计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
2026,35(2):165−174 [doi: 10.15888/j.cnki.csa.010081] [CSTR: 32024.14.csa.010081]
©中国科学院软件研究所版权所有.
E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

改进的轻量化 YOLO11 棉花病害检测①

蒋碧波, 汪明锐, 钱晓杭, 徐涵宇, 杨　超

(湖北大学 计算机学院, 武汉 430062)
通信作者: 杨　超, E-mail: stevenyc@hubu.edu.cn

摘　要: 棉花作为我国重要的经济作物, 其病害问题对产量和质量造成了显著影响, 快速准确地识别病害类型至关重要, 然而现有的目标检测模型大多侧重于提高检测精度而忽略检测效率, 这些模型通常存在着计算量大、参数量大、难以在资源受限的边缘设备上部署的问题. 本文针对这些问题提出一种改进的 YOLO11 算法——SDP-YOLO. 该算法以 StarNet 作为主干网络, 从而有效减少模型的参数量; 提出 DRBNCSPELAN4 模块代替颈部网络中的 C3K2, 强化特征中语义信息和位置信息, 提高模型特征提取能力; 提出轻量级部分卷积检测头 EPCD, 提高模型对重要特征的提取能力并且显著减少复杂度; 使用 Wise-IoU 边界损失函数, 提升网络边界框回归性能和对目标病害的检测效果. 实验结果表明, 改进后模型的参数量、浮点运算总数和模型大小相比原方法分别降低了 43.8%、96.9% 和 39.6%, 同时检测精度提升 1.3%, FPS 增加 40 帧, 显著提升了检测效率.

关键词: 目标检测; 作物病害检测; YOLO11; 轻量化

引用格式: 蒋碧波,汪明锐,钱晓杭,徐涵宇,杨超.改进的轻量化 YOLO11 棉花病害检测.计算机系统应用,2026,35(2):165–174. http://www.c-s-a.org.cn/1003-3254/10081.html

Cotton Disease Detection Based on Improved Lightweight YOLO11

JIANG Bi-Bo, WANG Ming-Rui, QIAN Xiao-Hang, XU Han-Yu, YANG Chao

(School of Computer Science, Hubei University, Wuhan 430062, China)

Abstract: Cotton is an important economic crop in China, and its diseases have a significant impact on yield and quality. Therefore, it is crucial to quickly and accurately identify the types of diseases. However, existing object detection models mostly focus on improving detection accuracy while neglecting detection efficiency. These models typically have large computational requirements and a large number of parameters, making it difficult to deploy them on resource-constrained edge devices. To address this issue, this study proposes an improved YOLO11 algorithm——SDP-YOLO. The StarNet is used as the backbone network structure to reduce the number of model parameters. A DRBNCSPELAN4 module is proposed to replace C3K2 in the neck network, enhancing the semantic and positional information within the features to improve the model's feature extraction capability. A lightweight partial convolution detection head, EPCD, is introduced to improve the model's ability to extract important features and significantly reduce complexity. The Wise-IoU bounding box loss function is used to improve the network's performance in bounding box regression and detection effectiveness for target diseases. Experimental results show that the improved model demonstrates significant reductions in various metrics: a 43.8% decrease in the number of parameters, a 96.9% decrease in the total floating-point operations, and a 39.6% decrease in model size, while increasing the detection accuracy by 1.3% and the FPS by 40 frames, significantly improving detection efficiency.

Key words: object detection; crop disease detection; YOLO11; lightweight

① 基金项目: 国家自然科学基金 (61977021)
收稿时间: 2025-07-14; 修改时间: 2025-09-05; 采用时间: 2025-09-22; csa 在线出版时间: 2025-12-31

棉花作为我国重要的经济作物, 广泛种植于全国多个区域, 其种植面积和产量均居世界前列. 然而, 棉花病害问题对农业经济产生了显著的负面影响. 据研究, 棉花枯萎病发病严重程度与单株籽棉产量呈显著负相关, 病株籽棉产量可比健康植株减产 18.9%–68.5%[1], 每年中国棉花因病虫害损失达 10%–20%[2], 根据 2020 年国家统计局的产量数据计算得每年总产量损失约 88.65 万吨. 在新疆等主要产区, 棉花病虫害造成的产量损失尤为严重, 2014–2018 年, 棉花虫害造成的产量损失占总产量损失的 70% 以上[3]. 然而, 现有的病害检测技术大多依赖深度学习模型, 这些模型虽然精度较高, 但计算复杂度高, 难以在资源受限的边缘设备上部署. 因此, 开发一种轻量化、高效率的棉花病害检测方法具有重要意义. 近年来, 计算机软硬件的日新月异, 使我们从过去在地头田间以人眼识别病害, 到现在上传图像就可根据其特征智能高效识别. 现代深度卷积神经网络系统往往消耗大量算力, 对设备平台提出了较高要求, 不适用于现场基于边缘计算的检测环境, 故轻量化卷积神经网络的应用需求广泛.

目前的目标检测算法主要分为两阶段算法和单阶段算法两大类. 两阶段算法中, R-CNN[4]系列算法较为突出, 它们依赖预先确定的候选区域来进行目标的分类和边界框回归. 单阶段算法, 如 SSD[5]和 YOLO[6]系列, 则直接在图像上完成目标的检测和分类任务, 不需要事先确定候选区域, 而是直接将目标定位问题转化为一个回归任务, 通过网络提取的特征直接预测病害的位置及类别, 这种方法具备高准确度、快速检测、训练周期短和较低的计算需求等优势. 例如, 刘阗宇等[7]采用 Faster R-CNN 提出一种基于卷积神经网络的病害检测方法, 其平均精度均值 (mean average precision, mAP) 为 75.52%. 齐永康[8]提出一种改进 YOLOv3 模型的轻量化检测方法, 用了一个轻量级的 MobileNetv1 网络来替换 YOLOv3 的主干网络实现轻量化, 与基准模型 YOLOv3 相比, mAP 提升 7%, 模型规模缩减至原模型的 40%. 聂忠强等[9]使用 DSConv 和 FEM 特征提取模块的组合来替换主干网络中的部分常规卷积进行轻量化改进, 引入结构化状态空间模型构建 CBAMamba 模块, 使用 Mamba 结构高效处理特征; 此后将检测头处的卷积替换为 RepConv, 并减小卷积层; 最后, 使用 W-IoU, 提高模型收敛速度, 进一步提升模型检测性能; 比原始 YOLOv8n 算法分别提升 1.6% 的 mAP@0.5 和 1.2% 的 mAP@0.5:0.95, 同时提升 8.0% 的 FPS 并降低 13.3% 的模型参数量. 郭文娟等[10]在 YOLOX-S 和 YOLOv7 算法中引入 SE、CBAM 和 ECA 注意力模块实现了检测速度、检测精度和模型体量之间的平衡, 能够在复杂背景下对棉花病害有较好的实时检测效果.

棉花病害检测领域, 在国内外还是一个较为空白的领域. 因此, 开发一种计算负担小、延迟低的轻量化检测方法变得尤为迫切, 同时也要确保检测精度不会受到太大影响. 为了应对这些挑战, 本研究选取 YOLO 系列中最新推出的 YOLO11 模型, 其具有精度高、参数量小和模型大小较小等优点. 在此基础上, 棉花病害检测方法得到改进和优化, 目的是为资源受限的边缘设备开发出一种快速且准确的病害检测模型.

1 YOLO11 算法

1.1 YOLO11 算法

YOLO11 是 YOLO 系列的最新版本, 在检测精度和速度上相较 YOLOv8 有显著提升. YOLO11 提供 5 种不同规模的模型, 分别是 n、s、l、m 和 x, 以适应不同的应用需求. 这些模型的网络复杂度逐渐增加, 从而提高检测的准确性, 但同时参数量和计算量也随之增加. 在这些模型中, YOLO11n 凭借最少的参数和最快的检测速度脱颖而出. 为了确保算法的高效执行和模型的紧凑性, 本研究选择 YOLO11n 版本.

YOLO11 相较 YOLOv8, 在主干网络、颈部网络、头部网络都有显著的改进. 首先, YOLO11 的主干网络和颈部网络中采用了 C3K2 模块, 这一设计通过引入可变的多尺度卷积核, 有效取代了 YOLOv8 中的 C2F 模块, 从而增强网络对图像中不同尺度物体的捕捉能力. 在主干网络中引入了类似注意力机制的 C2PSA 模块, 该模块位于 SPPF 模块之后, 进一步提升网络对图像特征的提取能力, 使得特征提取更加精细和高效. YOLO11 引入 YOLOv10 的头部设计理念, 采用基于无锚点的解耦头部网络. 这种设计将回归和分类任务分别交由标准卷积块和深度可分离卷积 (DWConv)[11]处理, 有效减少冗余计算, 提高模型的运行效率. YOLO-11n 的网络结构图如图 1 所示.

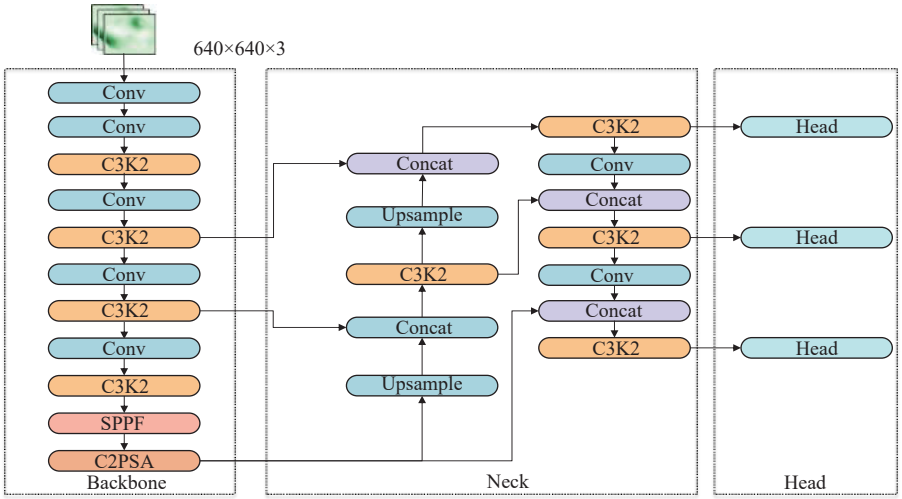

图 1 YOLO11n 网络结构图

2 YOLO11n 的轻量化改进——SDP-YOLO

针对棉花病害检测领域中模型参数多、模型文件体积大、实时检测性能不足以及检测效果不够精准等问题, 本研究提出一种基于轻量化改进的 YOLO11n 算法, 首先, 采用轻量级卷积神经网络 StarNet[12]作为主干网络, 显著降低模型的参数量和计算复杂度. 其次, 为增强网络对特征的提取和融合能力, 本研究汲取了 YOLOv9[13]的特征提取模块 RepNCSPELAN4 的架构优势, 并结合 UniRepLKNet[14]中扩张卷积思想, 设计了特征提取模块 DRBNCSPELAN4 (dilated reparam block nested cross-stage progressive enhanced layer aggregation network) 模块, 用以替换颈部网络中原有的 C3K2, 使模型能更有效地捕捉小目标病害的特征, 提高对细微病变的识别精度, 同时减少模型参数. 接着, 提出 EPCD (efficient partial convolution detection head), 这是一种基于部分卷积[15]的高效轻量化检测头, 以替代头部网络, 显著降低参数量, 大幅度提升了 FPS 并使模型更加精简高效. 最后, 为增强模型在边界框回归方面的表现, 引入 Wise-IoU 损失函数[16], 这是一种将分类信息融入 IoU 计算中的加权插值方法, 以提升模型的边界框回归能力, 有效解决类别不平衡问题, 从而提高检测准确性. 基于这些改进, 本文模型被命名为 SDP-YOLO 网络. SDP-YOLO 网络结构图如图 2 所示.

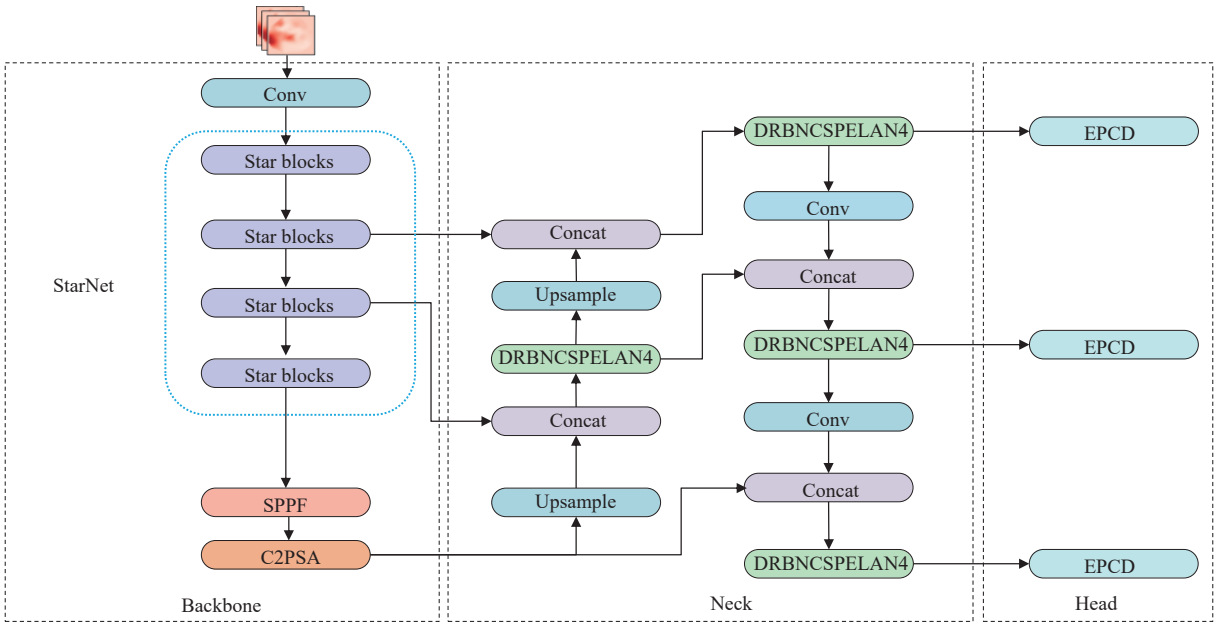

图 2 SDP-YOLO 网络结构图

2.1 StarNet 轻量化网络主干

StarNet 是微软在 2024 年 3 月推出的一种新型轻量化卷积神经网络架构. 在本研究中, StarNet 采用了四阶段的分层结构, 通过卷积层逐步降低特征图的分辨率, 并且通过在每个阶段将通道数量加倍, 以实现高效的特征提取. 其特征提取过程主要依赖多个“星型模块” Star block 中“星型运算”的堆叠, 无需复杂的网络结构或精细调整的超参数. 在单层卷积网络中星型运算的矩阵表达式为$w_1^{\mathrm{T}}x*w_2^{\mathrm{T}}x$. 其具体计算公式如式 (1)–(2) 所示:

$$\begin{aligned} & w_1^{\mathrm{T}}x*w_2^{\mathrm{T}}x \\ &= \left(\sum_{i=1}^{d+1} w_1^i x^i\right)*\left(\sum_{j=1}^{d+1} w_2^j x^j\right) \\ &= \sum_{i=1}^{d+1}\sum_{j=1}^{d+1} w_1^i w_2^j x^i x^j \\ &= \underbrace{\alpha_{(1,1)}x^1x^1+\cdots+\alpha_{(4,5)}x^4x^5+\cdots+\alpha_{(d+1,d+1)}x^{d+1}x^{d+1}}_{\frac{(d+2)(d+1)}{2}\text{items}} \end{aligned} \tag{1}$$

$$\alpha_{(i,j)}=\begin{cases} w_1^i w_2^j & \text{if } i==j \\ w_1^i w_2^j+w_1^j w_2^i & \text{if } i!=j \end{cases} \tag{2}$$

其中, i,j 是规定的通道标号, α 是每个项目的系数, “*”表示乘法. 由式 (1) 可以将“星型运算”看作一个由 $(d+2)(d+1)/2$ 个项组成的组合, 除最后一项外, 其余项均与 x 成非线性相关. StarNet 采用四阶段分层结构, 通过卷积层对特征图实现下采样, 并利用 Star block 模块进行分组高效的特征提取. 为进一步提升效率, 普通归一化层被替换为批量归一化层 (batch normalization), 并将其置于 DWconv 之后, 在推理阶段进行计算融合, 在不改变通道扩展因子的前提下 (始终为 4), 网络宽度在每个阶段都进行翻倍[17]. 此外, StarNet 作为新一代轻量化卷积神经网络, 延续了上一代轻量级网络 MobileNetV2[18]的设计思路, 将激活函数替换为 ReLU6. 与 GELU 相比, ReLU6 将输出值限制在 6 以内, 在一定程度上避免了梯度爆炸的问题. StarNet 通过“星型运算”实现高效的特征提取, 其优势在于能够在低维度的空间执行计算同时兼顾高维的特征信息, 无需产生额外计算开销. 用 StarNet 替换 YOLO11n 的主干网络后, 使模型瘦身, 更加简洁高效. 其网络结构如图 3 所示.

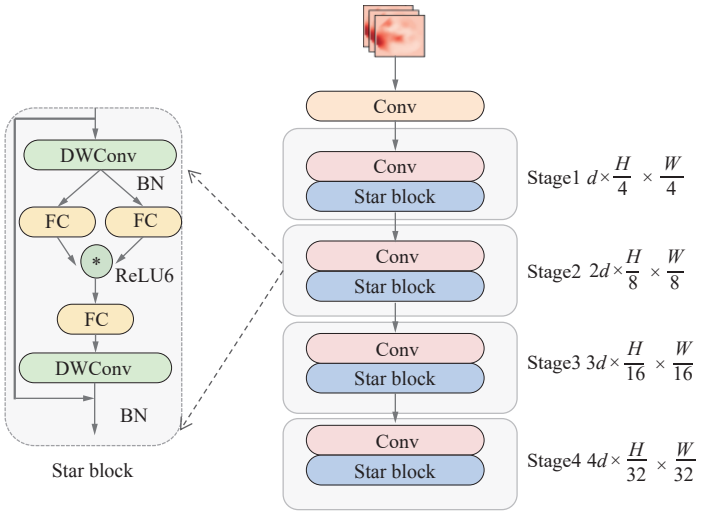

图 3 StarNet 结构图

2.2 DRBNCSPELAN4 特征提取模块

YOLO11 作为目标检测领域的最新模型, 其在效率、精确度等方面都取得显著的提升, 但是随着模型不断地增大和复杂化, 对于计算资源的需求也随之水涨船高. 针对这一问题, 本研究设计了 DRBNCSPELAN4 模块, 以此提高模型检测性能并让模型更加高效. 本研究采用了 YOLOv9 的 RepNCSPELAN4 特征提取模块的架构, 同时融合了扩张卷积思路, 进而设计出特征提取模块 DRBNCSPELAN4, RepNCSPELAN4 的 RepNCSP 模块将输入通道进行拆分并行处理, 在不牺牲检测精度的前提下减少前向传播的成本, 减少冗余参数, 提升通道信息利用效率. DRBNCSPELAN4 在

RepNCSP 的基础上, 采用扩张卷积思想, 通过大核卷积扩大感受野捕获长程依赖, 结合小核扩张卷积提升细节特征的捕获能力, 能够有效捕捉图像中的多尺度特征, 并减少计算量, 同时能够捕获更丰富的空间特征, 包括小尺度信息和稀疏模式. 其中对于一个小核卷积 $k\times k$ 和空洞率 r, 其等效非空洞卷积核大小见式 (3). 训练阶段使用普通大核卷积层作为主分支, 以此捕获更宽的上下文信息, 并使用多个不同空洞率的小核扩张卷积作为辅助分支, 最后通过 BN 层将所有分支的特征信息逐元素相加, 见式 (4), 以达到提升检测精度的效果. 而在推理阶段, 将扩张卷积核通过转置卷积和零填充, 重参数化后可以等效转换为单一的普通大核卷积层, 其公式见式 (5), 避免了额外的计算开销, 显著提升推理速度. DRBNCSPELAN4 模块结构如图 4 所示. Dilated reparam block 结构图如图 5 所示. 但由于 RepNCSPELAN4 结构中大量使用 CSP 风格重复的分支[19], 导致计算量有所上升, FPS 降低, 但考虑到能同时提升检测精度并降低参数量, DRBNCSPELAN4 仍是较为优秀的特征提取模块, 由此也引出了轻量化检测头 EPCD, 以解决计算量上涨和 FPS 下降的问题.

$$K_{\text{equiv}} = (k-1)\cdot r + 1 \tag{3}$$

$$W_{\text{final}} = W_{\text{large}} + \sum_i Pad \tag{4}$$

$$W' = \text{Conv}_{\text{transpose}}2d(W, I, \text{stride} = r) \tag{5}$$

其中, 式 (3) 中 k 为卷积核大小, r 为空洞率. 式 (4) 中 Pad 为 W_i' 对齐尺寸. 式 (5) 中 I 是值为 1 的张量 (单位核).

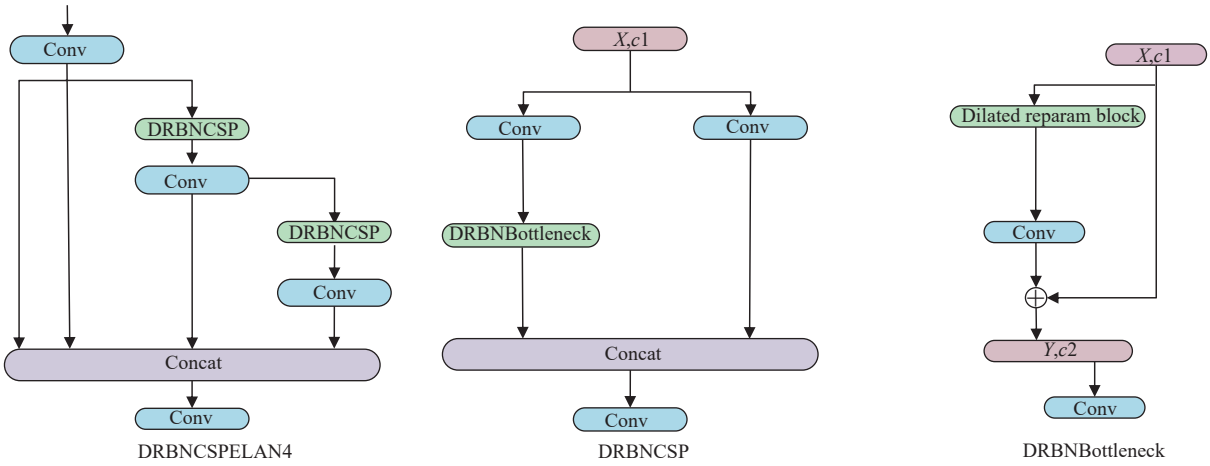

图 4 DRBNCSPELAN4 模块

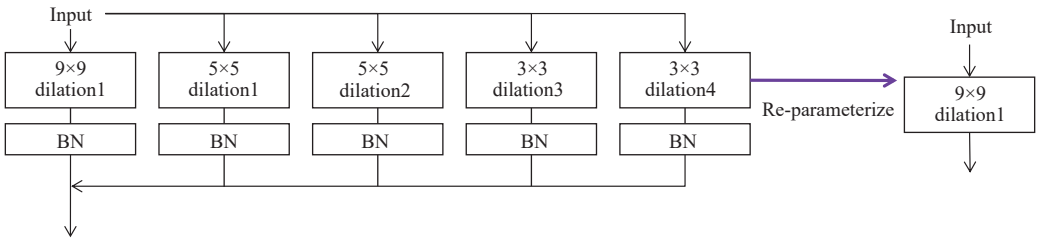

图 5 Dilated reparam block 模块

2.3 EPCD 轻量化检测头

YOLO11 在检测头设计方面相较 YOLOv8 实现了改进, 将普通卷积层替换为深度可分离卷积 (DWConv), 大幅降低模型的参数量和计算复杂度, 同时保证了检测性能, YOLO11 检测头结构图如图 6. 尽管 YOLO11 引入 DWConv 以实现轻量化, 但其检测头仍需进一步优化以提高效率. 为了进一步实现轻量化, 本文对 YOLO11 的检测头进行了重新设计, 提出了一种名为 EPCD (efficient partial convolution detection head) 的检测头结构. 其结构图如图 7 所示. 该结构使用部分卷积 (partial convolution, PConv) 替代普通卷积, 部分卷积将 1/4 的通道进行普通卷积操作, 另外 3/4 的通道保持

不变, 在卷积操作完成后再将通道进行拼接, 恢复原有通道数. 相当于仅进行了 1/4 的计算量, 能够有效减少计算冗余, 提高计算效率. 部分卷积的结构图如图 8 所示. 检测头的核心设计思路是用一个部分卷积层和一个 1×1 卷积层来替代 YOLO11 检测头中的普通卷积层以及深度可分离卷积层. 同时, 该结构不再将特征分开进行分类和边界框预测, 而是通过一条路径同时进行特征提取、分类和预测. 这种结构设计不仅减少了卷积层的数量, 还通过运用更轻量化的部分卷积, 有效减少计算量和参数量, 并提升 FPS, 让模型更适合在资源受限的设备上部署.

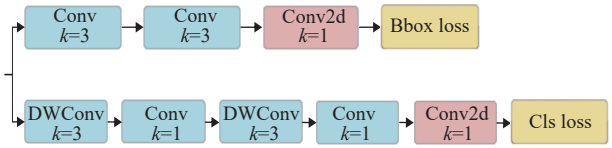

图 6 YOLO11 检测头结构图

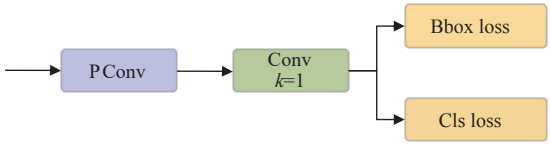

图 7 EPCD 结构图

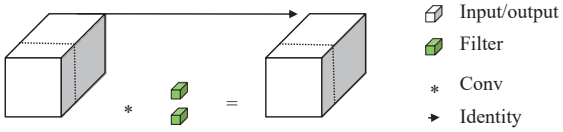

图 8 PConv 结构图

2.4 Wise-IoU 损失函数

在棉花病害检测任务中, 枯萎病的病理特征占比高, 平衡不同比例目标病害的检测是提升检测精度的关键. YOLO11 使用的损失函数为 CIoU loss[20], 其计算公式如式 (6) 所示. 其中, IoU 为预测框与棉花病害真实框的交并比, b^A 和 b^B 分别表示两框的中心点, ρ 表示两点之间的欧几里得距离, c 表示两框的最小外接矩形的对角线长度; α 为权重系数, 计算公式如式 (7); v 是衡量一致性的度量, 计算公式如式 (8).

$$L_{\mathrm{CIoU}} = 1 - IoU + \frac{\rho^2(b^A, b^B)}{c^2} + \alpha v \tag{6}$$

$$\alpha = \frac{v}{(1 - IoU) + v} \tag{7}$$

$$v = \frac{4}{\pi}\left(\arctan\frac{w^{\mathrm{gt}}}{h^{\mathrm{gt}}} - \arctan\frac{w^{\mathrm{pred}}}{h^{\mathrm{pred}}}\right) \tag{8}$$

CIoU loss 虽然采用了单调聚焦机制, 但没有平衡难度较大的样本与简单样本之间的关系, 因此当训练集包含低质量样本时, 模型的检测性能会下降, 本研究引入具有动态非单调聚焦机制的 Wise-IoU 损失函数. 通过使用“离群度” β 来替代传统的 IoU, 能够更好地识别并利用异常值来评估锚框的质量, 从而避免几何因素对模型检测精度的不必要惩罚[21]. Wise-IoU 通过将较小的梯度增益分配给高质量的锚框, 减弱高质量锚框之间的竞争; 同时, 减少低质量样本带来的有害梯度影响, 从而帮助模型更专注于质量较为平衡的样本, 提升模型的泛化能力. 计算公式如式 (9)–(11):

$$L_{\mathrm{WIoU}} = rR_{\mathrm{WIoU}},\ r = \frac{\beta}{\delta\alpha^{\beta-\delta}} \tag{9}$$

$$\beta = \frac{L^*_{\mathrm{IoU}}}{\overline{L_{\mathrm{IoU}}}} \in [0, +\infty) \tag{10}$$

$$R_{\mathrm{WIoU}} = \left(1 - \frac{W_i H_i}{S_u}\right)\exp\left(\frac{\left(x_p - x_{\mathrm{gt}}\right)^2 + \left(y_p - y_{\mathrm{gt}}\right)^2}{\left(W_g^2 + H_g^2\right)^*}\right) \tag{11}$$

其中, α 和 δ 为超参数, β 为离群度, IoU 的损失均值作为归一化因子, *为不参加反向传播的部分, (x_p, y_p) 代表预测框坐标, $(x_{\mathrm{gt}}, y_{\mathrm{gt}})$ 代表真实框坐标. H 和 W 分别代表两框的高和宽. Wise-IoU 的原理图如图 9 所示.

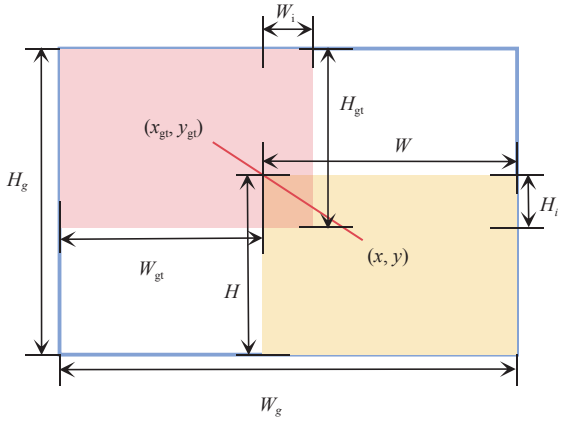

图 9 Wise-IoU 的原理图

3 实验分析

3.1 数据来源

本研究的数据集来自 Roboflow Universe 平台 zia-ilp5l 作者的公开棉花病害数据集, 其中共包含 6 种棉花图像, 涵盖健康叶片及 5 种病害类型, 具体包括枯萎病、卷曲病、灰霉病、褐斑病和黄萎病. 图像尺寸

统一为 640×640 像素. 数据集共包含 3 792 张图像样本, 其中枯萎病图像 652 张, 卷曲病图像 657 张, 灰霉病图像 678 张, 褐斑病图像 755 张, 黄萎病图像 691 张, 健康叶片图像 604 张. 数据集分为 3 个部分: 训练集 (train) 包含 3 708 张图片, 测试集 (test) 有 233 张, 验证集 (val) 为 232 张. 数据集样本如图 10 所示, 数据分布直方图如图 11 所示, 占比最大的褐斑病仅比占比最小的健康叶片多 3%.

图 10 数据集样本

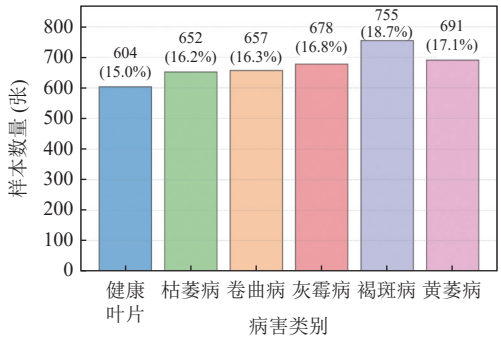

图 11 数据分布直方图

3.2 实验环境

实验环境为 64 位 Windows 10 专业版操作系统, 配备 NVIDIA GeForce RTX 4060 GPU (显存 8 GB) 和 16 GB 主机内存. 编程语言为 Python 3.10.15, 使用 CUDA v12.1 加速 GPU 计算, 并基于 PyTorch 2.4.1 深度学习框架进行模型训练. Torchvision 版本为 0.17.2+cu121. 实验过程中的超参数设置情况为: 输入图片尺寸为 640×640×3, batch 为 16, 优化器使用 SGD, 迭代轮次为 300 次, 禁用 mosaic 增强轮次数为 0, workers 为 4, lrf 为 0.01, weight_decay 为 0.000 5, momentum 为 0.937, warmup_momentum 为 0.8, lr0 为 0.01.

3.3 评价指标

本实验借助精确度 (P)、召回率 (R) 及平均精度均值 (mAP) 来衡量模型性能. AP@0.5 是在 IoU 阈值为 0.5 的情况下, 针对某一类别的特定样本计算的平均精确度. mAP@0.5 则是在 IoU 阈值为 0.5 的情况下, 对所有类别的精确度取平均值, 反映了模型精确度与召回率之间的关系. mAP@0.5 越高, 则说明模型在高召回率时仍能维持较高精确度. 而 mAP@0.5:0.95 是在 IoU 阈值在 0.5–0.95 范围内计算的平均精确度, 其反映了严苛环境下模型的检测性能[22]. 计算公式见式 (12)–(15).

$$P = \frac{TP}{TP+FP} \tag{12}$$

$$R = \frac{TP}{TP+FN} \tag{13}$$

$$AP = \int_0^1 P(R)\mathrm{d}R \tag{14}$$

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \tag{15}$$

3.4 不同目标检测算法间对比实验分析

为验证本研究算法相较于当前普遍在工农业使用的主流目标检测模型的优越性, 以浮点运算总数、参数量、模型大小、FPS 作为轻量化评价指标; 将精确度、召回率、mAP@0.5、mAP@0.5:0.95 作为评估模型检测性能的指标. 将本研究提出的算法 SDP-YOLO 与 Faster R-CNN、SSD、YOLOv3、YOLOv5n、YOLOv8n、YOLOv9t、YOLOv10n、YOLO11n、Hyper-YOLO[23]算法在相同的实验条件下进行对比, 实验结果如表 1 所示.

由于作物病害一般成片出现, YOLO 自带的可视化技术并不直观且很难体现模型改进前后的差异, 而深度学习中的热力图可视化技术, 是通过计算特征图与权重的加权和, 并根据结果生成对应的热力图. 可以通过颜色的深浅变化直观地展示模型对不同区域的关注程度, 直观地呈现模型的学习过程和检测效果. 通过热力图, 研究人员可以清晰地观察模型的内部运行机

制以及其对输入数据的敏感程度, 从而增强模型的可解释性, 进而明确模型在处理数据时的关键关注区域. 本研究利用 Grad-CAM[24]方法生成目标检测热力图. 实验从数据集中随机选取 6 种不同病害的图片, 并分别使用 YOLO11n 模型及本文算法进行棉花病害检测. 如图 12 所示.

表 1 对比实验结果

模型	*P*	*R*	*mAP*@0.5	*mAP*@0.5:0.95	FLOPs (G)	Parameters (M)	Size (M)	FPS (f/s)
Faster R-CNN	0.87	0.72	0.78	—	—	42	31.6	—
SSD	0.874	0.731	0.74	—	—	27	19.8	—
YOLOv3	0.871	0.767	0.863	0.637	18.9	12.1	23.3	260
YOLOv5n	0.974	0.76	0.882	0.713	7.1	2.52	5.1	224
YOLOv8n	0.955	0.907	0.9	0.727	8.1	3.01	6.0	253
YOLOv9t	0.952	0.816	0.907	0.732	7.6	1.97	4.5	95
YOLOv10n	0.881	0.811	0.869	0.69	6.5	2.23	5.5	213
YOLO11n	0.95	0.842	0.885	0.712	6.3	2.58	5.3	202
Hyper-YOLO	0.954	0.829	0.895	0.703	9.5	3.61	7.3	163
SDP-YOLO	0.963	0.843	0.898	0.718	3.2	1.45	3.2	240

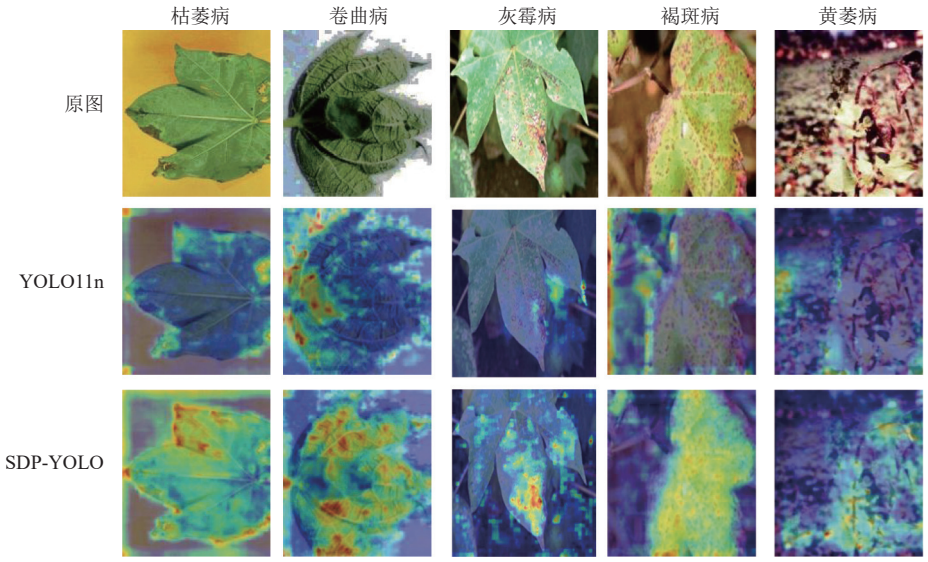

图 12 热力图可视化

如表 1 所示, Faster R-CNN 和 SSD 的计算量和参数量较大, 生成的模型文件体积也较大, 难以满足本数据集的轻量化实时检测需求. YOLOv9t 虽然在其他指标上表现优于其他模型, 但其浮点运算总数较高, 实时检测性能在对比模型中最差且多次训练的结果波动较大. YOLOv8n 的精度略高于 YOLOv5n, 但其参数量、浮点运算总数和模型大小均大于 YOLOv5n. YOLO11n 和 YOLOv10n 的轻量化指标相近, 但 YOLO11n 的检测精度更高, 约高出 3%. Hyper-YOLO 作为最新模型, 虽然在检测精度上有提升, 但其轻量化指标均低于 YOLO11n. 与原始 YOLO11n 相比, 改进的 SDP-YOLO 算法, 参数量、浮点运算总数和模型权重文件更小, 同时精度等指标均优于原始 YOLO11n.

在图 12 热力图可视化中, 原 YOLO11n 模型的热力值较低, 直观地表明其在某些区域存在漏检、误检问题. 尤其是在枯萎病和褐斑病上, 原 YOLO11n 没有精准识别病害. 而改进后的模型能够在语义信息较弱的目标中提取出更显著的特征, 热力值有显著提升, 并且误检漏检现象显著减少, 显示出更强的鲁棒性和泛化能力.

3.5 消融实验

为了验证本研究中算法改进模块的有效性和可行性, 以基准模型 YOLO11n 为参考. 通过精确度、召回率、*mAP*@0.5、*mAP*@0.5:0.95、浮点运算总数、每秒帧数、参数量和模型大小等指标进行评估, 相关实验结果已在表 2 中呈现.

表 2 消融实验结果

模型	*P*	*R*	*mAP*@0.5	*mAP*@0.5:0.95	FLOPs (G)	Parameters (M)	Size (M)	FPS (f/s)
YOLO11n	0.95	0.842	0.885	0.712	6.3	2.58	5.3	202
YOLO11n+StarNet	0.928	0.772	0.855	0.68	5	1.94	4.0	200
YOLO11n+StarNet+DRBNCSPELAN4	0.944	0.798	0.879	0.716	4.5	1.71	3.7	135
YOLO11n+StarNet+DRBNCSPELAN4+EPCD	0.946	0.811	0.884	0.709	3.2	1.45	3.3	240
YOLO11n+StarNet+DRBNCSPELAN4+EPCD+Wise-IoU	0.963	0.820	0.898	0.718	3.2	1.45	3.2	240

根据表 2 的实验结果, 将 YOLO11n 的原始主干网络替换为 StarNet 后, 通过星型运算在低维度空间融合高维特征, 无需复杂分支结构, 因此参数量和计算量显著降低, 更适配边缘设备. 模型的参数量、计算量和模型大小分别下降 25%、21%、25%, StarNet 的层数更少且星型运算兼顾低维度计算与高维特征融合, 在轻量化方面表现良好, 却使得特征提取在一定程度上更偏向于全局特征的整合, 而叶片病害特别是褐斑病斑点需要聚焦局部特征, 导致 *mAP*@0.5 下降 3%. 在改进主干网络的基础上, 用 DRBNCSPELAN4 替换 C3K2 特征提取模块, 该模块融合 YOLOv9 RepNCSPELAN4 的多分支并行处理和 UniRepLKNet 的扩张卷积思想, 可在不增加计算量的前提下, 提升对稀疏病变区域的细节捕捉能力, *mAP*@0.5 提升 1.3%, 且训练时通过 BN 层将多分支特征逐元素相加, 实现“全局+局部特征融合”; 推理时通过转置卷积重参数化, 将多分支转化为单一普通卷积, 可减少冗余计算, 使得浮点运算总数、参数量和模型大小分别下降 10%、12%、19%, 表明 DRBNCSPELAN4 在平衡模型性能和精确度方面具有显著优势, 但由于大量 CSP 结构的叠用导致 FPS 显著下降. 在此基础上, 加入轻量化检测头 EPCD, 替换 YOLO11 检测头的“深度可分离卷积+解耦”设计, 采用“部分卷积+1×1 卷积” 的一体化结构, 仅对 1/4 通道进行卷积操作, 其余 3/4 通道保持原值, 减少计算冗余, 卷积后通过通道拼接恢复维度, 特别是针对健康叶片, 大大提高了检测效率, 计算量仅为原始检测头的 27%; 除了 *mAP*@0.5:0.95 略有下降外, 模型的精确度和 *mAP*@0.5 分别提升了 0.04% 和 0.05%, 同时模型的参数量、浮点运算总数和大小分别减少了 15%、28% 和 11%, FPS 提升了 105 帧. 修改损失函数后, 通过离群度动态调整梯度分配, 将低质量样本的梯度权重降低, 同时将分类信息融入 *IoU* 计算, 减少类别混淆导致的边界框回归偏差, 使得模型的精确度、召回率、*mAP*@0.5 和 *mAP*@0.5:0.95 分别提升 1.7%、0.9%、1.4% 和 0.1%, 进一步提升模型检测性能.

综上, SDP-YOLO 模型相较原始 YOLO11n 基线模型, 精确度、*mAP*@0.5 和 *mAP*@0.5:0.95 分别提高 1.3%、0.8% 和 0.6%. 同时, 模型的浮点运算总数、参数量和模型权重文件大小分别降低 96.9%、43.8% 和 39.6%, FPS 提升 18%. 充分证明本文改进算法的有效性.

4 结论与展望

本文针对棉花病害多目标快速检测需求, 对 YOLO-11n 模型进行了轻量化优化改进. 具体而言, 研究将模型的主干网络替换为 StarNet 结构, 以减少模型的参数量、浮点运算总数和模型大小. 此外, 提出 DRBNCSP-ELAN4 特征提取模块, 提升模型特征提取能力、优化检测精度的同时, 减轻颈部网络的负担. 同时, 提出轻量级检测头 EPCD, 进一步降低模型参数量和浮点运算总数, 减小模型体积, 并提高检测精度. 本研究还引入了 Wise-IoU 损失函数, 以加权插值方式将分类信息融入 *IoU* 计算中, 解决了类别不平衡问题, 提升检测精度. 结合以上开发出一种更轻量化的棉花病害检测模型. 通过对实验结果的对比分析, 得出以下结论: 在同一环境中, 改进后的 SDP-YOLO 在棉花病害数据集中, 在不牺牲检测效果的同时模型更加轻量化.

本研究设置了 6 组热力图可视化对比实验, 以检验改进后 SDP-YOLO 模型检测的实际效果. 结果表明, SDP-YOLO 模型的热力值更加丰富, 减少了漏检可能, 其在棉花病害识别方面优于原始 YOLO11n.

未来的研究打算扩充棉花病害样本库, 并补充虫害样本信息. 此外, 计划对模型进行剪枝和知识蒸馏优化, 尝试将算法部署到边缘移动平台进行验证和改进.

参考文献

1 赵丽红, 冯自力, 冯鸿杰, 等. 棉花黄萎病对棉花单株产量和纤维品质的影响. 中国棉花, 2016, 43(7): 19–23.
2 李彩红. 棉花常见病害识别与防治. 湖南农业, 2023(10): 18–19.
3 明坤, 闫硕. 近几年我国棉花主要病虫害发生及防控情况分析. 棉花科学, 2020, 42(3): 13–19, 26.

4 Dong C, Zhang K, Xie ZY, *et al*. An improved cascade RCNN detection method for key components and defects of transmission lines. IET Generation, Transmission & Distribution, 2023, 17(19): 4277–4292.
5 Zhai SP, Shang DR, Wang SH, *et al*. DF-SSD: An improved SSD object detection algorithm based on DenseNet and feature fusion. IEEE Access, 2020, 8: 24344–24357. [doi: 10.1109/ACCESS.2020.2971026]
6 Jiang PY, Ergu D, Liu FY, *et al*. A review of YOLO algorithm developments. Procedia Computer Science, 2022, 199: 1066–1073. [doi: 10.1016/j.procs.2022.01.135]
7 刘阗宇, 冯全, 杨森. 基于卷积神经网络的葡萄叶片病害检测方法. 东北农业大学学报, 2018, 49(3): 73–83.
8 齐永康. 基于 MobileNet 与 YOLOv3 的路面障碍检测轻量化算法. 计算机系统应用, 2022, 31(2): 176–184. [doi: 10.15888/j.cnki.csa.008331]
9 聂忠强, 朱明. 基于轻量化 YOLOv8 模型的苹果快速识别. 计算机系统应用, 2025, 34(1): 200–210. [doi: 10.15888/j.cnki.csa.009749]
10 郭文娟, 冯全. 基于改进 YOLO 的棉花叶片病害检测. 干旱地区农业研究, 2024, 42(6): 195–205.
11 Han Q, Fan ZJ, Dai Q, *et al*. On the connection between local attention and dynamic depth-wise convolution. Proceedings of the 10th International Conference on Learning Representations. OpenReview.net, 2022.
12 Ma X, Dai XY, Bai Y, *et al*. Rewrite the stars. Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024. 5694–5703.
13 Wang CY, Yeh IH, Mark Liao HY. YOLOv9: Learning what you want to learn using programmable gradient information. Proceedings of the 18th European Conference on Computer Vision. Milan: Springer, 2025. 1–21.
14 Ding XH, Zhang YY, Ge YX, *et al*. UniRepLKNet: A universal perception large-kernel convnet for audio, video, point cloud, time-series and image recognition. Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024. 5513–5524.
15 Chen JR, Kao SH, He H, *et al*. Run, don't walk: Chasing higher FLOPS for faster neural networks. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 12021–12031.
16 Tong ZJ, Chen YH, Xu ZW, *et al*. Wise-IoU: Bounding box regression loss with dynamic focusing mechanism. arXiv:2301.10051, 2023.
17 方明, 张娇, 徐晶, 等. 基于改进 YOLOv8 的轻量化皮革缺陷检测方法. 电子测量技术, 2025, 48(1): 111–118.
18 Sandler M, Howard A, Zhu ML, *et al*. MobileNetV2: Inverted residuals and linear bottlenecks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 4510–4520.
19 邵佳慧, 姚百蔚, 田宏. 基于改进 YOLOv9 的黄瓜病害识别. 计算机系统应用, 2025, 34(7): 208–214. [doi: 10.15888/j.cnki.csa.009899]
20 Zheng ZH, Wang P, Liu W, *et al*. Distance-IoU loss: Faster and better learning for bounding box regression. Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York: AAAI Press, 2020. 12993–13000.
21 皮骏, 牛厚兴, 高志云. 融合 CA-BiFPN 的轻量化人体姿态估计算法. 图学学报, 2023, 44(5): 868–878.
22 马超伟, 张浩, 马新明, 等. 基于改进 YOLOv8 的轻量化小麦病害检测方法. 农业工程学报, 2024, 40(5): 187–195.
23 Feng YF, Huang JG, Du SY, *et al*. Hyper-YOLO: When visual object detection meets hypergraph computation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(4): 2388–2401. [doi: 10.1109/TPAMI.2024.3524377]
24 Selvaraju RR, Cogswell M, Das A, *et al*. Grad-CAM: Visual explanations from deep networks via gradient-based localization. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 618–626.

(校对责编: 李慧鑫)