig. 4　The network structure of this paper

本文通过构建融合 MBConv 与 Swin Transformer 模块的混合迭代网络，实现对中药材图像细粒度特征的提取. Swin Transformer 中将用基于窗口的多头自注意（W-MSA）与移位的基于窗口的多头自注意（SW-MSA）来替代了原生的 MSA 模块. 在 W-MSA 模块中，对大小为 $M\times M$ 的局部窗口进行自注意计算；另一方面，引入 SW-MSA，利用相对于输入移位的窗口配置，来实现跨窗口间的连接，增强模型的建模能力. 模块计算过程如下.

$$c^l = WMSA(LN(C^{l-1})) + C^{l-1} \tag{1}$$

$$C^l = MLP(LN(c^l)) + c^l \tag{2}$$

$$c^{l+1} = SWMSA(LN(C^l)) + C^l \tag{3}$$

$$C^{l+1} = MLP(LN(c^{l+1})) + c^{l+1} \tag{4}$$

其中，c^l 和 C^l 分别表示块 l 的 *WMSA* 模块与 *MLP* 模块的输出特征；同理，c^{l+1} 和 C^{l+1} 分别表示 $l+1$ 快的 *SWMSA* 模块与 *MLP* 模块的输出特征. *WMSA* 和 *SWMSA* 分别表示使用规则和移位窗口划分配置的基于窗口的多头部自注意，*MLP* 表示标准多头自注意力. 在本文的实验中，给定分辨率为 224×224 大小的输入图像，将输入图像中提取特征并将其送入到主干 MBConv 模块中，得到大小为 $H/4\times W/4$ 的输出 patches，随即将 patches 经由 PatchMerge 层运算. PatchMerge 层负责降采样与增维，通过对通道数翻倍，宽高减半，来实现将 2×2 组相邻的 patch 拼接起来；在降低分辨率的同时将嵌入维数加倍，并将其传输进第一个 Swin Transformer 模块中，以更好融合进 Swin Transformer 模块. 将生成的第一个基于注意力的特征映射，通过跳跃式连接，与 MBConv 模块计算的结果添加到现有的特征中，得到最大的分支特征图 F_s. 接着以相同的方式将每层 MBConv 模块所获取的不同尺度特征图与基于 Swin Transformer 块特征映射的更高级别的特征进行输出，得到 F_d.

Swin Transformer 通过多个阶段计算和更新特征映射；在每一阶段均与前者下采样之后的特征层进行关联融合；随后收集各个阶段的特征图，得到 3 个多尺度特征图. 通过结合两者对局部性与全局性信息的捕获能力，获取图像不同层级特征，融合浅层特征与高层语义特征进行建模分析. 浅层特征是直接来自图像像素，大多是针对视觉上的形状、颜色、纹理等特征，将其作为先验知识与高层级语义信息进行融合，既突出了传统图像处理技术中特征提取的要点，又完成了对多层级特征的整合与加强；最后将融合的特征经池化层、Dropout 层及全连接层后输出实现对细粒度图像的中药材不同产地的识别分类.

4.2　细粒度特征

受限于本文数据集数量,为了更好地获取图像的特征,MBConv 模块被引入该网络,其具体网络结构图如图 5 所示. 该模块先由 1×1 的卷积,对输入特征进行升维处理;再经过一个 3×3 大小的 Depthwise Conv 卷积与 SE 模块计算,减少参数量的同时增强对特征的关注;最后由 1×1 的普通卷积进行特征降维处理,连接 Dropout 层构成的"倒瓶颈"结构的卷积模块. 第一个 1×1 层用于增加维数并更加关注有效特征,最后一个 1×1 层用于降低维数,SE 模块的引入对特征进行注意力加权.

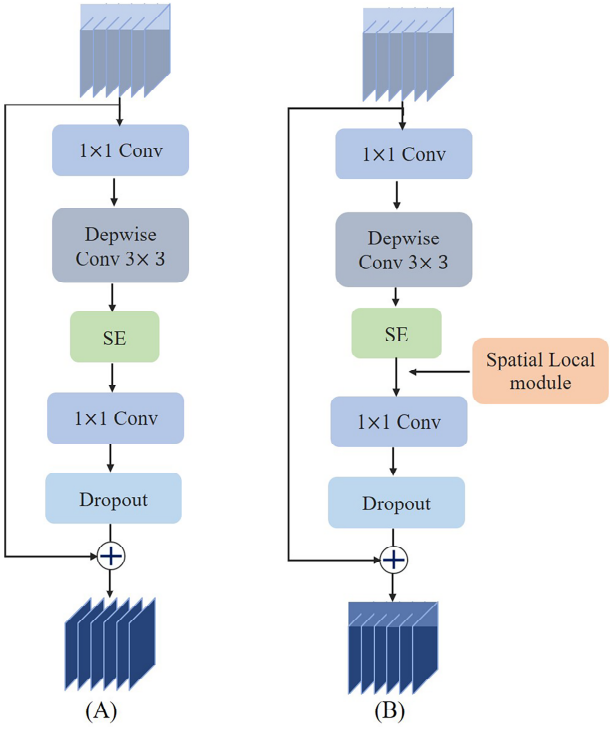

图 5　MBConv 模块结构对比图

(A)原 MBConv 模块网络结构图;(B)添加本文提出的空间注意力机制后的 MBConv 模块,新的空间注意力模块被添加于 SE 模块后

Fig. 5　MBConv module structure comparison

(A) The network structure diagram of the original MBConv module; (B) The newly proposed MBConv module, proposing that the new spatial attention module is added to the SE module

在原有的 MBConv 模块基础上,受 CBAM 混合域注意力机制[41]的启发,本文提出对 MBConv 模块的改进,即提出一种新的全局空间注意力模块应用于 SE 之后,形成通道注意力模块联合空间注意力模块的新结构,通过聚合特征映射的空间信息,压缩输入特征图的空间维数,以产生通道注意力图;在对输入特征通道进行加权后,进一步增加对空间方向的权重,本研究采用混合结构提取的特征图作为后续网络的输入,可以提高网络对图像特征的学习能力.

4.3　改进的注意力机制

空间注意力机制是指通过注意力机制,定位到原始图像在空间维度感兴趣的区域,增强对原始图片中空间信息的表现力,关注并保留关键特征信息,抑制不必要的特征,形成对空间的特征加权. 在本文的 MBConv 模块中,提出一种新的局部空间注意力模块(Spatial Local Attention Module)并将其融入 SE 模块之后,形成通道注意力混合空间注意力模块的新结构,以增强对图像细微特征的关注. 根据 CBAM 现有的空间注意力机制方法,其具体思想是采用平均池化与最大池化两种方式来获取同一像素点的每一个通道信息值,将输入图片从 $C\times H\times W$ 转换为 $1\times H\times W$,再将拼接后的特征经过卷积与上采样计算得到空间维度上的注意力权重值,从而实现对空间特征中每层像素点不同权重值的赋值. 该方法中平均池化操作是取池化区域中次重要的像素点,最大池化则是取池化区域中像素点的最大值,即关注度最高的点,放弃池化区域中不显著的值;由于细粒度图像类间差异较小,对于易混淆类别的判别通常依赖细微的特征差异,而池化操作往往更倾向于图像纹理高频信息的关注,对局部细节特征的保留效果较差. 基于此,本文提出一种新的局部空间注意力 SLA 模块(Spatial Local Attention Module),其网络结构如图 6 所示.

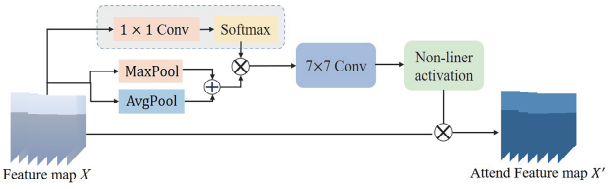

图 6　本文所提出的局部空间注意力模块 SLA 结构图

Fig. 6　The attention mechanism module SLA structure

该方法首先对输入 X 进行 1×1 *Conv* 的线性映射,进行通道数的压缩;再对线性变换后的特征向量进行 SoftMax[42]归一化处理以获取空间维度的 0-1 自注意力权重值;通过对获取的注意力权重平滑近似,将注意力进行分散,以减弱最大池化处理对局部特征的丢失,突出对图像低频区域的关注. 随后,将注意力权重与拼接的特征向量乘积,通过卷积与非线性激活 Sigmoid 函数运算,继而得到加权后的注意力特征值. 其过程数学表示过程为:

(1) 全局 attention：采用大小 1×1 卷积W_n与 SoftMax 函数相乘获取空间维度上的自注意力权重大小，然后再与池化操作处理后的W_m乘积以获得全局注意力特征.

$$W_m = \mathrm{Avg}(X_i) + \mathrm{Max}(X_i) \tag{5}$$

$$C(X_i) = \frac{e^{Z_i}}{\sum_{j=1} e^{Z_j}} \tag{6}$$

式中，X 表示输入特征；Avg 为平均池化；Max 表示最大池化，定义 $C(X)$为归一化因子，则Z^i为第 i 个节点的输出，j 为输出节点的个数，通过该操作计算获得[0−1]范围的自注意力权重. 由式(5)与式(6)则得到如下式(7).

$$Z' = \sum_{i=1} \frac{W_n \times X_i}{C(X_i)} \times W_m \tag{7}$$

式中，W_n为 1×1 的线性转换矩阵；Z'表示经全局注意计算后的输出.

(2) 特征提取：采用 7×7 的卷积W_v进行特征提取，再经由 Sigmoid 激活层函数运算，得到通道数为 1 的空间注意力权重.

$$Z'' = \sigma(Z' \times W_v) \tag{8}$$

式中，σ 表示非线性 Sigmoid 函数；Z''表示权重输出.

(3) 特征聚合：将注意力权重值Z''与原输入特征图 X 进行乘积，即可得到注意力加权后的特征图X'，通过计算不同空间维度的注意力权重值，来捕获空间维度间的依赖.

$$X' = Z'' \times X \tag{9}$$

由后续实验结果可知，该方法能增强对局部细粒度特征的感知，抑制图像中显著敏感区域，以解决当前空间注意力中池化操作更关注高频纹理而忽略局部细节的问题. 随后将该模块融入 MBConv 网络中，形成通道注意力混合空间注意力的新结构.

5　实验结果与分析

5.1　实现方式

本节介绍了所提出方法的实现训练设置，实验选用开源的 Pytorch 作为基础框架，基于 Pytorch1.8.1 与 Python3.9 进行开发，算法模型在 PC(Intel i7 处理器和 11 G 显存)上训练，显卡配置为 Nvidia GeForce RTX 2080 Ti，选用 AdamW[43]优化器对模型进行迭代优化. 学习率是衡量准确率的重要参数之一，本文采用余弦退火衰减策略，使学习率按照周期进行更替变化. 针对预训练阶段，最小学习率值为 0，初始学习率 blr 为 1e−3，训练批处理大小(batch-size)为 32，学习率权重衰减因子为 0.05，输入图像大小为 224×224，随机数种子为 1024，进行 200 次迭代，每次迭代对所有数据进行完整遍历，规定当达到 200 步时获得最终的训练模型，考虑到数据存在类间不平衡的情况，依然选用 FocalLoss 作为模型训练的损失函数，gamma 设置为 0.25，alpha 为 2.

5.2　鉴别结果分析

本文依据采用 80%的数据样本作为训练集，剩下的 20%被用来测试模型的性能. 本文提出的方法总体分类准确率为 89.306%，对于每一类具体分类结果如表 2 所示. 针对多分类研究，依旧使用准确率、召回率和特异性三个指标来评价模型的识别性能.

表 2　具体实验结果分析

Tab.2　The detailed experimental results of our model

Class	Precision	Recall	Specificity
(A)安徽	0.772	0.824	0.968
(B)河北承德	0.592	0.692	0.968
(C)河南安阳	0.951	0.893	0.993
(D)河南野生	0.909	0.923	0.987
(E)山西	0.954	0.973	0.992
(F)河南辉县	0.898	0.902	0.971
(G)山东潍坊	0.746	0.639	0.981
(H)山东平邑	0.646	0.708	0.971
(I)山东临沂	0.8	0.714	0.978
(J)江苏	0.947	0.923	0.996

从表 2 可以看出，针对不同产地的山楂鉴别中，(B)河北承德的识别精度最低为 0.592，其次是(H)山东平邑，两者的召回率明显高于识别的精确度，这表明所预测的阳性样本比实际的阳性数量高得多. (C)河南安阳，(D)河南野生，(E)山西，(F)河南辉县，(J)江苏这几类的识别分类表现较好. 另外本文还计算了混淆矩阵和 ROC 曲线，来衡量模型对不同产地山楂的鉴别性能.

图 7 与图 8 显示了每一类样本的混淆矩阵与 ROC 曲线变化. 图 7 中，从 1 到 10 的数字分别对应从 A 到 J 的不同类别，列表示预测标签，行表示真实标签. 行和列对应的值表示从真实数据中预测的正确类的数量. 图 8 中，从 0 到 9 的数字对应 A 到 J 的不同类别，以反映真阳性率和假阳性率的差异. ROC 曲线的范围为 0 ～ 1(1 为最佳，0 为最低).

从实验结果可看出,(B)河北承德的识别准确率最低,并将其识别为(G)山东潍坊、(H)山东平邑和(I)山东临沂的视觉特征具有相似性,他们的形态与纹理大多相近,它们来自山东省的不同城市,因此两者容易出现被混淆的情况.

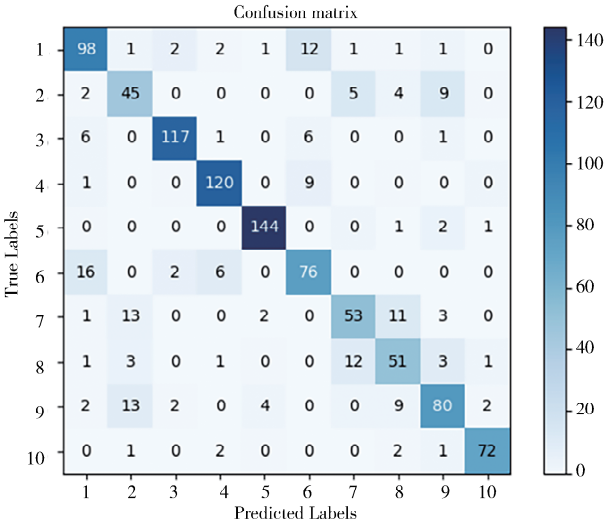

图 7　混淆矩阵的实验结果

Fig. 7　The experimental results of confusion matrix

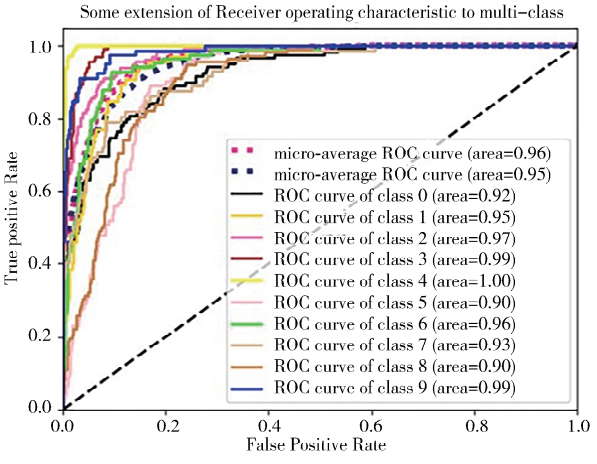

图 8　ROC 的实验结果

Fig. 8　The experimental results of Receiver Operating Characteristic (ROC)

5.3　不同 CNNs 方法的对比

为验证本文方法的性能,我们将本文方法与多个不同的 ConvNets 和 Transformers 进行了比较. 其中, VGG[44]、ResNet[44]、DenseNet[44]、MobileNet[44]、EfficientNet[45] 和 CoAtNet[46] 均选择 Focalloss 作为损失函数. 考虑评价模型的公平性,基线是基于改进的模型,其余训练的参数在对比实验中均保持不变. 针对不同产地山楂数据集的分类鉴别中,将本文方法与不同 ConvNets 方法的对比结果如表 3 所示.

表 3　针对不同卷积神经网络 ConvNets 方法实验对比结果

Tab. 3　The results of the comparison for the multiple different ConvNets

Name	Top-1 ACC /%	Params/M	FLOPs/G
VGG[44]	76.375	138.3	15.5
ResNet[44]	77.36	25.56	4.12
DenseNet[44]	80.251	14.15	3.42
EfficientNet[45]	83.347	5.29	0.01
MobileNet[44]	82.387	3.5	0.32
CoAtNet[46]	78.944	17.03	3.33
本文的方法	89.306	15.92	4.76

从表 3 可知,本文方法相较于其他方法有最高的鉴别准确率,相较 CoAtNet,提升 10.362%;相较 VGG,提升了 12.931%;相较 ResNet,提升了 11.946%;相较 EfficientNet,提升了 5.959%;从实验数据可看出,本文提出的方法有显著的提升效果,可获取图像更为丰富的细节特征. 另外,从表 3 实验对比的不同模型的参数量与模型复杂度可看出,传统卷积模型参数量较大、计算复杂度相对较大,且分类精度有限;而轻量级网络相较于前者,参数量得到极大降低的同时,准确度也有一定范围的增长,为本文实验选择 MBConv 模块提供理论基础.

5.4　不同 Transformers 方法的对比

同时,多种不同的基于 Transformers 算法也被用来与本文提出的方法进行比较,包括 ViT[26]、MaxViT[47]、Swin -Transformer[40]、FocalNet[48] 与 CMT[49] 等. 考虑评价模型的公平性,基线是基于改进的模型,其余训练的参数在对比实验中均保持不变.

表 4 显示将本文方法与当下最新的基于 Transformer 的方法进行对比;从实验结果可知,本文方法依然有最高的识别准确率,较 ViT 提升了 10.541%,MaxViT 提升了 9.06%,FocalNet 提升了 8.259%,相较 Swin Transformer,提升了 3.633%,依然是优于其他算法模型;另外,从分类精准度来看,本文所提出的模型方法相较于 ViT、CvT 等参数量有所降低,优势显著. 实验结果证明,本文所提出的分层级结构方法更适用于细粒度图像的分类识别问题,通过捕获细粒度图像多层次特征,让网络学习到更多有效特征细节,从而实现从视觉上的山楂产地精细分类.

表 4　基于不同 Transformers 网络结构实验对比结果

Tab. 4　The results of the comparison for the multiple different Transformers

Method	Top-1 ACC/%	Params/M	FLOPs/G
ViT[23]	78.765	86.57	16.86
MaxViT[47]	80.246	31.1	5.6
FocalNet[48]	81.047	27.67	4.41
Swin Transformer[40]	85.673	28.29	4.36
CMT[49]	82.347	9.51	0.62
CvT[50]	82.726	20.23	4.53
PVT[51]	78.620	24.52	3.81
本文方法	89.306	15.92	4.76

5.5　消融实验

5.5.1　注意力机制模块消融　为了验证本文提出的注意机制模块的有效性，我们在不同数据集上比较了不同的注意机制方法，空间注意模块(SPA)[38]和本文所提出的注意力机制模块(SLA). 本次对比试验采用控制变量法，消融中的基准模型是基于本文所提出的模型，其余训练参数在对比实验中保持不变. 基于不同注意力机制的不同产地山楂的鉴别结果如表 5 所示.

表 5　基于不同注意力机制的对比实验结果

Tab. 5　The comparative experimental results based on different attention mechanisms

注意力机制	Top-1 ACC/%	Params/M	FLOPs/G
without Attention	88.054	15.91	4.74
SPA	88.824	15.91	4.74
SLA	89.306	15.92	4.76

从表 5 可以看出，本文提出的注意力机制模型识别准确率最高 89.306%，相较于使用 SPA，本文的方法提升了 0.482%. 同时，与不添加空间注意力机制相比，可以确认加入空间注意机制模块后，模型的性能提高了 1.252%. 从参数量数值来看，参数量变化并不大，因此本文所提出的注意力机制具有良好的实用性. 此外，为了更好地验证两个注意模块的效果，我们进一步将不同注意力机制下的热力图显示在图 9 中. 图 9 中，每类的热图都是随机选择，其中第一个是原始图像；第二个是原 MBConv 模块的热图；第三个是加入 SPA 的热图；最后一个是本文提出的方法.

通过可视化的不同热图结果，可以清楚地观察到不同注意机制模块对图像关注区域的差异变化. 从图 9 可以看出，所有模型都可以聚焦在图像的感兴趣区域(ROI)上，与 SPA 相比，本文提出的注意力模块所涉及更深更广的 ROI 范围，关注区域更多；而对于识别准确率较低的(B)河北承德、(G)山东潍坊和(H)山东平邑，不同模型关注的 ROI 范围略有不同，证明不同注意力模型所聚焦的区域不一致，学习的特征也差异明显. 本文方法更容易获取到图像中具有区分性的区域，从而帮助网络学习到更加有效的特征，聚焦于图像局部区域的特征细节，用来完成对细粒度图像的分类与识别，提高模型的识别准确率.

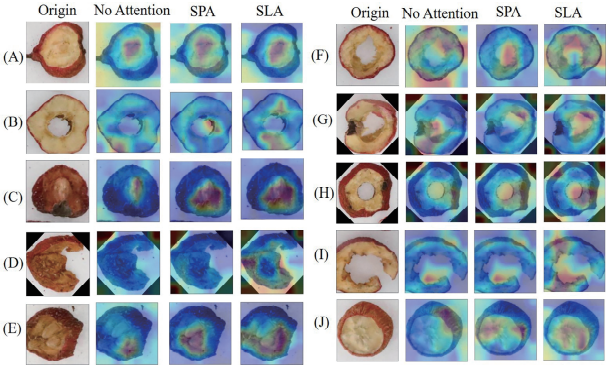

图 9　不同产地山楂的注意力可视化图

Fig. 9　The visualization of the different attention modules for each class

5.5.2　不同卷积模块消融　为验证使用 MBConv 模块的有效性，在本文实验中也选择使用不同类型卷积与 Swin Transformer 模块进行组合，考虑对比实验的公平性，每种卷积模块数量均与本文所使用的 MBConv 模块数量一致，其具体对比计算结果如表 6 所示.

表 6　基于不同卷积模块的对比实验结果

Tab. 6　The comparative experimental results based on different CNN modoule

CNN 模块	Top-1 ACC/%	Params/M	FLOPs/G
VGG[44]	86.431	31.42	10.51
ResNet[44]	87.272	56.34	4.57
DenseNet[44]	87.654	89.97	4.18
MobileNet[44]	87.817	10.83	3.94
本文方法	89.306	15.92	4.76

针对表 6 中，VGG 指基础卷积，ResNet 为残差卷积，DenseNet 指稠密连接卷积，MobileNet 则为深度可分离卷积模块；根据结果可知，本文的模型虽然不具备最优的参数量与计算复杂度，但其鉴别精准度依然为最高，所使用的 MBConv 模块能获取图像更为丰富的粒度特征，相较 ResNet，提高

2.304%;DenseNet,提高了 1.652%;相较于 MobileNet,提高了 1.489%,验证了所使用模块的科学性与合理性.

6 结　论

本文针对不同产地的中药材细粒度图像进行视觉鉴别.受 CoAtNet 与 Swin-Transformer 网络启发,结合 MBConv 模块中深度可分离卷积网络对局部信息建模的特点与 Swin Transformer 模块多层次结构可弥补网络非局部性损失的特性,本文提出一种新的混合深度神经网络模型,通过获取图像不同层级特征,将获取的形状、颜色与纹理等浅层特征作为先验知识与高层级语义信息进行特征融合,实现对不同产地山楂细粒度图像的精准鉴别.通过实验结果分析,本文的方法有最高的鉴别准确率为 89.306%,相较 CoAtNet,提升 10.362%;相较 ResNet,提升了 11.946%;相较 EfficientNet,提升了 5.959%;从实验数据可看出,本文提出的方法更有利于获取图像丰富的特征细节.从实验数据可看出,本文提出的方法能捕获图像丰富的细节特征,实现更多准确度识别.另外本文通过消融实验验证了提出的注意力模块有效性,实验结果表明,本文提出新的注意力机制相较于 SPA,提升了 0.482%;与不添加空间注意力机制相比,模型的性能提高了 1.252%.另外通过对不同类型卷积模块的消融,也证实了本文使用 MBConv 模块的科学性与合理性.实验结果证明,本文提出的方法通过注入浅层先验特征引导高层语义信息,获取图像更为丰富的细粒度特征细节,从而帮助网络学习到更加有效的特征,聚焦于图像局部区域的特征细节,用来实现对不同产地山楂细粒度图像的分类与识别,对模型性能提升有积极作用.

参考文献:

[1] Ministry of Public Health of the People's Republic of China. Pharmacopoeia of the People's Republic of China, part 1 [M]. Beijing: China Pharmaceutical Technology Press, 2015.[中华人民共和国卫生部. 中华人民共和国药典[M]. 北京:中国制药技术出版社, 2015.]

[2] Wu J Q, Peng W, Qin R X, *et al*. Crataegus pinnatifida: chemical constituents, pharmacology, and potential applications [J]. Molecules, 2014, 19: 1685.

[3] Liu C, He Q, Zeng L L, *et al*. Digestion-promoting effects and mechanisms of dashanzha pill based on raw and charred crataegi fructus [J]. Chem Biodivers, 2021, 18: e2100705.

[4] Li L, Yang S L, Liu Y J, *et al*. Preliminary study of odor change mechanism in Crataegi Fructus stir-fried process based on correlation analysis [J]. China J Chinese Mate Med, 2014, 39: 3283.

[5] Xue Q Q, Wang Y L, Fei C H, *et al*. Profiling and analysis of multiple constituents in Crataegi Fructus before and after processing by ultrahigh-performance liquid chromatography quadrupole time-of-flight mass spectrometry [J]. Rapid Commun Mass Sp, 2021, 35: e9033.

[6] Weon J B, Jung Y S, Ma C J. Quality analysis of chlorogenic acid and hyperoside in crataegi fructus [J]. Pharmacogn Mag, 2016, 12: 98.

[7] Yang X L, Sun X G, Zhou L J, *et al*. Study on the quality of Crataegi Fructus and its processed products from different origins[J]. Chinese J Inform TCM, 2022,18: 1.

[8] Yin F Z, Li L, Chen Y, *et al*. Quality control of processed Crataegi Fructus and its medicinal parts by ultra-high-performance liquid chromatography with electrospray ionization tandem mass spectrometry [J]. J Sep Sci, 2015, 38: 2630.

[9] Yu J S, Guo M Y, Jiang W J, *et al*. Illumina-based analysis yields new insights into the fungal contamination associated with the processed products of crataegi fructus [J]. Front Nutr, 2022, 9: 883698.

[10] Qin R X, Xiao K K, Li B, *et al*. The combination of catechin and epicatechin gallate from fructus crataegi potentiates beta-lactam antibiotics against methicillin-resistant staphylococcus aureus (MRSA) in vitro and in vivo [J]. Int J Mol Sci, 2013, 14: 1802.

[11] Fei C H, Dai H, Wu X Y, *et al*. Quality evaluation of raw and processed Crataegi Fructus by color measurement and fingerprint analysis [J]. J Sep Sci, 2018, 41: 582.

[12] Lee J J, Lee H J, Oh S W. Antiobesity effects of sansa (Crataegi fructus) on 3T3-L1 cells and on high-fat-high-cholesterol diet-induced obese rats [J]. J Med Food, 2017, 20: 19.

[13] Xie D S, Liu Y J, Yang S Y, *et al*. Processing degree of strychni semen based on combination of inside and outside analysis[J]. Chinese J Exper Tradit Med Form, 2016, 22: 1.[解达帅, 刘玉杰, 杨诗龙, 等. 基于"内外结合"分析马钱子的炮制火候

[J]. 中国实验方剂学杂志，2016，22：1.]

[14] Tan C Q, Wen C B, Wu C J. Research on recognition of traditional Chinese medicine pieces based on image processing technology [J]. Lishizhen Medicine and Materia Medica Research, 2018, 29: 1706. [谭超群，温川飙，吴纯洁. 基于图像处理技术的中药饮片识别研究[J]. 时珍国医国药，2018，29：1706.]

[15] Cai C Z, Yuan Q F, Xiao H Y, *et al*. Computer-aided classification and identification of traditional Chinese medicine prescriptions [J]. J Chongqing Univ(Nat Sci Ed), 2007, 29: 42. [蔡从中，袁前飞，肖汉光. 中药组方的计算机辅助分类与识别[J]. 重庆大学学报(自然科学版)，2007，29：42.]

[16] Y L C, Y B I, Hinton G. Deep learning [J], Nature, 2015, 521: 436.

[17] Spencer M, Eickholt J, Cheng J. A deep learning network approach to ab initio protein secondary structure prediction [J]. IEEE ACM T Comput Bi, 2015, 12: 103.

[18] Tan C Q, Wu C, Huang Y L, *et al*. Identification of different species of Zanthoxyli Pericarpium based on convolution neural network [J]. Plos One, 2020, 15: e0230287.

[19] Wang B, Li H, You J W, *et al*. Fusing deep learning features of triplet leaf image patterns to boost soybean cultivar identification [J]. Comput Electr Agr, 2022, 197: 106914.

[20] Zhang Y, Wen K Z, Min S Z, *et al*. Remote sensing image denoising based on attention mechanism and perceptual loss [J]. J Sichuan Univ (Nat Sci Ed), 2021, 58: 042001. [张意，阚子文，邵志敏，等. 基于注意力机制和感知损失的遥感图像去噪[J]. 四川大学学报(自然科学版)，2021，58：042001.]

[21] Lu Y M, Bu L M, Chen L, *et al*. Extracting clinical experiences from ancient literature of traditional chinese medicine [J]. J Sichuan Univ (Nat Sci Ed), 2022, 59: 023005. [卢永美，卜令梅，陈黎，等. 基于深度学习的中医古文献临床经验抽取[J]. 四川大学学报(自然科学版)，2022，59：023005.]

[22] Lu T, Yu F H, Xue C H, *et al*. Identification, classification, and quantification of three physical mechanisms in oil-in-water emulsions using AlexNet with transfer learning [J]. J Food Eng, 2021, 288: 110220.

[23] Jin R L, Qing L B, Wen H Q. Emotion recognition of the natural scenes based on attention mechanism and multi-scale network [J]. J Sichuan Univ(Nat Sci Ed), 2022, 59: 012003. [晋儒龙，卿粼波，文虹茜. 基于注意力机制多尺度网络的自然场景情绪识别[J]. 四川大学学报(自然科学版)，2022，59：012003.]

[24] Peng Z X, Pu Y F. Convolution neural network face recognition based on fractional differential [J]. J Sichuan Univ(Nat Sci Ed), 2022, 59: 012001. [彭朝霞，蒲亦非. 基于分数阶微分的卷积神经网络的人脸识别[J]. 四川大学学报(自然科学版)，2022，59：012001.]

[25] He Y, Yang P, Wang C S, *et al*. Research on financing websites identification based on deep neural network [J]. J Sichuan Univ(Nat Sci Ed), 2021, 58: 033003. [何颖，杨频，王丛双，等. 基于深度神经网络的配资网站识别研究[J]. 四川大学学报(自然科学版)，2021，58：033003.]

[26] Han K, Wang Y H, Chen H T, *et al*. A survey on vision transformer [J]. IEEE T Pattern Anal, 2022, 45: 87.

[27] Bazi Y, Bashmal L, Al Rahhal M M, *et al*. Vision transformers for remote sensing image classification [J]. Remote Sens-Basel, 2021, 13: 516.

[28] Kim K, Wu B C, Dai X L, *et al*. Rethinking the Self-Attention in vision transformers [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops [S. l.]: IEEE, 2021: 3071.

[29] Touvron H, Cord M, El-Nouby A, *et al*. Three things everyone should know about Vision Transformers [C]// European Conference on Computer Vision. Tel Aviv, Israel: Springer Nature Switzerland, 2022: 497.

[30] Tao O, Lin Z Z, Zhang X B, *et al*. Research on traditional Chinese medicine identification model based on texture feature parameters of sliced slice images [J]. World Science and Technology-Modernization of Traditional Chinese Medicine, 2014, 12: 2558. [陶欧，林兆洲，张宪宝，等. 基于饮片切面图像纹理特征参数的中药辨识模型研究[J]. 世界科学技术-中医药现代化，2014，12：2558.]

[31] Fataniya B, Modi U, Zaveri T. Automatic identification of licorice and rhubarb by microscopic image processing [J]. Procedia Comput Sci, 2015, 58: 723.

[32] Li Z. Feature extraction and recognition system for traditional Chinese medicine pieces [D]. Harbin Institute of Technology, 2013. [J 李震. 中药饮片特

征提取和识别系统[D]. 哈尔滨:哈尔滨工业大学, 2013.]

[33] Wang N, Lu W B, Ling X H, *et al*. Feature extraction and image recognition of Achyranthes bidentata and Cyathula officinalis[J]. China Pharmacy, 2017, 28, 12:1670.[王耐, 卢文彪, 凌秀华, 等. 牛膝和川牛膝药材的特征提取与图像识别[J]. 中国药房, 2017, 28, 12: 1670.]

[34] Zhu L H, Li X N, Zhang Y, *et al*. Chinese herbal medicine retrieval method based on shape features and texture features [J]. Comp Eng Des, 2014, 35, 11: 3903.[朱黎辉, 李晓宁, 张莹, 等. 基于形状特征及纹理特征的中药材检索方法[J]. 计算机工程与设计, 2014, 35, 11: 3903.]

[35] Tan X N, Wu W R, Liang W Q, *et al*. Application of customized AI training platform EasyDL in image classification of Qinghuizi and its counterfeit products [J]. Chin J Ethnomed Ethnopharm, 2022, 31: 40.[谭新宁, 吴文如, 梁婉晴, 等. 定制化 AI 训练平台 EasyDL 在青葙子及其混伪品图像分类中的应用[J]. 中国民族民间医药, 2022, 31: 40.]

[36] Wu C, Tan C Q, Huang Y L, *et al*. Intelligent identification of fritillary fritillary, hawthorn and pinellia decoction pieces based on deep learning algorithm [J]. China J Expe Trad Med Form, 2020, 26: 195.[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: 195.]

[37] Xu Y J, Yu J, Yu Y P, *et al*. Application of artificial intelligence in the field of identification of Chinese medicinal materials and slices [J]. Chin Arch Trad Chin Med, 2022, 40: 47.[徐雅静, 俞捷, 余远盼, 等. 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: 47.]

[38] Shi T T, Zhang X B, Guo L P, *et al*. Research on remote sensing identification method of wild-planted honeysuckle based on deep convolutional neural network [J]. China J Chin Mater Med, 2020, 45: 5658.[史婷婷, 张小波, 郭兰萍, 等. 基于深度卷积神经网络的仿野生种植金银花遥感识别方法研究[J]. 中国中药杂志, 2020, 45: 5658.]

[39] Han K, Wang Y H, Chen H T, *et al*. A survey on vision transformer [J]. IEEE T Pattern Anal, 2022, 45: 87.

[40] Liu Z, Lin Y T, Cao Y, *et al*. Swin transformer: hierarchical vision transformer using shifted windows [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Virtual: IEEE, 2021: 10012.

[41] Woo S H, Park J C, Lee J Y, *et al*. CBAM: convolutional block attention module [C]//Proceedings of the European conference on computer vision (ECCV). Munich, German: Springer Nature Switzerland, 2018: 3.

[42] Wu A M, Han Y H, Yang Y, *et al*. Convolutional reconstruction-to-sequence for video captioning [J]. IEEE T Circ Syst Vid, 2020, 30: 4299.

[43] Kingma D P, Ba J. Adam: a method for stochastic optimization [EB/OL]. [2022-12-01]. https://arxiv.org/abs/1412.6980.

[44] Li Y, Pang Y, Wang J, *et al*. Patient-specific ECG classification by deeper CNN from generic to dedicated [J]. Neurocomputing, 2018, 314: 336.

[45] Tan M X, Le Q V. EfficientNet: rethinking model scaling for convolutional neural networks[C]//International Conference on Machine Learning. [S.l]: PMLR, 2019: 6105.

[46] Dai Z, Liu H, Le Q V, *et al*. CoAtNet: marrying convolution and attention for all data sizes [J]. Adv Neural Inf Process Sys, 2021, 34: 3965.

[47] Tu Z Z, Talebi H, Zhang H, *et al*. MaxViT: multi-axis vision transformer [C]// Proceedings of the European Conference on Computer Vision (ECCV). Tel Aviv, Israel: Springer Nature Switzerland, 2022: 459.

[48] Yang J W, Li C Y, Gao J F. Focal modulation networks [J]. Adv Neural Inf Process Sys, 2022, 35: 4203.

[49] Guo J Y, Han K, Wu H, *et al*. CMT: convolutional neural networks meet vision transformers[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE, 2022: 12175.

[50] Wu H P, Xiao B, Codella N, *et al*. CvT: introducing convolutions to vision transformers[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Virtual: IEEE, 2021: 22.

[51] Wang W H, Xie E Z, Li X, *et al*. Pyramid vision transformer: a versatile backbone for dense prediction without convolutions[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). [S.l.]: IEEE, 2021: 568.