

DOI: 10.13733/j.jcam.issn.2095-5553.2025.03.037

郭敬涛, 吕凤, 章慧婷, 等. 基于迁移学习与轻量化 YOLOv5s 的草莓目标检测方法[J]. 中国农机化学报, 2025, 46(3): 253-260

Guo Jingtao, Lü Feng, Zhang Huiting, et al. Strawberry target detection method based on transfer learning and lightweight YOLOv5s [J]. Journal of Chinese Agricultural Mechanization, 2025, 46(3): 253-260

基于迁移学习与轻量化 YOLOv5s 的草莓目标检测方法^{*}

郭敬涛¹, 吕凤¹, 章慧婷¹, 杨彪², 刘大洋¹

(1. 东北林业大学计算机与控制工程学院, 哈尔滨市, 150040;

2. 商洛学院电子信息与电气工程学院, 陕西商洛, 726000)

摘要:为实现草莓采摘时精准检测,同时考虑到嵌入式设备内存小、计算能力低下,而当下目标检测模型参数量和计算量巨大的问题,提出一种基于 YOLOv5s 的轻量化网络模型。首先,对 YOLOv5s 进行轻量化处理,利用深度卷积(DWConv)替换普通卷积,同时用 C3Ghost 模块替换原网络模型中的 C3 模块,降低模型的复杂度。然后,为增强主干网络对特征信息的提取能力,加强输入特征图通道间的信息交互,在主干网络的 C3 模块中融合高效通道注意力(ECA)结构,在特征融合网络添加无参数注意力模块(SimAM),使网络聚焦更多的有效特征信息,达到不增加模型的参数量,同时又提升模型识别精度的目的。最后,结合迁移学习加快模型收敛速度并进一步提升模型检测精度。结果表明,轻量化后的网络模型体积减小 55.8%,计算量减少 55.1%,在自制草莓数据集上的平均精度均值 $mAP@0.75$ 达到 74.9%,比原模型提高 3.1%,单张图片平均推理时间仅 6.4 ms,能够实现在草莓采摘任务中的精准快速检测,为草莓生产智能化提供支持。

关键词:草莓目标检测;深度学习;注意力机制;轻量化模型;迁移学习

中图分类号:S126; TP391.4 **文献标识码:**A **文章编号:**2095-5553(2025)03-0253-08

Strawberry target detection method based on transfer learning and lightweight YOLOv5s

Guo Jingtao¹, Lü Feng¹, Zhang Huiting¹, Yang Biao², Liu Dayang¹

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

2. Electronic Information and Electrical Engineering College, Shangluo University, Shangluo, 726000, China)

Abstract: To achieve the accurate detection of strawberry in agricultural harvesting, a lightweight network model based on YOLOv5s is proposed considering the limited memory and low computational power of embedded devices, as well as the huge parameters and computational demands of current target detection models. First, the YOLOv5s structure is lightweight processed by replacing ordinary convolutions with depthwise convolutions (DWConv) and substituting the C3 module in the original network with the C3Ghost module to reduce the model complexity. Second, to enhance the ability of the backbone network to extract feature information and improve the interaction between channels in the input feature maps, an efficient channel attention (ECA) structure is integrated into the C3 module of the backbone network. Additionally, a parameter-free attention module (SimAM) is added to the feature fusion network, so that the model can focus on more effective feature information without increasing the number of parameters of the model while improving the recognition accuracy. Finally, transfer learning is combined to accelerate the convergence speed of the model and further improve the detection accuracy. The results indicate that the lightweight model reduces network size by 55.8% and computation by 55.1%. The $mAP@0.75$ tested on a custom strawberry dataset reaches 74.9%, which is 3.1% higher than that of the original model. The average inference time per image is only 6.4 ms. This enables accurate and fast detection in strawberry picking tasks and provides support for the intelligent production of strawberries.

Keywords: strawberry target detection; deep learning; attention mechanism; lightweight model; transfer learning

收稿日期:2023年9月14日 修回日期:2023年11月20日

^{*} 基金项目:国家自然科学基金(32202147);中国博士后基金面上项目(2021M690573);中央高校基本科研业务费专项资金(2572020BF05);陕西省科学技术协会青年人才托举计划项目(20220124)

第一作者:郭敬涛,男,1996年生,河南许昌人,硕士研究生;研究方向为农林业智能检测。E-mail: 17613563953@163.com

通讯作者:刘大洋,男,1990年生,吉林四平人,博士,副教授;研究方向为农林业智能检测。E-mail: ldy333ldy@163.com

0 引言

草莓是目前市场上较为常见的一种水果,因其丰富的营养和低含量的脂肪成分,有“水果皇后”的美誉,加上经常食用能够增强人体的抵抗力,因此备受市场喜爱^[1]。根据国家统计局数据显示,2014—2022年,我国草莓种植面积由 94.82 km² 增长至 130.24 km²^[2],但目前草莓采摘主要采取人工方式,自动化程度低、劳动强度大、采摘成本高且效率低,因此,研究草莓果实快速准确的检测方法,实现自动化采摘十分重要。

随着计算机视觉技术和深度学习的不断发展,尤其是一系列卷积神经网络模型的提出,将基于深度学习的目标检测技术应用于农业果实识别已经成为较热门的研究领域^[3-5]。Bargoti 等^[6]研究了将甜椒的 RGB 图像和 NIR 图像分别进行早期数据融合、晚期数据融合以及不进行数据融合情形下在 Faster R—CNN 模型中的检测性能,发现将两组信息进行晚期融合能够获得最高的 *F1* 值。闫建伟等^[7]改进 Faster R—CNN 中感兴趣区域池化为感兴趣区域校准,提高了模型检测精度,*F1* 值最高达 94.99%。以 Faster R—CNN 为主要代表的二阶段目标检测算法,优点是精度高,缺点是检测速度慢,以 YOLO 和 SSD^[8] 为代表的单阶段算法不需要得到建议框,在速度上有更大的优势,精度上也在逐渐赶上并超过二阶段算法。周桂红等^[9]在 YOLOv4 的特征提取层加入 scSE 注意力机制,替换 PANet 网络中的部分卷积为深度可分离卷积,结合基于阈值的边界框匹配合并算法,在苹果全景图像识别任务中的平均精度均值 *mAP* 达 95.06%,*F1* 值达 96%。Fan 等^[10]为解决夜间采集的图像光照不足的问题,提出一种结合 YOLOv5s 和暗通道增强的算法,在 4 种不同成熟度的草莓数据集上训练的 *mAP* 可以达到 90% 以上。

以上方法在实际果实目标检测任务中已经取得较好的效果,但由于卷积神经网络结构复杂且计算量大,不利于在嵌入式设备的部署。因此,对以上深度学习网络进行轻量化处理具有重要的现实意义。Fu 等^[11]提出一种权重大小为 27 MB 的基于 YOLOv3—tiny 模型的 DY3TNet 网络,对果园中的猕猴桃进行检测,平均检测精度达 90.05%,在 GPU 设备上单幅图像推理时间为 34 ms,实现了猕猴桃的快速检测。孙俊等^[12]利用轻量级网络 GhostNet 替换 YOLOv4—Tiny 模型的特征提取网络并添加注意力模块,并采用高效交并比损失作为边界框回归损失函数,最后得到的模型权重文件大小仅为 4.68 MB,在测试集上的 *mAP* 相较于原模型提高 5.77%,每张图片的平均推理时间

为 5.63 ms。陈仁凡等^[13]提出一种基于 YOLOv5s 的轻量级网络模型 YOLO—ODM,首先在主干网络中引入 Shuffle_Block 以实现轻量化,同时在颈部结构中使用全维度动态卷积模块提高模型信息提取能力,并进一步精简模型结构,最终改进后模型的 *mAP* 达 97.4%,模型体积为 7.79 MB,可快速准确地对温室环境下的草莓果实成熟度进行检测。

上述研究在平衡模型的轻量化与检测精度上已经取得较好的效果,但两者仍有一定的提升空间,且以往对草莓的研究侧重于对成熟果实的检测,缺少对草莓花期到成熟全过程的研究。为此,本文以温室草莓全生长阶段果实为研究对象,旨在探索低资源占用情况下,在保证检测速度的同时,提高对草莓果实的检测精度,为草莓的自动化生产提供一定的技术支持。首先,基于 YOLOv5s 网络模型,通过替换普通卷积为深度卷积(DWConv),引入 Ghostbottleneck 精简模型的结构;然后,在模型特征提取阶段融入高效通道注意力(ECA)结构,在特征融合网络添加无参数注意力模块(SimAM);最后,结合迁移学习策略加快模型的收敛速度,同时进一步提高模型的检测精度。

1 材料与方 法

1.1 数据集采集

数据集 1 为 PASCAL VOC 挑战赛的公共数据集,包括 VOC 2007 和 VOC 2012,其中标注的目标包含 vehicle、household、animal、person 4 个大类,这 4 个大类又细分为 20 个小类,因为该数据集目标种类比较丰富且图片数量适中,其作为大型目标检测比赛使用的专用数据集,在做消融试验时更具权威性,考虑到自身计算资源有限,所以选择该数据集作为预训练数据集。

数据集 2 的草莓图片一部分采集于河南省许昌市建安区五女店镇某草莓种植园,品种为“甜查理”;另一部分采集于陕西省杨凌区农业创新园内,品种为“章姬”,两个品种的草莓如图 1 所示。



(a) “甜查理”

(b) “章姬”

图 1 两个品种草莓图像

Fig. 1 Images of two varieties of strawberry

草莓种植方式为大棚地垄式,拍摄时间为 2023 年 1—2 月,这一期间包含草莓各个生长阶段的果实,在自然光照条件下拍摄,拍摄设备为华为 P30,拍摄距离在 30 cm 以内,图片保存格式为 jpg,“甜查理”草莓采集了

1 158 张,“章姬”草莓采集了 982 张,共计 2 140 张。

1.2 数据集构建

将数据集 1 中 2012 train+val 和 2007 train+val 作为训练集样本,2007 test 作为验证集样本,划分后的训练集和验证集的图片数量分别为 16 551 张和 4 952 张。

使用 LabelImg 工具对数据集 2 各个目标进行标注,格式为 YOLO,按照生长阶段不同,将果实目标分为花期、青果期、白果期、转色期和成熟期,分别对应的标签为 flower、green_fruit、white_fruit、color-turning 和 ripe,其中着色面积超过 80% 则被定义为成熟期。标注后的数据集按照 7 : 2 : 1 随机划分为训练集、验证集、测试集。为扩充数据集以增强模型的抗干扰能力,采用线上、线下结合的方式对训练集数据进行增强。考虑到图像采集装置在实际拍摄过程中可能产生噪声以及因为震动造成图像的模糊等情况,分别对数据集进行添加高斯噪声和运动模糊两种线下增强处理。另外考虑到光线的变化和采集角度对图片的影响,线上增强方式包含 HSV 增强、0°~180°随机旋转、平移、放缩等操作,最终上述处理后的图片再经过 Mosaic 增强,送入网络模型进行训练。数据集 2 各个类别目标的基本信息如表 1 所示。

表 1 草莓数据集分布信息

Tab. 1 Information on strawberry dataset distribution 张

数据集	目标类别				
	花期	青果期	白果期	转色期	成熟期
训练集	2 508	16 347	8 427	6 525	6 810
验证集	238	1 592	783	642	643
测试集	103	810	422	285	361
共计	2 849	18 749	9 632	7 452	7 814

1.3 训练策略

基于微调的二阶段学习策略,训练流程如图 2 所示。

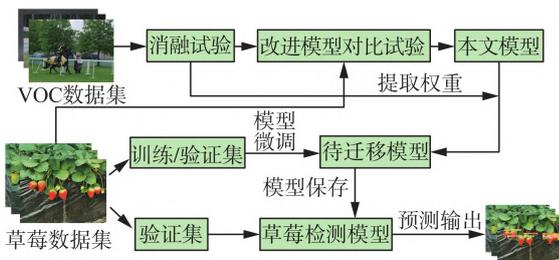


图 2 基于迁移学习策略的训练流程

Fig. 2 Training process based on transfer learning strategy

先将数据集 1 在不同网络模型上训练,迭代次数设置为 300, batch size 设置为 32,初始学习率为 0.001,使用余弦退火算法动态调整学习率,权重衰减

率为 0.000 5,动量为 0.937,采用随机梯度下降法优化参数。在数据集 1 上进行消融试验验证本文改进部分的效果后,在数据集 2 上进行改进模型的对比试验确定模型结构,接下来以数据集 1 训练的权重初始化网络参数,在数据集 2 上进行参数迁移和模型微调,迭代次数设置为 100,其余参数保持不变。

1.4 试验环境

试验环境基于 Pytorch 深度学习框架搭建,在 GPU 环境下训练测试,具体试验环境配置如表 2 所示。

表 2 试验环境配置

Tab. 2 Test environment configuration

试验环境	名称	参数
硬件	CPU	13th GenIntel(R) Core(TM) i9-13900K 3.00 GHz
	内存	32 G
	GPU	GeForce RTX 4090
	显存	24 G
	操作系统	Windows 11
软件	深度学习框架	Pytorch 1.13.0
	编程语言	Python 3.8.0
	CUDA	11.6

2 识别方法

2.1 YOLOv5s 网络介绍

YOLOv5s 模型的结构包括输入端,主干提取网络 Backbone,瓶颈网络 Neck 和检测网络 Detect。输入端图片尺寸默认为 640 像素 × 640 像素 × 3 通道。Backbone 是由 Conv、C3 和快速空间金字塔池化模块(SPPF)构成,用于输入图片的特征提取,其中 Conv 模块是 Conv2d()、BatchNorm2d()和 SiLU()的组合,简称为 CBS 操作;C3 模块将输入的特征图分为两个分支,分支 1 经过普通卷积 Conv 操作,分支 2 经过 CBS 和多个 Bottleneck 处理,最后将两路结果进行 Concat 作为模块的输出,残差结构在深层网络中很好地减少了特征图语义信息的丢失,在 Backbone 中 C3 模块 shortcut 设置为 True,而 Neck 中则设置为 False;SPPF 模块通过并列的 3 个最大池化操作与未经池化操作的通道进行 Concat 拼接,输出通道变为输入的 2 倍,极大地提升模型的感受野范围,有利于检测图像中不同大小的目标对象。Neck 网络采用 PANet 结构,深度卷积神经网络中,浅层网络的特征层包含的细节信息丰富但语义信息较少,深层网络的特征层细节信息不足但语义信息丰富,PANet 结构能够将浅层与深层网络特征进行融合提升检测的性能。最后将 Neck 网络中引出的 3 个尺寸的特征预测层送入

Detect 模块进行预测输出。

2.2 改进的 YOLOv5s 网络模型

改进后的网络模型如图 3 所示。(1)利用深度卷积代替普通标准卷积,以减少模型的参数量和计算量;(2)引入 GhostBottleneck 替换原 C3 模块的 Bottleneck,进一步实现模型的轻量化;(3)在模型的 Backbone 中融合一种轻量型有效通道注意力结构 ECA,并与 C3 模块相结合,加强 Backbone 对特征的提取能力;(4)在模型特征融合网络添加 SimAM,增强有效信息同时抑制无关特征的干扰。

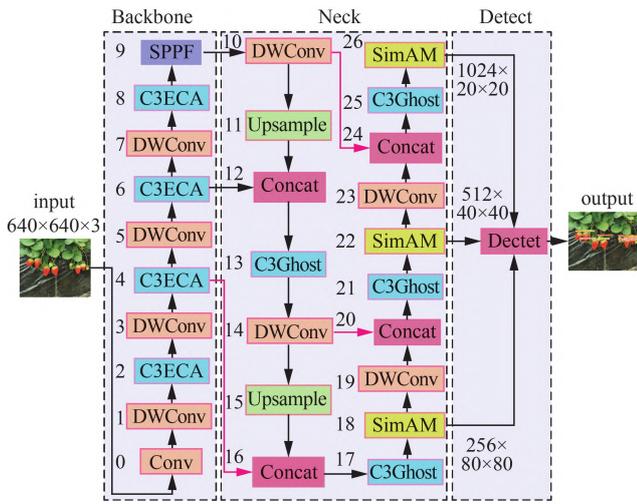


图 3 改进的 YOLOv5s 网络模型结构

Fig. 3 Improved YOLOv5s network model structure

2.2.1 利用 DWConv 代替普通卷积

如图 4 所示,对于输入尺寸为 $D_x \times D_y \times C_{in}$ 的特征图,在普通卷积中经过 C_{out} 个 3×3 的卷积核卷积计算后可以得到尺寸为 $D_x \times D_y \times C_{out}$ 的特征图的输出。普通卷积过程将 C_{out} 个 3×3 的卷积核分别与特征图的每个通道结合,得到 C_{out} 个通道的新特征图。假设 $padding = 1$,不考虑偏置,使用不同卷积的参数数量 P_1 和计算量 C_1 可以通过式(1)计算。

$$\begin{cases} P_1 = 3 \times 3 \times C_{in} \times C_{out} \\ C_1 = D_x \times D_y \times 3 \times 3 \times C_{out} \end{cases} \quad (1)$$

DWConv 是分组卷积^[14]的一种特例,即分组数和特征图的输入通道数 C_{in} 相等,使用 C_{out} 个 3×3 的卷积核与输入特征图的每个通道分别进行卷积,得到一个输入通道与输出通道相等的特征图。假设 $padding = 1$,不考虑偏置,使用不同卷积的参数数量 P_2 和计算量 C_2 可以通过式(2)计算。

$$\begin{cases} P_2 = 3 \times 3 \times C_{in} \\ C_2 = D_x \times D_y \times 3 \times 3 \times 3 \end{cases} \quad (2)$$

DWConv 与普通卷积的参数量与计算量之比为

$$K = \frac{P_2}{P_1} = \frac{C_2}{C_1} = \frac{1}{C_{out}} \quad (3)$$

由式(3)可知,DWConv 能够将模型体积和计算量压缩为原来的 $1/C_{out}$,大大降低了对部署设备计算能力的要求。

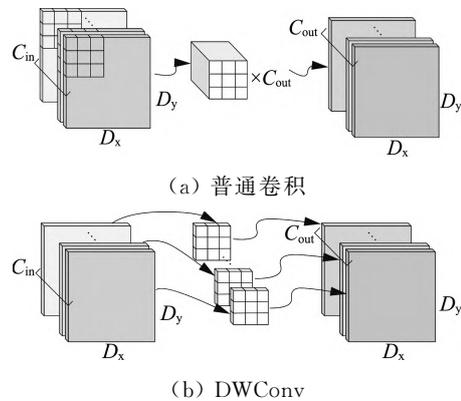


图 4 普通卷积与 DWConv 过程示意图

Fig. 4 Schematic diagram of ordinary convolution and DWConv process

2.2.2 融合 Ghostbottleneck 的 C3 模块

Ghostbottleneck 是一个即插即用的模块,其拥有基于 Ghost module^[15]搭建的残差网络结构,对传统卷积网络计算过程中特征图冗余的问题进行改进。在 Neck 网络部分的 C3 模块中引入 Ghostbottleneck。Ghost module 卷积过程如图 5 所示,首先,用步长为 1 的 1×1 卷积生成部分原始特征图,同时通道数减半;再用步长为 1 的 5×5 深度卷积逐通道生成原始特征图的“幻影”特征图来增加信道,在图 5 中表示为 $\varphi_1, \varphi_2, \dots, \varphi_k$ 等线性操作;最后,将原始特征图与恒等映射特征图拼接得到输出特征图,Identity 代表恒等映射。Ghost module 能够实现不改变输入输出特征图通道数的前提下,大幅度减小模型参数和计算量。

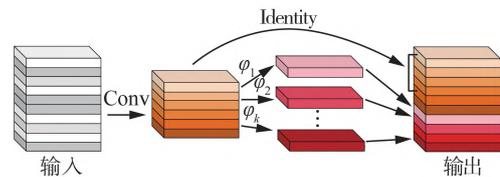


图 5 Ghost module 卷积过程

Fig. 5 Ghost module convolution process

Ghostbottleneck 借鉴 ResNet 中的残差瓶颈结构,如图 6 所示,当步长为 1 时,Ghostbottleneck 主要由两个堆叠的 Ghost module 组成,第一个 Ghost module 用作扩展层,增加通道数,第二个 Ghost module 减少通道数,用以和 shortcut 的通道数匹配,进而进行 Add 操作。当步长为 2 进行下采样时,在两个 Ghost module 之间增加一个步长为 2 的 DWConv,残差边为保持与另一路特征图尺寸和通道数一致,经过 DWConv 和 1×1 的普通卷积后与另一路输出做 Add 操作。

将 GhostBottleneck 替换 C3 模块中的 Bottleneck

成为 C3Ghost, 结构如图 7 所示, N 代表模型的深度, N 越大, 结构就越复杂。

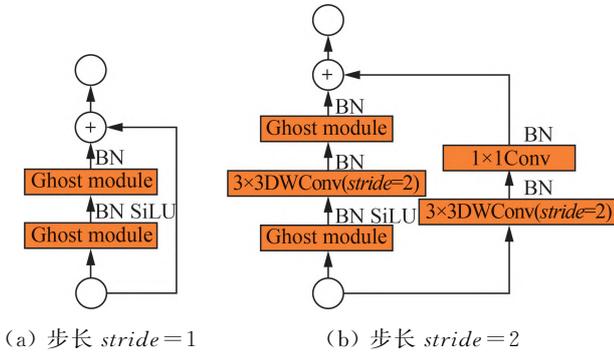


图 6 Ghostbottleneck 结构

Fig. 6 Ghostbottleneck structure

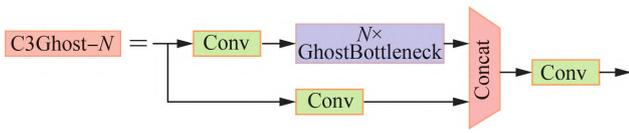


图 7 C3Ghost 模块结构

Fig. 7 C3Ghost module structure

2.2.3 引入 ECA 结构

在对 YOLOv5s 模型进行轻量化处理后, 模型的参数量和计算量减少, 其从输入的特征图提取的特征信息也会随之减少, 且由于 DWConv 只在特征图的各个通道分别卷积, 缺失了通道间的信息交互, 将导致识别精度有一定的下降。ECA 结构可以获取跨通道的信息, 提升通道间的关联性, 且只增加模型少量的参数, 就能获得较大的性能增益。ECA^[16] 结构如图 8 所示, 首先, 输入的特征图经过全局平均池化后维度由 $H \times W \times C$ 被压缩为 $1 \times 1 \times C$; 然后, 经过卷积核大小为 k 的自适应一维卷积实现跨通道信息的交互, k 的大小与通道数成正比, 由式(4)确定。

$$k = \varphi(C) = \left\lfloor \frac{\log_2 C}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}} \quad (4)$$

式中: C —— 通道数;

$\lfloor \cdot \rfloor_{\text{odd}}$ —— 取距离最近的奇数;

γ, b —— 常数, 设置为 2 和 1, 用于改变通道数 C 和卷积核大小和之间的比例。

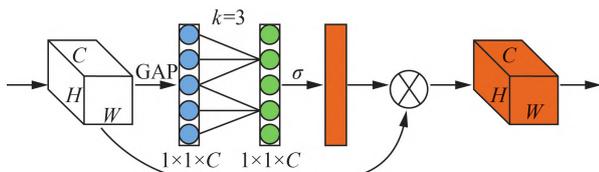


图 8 ECA 结构

Fig. 8 ECA structure

经过跨通道交互后的向量经过 Sigmoid 非线性激活函数, 得到每个通道的权值, 再与每个通道相乘, 形成最后的输出。

将 ECA 结构与 Bottleneck 结合嵌入 C3 模块中构成 C3ECA, 置于 Backbone 用于加强对输入特征图的特征信息提取能力, ECABottleneck 的结构见图 9。

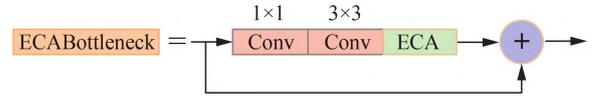


图 9 ECABottleneck 结构

Fig. 9 ECABottleneck structure

2.2.4 引入 SimAM 模块

SimAM^[17] 是中山大学在 2021 年提出的一种概念, 能够不额外增加学习参数的情况下更高效地评估特征权重, 实现目标的精确识别与定位, 提高基础模型的性能。在神经科学中, 信息丰富的神经元比周围神经元表现出更活跃的状态, 同时对周围神经元产生抑制。以此为基础, 每个神经元定义能量函数如式(5)所示。

$$\begin{cases} e_t(\omega_t, b_t, y, x_i) = (y_t - \hat{t})^2 + \frac{1}{M-1} \sum_{i=1}^{M-1} (y_o - \hat{x}_i)^2 \\ \hat{t} = \omega_t t + b_t \\ \hat{x}_i = \omega_t x_i + b_t \end{cases} \quad (5)$$

式中: t, x_i —— 目标神经元和同一通道内的其他神经元;

ω_t, b_t —— 权重、偏置项。

在式(5)中采用二值标签, 即令 $y_t = 1, y_o = -1$, 并添加正则项 $\lambda \omega_t^2$, 能量函数更新为式(6)。

$$e_t(\omega_t, b_t, y, x_i) = \frac{1}{M-1} \sum_{i=1}^{M-1} [-1 - (\omega_t x_i + b_t)]^2 + [1 - (\omega_t t + b_t)]^2 + \lambda \omega_t^2 \quad (6)$$

测量一个目标神经元与其他神经元的线性可分性可以评估该神经元的重要性, 而这等价于将上述能量函数最小化, 求出满足能量函数最小条件的解析解, 同时代入式(6)得到能量函数的最小值, 如式(7)所示。

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (7)$$

$$\text{其中 } \hat{\mu} = \frac{1}{M} \sum_{i=1}^M x_i, \hat{\sigma}^2 = \frac{1}{M} \sum_{i=1}^M (x_i - \hat{\mu})^2.$$

e_t^* 越低, 表明神经元 t 与其他神经元的区分度越高, 神经元的重要程度也越高。最后按照式(8)对特征进行增强处理。

$$\tilde{X} = \text{sigmoid}\left(\frac{1}{e_t^*}\right) \cdot X \quad (8)$$

式中: X —— 输入特征;

\tilde{X} —— 增强后的特征输出。

在网络的 Neck 部分加上 SimAM, 利用能量函数高效评估提取特征的重要性, 增强有效信息同时抑制无关特征的干扰, 提高模型定位能力与识别能力。

2.3 评价指标

采用的模型评价指标有 mAP 、单张图片推理时间 T 、 $F1$ 值、浮点运算数 $FLOPs$ 和模型体积。 T 在本地 GPU 设备上用测试集测得。相关评价指标的计算如式(9)~式(13)所示。

$$P = \frac{TP}{TP + FP} \times 100\% \quad (9)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (10)$$

$$AP = \int_0^1 P dR \quad (11)$$

$$mAP = \frac{\sum_{j=1}^N AP_j}{N} \quad (12)$$

$$F1 = 2 \times \frac{P \times R}{P + R} \quad (13)$$

式中： P 、 R ——精确率和召回率；

TP ——识别结果为真且实际也为真的数量；

FP ——识别结果为真但实际为假的数量；

FN ——识别结果为假但实际为真的数量；

AP ——单个目标类的平均精度；

mAP ——对所有目标类的 AP 求取的均值；

N ——所有目标类别的数量；

$F1$ ——精确度和召回率的调和平均数。

3 试验结果与分析

3.1 消融试验与改进模型对比试验

3.1.1 消融试验

为验证改进内容的有效性,在数据集 1 做相关消

融试验,结果如表 3 所示。

表 3 消融试验的 $mAP@0.5 : 0.95$ 对比
Tab. 3 $mAP@0.5 : 0.95$ comparison of ablation experiment

模型	DWConv+ C3Ghost	C3ECA	SimAM	$mAP@$ 0.5 : 0.95/%
	—	—	—	49.9
	✓			45.8
YOLOv5s	✓	✓		47.9
	✓		✓	46.1
	✓	✓	✓	48.4

由表 3 可知,只对原模型做轻量化后 $mAP@0.5 : 0.95$ 下降 4.1%;在 Backbone 中添加 ECA 结构后, $mAP@0.5 : 0.95$ 提高 2.1%;在 Neck 网络中添加 SimAM 模块后, $mAP@0.5 : 0.95$ 提高 0.3%;同时添加 ECA 结构和 SimAM 模块后,最终本文模型只比原模型的 $mAP@0.5 : 0.95$ 下降 1.5%。使用本次试验的模型初始化模型参数,可以在一定程度上优化接下来在数据集 2 上的试验结果。

3.1.2 改进模型对比试验

消融试验结果证明在网络模型中加入 ECA 注意力机制和 SimAM 模块的确能够提升模型的检测精度。

根据两者在网络模型中的不同位置设计 4 种不同的模型结构,在数据集 2 上做对比试验,结果如表 4 所示,由于 4 种结构的 $mAP@0.5$ 相差不大,不能体现出模型间的差距,所以选取 $mAP@0.75$ 为评价指标,从结果来看,本文模型有最高的 $mAP@0.75$ 为 71.9%,是当下条件下最优的结构。

表 4 4 种改进网络结构性能对比

Tab. 4 Performance comparison of four improved network structures

模型名称	Backbone	Neck	模型体积/MB	$FLOPs/G$	$mAP@0.75/%$
A(本文模型)	C3ECA	C3Ghost+SimAM	6.06	7.1	71.9
B	C3Ghost	C3ECA+SimAM	5.89	6.5	69.9
C	C3Ghost+SimAM	C3ECA	5.89	6.5	69.7
D	C3ECA+SimAM	C3Ghost	6.06	7.1	71.4

3.2 与原 YOLOv5s 在数据集 2 上的对比试验

在数据集 2 上的试验结果如表 5 所示,本文模型和原 YOLOv5s 相比,模型体积和计算量分别减少 55.8% 和 55.1%, $mAP@0.75$ 提高 0.1%,推理时间减少 3.5 ms。使用迁移学习策略后,与原模型相比,本文模型的 $mAP@0.75$ 提升 3.1%。图 10 为是否使用迁移学习时,模型在训练时的各损失值收敛曲线,可以发现使用迁移学习后 3 个损失值都收敛得更快更小,说明使用迁移学习相比于从头开始训练一个模型,预训练权重给

网络模型提供很好的初始化参数,能够一定程度上提升模型性能,加快训练时模型收敛速度。

表 6 为本次研究对草莓果实 5 个生长阶段检测的结果,在不比较模型间的差异时选用 $mAP@0.5$ 为衡量指标,5 个时期的 $mAP@0.5$ 和 $F1$ 值分别为 76.3% 和 72.5%、83.2% 和 81.5%、78.9% 和 77.6%、88.6% 和 84.6%、92.4% 和 89.4%,总体的 $mAP@0.5$ 和 $F1$ 值均大于 80%,基本满足草莓采摘时的精度要求。

表 5 改进方法与原模型在草莓目标检测任务中的性能对比

Tab. 5 Comparison of performance between improved method and original model in strawberry target

模型	$mAP@0.75/\%$	T/ms	$F1$ 值/ $\%$	模型体积/ MB	$FLOPs/\text{G}$
YOLOv5s	71.8	9.9	80.8	13.7	15.8
本文模型	71.9	6.4	80.6	6.06	7.1
本文模型+迁移学习	74.9	6.4	81.3	6.06	7.1

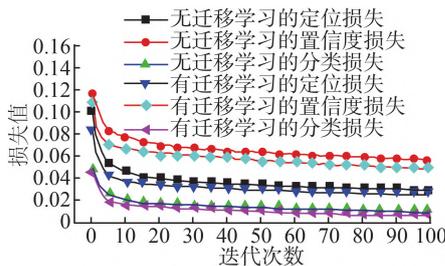


图 10 有无迁移学习 loss 值收敛过程对比

Fig. 10 Comparison of loss convergence process between transfer learning and without transfer learning

表 6 基于轻量化 YOLOv5s 网络和迁移学习的草莓目标检测结果

Tab. 6 Strawberry target detection results based on lightweight YOLOv5s network and transfer learning

生长阶段	数量/ 个	精度 $P/\%$	召回率 $R/\%$	$mAP@0.5/\%$	$F1$ 值/ $\%$
花期	238	84.4	63.5	76.3	72.5
青果期	1 592	82.6	80.4	83.2	81.5
白果期	783	76.7	78.5	78.9	77.6
转色期	642	85.5	83.8	88.6	84.6
成熟期	643	88.6	90.2	92.4	89.4
总计	3 898	83.5	79.3	83.9	81.3

3.3 与其他轻量化模型在数据集 2 上的对比试验

为进一步验证所提出模型的性能,在数据集 2 上又测试 YOLOv5m—Mobilenetv3、YOLOv5l—Shufflenetv2 和 YOLOv7—tiny 几个轻量化模型,训练迭代次数均设置为 100,其中 YOLOv5m—Mobilenetv3 和 YOLOv5l—Shufflenetv2 这两个模型是分别在 YOLOv5 模型基础上将 Backbone 替换为 Mobilenetv3—Large 和 Shufflenetv2,又调整模型的深度和宽度使模型的体积与本文模型体积相近而得到。对比结果如表 7 所示,本文模型的各项指标均最突出, $mAP @ 0.75$ 为 74.9%,比 YOLOv5l—Shufflenetv2 高出 5.9%, $F1$ 值最高为 81.3%,比 YOLOv7—tiny 高出 1.7%,平均单张图片推理时间为 6.4 ms,比 YOLOv5m—Mobilenetv3 快 16.1 ms。YOLOv7—tiny 是专门为边缘 GPU 设计的一种轻量

模型,使用 ReLU 作为激活函数,但在本次试验中表现出的性能却不如其他几个模型。

表 7 不同轻量化模型在草莓目标检测任务上的性能对比

Tab. 7 Performance comparison of different lightweight models in strawberry target detection

模型	$mAP@0.75/\%$	T/ms	$F1$ 值/ $\%$	模型体积/ MB	$FLOPs/\text{G}$
YOLOv5m—Mobilenetv3	71.2	22.5	81.1	8.15	7.2
YOLOv5l—Shufflenetv2	69.0	8.3	79.9	7.62	8.0
YOLOv7—tiny	69.1	19.1	79.6	11.7	13.1
本文模型+迁移学习	74.9	6.4	81.3	6.06	7.1

设定图片中草莓目标数目低于 5 个为稀疏,超过 15 个为密集,中间区间为适中,图 11 从左向右顺序展示了几种轻量化模型在 3 种情境的草莓图片目标检测任务中的可视化效果。在果实稀疏时每个模型均无漏检、错检,且本文模型有最高的置信度;在果实数量适中时 YOLOv5l—Shufflenetv2 和 YOLOv7—Tiny 均出现错检(图中箭头处有重复检测框);在果实密集时 YOLOv5m—Mobilenetv3、YOLOv5l—Shufflenetv2 和 YOLOv7—Tiny 均出现错检的情况,由此可以看出,本文模型在实际草莓目标检测任务中具有更好的性能。

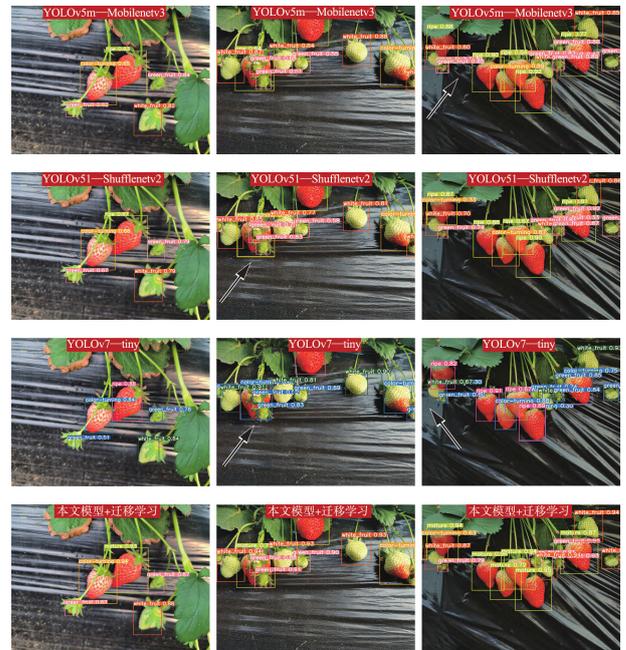


图 11 不同轻量化模型在草莓目标检测任务中的效果

Fig. 11 Effectiveness of different lightweight models in strawberry detection task

4 结论

提出一种基于 YOLOv5s 的轻量化网络模型,用

于草莓在机器采摘过程中的快速准确检测,为草莓智能化生产和管理提供相应的支持。

1) 利用深度卷积代替普通卷积,引入 Ghostbottleneck 精简 C3 模块,为加强特征图通道间的信息交互,在 Backbone 部分的 C3 模块中融合 ECA 结构,在特征融合网络加入 SimAM 模块,突出有效信息同时抑制无关信息。

2) 消融试验结果表明,改进模型的体积和计算量得到大幅度减少,但在大型目标检测数据集 PASCAL VOC 上, mAP 与原 YOLOv5s 模型相比略有下降。利用在 PASCAL VOC 数据集上训练的权重做迁移学习,加快模型训练时的收敛速度,同时进一步提高模型的 mAP 。在自制草莓数据集上进行验证,与原 YOLOv5s 模型对比,改进模型的 $mAP@0.75$ 提升 3.1%,单张图片推理时间减少 3.5 ms,与其他轻量化模型相比,在模型体积和 $FLOPs$ 相近的条件下,此模型识别精度最高、模型体积最小、检测速度最快。

参 考 文 献

- [1] 张晓慧. 草莓病害研究进展[J]. 安徽农学通报, 2018, 24(18): 52-57.
- [2] 2023—2029年中国草莓种植与深加工行业市场现状调查及投资方向研究报告[EB/OL]. https://www.chyxx.com/research/1135804.html?bd_vid=8237899329342221593, 2023-08-22.
- [3] 王卓, 王健, 王泉雄, 等. 基于改进 YOLOv4 的自然环境苹果轻量级检测方法[J]. 农业机械学报, 2022, 53(8): 294-302.
Wang Zhuo, Wang Jian, Wang Xiaoxiong, et al. Lightweight real-time apple detection method based on improved YOLOv4[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(8): 294-302.
- [4] 闫彬, 樊攀, 王美茸, 等. 基于改进 YOLOv5m 的采摘机器人苹果采摘方式实时识别[J]. 农业机械学报, 2022, 53(9): 28-38, 59.
Yan Bin, Fan Pan, Wang Meirong, et al. Real-time apple picking pattern recognition for picking robot based on improved YOLOv5m[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(9): 28-38, 59.
- [5] 宋怀波, 王亚男, 王云飞, 等. 基于 YOLOv5s 的自然场景油茶果识别方法[J]. 农业机械学报, 2022, 53(7): 234-242.
Song Huaibo, Wang Ya'nan, Wang Yunfei, et al. Camellia oleifera fruit detection in natural scene based on YOLOv5s[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(7): 234-242.
- [6] Bargoti S, Underwood J. Deep fruit detection in orchards[C]. 2017 IEEE International Conference on Robotics and Automation (ICRA), 2017: 3626-3633.
- [7] 闫建伟, 赵源, 张乐伟, 等. 改进 Faster R-CNN 自然环境下降识别刺梨果实[J]. 农业工程学报, 2019, 35(18): 143-150.
Yan Jianwei, Zhao Yuan, Zhang Lewei, et al. Recognition of rosa roxbunghii in natural environment based on improved Faster R-CNN[J]. Transactions of the Chinese Society of Agricultural Engineering, 2019, 35(18): 143-150.
- [8] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]. Computer Vision-ECCV 2016, 2016: 21-37.
- [9] 周桂红, 马帅, 梁芳芳. 基于改进 YOLOv4 模型的全景图像苹果识别[J]. 农业工程学报, 2022, 38(21): 159-168.
Zhou Guihong, Ma Shuai, Liang Fangfang. Recognition of the apple in panoramic images based on improved YOLOv4 model[J]. Transactions of the Chinese Society of Agricultural Engineering, 2022, 38(21): 159-168.
- [10] Fan Y, Zhang S, Feng K, et al. Strawberry maturity recognition algorithm combining dark channel enhancement and YOLOv5[J]. Sensors, 2022, 22(2): 419.
- [11] Fu L S, Feng Y L, Wu J Z, et al. Fast and accurate detection of kiwifruit in orchard using improved YOLOv3-tiny model[J]. Precision Agriculture, 2021, 22(3): 754-776.
- [12] 孙俊, 陈义德, 周鑫, 等. 快速精准识别棚内草莓的改进 YOLOv4-Tiny 模型[J]. 农业工程学报, 2022, 38(18): 195-203.
Sun Jun, Chen Yide, Zhou Xin, et al. Fast and accurate recognition of the strawberries in greenhouse based on improved YOLOv4-Tiny model[J]. Transactions of the Chinese Society of Agricultural Engineering, 2022, 38(18): 195-203.
- [13] 陈仁凡, 谢知, 林晨. 基于 YOLO-ODM 的温室草莓成熟度的快速检测[J]. 华中农业大学学报, 2023, 42(4): 262-269.
- [14] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [15] 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 (CVPR), 2020: 1577-1586.
- [16] Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 11531-11539.
- [17] Yang L, Zhang RY, Li L, et al. SimAM: A simple, parameter-free attention module for convolutional neural networks[C]. Proceedings of the 38th International Conference on Machine Learning, 2021: 11863-11874.