DOI:10. 13880/j. cnki. 65-1174/n. 2026. 21. 018 文章编号:1007-7383(2026)05-0529-10

基于 YOLOv8-MRF 的黄瓜生育期检测方法

陈言行[1],何青海[2*],郭洪恩[2],王琦[2],李天华[3],马本学[1]
(1 石河子大学机械电气工程学院,新疆 石河子 832003;2 山东省农业机械科学研究院养分资源高效利用全国重点实验室,山东 济南 250100;3 山东农业大学机械与电子工程学院,山东 泰安 271018)

摘要:针对温室环境下黄瓜生育期检测精度低、边缘设备难以兼顾检测精度与实时性等问题,本文提出一种基于 YOLOv8-MRF 的黄瓜生育期检测方法。首先,在骨干网络中引入了混合局部通道注意力机制(mixed local channel attention,MLCA),以增强模型的特征信息提取能力;其次,采用感受野坐标注意力卷积(Receptive field coordinates attention convolution,RFCAConv)融合多尺度上下文信息,以增强模型多尺度信息感知能力;随后,更换损失函数为 Focal-EIoU,增强困难样本的权重贡献,以提升模型收敛速度与检测准确度;最后,采用模型剪枝与模型量化技术降低模型计算量,并将 YOLOv8-MRF 模型部署至 RK3588 开发板。试验结果表明:YOLOv8-MRF 模型的准确率、召回率、平均精度均值与 F1 分数分别为 87.0%、83.8%、88.7%、85.3%;经 LAMP 剪枝与量化,模型参数量降至 1.12 M,权重仅 2.6 MB,减小了边缘设备的运算压力,模型部署至 RK3588 边缘设备后,平均精度均值达到 88.4%,平均帧率为 35 f/s,相较于 YOLOv8n 模型,其差值为 2.5%和 20 f/s,可实现黄瓜生育期的实时、精准检测,为设施农业黄瓜生育期模型的边缘部署与工程化应用提供了技术支撑。

关键词:生育期识别;边缘设备;LAMP 剪枝;模型量化
中图分类号:TP391. 4 **文献标志码**:A

Mobile device-based cucumber growth stage detection method based on YOLOv8-MRF

CHEN Yanxing[1], HE Qinghai[2*], GUO Hongen[2], WANG Qi[2], LI Tianhua[3], MA Benxue[1]
(1 College of Mechanical and Electrical Engineering, Shihezi University, Shihezi, Xinjiang 832003, China; 2 Shandong Academy of Agricultural Machinery Sciences, State Key Laboratory of Nutrient Use and Management, Jinan, Shandong 250100, China; 3 College of Mechanical and Electronic Engineering, Shandong Agricultural University, Taian, Shandong 271018, China)

Abstract: To address the issues of low detection accuracy for cucumber growth stages in greenhouse environments and the difficulty for edge devices to balance detection accuracy and real-time performance, this study proposes a cucumber growth stage detection method based on YOLOv8-MRF. Firstly, a mixed local channel attention mechanism (MLCA) is integrated into the backbone network to enhance the model's feature information extraction capability; secondly, receptive field coordinates attention convolution (RFCAConv) is used to fuse multi-scale contextual information, thereby enhancing the model's multi-scale information perception ability. Subsequently, the loss function is replaced with Focal-EIoU to enhance the weight contribution of hard samples, thereby improving the model's convergence speed and detection accuracy. Finally, model pruning and model quantization techniques are adopted to reduce the model's computational load, and the YOLOv8-MRF model is deployed onto an RK3588 development board. The experimental results show that the accuracy, recall rate, mean average precision and F1 score of the YOLOv8-MRF model are 87.0%, 83.8%, 88.7% and 85.3%, respectively. After LAMP pruning and quantization, the model's parameter count is reduced to 1.12 M with a weights of only 2.6 MB, reducing the computational pressure on edge devices. After the model is deployed onto the RK3588 edge device, the mean average precision reaches 88.4%, and the average frame rate is 35 f/s. Compared with the YOLOv8n model, the differences are 2.5% and 20 f/s, respectively. It enables real-time and accurate detection of cucumber growth stages, providing technical support for the edge deployment and engineering application of cucumber growth stage models in facility agriculture.

收稿日期:2025-12-08
基金项目:山东省重大科技创新工程项目(2022CXGC020708)
作者简介:陈言行(2000—),男,硕士研究生,专业方向为智能技术及应用。
*通信作者:何青海(1986—),男,高级工程师,从事设施装备与智慧节水方向的研究,e-mail:heqinghai1986@ 126. com。

Key words: growth stage recognition; edge devices; LAMP pruning; model quantization

黄瓜是我国重要瓜类蔬菜之一[1],据联合国粮食及农业组织统计,2023 年我国黄瓜种植面积达 $1.309\,4\times10^6$ hm^2,总产量超过 $8.021\,76\times10^7$ t,种植规模与产量均位居世界首位。果实生育期识别对于黄瓜水肥管理、环境调控及采收作业[2-3]具有重要指导意义。目前,黄瓜生育期识别依赖人工观测,存在效率低、主观性强的问题,难以满足现代农业实时、精准检测的需求。因此,研究黄瓜生育期识别方法,对于推动黄瓜现代化生产管理具有重要意义。

传统的果实生育期识别通常是获取果实颜色、纹理、光谱特征信息,并结合阈值分割或机器学习技术实现目标检测,这在背景简单、光照稳定的场景中具有一定有效性,但在自然环境下的泛化能力受限。Tan 等[4]提出一种结合颜色特征和方向梯度直方图的蓝莓果实生育期识别方法,通过提取方向梯度直方图与颜色特征信息,使用支持向量机与 K 最近邻分类算法分别检测蓝莓果实区域与果实生长期,该方法对于遮挡果实识别效果较差且耗时较长。另外,夏雪等[5]通过自适应 G-B 色差图与迭代阈值分割提取兴趣区,采用支持向量机算法实现幼果检测,模型仅针对非重叠或轻微遮挡的幼果期苹果。

近年来,以 YOLO 系列为代表的目标检测算法因其较高的检测精度与实时性,已被广泛应用于作物果实生育期识别领域。杨林等[6]基于 YOLOv5 算法,替换上采样模块并添加小目标检测层,实现了幼菇期与成熟期香菇的识别,识别准确率分别为 94.0%和 91.3%,平均帧率为 45.25 f/s,该模型未考虑生产场景光照变化影响。黄家才等[7]提出了一种轻量级 Mobile-YOLOv5s 草莓检测算法,通过替换 MobileNetV3 骨干网络、引入 Alpha-IoU 损失函数、采用 K-Means++重聚类 anchor 及增加检测头,使草莓检测精度提升至 99.5%,检测帧率提升至 44 f/s,模型大小缩减 41.5%,该方法在数据集规模与多场景泛化方面存在不足。Bai 等[8]提出 U-Net + YOLOv3 模型,采用 U-Net 模型对黄瓜图像进行分割,再利用 YOLOv3 模型对分割后的图像进行识别,黄瓜检测精度达 97.73%,但模型处理速度较慢,难以满足实时性要求。

针对上述问题,本文面向温室黄瓜生长场景,以果实生育期为研究对象,基于 YOLOv8n 模型结合混合局部通道注意力机制、感受野坐标注意力卷积、Focal-EIoU 损失函数,提出 YOLOv8-MRF 模型,采用模型剪枝与模型量化技术降低模型运算量,确保模型部署至 RK3588 开发板后实现黄瓜生育期的实时、精准检测。

1 材料与方法

1.1 数据采集

本文以广泛种植的长青 818 黄瓜品种为研究对象,数据采集于山东省济南市东方商人生态农场,采集设备为 Redmi K30 设备,其分辨率为 4 032×4 032 px,采集时间为 2024 年 5 月 15 日至 2024 年 7 月 10 日。

考虑到光照条件对数据质量的影响,每日采集时间定为上午 9:00—12:00 与下午 14:00—17:00。在不同光照条件下采用不同角度、距离采集数据,共采集 4 000 张图像,均保存为 JPG 格式,经人工筛选后保存 3 800 张原始数据。依据黄瓜的实际生长阶段,将果实分为子房发育期、生育期、膨大期和成熟期[9-10],各类别的图像数量分别为 896、1 026、1 004、874 张。黄瓜生育期样本如图 1 所示。



a—子房发育期 b—生长期 c—膨大期 d—成熟期

图 1 黄瓜生育期样本

Figure 1 Sample of cucumber growth period

1.2 数据集构建

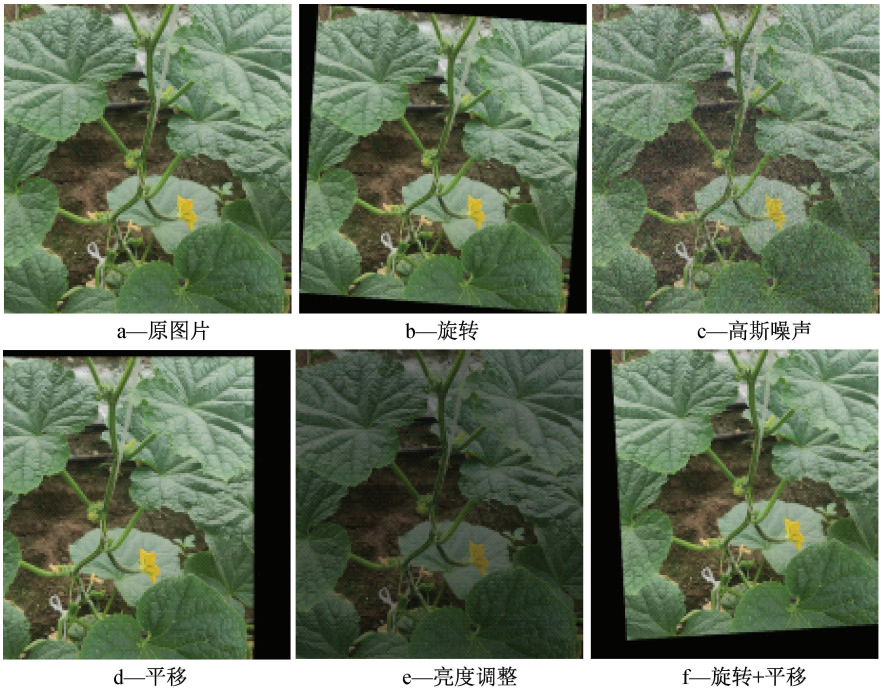
在温室环境下,光照变化、枝叶遮挡及果实重叠等因素会影响模型特征提取。为增强数据多样性,随机采用亮度调整、添加高斯噪声、旋转、平移等 1 种或几种数据增强方法,将数据集扩充至 9 500 张。干扰条件下黄瓜图像如图 2 所示,黄瓜数据增强示例如图 3 所示。



a—背景干扰　b—枝叶遮挡　c—果实重叠　d—逆光

图 2　干扰条件下黄瓜图像示例

Figure 2　cucumber images under interference conditions



a—原图片　b—旋转　c—高斯噪声

d—平移　e—亮度调整　f—旋转+平移

图 3　黄瓜数据增强示例

Figure 3　Cucumber data augmentation example

使用 Labelimg 软件标注果实生长期与位置,标注框尽可能贴近黄瓜轮廓,对于遮挡严重的黄瓜不标注,而小部分遮挡的黄瓜标注其未遮挡部分,标签保存为 txt 格式,然后按照 7:2:1将数据集随机划分为训练集、测试集、验证集。

1.3　试验平台与评价指标

1.3.1　试验平台和参数配置

模型训练硬件环境如下:Intel Core i5-13400F CPU 处理器,显卡型号为 NVIDIA GeForce RTX 3060,显存为 12 GB。

模型训练软件环境如下:Windows10 操作系统,深度学习框架为 Pytorch1.13.0,编程语言为 Python3.9.21,轮次为 150,批次为 8,输入图片为 640×640。

采用瑞芯微 RK3588 开发板为移动端硬件平台,该开发板采用 Debian11 操作系统,运行内存 16 G,存储容量 128 G,运算能力为 6TOPS。

1.3.2　评价指标

为衡量模型效果,本文参考文献[7]中的目标检测评价方法,采用准确率 P、召回率 R、平均精度均值 mAP、F1 分数衡量模型检测效果。其中,P 用于衡量模型检测结果中正确识别样本所占的比例,R 用于衡量实际目标被模型正确检测出的比例,F1 分数综合反映准确率与召回率之间的平衡关系,mAP 作为综合评价指标,用于衡量模型在不同类别的检测性能。

2　YOLOv8-MRF 模型

针对温室环境下黄瓜生育期检测精度低、边缘设备难以兼顾检测精度与实时性等问题,本文提出

一种 YOLOv8-MRF 黄瓜生育期识别模型,其网络结构如图 4 所示。首先,在骨干网络中添加 MLCA 注意力机制,以提升模型对局部空间细节与全局通道信息的提取能力。其次,将骨干网络的常规卷积替换为 RFCAConv,以增强模型对不同尺寸目标的感知能力,实现多尺度信息融合。随后,采用 Focal-EIoU 损失函数,增加困难样本的损失权重,以加速模型训练收敛。最后,采用模型剪枝与模型量化技术降低模型参数量与运算量,以降低边缘设备的运算压力。

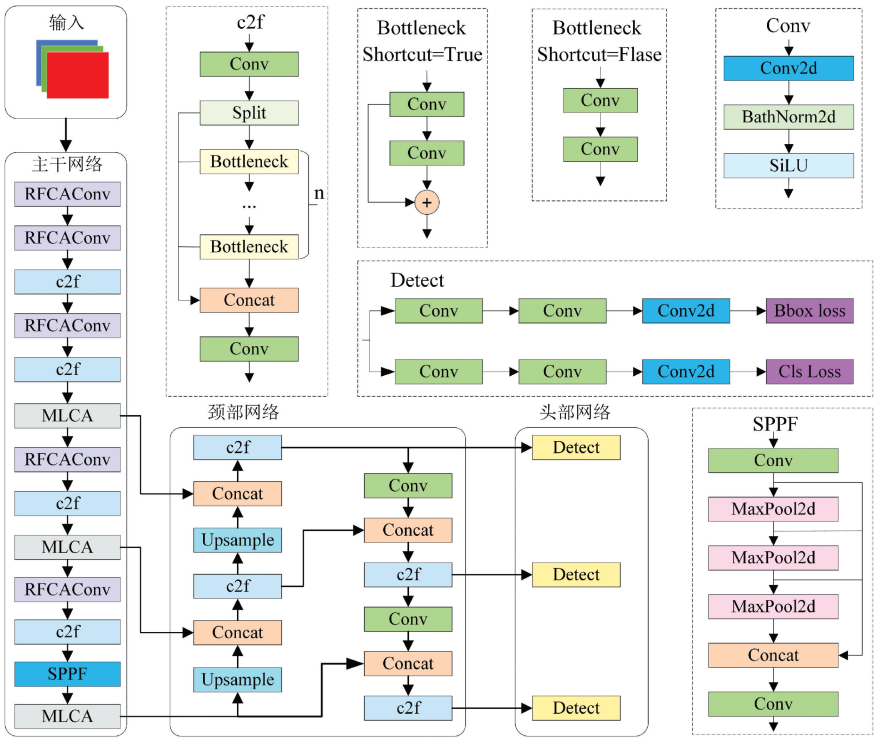


图 4 YOLOv8-MRF 模型结构

Figure 4 YOLOv8-MRF model structure diagram

2.1 MLCA 注意力机制

SE、ECA 等信道注意力机制关注信道特征信息,忽略了空间特征信息[11-12],混合局部通道注意力机制 MLCA 融合了通道信息、空间信息、局部通道信息与全局通道信息,在效率、性能和复杂性之间实现了更好的平衡[13],MLCA 网络结构如图 5 所示。

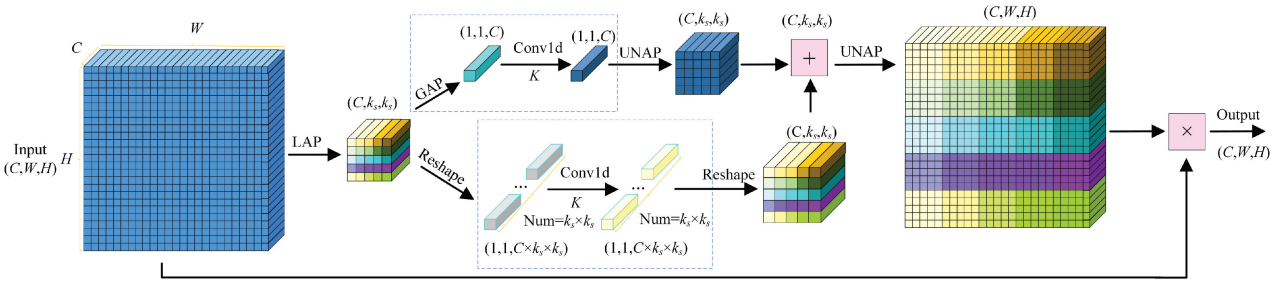


图 5 MLCA 模块结构

Figure 5 MLCA module structure diagram

首先使用 MLCA 模块对输入特征图进行局部平均池化(Local average Pooling,LAP),将特征图划分为若干个 $ks \times ks$ 的非重叠区域,以捕获各局部区域的空间信息,得到维度为 $C \times ks \times ks$ 的空间特征表达。然后采用双分支结构将输入特征转换为一维向量,获取全局通道信息与局部通道信息,经一维卷积后采用反池化或 Reshape 操作恢复输入特征分辨率,两分支特征经逐元素相加实现全局、局部通道信息融合。最后通过反池化操作恢复为 $C \times W \times H$ 原始空间维度,并与原始输入特征图逐元素相乘实现通道维度上的自适应加权,实现了局部与全局通道信息的混合建模。

Conv1d 一维卷积操作中,卷积核大小 k 依据通道数 C 设置,在关注通道维度上相邻特征之间的相

互作用时，仅对每个通道与其邻近 k 个通道的关系建模，实现跨通道局部信息的有效捕获。

为了增强对果实形态、纹理和颜色等特征的表达能力，以及在复杂温室环境下有效提升模型检测精度，本文研究引入 Wan 等[13]建立的 MLCA 模块，通过融合局部与全局通道信息，对通道特征进行自适应加权。

2.2　RFCAConv 模块

感受野注意力卷积（Receptive Field Attention Convolution，RFAConv）实现了关键区域的自适应聚焦，但通道特征的建模能力有限[14]。RFCAConv 在 RFAConv 的基础上融合 CA 注意力机制[15]，增强了模型通道维度的特征提取能力，RFAConv 网络结构如图 6 所示。

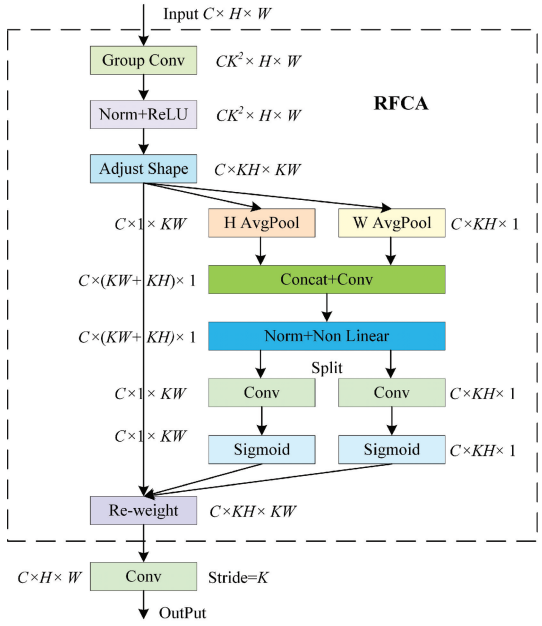


图 6　RFCAConv 结构

Figure 6　RFCAConv structure diagram

RFCAConv 采用分组卷积将输入通道划分为多感受野子空间，生成感受野空间特征，通过维度重组将其转换为 $C\times KH\times KW$ 结构，分离不同感受野区域，避免特征混淆。随后，采用双路平均池化提取垂直、水平方向特征，经拼接、卷积、归一化与非线性激活后，生成方向感知的注意力向量。通过参数解耦策略将注意力向量拆分为垂直、水平分支，采用卷积操作生成垂直、水平注意力图，并使用 Sigmoid 函数约束权重范围。最终将垂直、水平方向注意力与感受野空间特征进行自适应融合，通过卷积完成空间维度压缩与多尺度特征融合。

通过 RFCAConv 融合多尺度感受野与方向感知注意力，可强化对不同生育期黄瓜果实尺度差异及空间特征的表达能力，并改善模型在复杂背景下对黄瓜果实不同生育期的判别能力。

2.3　Focal-EIoU 损失函数

原 YOLOv8 模型采用 CIoU[16]作为边界框回归损失函数，CIoU 损失函数仅考虑重叠面积、中心点距离和纵横比 3 个因素，忽略了绝对尺寸差异，不利于网络训练。本研究采用 Focal-EIoU[17]作为边界框损失函数，引入宽度、高度的绝对值惩罚项与焦点调节机制，缓解了宽高比尺度不敏感与样本不平衡问题，Focal-EIoU 损失函数计算如图 7 所示。

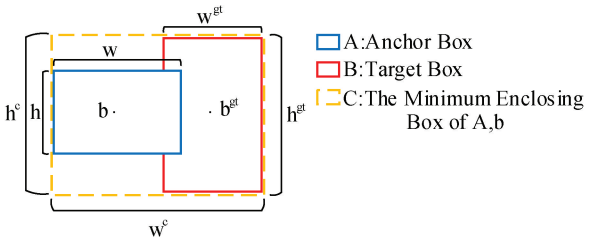


图 7　损失函数计算示意图

Figure 7　Schematic of the loss function calculation

Focal-EIoU 损失函数将损失函数分为三部分，分别为交并比损失 L_{IOU}、距离损失 L_{dis}、方位损失 L_{asp}。Focal-EIoU 损失函数采用宽度、高度的绝对值惩罚项替换纵横比，约束了预测框与真实框的绝对尺寸差异[18]，可解决 CIoU 在目标具有相同宽高比和尺寸差异较大场景下的优化失效问题。

由于目标在图像中通常较为稀疏，模型训练过程中存在样本不平衡问题，导致容易样本主导梯度优化过程，削弱了困难样本对模型参数更新的贡献，影响边界框回归的精度与稳定性。而 Focal-EIoU 损失函数引入基于交并比的焦点调节机制，通过指数函数抑制高交占比样本的梯度贡献和强化低交占比样本的贡献权重，可以有效平衡不同质量样本在训练过程中的贡献度差异，从而提升模型性能。Focal-EIoU 损失函数如式（1）、式（2）所示：

$$L_{\mathrm{EIoU}}=L_{\mathrm{IOU}}+L_{\mathrm{dis}}+L_{\mathrm{asp}},=1-\mathrm{IOU}(A,B)+\frac{\rho^2(b,b^{\mathrm{gt}})}{(h^{\mathrm{c}})^2+(w^{\mathrm{c}})^2}+\frac{\rho^2(h,h^{\mathrm{gt}})}{h^{\mathrm{c}}}+\frac{\rho^2(w,w^{\mathrm{gt}})}{w^{\mathrm{c}}},\tag{1}$$

$$L_{\mathrm{Focal-EIoU}}=\mathrm{IOU}(A,B)^{\gamma}L_{\mathrm{EIoU}},\tag{2}$$

式（1）、（2）中：$\mathrm{IOU}(A,B)$ 为 $|A\cap B|/|A\cup B|$；A、B 为预测框和真实框；$\rho(\cdot)$ 为中心点的欧氏距离；b 与 b^{gt} 为预测框与真实框的中心点坐标；W^{c} 与 h^{c} 为包围预测框与真实框的最小外接矩形的宽度和高度；w、h 为预测框的宽度、长度；w^{gt}、h^{gt} 为真实框的宽度、高度；γ 为抑制异常值程度的超参数。

Focal-EIoU 损失函数通过引入宽高绝对差异项,增强对果实尺度与形态变化的约束能力;通过基于交并比的焦点调节机制,提升低交占比样本的训练贡献,改善低交占比样本的回归效果,提高模型的检测精度。

2.4 LAMP 剪枝

YOLOv8-MRF 模型能提升检测精度,但增加了模型计算量与边缘设备的推理负担,所以为降低模型计算量,本文采用基于层自适应幅度剪枝(Layer-adaptive sparsity for the Magnitude-based Pruning, LAMP)[19]对 YOLOv8-MRF 模型进行模型轻量化处理。基于幅值的剪枝方法主要是根据幅值大小进行剪枝,由于不同层之间权重分布差异较大,容易导致不同层之间剪枝不均衡。而 LAMP 剪枝方法通过层幅值归一化机制评估每层权重的重要性,并考虑了每层权重的分布差异,从而可有效避免因层间权值分布差异导致的剪枝偏差,更准确地反映其在网络表达能力中的实际贡献,实现对权重重要性的自适应衡量,并有效降低模型冗余,同时最大程度保持网络的表达能力与检测性能。

LAMP 剪枝流程分为以下 4 个步骤:

(1)将网络中的每个权重张量展开为一维向量,并计算其对应的 LAMP 分数,以衡量权重的重要性。LAMP 分数综合考虑了权重在层内分布的相对幅值,为各层提供统一且可比较的剪枝依据。

(2)将同层权重按照 LAMP 分数进行升序排序。

(3)根据预设的剪枝率与 LAMP 分数,迭代删减 LAMP 分数最低的权重项,直至达到预定的剪枝率。

(4)对剪枝后的模型进行微调训练,以恢复部分因权重稀疏化而导致的性能退化,在保持模型精度的基础上,降低参数冗余与计算开销。

2.5 模型量化

为降低模型计算量、加速模型推理运算,本文采用模型量化技术对网络进行优化。模型量化是将网络的 32 位浮点型参数进行离散化处理,将其映射到整型参数的范围中,实现模型压缩与推理效率提升,其数学表达如式(3)所示:

$$X_{\text{int}} = \text{clamp}\left(\left\lfloor \frac{x}{s} \right\rceil + z, -2^{b-1}, 2^{b-1}-1\right), \quad (3)$$

式(3)中:X_{int} 为量化后的定点数值;x 为模型浮点参数数值;$\lfloor \cdot \rceil$ 四舍五入运算;s 为量化比例因子;z 为量化零点;b 为量化位宽,即定点数的二进制位数;clamp 为截断运算。

为保留模型特征参数的相对分布,量化比例因子 s 与线性非对称量化零点 z 根据浮点型参数量化范围与定点数数值范围 $[-2^{b-1}, 2^{b-1}-1]$ 计算,如式(4)、式(5)所示:

$$s = \frac{q_{\max} - q_{\min}}{(2^{b-1}-1) - (-2^{b-1})} = \frac{q_{\max} - q_{\min}}{2^{b-1}}, \quad (4)$$

$$z = (2^{b-1}-1) - \left\lfloor \frac{q_{\max}}{s} \right\rceil, \quad (5)$$

式(4)、(5)中:$q_{\max}$ 为浮点型参数最大值;$q_{\min}$ 为浮点型参数最小值;b 为量化位宽。

3 结果与分析

3.1 损失函数试验

为验证 Focal-EIoU 损失函数效果,本研究以 YOLOv8n 为基础模型,在相同试验条件下对 CIoU、SIoU、EIoU、Focal-EIoU 损失函数进行对比试验,试验结果见表 1、图 8。

表 1 损失函数性能对比

Table 1 Comparison of loss function performance

损失函数	准确率	召回率	平均精度均值	F1 分数
CIoU	83.7	81.9	86.1	82.8
SIoU	83.6	81.7	86.0	82.6
EIoU	84.0	82.1	86.4	83.0
Focal-EIoU	84.4	82.4	87.0	83.4

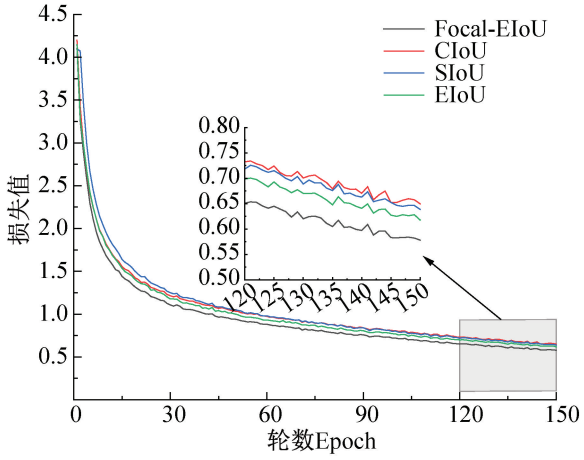


图 8 不同损失函数损失曲线

Figure 8 Loss curves for different loss functions

由表 1、图 8 可以看出:采用 SIoU 损失函数时,其损失曲线与 CIoU 函数近乎重合,模型平均精度均值仅 86.0%,相较于 CIoU 函数,其差值为-0.1%,SIoU 函数未能提升模型性能。EIoU 损失函数在回归阶段的收敛速度略快于 CIoU,整体损失值有所下降,但平均精度均值 mAP 差值仅为 0.3%,性能提升较为有限。Focal-EIoU 损失函数收敛趋势整体优于 CIoU、SIoU 和 EIoU 函数,与 CIoU 函数相比,其准确

率、召回率、mAP、F1 分数的差值分别为 0.7%、0.5%、0.9%、0.6%。这是因为 Focal-EIoU 函数通过焦点调节机制,自适应地调整不同样本的损失权重,强化困难样本的贡献权重,降低易分类样本的损失权重,提升模型在复杂目标样本上的鲁棒性与泛化能力,所以采用 Focal-EIOU 损失函数进行后续的改进试验。

3.2　不同剪枝方法结果对比

为比较不同剪枝策略在相同剪枝率下的性能差异,本研究将剪枝率设置为 2.0,对比 LAMP、Group Hessian、Group Taylor、L1 及 Random 等方法的剪枝效果。在统一剪枝率下,评估各方法在剪枝后参数量、浮点运算量、权重及 mAP 等指标上的表现,以确定最优的剪枝方案。各剪枝方法的剪枝效果见表 2。

表 2　剪枝策略性能对比

Table 2　Performance comparison of pruning strategies

剪枝方法	参数量	浮点运算量	权重	平均精度均值
LAMP	1.12	4.2	2.6	88.7
Group Hessian	1.47	4.2	3.2	87.8
Group Taylor	1.54	4.2	3.0	87.0
L1	1.82	4.2	4.0	87.0
Random	2.07	4.2	4.5	81.7

由表 2 可以看出:LAMP 方法通过归一化权重重要性实现自适应阈值分配,有效保留了关键特征通道,参数量仅为 1.12 M、权重 2.6 MB,保持了最高的检测精度,模型平均精度均值 mAP 为 88.7%。Group Hessian 通过二阶梯度信息评估网络权重重要性,虽能平衡剪枝与性能,但仍存在轻微精度损失,mAP 为 87.8%,相较于 LAMP 方法,其 mAP 差值为 -0.9%。Group Taylor 剪枝方法基于泰勒展开的一阶近似原理,通过评估权重剪枝后对网络损失函数的影响来衡量其重要性,但未考虑高阶信息变化,在非线性模型中精度损失较大,模型平均精度均值 mAP 仅为 87.0%。L1 剪枝通过基于权重幅值的简单筛选实现压缩,计算复杂度较低,精度下降较明显,模型平均精度均值 mAP 仅为 87.0%。Random 剪枝方法未考虑权重重要性,随机移除权重,导致特征表达受损最为严重,模型平均精度均值 mAP 降至 81.7%。

由以上结果可知:LAMP 方法在保持原有精度的同时,降低了模型计算量与参数量。因此,本研究采用 LAMP 剪枝方法进行模型轻量化处理。

为评估不同剪枝率对模型性能与模型压缩效果的影响,将剪枝率设置为 1.25~2.5,步长为 0.25,进行对比试验[20]。由表 3 可知:剪枝率为 2.0 时,模型参数量和运算量显著降低,模型平均精度均值维持原精度,但是剪枝率提升至 2.25、2.50 时,模型精度呈下降趋势。综合考虑模型轻量化效果与检测性能,采用剪枝率为 2.00 的 LAMP 剪枝方法,降低模型计算量。

表 3　不同剪枝率性能对比

Table 3　Performance comparison of different pruning rates

剪枝率	参数量	浮点运算量	权重	平均精度均值
1.25	1.95	6.7	4.2	88.9
1.50	1.47	5.6	3.2	88.8
1.75	1.25	4.7	2.8	88.8
2.00	1.12	4.2	2.6	88.7
2.25	1.02	3.7	2.4	88.3
2.50	0.95	3.3	2.2	88.0

3.3　消融试验

为验证各改进策略对原始模型的改进效果,分别将 Focal-EIoU 损失函数、RFCAConv、MLCA 注意力机制融合到 YOLOv8n 模型中,并在相同环境下进行 9 组消融试验,结果见表 4。

表 4　消融试验结果

Table 4　Results of ablation experiments

编号	模块		Focal-EIoU	剪枝	参数量	浮点运算量	平均精度均值	F1 分数
	MLCA	RFCAConv						
1	×	×	×	×	3.01	8.2	86.1	82.8
2	×	×	√	×	3.01	8.2	87.0	83.4
3	×	√	×	×	3.03	8.4	87.2	83.7
4	√	×	×	×	3.00	8.1	87.4	84.0
5	×	√	√	×	3.03	8.4	87.9	84.5
6	√	×	√	×	3.00	8.1	88.1	84.8
7	√	√	×	×	3.04	8.4	88.3	85.0
8	√	√	√	×	3.04	8.4	88.7	85.3
9	√	√	√	√	1.12	4.2	88.7	85.4

由表 4 可以看出:与原模型相比,模型 2 的 mAP、F1 分数差值分别为 0.9%、0.6%,其参数量与浮点运算量维持不变。

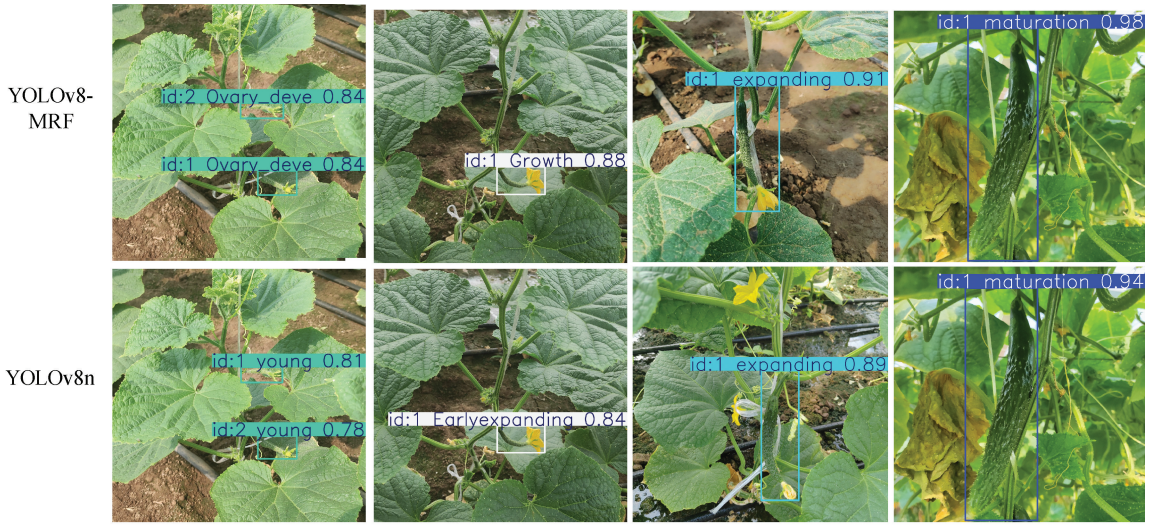
模型 3 采用 RFCAConv 替换骨干网络的常规卷积,引入空间感受野扩展机制与通道信息机制,实现空间、通道联合增强,强化了多尺度特征的建模能力,提升了对黄瓜果实不规则物体的检测效果。与原模型相比,模型 3 的 mAP、F1 分数差值分别为 1.1%、0.9%,模型参数量与浮点运算量仅增加 0.02 M 和 0.2 G。

模型 4 在骨干网络融合 MLCA 注意力,结合了通道、空间、局部、全局信息,提升了模型的特征提取能力,与原模型相比,其 mAP、F1 分数差值分别为 1.3%、1.2%,模型参数量与浮点运算量分别降低了 0.01 M 与 0.1 G。

模型 8 融合 MLCA 注意力、RFCAConv、Focal-EIoU 损失函数后,参数量与浮点运算量仅增加 0.03 M、0.2 G,mAP 与 F1 分数差值分别为 2.6%和 2.5%。

为了降低边缘设备的运算压力,模型 9 融合 MLCA 注意力机制、RFCAConv、Focal-EIoU 损失函数、LAMP 剪枝方法,并将剪枝率设置为 2,相较于基础模型 YOLOv8n,其模型参数量、浮点运算量降低了 1.89 M、4.0 G,其模型 mAP、F1 分数差值分别为 2.6%、2.6%。

模型检测效果如图 9 所示。



a—子房发育期 b—生长期 c—膨大期 d—成熟期

图 9 模型检测效果对比

Figure 9 Comparison of model checking effects

3.4 不同模型性能对比

为评估剪枝后 YOLOv8-MRF 模型的检测性能,本研究在相同试验条件下,基于测试集对 Faster R-CNN、SSD、YOLOv3-tiny、YOLOv5s、YOLOv7-tiny、YOLOv8n、YOLOv8-MRF 及剪枝后 YOLOv8-MRF 等模型进行对比分析,各模型性能的检测结果(表 5)显示:

表 5 不同模型在测试集上的检测性能

Table 5 Detection performance of different models on the test set

模型	准确率	召回率	平均精度均值	F1 得分	参数量	浮点运算量	权重
Faster R-CNN	65.1	62.3	61.5	63.7	41.2	208	160
SSD	80.1	57.2	67.1	66.7	61.0	23.9	91.6
YOLOv3-tiny	79.2	74.5	76.7	76.8	12.13	18.9	24.4
YOLOv5s	81	79.5	82.7	80.2	9.11	23.8	18.5
YOLOv7-tiny	83.1	80.7	85.3	81.8	7.9	13.2	12.5
YOLOv8n	83.7	81.9	86.1	82.8	3.01	8.2	6.3
YOLOv8-MRF	87.0	83.8	88.7	85.3	3.04	8.4	6.4
YOLOv8-MRF(剪枝)	87.1	83.8	88.7	85.4	1.12	4.2	2.6

Faster R-CNN 模型平均精度均值 mAP 仅为 61.5%,F1 得分为 63.7%且参数量和浮点运算量高达 41.2 M 和 208 G,该模型检测效果最差且计算开销较大,不利于轻量化部署。

SSD 模型其 mAP 和 F1 值分别为 67.1%和 66.7%,较 Faster R-CNN 略有提升,但参数量较高。

YOLOv3-tiny 模型对黄瓜果实生育期数据集检测性能较差,平均精度均值仅为 76.7%,低于其他 YOLO 系列模型,YOLOv3-tiny 模型参数量、浮点运算量分别为 12.13 M、18.9 G,边缘计算设备算力难以满足其需求。

YOLOv5s、YOLOv7-tiny 模型的 mAP 分别为 82.7%、85.3%,而基础模型 YOLOv8n 的 mAP 达到了 86.1%,与 YOLOv5s、YOLOv7-tiny 相比,其 mAP 差值分别为 3.4%、0.8%,得益于 YOLOv8 的轻量化网络结构设计,其模型参数量、浮点运算量仅 3.01 M、8.2 G,均远低于上述 YOLO 系列模型。

YOLOv8-MRF 模型的平均精度均值 mAP 为 88.7%,相较于 Faster R-CNN、SSD、YOLOv3-tiny、YOLOv5s、YOLOv7-tiny、YOLOv8n,其 mAP 差值分别为 27.2%、21.6%、12.0%、6.0%、3.4%、2.3%,检测性能显著增强。YOLOv8-MRF 模型参数量与浮点运算量仅为 3.04 M、8.4 G,相较于 YOLOv8n 模型略增 0.03 M、0.2 G。

采用 LAMP 剪枝后的 YOLOv8-MRF(剪枝)模型参数量、浮点运算量与模型权重分别为 1.12 M、4.2 G、2.6 M,与原 YOLOv8-MRF 模型相比,分别降低了 63%、50%、59%,而模型平均精度均值 mAP 仍

维持在 88.7%,F1 值为 85.4%,证明了该方法在模型压缩与精度保持方面的有效性与优越性。

3.5　边缘设备部署

将剪枝、量化后的 YOLOv8-MRF 模型部署至 RK3588 板卡,验证其检测效果,检测运行效果(图 10)显示:采用 16 位整型参数 RK3588 开发板可降低模型计算量,提升模型推理效率,但导致目标检测精度轻微下降。



图 10　模型运行效果

Figure 10　Model operation effect

由模型在各部署环境的检测效果(表 6)可以看出:YOLOv8-MRF 模型经 LAMP 剪枝与模型量化后,部署至 RK3588 板卡,其平均精度均值为 88.4%,平均帧率为 35 f/s。表明该模型具有较高精度与检测速度。

表 6　模型在各部署环境的检测效果

Table 6　Detection performance of the models across different deployment environments

模型	部署环境	平均精度均值	帧速率
YOLOv8n	RTX 3060 GPU	86.1	119
	RK3588	85.9	15
YOLOv8-MRF(剪枝)	RTX 3060 GPU	88.7	103
	RK3588	88.4	35

4　结论

(1)本文提出的 YOLOv8-MRF 模型依次融合 MLCA 注意力、RFCAConv 与 Focal-EIoU 损失函数,可以增强模型特征提取能力、多尺度信息感知能力,并平衡不同交占比样本的训练贡献,平均精度均值由 86.1%分别提升至 87.4%、88.3%、88.7%。

(2)YOLOv8-MRF 模型在测试集上的准确率、召回率、mAP 和 F1 分数分别达到 87.0%、83.8%、88.7%和 85.3%,与 YOLOv8n 模型相比,其差值分别为 3.3%、1.9%、2.6%和 2.5%,而模型参数量、计算量和权重规模仅增加 0.03 M、0.2 G 和 0.1 MB。

(3)通过模型剪枝与量化技术对 YOLOv8-MRF 模型进行压缩优化。剪枝后的 YOLOv8-MRF(剪枝)模型平均精度均值保持在 88.7%,其参数量、浮点运算量和权重规模分别降至 1.12 M、4.2 G 和 2.6 MB,较 YOLOv8n 模型分别降低了 62.8%、48.8%和 58.7%;部署至边缘设备后,量化模型的平均精度均值为 88.4%、平均帧率为 35 f/s,与 YOLOv8n 模型相比,其差值分别为 2.5%和 20 f/s,验证了该模型在精度与实时性方面的综合优势。

(4)本文提出的 YOLOv8-MRF 模型在结合模型剪枝与量化技术后,部署至边缘计算设备可实现较高的检测精度与实时性。该研究为黄瓜生育期识别模型在边缘计算设备上实现精准、实时检测提供了技术支撑,可促进农业智能检测技术在实际生产中的应用。

参考文献(References)

[1] 周海霞,杨凡,张果,等. 河南省设施黄瓜产业现状、存在问题与高质量发展对策[J]. 中国瓜菜,2025,38(4):198-205.
Zhou Haixia, Yang Fan, Zhang Guo, et al. The current situation, existing problems, and countermeasures for high quality development of facility cucumber industry in Henan province[J]. China Cucurbits and Vegetables, 2025, 38(4): 198-205.

[2] Hossain S A A M, Wang L X, Chen T T, et al. Leaf area index assessment for tomato and cucumber growing period under different water treatments[J]. Plant, Soil and Environment, 2017, 63: 461-467.

[3] 董适,赵国瑞,苟豪,等. 基于改进 RT-Detr 的黄瓜果实选择性采摘识别方法[J]. 农业工程学报,2025,41(1):212-220.
Dong Shi, Zhao Guorui, Gou Hao, et al. Identifying cucu-mber fruits during selective picking using improved RTDetr[J]. Transactions of the Chinese Society of Agricultural Engineering, 2025,41(1):212-220.

[4] Tan K Z, Lee W S, Gan H, et al. Recognising blueberry fruit of different maturity using histogram oriented gradients and colour features in outdoor scenes[J]. Biosystems Engineering, 2018, 176: 59-72.

[5] 夏雪,周国民,丘耘,等. 自然环境下果实作业机器人幼果期苹果侦测方法[J]. 中国农业科技导报,2018,20(5):64-74.
Xia Xue, Zhou Guomin, Qiu Yun, et al. Detection of young green apples for fruit robot in natural scene[J].

Journal of Agricultural Science and Technology,2018,20(5):64-74.

[6] 杨林,曾大鑫,边银丙,等.改进YOLOv5的香菇子实体生育期识别方法[J].农业工程学报,2024,40(9):182-189.
Yang Lin, Zeng Daxin, Bian Yinbing, et al. Recognizing fruiting body growth period of Lentinus edodes using improved YOLOv5[J]. Transactions of the Chinese Society of Agricultural Engineering, 2024, 40(9): 182-189.

[7] 黄家才,赵雪迪,高芳征,等.基于改进YOLOv5s的草莓多阶段识别检测轻量化算法[J].农业工程学报,2023,39(21):181-187.
Huang Jiacai, Zhao Xuedi, Gao Fangzheng, et al. Recognizing and detecting the strawberry at multi-stages using improved lightweight YOLOv5s[J]. Transactions of the Chinese Society of Agricultural Engineering, 2023, 39(21):181-187.

[8] Bai Y H, Guo Y X, Zhang Q, et al. Multi-network fusion algorithm with transfer learning for green cucumber segmentation and recognition under complex natural environment[J]. Computers and Electronics in Agriculture, 2022, 194: 106789.

[9] 高文瑞,徐刚,孙艳军,等.设施黄瓜果实发育过程中品质及形态变化规律的研究[J].金陵科技学院学报,2011,27(1):44-47.
Gao Wenrui, Xu Gang, Sun Yanjun, et al. The Fruit quality and morphological changes during the development process of cucumber fruits[J]. Journal of Jinling Institute of Technology, 2011,27(1):44-47.

[10] 杨悦.黄瓜果实发育过程纵横径变化规律及其器官形状相关性研究[D].杨凌:西北农林科技大学,2018.

[11] Wang Q L, Wu B G, Zhu P F, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[J]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2020: 11531-11539.

[12] Hu J, Shen L, Sun G, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.

[13] Wan D H, Lu R S, Shen S Y, et al. Mixed local channel attention for object detection[J]. Engineering Applications of Artificial Intelligence, 2023, 123(PC): 106442.

[14] Zhang X, Liu C, Yang D G, et al. RFAConv: innovating spatial attention and standard convolutional operation[EB/OL]. 2023-04-06[2025-06-25]. https://arxiv.org/abs/2304.03198.

[15] Hou Q B, Zhou D Q, Feng J S, et al. Coordinate attention for efficient mobile network design[EB/OL]. 2021-03-04[2025-06-25]. https://arxiv.org/abs/2103.02907.

[16] Zheng Z H, Wang P, Liu W, et al. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression[J]. Proceedings of the Aaai Conference on Artificial Intelligence, 2020, 34(7): 2993-13000.

[17] Zhang Y F, Ren W Q, Zhang Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506: 146-157.

[18] 刘诗怡,胡滨,赵春.基于改进YOLOv7的黄瓜叶片病虫害检测与识别[J].农业工程学报,2023,39(15):163-171.
Liu Shiyi, Hu Bin, Zhao Chun, et al. Detection and identi-fication of cucumber leaf diseases based improved YOLOv7[J]. Transactions of the Chinese Society of Agricultural Engineering, 2023,39(15):163-171.

[19] Lee J, Park S, Mo S, et al. Layer adaptive sparsity for the magnitude based Pruning[EB/OL]. arXiv, 2020[2025-12-08]. https://arxiv.org/abs/2010.07611.

[20] 施国英,纪嘉鹏,李天华,等.基于改进YOLOv8s的杭白菊检测与花期分类[J].农业工程学报,2025,41(7):192-199.
Shi Guoying, Ji Jiapeng, Li Tianhua, et al. Detecting chrysanthemum to classify flowering stages using improved YOLOv8s[J]. Transactions of the Chinese Society of Agricultural Engineering, 2025, 41(7): 192-199.

(责任编辑:张忠)