

杨国亮,盛杨杨,洪鑫芳,等. 基于改进 YOLO v5s 的果园番茄采摘检测方法[J]. 江苏农业科学,2024,52(24):187-195.

doi:10.15889/j.issn.1002-1302.2024.24.027

基于改进 YOLO v5s 的果园番茄采摘检测方法

杨国亮, 盛杨杨, 洪鑫芳, 余帅英, 张佳琦

(江西理工大学电气工程与自动化学院,江西赣州 341000)

摘要:针对复杂果园环境下,不同生长周期的番茄容易受叶片、藤蔓的遮挡,以及多果实之间相互遮挡进而导致的误检、漏检、难检问题,提出了一种基于改进 YOLO v5s 的番茄生长周期采摘实时检测方法。首先,设计多尺度反向自适应注意力模块 RAAM(Reversed Adaptive Attention Module),用于解决原模型连续下采样丢失上下文信息的问题,并可以提升特征图分辨率用于高层语义信息的加权融合。其次,在 Neck 中设计 C3DS-ST(C3 DySnakeConv-Swin Transformer)模块替换原有 C3,增大局部区域坐标特征感知能力,适应整体番茄结构形状特征。最后,设计新的边框损失函数 SWD_Loss,结合 BiFormer 注意力机制,使网络更加关注敏感区域,提高模型对遮挡部分的漏检率和多果实遮挡的检测精度。试验结果表明,改进后的模型在测试集上对番茄检测平均精度达到 93.7%,相比于原模型提升 2.1%,对图片中遮挡部分效果明显改善。同时,检测速度达到了 50 帧/s,具有非常良好的应用前景。

关键词:番茄检测;自适应 RAAM 模块;C3DS-ST;SWD_Loss;BiFormer

中图分类号:TP391.41 **文献标志码:**A **文章编号:**1002-1302(2024)24-0187-09

番茄是我国重要的蔬菜作物,其产量及需求量都相当巨大^[1]。番茄采摘目前仍以人工为主,工作强度大且成本高。果蔬采摘机器人也是近年来人工智能领域的热门话题,我国也相继开发出草莓、苹果等农业采摘机器人^[2-3]。精准的识别和定位是顺利采摘的前提,提高识别精度可有效减少未达到采摘标准的次数,提升采摘机器人的工作效率。在实际采摘条件下,番茄生长条件复杂,存在着枝叶遮挡、大小果重叠等多种干扰因素。然而,对于未成熟番茄误采摘会影响采摘效率,对成熟番茄不及时采摘会影响到果实品质 and 经济效益。所以,提高番茄的识别精度和速度,对农园采摘和果蔬实时运输都有十分重要的意义。

近年来随着计算机视觉技术的发展,深度学习在农业发展方面也随之进步,并展示出巨大的优越性^[4-5]。传统的目标检测算法可以分为 2 种:一种是 one-stage 目标检测法,其检测过程不产生候选框,直接输出分类结果和定位信息,如 YOLO、SSD

等^[6-7]。此类模型大多参数量少,识别速度较快,可以便捷地部署到机器人摄像头、智能手机等移动终端。王勇等针对不同成熟度苹果改进 YOLO v5,其检测精度达到 93.6%^[8];Fan 等针对夜间草莓误摘、识别精度低,提出了一种结合暗通道增强改进 YOLO v5 算法^[9]。另一种是 two-stage 目标检测法。该方法先对图像计算出候选区域,然后再对候选框中的目标进行分类,如 R-CNN、Faster R-CNN 等^[10]。此类算法模型参数量大,识别速度较慢,无法满足实际生产中实时性检测的需求,但是识别准确率较高。张文静等提出改进 Faster R-CNN 检测算法,单样本图像处理时间为 245 ms^[11]。岳有军等提出改进的 Cascade R-CNN 网络,用于增强对成熟绿果番茄与其他不同成熟阶段番茄的区分,识别精度达到 97.75%^[12]。

考虑到番茄生长环境复杂,需要对采摘新鲜度、品质实时监测等因素,所以在番茄目标检测时要重视其精度和速度。而一步走算法能满足实时性要求。本研究将基于 YOLO v5s 算法进行改进,在骨干网络中融合多尺度反向自适应注意力模块 RAAM,用于提升深层网络中特征图像分辨率使其更好地传递给后续网络,可有效提高网络特征融合泛化能力;将颈部网络原有结构 C3 替换为 C3DS-ST 模块,完成对局部番茄遮挡特征的充分提取,通过坐标映射学习方法让局部区域适应整体番茄结

收稿日期:2023-12-07

基金项目:江西省教育厅科技计划项目(编号:GJJ210861、GJJ200879)。

作者简介:杨国亮(1973—),男,江西宜春人,博士,教授,主要从事人工智能和模式识别研究。E-mail:ygliang30@126.com。

通信作者:盛杨杨,硕士研究生,主要从事人工智能和模式识别研究。E-mail:syy505093@163.com。

构特征,可有效解决自然果园中由于多种因素遮挡造成的像素少、轮廓纹理不充分等问题;并引入 BiFormer 注意力机制,改进边框损失函数 SWD 加强对预测框的约束能力。研究结果表明改进后的模型不仅提升了检测精度,还能有效解决生产实际中由于番茄果实扎堆密集、大小果不一、枝叶藤蔓遮挡产生的漏检、难检等问题。

1 材料与方法

1.1 数据集

本试验使用的番茄影像数据集主要来源于网

络公开数据集和大棚实地拍摄。为了符合番茄生长的真实环境和采摘遇到的实际问题,采集数据集时不仅关注了不同生长时期的各种图像,例如成熟、未成熟、过熟、腐烂等多个状态,而且还对密集、遮挡等情况进行收集(图 1)。使用 Labellmg 工具对图像标注时,尽可能将背景信息减少到最小,保证数据的准确性,还原实际生活中果园采摘的真实情况。经过处理后,得到图像 3 251 张,将数据集按照 8 : 2 比例划分,其中 2 610 张作为训练集,641 张作为测试集。



图1 部分样本数据

1.2 YOLO v5s 网络模型

YOLO v5s 算法是由 Ultralytics 公司于 2020 年 6 月公开发布,主要包括输入端、骨干网络(Backbone)、颈部网络(Neck)以及检测层(Head)4 个部分。输入端通常由 Mosaic 数据增强、自适应锚框计算、自适应图片缩放 3 部分组成。骨干网络主要由卷积(Conv)、瓶颈(BottleneckCSP)、C3 和空间金字塔(SPPF)构成,负责对图像从多方面提取特征。颈部网络使用特征金字塔网络(FPN)和金字塔注意力网络(PANet),用于增强图像特征的融合。检测层包括 GIoU_Loss 损失函数和非极大值抑制(NMS)来进行分类预测。

1.3 YOLO v5s 改进模型

针对 YOLO v5s 存在小目标、遮挡对象的漏检误检问题,本研究对其模型进行了结构改造,重新设计了 RAAM、C3DS-ST 等模块,并将其用于果园番茄检测,改进后模型如图 2 所示。

1.3.1 多尺度反向自适应注意力模块(RAAM)

目前,多尺度特征融合已经广泛应用于目标检测任务,用来提高多尺度目标的检测性能和识别精度。其中,特征金字塔(feature pyramid network, FPN)是最常见的用于改善多层信息的特征融合网络。但其最明显的缺点就是过于关注底层信息提取和优化,经过多层卷积后随着通道数减少会丢失大量值得关注的信息。为了缓解这一问题,Wang 等提出了自适应注意力模块(adaptive attention module, AAM)^[13]。原文中当输入图像经过多层卷积后生成了特征图 C5,它将作为输入传递给 AAM。首先,通过自适应池化层(下采样因子为 α_i)会获得 3 个不同尺度的特征图,然后将他们缩放到同一大小($C \times W \times H$)进行拼接得到空间权值图。其中, C 、 H 、 W 分别表示特征图通道数、高度以及宽度。最后将得到的空间权值图与之前通道融合后的输出特征图进行矩阵乘法,将他们各自分离融合到 C5 特征

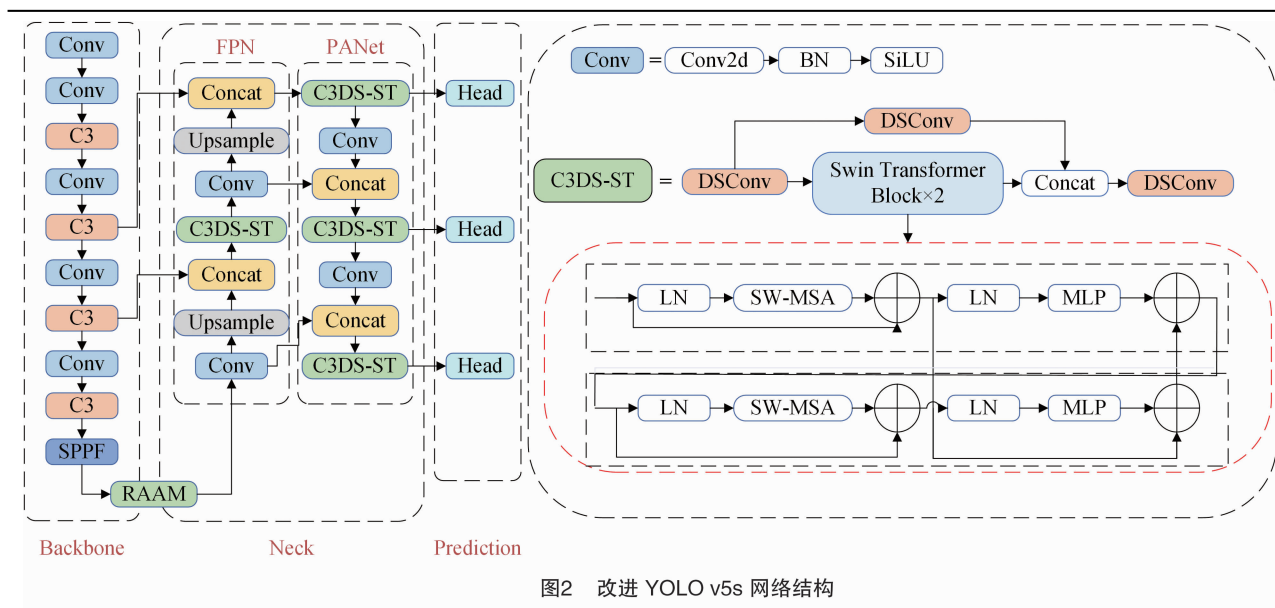


图2 改进 YOLO v5s 网络结构

图,聚合特征得到最后输出 P5。虽然原作者提出的 AAM 模块很大程度上缓解了由于特征通道数减少所造成的信息丢失问题,并通过自适应池化层丰富了多尺度信息维度,但是由于是低维度提取,得到的 3 种特征图包含了大量浅层信息,这会对后续网络提取的信息造成特征重叠和干扰。因此,这个方案对常规对象检测有所提升,但是对小目标、轮廓不清晰的遮挡部分并未产生良好的效果。因为这些目标本身在图像中就已经存在像素少、纹理缺失、模糊等问题,提取特征图 C1 到 C5 过程中,网络经过连续下采样提取到 C5 时输出尺度已经很小,继续下采样获得的特征图分辨率低、信息丢失严重。

本研究参考 Caballero 等提出的方法对 AAM 模块重新设计。Caballero 等提出一种新的高效亚像素卷积(sub-pixel conv)网络(图3)^[14],可以有效地将低分辨率图像恢复到高分辨率,缓解低通滤波和下采样造成的高频信息丢失;并且加入双层路由注意力机制(BiFormer Attention)对输出特征图进行加权融合,可以更好地提升网络对敏感区域的关注,抑制上下层无效信息的干扰。因为亚像素卷积网络是将不同尺度特征图进行上采样操作,刚好与 AAM 中下采样池化相反,因此改进后的模块称为反向自适应注意力模块 RAAM (Reversed Adaptive Attention Module),其模型结构如图 4 所示。

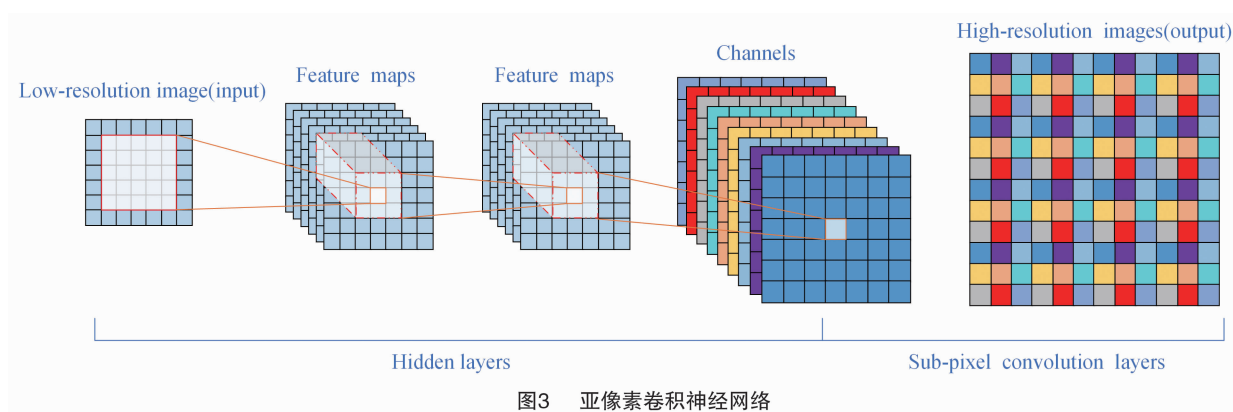


图3 亚像素卷积神经网络

首先由输入 C5 通过亚像素卷积层上采样(上采样因子为 β_i)得到 3 组不同分辨率的特征图 (Feature map - i),特征图尺寸(C_i, H_i, W_i)计算公式如下所示。经过对比试验当上采样因子 β_i 取(0.5, 1.5, 2)时效果最佳。

$$C_i = \frac{C}{\beta_i^2};$$

$$H_i = \beta_i \times H;$$

$$W_i = \beta_i \times W;$$

(1)

3 种不同尺度的特征图再经过 1×1 卷积将尺

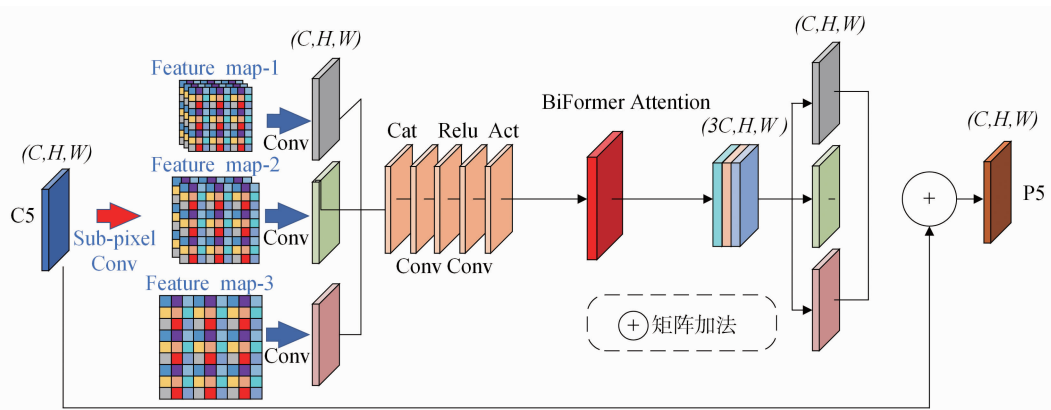


图4 反向自适应注意力模块(RAAM)结构

寸还原到 $(C \times H \times W)$,随后沿通道维度进行拼接。然后引入 BiFormer Attention 模块对有效信息进行增强,过滤冗余信息,最后分解成 3 个尺度都为 $(C \times H \times W)$ 的特征图与初始输入 C5 进行矩阵相加得到最后的特征映射 P5。新的 RAAM 模块通过在网络深层使用亚像素卷积网络进行上采样,可以有效解决因持续下采样通道数减少,图像中小目标以及局部遮挡区域像素点少产生的信息丢失、像素模糊等问题,生成高分辨率特征图传递给后续网络。

1.3.2 C3DS-ST 模块 Qi 等提出了一种动态蛇形卷积(dynamic snake convolution, DSConv),通过自适应聚焦细长和弯曲的局部结构来准确捕捉毛细

血管等管状结构的分割特征^[15]。在实际生长环境中番茄由于果实、枝叶遮挡从而导致漏检,是因为这些目标局部结构脆弱和全局形态缺失难以被模型所关注,其共同特征就是在图像中所占的比例很小,像素组成有限且难以捕捉。因此,迫切需要增强对局部影像结构的感知。基于此,本研究设计 C3DS-ST 模块用于解决这一问题,其结构如图 5 所示。其中,DSConv 是一种可变形卷积核,能够根据输入的大量番茄特征映射学习变形,动态地感知番茄的几何特征,更好地捕捉局部番茄的关键坐标特征,扩大自身感受野,以适应实际环境中番茄形态多变的结构(图 6)。

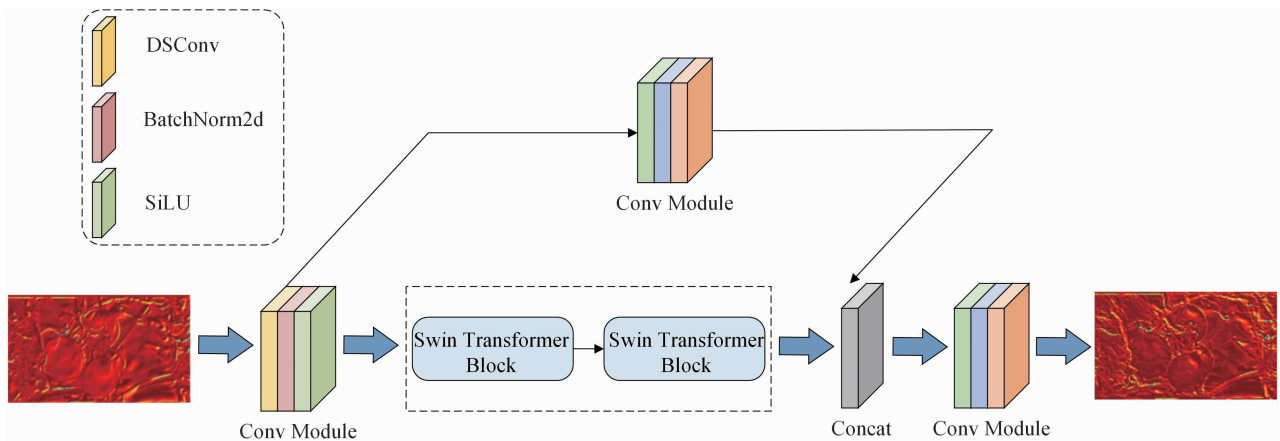


图5 C3DS-ST 模块结构

其中对于标准二维卷积坐标为 K ,中心坐标为 $K_i = (x_i, y_i)$,膨胀系数为 1 的 3×3 的卷积核可以表示为:

$$K = [(x-1, y-1), (x-1, y), \dots, (x+1, y+1)]. \quad (2)$$

采用图 6-a 中的迭代策略,对每个待处理的目标依次观察其位置。以坐标轴方向为例, K 中每个网格具体坐标表示为: $K_{i \pm c} = (x_{i \pm c}, y_{i \pm c})$,其中 $c =$

$(1, 2, 3, 4)$ 表示与中心网格坐标的水平距离,那么每个网格坐标 $K_{i \pm c}$ 的选择可以看作是对中心坐标 K_i 的一个累积过程。从中心位置 K_i 开始,下一步的位置取决于前一个网格的坐标,例如 K_{i+1} 与 K_i 相比,实际只增加了一个偏移量 $\Delta = \{\delta \mid \delta \in [-1, 1]\}$,从而确保卷积符合线性结构特征。图 6-b 在 x 轴、 y 轴方向表示为:

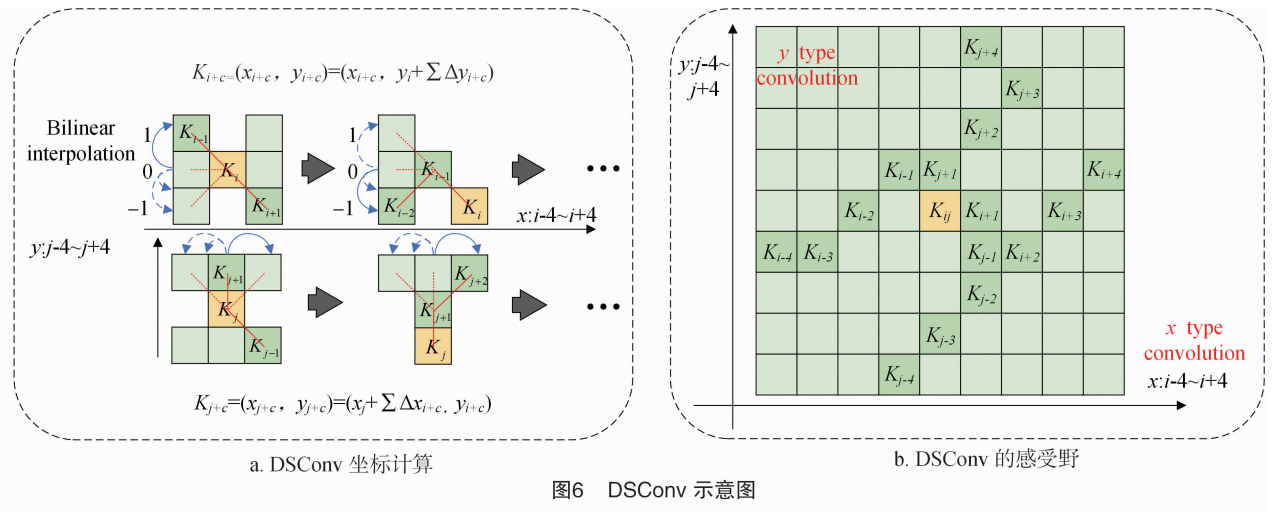


图6 DSConv 示意图

$$K_{i \pm c} = \begin{cases} (x_{i+c}, y_{i+c}) = (x_{i+c}, y_i + \sum_{i=c}^{i+c} \Delta y) \\ (x_{i-c}, y_{i-c}) = (x_{i-c}, y_i + \sum_{i=c}^i \Delta y) \end{cases}; \quad (3)$$

$$K_{j \pm c} = \begin{cases} (x_{j+c}, y_{j+c}) = (x_j + \sum_{j=c}^{j+c} \Delta x, y_{j+c}) \\ (x_{j-c}, y_{j-c}) = (x_j + \sum_{j=c}^j \Delta x, y_{j-c}) \end{cases}. \quad (4)$$

因为偏移量 Δ 通常是小数,所以双线性插值可以表示为:

$$K = K_K \cdot B(K', K) \cdot K'; \quad (5)$$

$$B(K, K') = b(K_x, K'_x) \cdot b(K_y, K'_y). \quad (6)$$

式中: K 表示分数阶位置, K' 枚举了所有积分空间位置, B 表示双线性插值。从图 6-b 中可以直观地看到经过 C3DS-ST 模块, 卷积可以覆盖 9×9 的感受野范围, 更好地感知番茄遮挡区域的局部关键信息, 以适应整体番茄的形状结构特征。同时为增强不同成熟度番茄的纹理特征差异, 提升模型泛化能力, 引入 Swin Transformer 模型中的 Swin Transformer Block 模块, 该模块在保留原 C3 融合残差特征基础上, 可以进一步对输入特征图进行独立窗口划分, 在各自窗口内独立计算以节省运算量, 之后采用滑动划分方式实现跨窗口之间的特征信息交换, 可以

增强不同成熟度番茄邻域间像素的纹理特征交互能力。

1.3.3 BiFormer 注意力机制 本研究在主干网络末端引入双层路由注意力机制 (Bi-level routing attention transformer, BiFormer), 其结构如图 7 所示。BiFormer 是一种新的动态稀疏注意力, 通过双层路由可以灵活地为键值对 (Key-value pairs, K-V) 提供算力分配和内容感知, 动态查询最相关 K-V 的一部分, 过滤大量不相关的 K-V 从而提升计算效率。BiFormer 注意力机制由 4 阶段金字塔构成, 依次重叠使用连续变换特征, 特征图尺寸依次变为之前的 $\frac{1}{2^i} H \times \frac{1}{2^i} W \times 2^i C (i=2, 3, 4, 5; j=0, 1, 2, 3)$ 。在每个 BiFormer Block 模块中, BRA 是核心构建块, 其结构如图 8 所示。

工作流程主要分 3 步实现, 第一步将输入特征图构建区域级有向图, 经过变形操作 (reshape) 将二维特征图 ($X \in R^{H \times W \times C}$) 完成线性映射 Q, K, V :

$$Q = X' W^q; \quad (7)$$

$$K = X' W^k; \quad (8)$$

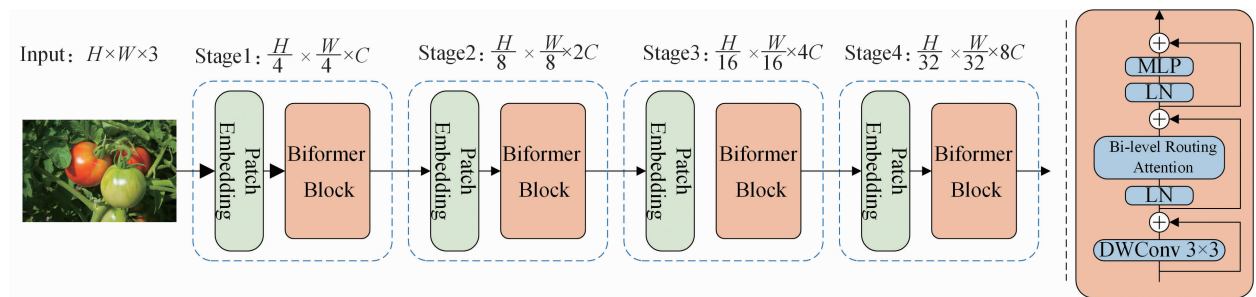


图7 BiFormer 注意力机制结构

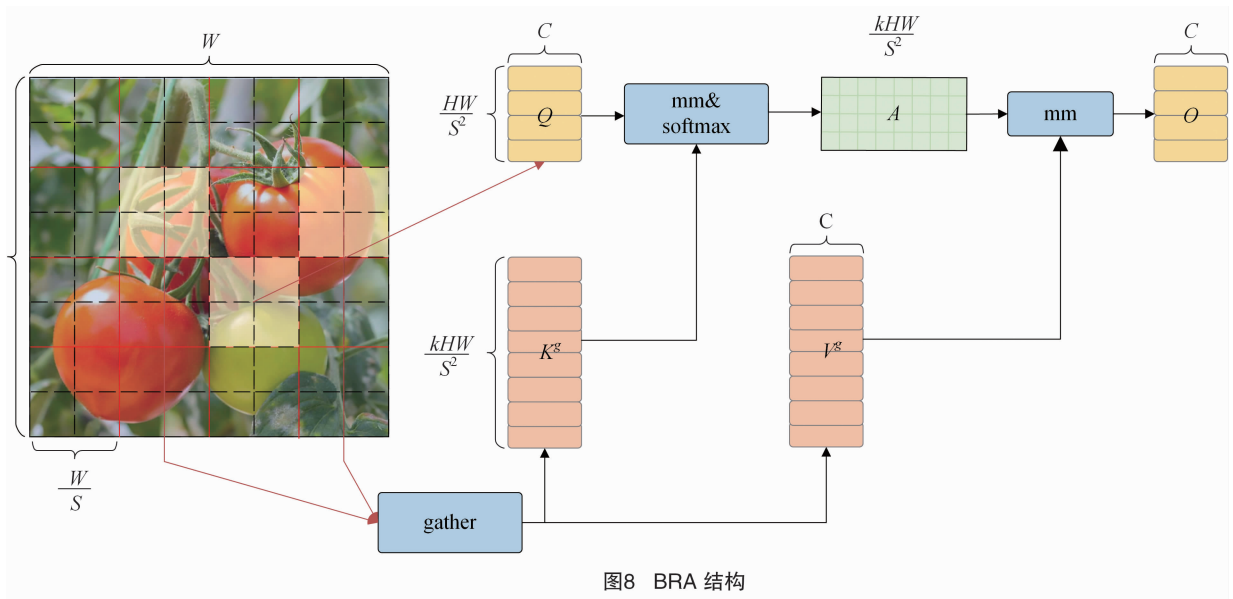


图8 BRA 结构

$$V = X^T W^T. \quad (9)$$

第二步通过有向图找到区域与区域之间的参与关系,针对 Q 、 K 分别计算每个网格平均值得到 Q^r 、 $K^r \in R^{S^2 \times C}$,进行矩阵乘法获得区域间亲和力邻接矩阵:

$$A^r = Q^r (K^r)^T. \quad (10)$$

矩阵 A^r 中的每个元素用来度量 2 个区域之间的语义相关程度。然后对每行使用 top-k 操作得到一个路由索引矩阵来获取每个区域与其他区域之间的前 K 个连接:

$$I^r = \text{topIndex}(A^r). \quad (11)$$

所以, I^r 的第 i 行包含第 i 个区域最相关的 k 个索引。第三步对于区域 i 中的每个查询标记,将根据索引向量 $I_{(i,1)}, I_{(i,2)}, \dots, I_{(i,k)}$ 实现对位于 k 个路由区域并集中的 $K-V$ 进行关注。对于 $K-V$ 的收集公式:

$$K^g = \text{gather}(K, I^r); \quad (12)$$

$$V^g = \text{gather}(V, I^r). \quad (13)$$

然后将注意力集中在收集到的 $K-V$ 上:

$$O = \text{Attention}(Q, K^g, V^g) + \text{LCE}(V). \quad (14)$$

式中: $\text{LCE}(V)$ 表示局部区域上下文增强项,函数 $\text{LCE}(\cdot)$ 使用深度卷积参数化,内核尺寸设为 5。

在 COCO 2017 数据集上,试验结果表明 BiFormer 注意力机制与一系列主流算法相比,不仅降低了算法参数量,还可以保持与目前主流的 Transformer 一样的精度水平,并在小目标检测方面展现出非常好的性能^[16]。BiFormer 的核心在于将特征图提取为区域间有向图路由,能够分布优化关

联内容选取,不再花费大量内存全局查询 $K-V$,而是更加精准地关注局部区域感受野,相较于大中目标检测,它对小目标尺寸区域抓取能力更强,优势更大。

1.3.4 改进边框损失函数 (SWD_Loss) YOLO v5s 原始模型是采用二元交叉熵损失函数和 Logits 函数计算目标得分概率。但是当数据集中包含大量小目标、遮挡对象时,其预测框往往会出现目标框内部。Wang 等提出一种新的度量方法称为归一化瓦瑟斯坦距离 (normalized Wasserstein distance, NWD),可以有效提升对缺乏外观信息的微小物体检测精度以及对微相素物体定位偏差的敏感性,并且在微小物体 (AI-TOD) 数据集上取得了不错效果^[17]。对于 2 个二维高斯分布 $\mu_1 = N(m_1, \Sigma_1)$ 和 $\mu_2 = N(m_2, \Sigma_2)$,则它们之间的 3 阶瓦瑟斯坦距离定义为:

$$W_2^2(\mu_1, \mu_2) = \|m_1 - m_2\|_2^2 + \text{Tr}[\Sigma_1 + \Sigma_2 - 2(\sum_1^{1/2} \sum_2^{1/2})^{1/2}]. \quad (15)$$

可以简化为:

$$W_2^2(\mu_1, \mu_2) = \|m_1 - m_2\|_2^2 + \left\| \sum_1^{1/2} - \sum_2^{1/2} \right\|_F^2. \quad (16)$$

式中: $\|\cdot\|_F$ 表示弗罗贝尼乌斯范数 (Frobenius norm)。对于真实框 $A = (cx_a, cy_a, w_a, h_a)$ 和预测框 $B = (cx_b, cy_b, w_b, h_b)$,高斯分布 N_a 和 N_b 进一步简化为:

$$W_2^2(N_a, N_b) = \left\| \left(\begin{bmatrix} cx_a, cy_a, \frac{w_a}{2}, \frac{h_a}{2} \end{bmatrix}^T, \begin{bmatrix} cx_b, cy_b, \frac{w_b}{2}, \frac{h_b}{2} \end{bmatrix}^T \right) \right\|_b^2. \quad (17)$$

因为 $W_2^2(N_a, N_b)$ 是距离度量单位, 最后归一化得到 NWD 公式, 其中 C 表示相关数据集的绝对平均尺寸:

$$NWD(N_a, N_b) = \exp\left(-\frac{\sqrt{W_2^2(N_a, N_b)}}{C}\right)。(18)$$

本研究借鉴其思想, 在保留原边框损失的情况下, 融入一部分瓦瑟斯坦距离系数 (θ), 可以根据待测数据集样本灵活控制系数 θ , 当数据集中常规对象居多, 可以减小系数 θ 提高 IoU 占比; 相反, 如果有大量缺乏外观信息的遮挡对象便可以增大系数 θ 提高对 NWD 占比。新的定位损失函数计算表达式如下所示, 并将其命名为遮挡瓦瑟斯坦距离损失 (shield Wasserstein distance, SWD_Loss)。针对本文数据集, 实验结果发现, 当 θ 取 0.5 时, 可以为 IoU 提供一个最佳的灵敏度阈值, 有效克服尺度敏感性和位置偏差平滑性, 提高图像中遮挡对象的漏检率。

$$SWD_Loss = (1 - \theta) \times IoU + \theta \times NWD。(19)$$

1.4 试验环境

本研究试验环境配置见表 1, 参数设置见表 2。此次试验于 2023 年 11 月 15 号在江西理工大学电气学院 415 实验室完成。

表 1 试验环境配置

参数	配置
操作系统	windows 11
CPU	Intel® Core™ i9 - 13900HX
GPU	GeForce RTX 4060
学习框架	PyTorch 2.0.0
CUDA 环境	CUDA 11.8
语言	Python 3.8

表 2 参数设置

参数	参数值
训练批次 (批)	300
批量大小	16
动量因子	0.937
初始学习率	0.01
权重衰减系数	0.000 5

1.5 试验评价指标

本研究采用查准率 P (precision)、召回率 R (recall)、平均精度 mAP (mean average precision) 和 FPS (每秒传输图像帧数) 作为试验的评价指标。计算表达式如下, 其中计算均值平均精度时将 IoU 阈值设为 0.5:

$$P = \frac{TP}{TP + FP} \times 100\%; (20)$$

$$R = \frac{TP}{TP + FN} \times 100\%; (21)$$

$$AP = \int_0^1 P(R) dR; (22)$$

$$mAP = \frac{1}{C} \sum_{i=1}^C AP_i。 (23)$$

式中: TP 表示预测为真的正样本; FP 表示预测为真实为假的负样本; FN 表示预测为假的实际为真的正样本; C 表示类别数, mAP 是由单类别精度 AP 相加后求均值得到。

2 试验结果与分析

2.1 消融试验

为了验证改进后的 YOLO v5s 模型对番茄成长周期采摘检测的有效性, 设计了 1 组消融试验。依次替换损失函数 SWD, 融入 RAAM、C3DS-ST、注意力 BiFormer, 具体试验结果如表 3 所示。其中“√”表示模型加入该模块。从表 3 中可以看出, 当各个模块单独使用时, 模型准确率和召回率会有一定的波动, 但是 mAP 均有所提升, 说明了各个模块均对提升网络检测性能有所帮助。当把优化的 4 个模块同时融入模型当中后, 与原 YOLO v5 模型相比准确率、召回率和 mAP 分别提升了 1.4、2.3、2.1 百分点, 效果较为明显。其中, 单独加入多尺度反向自适应注意力模块 (RAAM) 时对网络性能贡献最大, 由此可以验证在该模块提升特征图分辨率后, 对后续网络的计算提供了很大帮助。综上所述, 本研究提出的改进方案是有效的。

2.2 对比试验

此外, 还选取几种经典检测模型与改进 YOLO v5s 进行对比, 其中包括 SSD、Faster-RCNN、YOLO v3、YOLO v4-tiny, 并采用相同数据集和试验配置。由表 4 可知, 改进的模型在均值平均精度和召回率上均高于其他算法, 相比于 YOLO v5s, 分别提升了 2.1、2.3 百分点。虽然改进后的算法帧率有所下降, 但是仍满足实时检测的要求。

2.3 可视化分析

为了更直观地展示改进 YOLO v5s 模型的检测性能, 随机从验证集中挑取 4 张图片与原先模型进行横向对比。其中, 改进后的模型与原 YOLO v5s 模型均能完成正确检测, 但改进后的模型提供了更高的置信度, 效果如图 9 所示。另外, 针对之前 YOLO v5s 模型对番茄果实之间、枝叶遮挡的漏检问题, 改进后的 YOLO v5s 算法能准确检测出这些对象,

表 3 消融试验结果

模型	SWD	RAAM	C3DS	BiFormer	P (%)	R (%)	$mAP@0.5$ (%)
YOLO v5s					86.3	86.8	91.6
模型 1	✓				85.4	87.5	92.2
模型 2		✓			86.5	88.3	93.1
模型 3			✓		86.9	88.7	92.7
模型 4				✓	87.2	88.4	93.0
模型 5	✓	✓			87.9	88.7	93.2
模型 6	✓	✓	✓		87.6	88.9	93.5
模型 7	✓	✓	✓	✓	87.7	89.1	93.7

表 4 对比试验结果

模型	P (%)	R (%)	$mAP@0.5$ (%)	FPS (帧/s)
SSD	87.5	77.8	85.3	27
Faster-RCNN	88.2	80.9	84.3	14
YOLO v3	80.7	82.3	82.6	34
YOLO v4-tiny	90.4	87.5	90.1	48
YOLO v5s	86.3	86.8	91.6	52
改进 YOLO v5s	87.7	89.1	93.7	50

并保证较高的置信度,效果如图 10 所示。最后,为检验本研究模块对特征图丢失语义信息恢复问题,展示了 RAAM 模块加入前后的特征图以热力图可视化,试验结果如图 11 所示。经过对比可以发现,经过 RAAM 恢复的高分辨率特征图、全图区域特征得到了膨胀,番茄遮挡区域色调更暖,这表示该层有效恢复了部分丢失特征,对捕捉遮挡对象的局部关键信息提供了一定帮助。



图9 YOLO v5s 改进前后检测效果对比

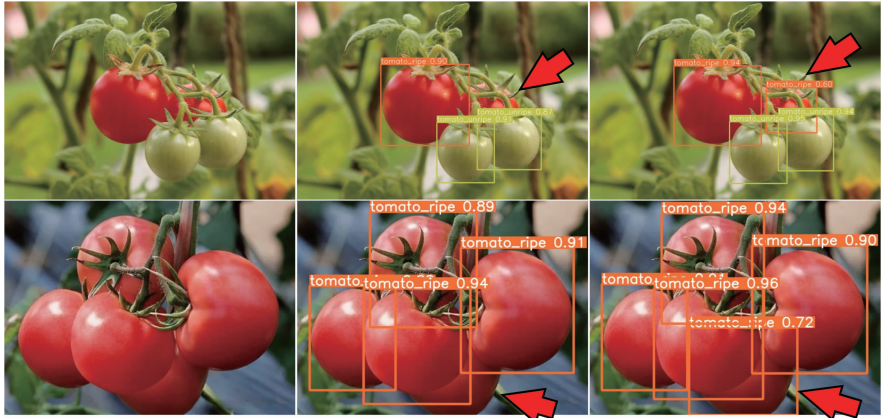
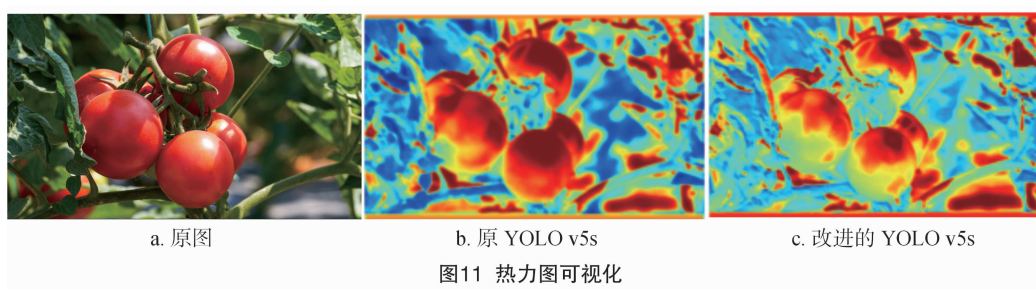


图10 YOLO v5s 改进前后漏检目标对比



3 讨论与结论

针对自然生长环境下番茄因果实、枝叶遮挡导致的误检、漏检问题,本研究提出了基于改进 YOLO v5s 的番茄检测模型。设计的多尺度反向自适应注意力模块(RAAM)提升了特征图像分辨率,有效缓解了由于连续下采样信息丢失的问题;设计新的C3DS-ST模块替换颈部C3,可以有效学习遮挡部分的坐标特征,完善局部信息的提取;提出新的边框损失函数SWD_Loss以及加入BiFormer注意力机制提升对敏感信息关注,过滤冗余信息提高计算效率。改进后的YOLO v5s模型与原算法相比,平均精度提升了2.1个百分点,不仅提高了原所有类别的置信度,而且还能提升对遮挡对象的辨识度,有效解决了实际生产中的番茄误检、漏检问题。虽然本模型一定程度上能够满足自然环境中机器人智能采摘番茄的需要,但是模型参数量有所增加。今后将通过引入模型剪枝和轻量化优化参数体积,方便部署到移动设备终端,为农业果蔬自动化采摘研究提供助力。

参考文献:

- [1]周明,李常保.我国番茄种业发展现状及展望[J].蔬菜,2022(5):6-10.
- [2]潘肖楠,张玥,耿宝龙,等.苹果采摘机器人的结构设计与分析[J].中国设备工程,2023(18):38-40.
- [3]王焱清,汤旸,杨光友.面向机器人柑橘采摘的控制系统设计与试验[J].中国农机化学报,2023,44(9):146-153.
- [4]张勤,庞月生,李彬.基于实例分割的番茄串视觉定位与采摘姿态估算方法[J].农业机械学报,2023,54(10):205-215.
- [5]宋怀波,尚钰莹,何东健.果实目标深度学习识别技术研究进展[J].农业机械学报,2023,54(1):1-19.
- [6]Redmon J,Divvala S,Girshick R,et al. You only look once:unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas:IEEE,2016:779-788.
- [7]Liu W,Anguelov D,Erhan D,et al. SSD:single shot MultiBox detector[M]//Leibe B,Matas J,Sebe N,et al. Lecture notes in computer science. Cham:Springer International Publishing,2016:21-37.
- [8]王勇,陶兆胜,石鑫宇,等.基于改进YOLO v5s的不同成熟度苹果目标检测方法[J].南京农业大学学报,2024,47(3):602-611.
- [9]Fan Y C,Zhang S Y,Feng K,et al. Strawberry maturity recognition algorithm combining dark channel enhancement and YOLO v5[J]. Sensors,2022,22(2):419.
- [10]Girshick R,Donahue J,Darrell T,et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus:IEEE,2014:580-587.
- [11]张文静,赵性祥,丁睿柔,等.基于Faster R-CNN算法的番茄识别检测方法[J].山东农业大学学报(自然科学版),2021,52(4):624-630.
- [12]岳有军,孙碧玉,王红君,等.基于级联卷积神经网络的番茄果实目标检测[J].科学技术与工程,2021,21(6):2387-2391.
- [13]Wang J F,Chen Y,Dong Z K,et al. Improved YOLO v5 network for real-time multi-scale traffic sign detection[J]. Neural Computing and Applications,2023,35(10):7853-7865.
- [14]Shi W Z,Caballero J,Huszár F,et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas:IEEE,2016:1874-1883.
- [15]Qi Y L,He Y T,Qi X M,et al. Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris:IEEE,2023:6047-6056.
- [16]Zou P,Yang K J,Liang C. Improving real-time detection of lightweight irregular driving behavior in YOLO v5[J]. Computer Engineering and Applications,2023,59(13):186-193.
- [17]Wang J W,Xu C,Yang W,et al. A normalized Gaussian Wasserstein distance for tiny object detection[EB/OL]. arXiv preprint,arXiv:2110.13389(2021-10-16)[2023-10-10]. <http://arxiv.org/abs/2110.13389v1>.