

基于改进 YOLO v8n 的轻量化茶叶嫩芽检测模型

张澎涛 于 帅 刘大洋 刘 振 张宪奇

(东北林业大学计算机与控制工程学院, 哈尔滨 150040)

摘要: 在智能采收名优茶过程中, 现有的目标检测算法存在对茶叶嫩芽检测精确率不高和推理速度较慢的问题, 导致在边缘计算设备上部署困难, 为此本文提出一种基于改进 YOLO v8n 的轻量化茶芽检测模型 YOLO-RET。通过引入重新设计的轻量化特征提取模块 (RGCSPELAN 模块) 和改进的多尺度特征融合金字塔结构 (EMBSFPN 结构), 在大幅降低模型参数量的前提下, 提升了对目标特征信息的提取与融合能力。引入新的损失函数 Focaler-IoU, 解决了样本分布不均衡问题, 进一步提高了模型检测精度和鲁棒性。试验结果表明, 与 YOLO v5n、YOLO v8n、YOLO v8s 和 YOLO v10n 相比, YOLO-RET 精确率分别提高 3.2、3.0、0.8、4.3 个百分点, 平均精度均值分别提高 2.7、3.2、0.4、3.1 个百分点。此外, YOLO-RET 模型参数量和浮点运算量相较原 YOLO v8n 模型分别降低 43% 和 2.2×10^9 。将算法模型移植到 ATK-DLRK3568 开发板上, 并进行量化处理以优化部署, 降低了对硬件资源的需求。YOLO-RET 模型在保持高识别精确率的同时, 提升了检测效率, 为在边缘计算端部署实时目标检测提供了一种高效、准确的解决方案。

关键词: 茶叶嫩芽检测; 目标检测; YOLO v8n; 轻量化

中图分类号: S24

文献标识码: A

文章编号: 1000-1298(2026)08-0214-12

OSID:



Lightweight Tea Bud Detection Model Based on Improved YOLO v8n

ZHANG Pengtao YU Shuai LIU Dayang LIU Zhen ZHANG Xianqi

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

Abstract: During the intelligent harvesting of premium tea, existing object detection algorithms face challenges of insufficient detection precision for tender tea shoots and slow inference speed, making deployment on edge computing devices difficult. To address these issues, YOLO-RET, a lightweight tea shoot detection model was proposed based on an improved YOLO v8n architecture. By introducing a redesigned lightweight feature extraction module (RGCSPELAN module) and an enhanced multi-scale feature fusion pyramid structure (EMBSFPN), the model significantly enhanced feature extraction and fusion capabilities while substantially reducing model parameters. Additionally, a novel loss function, Focaler-IoU, was incorporated to address sample distribution imbalance, further improving detection accuracy and robustness. Experimental results demonstrated that compared with YOLO v5n, YOLO v8n, YOLO v8s, and YOLO v10n, YOLO-RET achieved precision improvements of 3.2, 3.0, 0.8, and 4.3 percentage points respectively, with corresponding mAP@0.5 improvements of 2.7, 3.2, 0.4, and 3.1 percentage points. Furthermore, YOLO-RET reduced the parameter count by 43% and computational complexity by 2.2×10^9 compared with that of the original YOLO v8n. The algorithm was successfully deployed on an ATK-DLRK3568 development board with quantization optimization, which not only lowered hardware resource requirements but also maintained high recognition accuracy while enhancing inference efficiency and reducing the computational burden on edge devices. The research result can provide an efficient and accurate solution for real-time object detection deployment on edge computing platforms.

Key words: tea bud detection; object detection; YOLO v8n; lightweight

收稿日期: 2025-01-03 修回日期: 2025-03-03

基金项目: 国家自然科学基金项目(32202147)

作者简介: 张澎涛(1980—), 男, 副教授, 博士, 主要从事计算机视觉、智能检测与故障诊断技术研究, E-mail: zpt@nefu.edu.cn

通信作者: 刘大洋(1990—), 男, 副教授, 博士, 主要从事模式识别与智能系统、智能化检测技术研究, E-mail: liudayang@nefu.edu.cn

0 引言

茶作为全球消费量最大的天然饮品之一^[1],其市场正从传统形态向多元化衍生品持续拓展。茶叶可分为大宗茶和名优茶,名优茶采摘要求一般为茶叶的单芽或一芽一叶,大宗茶要求一般是一芽一叶或一芽多叶。人工采摘茶叶时,可基于形态特征实现嫩芽与嫩叶的精准区分,实施分级采摘。但人工作业效率低且成本逐年增高,不利于茶叶产业的可持续发展^[2]。目前,茶叶采摘技术与装备主要有基于往复式切割原理的大宗茶采收装备和基于芽叶智能识别的选择性名优茶采收装备^[3]。前者虽具有较高采收效率,但难以获得规格均一的原料,适应性受限。而后者对名优茶芽叶采摘位置精确率要求更高。茶产业正加速融合大数据、物联网、人工智能等新一代信息技术,通过数字化精准调控推动加工环节智能化转型,有效提升全产业链效率与产品附加值,为高质量发展注入新动能^[4]。因此,为了降低名优茶采摘过程的成本,并解决现有智能采摘设备茶叶识别精确率不足问题,研发先进的选择性智能采摘设备变得至关重要。研究的核心挑战在于如何精确地检测茶叶嫩梢,以确保采摘过程既高效又具有选择性。

目前茶叶嫩芽识别方法主要有基于颜色特征的传统图像处理与识别法、基于机器学习的识别方法和基于深度学习的识别方法^[5]。当前,茶叶嫩芽识别技术的研究正逐步从基于颜色特征的传统图像处理方法转向结合机器学习和深度学习的高级视觉检测系统。这些系统利用 RGB-D 多模态特征融合、区域亮度自适应校正以及超绿特征等技术进行精准的嫩芽检测。例如,WU 等^[6]提出采用 RGB-D 多模态特征融合进行茶叶嫩芽检测方法,利用红、绿、蓝、深(RGB-D)等多模态特征进行识别,增强了对茶叶嫩芽的识别能力;吕军等^[7]提出一种基于区域亮度自适应校正的茶叶嫩芽检测模型。对不同时间采集的茶叶图像进行灰度化,计算其平均灰度值,对平均灰度值偏离预设阈值的图像实施分块处理和局部区域伽马亮度自适应校正,以获取更为清晰的特征信息;在受控试验环境下展现出较高识别精确率,但在实际应用中,茶田环境的复杂性和多变性可能影响图像特征提取的有效性,进而影响到茶叶嫩芽识别的准确性。

在基于机器学习和基于深度学习的识别方法方面,通过引入机器学习算法,利用颜色、纹理和形态等多种特征信息,避免老叶、茶梗等元素对茶叶嫩芽图像识别或分割干扰^[8]。WANI 等^[9]使用 Faster R-

CNN 检测茶叶采摘区域,实现了 79% 的准确率和 90% 的召回率。朱红春等^[10]提出基于 Faster R-CNN 的目标检测算法,通过采集 72 幅茶园图像,利用 Faster R-CNN 与 VGG16 特征提取网络进行模型训练与测试,在检测精度和速度上均显著优于传统方法。虽然两阶段模型的准确性更高,但受限于移动端处理器的计算性能,难以直接应用于实际工程环境中。因此,在确保茶叶嫩芽识别准确率的前提下,对现有深度学习算法进行轻量化改进,并将其部署到移动端上,这一研究对于推动算法部署落地具有显著价值。为此,BAI 等^[11]基于 YOLO v5 网络,通过引入轻量模块和高效特征提取模块,修剪了模型的 Neck 网络,实现了模型的轻量化改进。王梦妮等^[12]研究对 YOLO v5s 在骨干网络中替换为空洞空间卷积池化金字塔结构,在模型 Neck 引入加权双向特征金字塔网络和卷积块注意力模块。黄家才等^[13]在 YOLO v4 网络模型的主干网络与 Neck 进行了改进,使用了 GhostNet 网络作为模型主干,并且使用 Ghost 卷积在 Neck 网络中进行卷积运算,在保持了一定精确率同时,进一步降低了模型的整体计算量与参数。ZHANG 等^[14]将 YOLO v4 结构中 CSPDarkNet53 替换为 MobileNet V2,使得改进后的网络参数量和模型内存占用量仅为原来的 16.76%、19.53%。LI 等^[15]在 YOLO v4 中引入 K-means 聚类算法和卷积注意力机制模块(CBAM),进行通道剪枝,与原模型相比,参数量和模型内存占用量分别下降 34% 和 47.6%。ZHANG 等^[16]提出一种基于 YOLO v4 的茶叶冠芽全天轻量级检测模型,将 YOLO v4 网络的主干和所有卷积更换为 MobileNetV3 轻量级神经网络和深度可分离卷积,改进后模型内存占用量为 11.78 MB,为原模型的 18.30%,检测速度提高 11.68 f/s,且检测精确率也有一定提高。

上述轻量化模型设计的核心在于采用高效的网络架构(如 MobileNet 系列、GhostNet 等),通过深度可分离卷积和 Ghost 模块减少参数量与计算量。同时,通过改进特征金字塔网络(FPN)架构,采用加权双向特征融合策略及空洞空间金字塔池化(ASPP)结构,在保证计算效率的前提下增强多尺度特征表征能力。此外,通过通道剪枝技术消除冗余参数,结合量化压缩与知识蒸馏方法,构建紧凑型网络结构。引入注意力机制(如 CBAM 模块),通过动态特征校准提升关键区域的特征响应,在可控计算量下优化检测精度。最后,针对 YOLO 架构的 Neck 部分通过移除冗余层等方式进行结构优化,减少整体计算量的同时不影响最终输出质量。上述改进措施旨在确

保模型在资源受限环境下高效运行的同时保持或提升性能。

综上所述,随着网络深度和模型参数量级的增加,在计算资源受限的可移动设备上部署深度学习模型面临严峻挑战。此外,茶叶生长环境存在显著复杂性,动态光照条件、复杂背景干扰及嫩芽形态的时变特征,共同构成识别难点。对基于图像处理的茶叶识别与检测技术而言,采取多信息融合方法获得更全面的茶叶嫩芽信息,可减少识别率低的现象发生^[17]。为解决上述问题,本文提出轻量化茶叶嫩芽检测模型 YOLO-RET,通过设计 RGCSPELAN 特征提取模块和 EMBSFPN 多尺度特征融合结构,大幅降低模型参数量并增强特征表达能力;引入 Focaler-IoU 损失函数解决样本分布不均衡问题;并在边缘计算设备上部署进行验证,为名优茶智能采摘系统提供可靠技术支持。

1 数据采集与增广

以单芽茶叶嫩芽为研究对象,聚焦名优茶识别场景的实际需求。图像数据采集于 2024 年 4 月浙江省丽水市松阳县裕溪乡青田圩茶园(北纬28°21',东经 119°40'),该时段对应清明节前后茶农采摘特

级/一级单芽名优茶的生产周期,该研究具有显著经济价值与产业指导意义。

为了丰富图像数据集多样性,采用智能手机(VIVO Y31S 型)采集视频数据,覆盖 3 种气象条件(阴天、雨天、晴天)及不同时段光照强度(低光:07:30—08:30/17:30—18:30;中高光:09:00—17:00)。通过多角度(俯视/侧视/远视/近视)、多背景复杂度(修剪/未修剪茶树)的拍摄策略,构建包含 1 920 像素×1 080 像素视频的原始数据集。基于 Python 脚本对视频进行帧采样,经质量筛选(排除模糊、重复及目标偏移图像)保留 1 863 幅原始图像。预处理阶段保留图像原始尺寸,模型训练时通过 YOLO v8 自适应图像缩放算法统一图像尺寸 640 像素×640 像素。经过筛选和预处理,共获得 1 863 幅高质量的茶叶嫩芽图像,构成本文数据集,如图 1 所示,第 1 行依次为雨后侧视、晴天侧视、雨后俯视、晴天俯视、背景简单和复杂背景未修剪情况下的茶叶嫩芽图像样本。如图 2 所示,采用 Labelme 工具对图像进行人工标注,生成 JSON 格式标注文件并转换为 YOLO v8 标准 TXT 格式。最终,获得总计 1 863 幅茶叶嫩芽图像,按比例 4:1 随机划分为训练集(1 491 幅)和验证集(372 幅),确保分布一致性。



图 1 不同场景茶芽图像
Fig. 1 Images of tea bud in different scenes

鉴于数据集规模较小,为了增强模型训练效果,对训练集实施 4 种数据增强操作,随机翻转、高斯模糊、亮度与对比度调整,以及自适应直方图均衡化

(图 1g~1l)。这些增强方法能够在训练过程中增加样本的多样性,从而提升模型的鲁棒性和适应性。通过增强,训练集图像扩展至 5 964 幅,扩充后数据



图 2 图像目标标记
Fig. 2 Images annotation

集被用于新设计模型的训练。

2 茶叶嫩芽检测算法及改进

2.1 YOLO v8 算法

YOLO v8^[18] 网络模型继承 YOLO 系列的单阶段目标检测方法,在继承的基础上进行了显著的优化和改进,以提高检测速度和精确率,源代码已在 GitHub 开源^[19]。YOLO v8 网络是通过单次前向传

播预测图像中所有目标的类别和位置,将目标检测任务视为一个回归问题,直接预测边界框和类别概率。

网络架构方面,YOLO v8 采用轻量级设计,减少了冗余计算。其 Backbone 部分基于 YOLO v7^[20] 的设计理念,Neck 集成路径聚合网络 (PANet),而 Head 引入解耦头设计,将分类和检测分离,增强了模型性能。损失函数方面,YOLO v8 使用多任务损失函数,结合分类和定位损失,并引入 IoU 损失处理重叠目标。不过,当预测框与真实框不相交时,IoU 损失无法计算梯度,因为它仅衡量重叠程度,而不能反映距离或长宽比差异。

2.2 茶叶嫩芽识别的 YOLO v8 改进

近年来,基于 YOLO 系列模型改进实时目标检测方法已成为目标检测领域的主流技术方案。鉴于茶叶嫩芽检测的特点和难度,在 YOLO v8n 模型基础上进行结构调整,提出 YOLO - RET 茶叶嫩芽检测模型,具体网络结构见图 3。

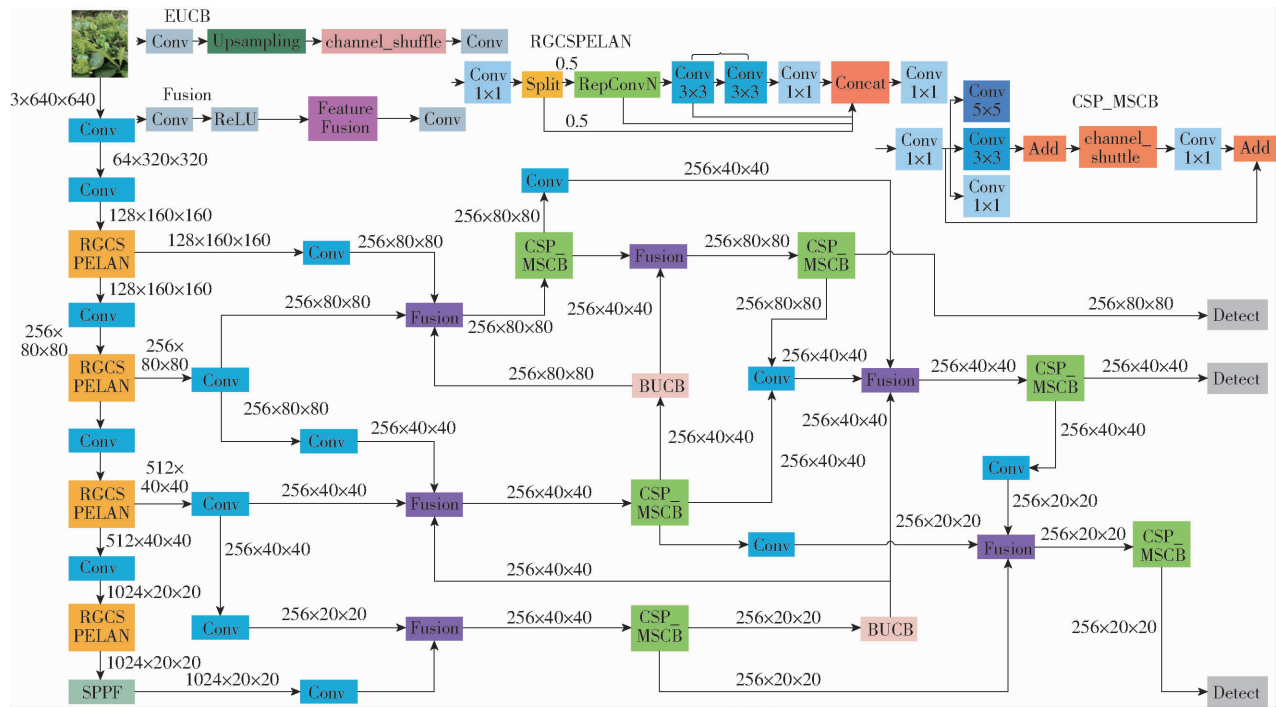


图 3 YOLO - RET 网络结构
Fig. 3 YOLO - RET network structure

2.2.1 RGCSPELAN 模块设计

在 GhostNet^[21] 的研究中,主流的卷积神经网络 (CNN) 计算,输出的特征图存在大量相似情况,表明特征映射有广泛的冗余。受 GhostNet 理念启发,并考虑到本研究面向复杂茶园场景,在同一视角下前景与背景中可能同时存在茶芽检测目标,且前后景不同会导致目标大小变化,从而增加识别难度。为更好地在复杂背景下提取识别对象的特征,本文旨在通过减轻模型参数量和计算量的同时增强特征

提取能力。为此,设计并引入新模块——RGCSPELAN,用于替代模型主干网络中的 C2f 模块,通过特征重校准在降低参数数量的同时增强多尺度特征提取能力,网络结构如图 4 所示。

然而,针对移除 C2f 模块带来的性能衰减,在梯度流通的关键路径上引入 RepConvN 模块,不仅改善梯度传递效果,确保训练过程的稳定性和效率,其多分支并行卷积架构还可提升特征融合效率,有助于进一步优化特征表示,提升模型的整体表现。

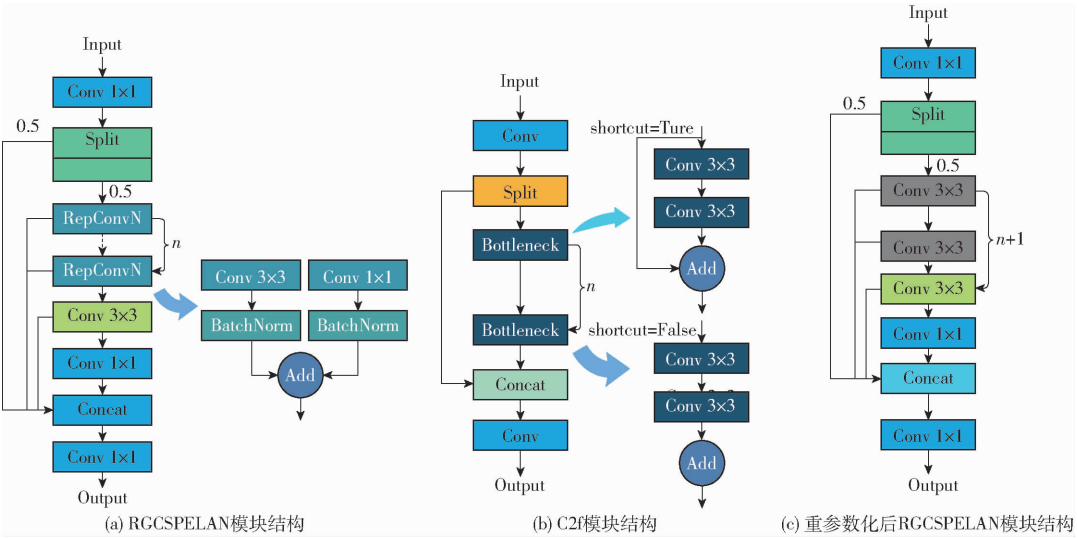


图 4 RGCSPELAN 模块网络结构

Fig. 4 Network architecture of RGCSPELAN module

RepConvN 模块是 RepConv 模块的一种变体,去掉恒等连接(即残差连接)。RepConvN 模块在 YOLO v7 和 YOLO v9^[22] 模型中均有应用,其内部数据流程如图 5 所示。

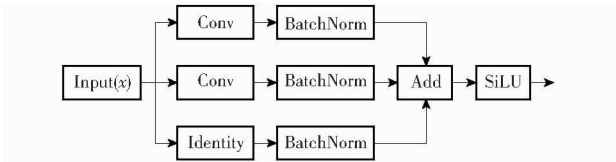


图 5 RepConvN 模块结构

Fig. 5 Architecture of RepConvN module

当恒等连接(Identity connection)不存在时,Identity 层权重与偏置被初始化为零张量,此时模块即为 RepConvN。相反,当 Identity 层权重与偏置参与融合时,模块则为 RepConv。通过融合 2 个卷积层的权重和偏置,能够构造出数学等效的单一卷积核,在推理阶段直接替代原始多分支结构,从而实现重参数化(Reparameterization)。重参数化技术通过将训练阶段的多路径结构转换为推理阶段的单路径等效结构,有效提升模型训练稳定性和泛化能力。

在深度学习中,为提高模型的训练稳定性和性能,通常会结合使用卷积层和批量归一化层(Batch normalization layer, BN)。在推理阶段,这 2 个层可以进行融合以减少计算量和内存占用量。

对于 1 个卷积层,计算式为

$$\text{Conv}(x) = W(x) + b \tag{1}$$

式中 W ——卷积核权重

b ——偏置

对于 1 个 BN 层,计算式为

$$\text{BN}(x) = \gamma \frac{x - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta \tag{2}$$

式中 γ, β ——学习参数

μ ——均值

σ^2 ——方差

ε ——微小常数,防止除零

将卷积层输出输入到 BN 层,可得

$$\text{BN}(\text{Conv}(x)) = \gamma \frac{W(x) + b - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta \tag{3}$$

化简后得

$$\text{BN}(\text{Conv}(x)) = \frac{\gamma W(x)}{\sqrt{\sigma^2 + \varepsilon}} + \left[\frac{\gamma(b - \mu)}{\sqrt{\sigma^2 + \varepsilon}} + \beta \right] \tag{4}$$

定义新的权重 W_{new} 和偏置 b_{new} 为

$$\begin{cases} W_{\text{new}} = \frac{\gamma W}{\sqrt{\sigma^2 + \varepsilon}} \\ b_{\text{new}} = \frac{\gamma(b - \mu)}{\sqrt{\sigma^2 + \varepsilon}} + \beta \end{cases} \tag{5}$$

将 2 层融合,得到 1 个新的卷积层为

$$\text{Conv}(x)_{\text{new}} = W(x)_{\text{new}} + b_{\text{new}} \tag{6}$$

为了满足轻量化的设计目标,以灵活控制模块的规模。如图 4a 所示,在 Split 模块中引入可调节的通道缩放因子 α (默认设置 0.5),将特征图按比例分为 2 路。主路进行后续特征提取,支路通过残差连接将原始输入特征与处理后特征在 Concat 层融合。该操作可以有效调节模块内部组件的宽度,从而直接影响整个模型参数数量和计算复杂度。

当任务性能达标时,降低 α 值可缩减模型参数量与计算量,实现轻量化部署且保持性能稳定。提升 α 值可扩展模型表征能力,为精度优化提供可调节的容量空间。由图 4 可知,RGCSPELAN 模块在 2 项关键处进行了优化,重新设计了 Split 机制,通过调整缩放因子参数,减少处理的信息流;将原 C2f 模块中 n 个串联的 Bottleneck 单元替换为 RepConvN 单元,将原串联的 2 个 3×3 卷积重构为 1 个 3×3

卷积和 1 个 1×1 卷积并行处理,降低计算量。尽管并行处理可能会对运行速度产生一定影响,但在推理阶段,RepConvN 通过结构重参数化技术等等效于单卷积层,实现计算效率优化。具体而言,原 C2f 模块中,存在 $2n$ 个 3×3 卷积实例,而改进后 RGCSPELAN 模块仅包含 $(n + 1)$ 个 3×3 卷积实例,如图 4c 所示。RGCSPELAN 模块通过结构优化,在降低计算复杂度的同时进一步压缩了模型参数量。

2.2.2 CSP_MSCB 模块设计

根据 Trident^[23] 网络的理论分析,目标尺度与有效感受野存在强相关性。为此,在 FPN 阶段设计针对对不同尺度特征层的不同尺寸卷积核,以匹配茶树芽点目标的尺度分布特性。

基于 Shuffle 的逐点分组卷积和通道重排操作的核心理念及其轻量化特性,设计 CSP_MSCB 模块。通过点卷积层扩展通道数,应用多尺度卷积与通道重排操作,再通过点卷积层压缩通道数。其中,分组卷积使用不同大小的卷积核 (1×1 、 3×3 和 5×5) 进行卷积操作,以提取不同尺度下的目标特征。分组卷积后,设置通道重排操作,提高计算效率的同时,还促进不同组特征图之间的交互,打破组间特征正交性,使得各组之间的信息流通更加顺畅,有助于提升模型表达能力和最终性能。

2.2.3 网络 Neck 结构设计与优化

原 YOLO v8 网络模型的 Neck 部分在特征融合时存在一定问题,尽管引入特征金字塔 (FPN)^[24] 和路径聚合网络 (PANet)^[25] 在一定程度上缓解了这一问题。然而,传统 Neck 结构如 FPN 及其变体,在特征信息传递过程产生梯度衰减,阻碍了 YOLO 系列网络进行更有效的信息融合^[26]。与传统 FPN 相比,重复加权双向特征金字塔网络 (BiFPN)^[27] 不仅在信息传递上更为全面,而且在网络结构上更加简洁高效。虽然 PANet 和 BiFPN 均采用双向路径设计,但 BiFPN 通过减少冗余计算和增加跳跃连接等方式进一步优化了网络结构,使得特征融合更加高效。此外,BiFPN 引入加权特征融合机制,解决了不同尺度特征贡献度不一的问题。

借鉴 MAF-YOLO^[28] 的多分支特征金字塔结构和 EMCAD^[29] 的跨尺度连接部分设计,构建高效多分支尺度特征金字塔网络 (Efficient multi-branch & scale FPN,EMBSFPN)。通过提取主干网络 P2 层浅层特征,增强小目标检测的特征表达能力,使网络更好地适应不同大小的目标或特征,丰富模型对多尺度上下文信息的理解,从而获得更加全面和准确的多尺度感知场信息。通过这种方式,模型展现出更强的鲁棒性和更高的精确率。模型 Neck 部分的信

息传递路径结构如图 6 所示。

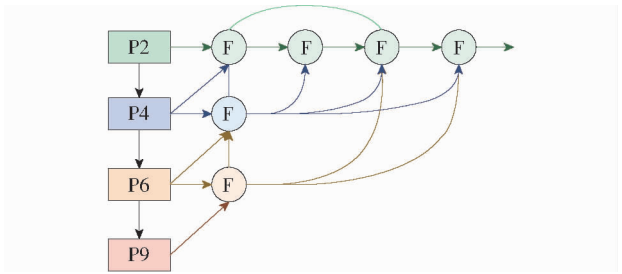


图 6 模型 Neck 部网络信息传递路径结构

Fig. 6 Information flow path structure in model neck

为实现模型轻量化目标,并受到 BiFPN 中多尺度特征加权融合策略的启发,将传统的特征拼接操作替换为对应元素求和操作,此外,采用高效上采样模块 (Efficient upsampling convolution block,EUCB)。通过结合上采样、深度可分离卷积、通道混洗 (Channel shuffle) 及点卷积操作,在保证检测精度的前提下维持高效的运算性能。EUCB 模块能够有效地放大特征图的尺寸,保持甚至增强特征的表达能力。

2.2.4 损失函数改进

在目标检测任务中,类别不平衡问题是普遍存在于计算机视觉领域的问题。在名优茶嫩芽等单一类别目标检测中,类别不平衡主要体现在前景 (目标对象) 与背景之间的比例失衡上。大多数区域都是背景,会导致模型过度拟合到背景特征,从而降低对实际目标的检测能力。CUI 等^[30] 研究了在类别不平衡的数据集中,某些类别的样本更容易被模型正确分类 (简单样本),而其他类别则包含更多难以分类的样本 (困难样本)。此外,SHRIVASTAVA 等^[31] 和 LIN 等^[32] 的研究强调困难样本对于模型训练的重要性,表明通过针对性地加强这些样本的学习,可以显著提升检测器的性能。

YOLO v8 网络使用 CIoU (Complete intersection over union) 损失函数优化边界框回归任务,旨在更准确地衡量目标检测框的匹配程度。尽管 CIoU 损失函数在多维度上提高了边界框回归的准确性,但在处理前景与背景之间的比例失衡问题上,并未进行专门优化。为此,采用 Focaler-IoU^[33] 损失函数替代 CIoU。通过线性区间映射的方法重新参数化 IoU 损失,从而更加聚焦于不同难度的回归样本。这种方法旨在解决目标检测中样本不均衡的问题,特别是在困难样本和简单样本在边界框回归中的分布对回归结果的影响。同时,Focaler-IoU 通过动态调整样本权重,使模型更加关注对定位精确率影响较大的样本,从而提升整体检测性能。

损失函数计算式为

$$IoU_{focaler} = \begin{cases} 0 & (IoU < d) \\ \frac{IoU - d}{u - d} & (d \leq IoU \leq u) \\ 1 & (IoU > u) \end{cases} \quad (7)$$

式中 $IoU_{focaler}$ ——重构后的 Focaler-IoU 值
 IoU ——原始的 IoU 值
 $d、u$ ——可调节阈值参数,取 $[0,1]$

使用线性区间映射的方法调整 IoU 损失,使其能够更加关注特定范围内的样本。通过调整 d 和 u 值,便可使 Focaler 关注焦点不同的回归样本,可以改变损失函数的焦点。损失函数定义为

$$L_{Focaler-IoU} = 1 - IoU_{focaler} \quad (8)$$

3 模型训练与测试

3.1 模型训练环境

使用数据集和基于改进 YOLO v8n 网络模型对茶叶嫩芽检测进行模型训练与验证。试验平台配置硬件为 Intel Core i7 - 13700K, 128 GB DDR5, NVIDIA RTX 4090 (24 GB); 软件为 Ubuntu 20.04 LTS, Python 3.9, PyTorch 2.3.0 + cu181。训练参数: 模型输入图像尺寸为 640 像素 × 640 像素, 批次大小 64, 模型初始学习率 lr0 为 0.001, 最终学习率 lrf 为 0.000 1, 权重衰减系数为 0.000 5, 动量为 0.937,

训练轮次为 300, 同时设置早停机机制训练轮数为 50, 使用自适应动量的随机优化算法 (Adam) 作为模型优化器, 开启马赛克数据增强功能。

3.2 评价指标

为了衡量模型有效性, 采用精确率 P 、召回率 R 、50% IoU 阈值平均精度均值 mAP@ 0.5、50% 到 95% IoU 阈值平均精度均值 mAP@ 0.5:0.95 和实时检测速度 FPS 作为评价指标, 实现对模型检测准确度、运行速度以及计算资源利用效率的综合考量。

图像实时处理总时间主要由平均预处理时间、平均推理时间和平均非极大值抑制时间组成。

3.3 模型训练结果与分析

3.3.1 消融试验

为了进一步验证每种改进方法对模型训练稳定性和有效性的影响, 进行消融试验设计。通过排列组合 3 种改进方法, 共设计 5 组试验, 所有试验数据均基于自建数据集的训练集和验证集进行训练和测试所得。为减少数据分布不平衡导致的训练偏差, 确保训练结果的正确性和稳定性, 每次试验使用相同样本集, 并重新按比例随机划分为训练集和测试集, 对使用 IoU 改进前后进行重复试验, 采用多次重复及动态划分方法减少因数据分割方式不同而引起的偏差, 提高结论的可信度。试验结果见表 1。

表 1 模型消融试验结果
Tab.1 Ablation test results

EMBSFPN	RGCSPELAN	IoU	模型参数量	浮点运算量	精确率/ %	召回率/ %	mAP@ 0.5/ %	mAP@ 0.5:0.95/ %	检测速度/ (f·s ⁻¹)	模型内存占 用量/MB
✓	✓	✓	3.005 × 10 ⁶	8.1 × 10 ⁹	77.3	76.4	81.9	46.2	322.58	5.93
			2.127 × 10 ⁶	7.1 × 10 ⁹	78.6	77.8	84.4	48.6	246.81	4.36
			2.197 × 10 ⁶	6.0 × 10 ⁹	76.6	80.0	83.7	47.3	372.74	4.45
			3.005 × 10 ⁶	8.1 × 10 ⁹	76.4	80.7	83.2	47.1	324.51	5.93
					78.8	79.1	84.7	48.7	218.86	3.59
✓	✓		1.710 × 10 ⁶	5.9 × 10 ⁹	76.5	81.1	84.9	48.8	227.67	3.59
					77.9	78.9	84.7	48.8	231.44	3.59
					77.7	79.5	85.0	48.7	222.60	3.59
					78.2	80.0	85.3	48.8	229.32	3.59
					79.1	79.1	85.2	48.7	232.56	3.59
✓	✓	✓	1.710 × 10 ⁶	5.9 × 10 ⁹	79.2	80.7	85.2	48.1	234.16	3.59
					80.3	79.2	85.1	48.3	222.86	3.59

注: ✓表示加入模块。

由表 1 可知, 与原模型相比, 仅引入单一改进的模型在召回率、mAP@ 0.5 和 mAP@ 0.5:0.95 均有提高, 但在识别精确率均值上略有波动。原模型的内存占用量 (5.93 MB) 与浮点运算量 (8.1 × 10⁹) 及检测速度都大于增加 RGCSPELAN 模块的模型 (4.45 MB, 6.0 × 10⁹)。当改进数量增加到 2 个时, 在召回率、mAP@ 0.5 和 mAP@ 0.5:0.95 上相较于

原模型均有提升。尽管改进后模型浮点运算量下降 2.2 × 10⁹, 模型内存占用量减少 39.46%, 但由于引入 EMBSFPN 使得模型架构由原 21 层增加至 32 层, 导致在模型内部增加了更多计算步骤。而且大量采用多尺度卷积和 RepConv 的结构性约束, 提升了每个步骤所需的并行度, 使处理单元效率下降。最终改进后的模型在所有指标上相比于原模型均有显著

提升。以改进后模型最差训练结果作为与原模型对比对象,在精确率、召回率、mAP@ 0.5 和 mAP@ 0.5:0.95 4 个指标上分别提升 3、2.8、3.2、2.1 个百分点。模型的参数量降低 43%,浮点运算量减少 2.2×10^9 。经优化后,在特定场景下检测速度出现衰减的情况下,模型整体准确率仍获得显著提升。

3.3.2 不同检测模型性能试验

将 YOLO-RET 网络与几种主流检测网络模型 (YOLO v5n、YOLO v7、YOLO v8n、YOLO v8s、YOLO

v10n、RT-DETR-L、Faster R-CNN-R50 和 EfficientNet) 作为对比进行测试,后 2 组模型基于 MMDetection3.1^[34] 开源项目实现,试验采用与 YOLO-RET 保持相同硬件配置训练环境,关键训练超参数保持统一配置。其中 RTDETR-L 因显存限制,批次大小设为 16 (其他模型为 32),试验结果如图 7 所示。由于设置了早停机制 (patience),各曲线长度存在差异。不同模型测试结果如表 2 所示。

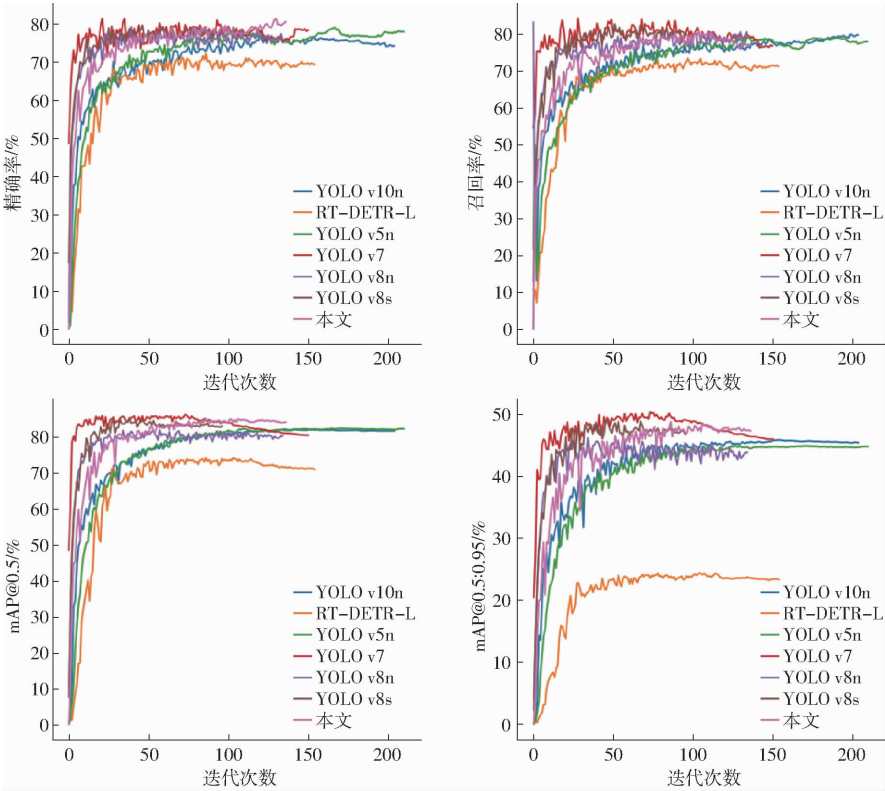


图 7 不同模型训练结果
Fig. 7 Training results comparison curves

表 2 不同模型试验结果

Tab. 2 Experimental results of different models								
模型	模型参数量	浮点运算量	精确率/ %	召回率/ %	mAP@ 0.5/ %	mAP@ 0.5:0.95/ %	检测速度/ (f·s ⁻¹)	模型内存占 用量/MB
YOLO v5n	2.503×10^6	7.10×10^9	77.1	77.9	82.4	44.9	312.50	9.85
YOLO v7	3.648×10^7	1.03×10^{11}	76.1	81.7	85.9	48.5	58.10	71.30
YOLO v8n	3.005×10^6	8.10×10^9	77.3	76.4	81.9	46.2	322.58	5.93
YOLO v8s	1.113×10^7	2.84×10^{10}	79.5	79.6	84.7	49.0	285.71	21.40
YOLO v10n	2.265×10^6	6.50×10^9	76.0	77.2	82.0	45.8	333.40	5.47
EfficientNet	1.998×10^7	7.64×10^{10}	74.8	73.5	81.7	44.5	97.70	15.80
Faster R-CNN-R50	4.112×10^7	9.10×10^{10}	77.1	78.3	84.0	45.7	77.60	107.10
RT-DETR-L	2.845×10^7	1.01×10^{11}	69.0	72.7	74.0	24.3	72.99	56.30
YOLO-RET	1.710×10^6	5.90×10^9	80.3	79.2	85.1	48.3	232.56	3.59

由图 7 可知,在自然环境中对茶叶嫩芽进行目标检测时,RT-DETR-L 网络的各项指标相对最低,表现出较高的误识别率和漏检概率,尤其对较小

或与背景相似度高的低对比度样本。YOLO v5n、YOLO v8n 和 YOLO v10n 模型在识别精确率和漏检率方面还有提升空间。相比之下,YOLO v7 和

YOLO v8s 由于网络结构更宽更深,在检测精确率上与 YOLO-RET 模型表现相近,但参数数量和计算复杂度显著高于改进的 YOLO-RET 模型。YOLO-RET 模型不仅在训练过程中表现出更小的波动,而且其训练时间也有所减少,表明模型的改进有效提升了精确率和训练稳定性。

由表 2 可知,YOLO v7 模型凭借其更宽更深的网络结构,在召回率、mAP@ 0.5 和 mAP@ 0.5:0.95 上分别比 YOLO-RET 模型高 2.5、0.8、0.2 个百分点。然而,YOLO-RET 模型的大小仅为 YOLO v7 的 5%。改进的 YOLO-RET 模型相比于 YOLO v5n、YOLO v8n、YOLO v8s 和 YOLO v10n,在精确率上分别提高 3.2、3.0、0.8、4.3 个百分点,在 mAP@ 0.5 上分别高 2.7、3.2、0.4、3.1 个百分点。在召回率方面,除略低于 YOLO v8s 0.4 个百分点外,比另外 3 个模型依次高 1.3、2.8、2.0 个百分点;同样在

mAP@ 0.5:0.95 上,除略低于 YOLO v8s 0.7 个百分点外,比另外 3 个模型依次高 3.4、2.1、2.5 个百分点。与 Faster R-CNN、EfficientNet 和 RTDETR-L 相比,YOLO-RET 模型在 mAP@ 0.5 分别提高 1.1、3.4、11.1 个百分点,在 mAP@ 0.5:0.95 分别提高 3.1、4.2、24.5 个百分点,同时模型的参数数量和浮点运算量显著减少。综合试验结果表明,尽管改进模型检测速度有所降低,但依然可以满足实时检测需求。

3.4 不同模型实际检测结果

为清晰且直观地比较各模型对茶叶嫩芽检测对象的预测结果,基于验证集的检测结果如图 8 所示。从左上到右下依次为 YOLO-RET、YOLO v5n、YOLO v8n、和 YOLO v10n 模型对不同图像的检测结果。图中蓝色框标出的目标为正确识别对象,红色框表示该模型未检测出的对象,即漏检对象,红色与蓝色框重叠表示检测错误的情况。

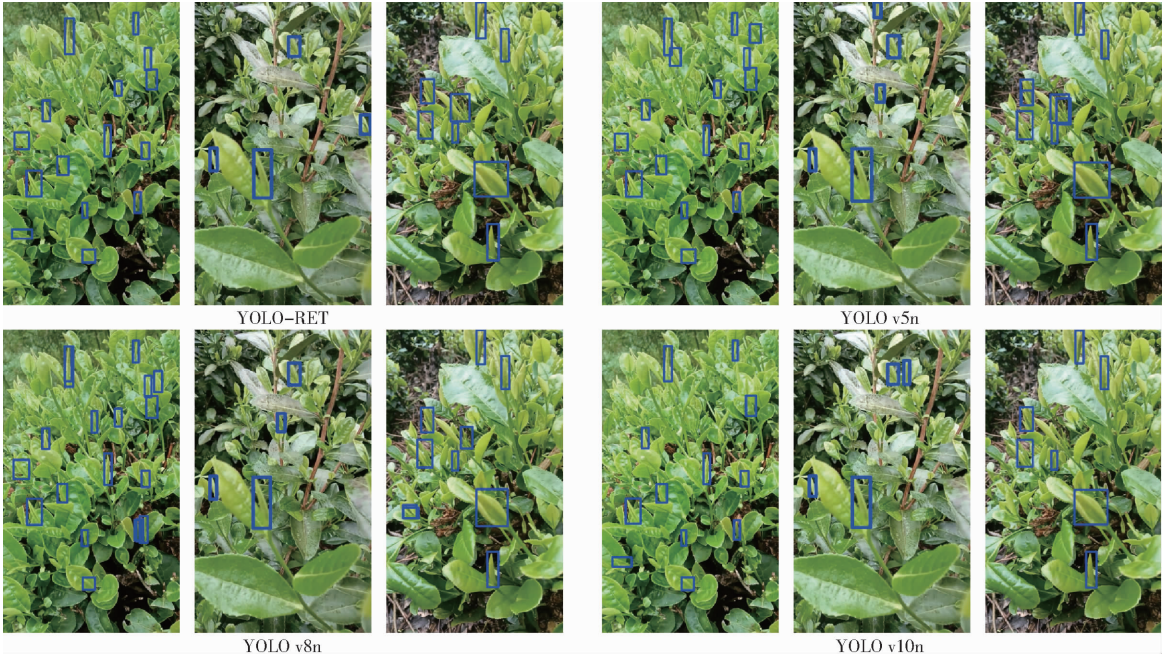


图 8 不同模型对茶叶嫩芽的实际检测结果

Fig. 8 Actual detection results of tea leaf buds by different models

由图 8 可知,在实际检测中存在漏检和误检的现象。相比之下,YOLO-RET 模型的检测效果最佳,漏检和误检最少,精确率最高,但仍有小部分漏检情况存在。与正确标签图像相比,YOLO v5n、YOLO v8s 和 YOLO v10n 模型均能够以较高精确率检测出大部分茶叶嫩芽,尽管也会有部分漏检和误检情况,但相对于本文 YOLO-RET 模型,检测精确率仍然较低,表明本文改进模型显著提高了模型检测性能。

为了深入分析改进模型对茶叶嫩芽检测的学习能力,选择梯度加权类激活映射 (Grad-CAM)^[35] 技术,对 YOLO v5n、YOLO v8n、YOLO v10n 和 YOLO-

RET 模型的可视化检测结果进行了输出,如图 9 所示。Grad-CAM 技术通过将目标概念的梯度流入模型的最后一个卷积层,生成定位映射并以权重形式重新展示出来。生成的热力图中,颜色表示权重,权重越小,相应位置区域颜色越蓝;权重越大,相应位置区域颜色越红,反映出该区域在茶叶嫩芽检测决策中的重要性。

由图 9 可知,YOLO v5n 和 YOLO v8n 模型视野范围较大,在识别小目标对象时,易受到周边信息影响,因此更易出现对小目标漏检的情况。YOLO v10n 模型则更倾向于关注茶树的叶子和部分与检测对象相似的茎,导致其在复杂背景下检测精确率

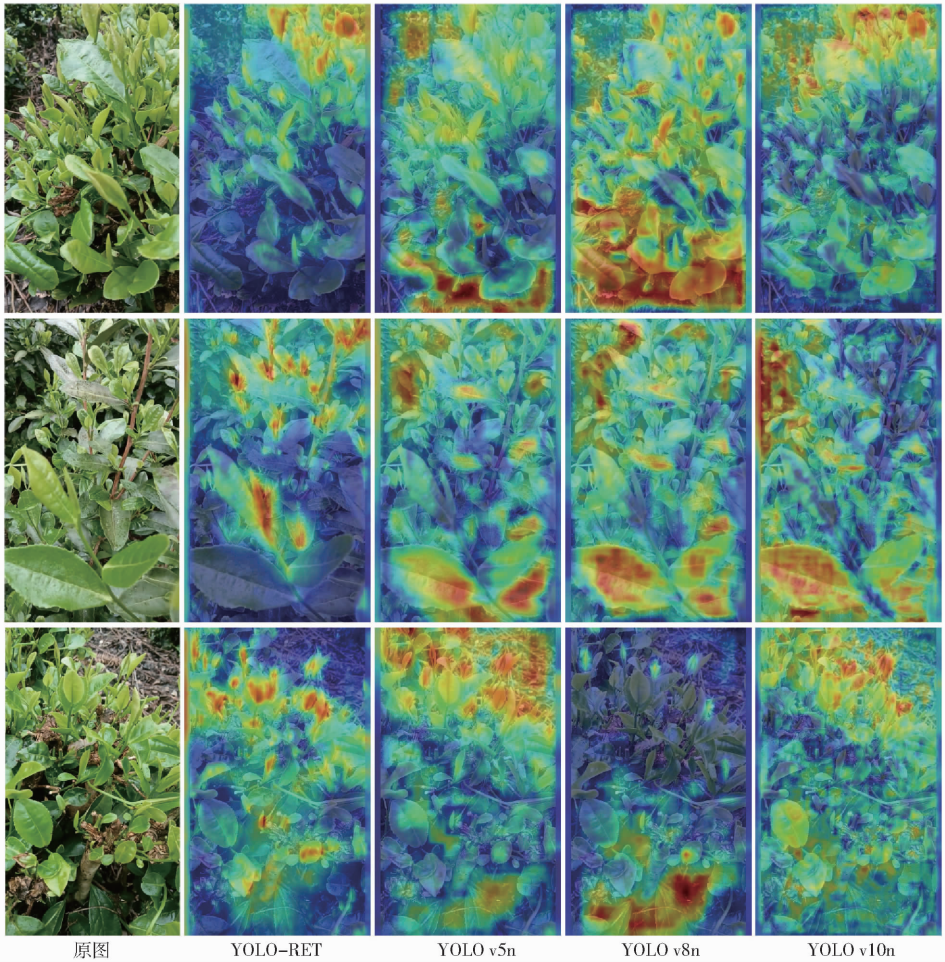


图 9 不同模型 Grad - CAM 输出结果

Fig. 9 Grad - CAM outputs of different models

下降。相比之下,本文 YOLO - RET 模型更专注于茶树的尖端嫩叶,受周围叶子和复杂背景图像的影响较小,同时也能够关注远景和近景的茶叶嫩芽,对远近大小不同的茶芽对象均保持较高的关注度。因此,热力图可视化验证了改进模型在茶叶嫩芽检测方面的潜在能力,证明其在复杂环境下具有更高的检测精确率和鲁棒性。

3.5 边缘设备部署

选择边缘设备开发平台为型号 ATK - DLRK3568 的瑞芯微 RK3568 芯片,配备了 4 颗主频为 2.0 GHz 的 64 位 ARM Cortex - A55 处理器,内置独立的神经网络处理单元(NPU),提供 1 TOPS 算力,配置 4 GB 的 LPDDR4 内存。

尽管 RK3568 开发板内置专用 NPU 加速模块,但直接部署基于 PyTorch 训练的 .pt 格式模型文件不仅无法充分利用这一硬件特性,还可能因性能不足而导致推理延迟。为解决这些问题,本文采用瑞芯微官方提供的 rknn - toolkit2 工具包,将 .pt 格式模型转换为开放标准的 ONNX 格式作为中间步骤,再通过优化和量化,最终导出 .rknn 格式模型。处

理后,模型能更好地适配 RK3568 的性能需求,确保高效运行。整个转换与量化流程如图 10 所示。

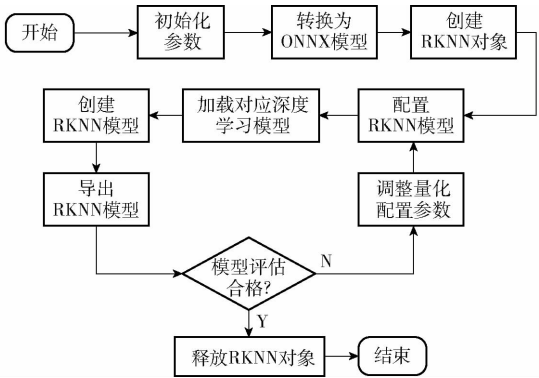


图 10 模型转换过程

Fig. 10 Process of model conversion

针对边缘设备算力约束,在量化模型与参数配置时需要考虑其性能限制。因此,本试验将批次大小设置为 1。通过对 16 位半精确率浮点数格式(FP16)和 8 位整数格式(INT8)的模型精度评估,探索部署后精度与推理速度的平衡方案。同时,为提升模型在终端设备上的推理效率,选择 C + + 编译环境进行部署。针对目标平台编写 C + + 代码,通

过交叉编译工具链生成适用于该硬件架构的执行文件,最终将编译好的程序部署到 RK3568 平台,可以在保持模型性能的前提下实现硬件兼容。

经优化压缩的模型关键参数指标如表 3 所示,FP16 量化操作导致模型体积增大,而 INT8 采用动态量化策略未发生显著变化。相比于 YOLO v8n 基准模型,本文模型在 INT8 格式下,模型内存占用量减少 20%,运行模型时设备的内存占用量降低 36.7%,但由于网络层数增加,导致推理检测速度下降 18.8 f/s,但仍能满足实时检测要求。YOLO - RET 模型经过量化转换导致 mAP@0.5 下降 8.5 个百分点,但相比于经过同样量化转换后 YOLO v8n 模型平均精确均值仍高 3.2 个百分点。使得检测速度和 mAP@0.5 能够满足在边缘设备上实时检测需求。

表 3 2 种模型部署后性能

指标	YOLO v8n		YOLO - RET	
	FP16	INT8	FP16	INT8
模型内存占用量/MiB	5.55	4.58	5.88	3.66
运行模型设备内存占用量/MiB	21.86	10.5	22.24	6.65
检测速度(f·s ⁻¹)	7.29	37.89	6.32	19.09
mAP@0.5/%	79.8	73.4	81.9	76.6

为验证转换后的 YOLO - RET 模型检测效果,本文在测试集中选取部分图像进行实际效果检测。如图 11 所示,根据检测目标的难易程度,在 INT8 格式下,图 11 中检测框上显示的置信度相比 FP16 格式最高降低近 20 个百分点。因此,在较高置信度阈值情况下,INT8 格式会出现明显的漏检情况。表明在选择量化数据格式时,需要根据边缘设备实际算力情况,权衡模型精确率与检测速度,以确保最佳的检测效果。

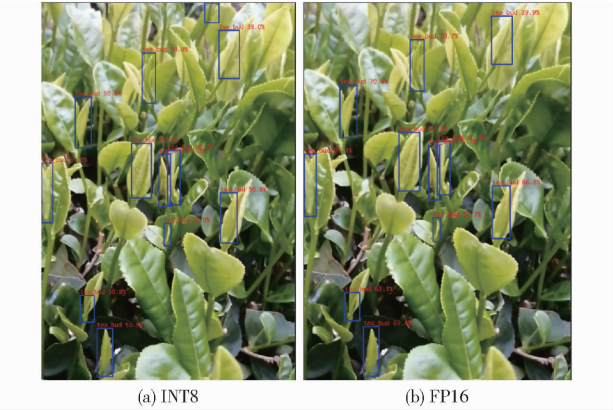


图 11 模型检测结果(边缘设备)

Fig. 11 Results of model inference(edge device)

4 结论

(1)提出了基于改进 YOLO v8n 的轻量化茶叶嫩芽检测模型 YOLO - RET,通过重新设计轻量化特征提取模块(RGCSPELAN 模块)和改进多尺度特征融合金字塔结构(EMBSFPN),在大幅降低模型参数的前提下显著提升了特征提取与融合能力。引入的 Focaler - IoU 损失函数有效解决样本分布不均衡问题,提高检测精度和模型鲁棒性。

(2)试验结果表明,与 YOLO v5n、YOLO v8n、YOLO v8s 和 YOLO v10n 相比,YOLO - RET 精确率分别提高 3.2、3.0、0.8、4.3 个百分点,mAP@0.5 分别提升 2.7、3.2、0.4、3.1 个百分点。模型参数量和浮点运算量较原 YOLO v8n 分别降低 43% 和 2.2 × 10⁹。Grad - CAM 热力图可视化分析证实,改进模型能更精准地聚焦于茶树尖端嫩芽区域,对复杂背景干扰具有更强的鲁棒性。

(3)在 ATK - DLRK3568 边缘计算平台上的成功部署进一步验证该模型实用价值,为资源受限环境下的实时目标检测提供了高效解决方案,为名优茶智能采摘系统提供了可靠的技术支持。

参 考 文 献

[1] AHAMED G J, LI X. Hormonal regulation of health-promoting compounds in tea (*Camellia sinensis* L.) [J]. Plant Physiology and Biochemistry: PPB,2022,185:390 - 400.

[2] 王明龙,徐焱,张智浩,等. 茶叶智能机械化采摘技术与装备研究进展[J]. 中国农机化学报,2024,45(9):305 - 310. WANG Minglong, XU Yao, ZHANG Zhihao, et al. Research progress of intelligent mechanized tea picking technology and equipment[J]. Journal of Chinese Agricultural Mechanization,2024,45(9): 305 - 310. (in Chinese)

[3] 郑航,傅童,薛向磊,等. 茶叶机械化采摘技术研究现状与展望[J]. 中国农机化学报,2023,44(9):28 - 35. ZHENG Hang, FU Tong, XUE Xianglei, et al. Research status and prospect of tea mechanized picking technology [J]. Journal of Chinese Agricultural Mechanization, 2023, 44(9): 28 - 35. (in Chinese)

[4] 高一聪,许晨,林琼,等. 茶叶生产装备自动化与智能化技术研究进展与展望[J]. 农业机械学报,2024,55(7):1 - 14. GAO Yicong, XU Chen, LIN Qiong, et al. Tea production equipment automation and intelligent technology research progress and prospects[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(7): 1 - 14. (in Chinese)

[5] 李杨,董春旺,陈建能,等. 茶叶智能采摘技术研究进展与展望[J]. 中国茶叶,2022,44(7):1 - 9. LI Yang, DONG Chunwang, CHEN Jianneng, et al. Research progress and prospects of intelligent tea picking technology[J]. China Tea, 2022, 44(7): 1 - 9. (in Chinese)

[6] WU Y X, CHEN J N, WU S K, et al. An improved YOLO v7 network using RGB - D multi-modal feature fusion for tea shoots detection[J]. Computers and Electronics in Agriculture,2024,216:108541.

- [7] 吕军, 方梦瑞, 姚青, 等. 基于区域亮度自适应校正的茶叶嫩芽检测模型[J]. 农业工程学报, 2021, 37(22): 278–285. LÜ Jun, FANG Mengrui, YAO Qing, et al. Detection model for tea buds based on region brightness adaptive correction[J]. Transactions of the CSAE, 2021, 37(22): 278–285. (in Chinese)
- [8] 张昆, 袁博涵, 崔静莹, 等. 基于计算机视觉的茶叶嫩芽识别方法研究进展[J]. 山东农业科学, 2024, 56(5): 163–170. ZHANG Kun, YUAN Bohan, CUI Jingying, et al. Research progress of tea bud recognition based on computer vision[J]. Shandong Agricultural Sciences, 2024, 56(5): 163–170. (in Chinese)
- [9] WANI J A, SHARMA S, MUZAMIL M, et al. Machine learning and deep learning based computational techniques in automatic agricultural diseases detection: methodologies, applications, and challenges [J]. Archives of Computational Methods in Engineering, 2022, 29: 641–677.
- [10] 朱红春, 李旭, 孟炆, 等. 基于 Faster R-CNN 网络的茶叶嫩芽检测[J]. 农业机械学报, 2022, 53(5): 217–224. ZHU Hongchun, LI Xu, MENG Yang, et al. Tea bud detection based on Faster R-CNN network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(5): 217–224. (in Chinese)
- [11] BAI Bingyi, WANG Junshu, LI Jianlong, et al. T-YOLO: a lightweight and efficient detection model for nutrient buds in complex tea-plantation environments[J]. Journal of the Science of Food and Agriculture, 2024, 104(10): 5698–5711.
- [12] 王梦妮, 顾寄南, 王化佳, 等. 基于改进 YOLO v5s 模型的茶叶嫩芽识别方法[J]. 农业工程学报, 2023, 39(12): 150–157. WANG Mengni, GU Jìnan, WANG Huajia, et al. Method for identifying tea buds based on improved YOLO v5s model[J]. Transactions of the CSAE, 2023, 39(12): 150–157. (in Chinese)
- [13] 黄家才, 唐安, 陈光明, 等. 基于 Compact-YOLO v4 的茶叶嫩芽移动端识别方法[J]. 农业机械学报, 2023, 54(3): 282–290. HUANG Jiakai, TANG An, CHEN Guangming, et al. Mobile recognition solution of tea buds based on Compact-YOLO v4 algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(3): 282–290. (in Chinese)
- [14] ZHANG M, XU S, SONG W, et al. Lightweight underwater object detection based on YOLO v4 and multi-scale attentional feature fusion[J]. Remote Sensing, 2021, 13(22): 4706.
- [15] LI F, JIANG Z, ZHOU S, et al. Spilled load detection based on lightweight YOLO v4 trained with easily accessible synthetic dataset[J]. Computers and Electrical Engineering, 2022, 100: 107944.
- [16] ZHANG Z, LU Y, ZHAO Y, et al. TS-YOLO: an all-day and lightweight tea canopy shoots detection model[J]. Agronomy 2023, 13: 1411.
- [17] 王文明, 肖宏儒, 陈巧敏, 等. 基于图像处理的茶叶智能识别与检测技术研究进展分析[J]. 中国农机化学报, 2020, 41(7): 178–184. WANG Wenming, XIAO Hongru, CHEN Qiaomin, et al. Research progress analysis of tea intelligent recognition and detection technology based on image processing[J]. Journal of Chinese Agricultural Mechanization, 2020, 41(7): 178–184. (in Chinese)
- [18] VARGHESE R, MENG S. YOLO v8: a novel object detection algorithm with enhanced performance and robustness[C]//2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). IEEE, 2024: 1–6.
- [19] GLENN J. YOLOv8 [EB/OL]. (2023) [2024-09-23]. Available online: <https://github.com/ultralytics/ultralytics>.
- [20] WANG C Y, BOCHKOVSKIY A, LIAO H. YOLO v7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[J]. arXiv Preprint arXiv:2207.02696, 2022.
- [21] HAN K, WANG Y, TIAN Q, et al. GhostNet: more features from cheap operations[J]. arXiv Preprint arXiv:1911.11907, 2019.
- [22] WANG C Y, YEH I H, LIAO H Y M. YOLO v9: learning what you want to learn using programmable gradient information[C]//European Conference on Computer Vision, 2024: 3–19.
- [23] LI Y, CHEN Y, WANG N, et al. Scale-aware trident networks for object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2019: 6053–6062.
- [24] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017: 936–944.
- [25] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 8759–8768.
- [26] WANG C C, HE W, NIE Y, et al. Gold-YOLO: efficient object detector via gather-and-distribute mechanism[J]. arXiv Preprint arXiv:2309.11331, 2023.
- [27] TAN M, PANG R, LE Q V. EfficientDet: scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020: 10777–10786.
- [28] XUE Y, JU Z, LI Y, et al. MAF-YOLO: multi-modal attention fusion based YOLO for pedestrian detection[J]. Infrared Physics & Technology, 2021, 118: 103906.
- [29] RAHMAN M M, MUNIR M, MARCULESCU R. EMCAD: efficient multi-scale convolutional attention decoding for medical image segmentation[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2024: 11533–11543.
- [30] CUI Y, JIA M, LIN T Y, et al. Class-balanced loss based on effective number of samples[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019: 9268–9277.
- [31] SHRIVASTAVA A, GUPTA A, GIRSHICK R. Training region-based object detectors with online hard example mining[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016: 761–769.
- [32] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318–332.
- [33] ZHANG H, ZHANG S. Focaler-IoU: more focused intersection over union loss[J]. arXiv Preprint arXiv:2401.10525, 2024.
- [34] MMDetection C. OpenMMLab detection toolbox and benchmark[EB/OL]. (2018) [2024-09-23]. <https://github.com/open-mmlab/mmdetection>.
- [35] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[C]//2017 IEEE International Conference on Computer Vision (ICCV). IEEE, 2017: 618–626.