

林祖香,王英东,马 荣,等. 基于 YOLO-L 的自然环境中澳洲坚果果实的检测和识别[J]. 江苏农业学报,2024,40(11):2102-2110.
doi:10.3969/j.issn.1000-4440.2024.11.014

基于 YOLO-L 的自然环境中澳洲坚果果实的检测和识别

林祖香¹, 王英东¹, 马 荣¹, 韦云松¹, 李子文¹, 李加强¹, 何 超²

(1.西南林业大学机械与交通学院,云南 昆明 650224; 2.德宏职业学院,云南 德宏 678400)

摘要: 针对自然环境下果实重叠、相互遮挡和目标小的澳洲坚果果实检测准确率低的问题,提出一种改进 YOLOv9 模型的识别方法(YOLO-L)。首先,引入 BiFormer 注意力机制,该机制通过双层路由注意力机制实现了动态、查询感知的稀疏注意力分配,能够很好地捕获特征表征,增强网络对全局特征的关注度;其次,采用 VoVGSCSP 模块代替 YOLOv9 中的 CBFuse 模块,提高了复杂场景下小目标的检测效果;最后,将 YOLOv9 模型默认的损失函数替换成排斥损失函数,解决了果实排列密集和漏检的问题,进一步提升了澳洲坚果果实检测的平均精度。通过消融试验和对比试验来验证模型的有效性,发现 YOLO-L 模型的平均精度均值、精确率、召回率和 F_1 值分别达到 96.2%、92.3%、88.2%和 90.2%。与 YOLOv9 模型相比,YOLO-L 模型的平均精度均值提升了 4.9 个百分点。总体而言,YOLO-L 模型能够在自然环境下准确识别被遮挡、重叠的澳洲坚果果实,且检测精度高。研究结果可为澳洲坚果产业的智能采摘提供有效的技术支持。

关键词: 图像处理;深度学习;YOLOv9 模型;澳洲坚果

中图分类号: TP301.6;S664

文献标识码: A

文章编号: 1000-4440(2024)11-2102-09

Macadamia (*Macadamia integrifolia* Maiden & Betcher) detection and recognition in natural environments based on YOLO-L

LIN Zuxiang¹, WANG Yingdong¹, MA Rong¹, WEI Yunsong¹, LI Ziwen¹, LI Jiaqiang¹, HE Chao²

(1.College of Mechanical and Transportation Engineering, Southwest Forestry University, Kunming 650224, China; 2.Dehong Vocational College, Dehong 678400, China)

Abstract: Aiming at the issue of low detection accuracy for macadamia nuts in natural environments due to overlapping, mutual occlusion, and small targets, an improved YOLOv9 model recognition method (YOLO-L) was proposed. Firstly, the BiFormer attention mechanism was introduced, which achieved dynamic and query-aware sparse attention allocation through the Bi-level routing attention mechanism. This mechanism was capable of effectively capturing feature representations and enhanced the network's focus on global features. Secondly, the VoVGSCSP module was used to replace the CBFuse module in YOLOv9, which improved the detection performance for small targets in complex scenes. Lastly, the default loss function of the YOLOv9 model was replaced with an exclusion loss function, which solved the problems of dense fruit arrangement and missed detections, and

further enhanced the average accuracy of macadamia nut detection. The effectiveness of the model was validated through ablation and comparative experiments. It was found that the mean average precision, precision, recall, and F_1 score of YOLO-L model reached 96.2%, 92.3%, 88.2%, and 90.2%, respectively. Compared with the YOLOv9 model, the mean average precision of the YOLO-L model was improved by 4.9 percentage points. Overall, the YOLO-L model

收稿日期:2024-08-23

基金项目:云南省高层次人才支持项目(YNWR-QNBJ-2018-066);云南省科技厅项目(202301BD070001-077)

作者简介:林祖香(1997-),女,云南保山人,硕士研究生,主要从事图像识别、目标检测方面的研究。(E-mail)2817702741@qq.com

通讯作者:马 荣,(E-mail)mr@swfu.edu.cn

can accurately identify occluded and overlapped macadamia nuts in natural environments with high detection accuracy. The research results can provide effective technical support for the intelligent harvesting in the macadamia industry.

Key words: image processing; deep learning; YOLOv9 model; macadamia (*Macadamia integrifolia* Maiden & Betcher)

澳洲坚果(*Macadamia integrifolia*)被誉为“坚果之王”,其果仁含有 17 种氨基酸、多种矿物质和大量蛋白质,是经济价值较高的食用干果之一^[1]。然而,澳洲坚果采摘过程依赖人工,效率低、速度慢且劳动强度大,因此自动化和智能设备化采摘是解决该问题的重要途径,但是复杂的生长环境和密集重叠的果实会影响视觉检测的准确性,造成智能采摘进展缓慢。因此,快速准确识别澳洲坚果果实是智能采摘的关键。

视觉识别在农业中的应用广泛,特别是对果实的识别和检测^[2]。传统机器视觉需要手动提取特征,鲁棒性差,难以应对复杂环境^[3-4]。相比之下,深度学习的目标检测算法因其优越的泛化能力而倍受关注^[5-6],该算法分为两阶段算法和单阶段算法^[7-8]。两阶段算法的精度高,但速度慢,单阶段算法具有更快的检测速度和更高的可拓展性,因此更适合实际应用^[9]。目前,作为单阶段深度学习算法的 YOLO 已成为农产品检测的主流方法^[10-13]。

自 YOLOv1 发布以来,YOLO 系列算法经历了多次迭代与优化。YOLOv1 首次解决了两阶段检测算法速度慢的问题^[14];YOLOv2 引入了新的 30 层架构、锚盒及批量标准化技术,显著提升了模型性能和收敛速度^[15];YOLOv3 通过增加 SPP 层和采用 Mish 激活函数,进一步提高了复杂环境下的识别精度^[16];YOLOv4 利用 CSP Darknet53、DenseNet 和 Leaky ReLU 等技术,增强了模型的鲁棒性和对小目标的检测能力^[17];YOLOv5 通过引入注意力机制(SE)和替换主干网络为 BottleneckCSP-2,提高了检测识别的精确率^[18];YOLOv6 实现了无锚设计,并引入自我关注机制、自适应训练技术,优化了训练过程^[19];YOLOv7 在 YOLOv6 基础上引入了 E-ELAN 模块,提高了推理效率;YOLOv8 则通过 C2f 模块将高级功能与上下文信息结合,提升了检测准确率^[20]。近年来,许多研究者采用 YOLO 对农产品果实进行检测识别,并尝试提升其检测性能。例如,利用添加了 BottleneckCSP-2、SE 模块的 YOLOv5 检测被遮挡苹果,召回率、精确率、平均精度均值和 F_1 值分别为 91.48%、83.83%、86.75%和 87.49%^[21];用结合 Transformer 注意力机制的 YOLOv7 检测荔枝果实,平

均精度均值达 93.6%^[22];通过深度可分离卷积、双路径注意力门模块来提升 YOLOv8 对大棚中番茄果实的检测精度,结果表明,平均精度均值达到 93.4%,且具有更低的损失^[23]。上述方法在农业领域的目标检测中取得了一定的成功,为实现智能采摘提供了有力支撑。

随着人工智能技术的进步,YOLOv9 模型在深度神经处理中减少信息丢失方面取得了重大进展,通过引入可编程梯度信息(PGI)和通用高效层聚合网络(GELAN)架构,确保了稳健可靠的梯度流动,有效防止了数据退化,实现了精确检测^[24]。尽管 YOLOv9 模型在番茄果实成熟度检测方面展现出了强大潜力^[25],但是在澳洲坚果果实的检测识别任务中,现有研究成果仍较为有限。因此,本研究选用 YOLOv9 模型用于检测识别澳洲坚果果实,并针对澳洲坚果果实检测识别中普遍存在的模型复杂度高、遮挡和密集小目标识别精度低等问题,提出 YOLO-L 模型。该模型通过引入 BiFormer 注意力模块,增强了模型对复杂场景中关键信息的捕捉能力;同时,采用 VoVGSCSP 模块替代原有结构,结合排斥损失函数(而非传统的 CIoU 损失函数),以进一步提高对密集重叠及被遮挡的小目标的检测精度。研究结果旨在为澳洲坚果产业的智能采摘提供更有效的技术支持。

1 材料与方法

1.1 澳洲坚果图像数据集的构建

1.1.1 数据采集 研究对象为 OC、H2、344 等不同品种的澳洲坚果,树龄均在 10~13 年。图像数据采集于 2023 年 8 月,分别来源于云南省德宏景颇族自治州的西山乡杏焕村、三台山德昂族乡和风平镇南相章村(23°50'N~25°20'N,97°31'E~98°43'E)。在不同天气和自然光照条件下,将相机与水平面的夹角设置为 -45°~45°,将采集到的图像统一保存为 JPG 格式,图像分辨率为 6 000×4 000 像素,图像数据统计的结果为无遮挡坚果图像 600 张,相互遮挡坚果图像 740 张,树叶遮挡坚果图像 660 张,共采集图像 2 000 张。采集到的部分澳洲坚果的图像见图 1。

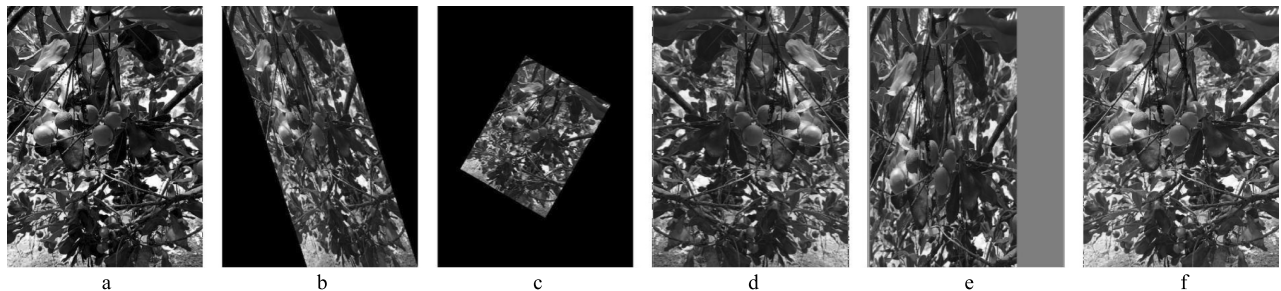


图1 澳洲坚果数据集的部分图像

Fig.1 Selected images of the macadamia nut dataset

1.1.2 数据集的构建 使用 labelImg 进行图像标注。数据分为健康果实、被蓟马和蜡象侵害的果实 3 个类别, 3 个类别分别表示为 health、thrips 和 shieldbug, 并将数据集按照 8 : 1 : 1 的比例划分为训

练集、验证集和测试集。如图 2 所示, 通过增强对比度、错切、仿射变换、翻转和平移等方法对原始图像进行了数据增强, 将数据集扩展到 12 000 张图像。



a: 增强对比度; b: 错切; c: 仿射变换; d: 翻转; e: 平移; f: HSV 增强(H 表示色相; S 表示饱和度; V 表示色明度)。

图2 数据增强图像

Fig.2 Data enhanced images

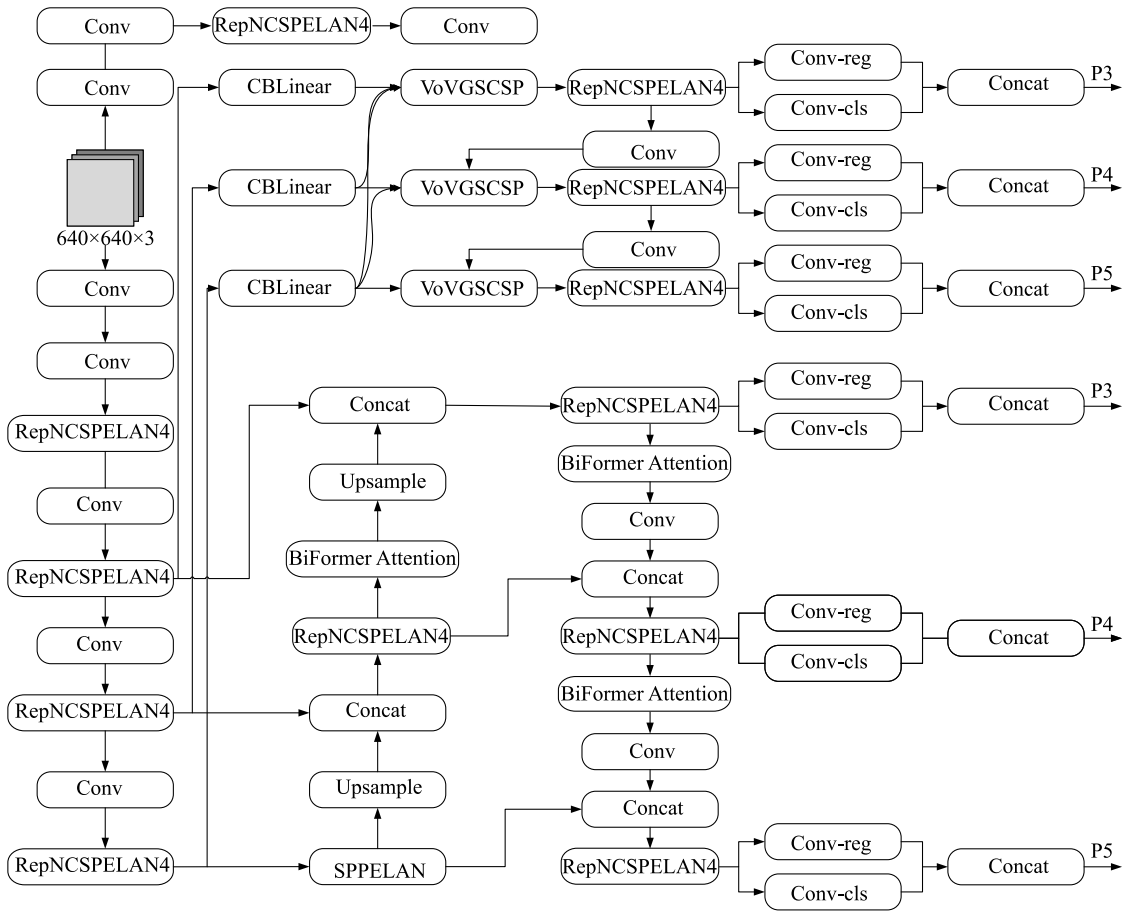
1.2 YOLOv9 算法的原理

YOLOv9^[24]在实时目标检测领域的表现出色, 通过引入可编程梯度信息(PGI)、广义高效聚合层(GELAN)两大核心技术, 显著提升了检测精度与效率。PGI 通过主分支、辅助可逆分支及多级辅助信息的集成, 有效解决了深度网络中信息丢失的问题, 确保了梯度流的可靠性; GELAN 结合 CSPNet 和 ELAN 的优势, 构建轻量、高效的网络架构, 实现了参数利用率与性能的提升。然而, 在检测密集、遮挡或重叠的小目标时, 特征融合的特征表征较弱, 容易导致误检、漏检, 从而影响模型的检测精度。为了进一步提升复杂林间背景下被遮挡的澳洲坚果果实的检测性能, 本研究重构了 YOLOv9 网络结构。在 Neck 网络的 RepNCSPELAN 模块之后, 特征融合和上采样操作之前引入 BiFormer 注意力机制^[26]。同时, 将网络结构中的 CBFuse 模块替换为 VoVGSCSP 模块, 并在 Head 网络的检测输出层引入排斥损失

函数^[27], 以提高模型在自然场景中对密集小目标和相互遮挡果实的检测鲁棒性。YOLO-L 模型的结构如图 3 所示。

1.3 YOLO 算法的改进

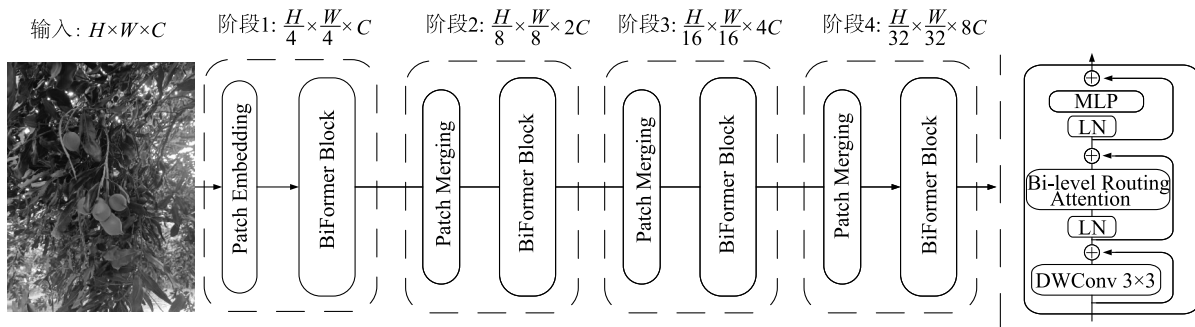
1.3.1 BiFormer BiFormer^[26]注意力机制通过结合区域级筛选和细粒度令牌级注意力计算, 有效减少了计算资源消耗, 同时保持了高效的特征捕捉能力。这种机制在粗粒度区域级别过滤掉不相关的键值对, 减少了细粒度级别上的详尽计算, 从而实现了计算资源的优化利用。BiFormer 的结构如图 4 所示, 其核心在于引入多个 BiFormer Block, 通过 Patch Nearing 步骤促进相邻模块之间的特征聚合, 增强特征表示的一致性、鲁棒性。在特征提取过程中, 首先使用 3×3 深度卷积进行隐式编码以捕捉相对位置信息。随后, 引入双层路由注意(BRA)模块用于捕捉不同位置之间的关系, 并通过多层感知器(MLP)模块实现全局位置的嵌入。



Conv:卷积模块;RepNCSPELAN4:泛化高效层聚合网络;CBLinear:1×1 卷积和线性变换操作;VoVGSCSP:VoVNet 和 CSPNet 的融合模块;Concat:特征连接模块;Upsample:上采样模块;SPPELAN:空间金字塔池化结构;BiFormer Attention:BiFormer 的注意力机制;Conv-reg:边界框回归(位置预测);Conv-cls:目标类别分类;P3、P4、P5 表示输出。

图 3 YOLO-L 算法的结构

Fig.3 Structure of the YOLO-L algorithm



H :高度; W :宽度; C :通道数;Conv:卷积模块;Patch Embedding:通过线性投影或卷积操作将这些小块映射到高维特征空间,形成初始的特征图;BiFormer Block:该结构的核心模块;LN:层归一化;Bi-level Routing Attention:双层路由注意力机制;DWConv:深度可分卷积;MLP:多层感知机;Patch Merging:将多个小块进行降采样,减少特征图的分辨率并增加通道数。

图 4 BiFormer 结构

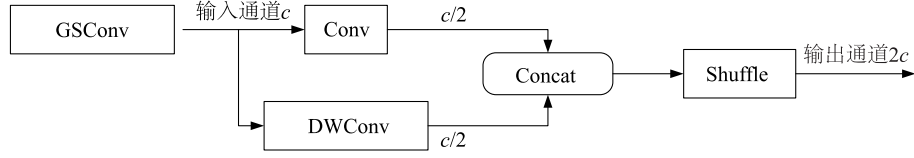
Fig.4 BiFormer structure

1.3.2 VoVGSCSP GSCConv 是 2022 年提出的一种新的轻量级卷积方法^[28]。如图 5 所示,输入特征图经过标

准卷积层后,生成 $c/2$ 个通道的特征图。同时经过深度卷积层,生成另外 $c/2$ 个通道的特征图。将 2 个 $c/2$

通道的特征图在通道维度上进行连接,形成 c 个通道的特征图,对连接后的特征图进行通道混洗操作,重新

排列通道顺序,输出 $2c$ 个通道的特征图。



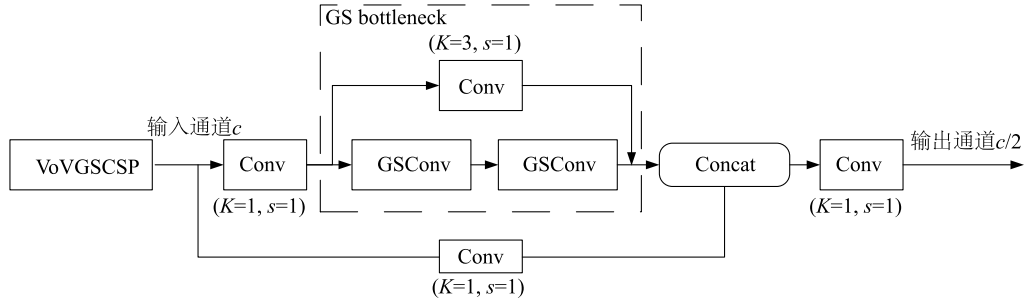
GSConv: 高效的卷积模块; Conv: 卷积模块; DWConv: 深度可分卷积; Concat: 通道拼接; Shuffle: 通道重排。

图 5 GSConv 的结构

Fig.5 Structure of GSConv

VoVGSCSP 是在 GSConv 的基础上引入 GS 瓶颈,并采用一次性聚合方法设计的跨阶段网络模块 (GSCSP),详见图 6。该模块在降低计算复杂度和

网络结构复杂性的同时,能够保持较高的精度和良好的特征重用率。



K : 卷积核大小; s : 步长; Conv: 卷积模块; GSConv: 高效的卷积模块; GS bottleneck: 瓶颈部分, 包含深度特征提取 (Conv) 和 Ghost 特征重建 (GSConv); Concat: 特征连接模块; VoVGSCSP: VoVNet 和 CSPNet 的融合模块。

图 6 VoVGSCSP 结构

Fig.6 VoVGSCSP structure

1.3.3 损失函数的优化 澳洲坚果果实排列密集, 所处环境复杂且遮挡问题严重, 传统损失函数在处理质量较差的实例或具有密集对象的数据集时表现出局限性。因此, 采用 Wang 等^[27] 基于预测框与真实框排斥吸引的 Repulsion Loss 来解决这些限制[排斥损失函数可定义为公式(1)]。该排斥损失函数对自然场景中密集小目标的检测具有更好的鲁棒性, 克服了传统损失函数带来的挑战。

$$L = L_{Attr} + \alpha L_{RepGT} + \beta L_{RepBox} \quad (1)$$

式中, α 和 β 表示平衡辅助损失项的权重; L_{Attr} 表示吸引项; L_{RepGT} 和 L_{RepBox} 表示排斥项。

具体而言, L_{Attr} 的定义见公式(2), 用于指导模型预测的边界框与真实目标框之间的对齐, 以提高检测性能和准确性。 L_{RepGT} 、 L_{RepBox} 是排斥项, 其定义分别见公式(3)、公式(4)。 $Smooth_{ln}$ 是一个平滑函数[定义见公式(5)], 用于计算平滑的自然对数, 这种平滑处理的目的是保持损失函数在 x 接近 σ 时

的连续性, 避免出现梯度消失或梯度爆炸的情况, 从而更稳定地训练模型。

$$L_{Attr} = \frac{\sum_{P \in p^+} Smooth_{Ll}(B^P, G_{Attr}^P)}{|p^+|} \quad (2)$$

$$L_{RepGT} = \frac{\sum_{P \in p^+} Smooth_{ln}[IoU(B^P, G_{Rep}^P)]}{|p^+|} \quad (3)$$

$$L_{RepBox} = \frac{\sum_{i \neq j} Smooth_{ln}[IoU(B^{P_i}, B^{P_j})]}{\sum_{i \neq j} 1 [IoU(B^{P_i}, B^{P_j}) > 0] + \varepsilon} \quad (4)$$

$$Smooth_{ln} = \begin{cases} -\ln(1-\sigma), & x \leq \sigma \\ \frac{x-\sigma}{1-\sigma} - \ln(1-\sigma), & x > \sigma \end{cases} \quad (5)$$

式中, $\sum_{P \in p^+}$ 表示对正样本候选框 P 进行求和; $Smooth_{Ll}(B^P, G_{Attr}^P)$ 表示使用 $Smooth_{Ll}$ 平滑函数计算候选框 P 与其指定目标 G_{Attr}^P 之间的损失; $|p^+|$ 表示正样本候选框的数量; $Smooth_{ln}[IoU(B^P, G_{Rep}^P)]$ 表示使用 $Smooth_{ln}$ 平滑函数计算候选框 P 与其目标指定

的排斥目标 G_{Rep}^P 之间的交并比 (IoU); $\Sigma_{i \neq j}$ 表示对所有的边界框 P_i, P_j 进行求和; $IoU(B^i, B^j)$ 表示预测框 P_i 和 P_j 的交并比; ε, σ 表示常数。

1.4 模型训练

1.4.1 训练平台及策略 操作系统为 Windows10, CPU 为 Intel Core i7-9700@ 3.00 GHz, 运行内存为 64 GB; 显卡 (GPU) 为 NVIDIA GeForce RTX 2080 SUPER, 显存为 8 GB; 深度学习框架为 PyTorch 1.12.1; Python 版本号为 3.8; CUDA 的版本号为 CUDA 10.2; cuDNN 的版本号为 cuDNN 8.6.0。

采用权重文件 yolov9.pt 进行训练。以自创澳洲坚果图像数据集作为训练数据集, 输入图像大小为 640×640 像素, 经过多次调整和测试, 设置训练批次为 2, 迭代次数为 120, 初始学习率为 0.1, 动量因子为 0.9。通过消融试验和对比试验充分验证 YOLO-L 模型的性能。

1.4.2 评价指标 评估指标包括精确率 (P)、召回率 (R)、 F_1 值^[29]、平均精度均值 (mAP)^[30] 和每秒帧数 (FPS)。

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N} \quad (8)$$

$$F_1 \text{ 值} = \frac{P \times R \times 2}{P + R} \quad (9)$$

$$FPS = \frac{TotalTime}{FigureNumber} \quad (10)$$

式中, TP 、 FN 、 FP 分别是真阳性、假阴性、假阳性; mAP 为所有类别的平均精度均值; N 是类别的数量; AP_i 是第 i 个类别的平均精度; F_1 值是模型检测准确率的综合评估指标, 是精确率、召回率的调和平均值, F_1 值越高, 表示模型越稳定; FPS 表示每张图片检测所用的时间; $TotalTime$ 指检测全部验证图像所用时间; $FigureNumber$ 表示验证图像总数。

2 结果与分析

2.1 使用不同注意力机制的模型的检测性能的对比

以 YOLOv9 为基础模型, 在 Neck 网络的相同部位引入不同的注意力机制, 有助于模型更有效地提取关键特征, 从而提高目标检测框的准确性、精确度。

由表 1 可以看出, YOLOv9 模型使用 BiFormer 注意力机制的效果要优于使用 SimAM、CBAM 的效果。具体而言, YOLOv9-BiFormer 模型的精确率、召回率、平均精度均值和 F_1 值都较另外 2 个模型得到了大幅度提升, 主要原因是 BiFormer 提高了算法捕获长距离上下文依赖关系的能力, 并且有效提升了澳洲坚果果实识别检测的准确率。同时, 稳定上升的数据进一步验证了该试验方法的可行性、有效性。

表 1 使用不同注意力机制的模型训练的结果

Table 1 The training results of models using different attention mechanisms

模型	平均精度均值 (%)	精确率 (%)	召回率 (%)	F_1 值 (%)
YOLOv9-SimAM	90.1	87.2	84.3	85.7
YOLOv9-CBAM	86.0	82.0	80.2	81.1
YOLOv9-BiFormer	96.4	95.3	93.0	94.1

F_1 值是模型检测准确率的综合评估指标, 是精确率、召回率的调和平均值, F_1 值越高, 表示模型越稳定。

YOLOv9-BiFormer 模型和原始 YOLOv9 模型迭代训练 120 次得到的精确率、召回率曲线如图 7 所示。可以看出, YOLOv9-BiFormer 模型的精确率在训练初期即展现出优势, 并且随着迭代次数的增加持续上升, 最终趋于平稳, 表明该模型在目标检测时具有相对更高的预测准确性、稳定性。图 7b 则揭示了召回率的变化趋势, 可以看出, YOLOv9-BiFormer 模型的召回率曲线相较于原始模型更为平稳, 且在训练后期振荡减少, 显示其全面捕捉目标对象的能力更强。综上所述, YOLOv9 模型通过引入 BiFormer 注意力机制, 在精确率、召回率上均实现了显著提升, 为目标检测带来了更加优异的表现。

2.2 YOLOv9 模型消融试验结果

为了验证本研究提出的 YOLO-L 模型的性能提升效果, 将 YOLO-L 模型和 YOLOv9 模型进行对比, 以验证每步改进的有效性。由表 2 可以看出, 与原始的 YOLOv9 模型相比, 版本 1 通过引入 BiFormer 注意力机制, 提升平均精度均值至 96.4%, 提升幅度达 5.1 个百分点, 表明模型捕捉全局信息的能力得到了增强。版本 2 引入的 VoVGSCSP 模块, 成功将参数量压缩至 43.1 M。版本 3 引入的排斥损失函数, 对自然场景中的密集小目标检测具有更好的鲁棒性, 克服了传统损失函数带来的挑战。版本 4 结合上述的 3 个模块, 相比于 YOLOv9 实现了平均精度均值、精确率、召回率和 F_1 值的全面提升, 分别达到 96.2%、92.3%、88.2% 和

90.2%。试验结果表明,在复杂背景和有遮挡的情况下,YOLO-L 模型依然能够保持较高的检测精度和出色的泛化性能,同时有效降低了错检率、漏检率。

2.3 不同模型检测结果及分析

对 Faster-RCNN、SSD、YOLOv5、YOLOv8、YOLOv9 和 YOLO-L 进行了多组对比试验。试验时使用相同设备、数据集和数据增强方法,保持训练集、测试集的比例相等,从而保证对比结果的可靠性。由表 3 可以看出,在相同试验条件下,本研究提出的 YOLO-L 模型具

有 96.2%的平均精度均值。具体而言,YOLO-L 模型的平均精度均值分别较 Faster-RCNN、SSD、YOLOv5、YOLOv8、YOLOv9 模型提高了 7.5 个百分点、14.0 个百分点、9.1 个百分点、6.0 个百分点、4.9 个百分点。YOLO-L 模型的检测速度分别比 Faster-RCNN、SSD、YOLOv5、YOLOv8、YOLOv9 模型提高了 190.58%、41.65%、61.20%、155.81% 和 23.04%。结果表明,YOLO-L 模型在识别精度、检测速度方面的表现比其他模型出色。

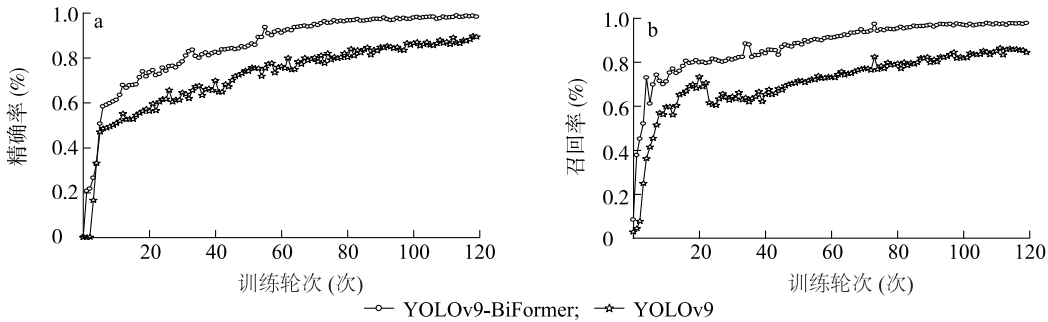


图 7 YOLOv9-BiFormer 模型和 YOLOv9 模型的精确率、召回率曲线
Fig.7 Precision and recall curves for YOLOv9-BiFormer and YOLOv9 models

表 2 改进模型消融试验的准确性和性能

模型	BiFormer	VoVGSCSP	Repulsion Loss	平均精度均值 (%)	精确率 (%)	召回率 (%)	F_1 值 (%)	参数量 (M)
YOLOv9	×	×	×	91.3	87.8	82.6	85.1	288.5
版本 1	√	×	×	96.4	95.3	93.0	94.1	345.3
版本 2	×	√	×	87.2	82.7	80.5	81.6	43.1
版本 3	×	×	√	95.6	90.2	89.3	89.7	288.5
版本 4	√	√	√	96.2	92.3	88.2	90.2	73.8

F_1 值见表 1 注。

表 3 不同模型的评价指标

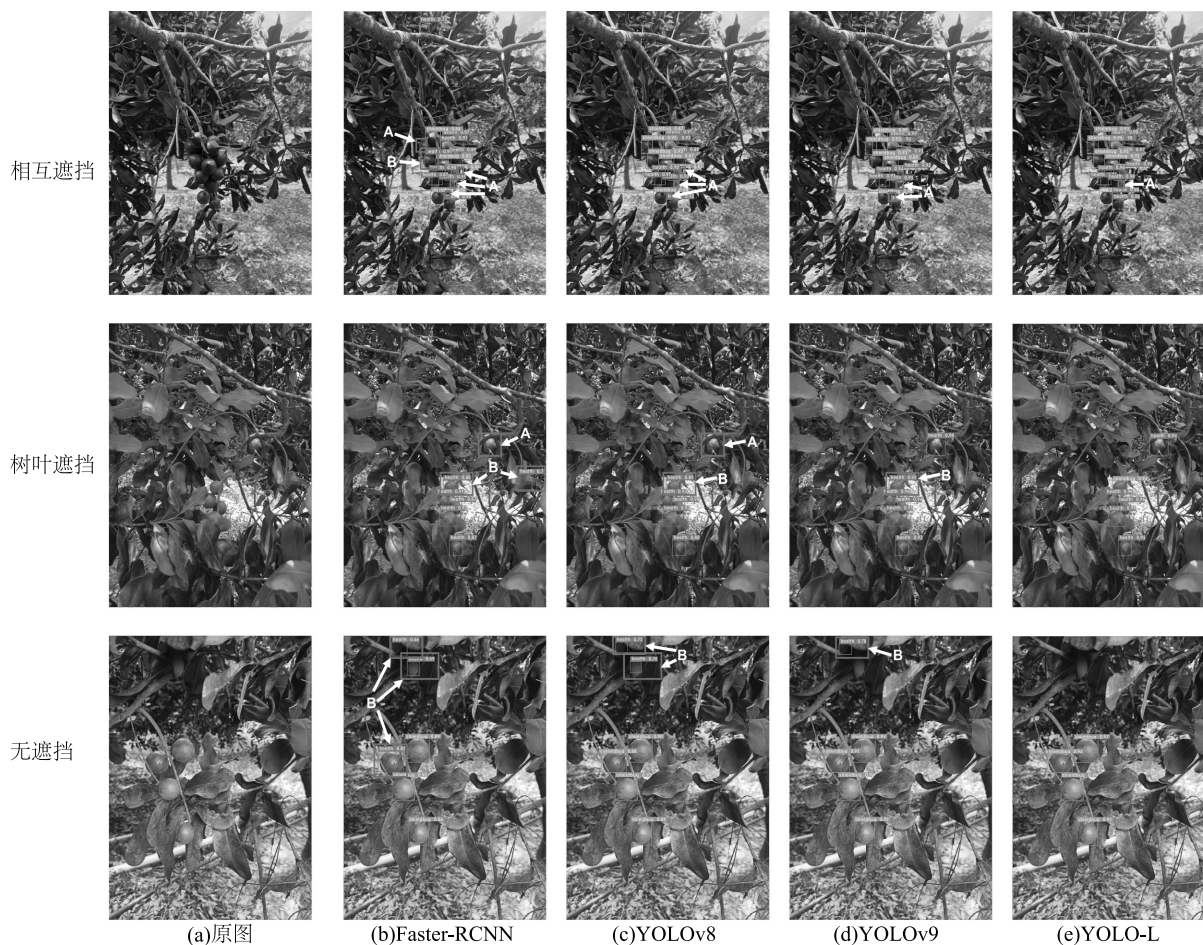
模型	$mAP_{0.50}$ (%)	$mAP_{0.50-0.95}$ (%)	精确率 (%)	召回率 (%)	F_1 值 (%)	检测速度 (帧, 1 s)
YOLOv5	87.1	64.0	87.2	82.2	84.6	80.36
YOLOv8	90.2	82.3	90.5	89.1	89.8	50.64
YOLOv9	91.3	86.7	87.8	82.6	85.1	105.28
YOLO-L	96.2	92.0	92.3	88.2	90.2	129.54
Faster-RCNN	88.7	81.6	87.5	70.6	78.1	44.58
SSD	82.2	59.5	79.7	64.2	71.1	91.45

$mAP_{0.50}$ 表示交并比(IoU)=0.50 时的平均精度均值, $mAP_{0.51-0.95}$ 表示 IoU 范围为 0.51~0.95 时的平均精度均值。 F_1 值见表 1 注。

图 8 展示了不同模型的可视化结果,图 8a 为随机选择的图片。选择检测结果较好的模型(Faster-RCNN、YOLOv8、YOLOv9、YOLO-L)的可视化结果进行对比分析。由图 8b 可以看出,Faster-RCNN 模型

在不同场景中的表现不佳,检测结果存在不稳定、易发生错检或漏检的情况。由图 8c 可以看出,在检测过程中,由于 YOLOv8 模型的背景颜色与果实颜色相近,因此偶尔会出现漏检的情况,但是其整体目标检测置信度明显高于 Faster-RCNN 模型。由图 8d 可以看出,YOLOv9 模型的检测效果较为理想,但仍存在目标置信度不够高、小目标的检测效果较差、偶

尔将绿叶检测为健康果的问题。由图 8e 可以看出,YOLO-L 模型的检测效果表现突出。从上述检测结果可以看出,在自然背景下,目标检测的置信度相对更高,预测模型的匹配能力也得到了提升,并且几乎没有出现漏检实例。总体而言,YOLO-L 模型在自然光照等背景条件下能够精确识别密集、被遮挡以及



A 表示漏检;B 表示错检;其余检测框内均为正常检测结果。

图 8 不同模型检测的可视化结果

Fig.8 Visualization results of different model detection

3 结论

在对澳洲坚果的实际检测过程中,YOLO-L 模型表现出了优异的性能。具体而言,BiFormer 注意力机制增强了网络对全局特征的注意力,提高了模型性能;VoVGSCSP 模块提高了复杂场景下小目标的检测效果;排斥损失函数对提高被遮挡目标的检测精度、加速网络收敛起着关键作用。

通过消融试验、对比试验,得出以下结论:(1)对

BiFormer 与 SimAM、CBAM 注意力机制进行对比发现,BiFormer 表现出更优越的性能。YOLOv9-BiFormer 的平均精度均值、精确率、召回率和 F_1 值分别达到 96.4%、95.3%、93.0%和 94.1%。(2)消融试验结果表明,YOLO-L 模型平均精度均值、精确率、召回率和 F_1 值分别为 96.2%、92.3%、88.2%和 90.2%。值得一提的是,YOLO-L 模型在减少模型训练参数的同时,显著提高了对密集重叠果实的检测效果。(3)对比试验结果表明,相较于 Faster-RCNN、SSD、YOLOv5、

YOLOv8 和 YOLOv9 模型, YOLO-L 模型的平均精度均值(交并比=0.50)分别提升了 7.5 个百分点、14.0 个百分点、9.1 个百分点、6.0 个百分点和 4.9 个百分点。本研究结果不仅为自然环境下澳洲坚果果实的智能检测提供了高效准确的解决方案,也为澳洲坚果产业的智能采摘提供了更有效的技术支持。

参考文献:

- [1] MERELES CEUPPENS L G, FERRO E A, ALVARENGA SOSA N L, et al. Chemical composition of *Macadamia integrifolia* (Maiden and Betche) nuts from Paraguay[J]. International Food Research Journal, 2017, 24(6):2599-2608.
- [2] 苏继龙,刘远力,吴金东. 香蕉采摘机械的研究现状与发展趋势[J]. 中国农业大学学报,2019,24(7):131-137.
- [3] MOREIRA G, MAGALHÃES S A, PINHO T, et al. Benchmark of deep learning and a proposed HSV colour space models for the detection and classification of greenhouse tomato[J]. Agronomy, 2022,12(2):356.
- [4] 柳 军,孔 杰,皮 杰,等. 果蔬分拣设备研究现状及发展趋势[J]. 中国农机化学报,2024,45(8):120-125.
- [5] 邢卓冉,丁松爽,张 凯,等. 计算机视觉与深度学习技术在烟叶生产上的研究进展[J/OL]. 中国农业科技导报,2024:1-11 [2024-08-01]. <https://doi.org/10.13304/j.nykjdb.2023.0379>.
- [6] 廖 越,李智敏,刘 偲. 基于深度学习的人—物交互关系检测综述[J]. 中国图象图形学报,2022,27(9):2611-2628.
- [7] DONG C, ZHANG Z W, YUE J, et al. Automatic recognition of strawberry diseases and pests using convolutional neural network[J]. Smart Agricultural Technology, 2021,1:100009.
- [8] ZHENG Z H, XIONG J T, LIN H, et al. A method of green *Citrus* detection in natural environments using a deep convolutional neural network[J]. Frontiers in Plant Science, 2021,12:705737.
- [9] 金 梅,李义辉,张立国,等. 基于注意力机制改进的轻量级目标检测算法[J]. 激光与光电子学进展,2023,60(4):385-392.
- [10] 刘 芳,刘玉坤,林 森,等. 基于改进型 YOLO 的复杂环境下番茄果实快速识别方法[J]. 农业机械学报,2020,51(6):229-237.
- [11] 侯依廷,饶 元,宋 贺,等. 复杂大田场景下基于改进 YOLOv8 的小麦幼苗期叶片数快速检测方法[J]. 智慧农业(中英文),2024,6(4):128-137.
- [12] 吴小燕,郭 威,朱轶萍,等. 基于改进 YOLOv8s 的大田甘蓝移栽状态检测算法[J]. 智慧农业(中英文),2024,6(2):107-117.
- [13] 赵鹏飞,钱孟波,周凯琪,等. 改进 YOLOv7-Tiny 农田环境下甜椒果实检测[J]. 计算机工程与应用,2023,59(15):329-340.
- [14] AHMAD T, MA Y L, YAHYA M, et al. Object detection through modified YOLO neural network[J]. Scientific Programming, 2020, 2020:8403262.
- [15] SAPKOTA R, QURESHI R, CALERO M F, et al. YOLOv10 to its genesis: a decadal and comprehensive review of the you only look once (YOLO) series [EB/OL]. (2024-07-26) [2024-08-16]. <https://arxiv.org/pdf/2406.19407>.
- [16] LAWAL M O. Tomato detection based on modified YOLOv3 framework[J]. Scientific Reports, 2021,11(1):1447.
- [17] CHEN W K, LU S L, LIU B H, et al. Detecting *Citrus* in orchard environment by using improved YOLOv4[J]. Scientific Programming, 2020, 2020:8859237.
- [18] YAO J, QI J M, ZHANG J, et al. A real-time detection algorithm for kiwifruit defects based on YOLOv5[J]. Electronics, 2021, 10(14):1711.
- [19] ZHANG Z X, LU X Q, CAO G J, et al. ViT-YOLO: transformer-based YOLO for object detection [C]//IEEE. 2021 IEEE/CVF International Conference on Computer Vision Workshops. Montreal, BC, Canada: IEEE, 2021:2799-2808.
- [20] TERVEN J, CORDOVA-ESPARZA D M, ROMERO-GONZÁLEZ J A. A comprehensive review of YOLO architectures in computer vision: from YOLOv1 to YOLOv8 and YOLO-NAS[J]. Machine Learning and Knowledge Extraction, 2023,5(4):1680-1716.
- [21] YAN B, FAN P, LEI X Y, et al. A real-time apple targets detection method for picking robot based on improved YOLOv5[J]. Remote Sensing, 2021,13(9):1619.
- [22] LI C, LIN J Q, LI Z, et al. An efficient detection method for *Litchi* fruits in a natural environment based on improved YOLOv7-Litchi[J]. Computers and Electronics in Agriculture, 2024,217:108605.
- [23] YANG G L, WANG J X, NIE Z L, et al. A lightweight YOLOv8 tomato detection algorithm combining feature enhancement and attention[J]. Agronomy, 2023,13(7):1824.
- [24] WANG C Y, YEH I H, MARK LIAO H Y. YOLOv9: learning what you want to learn using programmable gradient information [M]. Cham: Springer Nature Switzerland, 2024:1-21.
- [25] BAKIRCI M, BAYRAKTAR I. YOLOv9-enabled vehicle detection for urban security and forensics applications [C]//IEEE. 2024 12th International Symposium on Digital Forensics and Security, San Antonio, TX, USA: IEEE, 2024:1-6.
- [26] LIU S H, YANG Y C, JING X J, et al. BiFormer: an end-to-end deep learning approach for enhanced image-based photoplethysmography and heart rate accuracy [M]. Amsterdam: IOS Press, 2024: 205-214.
- [27] WANG X L, XIAO T T, JIANG Y N, et al. Repulsion loss: detecting pedestrians in a crowd [C]//IEEE. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018:7774-7783.
- [28] LI H L, LI J, WEI H B, et al. Slim-neck by GSCnv: a lightweight-design for real-time detector architectures [J]. Journal of Real-Time Image Processing, 2024,21(3):62.
- [29] GOUTTE C, GAUSSIER E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation [M]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005:345-359.
- [30] FLACH P, KULL M. Precision-recall-gain curves: PR analysis done right [J]. Advances in Neural Information Processing Systems, 2015, 28:838-846.

(责任编辑:徐 艳)