

引用格式: 吴迪, 肖衍, 沈学军, 等. 基于改进 Res2Net 与迁移学习的水果图像分类 [J]. 电子科技大学学报, 2025, 54(1): 62-71.

WU D, XIAO Y, SHEN X J, et al. Fruit image classification based on improved Res2Net and transfer learning[J]. Journal of University of Electronic Science and Technology of China, 2025, 54(1): 62-71.

基于改进 Res2Net 与迁移学习的水果图像分类



吴迪^{1,2*}, 肖衍¹, 沈学军¹, 万琴^{1,2}, 陈子涵¹

(1. 湖南工程学院 电气与信息工程学院, 湘潭 411100; 2. 湖南大学 机器人视觉感知与控制技术国家工程研究中心, 长沙 410082)

摘要: 针对传统水果图像分类算法特征学习能力弱和细粒度特征信息表示不强的缺点, 提出一种基于改进 Res2Net 与迁移学习的水果图像分类算法。首先, 针对网络结构, 在 Res2Net 的残差单元中引入动态多尺度融合注意力模块, 对各种尺寸的图像动态地生成卷积核, 利用 meta-ACON 激活函数优化 ReLU 激活函数, 动态学习激活函数的线性和非线性, 自适应选择是否激活神经元; 其次, 采用基于模型迁移的训练方式进一步提升分类的效率与鲁棒性。实验结果表明, 该算法在 Fruit-Dataset 和 Fruits-360 数据集上的测试准确率相比 Res2Net 提升了 1.2% 和 1.0%, 召回率相比 Res2Net 提升了 1.13% 和 0.89%, 有效提升了水果图像分类性能。

关键词: 图像分类; Res2Net; 动态多尺度融合注意力; 激活函数; 迁移学习

中图分类号: TP391.4

文献标志码: A

DOI: 10.12178/1001-0548.2023258

Fruit image classification based on improved Res2Net and transfer learning

WU Di^{1,2*}, XIAO Yan¹, SHEN Xuejun¹, WAN Qin^{1,2}, and CHEN Zihan¹

(1. Institute of Electrical and Information Engineering, Hunan Institute of Engineering, Xiangtan 411100, China;

2. National Engineering Research Center of RVC, Hunan University, Changsha 410082, China)

Abstract: Aiming at the shortcomings of the traditional fruit image classification algorithm with weak feature learning ability and weak representation of fine-grained feature information, this paper proposes a fruit image classification algorithm based on improved Res2Net with migration learning. First, for the network structure, a dynamic multi-scale fusion attention module is introduced into the residual unit of Res2Net to dynamically generate convolution kernels for images of various sizes, optimize the ReLU activation function by using the meta-ACON activation function, and dynamically learn the linearity and nonlinearity of the activation function to adaptively choose whether to activate the neurons or not; second, a training method based on model migration is used to further improve the efficiency and robustness of classification. The experimental results show that the algorithm proposed in this paper improves the test accuracy on Fruit-Dataset and Fruits-360 dataset by 1.2% and 1% compared with Res2Net, and the recall rate improves by 1.13% and 0.89% compared with Res2Net, which effectively improves the performance of fruit image classification.

Key words: image classification; Res2Net; dynamic multi-scale fusion attention; activation function; transfer learning

水果分类技术一直是国内外农业信息化领域的研究热点问题。然而, 在实际应用场景中, 光照、阴影、模糊等^[1-2] 外界因素会影响图片的质量,

给特征提取带来困难^[3]。因此, 如何提取出更有效和更有代表性的特征对水果图像分类具有重要意义。

收稿日期: 2023-10-13

基金项目: 国家重点研发计划 (2020YFB1713600); 国家自然科学基金 (62476084); 湖南省教育厅重点项目 (24A0528); 湖南省自然科学基金 (2022JJ30198); 湖南省研究生科研创新项目 (YC202228)

作者简介: 吴迪, 博士, 副教授, 主要从事多模态融合行人再识别和信息融合理论与应用等方面的研究。

*通信作者 E-mail: wudi6152007@163.com

水果图像分类算法主要分为基于传统机器学习和基于深度学习的方法^[4]。基于传统机器学习算法主要有随机森林、最近邻算法和支持向量机等^[5], 若简单地提取水果的颜色、几何、纹理等粗粒度特征, 传统机器学习算法只能较好地识别类间差异较大的水果, 在面对细粒度水果图像时, 准确率较低且需要大量人工标注。为了减少人工标注带来的人力物力消耗和提升分类性能, 基于深度学习的水果图像分类方法受到越来越多的关注。文献 [6] 基于 GoogLeNet 模型对香蕉进行分类, 该结构同时使用多个不同大小的卷积核和池化操作, 从而可以有效地提取图像中不同级别的特征, 但这可能导致特征图的稀疏性增加, 使得网络难以提取到细节信息。文献 [7] 基于 VGG-16 模型对水果进行分类, 该模型使用多个卷积层和池化层的堆叠, 以及小尺寸的卷积核, 能够更好地捕捉图像的细节特征, 但较深的网络结构会导致梯度消失问题, 无法提取更为复杂的特征。文献 [8] 基于 ResNet 模型对椰枣进行分类, 通过残差连接将信息直接从浅层传递到深层, 能够保持和增强原始特征的传递, 这种特征传递机制有助于网络学习更准确、更具区分性的特征, 但对于多尺度的特征提取能力较差。文献 [9] 基于 Inception-ResnetV2 模型对橄榄果实进行分类, 通过结合 Inception 模块和残差模块, 可以捕捉多个不同尺度的特征和解决深层网络中的梯度消失问题, 能够学习到更复杂的图像特征, 但网络的计算复杂度相对较高。因此, 在保证现有神经网络模型性能相同的情况下, 研究人员不再追求网络层数的堆叠, 而是在网络结构设计和训练方式等方面进行改进。

近年来, 注意力机制广泛应用于计算机视觉。文献 [10] 以 ResNet 为骨干网络, 提出将多个注意力在不同网络之间迁移的方法, 能够识别更有判别力的特征而不增加计算复杂度。文献 [11] 以 DenseNet 为骨干网络, 使用 CBAM 混合注意力机制, 同时从通道和空间层面加强对关键特征的提取, 提高了水果分类的准确率。文献 [12] 使用注意力机制对 ShuffleNet 进行改进, 通过分组卷积和通道重排实现了较低的计算复杂度, 提高了水果识别的准确率。不同大小的卷积核对于不同尺度的水果图像会有不同的效果, 虽然可以利用注意力机制增加多个卷积核来适应不同尺度的水果图像, 但是一旦训练完成后, 参数就固定, 每个卷积核的权重相同。因此, 如何利用注意力

机制来解除参数固定的限制, 强调卷积核的重要性, 针对不同尺度水果图像动态生成卷积核是亟须解决的问题。

文献 [13] 是最早应用于卷积神经网络的非线性饱和激活函数, 但是它们都存在软饱和问题, 即当参数在训练过程中处于饱和区时, 会出现梯度消失的现象。基于此, 文献 [14] 提出了 ReLU 函数, 它是非饱和激活函数, 可以缓解饱和函数存在梯度消失的问题。由于 ReLU 将所有负值取为 0, 导致模型在训练过程中很容易出现“神经元死亡”, 从而致使模型分类效果下降。针对此问题, 文献 [15] 提出了 meta-ACON 激活函数, 它可以自适应地选择是否激活神经元, 以提升模型分类能力。

迁移学习使用预训练模型, 能有效提升图像分类的效率和精度。文献 [16] 采用 Net2Net 训练策略, 将一个网络的知识迁移到另一个网络, 大多数卷积层中的参数量减少到原来的 30%, 提高了训练效率。文献 [17] 基于 MobileNetV2 模型, 通过深度可分离卷积和特殊的残差块设计, 采用迁移学习训练方式进行分类, 提高了水果分类效率和准确率。文献 [18] 基于 ResNet 模型, 采用迁移学习的训练方式来识别水果的新鲜度, 达到了较好的分类效果。由此可见, 改进网络结构和改变训练方式是提升水果图像分类性能的一种亟待探索的方法。

基于此, 本文提出一种基于改进 Res2Net 与迁移学习的水果分类方法。首先, 在 Res2Net^[19] 中, 加入多尺度融合注意力机制 (selective kernel network, SKNet)^[20], 对不同尺度图像自适应地选择合适的卷积核, 提取图像关键特征, 用 meta-ACON 激活函数优化 ReLU 激活函数, 自适应学习是否激活神经元; 其次, 采用迁移学习的训练方式, 解决数据集间因内容差异而特征识别力低的问题, 避免了从头开始训练模型的缺点, 减少训练时间。

1 基于改进 Res2Net 与迁移学习的图像分类算法

为提高传统水果图像分类算法细粒度特征信息的表示能力, 深度学习展现出其独特的优势。在此基础上, 本文从特征提取网络设计和训练方式两个方面出发, 以迁移学习为切入点, 对水果图像分类任务进行研究。

1.1 本文所提算法网络结构

Res2Net 通过提高模型的多尺度特征学习能

力, 能够有效解决水果图像分类任务, 但是该算法引入了多个子路径和残差组, 以及不同尺寸的卷积核, 导致计算复杂性增加。此外, 在数据集较小的情况下, 没有足够的训练数据来有效地训练模型, 可能导致模型过拟合, 因此, 本文对 Res2Net 算法进行改进, 如图 1 所示。图中的 Conv 代表卷积层, BN 代表批量归一化层, ACON 代表激活函数, Max pooling 代表最大池化, GAP 代表全局平均池化, FC 代表全连接层, stage 代表模型内部 1 个大模块, 它由多个基本单元构成。首先, 本文根据多尺度融合注意力具备动态选择和融合不同尺度的卷积核的能力,

改进了 Res2Net 的多尺度网络, 从而有效提升网络的特征表达能力。然后, 用 meta-ACON 激活函数替换多尺度融合注意力模块中的 ReLU 激活函数, 可以动态学习激活函数的线性和非线性, 自适应地选择是否激活神经元, 进一步提升 Res2Net 的非线性表达能力。最后, 使用迁移学习的训练方式, 通过在预训练模型的基础上进行微调, 使模型更快地收敛到新任务, 从而减少训练时间和资源消耗。改进的模型结构由 3 个 3×3 卷积, 3×3 最大池化, 分组为 3、4、6、3 的动态分组残差单元结构, 1 个全局平均池化层, 1 个全连接层, 以及 1 个 softmax 分类层组成。

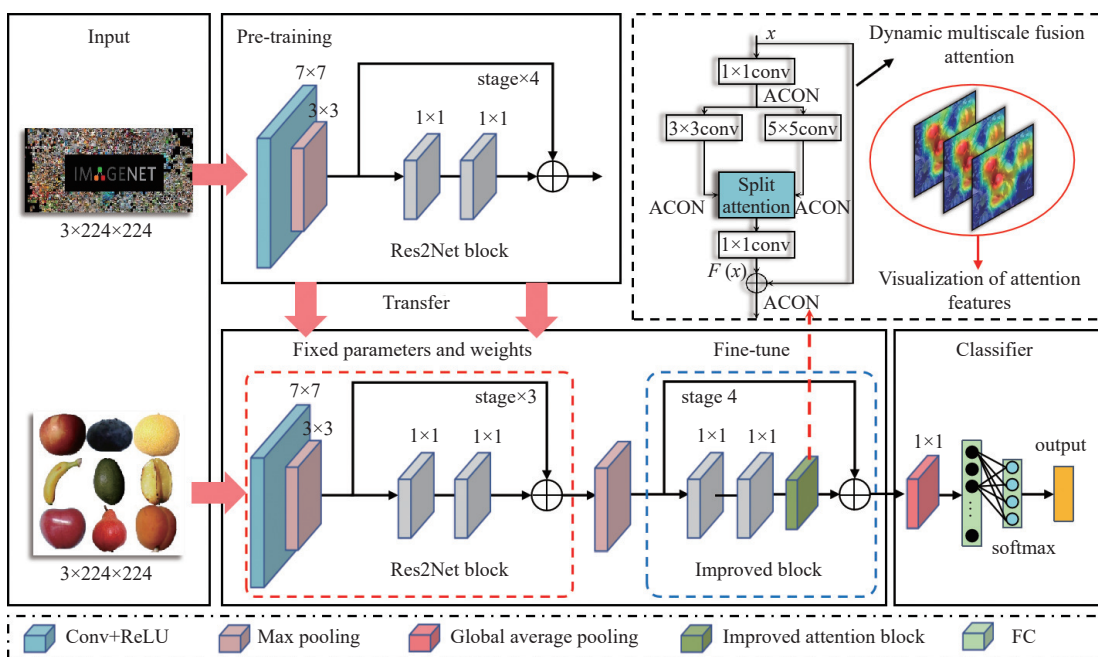


图 1 本文算法网络结构

1.2 动态多尺度融合注意力

为进一步提升网络对关键特征的提取能力, 本文提出了一种新的动态多尺度融合注意力。首先, 对多尺度融合注意力网络中的单元进行改进, 相较于传统单卷积核提取特征, 通过选取 2 个或多个不同大小的卷积核来提取多尺度特征^[21], 并进行特征融合; 其次, 用 meta-ACON 激活函数优化常用的 ReLU 激活函数。动态多尺度注意力能够自适应地选择合适权重来处理不同尺度的特征, 进一步提高了卷积核提取特征的能力, 并利用 meta-ACON 的非线性表达能力, 降低 ReLU 的稀疏性, 避免特征图中积累更多的特征冗余, 来提取水果图像特征的细粒度信息。具体结构如图 2 所示。

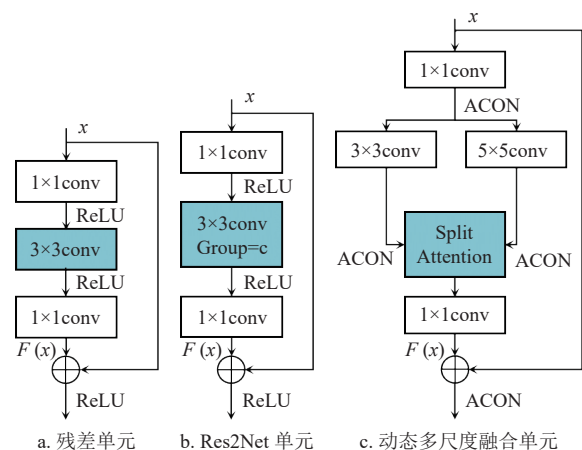


图 2 动态多尺度融合注意力结构

1.2.1 多尺度融合注意力

多尺度融合注意力能够自适应地选择合适权重来处理不同尺度特征, 提高了模型对细粒度和多尺度信息的学习表达能力。其中, 多尺度融合注意力网络是用于关注细粒度特征的一种通道注意力机制, 能够根据输入图像特点自适应地选择卷积核尺寸, 增大网络的感受野, 以获取图像中不同的特征信息, 网络结构如图 3 所示。

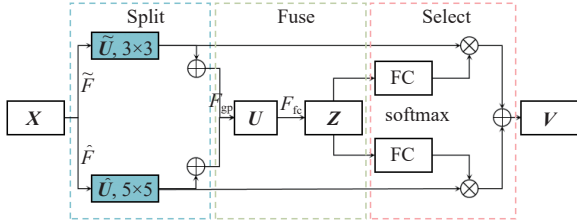


图 3 多尺度融合注意力结构图

Fuse 将 \hat{U} 和 \tilde{U} 融合得到 U , 将 U 通过全局平均池化获得特征向量 S 以统计每个通道携带的特征信息, 再将该特征向量输入至全连接层得到向量 Z 。Select 对向量 Z 通过 softmax 函数预测出不同尺度的通道权重特征层 a 和 b , 即 \hat{U} 和 \tilde{U} 的注意力机制, 对 \hat{U} 和 \tilde{U} 进行通道特征信息加权操作, 得到最终的特征信息映射向量 V , 以加强必要特征信息, 抑制非必要特征信息。

1.2.2 Meta-ACON 激活函数

传统 Res2Net 模型中使用了 ReLU 激活函数, 具有运算速度快、可以有效缓解过拟合的特点, 计算公式如下:

$$f(x) = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (1)$$

由式 (1) 可以看出, 当神经元的输入为负值时, 计算的输出全部为 0。在模型反向传播根据梯度更新参数时, 由于梯度被置 0, 因此该神经元参数可能不会进行激活更新, 从而出现神经元失活现象。为了解决该问题, 本文使用 meta-ACON 激活函数代替 ReLU 激活函数, 其公式如下:

$$Z_{\beta}(x_1, x_2, \dots, x_n) = \frac{\sum_{i=1}^n x_n e^{\beta x_i}}{\sum_{i=1}^n e^{\beta x_i}} \quad (2)$$

式中, β 表示连接系数。当 β 趋于无穷大时, 则有 Z_{β} 趋于最大值, 当 β 趋于零时, 则有 Z_{β} 趋于均值。当切换因子 β 控制激活为非线性或线性时, meta-ACON 能够动态地学习激活函数的线性和非线性, 它可以自适应地学习是否激活神经元, 避免提取到

冗余特征, 如图 4 所示。

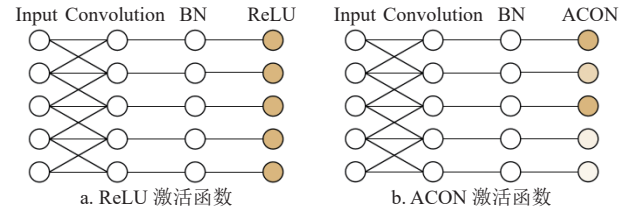


图 4 meta-ACON 激活函数原理图

1.3 基于模型迁移的训练方式

基于模型迁移的训练方式是迁移学习^[22]的常见方法, 它通过已有的预训练模型, 将 1 个或多个源领域上学到的知识应用于目标领域的任务中。具体训练步骤为: 1) 用 ImageNet 数据集训练源域网络模型, 使源域网络的参数通过训练不断优化, 直到源域网络模型的分效果达到最优; 2) 将源域网络训练得到的参数迁移到目标域网络模型, 对目标域模型的参数进行初始化; 3) 用水果数据集对目标域网络进行微调, 从而得到适用于水果图像分类的网络模型。

在微调阶段, 1) 将目标域网络的特征提取部分进行冻结, 并用水果图像对分类部分进行训练, 直到训练集和验证集的准确率和损失函数基本收敛; 2) 对特征提取部分的前端网络进行冻结, 并用水果图像对未冻结的网络层进行训练, 最终达到模型收敛。模型迁移完成后, 用水果图像进行微调, 可以让目标域网络学习到特定的知识, 这样就完成了从一般领域到特定领域的迁移, 最终得到针对水果图像的分类模型, 具体流程如图 5 所示。

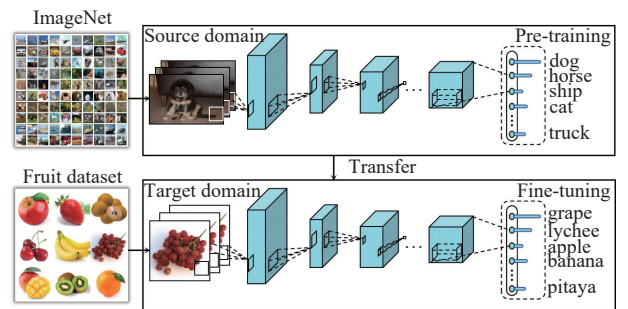


图 5 迁移学习具体流程

2 实验方法

2.1 实验环境

本文所有模型的训练和测试过程在 Microsoft Windows 11 操作系统下进行, CPU 型号为 13th Gen Intel (R) Core (TM) i9-13900KF, GPU 型号为 GeForce RTX 4090, GPU 可用显存为 24 G。

2.2 超参数设置和评价指标

为对比不同模型的性能,对模型训练时的超参数进行设置。通过分析训练过程和分类结果,发现影响算法分类性能的因素主要包括学习率、梯度优化算法和损失函数等,超参数设置如表 1 所示。

表 1 超参数设置

超参数	设置
学习率	0.001
Batch Size	32
学习率衰减	0.000 001
梯度优化算法	SGD
Epoch	100

在图像分类任务中,评价指标包括训练过程中的准确率、精度、召回率、F1 分数和混淆矩阵,模型在测试集的准确率和 loss 的曲线图等可用于衡量模型的性能。计算公式为:

$$ACC = \frac{TP + TN}{TP + FP + FN + TN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (6)$$

式中,ACC 表示准确率,即所有的样本中正确判断出来为正例的个数占所有样本的比;Precision 表示精度,即检测出的正例中真正是正例的个数占检测出的正例的比例;Recall 表示召回率,即所有的样本中被判断为正例的占所有正例的比例;F1 值则表示 Precision 和 Recall 的调和平均数。其中 TP 代表正类判定为正类的图像数目;FP 代表负类判定为正类的图像数目;FN 代表正类判定为负类的图像数目;TN 代表负类判定为负类的图像数目。

2.3 数据集

Fruit-Dataset 数据集是一个用于识别水果的小型数据集。该数据集一共包含 9 个类别(芒果、香蕉、樱桃、草莓、鹰嘴豆、葡萄、猕猴桃、橙子和苹果)的 360 张 RGB 彩色图片,每个类别有 40 张。本文采用数字图像处理算法中的旋转、平移、高斯模糊和对比度变换等数据增强技术对初始样本集进行扩充。扩充的图像模仿了不同天气、不同拍

摄角度、不同相对位置、不同光照强度等条件下的水果图像。采用扩充后的数据训练卷积神经网络能够有效地防止过拟合,并提高模型在复杂条件下的抗干扰性能。Fruit-Dataset 的图片样例如图 6 所示。Fruits-360 数据集是常见的水果数据集,包含 131 类水果和蔬菜的图片共 90 483 张,同一水果的不同品种被存储为不同类别。为满足本文算法对图像分类信息的学习需求,保证随机性和模型对未知数据的泛化能力,以 8:2 的比例随机划分训练集和测试集。



a. Fruit-Dataset 数据集图像



b. Fruits-360 数据集图像

图 6 水果数据集图像

3 结果与讨论

3.1 定量分析

3.1.1 分类性能对比

为了验证本文提出算法的有效性,将本文提出的算法与其他几种具有代表性的图像分类算法进行比较,并通过分类准确率、精确率、召回率和 F1 分数来分析实验结果。表 2 和表 3 分别展示了不同算法在 Fruit-Dataset 数据集和 Fruits-360 数据集上的定量实验结果,其中最优结果被加粗显示。观察表 2 和表 3 中结果可知,本文所提的算法在 Fruit-Dataset 和 Fruits-360 数据集中都取得最佳分类准确率,分别超过次优结果 1.2% 和 1.0%;训练精度相比次优模型分别超过 0.99% 和 0.38%;召回率分别超过次优结果 1.13% 和 0.89%;F1 分数领先次优模型 1.03% 和 1.99%。分类效果都有

提高, 验证了本文提出的算法在对水果图像分类的有效性。

表 2 不同算法在 Fruit-Dataset 数据集上实验结果对比

Model	Accuracy/%	Precision/%	Recall/%	F1
AlexNet ^[23]	77.13	82.48	73.42	0.776 9
VGG-16 ^[7]	91.42	90.88	91.32	0.913 6
ResNet-50 ^[10]	95.81	96.81	93.32	0.953 6
ResNext-50 ^[24]	96.46	96.83	96.89	0.969 3
ResNeSt-50 ^[25]	96.93	96.90	96.67	0.968 5
Res2Net-50 ^[19]	97.12	97.25	97.02	0.971 3
InceptionV2 ^[26]	95.63	96.97	94.79	0.951 3
InceptionV3 ^[27]	95.81	97.03	94.89	0.958 4
MobileNetV2 ^[17]	96.13	97.12	95.47	0.964 5
MobileNetV3 ^[28]	96.24	97.16	95.56	0.968 9
ShuffleNetV1 ^[29]	96.84	97.13	96.83	0.968 7
ShuffleNetV2 ^[12]	96.98	97.18	96.86	0.969 1
EfficientNetV1 ^[30]	97.05	97.15	96.95	0.970 8
EfficientNetV2 ^[31]	97.11	97.24	97.00	0.971 1
本文方法	98.32	98.24	98.15	0.981 6

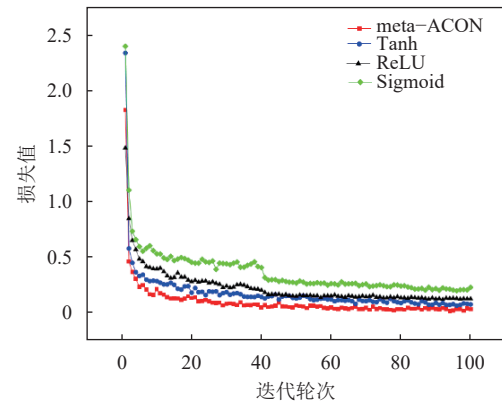
表 3 不同算法在 Fruits-360 数据集上实验结果对比

Model	Accuracy/%	Precision/%	Recall/%	F1
AlexNet ^[23]	77.04	81.45	72.63	0.713 4
VGG-16 ^[7]	90.16	89.52	90.54	0.899 7
ResNet-50 ^[10]	94.35	94.96	92.54	0.921 5
ResNext-50 ^[24]	95.14	95.06	93.89	0.936 1
ResNeSt-50 ^[25]	95.87	96.45	95.67	0.951 3
Res2Net-50 ^[19]	96.14	96.67	96.23	0.946 5
InceptionV2 ^[26]	94.31	96.29	93.54	0.930 4
InceptionV3 ^[27]	94.35	96.37	93.61	0.931 5
MobileNetV2 ^[17]	95.56	96.57	94.67	0.951 7
MobileNetV3 ^[28]	96.13	96.65	95.26	0.950 9
ShuffleNetV1 ^[29]	96.04	96.74	95.93	0.961 3
ShuffleNetV2 ^[12]	96.11	96.81	96.07	0.962 7
EfficientNetV1 ^[30]	95.28	96.42	95.09	0.954 3
EfficientNetV2 ^[31]	95.61	96.64	95.27	0.955 9
本文方法	97.14	97.05	97.12	0.966 4

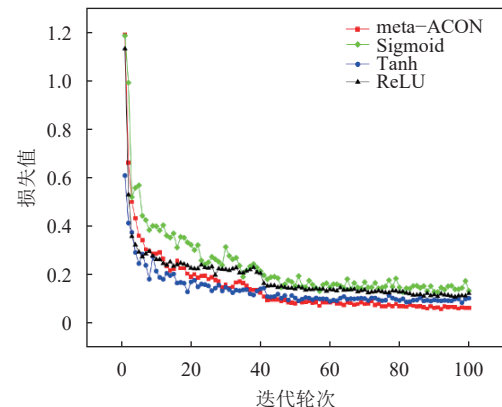
3.1.2 不同激活函数对比实验

为了验证不同激活函数对本文提出算法的影响, 利用验证集进行了对比实验。图 7 是 4 种激活函数在 Fruit-Dataset 和 Fruits-360 数据集所得到的训练集误差曲线, 从图中可以看出, 相对于其他 3 个激活函数, meta-ACON 激活函数误差率最小, 损失函数收敛曲线下降后更平稳, 在前 10 个迭代过程中, 所有激活函数的损失曲线相差不大, 但随着迭代次数增加, 训练集损失值逐渐下降到最低, 并趋于稳定。因此, meta-ACON 函数的误差值较

小、训练的性能更稳定, 在水果图像分类时具有更高的相对精度。



a. Fruit-Dataset 上训练损失曲线

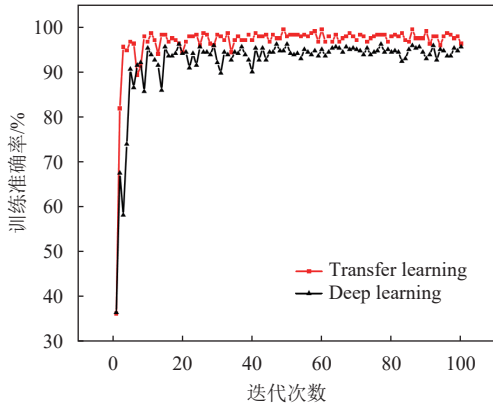


b. Fruits-360 上训练损失曲线

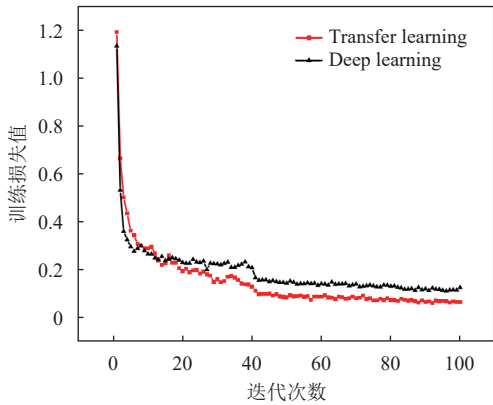
图 7 不同激活函数训练损失对比实验

3.1.3 不同训练方法对比实验

为了验证本文改进的训练方式的有效性, 在验证集上进行对比实验。本文所设计的网络在训练时采用了迁移学习的训练方式, 与以往从头开始训练网络的方法不同, 这种训练方式摆脱了深度学习训练时参数量大导致的耗时问题, 可以缩短网络收敛时间, 并提升训练效率。训练准确率和损失曲线如图 8 所示。由图 8a 可知, 本文提出的算法采用迁移学习的训练方法在训练收敛时准确率达到 98% 左右, 采用深度学习方法在训练收敛时准确率达到 95% 左右; 由图 8b 可知, 在前 20 个迭代过程中, 迁移学习与深度学习的效果相差不大, 但随着迭代次数的增加, 通过采用迁移学习的训练方式, 本文所提的算法能够根据目标任务的数据来微调模型参数, 让模型适应目标任务的特征分布, 在训练收敛时损失下降至 0.05 左右, 采用深度学习方法在训练收敛时损失下降至 0.10 左右。由此可以说明, 迁移学习可以训练出一个具有良好泛化能力的模型。



a. 训练准确率曲线



b. 训练损失曲线

图 8 不同训练方法的训练准确率和训练损失对比

3.1.4 消融实验

为了验证本文提出的算法对水果分类的有效性，需要对模型进行消融实验。实验中，分别使用 Res2Net50、多尺度融合注意力和 meta-ACON 激活函数在 Fruit-Dataset 和 Fruits-360 数据集上训练，并采用准确率和训练参数量作为评价指标，结果如表 4 所示。从表 4 中可以看出，本文所提方法中几个部分对最终结果都有贡献，在基础模型上添加注意力模块可以将准确率分别提升 0.34% 和 0.09%，在基础模型上改变激活函数可以将准确率分别提升 0.84% 和 0.44%，在基础模型上同时添加注意力和改变激活函数可以将准确率分别提升 1.13% 和 0.60%。这也证明了本文改进的网络结构中每个部分的必要性。

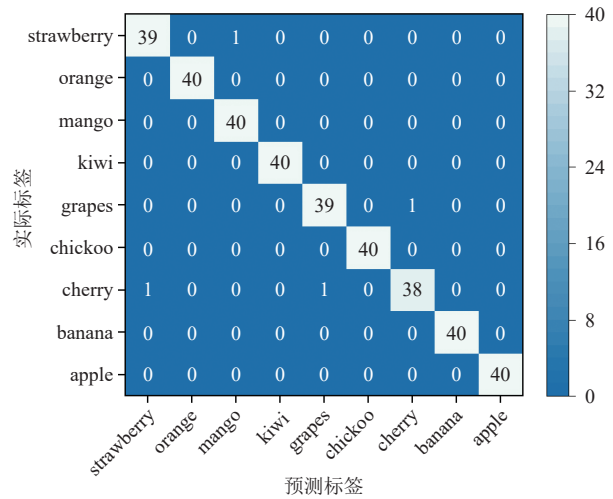
表 4 本文算法在 Fruit-Dataset 和 Fruits-360 数据集上的消融实验

	多尺度融合			准确率/%	
	注意力	meta-ACON	参数量/M	Fruit-Dataset	Fruits-360
Res2Net50	×	×	23.67	97.19	96.54
	√	×	24.12	97.53	96.43
	×	√	23.85	98.03	96.98
	√	√	24.36	98.32	97.14

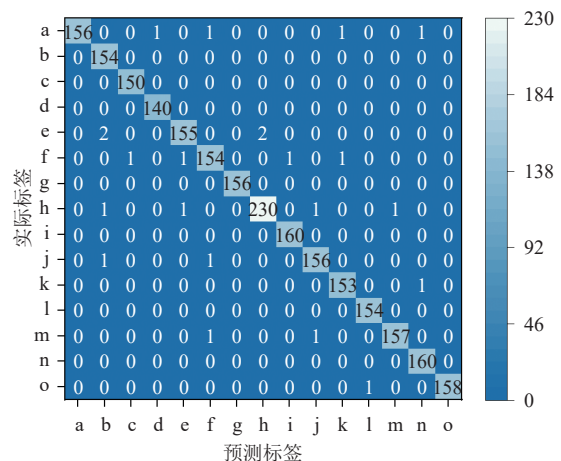
3.2 定性分析

3.2.1 混淆矩阵分析

为了进一步验证本文模型在 Fruit-Dataset 数据集和 Fruits-360 数据集上的分类性能，分别在 Fruit-Dataset 上进行 9 种不同类别水果的对比和在 Fruits-360 数据集上选取 15 种不同类型的苹果进行对比，结果如图 9 所示。在图 9a 中，横纵坐标代表 9 个类别的水果。从混淆矩阵可以看出，本文方法用于识别芒果、香蕉、猕猴桃、鹰嘴豆、橙子和苹果都能达到 100% 的分类准确率，但在识别草莓、樱桃和葡萄时出现个别差错，可能是因为存在树叶等严重背景干扰。在图 9b 中，横纵坐标分别代表 15 种不同类型的苹果，本文算法通过克服类内差异，在 h 类型苹果上最高达到 98.29% 的分类准确率。



a. 9 种不同种类水果



b. 15 种不同类型苹果

图 9 验证集混淆矩阵

3.2.2 可视化分析

为了更直观地验证本文算法对于水果图像分类的有效性，本文选取了 Fruit-Dataset 和 Fruits-360

验证集的图片在空间位置上与卷积层的特征映射进行对应, 利用类激活可视化方法, 将模型最后一个卷积层在水果上的注意力可视化, 更清晰地说明注意力机制在模型判别时所关注的水果特征。如图 10 所示, 以 Fruit-Dataset 数据集中的苹果和香蕉类别图像为例, 传统 CNN 网络特征激活区域范围较大,

导致图像中一些无关的背景信息干扰后续分类。然而, 本文使用注意力后的模型对于水果判别性区域识别更精准, 不会过多关注背景信息, 减少了背景噪声干扰, 模型的代表能力更强。图 11 为 Fruits-360 数据集中芒果和橘子类别图像, 本文方法相比传统 CNN, 也能更精准地识别更有判别性的特征。

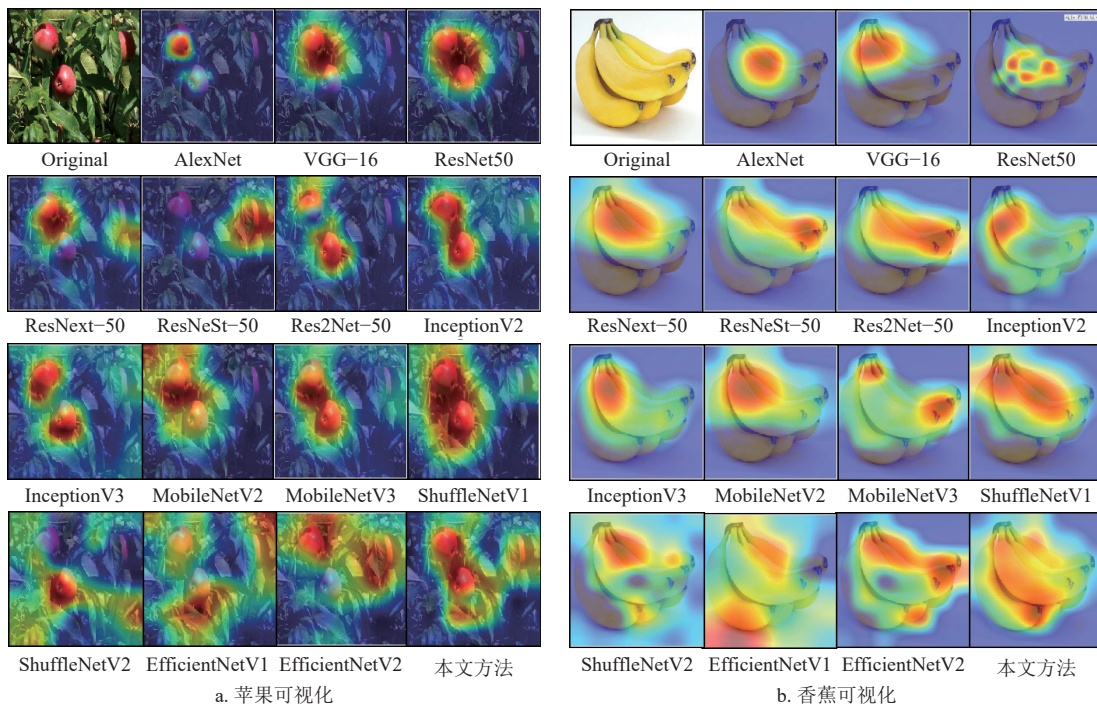


图 10 不同方法在 Fruit-Dataset 数据集上的可视化对比

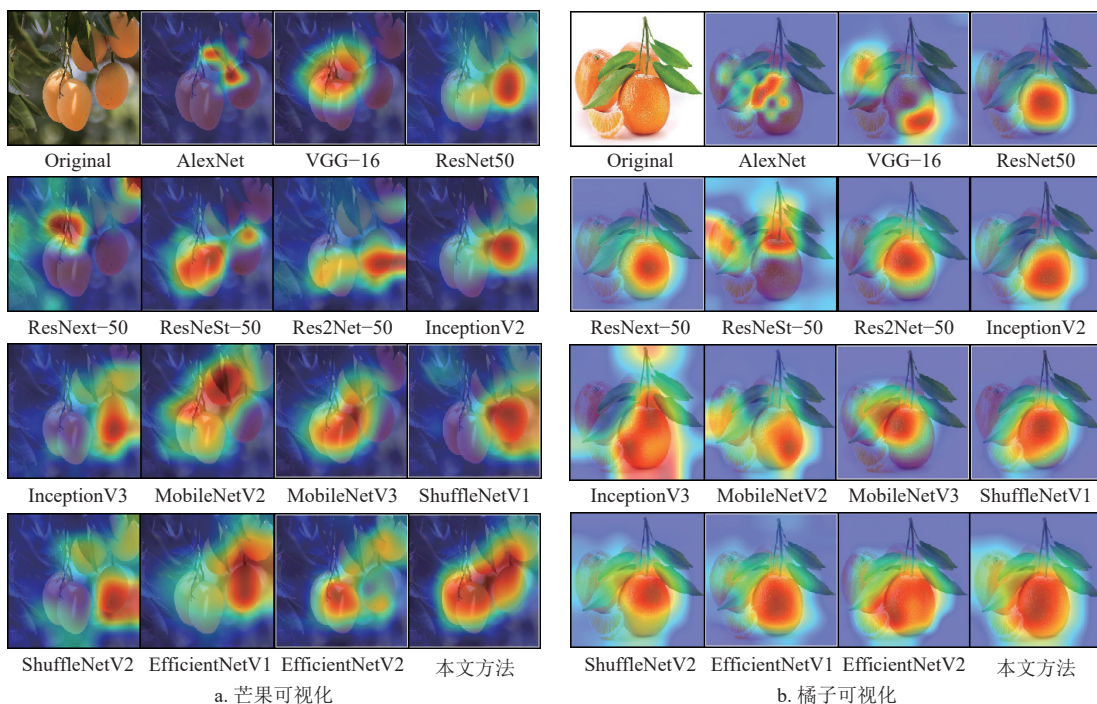


图 11 不同方法在 Fruits-360 数据集上的可视化对比

4 结束语

本文提出了一种基于改进 Res2Net 和迁移学习的图像分类算法。首先,在特征提取网络设计方面,提出新的动态多尺度融合注意力机制,在 Res2Net 基础上,通过引入该注意力机制来动态生成卷积核,以提高图像细粒度特征信息表达能力,加强对图像关键特征的提取。通过 meta-ACON 激活函数优化 ReLU 激活函数,动态地学习激活函数的线性和非线性,提高模型的非线性表达能力。其次,在训练方式方面,本文采用基于模型迁移的网络训练方式,可以将从源域学习到的知识迁移到目标域,辅助目标域任务优化,缓解小数据集上数据样本不足的问题。通过在公开数据集 Fruit-Dataset 和 Fruits-360 上进行实验,本文所提算法测试准确率分别达 98.32% 和 97.14%,相比于其他算法获得了更好的效果。

在未来的研究中,将迁移学习的理念与深度模型的方法进行结合,可以适当增加输入数据集的大小以提升精度,或者通过优势组合,继续优化模型,进一步提高对特征提取感受野的范围大小等。

参考文献

- [1] 刘祎,刘宇航,颜溶樑,等.基于细节增强的多能 DR 图像融合网络[J].*电子科技大学学报*,2023,52(3):379-389.
LIU Y, LIU Y H, YAN R B, et al. Detail enhancement-based fusion network for multi-energy digital radiography images[J]. *Journal of University of Electronic Science and Technology of China*, 2023, 52(3): 379-389.
- [2] GULZAR Y. Fruit image classification model based on MobileNetV2 with deep transfer learning technique[J]. *Sustainability*, 2023, 15(3): 1906.
- [3] SINGH GILL H, IBRAHIM KHALAF O, ALOTAIBI Y, et al. Fruit image classification using deep learning[J]. *Computers, Materials & Continua*, 2022, 71(3): 5135-5150.
- [4] REN S, CAO X, WEI Y, et al. Global refinement of random forest[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.]: IEEE, 2015: 723-730.
- [5] ABEYWICKRAMA T, CHEEMA M A, TANIAR D. K-nearest neighbors on road networks: A journey in experimentation and in-memory implementation[J]. *Proceedings of the VLDB Endowment*, 2016, 9(6): 492-503.
- [6] NI J, GAO J, DENG L, et al. Monitoring the change process of banana freshness by GoogLeNet[J]. *IEEE Access*, 2020, 8: 228369-228376.
- [7] HOSSAIN M S, AL-HAMMADI M, MUHAMMAD G. Automatic fruit classification using deep learning for industrial applications[J]. *IEEE Transactions on Industrial Informatics*, 2019, 15(2): 1027-1034.
- [8] FAISAL M, ALBOGAMY F, ELGIBREEN H, et al. Deep learning and computer vision for estimating date fruits type, maturity level, and weight[J]. *IEEE Access*, 2020, 8: 206770-206782.
- [9] PONCE J M, AQUINO A, ANDÚJAR J M. Olive-fruit variety classification by means of image processing and convolutional neural networks[J]. *IEEE Access*, 2019, 7: 147629-147641.
- [10] ZAGORUYKO S, KOMODAKIS N. Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer [EB/OL]. [2023-09-29]. <https://arxiv.org/abs/1612.03928>.
- [11] XUE G, LIU S F, MA Y C. A hybrid deep learning-based fruit classification using attention model and convolution autoencoder[J]. *Complex & Intelligent Systems*, 2023, 9(3): 2209-2219.
- [12] GHOSH S, MONDAL M J, SEN S, et al. A novel approach to detect and classify fruits using ShuffleNet V2[C]//*Proceedings of the IEEE Applied Signal Processing Conference*. Kolkata: IEEE, 2020: 163-167.
- [13] QIAO J, WANG G, LI W, et al. An adaptive deep Q-learning strategy for handwritten digit recognition[J]. *Neural Networks: The Official Journal of the International Neural Network Society*, 2018, 107: 61-71.
- [14] GLOROT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks[J]. *Journal of Machine Learning Research*, 2011, 15: 315-323.
- [15] MA N, ZHANG X, LIU M, et al. Activate or not: Learning customized activation[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. New York: IEEE, 2021: 8032-8042.
- [16] CHEN T Q, GOODFELLOW I, SHLENS J. Net2Net: Accelerating learning via knowledge transfer[EB/OL]. (2015-11-18)[2023-9-29]. <https://arxiv.org/abs/1511.05641>.
- [17] YASIR H, SHARYAR W, ARJUMAND B S, et al. Smart seed classification system based on MobileNetV2 architecture[C]//*Proceedings of the 2022 2nd International Conference on Computing and Information Technology*. Cox's Bazar: IEEE, 2022: 217-222.
- [18] KAZI A, PANDA S P. Determining the freshness of fruits in the food industry by image classification using transfer learning[J]. *Multimedia Tools and Applications*, 2022, 81(6): 7611-7624.
- [19] 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*, 2021, 43(2): 652-662.
- [20] LI X, WANG W, HU X, et al. Selective kernel networks[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Long Beach:

- IEEE, 2019: 510-519.
- [21] GLOROT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks[C]//Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. Fort Lauderdale: PMLR, 2011, 15: 315-323.
- [22] HAN D M, LIU Q G, FAN W G. A new image classification method using CNN transfer learning and web data augmentation[J]. *Expert Systems with Applications*, 2018, 95: 43-56.
- [23] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2012, 60: 84-90.
- [24] XIE S N, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2017: 5987-5995.
- [25] ZHANG H, WU C R, ZHANG Z Y, et al. ResNeSt: Split-attention networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. New York: IEEE, 2022: 2735-2745.
- [26] 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: ACM, 2015: 448-456.
- [27] SZEGEDY C, VANHOUCHE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 2818-2826.
- [28] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3[C]//Proceedings of the IEEE International Conference on Computer Vision. Seoul: IEEE, 2019: 1314-1324.
- [29] ZHANG X, ZHOU X, LIN M, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 6848-6856.
- [30] HUMPHREY E J, BELLO J P. Rethinking automatic chord recognition with convolutional neural networks[C]//Proceedings of the 11th International Conference on Machine Learning and Applications. New York: IEEE, 2012: 357-362.
- [31] TAN M, LE Q. EfficientNetV2: Smaller models and faster training[C]//Proceedings of the International Conference on Machine Learning. Vienna: ACM, 2021: 10096-10106.

编辑 张莉