doi:10.6041/j.issn.1000-1298.2024.04.021

基于多器官特征融合的枣品种识别方法

许　楠[1]　苑迎春[1]　雷　浩[1]　孟　惜[2]　何振学[1]

(1.河北农业大学信息科学与技术学院,保定 071001;2.石家庄学院未来信息技术学院,石家庄 050035)

摘要:针对自然场景下的枣品种识别问题,以枣果为研究对象的机器视觉技术已成为枣品种精准识别的主流方法之一。针对枣品种存在类间差异小、类内差异大的问题,提出了一种基于多器官特征融合的枣品种识别方法。首先利用 YOLO v3 检测算法将采集的自然场景图像中的枣果和叶片器官分割提取,提出了基于笛卡尔乘积构建两器官组合对的枣品种多样本数据集,然后基于 EfficientNetV2 网络模型,设计了能够充分学习两器官特征相关性的融合策略来提升模型性能,引入了逐步迁移训练方式以提升枣品种识别效率。最后,在构建的包含 20 个枣品种数据集上进行了大量实验,得到 97.04% 的识别准确率,明显优于现有研究结果,并且在训练时间和收敛速度上,本方法也有一定提升。结果表明该方法能够有效融合枣品种枣果和叶片器官的特征信息,可为其他品种识别研究提供参考。

关键词:枣品种识别;笛卡尔乘积;特征融合;迁移学习;YOLO v3

中图分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1000-1298(2024)04-0213-08　　**OSID**:

Jujube Variety Recognition Method Based on Multi-organ Feature Fusion

XU Nan[1]　YUAN Yingchun[1]　LEI Hao[1]　MENG Xi[2]　HE Zhenxue[1]

(1. *College of Information Science and Technology*, *Hebei Agricultural University*, *Baoding* 071001, *China*

2. *College of Future Information Technology*, *Shijiazhuang University*, *Shijiazhuang* 050035, *China*)

Abstract: Aiming at the problem of jujube variety identification in natural scenes, machine vision technology with jujube fruit as the research object has become one of the mainstream methods for accurate identification of jujube varieties. However, due to the small inter-class difference and large intra-class difference of jujube varieties, it is difficult for a single organ to fully express the different characteristics of jujube varieties. A method of jujube varieties recognition based on multi-organ feature fusion was proposed. Firstly, the YOLO v3 detection algorithm was used to segment and extract the jujube fruit and leaf organs in the collected natural scene images, and a multi-sample dataset of jujube varieties based on Cartesian product was proposed to construct two organ combination pairs, and then based on the EfficientNetV2 network model, a fusion strategy that can fully learn the correlation between the characteristics of the two organs was designed to improve the model performance, and a stepwise transfer training method was introduced to improve the recognition efficiency of jujube varieties. Finally, a large number of experiments were carried out on the constructed dataset containing 20 jujube varieties, and the recognition accuracy of 97.04% was obtained, which was significantly better than that of the existing research results, and the training time and convergence speed of the proposed method were also improved. The results showed that this method can effectively integrate the characteristic information of jujube fruit and leaf organs of jujube cultivars, which can provide valuable reference for other variety identification research.

Key words: jujube variety recognition; Cartesian product; feature fusion; transfer learning; YOLO v3

收稿日期:2023-08-25　修回日期:2023-09-28

基金项目:国家自然科学基金项目(62102130)和河北省自然科学基金项目(F2020204003)

作者简介:许楠(2000—),男,硕士生,主要从事图像处理研究,E-mail:502833816@qq.com

通信作者:苑迎春(1970—),女,教授,博士,主要从事智能信息处理与大数据技术研究,E-mail:nd_hd_yyc@163.com

0 引言

我国是枣产量第一大国,据相关研究统计有900余种[1-2]。基于机器视觉技术实现自动、精准识别枣品种是实现枣产业自动化管理的前提,对枣果的病害防治、产量预测、成熟度预测、自动采摘、分选加工以及品质分级等环节都具有重要的意义。

目前,在枣品种识别领域,通常将枣果作为识别枣品种的主要器官。文献[3]在高光谱技术基础上,结合自适应重加权采样和支持向量机模型,实现了3种红枣的分类,识别准确率达到91.2%。文献[4]通过Stacking模型集成卷积神经网络VGG16、ResNet50和Densenet121,实现了5个红枣品种的分类,识别准确率达到92.38%。上述研究均取得了较高的识别率,但考虑的枣品种数量较少。文献[5]考虑自然场景下20个枣品种的识别研究,在VGG16网络基础上,同时从形状和细粒度两个角度学习枣果图像特征,并采用Coupled cluster loss和Focal loss两种损失函数对模型优化,在20种枣数据集上识别准确率为84.16%。文献[6]同样考虑20个枣品种识别问题,在MobileNetV3网络基础上,采用多任务学习方法同时对枣果进行形状分类和细粒度分类,并将枣果的细粒度特征融合到形状分类任务中,通过Reduce focal loss来进行优化,而整个枣品种分类任务采用Arcface和Softmax loss损失函数联合进行优化。该方法在20种枣的数据集上识别准确率为91.79%,有效提升了枣品种识别准确率。

针对大量枣品种间存在的类间相似性和类内差异性问题,仅考虑枣果单一器官特征构建识别模型很难提升大量品种的识别准确率。近年来,特征融合技术在农作物识别领域被广泛应用,一些学者在苹果[7-8]、油桃[9]、柑橘[10]、稻谷[11]、大豆[12]等农作物的识别研究表明,特征融合方法能够综合利用多种图像特征,实现多特征优势互补,获得更加稳定和准确的识别结果。随着卷积神经网络等深度学习技术的发展,特征融合技术使农作物分类识别的性能得到进一步提升,因此基于多器官特征融合的可行性被广泛探究。文献[13]通过融合花冠和花蕊两个器官的颜色、形状、纹理等3个特征,实现了花卉的分类识别,在Oxford flower花卉数据库上识别准确率相比单器官提升1倍多。文献[14]通过融合树叶和树皮特征,在17个树种的数据集上识别准确率达到87.86%,实验结果明显优于单器官和传统方法。文献[15]通过融合奶牛头部、躯干和腿3个器官特征,在93头牛的数据集上识别准确率达到98.36%,证明了多器官融合比任意单器官以及任意两种器官组合的效果更优。文献[16]同时考虑植物的茎、叶、花和果实等器官,提出了一种植物自动识别方法,并探索了能包含更多特征信息的器官组合,进而提高识别精度,在包含101个植物物种的PlantCLEF数据集上,融合所有器官特征的识别算法获得最高97%的识别准确率。此外,文献[17]也在该数据集上选取12种图像丰富的植物做了相关研究,其通过融合叶、花、果实和整个植物的分类结果,识别准确率达到91.23%,有效提升了模型分类性能。上述研究表明,在动物和植物识别领域,融合多个器官特征的分类识别都取得了较单器官更优的识别结果。

事实上,植物本身就是由多种器官组成的复杂实体,其中大多数植物叶片包含大量具有识别能力的特征[18]。文献[19]在枣品种识别研究中,同时考虑枣果和叶片两个器官的特征信息,提取了枣果、叶片形状以及叶片纹理特征,在20个枣品种上识别准确率达到91.68%,相较于单器官识别算法准确率提升近8%,表明了多器官识别的有效性。

为充分发挥多器官特征优势,本文在文献[19]研究基础上,对其构建的数据集、训练方式以及融合策略进行改进,提出一种新的多器官枣品种识别算法。首先,对获取的枣果和叶片图像,依据笛卡尔积原理构建多器官枣品种分类数据集,即将每个品种的所有枣果图像与该品种所有叶片采用全配对方式,形成20个品种的枣果叶片组合集。然后,根据多分支网络模型出现的正负样本失衡问题,提出采用逐步训练和逐步迁移学习的方法训练模型,加快模型收敛速度,以解决由小样本和样本不均衡带来的准确率震荡问题。再次,为解决多器官特征融合时特征之间的互相干扰问题,设计一种强化特征相关性、弱化特征干扰的融合策略,以提升模型的分类性能。此外,本文还提出使用YOLO目标检测算法对原始图像中的枣果和叶片目标进行智能定位和裁剪,为构建样本丰富、多器官全组合的数据集提供自动化工具。

1 实验数据集

1.1 枣图像采集

本文枣图像数据与文献[19]一致,拍摄于河北省沧县国家枣树良种繁育基地,拍摄设备为安卓手机和尼康D7500型单反数码相机,分辨率分别为3 024像素×4 032像素和6 000像素×4 000像素。在晴天、阴天、顺光和逆光等自然环境下,采集不同

成熟期、不同遮挡情况以及不同光线的 20 个枣品种图像，共 3 881 幅，如表 1 所示。

表 1　枣品种数据集分布

Tab. 1　Distribution of jujube variety dataset

序号	枣品种	原始图像数量/幅	分割后枣果图像数量/幅	分割后叶片图像数量/幅	训练集图像数量/组	测试集图像数量/组
1	茶壶枣	228	119	104	8 064	460
2	朝阳圆枣	165	107	73	5 074	294
3	大荔龙枣	197	107	80	5 504	336
4	大木枣	172	88	84	4 828	272
5	短里长红	158	89	75	4 320	255
6	官滩枣	263	159	120	12 288	744
7	骨头小枣	162	99	74	4 800	266
8	葫芦长红	172	68	103	4 565	260
9	晋枣	161	95	68	4 180	247
10	辣椒枣	220	95	127	7 752	475
11	磨盘枣	181	50	132	4 240	260
12	酥圆铃	198	94	93	5 700	324
13	襄汾圆枣	200	110	94	6 688	396
14	小枣	216	177	50	5 680	350
15	小紫枣	159	104	77	5 208	300
16	宣城圆枣	154	86	60	3 312	204
17	圆铃小枣	199	118	84	6 460	368
18	圆铃枣	177	101	73	4 779	280
19	月芽枣	241	128	96	7 931	475
20	早躺熟枣	258	125	121	9 700	600
总计		3 881	2 119	1 788	121 073	7 166

每幅图像可能包含单个枣果、单个叶片、多个分离或重叠枣果以及枣果和叶片组合等复杂情况，如图 1 所示。

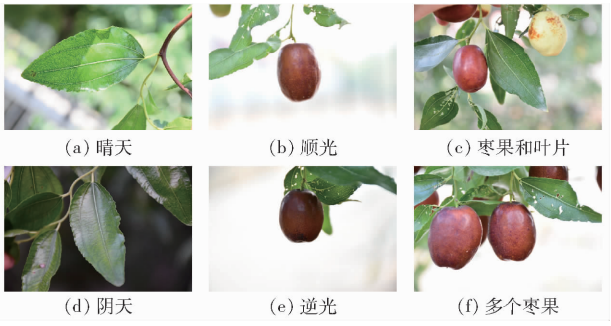

图 1　大荔龙枣不同场景原始图像

Fig. 1　Original images of different scenes of Dalilongzao

1.2　多器官数据集构建

基于多器官的品种识别模型与传统单分支卷积神经网络模型不同，其需要同时输入多个具有相同标签的器官图像，因此需构建适用于多器官分类模型的数据集，即将同品种枣果图像和叶片图像形成组合对，完成数据集的构建过程。为实现自动化数据集构建，丰富枣品种样本数量，有效刻画类间差异性和类内相似性特征，本文提出了如下的数据集构建方法。

(1) 使用微调过的 YOLO 目标检测算法对原始图像中的枣果和叶片器官进行自动分割裁剪，分别得到枣果和叶片图像，数量如表 1 所示。

(2) 对表 1 中第 4 列枣果和第 5 列叶片图像进行归一化处理，即按照长边等比缩放至 224 像素 × 224 像素，剩余像素填充为 0，以保证输入模型的图像不会扭曲变形。

(3) 分别将训练集和验证集中枣果和叶片图像数据集以笛卡尔积形式进行组合配对封装，形成多器官枣品种分类数据集。假设 $J_i(1 \leqslant i \leqslant 20)$ 表示本文研究的任意一个枣品种，则所构建的多器官数据集输入到模型的图像集合 I 计算公式为

$$I = Z_{J_i} Y_{J_i} = \{(z, y) \mid z \in Z_{J_i}, y \in Y_{J_i}; i = 1, 2, \cdots, 20\} \tag{1}$$

式中　Z_{J_i}——第 J_i 类枣品种所有枣果图像的集合

Y_{J_i}——第 J_i 类枣品种所有叶片图像的集合

z——枣果图像

y——叶片图像

(4) 将训练集和验证集均通过步骤 (3) 进行封装，最后得到输入到模型的训练集和测试集图像组数如表 1 中最后两列所示。

该数据集构建方法为利用自动化方法进行器官分割提取，并运用笛卡尔积思想构建器官样本数据组合对，是一种新的图像扩增方法，不仅扩大了数据集样本数量，也提升了数据集构建效率。同时这种配对方式也能充分利用同类枣品种中不同器官之间的关联关系，为模型训练、性能提升提供了有效的样本数据。

2　枣品种识别方法

本文提出的枣品种识别方法针对自然场景下的枣图像，以自动化数据集构建为基础，基于深度学习网络实现多分支的器官特征提取，经过特征融合完成品种识别，总体识别流程如图 2 所示。

首先将原始图像输入到训练好的 YOLO 模型，通过目标检测自动定位枣果和叶片器官，实现器官的自动分割和裁剪。其次，将分割得到的枣果和叶片图像分别输入到两个独立的卷积神经网络，通过训练进行特征提取。然后，将提取到的枣果和叶片特征通过融合策略进行特征融合。最后，将融合后的特征输入到分类层，实现枣品种分类识别。

2.1　器官图像的自动分割

利用目标检测方法从原始图像中分割提取并完

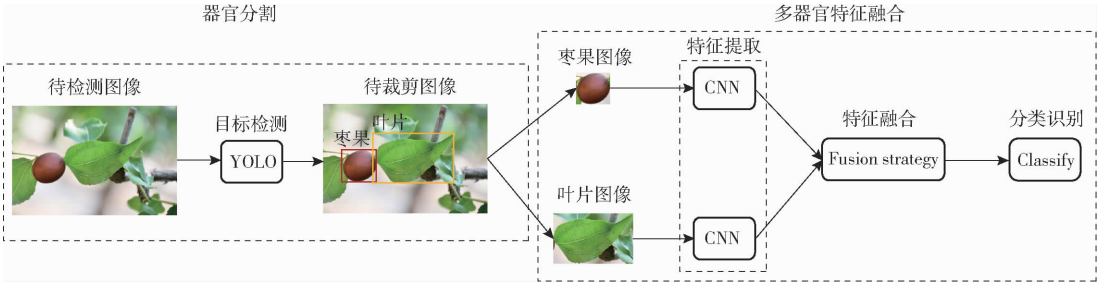

图 2 枣品种识别流程

Fig. 2 Jujube variety recognition process

成标注,包含器官图像检测和分类标签自动生成两个阶段。

在目标检测阶段,选取 YOLO v3[20] 目标检测模型用于枣果和叶片器官的自动定位和裁剪。为训练 YOLO v3 的目标检测性能,使用 LableImg 标注枣果和叶片发育良好、纹理清晰且形状完整的图像,将标注图像的 80% 用来微调模型参数,其余 20% 用来验证模型性能。图 3 为目标检测结果示例。



图 3 目标检测结果示例

Fig. 3 Example of object detection results

在分类标签自动生成阶段,通过设定概率阈值的方法完成自动裁剪和分类标注。假设枣果集合为 Z,叶片集合为 Y,计算公式为

$$Z=\{D_i \mid P_{D_i}>P_t, L_{D_i}=\mathrm{zg}; i=1,2,\cdots,n\} \quad (2)$$

$$Y=\{D_i \mid P_{D_i}>P_t, L_{D_i}=\mathrm{yp}; i=1,2,\cdots,n\} \quad (3)$$

式中 D_i——第 i 个目标检测结果

P_{D_i}——第 i 个目标检测结果概率

L_{D_i}——第 i 个目标检测结果目标名称

P_t——概率阈值

n——目标检测结果总数

其中,zg 表示枣果,yp 表示叶片。此方法会将每幅图像的定位检测结果自动分配到对应的枣果集合 Z 或叶片集合 Y 中。为保证分类图像数据集的质量,根据经验将 P_t 值设为 0.6。实验表明该经验值不仅能有效抑制原始图像中复杂背景的干扰,防止网络学习到无用特征,还可以直接用于新增枣类,为枣品种数据集构建提供一种通用方法。

2.2 多器官特征融合的枣品种识别

在多器官枣品种识别中,CNN 模型选取和特征融合策略是提升模型性能的关键,本文基于多器官特征融合提出的枣品种识别模型如图 4 所示。该模型主要由特征提取层、特征融合层以及分类层 3 部分组成。首先,在特征提取层使用 EfficientNetV2 的 Stage1 ~ Stage6 分别提取出枣果和叶片器官特征。然后,在特征融合层通过提出的拼接卷积融合(Concat convolution fusion,CCF)模块将这两种器官的特征进行融合。最后,在分类层使用融合后的特征进行 20 个枣品种的分类识别。

(1)多器官特征提取

在特征提取阶段,枣果和叶片两个分支均采用

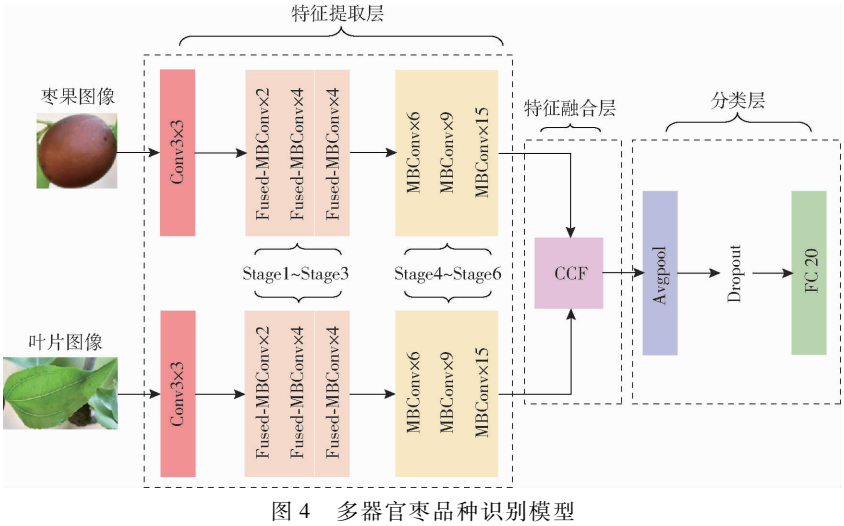

图 4 多器官枣品种识别模型

Fig. 4 Model for identifying multiple organ jujube varieties

EfficientNetV2[21]作为基础网络。该网络于 2021 年提出，并在 ImageNet 数据集[22]上取得了最优结果。该网络在 MBConv 模块的基础上引入 Fused-MBConv 模块，充分利用现有加速器，并使用较小的卷积核和扩展因子以及通过移除一些不必要的卷积层，以降低访问内存的开销，使模型训练速度得到提升的同时降低了模型参数。

在该网络结构中，第 1 层是卷积核尺寸为 3×3 的卷积层，并使用 SiLU 激活函数和 BN 层[23]。其次是 Stage1 ~ Stage3，每个 Stage 会重复堆叠 Fused-MBConv 模块，堆叠次数分别为 2、4、4 次。该模块主要由 1 个 3×3 的卷积层、1 个起降维作用的 1×1 卷积层和 1 个 Dropout 层组成，结构如图 5 所示。然后是 Stage4 ~ Stage6，每个 Stage 会重复堆叠 MBConv 模块，堆叠次数分别为 6、9、15 次。该模块主要由 1 个起升维作用的 1×1 卷积层、1 个 3×3 的 Depthwise Conv[24]、1 个 SE 模块[25]、1 个起降维作用的 1×1 卷积层和 1 个 Dropout 层组成，结构如图 6 所示。

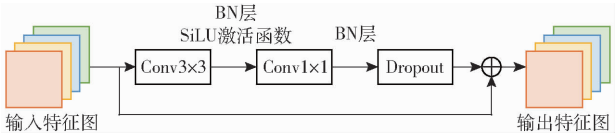

图 5　Fused-MBConv 模块

Fig. 5　Fused-MBConv module

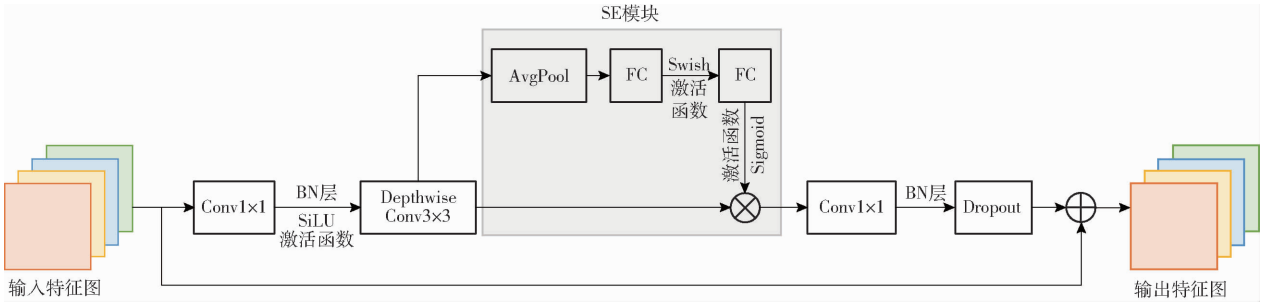

图 6　MBConv 模块

Fig. 6　MBConv module

此外，Fused-MBConv 模块和 MBConv 模块中的 Dropout 层为 Stochastic Depth[26]，即会随机丢掉整个模块的主分支，可以有效减小网络深度。

(2)多器官特征融合

在特征融合阶段，充分利用获得的多个器官特征是关键。本文提出的 CCF 模块由 1 个 Concat 模块和 1 个 1×1 的卷积模块组成，具体如图 7 所示。Concat 模块将枣果和叶片两个输入特征图在深度方向进行拼接，卷积模块则在多个方向上对每个元素进行更深一步融合，使拼接后的特征图在进行降维的同时能够充分学习到两个不同器官之间的相关性，进而提升模型分类性能。

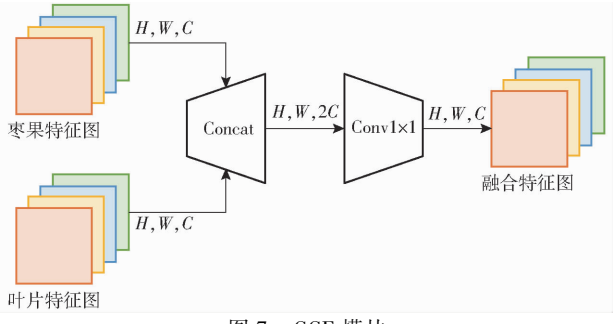

图 7　CCF 模块

Fig. 7　CCF module

(3)枣品种分类识别

在分类识别阶段，对融合特征图进行合理的下采样以去除冗余信息是关键。本文先通过参数为 1×1 的全局自适应平均值池化层将融合后的特征图池化成 1×1×C 的尺寸，以便展平后进行全连接操作。然后使用参数为 0.2 的 Dropout 层随机丢弃一部分神经元，防止模型出现过拟合。最后为输出通道数为 20 的全连接层以及 Softmax 函数，用于输出 20 种枣最终的概率分布。

2.3　训练策略

由于枣器官形状和外观差异较大，若同时学习不同器官的特征，容易导致模型正负样本失衡，训练难度极大。为充分发挥多器官的特征优势，本文提出了一种新的训练策略，即逐步迁移训练。该策略以迁移学习为基础，通过多次迁移和多次训练交替进行的方式，将整个模型的训练划分为 3 小部分逐步完成。训练步骤如图 8(灰色表示未训练、虚线表示不训练)所示，先训练枣果分支，再训练叶片分支，最后训练融合层与分类层。

(1)训练枣果分支。如图 8a 所示，通过迁移 EfficientNetV2 模型在 ImageNet 上的预训练权重，加载卷积部分的参数，单独训练枣果分支，其他部分学习率置零。

(2)训练叶片分支。如图 8b 所示，同样使用迁移学习的方式加载 ImageNet 预训练权重，单独训练叶片分支，其他部分学习率置零。

(3)训练融合层与分类层。如图 8c 所示，迁移

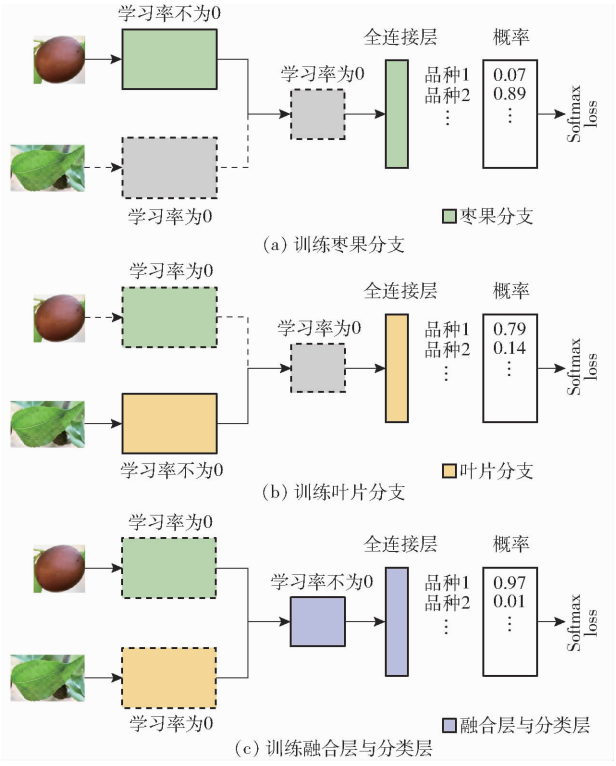

图 8 训练步骤图

Fig. 8 Training step diagram

前面两步训练好的枣果和叶片分支权重,并通过学习率置零的方式冻结两路参数,只训练融合模块中的卷积层以及最后的全连接层。

该方法通过逐步迁移和逐步训练相结合的学习方式,在每一部分的训练中,均通过加载预训练模型来提高模型的收敛速度、降低网络训练难度。而迁移学习作为一种优化,可以将卷积神经网络在一个任务上学习到的模型结构、参数权重等迁移到另一个任务中,能有效提高模型的建模速度、降低网络的训练时间,还可以解决枣图像数据不够庞大以及数量不均衡带来的准确率震荡问题,进而提升模型的稳定性和识别准确率。

3 实验

实验采用的硬件环境为 Intel Core i7 - 12700F 2.10 GHz 处理器、32 GB 内存和 NVIDIA GeForce RTX 3060Ti 显卡,软件环境为 Windows 10 操作系统、Pycharm 开发环境、Python 3.7 编程语言和 Pytorch 1.13.1 深度学习框架。模型的批量大小设置为 32,学习率设置为 0.000 1,并使用 Adam 优化器进行参数更新。

为更好地评价所提出算法的性能,本文主要选用准确率作为评价指标。同时,为检验本文模型的训练效率,还增加了训练时间和收敛时间两个评价指标。其中,训练时间为模型训练 1 个周期所需的时间。收敛时间为模型收敛时训练的周期数量。

3.1 枣品种识别结果分析

为验证本文提出的多器官枣品种识别模型的性能,对近年来枣品种识别研究中选取品种基本一致的 5 个识别方法进行了比较。本文提出的方法与 5 个枣品种分类识别方法在相同评估指标上的对比结果如表 2 所示。

表 2 现有枣品种识别方法对比

Tab. 2 Comparison of research on identification of existing jujube varieties

方法	准确率/%	种类数	器官选取
高光谱技术[3]	91.20	3	枣果
集成学习[4]	92.38	5	枣果
细粒度[5]	84.16	20	枣果
多任务细粒度[6]	91.79	20	枣果
三分支并行网络[19]	91.68	20	枣果、叶片
本文方法	97.04	20	枣果、叶片

由表 2 可知,在识别器官选取上,选择枣果作为枣品种识别器官取得的最好效果为 92.38%,但其针对枣品种数量仅为 5 种。文献[19]同时考虑枣果和叶片器官的特征信息,在 20 种枣上识别准确率为 91.68%。与上述文献相比,本文方法在识别种类数和识别准确率上都取得了最好的效果,且在相同枣品种数量情况下,识别准确率比文献[19]有明显提升。一方面,本文算法在预处理时对原始枣果图像进行了分割和裁剪,排除了由复杂背景带来的干扰;另一方面,提出了优化多器官的特征融合方法和训练方式,充分发挥了枣果和叶片的特征优势。

3.2 单器官与多器官在不同模型上的识别结果

为验证多器官特征融合的性能,本文选取 AlexNet[27]、VGG16[28]、ResNet18[29]和 GoogleNet[30] 4 个经典的卷积神经网络作为基础网络,对单器官与多器官的识别结果进行了对比和分析。其中,单器官识别结果是直接使用原模型训练和验证的结果,多器官识别结果是将本文方法中特征提取部分的网络结构替换为其他基础网络的卷积层后进行训练得到的验证结果,如表 3 所示。

由表 3 可知,选用 EfficientNetV2 作为基础网络的方法取得了 97.04% 的最高准确率,表明了 EfficientNetV2 网络比其他网络具有明显优势。从器官选取看,综合多器官的识别结果均比任一单器官识别结果有明显提升。因此,融合多种器官的特征可以弥补单器官特征不足,为枣品种识别提供了更多具有辨别能力的特征信息。若从单一器官的角

表 3　单器官与多器官在不同模型上的识别准确率

Tab. 3　Recognition accuracy of single and multiple organs on different models　%

基础网络	器官		
	枣果	叶片	枣果 + 叶片
AlexNet	75. 90	62. 18	84. 93
VGG16	78. 31	64. 47	89. 83
ResNet18	86. 51	76. 50	95. 33
GoogleNet	86. 99	76. 79	95. 58
EfficientNetV2	89. 88	79. 08	97. 04

度来看，枣果和叶片器官的特征均具有辨别能力，但枣果明显拥有比叶片更好的识别效果，说明枣果的类间差异性大、类内差异性小，对枣品种的识别贡献度更高。

3.3　不同特征融合方法在不同模型上的识别结果

为验证特征融合方法的性能，本文尝试了最大值融合（Max）、加法融合（Add）和拼接融合（Concat）3 种融合方法，并将这些方法与不同基础模型结合进行对比实验，结果如表 4 所示。

表 4　不同特征融合方法在不同模型上的识别准确率

Tab. 4　Recognition accuracy of different feature fusion methods on different models　%

融合方法	基础网络				
	EfficientNetV2	GoogleNet	ResNet18	VGG16	AlexNet
Max	96. 48	95. 35	94. 98	85. 95	80. 28
Add	96. 75	95. 46	94. 89	86. 09	80. 38
Concat	96. 82	95. 38	95. 19	87. 97	84. 79
CCF	97. 04	95. 58	95. 33	89. 83	84. 93

由表 4 可知，本文所提出的融合方法相较于其他融合方法，在所有基础模型上的识别效果均有提升，说明此方法能够有效结合不同器官特征之间的差异性特征，使枣品种的特征类内聚合、类间分散，进而降低枣品种间的歧义性，提升枣品种识别模型的分类性能。另外，结合表 3 和表 4 实验结果，本节实验结果又对 3. 2 节结果进行了进一步验证，即不论选用何种基础网络以及何种融合方法，多器官识别结果均优于任一单器官的识别结果。

3.4　消融实验

为验证本文提出的训练方法以及构建数据集方法的有效性，对不同训练方式和构建数据集方式的组合进行了对比分析，并且在不同网络模型上做了消融实验。其中，直接迁移学习是指迁移基础模型在 ImageNet 上的部分预训练参数应用到本文模型的特征提取部分，并直接训练多器官枣品种识别模型。传统图像扩增是指在划分完训练集和测试集后，将枣果和叶片训练集图像通过亮度增强、对比度增强、饱和度增强、高斯模糊、高斯噪声、水平翻转、垂直翻转以及随机旋转 90°等 8 种方式扩增，然后再归一化和数据封装后输入到模型进行训练和验证。输入到模型的图像组数为当前种类枣果和叶片图像数量最大值，且在封装时，图像数量较少的器官将被循环取用。表 5 为不同训练和构建数据集方式的组合在不同模型上的消融实验结果。

表 5　不同训练和构建数据集方式的组合在不同模型上的识别准确率

Tab. 5　Recognition accuracy of different combinations of training and dataset construction methods on different models　%

基础网络	数据集方式			
	传统图像扩增 + 直接迁移学习	传统图像扩增 + 本文训练方式	本文构建方式 + 直接迁移学习	本文构建方式 + 本文训练方式
AlexNet	79. 73	80. 18	83. 48	84. 93
VGG16	85. 59	86. 71	87. 55	89. 83
ResNet18	85. 34	88. 54	91. 80	95. 33
GoogleNet	86. 94	92. 34	89. 51	95. 58
EfficientNetV2	90. 77	94. 37	93. 50	97. 04
平均值	85. 67	88. 43	89. 17	92. 54

由表 5 可得，本文提出的训练方法和数据集构建方法在 5 个不同基础网络上均有明显提升。若仅使用本文提出的训练方式，能平均提升 2. 76 个百分点。若仅使用本文提出的数据集构建方式，能平均提升 3. 50 个百分点。若同时使用两种方法，能平均提升 6. 87 个百分点。由此分析可得，无论将何种网络作为基础网络，本文提出的训练方法和数据集构建方法均能达到最优的识别结果，且两种方法对模型性能的提升是正相关的。

此外，由表 5 可知，在 ResNet18 和 GoogleNet 作为基础网络时，使用传统图像扩增 + 直接迁移学习方法的多器官枣品种识别模型达到的准确率分别为 85. 34% 和 86. 94%，而表 3 中对应模型的枣果单器官准确率却为 86. 51% 和 86. 99%，略高于多器官的识别结果。通过分析可得，出现该问题的原因在于叶片单器官在两个基础网络上的准确率较低，因此融合时枣果信息受到了叶片信息的干扰，导致最终模型性能的降低。由此可见，只有使用适合于多器官的训练方式和构建数据集方法，才能充分利用多器官的特征优势，使模型达到正向的识别结果。

3.5　不同训练方式和构建数据集方式性能对比

为客观比较本文提出的训练方式和构建数据集方式对模型整体性能的影响，以 EfficientNetV2 作为基础模型，比较了不同训练方式与不同数据集构建方式在训练时间与收敛速度上的表现，结果如表 6

表 6 不同训练方式与构建数据集方式性能比较

Tab. 6 Performance comparison between different training methods and dataset construction methods

训练方式	传统图像扩增			本文构建方式		
	准确率/%	训练时间/s	收敛速度/轮	准确率/%	训练时间/s	收敛速度/轮
直接迁移学习	90.77	301 ±2	10	93.50	2 030 ±25	4
本文训练方式	94.37	89 ±4	5	97.04	481 ±10	2

所示。

由表 6 可知,本文提出的方法不仅在准确率上取得最优,而且在训练时间和收敛速度上也表现良好。分析可得,本文提出的训练方式与直接迁移学习相比,训练速度平均快 3.22 倍,收敛速度均快1 倍。此外,由表 1 可得,本文构建的数据集输入到模型进行训练的图像组数高达 121 073,而采用传统图像扩增方式输入到模型进行训练的图像组数仅为 16 416,数量高出近 7 倍,但训练时间仅平均高出 4.4 倍,且收敛速度平均提升 1.5 倍。因此,若将训练时间平均到每幅图像或每个批次而言,本文所提出的方法在训练时间和收敛速度上仍表现最优。

4 结束语

提出了一种基于多器官特征融合的枣品种识别方法。首先,使用 YOLO 模型在包含复杂背景的原始图像中自动分割出枣果和叶片器官,完成数据集的自动构建。然后,使用微调后的 EfficientNetV2 模型分别提取这两个器官的特征,并通过 CCF 特征融合模块进行特征融合。最后,使用融合后的特征对枣品种进行分类。实验结果表明,在包含 20 个枣品种的数据集上,该方法获得 97.04% 的识别准确率,明显优于现有研究结果。在模型训练时间和收敛速度上,该方法也表现最优,充分展现了多器官特征融合方法的优势。

参 考 文 献

[1] 刘孟军,汪民. 中国枣种质资源[M]. 北京:中国林业出版社,2009:9 - 11.

[2] 曲泽洲,王永惠. 中国果树志·枣卷[M]. 北京:中国林业出版社,1993:5 - 22.

[3] 王秉聪. 基于高光谱图像技术的红枣品种智能检测模型[J]. 农业装备技术,2023,49(2):10 - 12.

[4] 余游江,喻彩丽,尚远航,等. 基于 Stacking 模型的红枣品种分类识别[J]. 北方园艺,2022(8):139 - 148.

[5] MENG X, YUAN Y C, TENG G F, et al. Deep learning for fine-grained classification of jujube fruit in the natural environment [J]. Journal of Food Measurement and Characterization, 2021, 15(5): 4150 - 4165.

[6] ZHANG R C, YUAN Y C, MENG X, et al. A multitask model based on MobileNetV3 for fine-grained classification of jujube varieties[J]. Journal of Food Measurement and Characterization, 2023, 17(5): 4305 - 4317.

[7] 刘媛媛. 基于特征融合的苹果分类方法研究[D]. 福州:福建师范大学,2020.
LIU Yuanyuan. Research on apple classification approach based on feature fusion[D]. Fuzhou: Fujian Normal University, 2020. (in Chinese)

[8] 雷欢,焦泽昱,马敬奇,等. 基于多特征融合与 SVM 的苹果品种快速识别算法[J]. 自动化与信息工程,2020,41(4):13 - 17.
LEI Huan, JIAO Zeyu, MA Jingqi, et al. Fast recognition algorithm of apple varieties based on multi feature fusion and SVM [J]. Automation & Information Engineering, 2020, 41(4): 13 - 17. (in Chinese)

[9] 苗荣慧,黄锋华,杨华,等. 基于光谱和 Gabor 纹理信息融合的油桃品种识别[J]. 江苏农业科学,2019,47(6):174 - 178.

[10] 王佳琪. 基于深度学习的果实检测跟踪系统的研究与实现[D]. 北京:北京工业大学,2021.
WANG Jiaqi. Research and implementation of fruit detection and tracking system based on deep learning[D]. Beijing: Beijing University of Technology, 2021. (in Chinese)

[11] 李秀昊,马旭,李泽华,等. 基于特征融合和 SVM 的稻谷品种识别[J]. 中国农机化学报,2019,40(7):97 - 102.
LI Xiuhao, MA Xu, LI Zehua, et al. Identification of rice variety based on multi-feature fusion and SVM[J]. Journal of Chinese Agricultural Mechanization, 2019, 40(7): 97 - 102. (in Chinese)

[12] 游嘉伟. 基于深度学习的大豆叶片模式品种识别方法[D]. 南京:南京财经大学,2021.
YOU Jiawei. Cultivar recognition methods of soybean leaf image based on deep learning[D]. Nanjing: Nanjing University of Finance and Economics, 2021. (in Chinese)

[13] 刘欣悦. 基于移动端的多特征花卉识别系统[D]. 长春:吉林大学,2019.
LIU Xinyue. Muti-feature flower recognition system based on mobile terminal[D]. Changchun: Jilin University, 2019. (in Chinese)

[14] ZHAO Y F, GAO X, HU J F, et al. Tree species identification based on the fusion of bark and leaves[J]. Mathematical Biosciences and Engineering, 2020, 17(4): 4018 - 4033.

[15] HU H Q, DAI B S, SHEN W Z, et al. Cow identification based on fusion of deep parts features[J]. Biosystems Engineering, 2020, 192: 245 - 256.

(下转第 240 页)

regression[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 658 - 666.
[23] CAO Y, XU J, LIN S, et al. Gcnet: non-local networks meet squeeze-excitation networks and beyond[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019: 1971 - 1980.
[24] LI X, WANG W, HU X, et al. Selective kernel networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 510 - 519.
[25] LI Y, YAO T, PAN Y, et al. Contextual transformer networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(2): 1489 - 1500.
[26] 方鹏,郝宏运,李腾飞,等.基于注意力机制和可变形卷积的鸡只图像实例分割提取[J].农业机械学报,2021,52(4):257 - 265.
FANG Peng, HAO Hongyun, LI Tengfei, et al. Instance segmentation of broiler image based on attention mechanism and deformable convolution[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(4): 257 - 265. (in Chinese)
[27] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]// Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3 - 19.
[28] YANG L, ZHANG R Y, LI L, et al. Simam: a simple, parameter-free attention module for convolutional neural networks[C]// International Conference on Machine Learning, 2021: 11863 - 11874.
[29] 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.
[30] HAN K, WANG Y, TIAN Q, et al. Ghostnet: more features from cheap operations[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1580 - 1589.

(上接第 220 页)

[16] RZANNY M, MÄDER P, DEGGELMANN A, et al. Flowers, leaves or both? How to obtain suitable images for automated plant identification[J]. Plant Methods, 2019, 15(1): 77.
[17] 罗娟,蔡骋.多线索植物种类识别[J].计算机工程与应用,2020, 56(5):160 - 165.
LUO Juan, CAI Cheng. Multi-cue plant species identification[J]. Computer Engineering and Applications, 2020, 56(5): 160 - 165. (in Chinese)
[18] LEE S H, CHAN C S, MAYO S J, et al. How deep learning extracts and learns leaf features for plant classification[J]. Pattern Recognition,2017, 71: 1 - 13.
[19] 孟惜,苑迎春,刘博,等.融合枣果与叶片信息的枣品种识别方法研究[J].河北农业大学学报,2022,45(2):107 - 113.
MENG Xi, YUAN Yingchun, LIU Bo, et al. Classification of jujube by integrating fruit with leaf information[J]. Journal of Hebei Agricultural University, 2022, 45(2): 107 - 113. (in Chinese)
[20] REDMON J, FARHADI A. YOLO v3: an incremental improvement[J]. arXiv Preprint arXiv:1804.02767, 2018.
[21] TAN M, LE Q. EfficientNetV2: smaller models and faster training[C]// International Conference on Machine Learning. PMLR, 2021: 10096 - 10106.
[22] DENG J, DONG W, SOCHER R, et al. Imagenet: a large-scale hierarchical image database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009: 248 - 255.
[23] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]// International Conference on Machine Learning. PMLR, 2015: 448 - 456.
[24] SIFRE L. Rigid-motion scattering for image classification[D]. Paris: Ecole Polytechnique, 2014.
[25] 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.
[26] HUANG G, SUN Y, LIU Z, et al. Deep networks with stochastic depth[C]// Computer Vision-ECCV 2016: 14th European Conference, 2016: 646 - 661.
[27] ALEX K, ILYA S, GEOFFREY E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017,60(6):84 - 90.
[28] SIMONYAN K, ZISSERMA A. Very deep convolutional networks for large-scale image recognition[J]. arXiv Preprint arXiv: 1409.1556, 2014.
[29] 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.
[30] 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.