

黄晓文, 胡峻峰, 朱昊, 等. 基于多尺度特征融合的食用葡萄细粒度分类方法 [J]. 智能计算机与应用, 2024, 14(4): 146-150.
DOI: 10.20169/j.issn.2095-2163.240422

基于多尺度特征融合的食用葡萄细粒度分类方法

黄晓文¹, 胡峻峰¹, 朱昊¹, 李柏聪¹, 贾梦实²

(1 东北林业大学 计算机与控制工程学院, 哈尔滨 150040; 2 东北林业大学 家居与艺术设计学院, 哈尔滨 150040)

摘要: 中国葡萄种植规模逐年扩大, 但由于混种、串粉等导致了葡萄品种混杂, 葡萄的分类筛选工作成为限制农业生产效率的重要因素。为了提高生产中葡萄品种分类精度, 本文基于计算机视觉技术和深度学习理论, 设计了多尺度特征融合网络 (Multi-Scale Feature Fusion Network, MSFF-Net)。首先通过混合注意力模块改进 ResNet34 残差模块, 提高模块关键特征提取能力; 然后采用双线性结构, 改进网络提取葡萄图像的二阶特征; 最后通过融合二阶特征和交叉双线性特征, 建立了一种基于多尺度特征融合网络的食用葡萄分类模型。在葡萄分类试验中, 达到了 96.97% 的识别精度, 相较于对照模型, 取得了更高的精度。

关键词: 卷积神经网络; 葡萄; 注意力机制; 细粒度分类; 特征融合

中图分类号: TP391.4; S663.1

文献标志码: A

文章编号: 2095-2163(2024)04-0146-05

Fine-grained classification method of edible grapes based on multi-scale feature fusion

HUANG Xiaowen¹, HU Junfeng¹, ZHU Hao¹, LI Baicong¹, JIA Mengshi²

(1 College of Computer and Control Engineering, Northeast Forestry University, Harbin 150040, China;

2 College of Home and Art Design, Northeast Forestry University, Harbin 150040, China)

Abstract: The scale of grape planting in China has been expanding year by year, but the mixing and cross-pollination of grape varieties have led to a mixture of grape varieties. The classification and screening of grape varieties have become an important factor limiting agricultural production efficiency. In order to improve the accuracy of grape variety classification in production, this paper proposes a Multi-Scale Feature Fusion Network (MSFF-Net) based on computer vision technology and deep learning theory. Firstly, the mixed attention module is used to improve the residual module of ResNet34, enhancing the key feature extraction ability of the module. Then, a bilinear structure is adopted to extract second-order features of grape images from the network. Finally, a model for classifying edible grapes based on MSFF-Net is established by fusing second-order features and cross-bilinear features. In grape classification experiments, an accuracy of 93.66% was achieved, which is higher than that of the control model.

Key words: Convolutional Neural Network; grape; attention mechanism; fine-grained classification; feature fusion

0 引言

中国是世界上最大的食用葡萄种植国, 截至 2020 年, 中国葡萄种植面积超过 73.07 万 hm^2 , 且近年来中国葡萄种植规模仍在持续增加^[1]。但由于没有规范标准的管理, 因此出现混种、串粉等问题, 影响了采摘效率^[2]。伴随机器学习的发展, 越来越多科研人员将其应用在农业领域, 不仅提高了农作物分类的准确率^[3], 还极大程度地提高了劳动生产率, 解放了生产力^[4]。

传统机器学习方法受早期设备算力的限制, 无法并行处理大量的数据^[5], 因此需要拥有专业知识的工作人员手动处理数据^[6]、筛选特征^[7]、设计网络结构^[8]。赵杰文等^[9]选定径向基函数作为支持向量机 (SVM) 核函数, 建立高精度的苹果近红外光谱分类模型, 对不同品种苹果分类模型的回判识别率和预测识别率均达到 100%。张伏等^[10]提出了基于 SVM 算法的千禧番茄光谱特征识别模型, 对测试集番茄分类准确率达到 81.25%。深度学习网络相较传统机器学习, 省去了手动筛选特征、设计网络

作者简介: 黄晓文 (1995-), 男, 硕士研究生, 主要研究方向: 模式识别。

通讯作者: 胡峻峰 (1980-), 男, 博士, 副教授, 硕士生导师, 主要研究方向: 模式识别。Email: nefuhjf@126.com

收稿日期: 2023-12-06

的过程,不仅节约了检测和分类的时间,而且模型精度和泛化性得到了极大幅度的提升。岳有军等^[11]改进了 VGG-16 网络,对苹果病害外观识别准确率达到了 99.25%。

计算机视觉算法的应用,大大提高了农产品分类的效率与准确率^[12],但将其应用于葡萄分类的研究并不多。因此,本研究针对食用葡萄的分类问题,提出了改进的多尺度特征融合网络 (Multi-Scale Feature Fusion Network, MSFF-Net)。该网络基于 ResNet34,以双线性特征融合的方式,提高了模型的细粒度特征提取能力,并改进了模型的注意力机制,提高模型的葡萄分类效果。

1 材料和方法

1.1 特征融合分类网络构建

2015 年,He 等^[10,13]提出残差网络 (Residual Network, ResNet)。通过在残差块中引入捷径,有效解决深层网络的退化问题。其中 ResNet34 为 ResNet 系列网络中,网络层数比较适中的一个版本,其结构如图 1 所示。

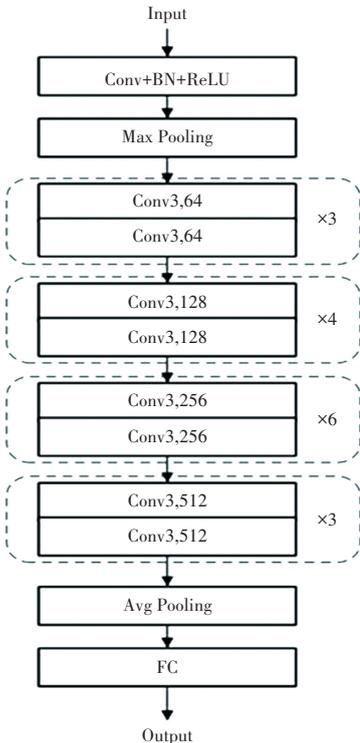


图 1 ResNet34 结构图

Fig. 1 Structure of ResNet34

ResNet34 具有收敛速度快、精度高等特点,能够有效地避免训练过程中的过拟合现象,在计算机视觉领域应用广泛。

1.2 改进混合注意力机制

在网络中添加注意力机制能够使其更关注输入图像的细节信息,有效提高模型的辨识能力。混合注意力模块 (Convolutional Block Attention Module, CBAM)^[14]通过将特征图先后通过通道注意力模块 CAM^[15]和空间注意力模块 SAM^[16]来获取包含了通道及空间维度的混合注意力,其结构如图 2 所示。在 CAM 与 SAM 采用串联连接的情况下,由于 SAM 模块接收到的是经由 CAM 通道注意力权重修饰过的特征图,导致后者不能有效地提取到原始图像的空间注意力特征,使分类效果不够理想。

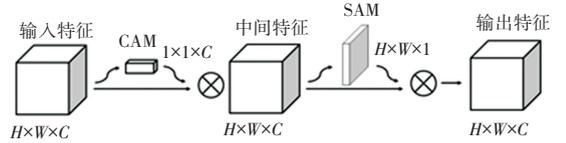


图 2 CBAM 模块

Fig. 2 CBAM Module

并联混合注意力模块 (PACBAM) 为避免两个阶段的注意力机制互相影响,采用 CAM 和 SAM 并行连接的方式,两个子模块分别学习原始输入特征图,同时产生通道阈和空间阈对应的权重,避免了串行连接所带来的识别干扰问题。PACBAM 模块的整体结构如图 3 所示。

PACBAM 的计算过程如下:

步骤 1 输入特征图 F , 分别经过空间维度的池化操作,得到 F_{CAvg} 和 F_{CMax} , 并分别送入全连接层 FC_1 、 FC_2 进行降维,并将得到的非线性特征进行相加,得到通道注意力特征图 A_c :

$$A_c = FC_2(FC_1(F_{CMax})) + FC_2(FC_1(F_{CAvg})) \quad (1)$$

步骤 2 输入特征图 F 分别经过通道维度的池化操作,得到通道数为 1 的特征图 F_{SAvg} 和 F_{SMax} , 之后进行通道拼接,再经过卷积降维后得到空间注意力特征图 A_s :

$$A_s = Conv(F_{SAvg} + F_{SMax}) \quad (2)$$

步骤 3 将 A_c 、 A_s 直接与原始输入特征图 F 进行元素乘法运算,得到 PACBAM 模块输出的加权特征图 F_{PA} :

$$F_{PA} = A_s \otimes A_c \otimes F \quad (3)$$

本文通过在 ResNet34 网络的残差块中加入 PACBAM 模块,得到了改进的混合注意力网络 (ResNet34+PACBAM)。ResNet34+PACBAM 的网络结构如图 4 所示。

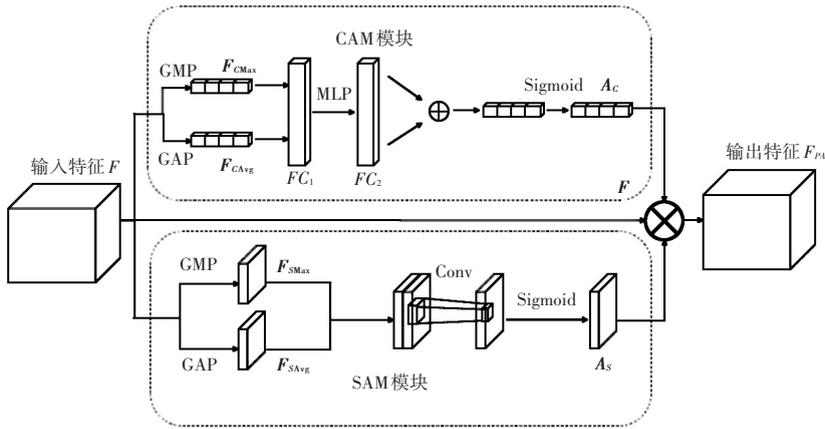


图 3 PACBAM 模块

Fig. 3 PACBAM Module

双线性融合操作。将网络左侧的加性融合特征 F_1 与 Stream A 中输出特征 f_a 进行外积相乘的操作, 得到交叉双线性特征 B_1 。同理得到 Stream B 的交叉双线性特征 B_2 。计算公式如下:

$$\begin{cases} B_1(F_1, f_a) = F_1^T \otimes f_a \\ B_2(F_2, f_b) = F_2^T \otimes f_b \\ B_0(f_a, f_b) = f_a^T \otimes f_b \end{cases} \quad (5)$$

将上述得到的交叉双线性特征向量送入级联融合层。Concat 操作能够加强不同通道之间的特征联系, 实现特征信息互补。本文利用级联融合, 将得到的特征 B_1, B_2 与 B_0 组不同双线性向量进行通道拼接得到特征 B 。

$$B = [B_0, B_1, B_2] \quad (6)$$

将拼接后的特征向量 B 降维后, 送入全连接层, 最终由分类器输出分类结果。

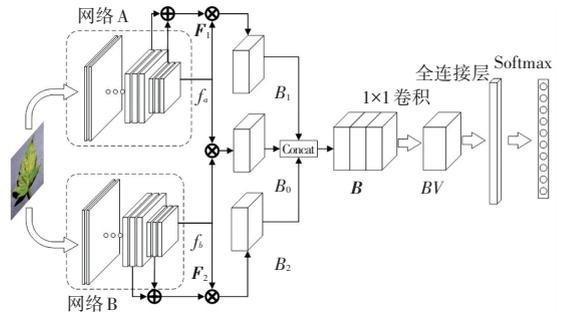


图 5 MSFF-Net 的结构

Fig. 5 Structure of MSFF-Net

2 试验与分析

2.1 食用葡萄类别数据集

2.1.1 数据集的采集

葡萄类别数据集采集自山东省地区葡萄采摘

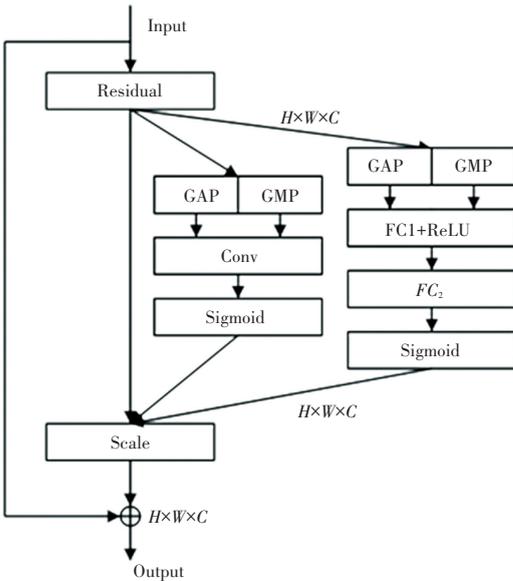


图 4 ResNet34+PACBAM 模块

Fig. 4 ResNet34+PACBAM Module

1.3 多尺度特征融合网络

本文基于改进混合注意力机制的残差神经网络 (ResNet34+PACBAM), 针对不同层次特征的深度通道拼接操作, 有效融合不同层次多尺度的特征信息^[17], 设计了多尺度特征融合网络 MSFF-Net。

如图 5 所示, 首先提取网络 A 和网络 B 中不同尺度的特征向量进行融合。将卷积网络 Stream A 的输出特征 f_1 和 f_2 经过 Add 操作, 得到加性融合特征 F_1 ; 将卷积网络 Stream B 中输出特征 f_3 和 f_4 经过 Add 操作, 得到融合特征 F_2 。计算公式如下:

$$\begin{cases} F_1(f_1, f_2) = f_1 + f_2 \\ F_2(f_3, f_4) = f_3 + f_4 \end{cases} \quad (4)$$

分别对上一阶段提取的特征向量 F_1 和 F_2 进行

园,其中包括红巴拉多(Balado Red)、金优 II 号(Jinyou II)、巨峰(Kyoho)、玫瑰香(Muscat)、妮娜皇后(Queen Nina)、阳光玫瑰(Shine Muscat)及甜蜜蓝宝石(Sweet Sapphire)等 7 种葡萄类别。数据集使用专业相机采集,共包含了 7 990 张分辨率为 4 480×2 016 像素的葡萄图像。数据集包含了不同成熟期、不同光照情况、以及遮挡的葡萄样本。

2.1.2 数据预处理

通过数据增强对图像进行预处理,可以提高数据集质量。首先从图像中裁切出识别目标葡萄图像。为了提高模型的泛化性及抗干扰性,采用镜面翻转、旋转裁切、添加高斯噪声、Cutout 等方法,对数据集进行增强。增强效果如图 6 所示。

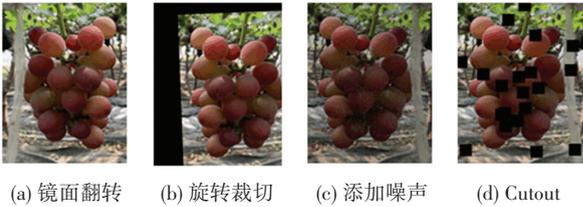


图 6 数据增强效果展示

Fig. 6 Display of data augmentation effects

经过数据增强后,最终得到 23 970 张葡萄图像,以 8 : 2 的比例将数据集进行划分,其中训练集 6 394 张、测试集 1 596 张。各类别数据分布见表 1。

表 1 数据集基本信息

Table 1 Basic information of a dataset

类别	训练集	测试集	合计
红巴拉多	2 760	690	3 450
金优 II 号	4 944	1 233	6 177
巨峰	3 639	909	4 548
玫瑰香	1 212	303	1 515
妮娜皇后	2 430	606	3 036
阳光玫瑰	1 134	282	1 416
甜蜜蓝宝石	3 063	765	3 828
合计	19 182	4 788	23 970

2.2 试验环境及参数设定

本试验环境设置:CPU 为 i7 13 700 k,显卡为 NVIDIA GeForce RTX 4090,服务器内存为 128 GB,GPU 显存为 24 GB,操作系统版本为 Ubuntu 22.04.2 LTS。搭建卷积神经网络模型时,使用 Python3.10.10 深度学习框架,CUDA 版本为 v12.1。模型参数见表 2。

表 2 模型参数设置

Table 2 Model parameter settings

参数名称	参数值
图像尺寸	224×224
批尺寸	32
学习率	0.001
优化器	SGD
训练周期	150

2.3 评价指标

本文采用分类准确率(Accuracy, Acc)作为模型性能的评价指标,计算公式如下:

$$Acc = \frac{N_R}{N} \times 100\% \quad (7)$$

其中, N_R 为食用葡萄测试集中正确分类的样本数, N 为测试集中样本总数。

2.4 试验结果及分析

2.4.1 模型改进对比试验与分析

为了验证本文提出的多尺度特征融合网络在食用葡萄类别分类任务中的有效性,将多尺度特征融合网络 MSFF-Net 分别与 ResNet34、改进混合注意力网络 ResNet34+CBAM、ResNet34+PACBAM 进行比较。不同模型的训练准确率曲线如图 7 所示,试验的结果数据见表 3。

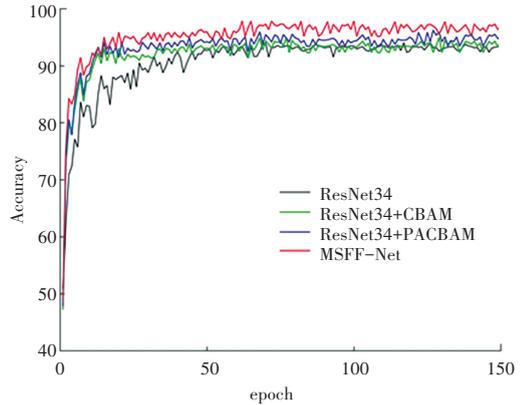


图 7 不同模型训练曲线

Fig. 7 Training curves for different models

表 3 不同模型分类准确率

Table 3 Accuracy of different models

模型	Acc/ %
ResNet34	93.38
ResNet34+CBAM	94.27
ResNet34+PACBAM	95.02
MSFF-Net	96.97

在葡萄分类准确率方面,MSFF-Net 模型充分

利用残差网络不同层次的特征映射,提高网络对特征信息的利用,分类准确率为 96.97%,相比 ResNet34 提高 3.59%,相比改进混合注意力网络的 ResNet34 + CBAM 提高 2.7%,相比 ResNet34 + PACBAM 提高 1.95%。因此,本文提出的 MSFF-Net 模型对食用葡萄分类效果优于其他对比模型。

2.4.2 混淆矩阵分析

通过混淆矩阵,可以分析网络对具体类别的识别效果,分析误差产生原因。MSFF-Net 对食用葡萄分类的混淆矩阵如图 8 所示。

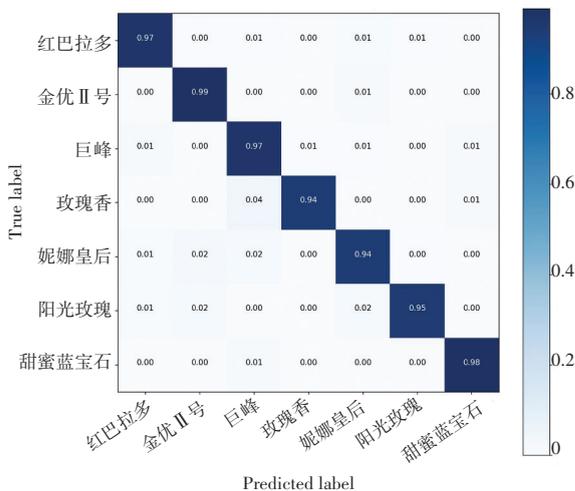


图 8 MSFF-Net 混淆矩阵

Fig. 8 Confusion matrix of MSFF-Net

MSFF-Net 对红巴拉多、金优 II 号、巨峰、甜蜜蓝宝石的分类效果较好,准确率均在 97% 以上;由于巨峰与玫瑰香外形较为近似,有 4% 的玫瑰香图像被识别为巨峰;妮娜皇后、阳光玫瑰识别效果相对较差,分别为 94%、95%。

3 结束语

本文提出了一种通过融合不同尺度的特征信息,提高番分类效果的网络模型 MSFF-Net,该模型以改进混合注意力网络 ResNet34 + PACBAM 为基础,采用多尺度特征融合方法获取更具判别性的特征信息,实现了对细粒度特征提取效果的增强。经在食用葡萄类别数据集上进行试验,验证了本模型的有效性。结果证明,该模型有效地融合了不同网络层级的特征信息,在拓宽了模型感受野的同时,提

高了模型细粒度特征提取能力,在葡萄种类识别任务上具有出色的表现。在未来的工作中,将进一步优化网络,促进机器视觉在葡萄生产中的实际应用。

参考文献

- [1] 汪丽华. 鲜食葡萄优质丰产栽培[J]. 云南农业, 2022, 400(5): 61-64.
- [2] 胡心意, 戴芬, 赵学平, 等. 鲜食葡萄全产业链标准化探索与实践[J]. 农产品质量与安全, 2023, 126(6): 10-13.
- [3] 蒙贺伟, 李成松, 坎朵, 等. 新疆酿酒葡萄振动采收装置设计与试验[J]. 农机化研究, 2017, 39(10): 42-45, 62.
- [4] LUO Lufeng, TANG Yunchao, ZOU Xiangjun, et al. Vision-based extraction of spatial information in grape clusters for harvesting robots[J]. Biosystems Engineering, 2016, 151: 90-104.
- [5] LIU S, WHITTY M. Automatic grape bunch detection in vineyards with an SVM classifier[J]. Journal of Applied Logic, 2015, 13(4): 643-653.
- [6] 陈英, 李伟, 张俊雄. 基于图像轮廓分析的堆叠葡萄果粒尺寸检测[J]. 农业机械学报, 2011, 42(8): 168-172.
- [7] 侯森, 刘陈帅, 王保荣. 基于灰色关联分析的 SOM 神经网络在葡萄分类中的应用[J]. 智能计算机与应用, 2017, 7(6): 42-46.
- [8] 张焱伟. 计算机视觉系统在番茄品质识别与分类中的研究[D]. 保定: 河北农业大学, 2005.
- [9] 赵杰文, 呼怀平, 邹小波. 支持向量机在苹果分类的近红外光谱模型中的应用[J]. 农业工程学报, 2007(4): 149-152.
- [10] 张伏, 王新月, 崔夏华, 等. 可见/近红外光谱结合 GWO-SVM 对千禧番茄的分类鉴别[J]. 光谱学与光谱分析, 2022, 42(10): 3291-3297.
- [11] 岳有军, 田博凯, 王红君, 等. 改进 VGG 模型在苹果外观分类中的应用[J]. 科学技术与工程, 2020, 20(19): 7787-7792.
- [12] 李惠鹏, 李长勇, 李贵宾, 等. 基于深度学习的多品种鲜食葡萄采摘点定位[J]. 中国农机化学报, 2022, 43(12): 155-161.
- [13] HE K M, ZHANG X Y, 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.
- [14] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module [C]// Proceedings of the 15th European Conference. Munich: Springer. 2018: 3-19.
- [15] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7132-7141.
- [16] ZHU X, CHENG D, ZHANG Z, et al. An empirical study of spatial attention mechanisms in deep networks [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 6688-6697.
- [17] LIN Z, MU S, SHI A, et al. A novel method of maize leaf disease image identification based on a multichannel convolutional neural network[J]. Transactions of the ASABE. 2018, 61(5): 1461-1474.