

基于改进 YOLOv5s 的莴笋芯部检测算法

代尹翹, 肖武龙, 李柏林, 李立

(西南交通大学机械工程学院, 四川 成都 610031)

摘要: 精确的作物行检测作为智能化农业的一项重要技术, 对于无人收获装置的导航和采摘具有重要意义。对莴笋生长过程中歪斜、移位和倒伏等因素导致作物行提取不准确的问题, 将其转化为莴笋芯部区域的目标检测问题, 提出一种以成熟期莴笋芯部为目标的目标检测算法。该算法基于广泛采用的目标检测框架 YOLOv5s, 通过在主干网络中嵌入动态卷积模块, 以动态感知的方式过滤特征图中的背景干扰, 在局部区域保留重要细节特征, 从而增强网络对莴笋芯部特征的学习能力。同时, 在网络的特征金字塔网络(FPN)结构中引入基于空洞卷积和权值共享的多尺度融合模块, 确保网络经过多次下采样后能够有效保留目标结构信息, 有利于对莴笋芯部这类小目标的检测。此外, 引入 CARAFE 上采样操作充分利用特征提取过程中的上下文信息, 增强网络对小目标特征的提取能力。进一步, 基于 Wasserstein 距离和 SIoU 提出一种新的损失函数, 解决了传统 IoU 方法对小目标位置敏感的问题, 并加快了网络拟合速度。实验结果表明, 改进算法对莴笋芯部提取的平均精确度和召回率分别达到了 0.586 和 0.574, 较于 YOLOv5s 提高了 6.1 和 6.3 个百分点。网络检测出莴笋芯部坐标信息后, 采用最小二乘法将坐标点进行直线拟合, 得到莴笋作物行中心线。该算法使原始 YOLOv5s 模型在不同光照条件下对莴笋芯部的漏检问题得到明显改善, 从而能够提取出更加准确的作物行中心线。

关键词: 目标检测网络; 小目标检测; 注意力机制; 多尺度特征融合; 作物行中心线

中图分类号: TP391.4

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0069727

Lettuce Core Detection Algorithm Based on Improved YOLOv5s

DAI Yinqiao, XIAO Wulong, LI Bailin, LI Li

(School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, Sichuan, China)

【Abstract】 Precise crop row detection is crucial in intelligent agriculture because it significantly affects the navigation and harvesting capabilities of unmanned harvesters. Crop row extraction accuracy is affected by factors such as slanting, displacement, and lodging during lettuce growth. This study transforms this issue into a target detection problem focusing on the core area of mature lettuce and proposes a target detection algorithm. This algorithm is based on YOLOv5s, a widely adopted target detection framework, and incorporates a dynamic convolution module into its backbone network. By dynamically filtering out background interference from feature maps, it preserves important detail features in local areas, thereby enhancing the network's ability to learn the features of the lettuce core. Additionally, the Feature Pyramid Network (FPN) structure introduces a multiscale fusion module based on dilated convolution and weight sharing, ensuring effective retention of target structural information after multiple downsampling processes, which is beneficial for detecting small targets such as lettuce cores. Furthermore, the CARAFE upsampling operation is introduced to fully utilize the contextual information during the feature extraction process, thereby enhancing the network's ability to extract small target features. Moreover, a new loss function based on the Wasserstein distance and SIoU is proposed to address the sensitivity of traditional IoU methods to the positions of small targets and accelerate the fitting speed of the network. Experimental results demonstrate that the improved algorithm achieves an average precision and recall of 0.586 and 0.574 for lettuce core extraction, representing increases of 6.1 and 6.3 percentage points compared to those achieved by YOLOv5s, respectively. After detecting the coordinates of the lettuce core, the algorithm uses the least squares method to fit the coordinate points into a straight line, thereby obtaining the central line of the lettuce crop row. Experimental results indicate that this algorithm significantly improves the performance of the original YOLOv5s model in detecting lettuce cores under different lighting conditions, thereby enabling a more accurate extraction of the crop row centerline.

【Key words】 object detection network; small target detection; attention mechanism; multi-scale feature fusion; crop row centerline

基金项目: 四川省科技厅重点研发项目(2023YFG0047)。

作者简介: 代尹翹, 男, 硕士研究生, 主研方向为机器视觉、深度学习; 肖武龙, 硕士研究生; 李柏林、李立(通信作者), 教授、博士。

收稿日期: 2024-04-11

修回日期: 2024-07-13

E-mail: lili@swjtu.edu.cn

0 引言

作物行感知技术是实现农机自动导航和精确定位的关键技术之一。依据作物行检测信息,通过监测行进方向和相对偏移量,能够引导农机完成自主作业,提高生产效率和效益。

近年来,国内外研究者在作物行感知领域进行了大量的研究。根据处理对象,主要分为基于图像和三维点云处理的感知方法^[1]。点云处理方法主要利用作物的高程信息,如 ZHANG 等^[2]采用 k -means 算法对点云进行聚类,利用最小二乘法拟合玉米行的中心线,贺静等^[3]针对水田硬底层高低不平的问题,采用动态分割阈值提取作物行点云。尽管三维点云方法具有准确度高的优势,当作物密集、行间粘连严重时(莴笋等作物),高程信息缺失则会导致此类方法无法有效提取作物行信息。

相较于点云处理方法,基于图像的作物行感知方法主要利用图像中目标的形状、纹理以及颜色等特征。面对作物稀疏、背景干扰少的识别场景,Hough 变换是一种高效的作物行提取方法。随着深度学习的发展,许多学者致力于将其应用于复杂场景下的作物行检测。如 HUANG 等^[4]创建早期水稻图像数据库,采用 ResNet 深度学习方法识别水稻秧苗,王珊珊等^[5]将水稻行识别问题转化为行间连续分布植株的特征点识别,并采用 Faster R-CNN 算法回归秧苗的特征点。针对可用特征单一导致秧苗识别在复杂田间环境下识别性能不佳的问题,HE 等^[6]提出利用高斯热力图引导点回归模型的学习,BAH 等^[7]则结合两个不同卷积神经网络(CNN)的识别结果进行最终决策。然而,利用特征点回归拟合作物行需要株间间隔清晰且与目标相似的背景干扰较少,因此点回归算法大多应用于作物苗期。莴笋采摘发生在其成熟时期,叶片茂盛株间粘连严重、背景干扰物较多,采用点回归方法的预测结果难以满足实际需求。此外,由于莴笋生长期存在倒伏、移位,因此采用语义分割算法对株行进行检测,无法解决倒伏所致分割区域偏移的问题。

针对莴笋采摘的作物行检测,将其转化为莴笋芯部区域的目标检测问题,采用目标检测方法对芯部进行检测有利于后续直线拟合时对离群的噪声点进行过滤,使算法能够更好地抵抗倒伏、移位等因素的影响。不同于点回归方法在语义模糊区域预测高置信度点,目标检测算法趋向于在单个莴笋株内回归芯部区域,能更好地避免网络训练的过拟合。尽管近些年来目标检测算法得到了长足的发展,针对

莴笋芯部的目标检测依然存在以下挑战:

1) 莴笋芯部所占图像区域少且存在遮挡等问题,这导致用于识别的可提取特征信息少,属于小目标检测问题。改善高层特征的感受野,增强模型特征提取能力,被证明是提高小目标检测性能的有效方法。徐光达等^[8]通过多路径特征传输对主干网络中的多尺度特征融合方法进行改进,改善了算法对航拍图像中细小物体的检测性能。LIU 等^[9]采用两个并联的特征金字塔网络(FPN)构造更丰富的多尺度特征表示,但该方法也显著增加了网络的计算开销,使得算法难以被部署到嵌入式设备当中。庄集超等^[10]通过采用可变形密集卷积神经网络和密集连接的方式,增强了特征的语义表达能力。赵楚等^[11]在 Faster R-CNN 的基础上引入了可变形卷积和 Rank&Sort Loss,有效提高了小目标检测精度,但这些方法也增加了整个检测流程的复杂性,并降低了推理速度。赵继达等^[12]基于 YOLOv8 网络设计空间-通道滤波模块,增强目标的特征信息,滤除无用信息。但是当特征信息较少时,过滤信息操作可能会丢失特征,导致检测精度降低。

2) 莴笋芯部检测同时受到农田环境中光照、背景干扰物等因素的影响。注意力机制模仿人眼使网络学习聚焦于目标区域,被广泛地用于检测算法中用于抵抗识别干扰。ZHENG 等^[13]为了降低青枣在严重遮挡条件下的漏检率,通过嵌入注意力机制的方式强化前景特征的表达。袁天乐等^[14]通过嵌入 Transformer 多头注意力机制实现对轴承表面缺陷检测性能的提升。尽管嵌入此注意力机制获得了准确率提升,但改进后算法的检测速度明显降低。

针对莴笋芯部检测因目标尺度小,且同时受到环境及背景干扰引起的漏检和误检率较高的问题,本文基于检测性能和预测时间取得较好平衡的目标检测网络 YOLOv5s,提出如下改进:

1) 改变原始主干网络中卷积层的结构,在输出端嵌入动态稀疏注意力机制,提出全新的卷积模块 Bi-CBS。该模块具备动态查询性,通过查询与目标区域相似度高的区域,实现更灵活的计算分配,同时聚焦于细粒度特征,使模型在性能和计算开销间取得良好平衡。

2) 改进其特征融合方法,提出一种基于空洞卷积和权值共享的多尺度融合方法,在增大网络中特征图感受野的同时减少下采样过程中对特征结构信息的损失,从而减少模型对芯部的漏检。在颈部(Neck)部分,引入预测内核和特征重组的上采样操作(CARAFE)^[15],在不同层级间更有效地聚合上下

文信息。

3)提出一种全新的损失函数,将基于归一化高斯 Wasserstein 距离(NWD)^[16]的损失函数和基于 SIoU^[17]的损失函数相结合,计算回归框之间二维高斯分布的 Wasserstein^[18]距离作为其相似度并限制网络回归方向,解决莴笋芯部识别过程中回归框位置不准确与拟合速度较慢的问题。

1 所提方法

1.1 基于动态稀疏注意力机制的卷积模块

在彩色莴笋图像中,莴笋芯部包含的像素较少,而且在识别过程中会受外围莴笋叶片的干扰,因此本文设计了一种基于动态稀疏注意力机制的卷积模块,加入到 YOLOv5s 主干网络中。图 1 所示为两类注意力机制实现效果的对比(彩色效果见《计算机工程》官网 HTML 版,下同),其中,五角星为标签的莴笋芯部特征,三角形为待匹配的莴笋芯部特征,阴影部分表示参与特征匹配的区域。目前常用的注意力机制有 SE^[19]和 CA^[20],都是在一个全局范围获取重点关注目标,如图 1(a)所示。而本文用到的动态稀疏注意机制(BRA)^[21]通过特征匹配的方式实现粗粒度区域中不相关内容的过滤,在获取注意图的同时节省计算开销。如图 1(b)所示,对于输入特征和标签区域,动态稀疏注意机制构造一个局部关联图(黄色阴影部分),并对其进行修剪,获取与目标区域相似度高的其他区域信息,从而减少不相关特征对于检测的影响。

为了动态提取特征相关区域,本文通过邻接矩阵构建区域级注意力关系图,定量表示不同区域间的相关性,最后根据区域路由索引矩阵,提取出最相关的区域,BRA 注意力机制的伪代码流程如算法 1 所示。

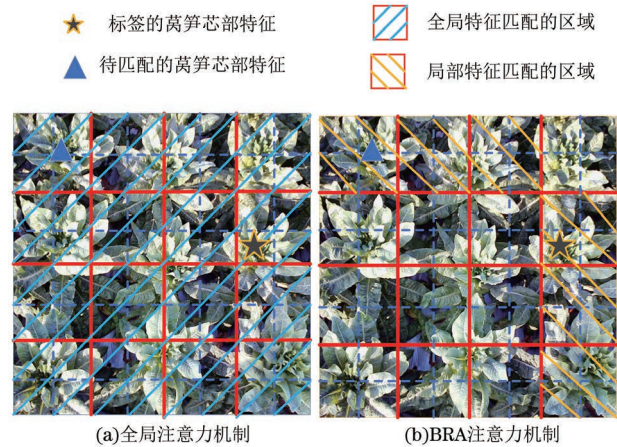


图 1 注意力机制实现效果对比

Fig.1 Comparison of attention mechanism implementation effects

算法 1 Pseudocode of BRA

输入 features(H, W, C), Assume $H = W$

输出 features(H, W, C)

S: square root of number of regions

k: number of regions to attend

1. $x = \text{patchify}(\text{Input}, \text{patch_size} = H//S)$

2. $\text{query}, \text{key}, \text{value} = \text{linear_qkv}(x). \text{chunk}(3, \text{dim} = -1)$

3. $\text{query_r}, \text{key_r} = \text{query}. \text{mean}(\text{dim} = 1), \text{key}. \text{mean}(\text{dim} = 1)$

4. $A_r = \text{mm}(\text{query_r}, \text{key_r}. \text{transpose}(-1, -2))$

5. $I_r = \text{topk}(A_r, k). \text{index}$

6. $\text{key_g} = \text{gather}(\text{key}, I_r)$

7. $\text{value_g} = \text{gather}(\text{value}, I_r)$

8. $A = \text{bmm}(\text{query}, \text{key_g}. \text{transpose}(-2, -1))$

9. $A = \text{softmax}(A, \text{dim} = -1)$

10. $\text{output} = \text{bmm}(A, \text{value_g}) + \text{dwconv}(\text{value})$

11. $\text{Output} = \text{unpatchify}(\text{output}, \text{patch_size} = H//S)$

bmm: batch matrix multiplication; mm: matrix multiplication; dwconv: depthwise convolution.

BRA 注意力机制原理如图 2 所示。

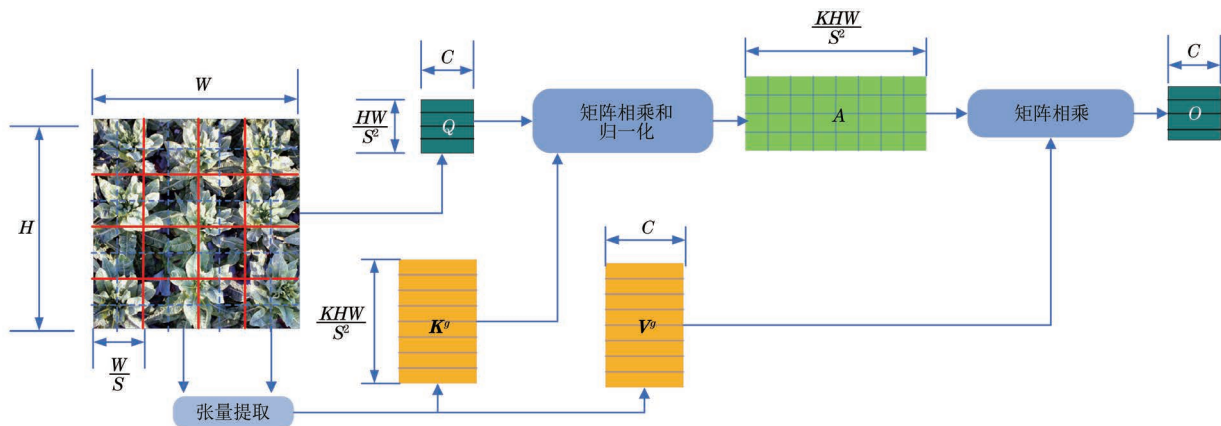


图 2 BRA 注意力机制原理

Fig.2 BRA attention mechanism principle

对于输入特征图 $X \in \mathbb{R}^{H \times W \times C}$, 将其划分为 $S \times S$ 个不重叠的区域, 提取区域特征并重组得到特征 $X^r \in \mathbb{R}^{S^2 \times (HW/S^2) \times C}$ 。以线性映射的方式分别推导出不同特征映射图 $Q, K, V \in \mathbb{R}^{S^2 \times (HW/S^2) \times C}$, 具体表达式如下:

$$Q = X^r \times W^q \quad (1)$$

$$K = X^r \times W^k \quad (2)$$

$$V = X^r \times W^v \quad (3)$$

式中: W^q, W^k, W^v 分别为不同投影权值矩阵。

为了构建区域级注意力关系图, 首先导出特征映射图 Q 和 K 中区域级分量 Q^r 和 K^r ($Q^r, K^r \in \mathbb{R}^{S^2 \times C}$) 并施加平均操作, 分别作为查询 (queries) 和键 (keys)。利用归一化点积计算得到区域间的亲和力图 $A^r \in \mathbb{R}^{S^2 \times S^2}$:

$$A^r = Q^r (K^r)^T \quad (4)$$

式中: A^r 中的元素表示两个区域在语义上的潜在关联程度。因此可以利用 topkIndex 算法修剪 A^r 中的元素, 就能保留与查询区域相关性最大的区域表示为 top- k 。具体的, 对 A^r 中的各行实行 topkIndex 操作, 从而获取对应的索引矩阵 I^r 。

$$I^r = \text{topkIndex}(A^r) \quad (5)$$

式中: topkIndex 是检索 top- k 的函数。

I^r 的每一行包含了 k 个与对应区域相关性最高的区域索引。根据索引提取对应查询和键值张量:

$$K^g = \text{gather}(K, I^r) \quad (6)$$

$$V^g = \text{gather}(V, I^r) \quad (7)$$

式中: gather 是根据索引矩阵 I^r 从输入张量中提取数据并生成新的组合张量的函数。

最后, 利用查询和键值矩阵构建注意力机制:

$$O = \text{Attention}(Q, K^g, V^g) + \text{LCE}(V) \quad (8)$$

式中: Attention 是多头自注意力机制函数; LCE 是基于深度卷积的局部上下文增强函数。

本文改变原始主干网络中卷积层的结构, 在输出端嵌入动态稀疏注意力机制, 提出全新的卷积模块 Bi-CBS, 其结构如图 3 所示。Conv $_{3 \times 3}$ 、BN、SiLU 分别表示采用 3×3 卷积核的卷积、特征归一化层以及 SiLU 激活函数。

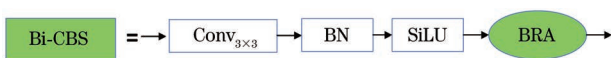


图 3 Bi-CBS 卷积模块结构

Fig.3 Bi-CBS convolutional module structure

1.2 基于空洞卷积和权值共享的多尺度特征融合方法

针对莠笋芯部检测, 本文通过在 YOLOv5s 的

空间金字塔池化结构 (SPPF) 中引入感受野增强 (RFE) 模块^[22], 旨在扩大高层特征的感受野, 增强小目标特征的响应。该模块基于空洞卷积和权值共享的设计理念, 通过多尺度融合的方式, 在增大特征图感受野的同时保持分辨率, 从而提高小目标检测的准确性。

如图 4 所示, REF 采用 3 个卷积模组分别处理输入特征以提取包含不同大小感受野的特征。卷积模组由 2 个采用 1×1 卷积核的卷积层和 1 个扩张率不同的 3×3 卷积层组成 (三支中扩张率分别为 1、2 和 3)。在 3 个卷积模组中, 所有的分支都有共同的权值。首先这样做减少了参数的数量, 从而降低潜在过拟合的风险。其次共享权值可以充分利用每个样本的信息, 平衡不同分支的表示。短接层被嵌入用于防止训练过程中梯度爆炸问题。

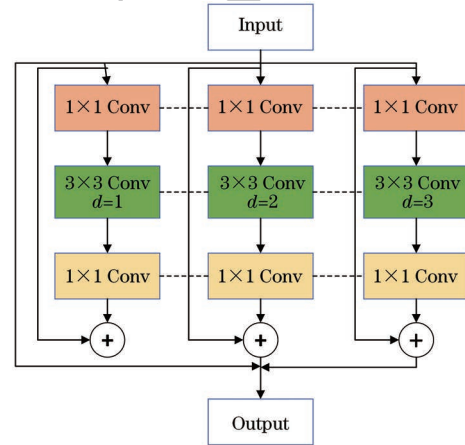


图 4 RFE 模块结构

Fig.4 RFE module structure

感受野的大小不同意味着捕获远程依赖的能力不同, 因此本文在设计 SPPF 模块时引入 RFE 模块, 充分利用扩展卷积对特征图中的感受野进行分析, 利用不同扩展率的卷积操作进一步捕获不同范围的依赖信息。此外, 由于所有分支共享权值, 减少了该模块的计算开销, 从而降低潜在过拟合的风险。RFE-SPPF 的结构如图 5 所示。

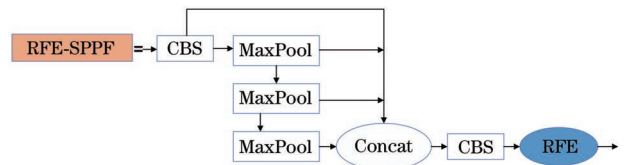


图 5 多尺度融合模块结构

Fig.5 Multi-scale fusion module structure

1.3 预测内核和特征重组的上采样算子

莠笋图像的特征提取受到莠笋芯部区域尺寸小以及图像分辨率低等因素影响, 导致能够用于莠笋芯部识别的有效信息较少。在这种情况下, 有效地

聚合上下文信息,理解场景与目标之间的关系对于莠笋芯部检测至关重要。本文在 YOLOv5s 的颈部模块中引入内容感知特征重组上采样算子替代最近邻插值(Nearest)上采样操作。最近邻插值仅仅依据邻域信息进行上采样操作,未能有效地利用图像的结构信息。因此,本文引入了基于上下文信息的特征重构模块 CARAFE,其首先通过采样核预测函数自适应求取不同上采样核尺寸,然后根据自适应核尺寸对该级特征进行上采样。由于该方法在采样核位置和尺度上的灵活性,能够捕获更多的上下文信息,有利于莠笋芯部的检测。

如图 6 所示,对于任意给定的输入特征 X ,假设上采样率为 σ ,输出特征图 X_o 。 X_o 中任意像素点位置表示为 $l_o = (i_o, j_o)$,其对应于输入特征图的像素位置表示为 $l = (i, j)$,其中 $i = i_o/\sigma, j = j_o/\sigma$ 。此外本文用 $N(X_l, k)$ 表示以像素位置 l 为中心、面积为 $k \times k$ 的领域。利用上采样核预测模块 ψ 预测 $N(X_l, k)$ 区域的上采样核 W_l ,该过程如式(9)所示:

$$W_{l_o} = \psi(N(X_l, k_{\text{encoder}})) \quad (9)$$

式中: k_{encoder} 表示上采样过程中采用卷积操作的卷积核尺寸。

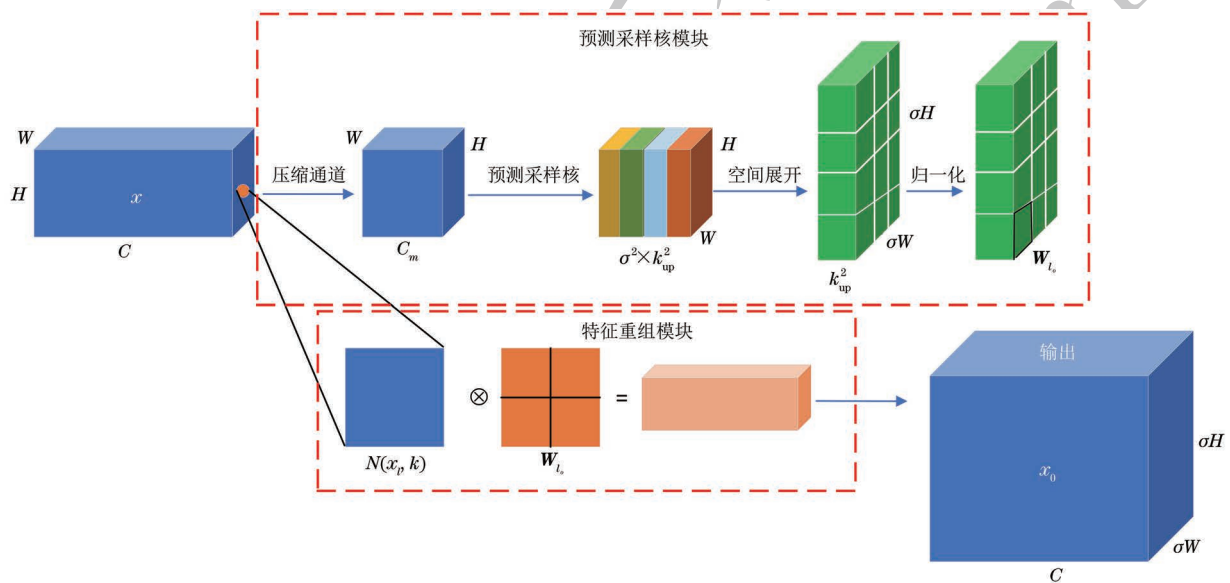


图 6 CARAFE 上采样算子原理

Fig.6 CARAFE upsampling operator principle

对于给定的输入特征 $X \in \mathbb{R}^{H \times W \times C}$,首先将其通道数量压缩至 C_m 以减少后续过程计算开销。对于压缩后的特征图,利用 $k_{\text{encoder}} \times k_{\text{encoder}}$ 卷积核对其实施卷积操作,其输出特征图的通道数量为 $k_{\text{up}}^2 \times \sigma^2$,其中, σ 为上采样率, k_{up} 表示预测采样核尺寸,两者都为超参数。增加 k_{encoder} 和 k_{up} 可以扩大特征的接受域,在更大的区域利用上下文信息,这对上采样具有重要意义。然而,计算复杂度也会随着内核大小的增长而增长。因此在第 2.3.2 节中设置实验来确定 k_{encoder} 和 k_{up} 的数值,以保证模型在性能和效率之间的良好权衡。

将卷积输出的特征图展开并对展开后的元素进行归一化处理,得到包含采样核的矩阵 $X_w \in \mathbb{R}^{\sigma H \times \sigma W \times k_{\text{up}}^2}$ 。

依据所获取的自适应采样核对输入特征 $X \in \mathbb{R}^{H \times W \times C}$ 进行重组,其过程可表示为:

$$X = \varphi(N(X_l, k_{\text{up}}), W_{l_o}) \quad (10)$$

式中: φ 为特征重组模块,该模块利用预测得到的上采样核矩阵 W_{l_o} 与输入特征图中的邻域 $N(X_l, k_{\text{up}})$ 作点积,完成该区域的上采样。

在该方法中,输入特征图的尺寸为 $H \times W \times C$,上采样率为 σ ,卷积核尺寸 k_{encoder} 预测采样核尺寸为 k_{up} 。由于上采样核预测模块需要对每个像素点进行自适应计算,预测过程的时间复杂度为 $O(H \times W \times C \times k_{\text{encoder}}^2)$ 。特征重组过程涉及 $k_{\text{up}}^2 \times \sigma^2$ 个卷积操作,其时间复杂度为 $O(\sigma^2 \times H \times W \times C \times k_{\text{up}}^2)$ 。因此,总的时间复杂度为 $O(H \times W \times C \times k_{\text{encoder}}^2 + \sigma^2 \times H \times W \times C \times k_{\text{up}}^2)$ 。

在空间复杂度方面,主要体现在存储采样核矩阵和展开后的特征图。上采样核矩阵的尺寸为 $H \times W \times \sigma^2 \times k_{\text{up}}^2$,特征图的尺寸为 $H \times W \times C \times \sigma^2$ 。因此,总的空间复杂度为 $O(H \times W \times \sigma^2 \times k_{\text{up}}^2 + H \times W \times C \times \sigma^2)$ 。

由于上采样核是由输入特征的具体内容预测而

来的,输入特征图的不同位置生成不同的采样核,因此 $N(X_l, k_{up})$ 区域内的每个像素对于输出特征图的每个像素贡献是不同的,局部区域中目标点信息和其周围场景信息的联系能够得到更多关注,从而实现基于特征上下文信息的上采样操作。

1.4 基于二维高斯分布的目标检测评估方法

传统 IoU 损失通常被引入来消除回归训练与测试间的性能差距。然而,当真实框被预测框完全覆盖时,IoU 损失趋向稳定,此类预测则无法为网络训练提供有效的反向传播,这在目标尺度较小时尤为明显。此外,尺度较小的预测框的微小坐标偏差会导致其与目标框间变得毫无重合,这易使得网络训练出现波动而难以拟合。针对小尺度莴笋芯部的目标检测,本文基于归一化高斯 Wasserstein 距离(NWD)来度量目标框与预测框之间的相似性。

该方法将预测框与目标框建模为二维高斯分布 $\mathcal{N}_p$ 和 $\mathcal{N}_g$,并归一化高斯 Wasserstein 距离(NWD, N_{NWD})为:

$$N_{NWD}(\mathcal{N}_p, \mathcal{N}_g) = \exp\left(-\frac{\sqrt{W_2^2(\mathcal{N}_p, \mathcal{N}_g)}}{C}\right) \quad (11)$$

式中: C 为超参数; $W_2^2(\cdot)$ 为距离度量。 $W_2^2(\cdot)$ 定义为:

$$W_2^2(\mathcal{N}_p, \mathcal{N}_g) = \left\| \left[cx_p, cy_p, \frac{\omega_p}{2}, \frac{h_p}{2} \right]^T, \left[cx_g, cy_g, \frac{\omega_g}{2}, \frac{h_g}{2} \right]^T \right\|_2^2 \quad (12)$$

式中: x_p, y_p, ω_p, h_p 和 x_g, y_g, ω_g, h_g 分别表示预测框与目标框的位置和形状参数。

依据归一化高斯 Wasserstein 距离,本文将莴笋芯部目标检测的回归损失重置为:

$$L = 1 - N_{NWD}(\mathcal{N}_p, \mathcal{N}_g) \quad (13)$$

由于莴笋芯部不全是小尺度目标,完全基于 NWD 损失将会显著增加模型的拟合所需时间。本文进一步引入 SIoU,通过引入回归框的向量角度,能够限制回归方向,提升回归的速度和准确性,加速网络的收敛。最终本文提出损失函数为:

$$L_{NS} = \lambda L + (1 - \lambda) L_{SIoU} \quad (14)$$

式中: λ 为超参数, $\lambda \in (0, 1)$ 。

将 L_{NS} 引入到 YOLOv5s 的 Head 网络部分,能够增强模型在边界框回归阶段的位置准确性。面对莴笋芯部检测场景时, L_{NS} 可以确保边界框回归的高质量输出和网络快速拟合。

1.5 算法整体架构

本文所提算法整体架构如图 7 所示,包括 3 个

最主要的部分:特征提出部分(BackBone),多尺度特征融合部分(Neck)和分类、检测头部分(Head)。主干网络主要用于下采样输入图像并提取多层级特征。颈部部分主要通过多尺度特征融合来获取更大的感受野。检测头主要用于预测目标位。在特征提取部分,本文算法将原网络的第 4 层和第 6 层的 CBS(卷积+正则化+SiLU 激活函数)替换为 Bi-CBS(卷积+正则化+SiLU 激活函数+BRA 注意力机制),BRA 的动态稀疏性使得网络在提取特征时会自动过滤掉干扰信息,提高特征匹配的效率和准确率。用 Bi-CBS 替换第 4 层和第 6 层的 CBS,首先是为了避免将网络的卷积层全面替换导致的计算开销过大,其次在网络的浅层应用此模块(第 4 层和第 6 层),是因为网络的浅层包含更多的细节特征,从而能充分地利用 BRA 的动态稀疏性来进行特征匹配。此外,将原始 SPPF 模块替换为具有多尺度特征提取能力的 RFE-SPPF 模块,该模块在空洞卷积设计的基础上加入了权值共享连接,从而能够在增大特征图感受野的同时不会丢失语义细节特征,助力小目标检测。

在多尺度特征融合部分(Neck),将原始的上采样模块替换成 CARAFE 上采样模块,该模块能够生成自适应的上采样核,从而更有效地捕获上下文信息,提高模型聚合上下文信息的能力,最终增强网络对小目标的检测效果。

在分类、检测头部分(Head),提出了损失函数 L_{NS} 。该损失函数融合了 NWD 损失函数和 SIoU 的损失函数的优点,先计算边界框之间归一化高斯分布的 Wasserstein 距离作为评价指标;然后考虑回归的向量角度,限制参数的回归方向。该方法能够有效地降低回归框对小目标位置偏移的敏感度,从而获得更加准确的回归框,同时能大大提高网络的拟合速度。

2 实验数据与设置

2.1 图像采集处理与数据集构建

实验中所采集的莴笋图像源自四川省成都市周边的莴笋智能采摘技术试验示范基地,采集时间为 2023 年 4 月,该时间段为莴笋成熟期,适宜进行采摘。采用 Intel RealSense L515 LiDAR 相机进行图像采集,该相机固定在小车上方,摄像头距离地面 180 cm,光轴与水平线夹角为 $80^\circ \sim 90^\circ$,实现了俯视拍摄。图像格式为 PNG,大小为 1280×720 像素,共采集了强光和弱光条件下的莴笋种植图像 800 张。600 张图像用于识别算法的训练和验证,其

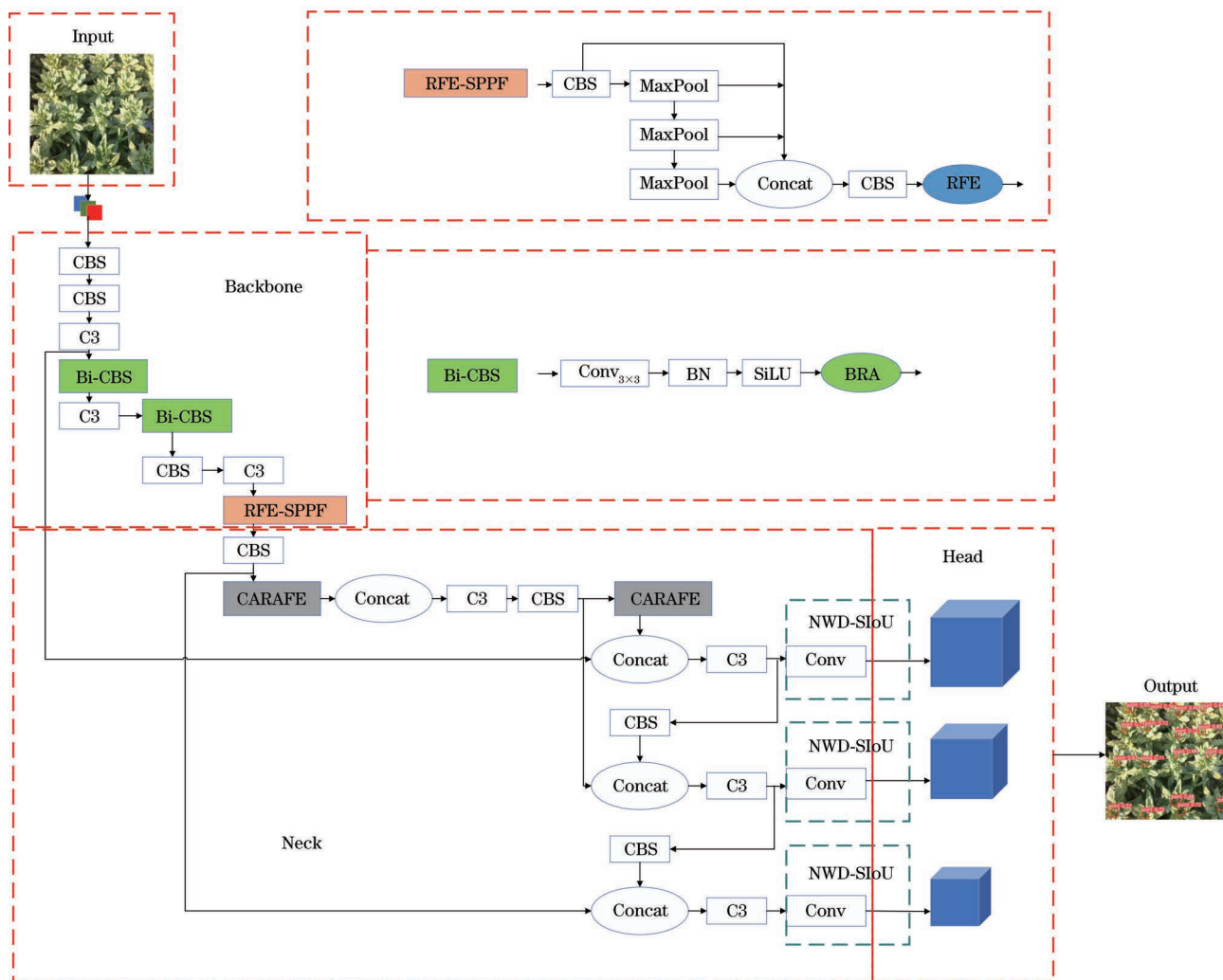


图 7 改进模型的结构

Fig. 7 Structure of improved model

中,450 张用于训练,150 张用于验证。另外,还有 200 张未标记的图像用于测试算法的识别效果。研究的目标是识别莴笋的中心顶端,即莴笋芯部。鉴于原图中莴笋芯部特征不够明显,将原图放入模型训练可能导致过拟合问题。因此,本文采用了对比度自适应直方图均衡 CLAHE 方法进行图像处理,该方法能够增强对比度并抑制噪声。图像处理前后的对比如图 8 所示。图像标注使用 LabelMe 软件进行手动标注莴笋芯部。

2.2 实验设置及实验评价

在实验环境方面,本文选择了 PyTorch 深度学习框架作为软件工具,并搭建了一台配备 Intel® Core™ I5-9400F CPU@2.90 GHz、NVIDIA GeForce GTX 1080 GPU、32 GB 运行内存的工作站作为硬件设备。实验设置了 300 个训练周期(Epochs)以确保模型充分收敛。单次处理图像数量(batchsize)合理设定为 8,以平衡计算效率与内存占用;学习率为 0.01,可以帮助模型更快地收敛,同时避免在最优解附近震

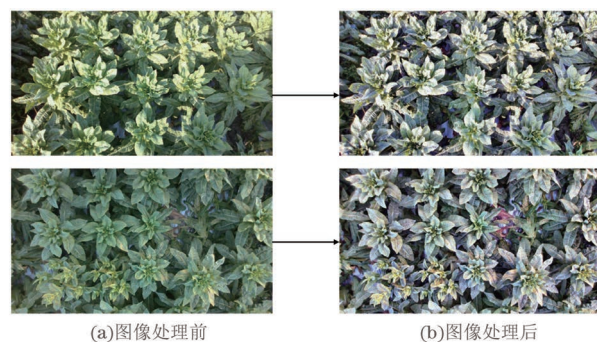


图 8 图像处理前后对比

Fig. 8 Comparison before and after image process

荡;学习率衰减采用余弦退火衰减算法;在优化算法方面,考虑到大数据量处理的需求以及训练过程中的梯度更新特性,本文采纳随机梯度下降(SGD)作为优化器以驱动整个神经网络训练流程。

为了保证实验的公平性,每次实验的训练轮数均设置为 300 Epochs,并选取了召回率(Recall)、平均精确度(Average Precision)、参数量(Params)、模型复杂度(GFLOPs)、帧速率(FPS)5 个评估指标来

评估模型的性能。

召回率:指所有的正样本中正确识别的概率。

平均精确度:以精确率为纵坐标、召回率为横轴坐标的二维曲线下的面积,综合了精确率和召回率,能够更全面地衡量模型的性能。

模型参数量:指在模型中所有可训练权重参数的总数。模型参数量大小直接影响模型的复杂度和对硬件资源的需求。

模型复杂度:代表模型在推理过程中 1 s 内执行的浮点运算次数,可用来评估模型的计算复杂度。

帧速率:指模型每秒钟能够处理并输出检测结果的静止图像帧的数量,帧速率的值越高,模型的实时响应能力越强,运行效率和数据处理吞吐量也就越高,也就能更好地兼容和整合其他实时系统。

2.3 对比实验与结果分析

2.3.1 Bi-CBS 模块性能对比实验

表 1 所示的是通过在原始网络上使用不同数量的 Bi-CBS 卷积模块进行训练,研究其对模型性能和计算开销的影响。实验基线模型使用 YOLOv5s,在其主干网络的浅层(1、2、4、6 层)分别使用 1、2、3、4 个 Bi-CBS 模块和单独使用 BRA 注意力机制。

表 1 不同 Bi-CBS 模块数量对模型性能的影响

Table 1 Effect of different Bi-CBS module numbers on model performance

数量	Average Precision	Recall	GFLOPs
0	0.525	0.511	15.8
1	0.536	0.528	16.9
2	0.544	0.529	17.8
3	0.551	0.548	18.9
4	0.549	0.537	20.3
+BRA	0.499	0.508	17.3

根据 Bi-CBS 模块使用位置和使用数量的不同,整个模型的检测精度和计算成本也不同。综合考虑算法精度、复杂度,实验结果表明,将 Bi-CBS 模块替换原主干网络中的第 4 层和第 6 层卷积模块的综合效果最好,此时模型的 Average Precision 和 Recall 分别达到 54.4%,和 52.9%,GFLOPs 仅增加 2.0。将 Bi-CBS 模块替换原来的第 2 层、第 4 层和第 6 层卷积,虽然模型的 Average Precision 和 Recall 分别达到了 55.1%和 54.8%,但是 GFLOPs 也有 18.9,较原模型增加了 19.6%,大大提高了计算开销。同时将 Bi-CBS 替换主干网络中的卷积比单独使用 BRA 注意力机制计算开销更小且检测效果更佳,表明 Bi-CBS 更好地使网络与 BRA 模块进行融合。

综上所述,此方法在增加少量计算开销的同时,

能够充分利用 BRA 的动态稀疏性,聚焦有价值的特征信息,增强网络的特征匹配能力。

2.3.2 CARAFE 上采样算子性能对比实验

表 2 展示了不同上采样方法对模型性能的影响。基线采用的是 Nearest 上采样。在 CARAFE 中有编码器内核大小 k_{encoder} 和重组内核大小 k_{up} 两个参数,它们对网络是否能够充分有效地利用上下文信息至关重要。为了验证该上采样能够提升模型的小目标检测性能,并设置最佳的 k_{encoder} 与 k_{up} 值,将对使用不同的上采样操作的相同模型进行对比,实验基础模型使用 YOLOv5s,训练 300 轮。实验结果如表 2 所示。

表 2 上采样操作对比实验

Table 2 Comparison experiment of upsampling operations

上采样	Average Precision	Recall	GFLOPs	Params
Nearest	0.525	0.511	15.8	7 012 822
Bilinear	0.537	0.533	15.8	7 012 822
CARAFE(1,3)	0.532	0.513	15.8	7 042 206
CARAFE(3,5)	0.543	0.536	16.0	7 152 926
CARAFE(5,7)	0.529	0.518	17.1	7 665 118

由表 2 可知,当上采样操作采用 CARAFE(3,5)时模型的检测性能可达到最优水平,这时模型的 Average Precision 和 Recall 较基础模型的 Nearest 上采样方法提高了 1.8 和 2.5 百分点。而模型的 GFLOPs 和 Params 只增加了 0.2 和 140 104,并没有大幅增加模型的计算开销。CARAFE 通过生成自适应的上采样核,能够在一个可动态调节的大尺度上聚合上下文信息,使网络能够发掘出更多特征,最终增加模型的检测精度。

2.3.3 综合对比实验

为了体现本文所提算法在检测精度和速度上的优越性,将其与目前主流的目标检测模型进行对比,在同样的实验环境与参数的条件下对莒县数据集进行检测,模型对比实验结果如表 3 所示。由表 3 可知,与 YOLO 系列算法相比,本文算法在检测精度和速度上效果优异。YOLOv5n^[23] 的 Average Precision 和 Recall 仅有 46.2%和 47.3%,无法满足检测准确性的要求;YOLOv7^[24] 的 Average Precision、Recall 和 FPS 均低于 YOLOv5s;YOLOv8s^[25] 和 YOLOv9-c^[26] 的 Average Precision 和 Recall 略高于 YOLOv5s,将本文设计的模块与其融合后虽然 Average Precision 略有提高,但是模型的推理速度远远不及本文所提算法,均不足 30 帧/s,达不到检测的实时性要求;综上所述,本文算法采用 YOLOv5s 作为基线模型进行研究。

表 3 模型对比实验

Table 3 Model comparison experiments

Model	Average Precision	Recall	FPS/ (帧·s ⁻¹)	Params/ 10 ⁶
YOLOv5s	0.525	0.511	61.4	7.0
YOLOv5n	0.462	0.473	70.9	3.6
YOLOv7	0.493	0.486	41.2	13.5
YOLOv8s	0.541	0.533	49.8	11.1
YOLOv8s+ 所提模块	0.584	0.576	24.3	9.4
YOLOv9-c	0.539	0.524	38.6	51.0
YOLOv9-c+ 所提模块	0.579	0.568	18.6	38.7
Faster R-CNN	0.496	0.504	7.25	41.3
Cascade R-CNN	0.503	0.513	6.37	69.4
SSD300	0.502	0.482	39.8	34.3
RT-DETR	0.527	0.508	13.5	42.7
ours	0.586	0.574	42.9	2.3

将本文算法与当前主流的两阶段检测器 Faster R-CNN^[27]、Cascade R-CNN^[28] 进行对比,Faster R-CNN、Cascade R-CNN 检测精度有所下降,而且伴随着更大的计算开销和更耗时的推理速度。因为更深的模型结构意味着更多的参数和更复杂的特征提取机制,训练时往往会丢失部分特征。由于莴笋芯部可利用的特征少,只是增大网络层数反而难以收敛,导致检测精度降低。

单阶段检测器 SSD^[29] 由于其多尺度预测需要生成大量的预测框,因此检测速度较慢。RT-DETR^[30] 作为最新的端到端目标检测器,对 DETR 系列算法进行了多方面的改进和优化。该模型引入了高效的混合解码器,并通过优化 token 和多尺度特征的采样利用,进一步提升了目标检测性能。然而,在本文实验的莴笋芯部检测任务中,RT-DETR 未能表现出优异的效果。首先因为以 Transformer 为架构的模型通常在更大的数据集上才能充分发挥其优势,对于较小规模的数据集任务,以 YOLO 为基础的模型通常是

一个更好的选择^[31];其次基于 Transformer 架构的 RT-DETR 对全局上下文信息过于敏感,可能会忽略局部细节,从而无法有效提取小目标的特征。

综上本文所提模型在平均精度、召回率和推理速度上都具有较大的优势。Average Precision 和 Recall 指标分别为 58.6%、57.4%,均优于 YOLOv5s、YOLOv7、YOLOv8s、Faster-RCNN 等主流目标检测模型。所提模型的检测速度为 42.9 帧/s,可以满足莴笋芯部检测对实时性要求,从而实现了检测精度和速度之间最佳权衡;而且所提模型的模型参数量仅有 2.3×10^6 ,明显优于其他目标检测器,在提高检测精度的基础上进一步实现了轻量化,表现出卓越的性能。

YOLOv5s 和本文所提模型的边界框损失(Box Loss)随训练轮次(epoch)变化的趋势如图 9 所示。随着训练进程的推进,前 50 个 epoch 内两个模型的 Box Loss 均快速下降。可以看出,本文所提模型的 Box Loss 下降速度明显更快,呈现出更加陡峭的下降曲线,表明其具有更高的网络拟合效率和学习速度。

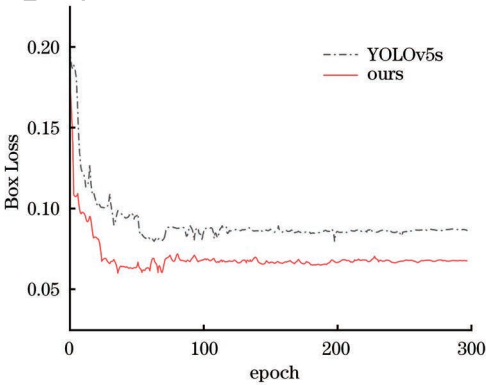


图 9 边界框损失曲线对比

Fig.9 Comparison of bounding box loss curves

2.3.4 消融实验

通过对每个改进模块进行消融实验来验证所提方法的有效性,实验结果如表 4 所示。在该实验中基线模型采用 YOLOv5s。

表 4 消融实验结果

Table 4 Ablation experiment results

基线	Bi-CBS	RFE	CARAFE	L_{NS}	Average Precision	Recall	GFLOPs
✓	×	×	×	×	0.525	0.511	15.8
✓	✓	×	×	×	0.544	0.529	17.8
✓	×	✓	×	×	0.568	0.564	16.6
✓	×	×	✓	×	0.543	0.536	16.0
✓	×	×	×	✓	0.548	0.535	15.8
✓	✓	✓	×	×	0.572	0.549	18.0
✓	✓	✓	✓	×	0.581	0.567	18.3
✓	✓	✓	✓	✓	0.586	0.574	18.3

将基线模型主干中的卷积模块替换为 Bi-CBS 卷积模块后,虽然 GFLOPs 增加了 2.0,但是 Average Precision 提升了 1.9 百分点,说明 Bi-CBS 卷积模块可以滤除与目标不相关的区域并提高对有价值区域的关注,增强模型对小目标定位的能力;在基线模型的多尺度融合网络上添加 RFE 模块后, Average Precision 和 Recall 指标上的提升最大,分别提高了 4.3、5.3 百分点,说明 RFE 感受野增强模块在增大特征图感受野的条件下没有丢失特征图细节特征,从而提升模型检测小目标的效果;将基线模型的上采样操作替换成 CARAFE 后, Average Precision 和 Recall 分别提高 1.8 和 2.5 百分点,证明了 CARAFE 上采样算子所具备的融合上下文信息的能力对于检测任务是有效的;将基线模型的损失函数替换为 L_{NS} ,其 Average Precision 和 Recall 分别提高 2.3 和 2.4 百分点;将 Bi-CBS、RFE 同时作用于主干网络和特征融合模块中, Average Precision 和 Recall 在 YOLOv5s 的基础上分别提高 4.7 和 3.8 百分点,在其上再嵌入 CARAFE 上采样算子使模型的 Average Precision 和 Recall 较只融合 Bi-CBS、RFE 两

个模块的模型进一步提升了 0.9 和 1.8 百分点。最后把设计的 4 个模块都应用到基线模型上后,模型的检测性能达到最佳, Average Precision 和 Recall 较基线模型提升 6.1 和 6.3 百分点, GFLOPs 仅增加 2.5。实验结果表明,本文所提出的算法能更加精确地实现莴笋芯部检测任务。

将 YOLOv5s、YOLOv7、YOLOv8s 和改进模型在强光和弱光条件下的测试效果进行对比,如图 10、图 11 所示。图中圆形框为漏检目标。在图 10(a)、图 10(b)、图 10(c)中,各模型对后排莴笋芯部均存在漏检现象,且检测出的目标置信度也很低,个别位置还出现重复检测的情况。后排莴笋芯部较小,原模型对其检测效果不佳。从图 10(d)中可以看出,改进后的模型将后排莴笋芯部全部识别出来,而且检测的置信度也整体高于原模型。在图 11(a)、图 10(b)、图 10(c)中,各模型对图像边缘的目标检测效果不够理想,而改进算法检测出更多目标。综合检测效果对比以及上述分析可知,与现有算法相比,改进算法对小目标的检测效果得到明显提升,有效改善了现有算法在不同场景下存在的漏检问题。

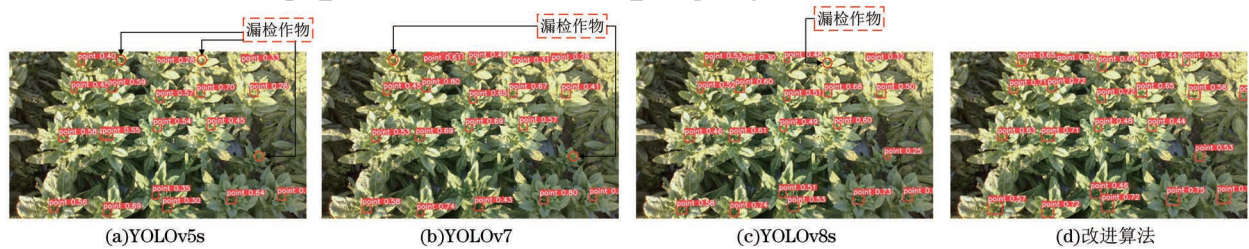


图 10 强光条件下检测效果对比

Fig. 10 Comparison of detection effects under strong light conditions

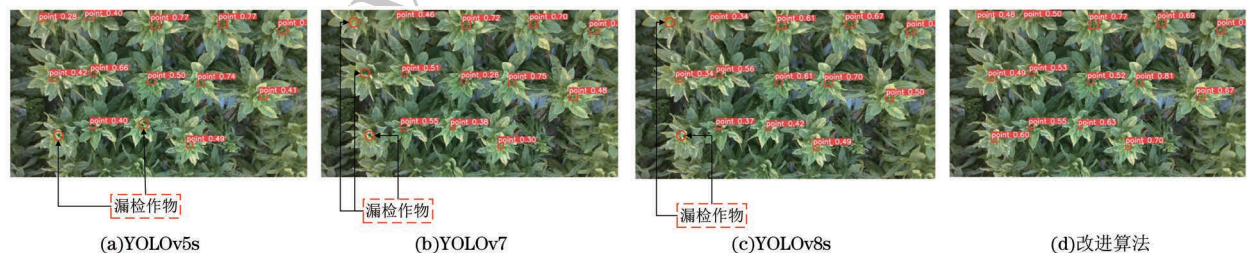


图 11 弱光条件下检测效果对比

Fig. 11 Comparison of detection effects under low light conditions

2.3.5 K 折交叉验证实验

K 折交叉验证是一种常见的模型评估方法,旨在评估模型在训练数据上的表现。该方法通过将原始数据集分割成 K 个子集来实施。在 K 次训练和验证中,每次会将其中一个子集用作验证集,而剩下的 K-1 个子集用作模型训练,一共进行 K 次。这种实验设计能够充分利用数据降低过拟合的风险,并提供对模型泛化能力的更可靠评估。通过这种方式,模型在不同的数据划分上得到测试,确保了评估

结果的稳健性。本文实验取 $K=5$,将每次训练模型的平均精确度作为评估指标,并取 5 次实验结果的平均值作为模型最终的平均精确度。本实验比较了不同模型在 K 折交叉验证下的性能表现,实验结果如表 5 所示。由表 5 可知,本文所提模型的平均精确度为 0.594,较 YOLOv5s 提高了 7.1 百分点,说明本文所提模块显著提升了模型检测的准确性。同时,所提模型的平均精确度也高于 YOLOv8s 和 FasterR-CNN 这些目前主流的一阶段和二阶段目标检测器,表现出

所提模型在检测精确度上的卓越性能。

表 5 K 折交叉验证实验

Table 5 K-fold cross-validation experiment

K	所提模型	YOLOv5s	YOLOv8s	FasterR-CNN
1	0.588	0.531	0.546	0.498
2	0.584	0.523	0.542	0.502
3	0.581	0.515	0.538	0.491
4	0.602	0.519	0.529	0.487
5	0.613	0.526	0.553	0.513
Average	0.594	0.523	0.542	0.498

2.3.6 作物行中心线提取实验

在得到每株莴笋的中心坐标后,需要将识别的坐标点拟合成直线。沿图像纵向设置条状感兴趣区域(ROI),提取同一 ROI 内的莴笋中心坐标为一组,保证同一列作物的中心坐标点归为一组,同时可以过滤离群特征点对提取结果的干扰。最后采用最小二乘法将同一组的作物中心点拟合成直线,即为该作物行的中心线。

表 6 展示的是检测结果对作物行中心线提取的定量影响。实验采用平均像素偏差以及平均前向角偏差作为评价指标。像素偏差是提取作物行中心线

与真实作物行中心线行中点坐标之间的像素偏移量;前向角偏差是提取作物行中心线与真实作物行中心线的偏差角度。在实验中,在所提取的作物行中心线上以平均间隔选取 10 个基准点计算像素偏差和前向角偏差,并求取平均值作为像素偏移量和前向角偏差的比较标准。结果表明,基线模型在不同光照条件下提取作物行中心线的平均像素偏差分别为 4.23、6.52 像素,其平均前向角偏差分别为 0.82° 和 1.81° ;而所提模型在不同光照条件下提取作物行中心线的平均像素偏差分别为 0.78、0.56 像素,平均前向角偏差分别为 0.37° 和 0.24° ,较基线模型均有大幅下降。

表 6 不同模型检测结果对提取作物行中心线的影响

Table 6 The impact of different model detection results on extracting crop row center lines

光照条件	模型	平均像素偏差/像素	平均前向角偏差/ $^{\circ}$
强光	基线	4.23	0.82
强光	ours	0.78	0.37
弱光	基线	6.52	1.81
弱光	ours	0.56	0.24

基于原始模型和改进模型作物行中心线提取可视化结果如图 12 所示。其中,红线代表实际作物行

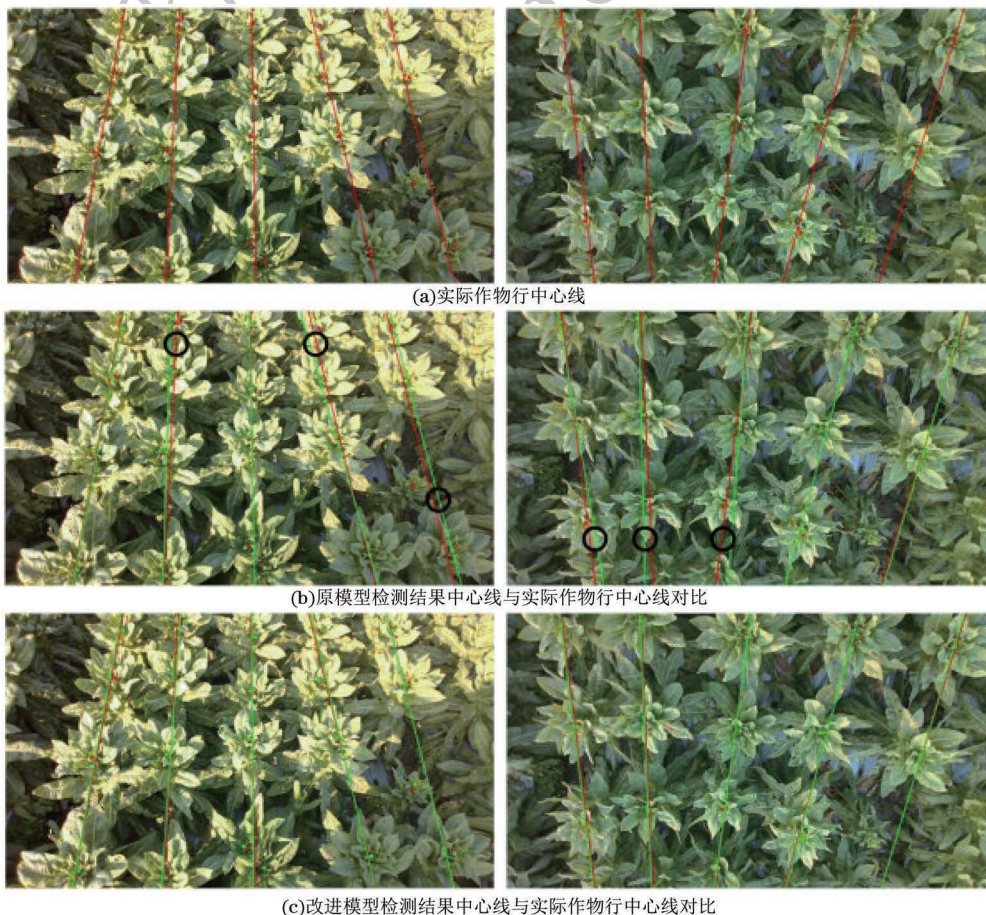


图 12 作物行中心线与实际作物行中心线效果对比

Fig.12 Comparison of the effect between the crop row center line and the actual crop row center line

中心线,绿线代表由模型检测结果提取的作物行中心线。实际作物行中心线是通过实际莴笋芯部坐标提取得到的。图 12(b)所示由原模型检测结果提取的作物行中心线每一列都与实际中心线有所偏差,特别是图中圈出的部分,偏差比较严重。图 12(c)所示由改进模型检测结果提取的作物行中心与实际情况偏差很小,可见改进模型检测结果能提取出更加准确的作物行中心线。

3 结束语

本文基于 YOLOv5s 提出一种改进的莴笋芯部检测算法。首先基于 BRA 注意力机制设计 Bi-CBS 卷积模块,并替换原始网络主干中的卷积模块,以保留细粒度的细节信息,增强模型对小目标检测的能力;然后设计基于空洞卷积和权值共享的多尺度融合模块,在增大特征图感受野的情况下不丢失分辨率有利于小目标的检测;另外引入拥有更大感受野的 CARAFE 上采样操作,用于在更大接受域中聚合上下文信息;最后提出一种损失函数解决传统 IoU 方法对小目标的位置偏差非常敏感的问题,以加快网络拟合速度。实验结果表明,改进算法相较于基线模型平均精确度和召回率分别提升了 6.1、6.3 百分点,解决了基线模型 YOLOv5s 在莴笋芯部检测中精度低、边界框回归不稳定等问题。本文模型的各项评估指标均优于其他对比目标检测模型,拥有优良的莴笋芯部检测效果,能提取出更加精确的作物行中心线,为智能农机提供更好的导航信息,助力农机的精确导航。在未来的工作中,将对改进网络模型的模型规模及推理速度进一步优化,期望在保证模型准确性和鲁棒性的同时,大幅提升模型在计算资源受限的农业设备上的运行效率和实时性,从而推动智能农业技术的广泛应用。

参考文献

- [1] 万欢, 欧媛珍, 管宪鲁, 等. 无人农机作业环境感知技术综述[J]. 农业工程学报, 2024, 40(8): 1-18.
WAN H, OU Y Z, GUAN X L, et al. Review on operational environment awareness technology of unmanned agricultural machinery [J]. Transactions of the Chinese Society of Agricultural Engineering, 2024, 40(8): 1-18. (in Chinese)
- [2] ZHANG S L, MA Q L, CHENG S K, et al. Crop row detection in the middle and late periods of maize under sheltering based on solid state LiDAR [J]. Agriculture, 2022, 12(12): 2011.
- [3] 贺静, 何杰, 罗锡文, 等. 基于多传感器融合的水稻行识别与跟踪导航研究[J]. 农业机械学报, 2022, 53(3): 18-26, 137.
HE J, HE J, LUO X W, et al. Research on rice row Recognition and tracking navigation based on multi-sensor fusion [J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(3): 18-26, 137. (in Chinese)
- [4] HUANG S P, WU S H, SUN C, et al. Deep localization model for intra-row crop detection in paddy field [J]. Computers and Electronics in Agriculture, 2020, 169: 105203.
- [5] 王姗姗, 余山山, 张文毅, 等. 基于特征点邻域 Hough 变换的水稻秧苗行检测[J]. 农业机械学报, 2020, 51(10): 18-25.
WANG S S, YU S S, ZHANG W Y, et al. Rice seedling row detection based on feature point neighborhood Hough transform [J]. Journal of Agricultural Machinery, 2020, 51(10): 18-25. (in Chinese)
- [6] HE R R, LUO X W, ZHANG Z G, et al. Identification method of rice seedlings rows based on Gaussian heatmap[J]. Agriculture, 2022, 12(10): 1736.
- [7] BAH M D, HAFIANE A, CANALS R. CRownNet: deep network for crop row detection in UAV images [J]. IEEE Access, 2020, 8: 5189-5200.
- [8] 徐广达, 毛国军. 基于多层次特征融合的无人机航拍图像目标检测[J]. 计算机科学与探索, 2023, 17(3): 635-64.
XU G D, MAO G J. Target detection in UAV aerial images with multi-level feature fusion [J]. Journal of Frontiers of Computer Science & Technology, 2023, 17(3): 635-645. (in Chinese)
- [9] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 8759-8768.
- [10] 庄集超, 郭保苏, 吴凤和. 基于可变形密集卷积神经网络的布匹瑕疵检测[J]. 计量学报, 2023, 44(2): 178-185.
ZHUANG J C, GUO B S, WU F H. Fabric defect detection based on deformable dense convolutional neural network [J]. Acta Metrologica Sinica, 2023, 44(2): 178-185. (in Chinese)
- [11] 赵楚, 段先华, 苏俊楷. 改进 Faster RCNN 的瓷砖表面瑕疵检测研究[J]. 计算机工程与应用, 2023, 59(14): 201-208.
ZHAO C, DUAN X H, SU J K. Research on improved Ceramic tile surface defect detection by Faster RCNN [J]. Computer Engineering and Applications, 2023, 59(14): 201-208. (in Chinese)
- [12] 赵继达, 甄国涌, 储成群. 基于 YOLOv8 的无人机图像目标检测算法[J]. 计算机工程, 2024, 50(4): 113-120.
ZHAO J D, ZHEN G Y, CHU Q. The UAV image target detection algorithm based on YOLOv8 [J]. Computer Engineering, 2024, 50(4): 113-120. (in Chinese)
- [13] ZHENG Z Z, HU Y H, QIAO Y C, et al. Real-time detection of winter jujubes based on improved YOLOX-nano network [J]. Remote Sensing, 2022, 14(19): 4833.
- [14] 袁天乐, 袁巨龙, 朱勇建, 等. 基于改进 YOLOv5 的推力球轴承表面缺陷检测算法[J]. 浙江大学学报(工学版), 2022, 56(12): 2349-2357.
YUAN T L, YUAN J L, ZHU Y J, et al. Thrust ball bearing surface defect detection algorithm based on improved YOLOv5 [J]. Journal of Zhejiang University (Engineering Edition), 2022, 56(12): 2349-2357. (in Chinese)
- [15] WANG J Q, CHEN K, XU R, et al. CARAFE: content-aware reassembly of features [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2019: 3007-3016.
- [16] ZHOU Y T, CAO K Y, LI D, et al. Fine-YOLO: a simplified X-ray prohibited object detection network based on feature aggregation and normalized Wasserstein distance [J]. Sensors, 2024, 24(11): 3588.
- [17] ZHANG Q H, BAO X, SUN S T, et al. Lightweight network for small target fall detection based on feature fusion

- and dynamic convolution [J]. *Journal of Real-Time Image Processing*, 2024, 21(1): 17.
- [18] SHEN J, QU Y R, ZHANG W N, et al. Wasserstein distance guided representation learning for domain adaptation [J]. *Artificial Intelligence*, 2018, 32(1): 328-336.
- [19] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2018: 7132-7141.
- [20] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2021: 13713-13722.
- [21] ZHU L, WANG X J, KE Z H, et al. BiFormer: vision transformer with bi-level routing attention [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2023: 10323-10333.
- [22] YU Z P, HUANG H B, CHEN W J, et al. YOLO-FaceV2: a scale and occlusion aware face detector [J]. *Pattern Recognition*, 2024, 155: 110714.
- [23] JOCHER G. Ultralytics/YOLOv5: v7.0-YOLOv5 SOTA realtime instance segmentation (v7.0) [EB/OL]. [2024-03-21]. <https://ui.adsabs.harvard.edu/abs/2022zndo...3908559J/abstract>.
- [24] WANG C Y, BOCHKOVSKIY A, LIAO H M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE Press, 2023: 7464-7475.
- [25] JOCHER G, CHAURASIA A, QIU J. Ultralytics YOLOv8 (8.0.0) [EB/OL]. [2024-03-21]. <https://github.com/ultralytics/ultralytics>.
- [26] AN R, ZHANG X C, SUN M P, et al. GC-YOLOv9: Innovative smart city traffic monitoring solution [J]. *Alexandria Engineering Journal*, 2024, 106: 277-287.
- [27] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 39(6): 1137-1149.
- [28] CAI Z W, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE Press, 2018: 6154-6162.
- [29] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector [M]. Berlin, Germany: Springer, 2016.
- [30] ZHAO Y A, LV W Y, XU S L, et al. DETRs beat YOLOs on real-time object detection [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: IEEE Press, 2024: 16965-16974.
- [31] ZHENG Q H, TIAN X Y, YU Z G, et al. DL-PR: generalized automatic modulation classification method based on deep learning with priori regularization [J]. *Engineering Applications of Artificial Intelligence*, 2023, 122: 106082.

文字编辑 索书志
栏目编辑 宋 圆