张帅平,时雷,郑光,等.基于剪枝和知识蒸馏的YOLOv8轻量化苹果叶片病害检测方法[J].华中农业大学学报,2026,45(3):98-114.
DOI:10.13300/j.cnki.hnlkxb.2026.03.009

# 基于剪枝和知识蒸馏的YOLOv8轻量化苹果叶片病害检测方法

张帅平,时雷,郑光,王慧新,尹飞

河南农业大学信息与管理科学学院,郑州 450046

**摘要** 为解决现有苹果叶片检测方法在准确率、实时性及复杂噪声环境下鲁棒性等方面的问题,基于YOLOv8n提出一种轻量化、实时苹果叶片病害检测模型SMPD-YOLO。首先,将SPPFCSPC(spatial pyramid pooling fast cross stage partial CSP)特征金字塔模块嵌入到主干网络,增强特征融合能力。其次,引入MPD-IoU(minimum point distance-IoU)作为边界框回归损失函数,提高模型的精度和收敛速度。再次,通过基于层自适应幅度剪枝(layer adaptive magnitude-based pruning,LAMP)进一步压缩模型体积、降低浮点运算速度。最后,采用通道级知识蒸馏(channel-wise knowledge distillation,CWD)策略,提升检测性能。实验结果显示,改进后SMPD-YOLO模型在IoU(intersection over union)阈值大于0.5时的平均精度均值和帧率分别达到90.20%和133.3帧/s,模型权重、浮点运算速度分别为5.0 MB、$7.3\times10^{9}s^{-1}$。此外,改进模型在强光、弱光及图像模糊复杂噪声环境下仍能展现出优异的鲁棒性。结果表明,SMPD-YOLO模型兼具高准确性、轻量化和实时性,可在资源受限设备上实现高效叶片病害检测。

**关键词** 苹果;叶片病害检测;YOLOv8;轻量化;自适应幅度剪枝;知识蒸馏

**中图分类号** TP391.4;S661.1 **文献标识码** A **文章编号** 1000-2421(2026)03-0098-17

苹果叶片的健康状况直接影响其果实的发育与营养输送,进而影响苹果的品质和产量,但苹果生长发育过程中极易感染多种叶片病害[1]。因此,快速、准确地识别苹果叶片病害具有重要意义。传统苹果叶片病害诊断主要依赖人工经验,往往存在主观性强、耗时、成本高且易受外界环境干扰等问题[2]。早期的检测方法主要基于机器学习,通过手工提取颜色、纹理、几何形状等特征,并利用特定的特征向量进行模型训练和病害分类[3]。但该类方法对数据多样性较为敏感,且特征提取过程繁琐,难以满足大规模应用需求。近年来,基于深度学习的计算机视觉技术快速发展,为病害检测提供了更高效、更精准的自动化解决方案[4]。

基于深度学习的目标检测算法主要分为两阶段和单阶段检测算法。前者以R-CNN系列为代表,检测流程包括生成候选区域、目标分类和回归,如Faster R-CNN[5]等。例如,姜晟等[6]通过优化特征提取网络,引入特征金字塔网络和Rank&Sort(RS)损失函数,提出一种基于Faster R-CNN改进的识别算法,在茶叶病害数据集上平均精度均值为88.06%。然而该方法结构复杂、推理速度慢,且在藻斑病等特定病害上识别性能仍不理想,反映出其在模型效率与泛化能力方面的局限。

单阶段检测算法,如SSD、DETR系列和YOLO系列等,则是直接从图像中提取目标类别及位置信息,无需额外候选区域提取步骤,检测速度更快,计算效率更高,为实时检测任务提供解决方案。Li等[7]提出一种多尺度荔枝叶病害检测模型IMLL-DERT,在5种常见荔枝叶片病害数据上,模型平均精度、模型权重和浮点运算速度分别为84.40%、37.5 MB和$79.1\times10^{9}s^{-1}$,帧率为74.9帧/s。Yao等[8]针对RT-DETR检测方法模型参数量大、多目标识别不准确的问题,提出一种改进RT-DETR,引入注意力机制,通过SCConv优化特征融合,并对骨干网络

收稿日期:2025-09-25
基金项目:河南省科技攻关项目(242102521027);河南省科技研发计划联合基金项目(222301420113)
张帅平,E-mail:zsp2292@stu.henau.edu.cn
通信作者:尹飞,E-mail:yin.fei@henau.edu.cn

进行重构。实验结果表明，该模型在玉米病害的数据集上 mAP@0.5、参数量和帧率分别为 92.0%、$23.8\times10^6$和 89.3 帧/s。此类方法虽在精度上超越两阶段算法，但其解码器需通过序列化计算逐步生成预测，该过程难以并行优化，导致其推理速度通常不及采用单次前向传播的 YOLO 系列。因此，在实时检测场景中，YOLO 系列仍更具优势。

在 YOLO 系列中，研究者主要集中于通过引入轻量主干网络与注意力机制以优化效率。例如，赵嘉威等[9]基于 YOLOv4，将特征提取网络替换为 MobileNetV3 网络，引入 CBAM 注意力机制，并通过 *K*-means 聚类算法更新锚框信息，提出了一种 MC-YOLOv4 模型；实验结果表明，该模型对苹果叶片 5 种病害的平均精度为 97.25%，单张图像检测时间为 13.3 ms，模型权重为 55.5 MB。李昱达等[10]提出一种基于 YOLOv5 的改进苹果叶片病害检测算法 YOLOv5-CSEP，平均精度为 87.90%，模型权重为 6.89 MB，浮点运算速度为 $14.1\times10^9 s^{-1}$。袁杰等[11]提出一种改进的 YOLOv7 模型 YOLOv7_ECA，实验结果表明，在 4 种苹果叶片病害上，模型 mAP@0.5、mAP@0.5:0.95 分别为 90.50%、62.10%，单张图像检测时间为 12 ms。张惠莉等[12]将 GhostNetV2 特征提取网络引入 YOLOv8，并将 SPPFCSPC 特征金字塔池化、GAM 注意力机制以及 Focal-EIoU 损失函数，提出一种用于葡萄叶片病害识别的 GSGF 模型，平均精度达 97.10%，模型权重为 49.3 MB，推理时间为 43.7 ms。以上成果虽在特定任务中取得高精度，但普遍存在模型体积大、推理延迟高或计算量偏大的问题，反映出 YOLO 系列在轻量化与高精度之间尚未取得理想平衡。

部分研究虽在轻量化方面有所推进，如曾晏林等[13]针对苹果叶部病害检测提出一种融合自注意力机制和 Transformer 模块的模型 BCE-YOLOv5，浮点运算速度和模型权重分别减少至 $15.1\times10^9 s^{-1}$和 13.5 MB，3 种苹果叶片病害的检测平均精度为 91.10%。黄政等[14]针对梨树叶片病害检测精度低、参数量大问题，提出一种基于 YOLOv8n 的梨树叶片病害检测模型，在 DiaMOS plant 数据集上，mAP@0.5 为 90.10%，浮点运算速度为 $7.0\times10^9 s^{-1}$，参数量为 $2.4\times10^6$。Xue 等[15]在 YOLOv5s 的基础上集成 ACmix 和 CBAM 注意力机制，并用 RFB 模块取代 SPPF 模块，提出一种茶叶病害检测模型 YOLO-Tea，模型平均精度为 79.3%，模型权重 15.6 MB。但这些方法或未报告完整速度指标，或在精度上仍有提升空间，显示出当前研究在系统性地评估与优化综合性能方面的不足。

综上所述，现有研究在苹果叶片乃至多种作物的病害检测精度与效率的协同优化方面仍存在明显局限：高精度模型往往模型体积大、检测速度慢；轻量化模型则在精度保持、检测速度方面存在妥协；多数方法未能同时在精度、速度、模型权重与计算开销等多个维度上取得良好平衡。此外，相关研究更普遍缺乏对模型在复杂自然背景（如光照变化）下鲁棒性的系统评估，制约了技术在实际场景中的部署与应用。因此，探索一种能够同时兼顾高精度、高效率与强鲁棒性的检测模型，已成为推动该技术从实验室走向实际应用的迫切需求。为应对这一挑战，本研究聚焦于苹果这一重要经济作物，基于 YOLOv8n 提出一种轻量化、实时检测模型 SMPD-YOLO，通过 4 项核心改进构建了一个协同优化的轻量级检测模型，以期提升模型在复杂自然场景下的综合检测性能。

## 1　材料与方法

### 1.1　数据来源

本研究采用 3 个公开数据集与 1 个自主采集数据集，其中公开数据集包括 Plant Pathology 2020-FGVC7[16]、Plant Pathology 2021-FGVC8[17]和 PlantDoc[18]，共包含 4 种苹果叶片病害：灰斑病、白粉病、锈病和黑星病。为增强模型的泛化能力与对复杂环境的适应性，本研究自主采集了苹果叶片病害图像 1 002 张，采集于河南省荥阳市王村镇国盛苹果生态园，拍摄时间为 2024 年 5—8 月（08:00—18:00），拍摄设备为 Redmi k40pro，距离叶片 10～50 cm。针对果园实际种植场景中病害检测的复杂性，本研究挑选不同光照条件、复杂背景、病害程度、发病阶段及拍摄角度的病害图像，构建了包含 2 128 张苹果叶片病害图像的数据集，其中灰斑病 501 张、锈病 509 张、白粉病 603 张、黑星病 515 张。

### 1.2　图像预处理

根据不同病害的视觉形态学特征，本研究对数据集标注策略进行了针对性设计，旨在使标注结果更符合病理学事实与实际诊断逻辑，具体如下：

对于锈病与灰斑病，其病斑通常与健康叶片组织边界清晰、轮廓突出，且分布相对独立。对此类病害，本研究采用病斑级标注，即为每一个可分辨的独立病斑绘制边界框。这种精细化的标注有助于模型

学习定位和识别具体的病斑实例。对于白粉病与黑星病，其病斑在早期或中期常表现为弥散性分布，病斑与健康组织间边界模糊、对比度低，致使难以精确分割单个病斑。若进行病斑级标注，则会引入显著的主观误差。因此，本研究对此类病害采用叶片级标注，即标注整个患病叶片区域。该策略与农业专家在实际诊断中通观全叶、依据整体症状进行判断的逻辑保持一致，能为模型提供更可靠的全局特征监督信号。

所有图像均使用LabelImg软件进行标注，标注文件保存为‘txt’格式，共8 608个患病区域标注。按8∶2的比例将标注的数据集随机划分为训练集（1 703张图像）和测试集（425张图像）。此外，本研究对测试集进行了亮度增强、亮度减弱以及图像模糊处理，用于评估模型的鲁棒性，分别生成425张图像，总计1 275张，具体示例如图1所示。每种病害类别的标注示例如图2所示，数据集详细信息见表1。值得注意的是，由于采用了与病害形态相适应的分级标注策略，不同类别间的标注实例数量存在显著差异，这种分布真实地反映了不同病害在自然条件下的表现形式，确保了标注的准确性与科学性。




图1 不同噪声条件下的图像示例

Fig. 1 Image examples of different noise conditions

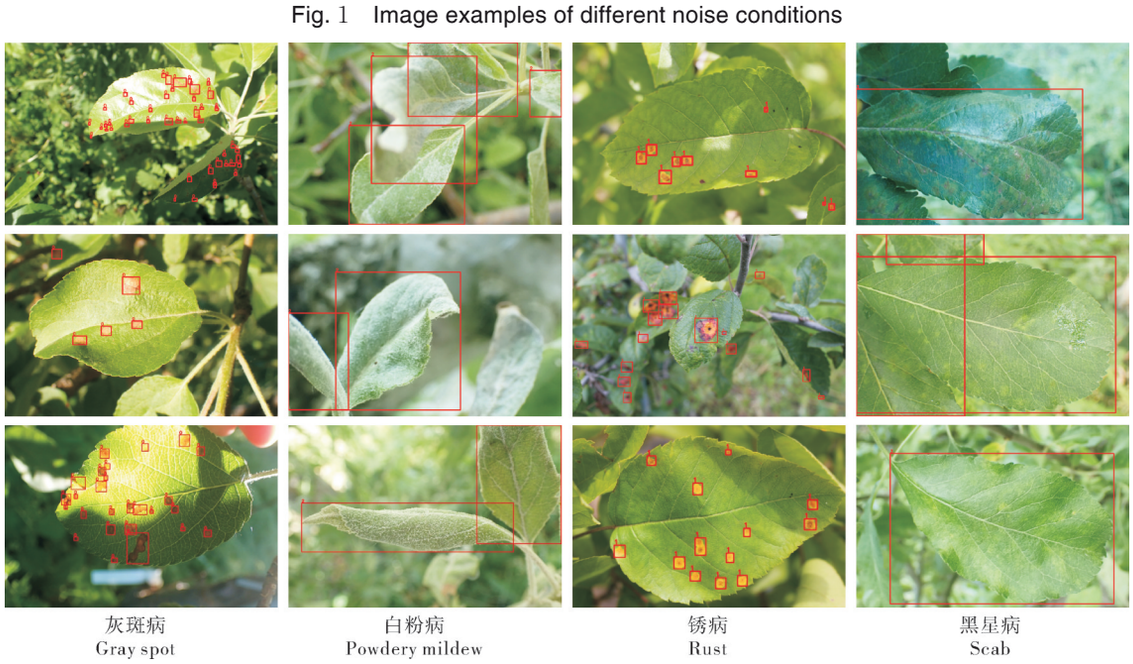



图2 病害图像的标注示例

Fig. 2 Example of annotated disease images

为提升模型鲁棒性与泛化能力，避免过拟合，本研究采用了在线数据增强与离线数据增强相结合的策略。在线数据增强可在训练过程中动态生成样本，每轮迭代均提供不同的图像变化，有效防止过拟合现象。离线数据增强则在训练前对原始数据进行扩展，通过生成大量增强图像丰富训练集。本研究采用随机翻转、随机旋转、颜色抖动、高斯模糊以及马赛克增强等数据增强操作，图3展示了部分增强图像样本示例。

### 1.3 YOLOv8网络

YOLOv8是YOLO系列中的前沿目标检测算法，标志着该领域的重要突破[19]。YOLOv8包括

表 1　数据集信息

Table 1　Dataset information

| 病害类型 Disease type | 标注策略 Labeling strategy | 图像数量 Number of images | 标注数量 Number of annotations | 增强后的训练集图像数量 Number of augmented training images | 增强后的标注数量 Number of annotations after augmentation | 测试集图像数量 Number of test images |
|---|---|---|---|---|---|---|
| 灰斑病 Gray spot | 病斑级 Lesion-level | 501 | 3 606 | 1 257 | 7 047 | 101 |
| 白粉病 Powdery mildew | 叶片级 Leaf-level | 603 | 1 173 | 1 300 | 2 528 | 121 |
| 锈病 Rust | 病斑级 Lesion-level | 509 | 3 196 | 1 277 | 7 018 | 101 |
| 黑星病 Scab | 叶片级 Leaf-level | 515 | 633 | 1 275 | 1 567 | 102 |
| 总计 Total | | 2 128 | 8 608 | 5 109 | 18 160 | 425 |

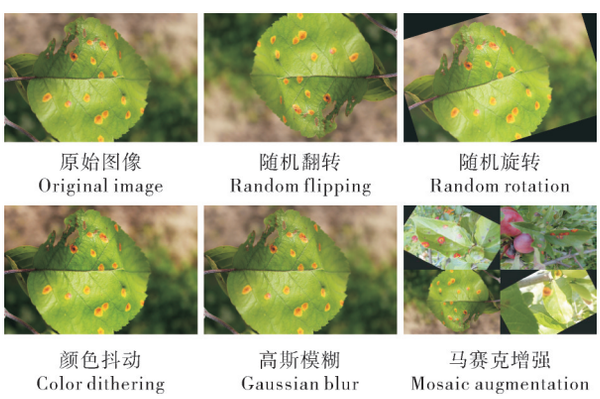



图 3　图像数据增强示例

Fig. 3　Examples of augmented images

5 种不同大小的版本：YOLOv8n、YOLOv8s、YOLOv8m、YOLOv8l 和 YOLOv8x。随着模型规模的增大，网络深度和检测精度逐步提升。其中，YOLOv8n 具有紧凑的模型结构和较浅的网络深度，在保证较高检测精度的同时，满足轻量级部署与实时检测需求。因此，本研究选择 YOLOv8n 作为基线模型进行优化。

YOLOv8 由 4 个核心模块组成：输入（Input）、骨干网络（Backbone）、颈部网络（Neck）和预测头（Head）[20]。首先，输入模块负责对图像进行预处理，包括图像缩放、归一化和数据增强。骨干网络提取输入 RGB 图像中的特征，是模型的核心部分[21]。颈部网络位于骨干网络和预测头之间，采用特征金字塔网络（feature pyramid networks，FPN）[22]和路径聚合网络（path aggregation network，PAN）[23]进行多层次特征融合，从而增强对不同尺度目标的检测能力[24]。预测头则负责最终的目标识别与定位，利用多个检测层对不同尺寸的目标进行预测。YOLOv8 的 C2f 模块借鉴了 YOLOv7[25]中 ELAN 结构，通过优化梯度路径设计，增强了模型对丰富梯度流信息的捕获能力。此外，YOLOv8 采用 SPPF（spatial pyramid pooling fast）模块，对 SPP[26]结构进行了优化，通过减少网络层数提高特征融合效率。同时，该模型采用无锚点（anchor-free）检测机制，并使用任务对齐分配器进行正负样本匹配，结合多尺度下采样特征进行目标预测，从而在提升检测精度的同时实现轻量化[27]。

### 1.4　轻量化苹果叶片病害的实时检测模型

苹果叶片病害检测面临模型体积大、计算成本高以及检测性能不佳等问题。针对这些问题，本研究提出了一种轻量化、实时的检测模型 SMPD-YOLO，其构建流程如图 4 所示。首先，将 YOLOv8n 中的 SPPF 模块替换为 SPPFCSPC 模块，引入额外的卷积与级联结构，显著提升模型的多特征融合能力，提高了检测精度。其次，采用 MPD-IoU 损失函数作为边界框回归损失函数，以增强目标尺寸变化的适应性，从而实现模型收敛速度和检测准确率的提升。然后，使用 LAMP 剪枝策略，剪除模型冗余的参数和权重，在保证较高模型精度的同时，实现模型规模的大幅缩减并提升检测速度。最后，在第 12、15、18、21 特征层上应用通道级知识蒸馏，进一步优化模型的检测性能。

### 1.5　SPPFCSPC 模块

在苹果叶片病害检测任务中，由于传统卷积神经网络通常要求输入图像为固定尺寸，往往需要对原始图像进行裁剪或缩放等预处理操作。此类操作极易导致图像失真，尤其是改变病斑的原始形态与比例，从而直接影响检测模型的精度。为解决这一问题，支持任意尺寸输入的空间金字塔方法被提出，如 SPP、SPPF、ASPP[28]、SPPCSPC[29]、SP-

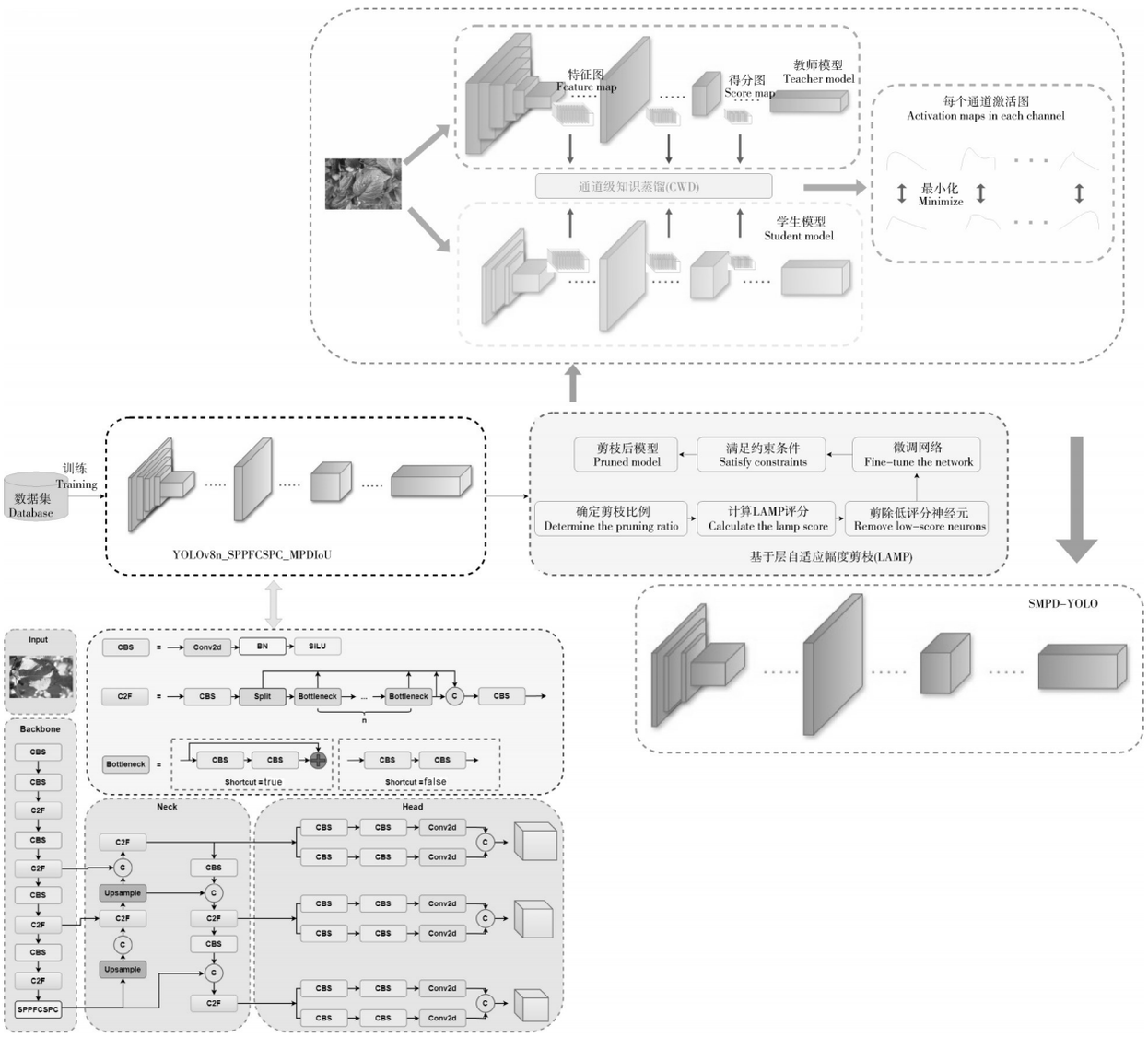



图4 SMPD-YOLO模型结构设计图

Fig. 4 Architecture of the SMPD-YOLO model

PFELAN、SPPFCSPC[30]等。此类结构能够有效避免预处理造成的信息丢失，并实现高效的多尺度特征融合。针对苹果叶片病斑在形态、尺寸及边界清晰度上的显著差异，SPPFCSPC模块展现出独特优势。该模块融合了SPPF的串行快速池化与SPPCSPC的特征精简能力，其核心机理在于：通过串行多尺度池化构建多重感受野，使网络能够同步捕捉从微小病斑点到大病斑区域的所有尺度特征，直接应对叶片病害的尺度多样性挑战；同时借助跨阶段部分连接结构，在减少特征冗余的同时，充分融合浅层纹理细节与深层抽象语义，显著增强了对边界模糊病斑的特征表达能力。如图5所示，该模块采用3个独立的最大池化层，在保持感受野的同时优化计算效率；较小池化核减少计算量，较大池化核捕获关键信息，从而在加速计算的同时保证苹果叶片病害检测精度。其计算过程如下公式所示：

$$S_1(F)=\mathrm{MaxPool}_{k=5}^{p=2}(F) \tag{1}$$

$$S_2(S_1)=\mathrm{MaxPool}_{k=5}^{p=2}(S_1) \tag{2}$$

$$S_3(S_2)=\mathrm{MaxPool}_{k=5}^{p=2}(S_2) \tag{3}$$

$$R=S_1 \circledast S_2 \circledast S_3 \tag{4}$$

其中，$F$表示输入特征层，$S_1$、$S_2$和$S_3$分别表示最小、中等和最大池化核的结果，$R$为最终输出，$\circledast$表示张量拼接。

### 1.6 MPD-IoU损失函数

YOLOv8采用的CIoU[31](complete IoU)损失函数在综合考虑边界框重叠度和长宽比的基础上，有效提升目标定位精度。然而，其惩罚项中的长宽比度量在预测框与真实框长宽比相近但绝对尺寸差异

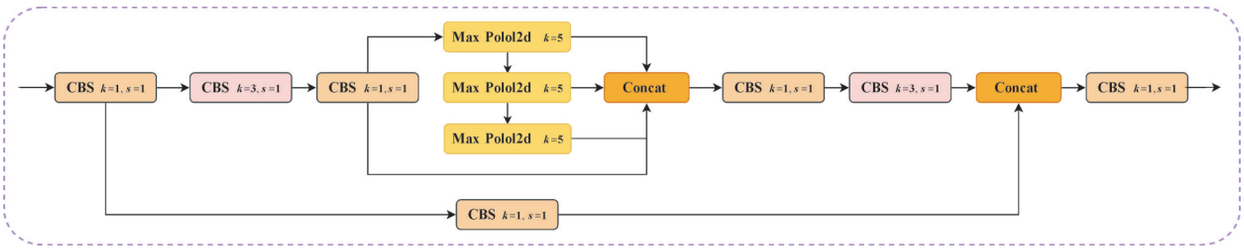



图 5　SPPFCSPC 模块结构图

Fig. 5　Structure of the SPPFCSPC module

较大时所提供的梯度信息会减弱，这一局限性导致其在微小病斑的精准定位上表现不佳，可能导致模型收敛速度下降与检测精度损失。在苹果叶片病害检测中，不同病害的形态与尺度差异显著（例如白粉病病斑与锈病病斑），对边界框回归损失函数的多尺度适应能力提出了更高要求。为此，本研究引入 MPD-IoU[32]损失函数。该函数的核心原理在于直接最小化预测框与真实框中心点距离与宽高尺寸的差异，通过一种更为简洁统一的度量方式，规避了 CIoU 在特定尺度下的优化困境。MPD-IoU 不仅保留了 IoU 对重叠度的评估优势，更重要的是，其通过独立且明确地优化中心点坐标与宽高参数，为模型提供了更直接、更稳定的梯度方向。这一特性使其能够平等而有效地处理不同尺度的病斑：无论是对于中心点定位极其敏感的小病斑，还是对宽高尺寸准确性要求高的大病斑，MPD-IoU 都能实现更精准的引导。图 6 展示了 MPD-IoU 在预测边界框（红色）与真实边界框（黄色）之间的对比情况，其具体计算公式如下：

$$d_1^2=(x_1^{\mathrm{prd}}-x_1^{\mathrm{gt}})^2+(y_1^{\mathrm{prd}}-y_1^{\mathrm{gt}})^2 \tag{5}$$

$$d_2^2=(x_2^{\mathrm{prd}}-x_2^{\mathrm{gt}})^2+(y_2^{\mathrm{prd}}-y_2^{\mathrm{gt}})^2 \tag{6}$$

$$w_{\mathrm{gt}}=x_2^{\mathrm{gt}}-x_1^{\mathrm{gt}},h_{\mathrm{gt}}=y_2^{\mathrm{gt}}-y_1^{\mathrm{gt}},$$

$$w_{\mathrm{prd}}=x_2^{\mathrm{prd}}-x_1^{\mathrm{prd}},h_{\mathrm{prd}}=y_2^{\mathrm{prd}}-y_1^{\mathrm{prd}} \tag{7}$$

$$J_{\mathrm{MPDIoU}}=J_{\mathrm{IoU}}-\frac{d_1^2}{h_{\mathrm{gt}}^2+w_{\mathrm{gt}}^2}-\frac{d_2^2}{h_{\mathrm{prd}}^2+w_{\mathrm{prd}}^2} \tag{8}$$

$$L_{\mathrm{MPDIoU}}=1-J_{\mathrm{MPDIoU}} \tag{9}$$

其中，$(x_c^{\mathrm{gt}},y_c^{\mathrm{gt}})$和$(x_c^{\mathrm{prd}},y_c^{\mathrm{prd}})$分别为真实边界框和预测边界框的中心点坐标；$w_{\mathrm{gt}}$和$h_{\mathrm{gt}}$分别为真实边界框的宽度和高度，而$w_{\mathrm{prd}}$和$h_{\mathrm{prd}}$则分别表示预测边界框的宽度和高度。

### 1.7　基于层自适应幅度剪枝

深度神经网络中通常存在大量冗余参数和权重，导致计算资源浪费与推理延迟[33-34]。为实现苹果叶片病害的实时检测，本研究采用基于层自适应幅度剪枝（LAMP）[35]方法。该方法并非无差别地剪

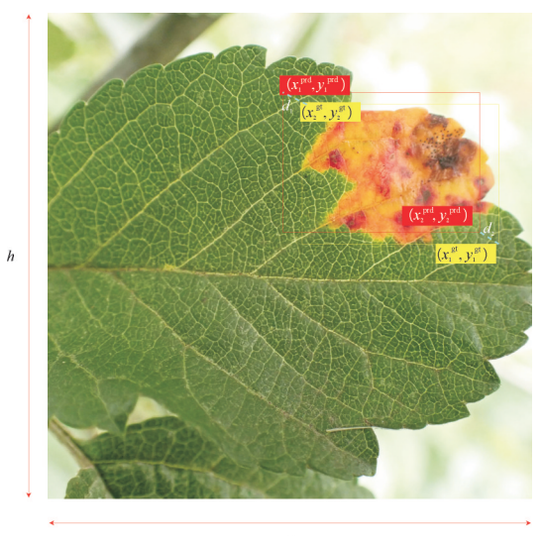



图 6　MPD-IoU 的计算原理

Fig. 6　Schematic calculation of MPDIoU

枝，而是依据其核心的 LAMP 评分，精准识别并剪除对叶片病害特征贡献微弱的冗余权重，从而在大幅减少模型参数量与计算开销的同时，最大限度地保留对多尺度、多形态病斑的关键判别能力。图 7 展示了模型剪枝前后通道数量变化情况。基于层自适应幅度剪枝方法主要包括以下步骤：

1）确定剪枝比例。首先设定一个全局剪枝目标，随后根据网络中每一层对叶片病害特征的敏感度与重要性（例如，负责捕捉细微纹理的浅层卷积与负责抽象语义的深层卷积具有不同的容剪枝能力），动态分配各层的具体剪枝比例，实现结构化压缩。

2）剪枝决策。计算每个权重的 LAMP 评分［如公式（10）所示］，并根据权重的绝对值排序。按照每层的剪枝比例，从排序列表中选择需剪除的权重，以确保剪枝操作的合理性和有效性。这种策略能系统性地消除冗余连接，同时保护那些对识别典型病斑（如白粉病的粉状纹理、锈病的斑点轮廓）至关重要的特征通路。

3）微调。剪枝操作不可避免地会改变模型原有

的结构。因此，必须在剪枝后使用苹果叶片病害数据集对模型进行微调，以重新校准剩余权重，恢复其检测性能，确保其对各类病害的识别精度不出现显著下降。

通过基于层自适应幅度剪枝比例的调整和基于权重评分的剪枝策略，在减少参数量和计算成本的同时，仍能维持模型的检测性能。其中权重评分的计算公式如下：

$$\mathrm{score}(a;w)=\frac{(w_{|a|})^2}{\sum_{b\geqslant a}(w_{|b|})^2} \tag{10}$$

该公式表示权重张量$w$关于指标$a$的LAMP评分。其中，$w$表示权重张量，$w_{|a|}$和$w_{|b|}$分别表示与索引映射相关的项。索引$a$和$b$对应于与权重张量相关的升序索引映射。

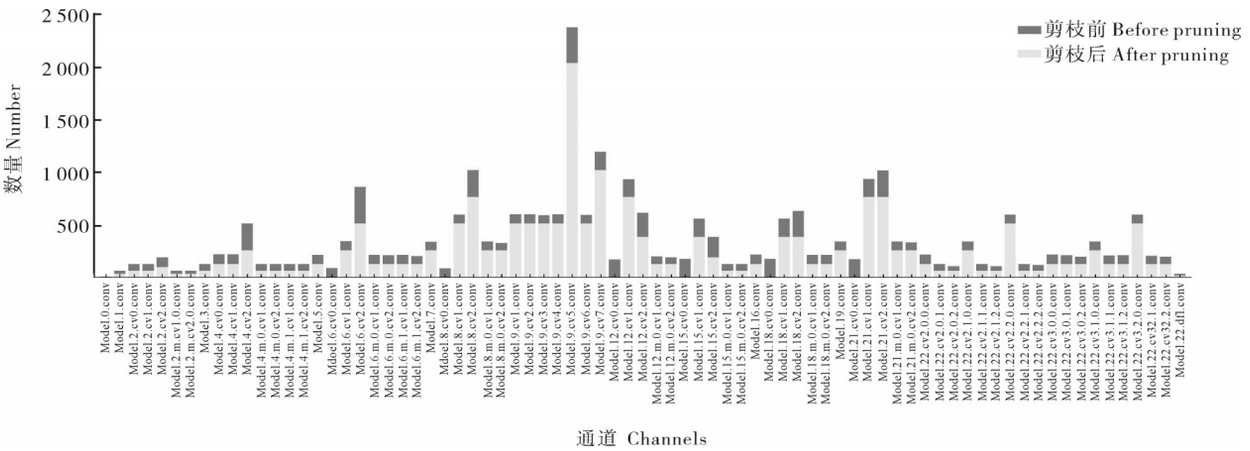



**图7 剪枝前后的通道数量**

Fig. 7 Number of channels before and after pruning

### 1.8 通道级知识蒸馏

剪枝策略虽能大幅缩减模型尺寸与计算量，但不可避免地会削弱模型的识别能力，尤其可能损害其对形态多样、尺度各异的叶片病斑细微特征的捕捉能力。为精准弥补这一性能损失，本研究在剪枝后引入知识蒸馏框架，将剪枝得到的轻量模型作为学生模型，以性能完备的YOLOv8s作为教师模型，进行特征层面的引导与强化。

知识蒸馏(knowledge distillation)是一种有效提升轻量化目标检测算法性能的方法。通过让较小的学生网络学习较大的教师网络的知识，使得学生模型的输出能够尽可能接近教师模型，从而在不改变计算效率的前提下，提升检测性能。传统的知识蒸馏方法，如Fitnets[36]，主要应用于图像分类任务，即针对整幅图像进行全局预测。然而，目标检测为密集预测任务，需要对每个像素或小区域进行预测，其任务特点与图像分类存在显著差异。在密集预测任务中，现有方法通常通过在空间维度对齐教师与学生网络的激活图进行知识蒸馏，这种方式可能引入冗余信息，限制了学生模型的性能提升。为解决这一问题，Shu等[37]提出了通道级知识蒸馏(CWD)方法。与基于空间对齐的方法不同，CWD将各通道的激活值归一化为概率分布，并通过最小化学生模型与教师模型之间的Kullback-Leibler(KL)散度，实现了高效知识迁移。这种通道级的匹配机制，其优势在于它能让学生模型自动继承教师网络对病害敏感通道的先验知识。在苹果叶片病害任务中，这意味着学生模型能更专注于学习那些专门用于响应特定病斑纹理(如白粉病的粉状特征)、颜色异常(如灰斑病的褐色斑块)或边缘变化的特征通道，同时忽略与健康叶片背景相关的冗余通道。通过这种方式，CWD能有效增强轻量化学生模型对于小型病斑和边界模糊区域的特征学习能力，从而在保持高计算效率的同时，显著提升检测精度与模型鲁棒性，确保轻量化模型在实际复杂环境中依然可靠。

设教师网络和学生网络分别为t和s，其对应的激活值分别记为$y^t$和$y^s$，则CWD的损失函数表达如下：

$$\varphi(\phi(y^t),\phi(y^s))=\varphi(\phi(y_c^t),\phi(y_c^s)) \tag{11}$$

其中，$\phi(\cdot)$用于将激活值转换为概率分布，计算公式如下：

$$\phi(y_c)=\frac{\exp(y_{c,i}/T)}{\sum_{i=1}^{w\times H}\exp(y_{c,i}/T)} \tag{12}$$

$c=1,2,\cdots\cdots,n$，$C$代表通道索引，$i$表示空间坐标，$T$为温度参数。较高的$T$会使概率分布更平滑，从而使得模型在每个通道上关注更广泛的空间区

域。通过Softmax归一化，能够消除教师网络与学生网络之间的规模差异影响，而KL散度可以用于衡量教师与学生网络在通道分布上的差异，其计算公式如下：

$$\varphi(y^{t},y^{s})=\frac{T^{2}}{C}\Sigma_{c=1}^{C}\Sigma_{i=1}^{w\times H}\phi(y_{c,i}^{t})\log(\frac{\phi(y_{c,i}^{t})}{\phi(y_{c,i}^{s})}) \quad (13)$$

KL散度是一种非对称度量，能够促使学生网络更有效地学习教师网络的关键特征区域，从而优化密集预测任务中的蒸馏过程。

### 1.9　实验环境、参数配置与训练策略

实验硬件环境如下：操作系统为Windows 11，内存为16 GB，GPU为NVIDIA GeForce RTX 4060（8 GB），CPU为13代Intel Core i9-13900HX。实验软件环境为：Pytorch深度学习框架、Python 3.10、CUDA 11.8和cuDNN 8.5.0。模型训练参数详见表2。

由于采用了分级标注策略，数据集中各类别的标注实例数量存在显著差异。类别不平衡问题可能导致模型训练过程偏向于实例数量多的类别（如灰斑病）。为确保模型对所有病害类别都具有均衡识别能力，本研究在训练过程中采用了类别平衡采样器。根据各病害类别数量计算样本权重，创建自定义采样器，并定义加权损失，以确保在每个训练批次中，来自每一个类别的样本都能以近似相等的概率被选中，避免了模型优化过程被多数类样本所主导。

**表2　训练参数设置**

**Table 2　Training parameter settings**

| 训练参数 Training parameter | 数值 Value |
|---|---|
| Batch | 16 |
| Epoch | 250 |
| Patience | 50 |
| Workers | 8 |
| Imgsz | 640 |
| Close_mosaic | 10 |
| Momentum | 0.937 |
| Weight_decay | 0.000 5 |

### 1.10　评价指标

本研究采用精确率（precision，$P$）、召回率（recall，$R$）、$F_1$-score、平均精度（average precision，AP，公式中记作$P_A$）、平均精度均值（mean average precision，mAP，公式中记作$P_{mA}$）、模型权重（model size）、浮点运算速度（floating point operations per second，FLOPs）和帧率（frames per second，FPS）对模型性能进行综合评估。其中，$F_1$-score为精确率和召回率的调和平均值，而平均精度基于精确率-召回率（precision-recall，PR）曲线计算，表示PR曲线下的面积，评估单类病害检测性能。具体而言，mAP@0.5表示在IoU阈值设定为0.5时计算的平均精度，而mAP@0.5：0.95表示在IoU阈值从0.5到0.95（步长为0.05）范围内计算所得的平均精度均值。FPS（公式中记作$F_{PS}$）反映模型的推理速度与实时性。评价指标计算公式如下：

$$P=\frac{T_P}{T_P+F_P} \quad (14)$$

$$R=\frac{T_P}{T_P+F_N} \quad (15)$$

$$F_1=\frac{2\times P\times R}{P+R} \quad (16)$$

$$P_A=\int_0^1 P(R)\mathrm{d}R \quad (17)$$

$$P_{mA}=\sum_{i=1}^{n}P_{A_i} \quad (18)$$

$$F_{PS}=\frac{N}{T} \quad (19)$$

式（14）～（19）中：$T_P$、$F_P$、$F_N$分别为正确检测的正样本数、错误检测的正样本数和错误检测的负样本数；$n$代表测试集中的类别数。$N$为处理的帧数，$T$为总处理时间。

## 2　结果与分析

### 2.1　SMPD-YOLO与基线模型的检测性能对比

在训练过程中，基线模型与SMPD-YOLO模型采用相同的超参数，以确保对比结果的公平性。由表3可知，SMPD-YOLO模型在mAP@0.5上较YOLOv8n提升1.80百分点，并在所有病害类别中均取得更优的检测性能。在灰斑病、锈病、白粉病和黑星病的检测中的$P_A$值分别提升了1.80、0.20、3.50和1.60百分点。

**表3　YOLOv8n与SMPD-YOLO的检测性能**

**Table 3　Comparison of YOLOv8n and SMPD-YOLO**　%

| 模型 Models | $P_A$ | | | | mAP@0.5 |
|---|---|---|---|---|---|
| | 灰斑病 Gray spot | 锈病 Rust | 白粉病 Powdery mildew | 黑星病 Scab | |
| YOLOv8n | 88.00 | 90.10 | 83.20 | 92.50 | 88.40 |
| SMPD-YOLO | 89.80 | 90.30 | 86.70 | 94.10 | 90.20 |

由图8可知，SMPD-YOLO模型（蓝色曲线）在召回率、mAP@0.5和mAP@0.5：0.95上始终优于基线模型（红色曲线），同时保持了相近的精度。此外，该模型在整个训练过程中表现出更高的稳定性和更

优的整体性能。尽管2种模型都能较好地收敛，但SMPD-YOLO在训练早期（前50轮）表现出更高的稳定性，而基线模型在初期阶段表现出明显波动。

PR曲线和$F_1$-confidence曲线进一步验证了SMPD-YOLO在不同IoU阈值下的优越性，尤其是在高召回率和高置信度场景下（图9）。此外，与YOLOv8n相比，SMPD-YOLO模型生成的Grad-CAM热力图能够更准确地突出各类病害的关键特征（图10），表明其在目标识别能力方面更具优势。

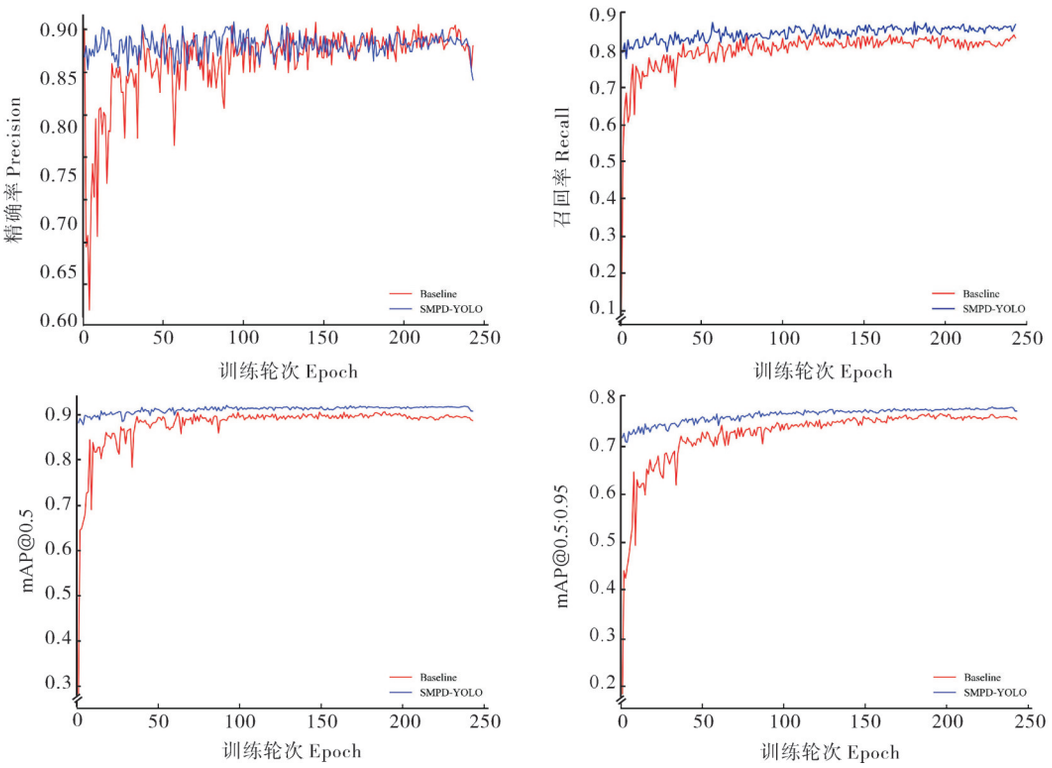



图8 YOLOv8n与SMPD-YOLO训练曲线

Fig. 8 Training curves of YOLOv8n and SMPD-YOLO

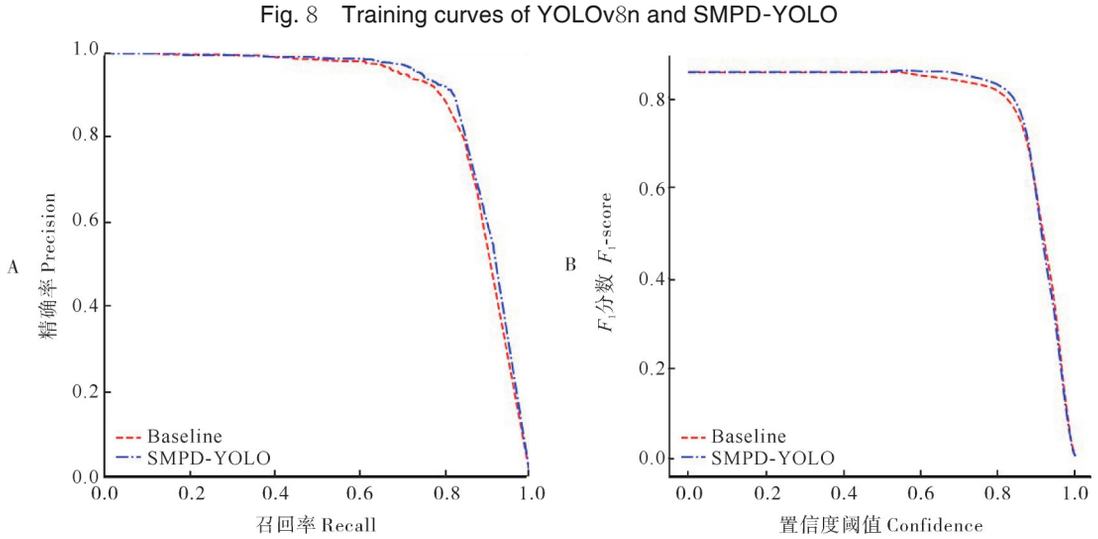



图9 YOLOv8n与SMPD-YOLO的PR（A）和$F_1$-confidence（B）曲线

Fig. 9 Precision-recall (A) and $F_1$-confidence (B) curves of YOLOv8n and SMPD-YOLO

## 2.2 消融实验

由表4可知，SPPFCSPC模块和MPD-IoU损失函数显著提升了模型的mAP@0.5。SPPFCSPC模块通过多尺度池化和特征融合增强了模型捕捉更丰富层次特征的能力；MPD-IoU损失函数则改善了边界框的对齐性；采用LAMP剪枝方法，剪除冗余的参

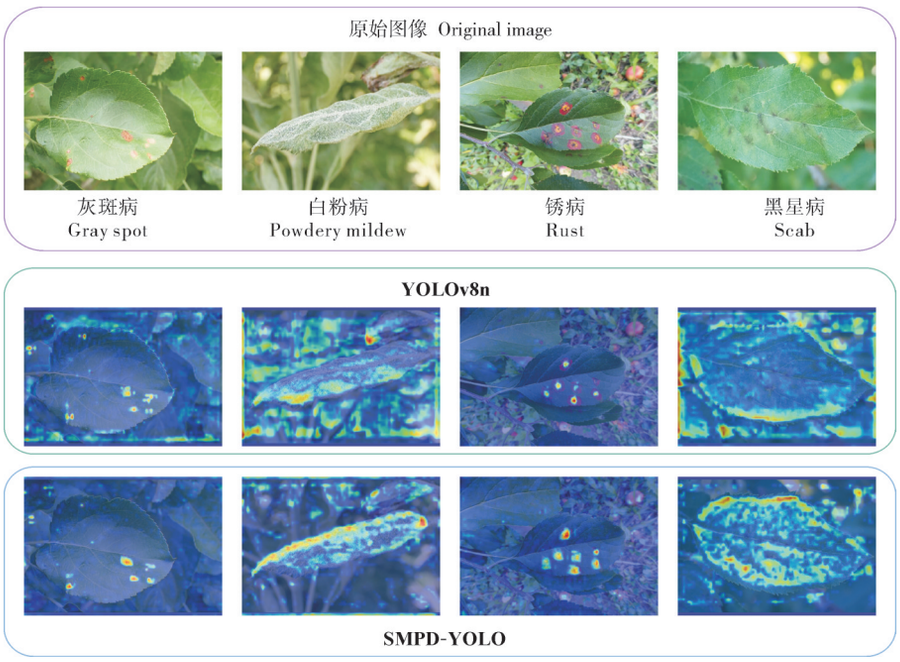



图 10　YOLOv8n 与 SMPD-YOLO 模型的热力图

Fig. 10　Comparison of heat maps of YOLOv8n and SMPD-YOLO models

数和权重，而后对模型进行微调，保持较高精度的同时，模型权重和参数量分别减少了原始模型的 20.64% 和 28.13%。将剪枝后模型作为学生模型，YOLOv8s 作为教师模型，采用通道级知识蒸馏进一步提升模型的召回率和 $F_1$-score，并使 SMPD-YOLO 的 mAP@0.5 提升至 90.20%，同时保持 5.0 MB 的模型权重。说明两者具有显著的互补与协同作用：剪枝负责模型结构压缩，而蒸馏通过通道注意力引导特征重建。

LAMP 剪除冗余通道，对模型特征表达能力造成非均衡的、结构性的破坏。模型的某些特征提取器被直接移除，改变了整个网络的数据流。保留下来的通道，其接收到的信息已经发生了变化，内部连接也被稀疏化，破坏了一些提取特定特征的通道间的合作，导致其提取的特征图变得粗糙、不完整，对细节和噪声的区分能力下降。CWD 让学生模型学习每个通道的概率分布，同教师模型关注图像中同样重要的语义信息；CWD 会促使学生模型保留的功能相近通道发生相应调整，并通过模仿其他通道方差分布，促使学生模型更高效地利用剩余、有限的通道容量；此外 CWD 中的 Softmax 归一化和 KL 散度，增强了教师模型中关键特征区域的指导作用，学生模型会优先学习这些关键区域的特征，从而自动忽略背景，提升了鲁棒性。

表 4　消融实验

**Table 4　Ablation experiments**

| SPPFCSPC | MPD-IoU | LAMP | CWD | 精确率/% Precision | 召回率/% Recall | $F_1$ 分数/% $F_1$-score | mAP@0.5/% | mAP@0.5:0.95/% | 浮点运算速度/($10^9$ $s^{-1}$) FLOPs | 模型权重/MB Model weights | 参数量/$10^6$ Parameters |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 88.80 | 82.40 | 85.48 | 88.40 | 77.90 | 8.1 | 6.3 | 3.2 |
| √ | | | | 88.30 | 83.50 | 85.83 | 89.10 | 78.40 | 18.3 | 35.4 | 17.0 |
| | √ | | | 90.20 | 82.10 | 85.96 | 88.80 | 78.40 | 8.1 | 6.3 | 3.2 |
| √ | √ | | | 90.60 | 82.40 | 86.31 | 89.20 | 77.80 | 18.3 | 35.4 | 17.0 |
| √ | √ | √ | | 88.50 | 82.30 | 85.29 | 88.30 | 77.50 | 7.3 | 5.0 | 2.3 |
| √ | √ | √ | √ | 88.30 | 86.20 | 87.24 | 90.20 | 79.50 | 7.3 | 5.0 | 2.3 |

## 2.3　SMPD-YOLO 与主流模型的检测性能对比

在相同的实验环境、超参数和数据集下，改进的 SMPD-YOLO 模型与多种主流目标检测模型进行性能对比（表 5）。Faster R-CNN 和 YOLOX-s 的

mAP@0.5和$F_1$-score均未超过70%，同时模型权重、参数量和FLOPs较大，且FPS较低，难以满足实时检测需求。尽管YOLOv8m在mAP@0.5上表现较好，但其在所有YOLO系列模型中FPS最低（42.4帧/s），且参数量和FLOPs较大；YOLOv5s的FPS相对较高，但在$F_1$-score、mAP@0.5、模型权重、参数量和FLOPs多个指标上均不及SMPD-YOLO；YOLOv8s在所有模型中$F_1$-score最高，但其FPS较低，且参数量、模型权重和FLOPs较大；YOLOv7-tiny的FPS达131.6帧/s，mAP@0.5和$F_1$-score略高于YOLOv8n，但其参数量、模型权重约是YOLOv8n的2倍；YOLOv9t和YOLOv10n的mAP@0.5最高，YOLOv11n、YOLOv12n、YOLOv13n的FLOPs最低，但在其他性能均不及SMPD-YOLO模型。相比之下，改进的SMPD-YOLO模型在FPS（133.3帧/s）、mAP@0.5（90.20%）、模型权重（5.0 MB）、参数量（$2.3\times10^6$）和FLOPs（$7.3\times10^9$ $s^{-1}$）之间实现了最佳平衡，表现出更优的实时性与综合性能。

**表5 不同检测模型性能对比**

**Table 5 Comparison of performance of different detection models**

| 模型 Models | $F_1$-score/% | mAP@0.5/% | 模型权重/MB Model weights | 参数量/$10^6$ Parameters | 浮点运算速度/$10^9$ $s^{-1}$ FLOPs | 帧率/(帧/s) FPS |
|---|---|---|---|---|---|---|
| Faster R-CNN | 49.02 | 64.50 | 108.0 | 41.5 | 386.3 | 11.0 |
| YOLOv5s | 85.81 | 89.30 | 14.4 | 7.0 | 15.8 | 111.1 |
| YOLOX-s | 64.60 | 85.10 | 68.5 | 8.9 | 26.8 | 31.2 |
| YOLOv7-tiny | 86.21 | 90.00 | 12.3 | 6.0 | 13.0 | 131.6 |
| YOLOv8n | 85.48 | 88.40 | 6.3 | 3.2 | 8.1 | 130.1 |
| YOLOv8s | 87.95 | 90.20 | 22.5 | 11.0 | 28.4 | 113.2 |
| YOLOv8m | 86.99 | 89.60 | 52.1 | 25.0 | 78.7 | 42.4 |
| YOLOv9t | 85.73 | 91.00 | 6.1 | 2.6 | 10.7 | 84.75 |
| YOLOv10n | 85.39 | 91.00 | 5.8 | 2.7 | 8.2 | 113.0 |
| YOLOv11n | 85.60 | 89.00 | 5.5 | 2.6 | 6.3 | 113.9 |
| YOLOv12n | 86.76 | 89.60 | 5.5 | 2.6 | 6.3 | 101.4 |
| YOLOv13n | 85.89 | 89.20 | 5.4 | 2.5 | 6.2 | 104.0 |
| SMPD-YOLO | 87.24 | 90.20 | 5.0 | 2.3 | 7.3 | 133.3 |

由图11可知，在强光下，SMPD-YOLO检测到的灰斑病病变数量最多，而其他模型存在较高的漏检率；在自然光照条件下对白粉病的检测中，其他模型的漏检率仍然较高；在锈病检测任务中，各模型表现相近，漏检率相当；而在黑星病检测中，YOLOv7-tiny漏检率较高。SMPD-YOLO在处理复杂检测场景（如灰斑病和白粉病等细微纹理变化）时表现出明显优势，这主要得益于SPPFCSPC模块的集成及MPD-IoU损失函数的应用，这两者共同促进了多尺度特征融合并加速了收敛。此外，LAMP方法在保持较高检测精度的同时减少了计算成本，通道级知识蒸馏通过关注目标相关区域进一步增强了模型特征提取能力。

## 2.4 不同特征融合方法与损失函数对比

为验证SPPFCSPC特征融合方法与MPD-IoU边界框回归损失函数在YOLOv8n苹果叶片病害检测中的有效性，分别对比了不同特征融合方法与边界框回归损失函数的性能。由表6可知，在不同融合方法中，YOLOv8n的SPPF模块在精确率上优于SPPFELAN、SPPCSPC与SPPFCSPC。SPPFCSPC虽在精确率上略低于SPPF，但其召回率、mAP@0.5均优于其他特征融合方法，且FLOPs较低。SPPF仅通过多尺度池化扩展感受野，但深浅层特征交互有限；SPPFELAN引入多层特征叠加，增强了语义信息，但跨尺度融合能力仍有限；SPPCSPC结合SPP与CSP分支，改善了多尺度特征表达与梯度流动。相比之下，SPPFCSPC通过CSP分支与快速卷积路径实现跨层、跨尺度的充分特征融合，既保留了浅层纹理信息，又提取了深层语义特征，因此，SPPFCSPC在苹果病害检测中性能更佳。

由表7可知，在边界框回归方面，YOLOv8n的CIoU损失考虑了中心距离、宽高比及重叠度，但在不规则小目标定位上仍存在偏差，因此精确率略低；Wise-IoU优化梯度分布[38]，Focaler-IoU强调难检测样本[39]，但两者对边界模糊、大小差异显著的病斑表现不足，因此检测性能较差。相比之下，MPD-IoU通过多维惩罚策略综合约束重叠度、中心距离、宽高比及方向匹配，显著增强了多尺度病斑的定位能力，




黄色圆圈代表漏检。Yellow circle represents missed detections.

**图 11　不同模型病害检测结果**

Fig. 11　Detection results of different models

从而使模型对多尺度苹果叶片病害检测中定位更准确、漏检率更低，整体性能上优于其他损失函数。

**表 6　不同特征融合方法对比结果**

**Table 6　Comparison results of different feature fusion methods**　%

| 模型 Model | 精确率 Precision | 召回率 Recall | mAP@0.5 |
|---|---|---|---|
| YOLOv8n_SPPF | 88.80 | 82.40 | 88.40 |
| YOLOv8n_SP-PFELAN | 88.10 | 82.30 | 87.70 |
| YOLOv8n_SPPC-SPC | 87.30 | 82.10 | 88.20 |
| YOLOv8n_SPPF-CPSC | 88.30 | 83.50 | 89.10 |

**表 7　不同边界框回归损失函数对比结果**

**Table 7　Comparison results of different bounding box regression loss functions**　%

| 模型 Model | 精确率 Precision | 召回率 Recall | mAP@0.5 |
|---|---|---|---|
| YOLOv8n_CIoU | 88.80 | 82.40 | 88.40 |
| YOLOv8n_Wise-IoU | 88.30 | 82.00 | 87.50 |
| YOLOv8n_Focaler-IoU | 88.50 | 81.70 | 88.00 |
| YOLOv8n_MPD-IoU | 90.20 | 82.10 | 88.80 |

## 2.5　LAMP 对比

为探究 LAMP 方法在不同剪枝加速比下对模型权重的影响，本研究开展了对比试验。其中，加速比指剪枝前后模型计算成本之间的比值。由表 8 可知，随着加速比的增加，模型权重逐渐减小。加速比为 2.0 时，模型权重缩小至 7.2 MB（大于 YOLOv8n），但检测性能略有下降。随着加速比的进一步增加，剪枝后的模型权重均小于 YOLOv8n，同时 FPS 持续提升。当加速比为 3.0 时，模型达到最佳的 $F_1$-score 和 mAP@0.5。

随后，将通道级知识蒸馏应用于所有剪枝后的模型（加速比 2.0 除外），教师模型均为 YOLOv8s。由表 9 可知，当加速比为 2.5 时，模型性能最佳，$F_1$-score 达到 87.24%，mAP@0.5 为 90.20%，mAP@0.5∶0.95 为 79.50%。而应用其他加速比的模型 $F_1$-score、mAP@0.5 和 mAP@0.5∶0.95 提升较少。

由图 12 可知，LAMP 剪枝模型（加速比 2.5）与 SMPD-YOLO 模型（剪枝后应用通道级知识蒸馏的模型）在训练期间的 mAP@0.5 始终优于基线模型。其中，SMPD-YOLO 模型在边界框回归损失的收敛速度方面表现更为出色，训练至第 100 轮时，其损失值已接近加速比为 4.0 模型的水平，最终实现所有模型中最低的回归损失。因此，对加速比为 2.5 的剪枝模型进行通道级知识蒸馏，有效实现了精度与损失之间的平衡。

表8 不同加速比对比试验结果

**Table 8 Comparison results of different speed-up ratios**

| 加速比 Speed-up ratios | $F_1$分数/% $F_1$-score | mAP@0.5/% | 模型权重/MB Model weights | 帧率/(帧/s) FPS |
|---|---|---|---|---|
| 2.0 | 85.44 | 88.30 | 7.2 | 128.8 |
| 2.5 | 85.29 | 88.30 | 5.0 | 133.3 |
| 3.0 | 85.77 | 89.00 | 3.8 | 135.2 |
| 3.5 | 85.62 | 88.70 | 3.0 | 135.6 |
| 4.0 | 85.34 | 88.20 | 2.5 | 136.6 |
| 4.5 | 84.77 | 88.00 | 2.2 | 140.1 |

表9 不同加速比蒸馏对比试验结果

**Table 9 Distillation results after pruning with different speed-up ratios**

| 蒸馏模型 Distillation models | $F_1$分数/% $F_1$-score | mAP@0.5/% | mAP@0.5:0.95/% | 模型权重/MB Model weights | 帧率/(帧/s) FPS |
|---|---|---|---|---|---|
| 2.5_CWD | 87.24 | 90.20 | 79.50 | 5.0 | 133.3 |
| 3.0_CWD | 85.48 | 89.60 | 79.00 | 3.8 | 138.2 |
| 3.5_CWD | 86.19 | 89.10 | 78.50 | 3.0 | 138.5 |
| 4.0_CWD | 85.57 | 88.90 | 78.10 | 2.5 | 137.3 |
| 4.5_CWD | 85.97 | 89.00 | 77.50 | 2.2 | 136.9 |

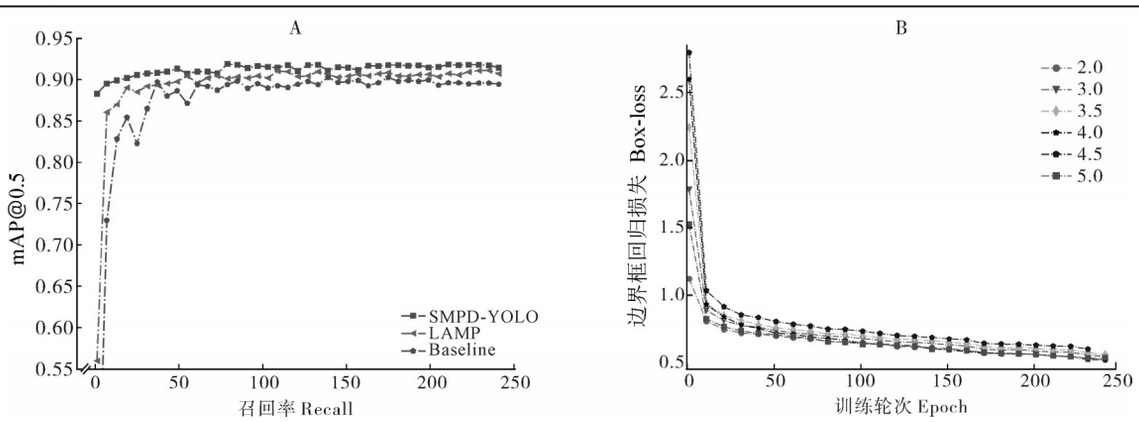



图12 不同模型的mAP@0.5(A)和边界框回归损失(B)曲线

Fig. 12 mAP@0.5 (A) and box-loss (B) curves of different models

## 2.6 知识蒸馏对比

实验采用通道级知识蒸馏(CWD),以YOLOv8s作为教师模型,并以加速比为2.5的剪枝模型作为学生模型。为验证CWD方法的优势,本研究将其与另外2种知识蒸馏方法进行对比,分别为MGD(matching guided distillation)[40]和Mimic[41],其中学生模型的剪枝加速比均为2.5。由表10可知,CWD方法在性能方面显著优于MGD和Mimic。Mimic方法未考虑通道间的特征关系,导致其在复杂的视觉检测任务中的表现较差;MGD方法依赖掩码策略,使学生模型在捕获全局信息方面受到一定限制。相比之下,CWD通过关注目标的关键特征并增强跨通道的信息传递,进而显著提升了模型的检测性能。综上,CWD能够更有效地引导学生模型学习教师模型的

表10 不同知识蒸馏对比实验结果

**Table 10 Comparison of different knowledge distillation methods**

| 蒸馏方法 Distillation methods | $F_1$-score/% | mAP@0.5/% | mAP@0.5:0.95/% | 模型权重/MB Model weights | 帧率/(帧/s) FPS |
|---|---|---|---|---|---|
| CWD | 87.24 | 90.20 | 79.50 | 5.0 | 133.3 |
| MGD | 86.21 | 89.00 | 78.40 | 5.0 | 130.0 |
| Mimic | 86.34 | 89.20 | 78.50 | 5.0 | 131.9 |

知识，从而实现更优的检测性能。

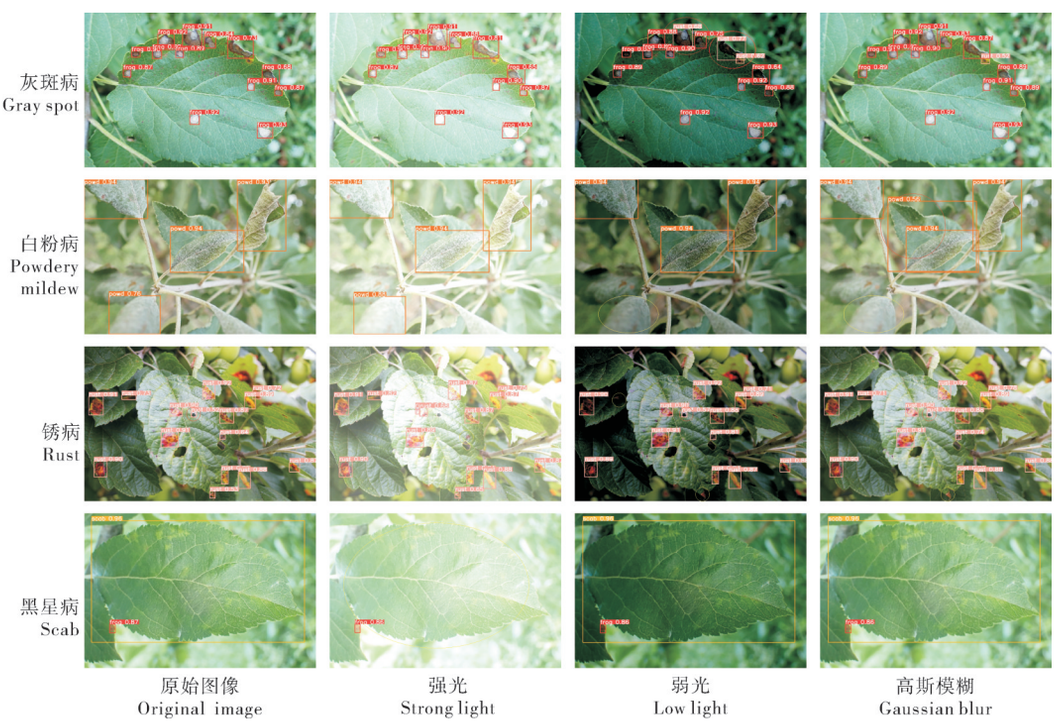
## 2.7　鲁棒性测试

在实际应用中，强光、弱光和图像模糊等环境噪声因素对苹果叶片病害的检测带来了严峻挑战。因此本研究对测试集图像进行处理，验证3种不同噪声条件下的鲁棒性。由表11和图13（黄色圆圈和红色圆圈处分别代表漏检和误检病斑）可知，在4类病害的检测中，SMPD-YOLO模型mAP@0.5均高于84%，即使在复杂环境干扰下，依然表现出优异的检测性能。具体来看，在强光条件下，由于强光会削弱病灶的颜色特征，导致锈病与黑星病的召回率在所有干扰条件中最低。在弱光环境中，灰斑病的病斑和白粉病的纹理特征不明显；而在图像模糊条件下，灰斑病与白粉病的病灶区域与背景对比度降低，轮廓边缘不清晰。因此，在弱光和高斯模糊干扰条件下，灰斑病与白粉病的召回率相对较低。总体而言，在所有干扰条件下，4种病害的平均召回率均高于77%，表明SMPD-YOLO具备较强的环境适应能力，在多种噪声干扰下依然能够稳定实现苹果叶片病害的精准检测，展现出良好的鲁棒性和应用前景。

表 11　不同噪声条件下的检测结果

**Table 11　Detection results under different noise conditions**　%

| 病害类型 Disease type | 召回率 Recall | | | mAP@0.5 | | |
|---|---|---|---|---|---|---|
| | 强光 Strong light | 弱光 Low light | 高斯模糊 Gaussian blur | 强光 Strong light | 弱光 Low light | 高斯模糊 Gaussian blur |
| 所有病害 All diseases | 77.60 | 81.00 | 82.70 | 85.40 | 87.40 | 89.00 |
| 灰斑病 Gray spot | 78.50 | 78.90 | 81.60 | 86.40 | 87.40 | 89.00 |
| 白粉病 Powdery mildew | 81.50 | 79.20 | 81.70 | 85.00 | 84.00 | 86.50 |
| 锈病 Rust | 74.70 | 80.00 | 84.60 | 84.90 | 87.00 | 90.50 |
| 黑星病 Scab | 75.80 | 85.90 | 82.80 | 85.30 | 90.70 | 90.00 |




黄色圆圈和红色圆圈处分别代表漏检和误检病斑。Yellow circle and the red circle respectively represent the missed detection and false detection of the diseased spots.

图 13　不同噪声条件下结果对比

Fig. 13　Comparison image of results under different noise conditions

## 2.8　不同硬件平台的帧率稳定性分析

为验证模型在不同硬件平台上的实时性能，本研究分别在GPU（GeForce RTX 4060，8 GB显存、NVIDIA GeForce RTX 3060，12 GB显存）与CPU（Intel Core i9-13900HX）环境下，对1 000张测试图像进行测试，计算平均帧率（FPS），并剔除前5帧

的预热时间以减少初始化影响。由表12可知，SMPD-YOLO模型在2类GPU上的平均FPS均超过100帧/s，且波动范围较小，表现出较高的推理效率与稳定性；在CPU环境下，平均FPS达到(30.1±1.6)帧/s，同样具备良好的实时处理能力与帧率稳定性。

表12 不同硬件平台的帧率对比

**Table 12 Comparison of frame rates across different hardware platforms** 帧/s

| 平台类型 Platform type | 处理器 Processor | 平均FPS Average FPS | 波动范围 Fluctuation range |
|---|---|---|---|
| GPU | NVIDIAGeForceRTX4060(8 GB) | 133.3 | ±2.8 |
| GPU | NVIDIAGeForceRTX3060(12 GB) | 102.4 | ±2.1 |
| CPU | IntelCorei9-13900HX | 30.1 | ±1.6 |

## 3 讨 论

苹果叶片病害的快速准确的识别对果业经济发展具有重要意义，但现有研究在实际果园场景下的检测精确率、轻量化与实时性之间难以实现良好平衡。针对果园场景下的苹果叶片病害检测任务，本研究提出了一种轻量化、实时的检测模型SMPD-YOLO。结果表明，该模型在检测精度、速度、模型权重、FLOPs与复杂环境下的鲁棒性之间取得了良好平衡。与最新的YOLOv11n、YOLOv12n、YOLOv13n模型相比，其FLOPs提升至$7.3\times10^{9}s^{-1}$，且SMPD-YOLO综合性能在其他所有评估指标中均达到最优。与YOLOv8n相比，其在召回率、mAP@0.5和mAP@0.5:0.95上分别提升了3.80、1.80和1.60百分点，同时模型权重、参数量与FLOPs显著降低，FPS高达133.3帧/s。此外，SMPD-YOLO在各类噪声干扰下具备良好的鲁棒性。

尽管本研究在苹果叶片病害检测中取得一定进展，但仍存在若干局限性，有待后续进一步改进与完善，未来研究将从以下几个方面展开：一是结合神经网络量化、结构重参数化等技术，进一步优化模型压缩策略，实现精度与效率的极致平衡；二是扩充数据集，重点构建大规模、细粒度标注的跨物种、跨环境病害数据集，涵盖早期病斑与复杂背景，同时研究自动化标注手段，降低训练成本，为模型泛化性奠定基础；三是拓展模型的跨作物泛化能力，探索轻量化的少样本学习范式，构建多作物、多病害的统一检测框架，增强模型的可迁移性和通用性；四是结合时序与多模态信息，探索模型在不同生长阶段及多模态数据融合下的稳健性；五是实现移动端和边缘端实时部署，实现低功耗、高精度的智能病害诊断系统，为农业智能化提供可落地的技术支撑。

## 参考文献 References


[1] SANGEETHA K, VISHNU RAJA P, RIMA P, et al. Apple leaf disease detection using deep learning[C]//2022 6th International Conference on Computing Methodologies and Communication (ICCMC). March 29-31, 2022, Erode, India. Erode: IEEE, 2022: 1063-1067.

[2] KHAN A I, QUADRI S M K, BANDAY S, et al. Deep diagnosis: a real-time apple leaf disease detection system based on deep learning [J/OL]. Computers and electronics in agriculture, 2022, 198: 107093 [2025-09-25]. https://doi. org/10.1016/j.compag.2022.107093.

[3] LI J W, QIAO Y L, LIU S, et al. An improved YOLOv5-based vegetable disease detection method [J/OL]. Computers and electronics in agriculture, 2022, 202: 107345 [2025-09-25]. https://doi.org/10.1016/j.compag.2022.107345.

[4] ZHANG S W, WANG D W, YU C Q. Apple leaf disease recognition method based on Siamese dilated Inception network with less training samples [J/OL]. Computers and electronics in agriculture, 2023, 213: 108188 [2025-09-25]. https://doi.org/10.1016/j.compag.2023.108188.

[5] 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, 2017, 39(6): 1137-1149.

[6] 姜晟，曹亚芃，刘梓伊，等. 基于改进Faster RCNN的茶叶叶部病害识别[J]. 华中农业大学学报，2024，43(5)：41-50. JIANG S, CAO Y P, LIU Z Y, et al. Recognition of tea leaf disease based on improved Faster RCNN [J]. Journal of Huazhong Agricultural University, 2024, 43(5): 41-50 (in Chinese with English abstract).

[7] LI Z S, SHEN Y, TANG J S, et al. IMLL-DETR: an intelligent model for detecting multi-scale litchi leaf diseases and pests in complex agricultural environments [J/OL]. Expert systems with applications, 2025, 273: 126816 [2025-09-25]. https://doi.org/10.1016/j.eswa.2025.126816.

[8] YAO J B, ZHU Z H, YUAN M Q, et al. The detection of maize leaf disease based on an improved real-time detection transformer model [J/OL]. Symmetry, 2025, 17 (6): 808 [2025-09-25]. https://doi.org/10.3390/sym17060808.

[9] 赵嘉威，田光兆，邱畅，等. 基于改进YOLOv4算法的苹果叶片病害检测方法[J]. 江苏农业科学，2023，51(9)：193-199. ZHAO J W, TIAN G Z, QIU C, et al. Apple leaf disease detection method based on improved YOLOv4 algorithm [J]. Jiangsu agricultural sciences, 2023, 51 (9): 193-199 (in Chinese with English abstract).

[10] 李昱达,吴正平,孙水发,等.改进YOLOv5算法的多类苹果叶片病害检测[J].中国农机化学报,2024,45(12):230-237. LI Y D,WU Z P,SUN S F,et al.Multi-species apple leaf disease detection with improved YOLOv5 algorithm[J].Journal of Chinese agricultural mechanization,2024,45(12):230-237 (in Chinese with English abstract).

[11] 袁杰,谢霖伟,郭旭,等.基于改进YOLOv7的苹果叶片病害检测方法[J].农业机械学报,2024,55(11):68-74. YUAN J,XIE L W,GUO X,et al.Apple leaf disease detection method based on improved YOLOv7[J].Transactions of the CSAM,2024,55(11):68-74 (in Chinese with English abstract).

[12] 张惠莉,代晨龙,任景龙,等.基于YOLOv8-GSGF模型的葡萄病害识别方法研究[J].农业机械学报,2024,55(11):75-83. ZHANG H L,DAI C L,REN J L,et al.Grape disease identification method based on YOLOv8-GSGF model[J].Transactions of the CSAM,2024,55(11):75-83 (in Chinese with English abstract).

[13] 曾晏林,贺壹婷,蔺瑶,等.基于BCE-YOLOv5的苹果叶部病害检测方法[J].江苏农业科学,2023,51(15):155-163. ZENG Y L,HE Y T,LIN Y,et al. Apple leaf disease detection method based on BCE-YOLOv5 [J]. Jiangsu agricultural sciences,2023,51(15):155-163 (in Chinese with English abstract).

[14] 黄政,张涛,孔万仔,等.基于YOLOv8n的梨树叶片病害检测模型[J].湖南农业大学学报(自然科学版),2025,51(2):113-121. HUANG Z,ZHANG T,KONG W Z,et al.Detection model of pear leaf disease based on YOLOv8n[J].Journal of Hunan Agricultural University (natural sciences edition),2025,51(2):113-121 (in Chinese with English abstract).

[15] XUE Z Y,XU R J,BAI D,et al.YOLO-tea:a tea disease detection model improved by YOLOv5[J/OL].Forests,2023,14(2):415[2025-09-25].https://doi.org/10.3390/f14020415.

[16] THAPA R,ZHANG K,SNAVELY N,et al.The plant pathology challenge 2020 data set to classify foliar disease of apples[J/OL].Applications in plant sciences,2020,8(9):e11390[2025-09-25].https://doi.org/10.1002/aps3.11390.

[17] THAPA R,ZHANG K,SNAVELY N,et al. Plant pathology 2021-FGVC8[EB/OL]. [2025-09-25]. https://kaggle.com/competitions/plant-pathology-2021-fgvc8.

[18] SINGH D,JAIN N,JAIN P,et al.PlantDoc:a dataset for visual plant disease detection[C]//Proceedings of the 7th ACM IKDD CoDS and 25th COMAD. January 5-7,2020,Hyderabad India.New York:ACM,2020:249-253.

[19] HUSSAIN M.YOLO-v1 to YOLO-v8,the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection[J/OL].Machines,2023,11(7):677[2025-09-25].https://doi.org/10.3390/machines11070677.

[20] LIU L Z,LI P R,WANG D H,et al.A wind turbine damage detection algorithm designed based on YOLOv8[J/OL].Applied soft computing,2024,154:111364 [2025-09-25]. https://doi.org/10.1016/j.asoc.2024.111364.

[21] SANDHYA,KASHYAP A.A novel method for real-time object-based copy-move tampering localization in videos using fine-tuned YOLO V8[J/OL].Forensic science international:digital investigation,2024,48:301663[2025-09-25].https://doi.org/10.1016/j.fsidi.2023.301663.

[22] LIN T Y,DOLLÁR P,GIRSHICK R,et al. Feature pyramid networks for object detection[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). July 21-26,2017,Honolulu,HI,USA. Honolulu:IEEE,2017:2117-2125.

[23] LIU S,QI L,QIN H F,et al.Path aggregation network for instance segmentation[PP/OL].arXiv,2018:1803.01534[2025-09-25].https://doi.org/10.48550/arXiv.1803.01534.

[24] WANG X Q,GAO H B,JIA Z M,et al.BL-YOLOv8:an improved road defect detection model based on YOLOv8[J/OL].Sensors,2023,23(20):8361[2025-09-25].https://doi.org/10.3390/s23208361.

[25] WU T Y,DONG Y K.YOLO-SE:improved YOLOv8 for remote sensing object detection and recognition[J/OL].Applied sciences,2023,13(24):12977[2025-09-25].https://doi.org/10.3390/app132412977.

[26] HE K M,ZHANG X Y,REN S Q,et al.Spatial pyramid pooling in deep convolutional networks for visual recognition[J].IEEE transactions on pattern analysis and machine intelligence,2015,37(9):1904-1916.

[27] TERVEN J,CÓRDOVA-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.

[28] CHEN L C,PAPANDREOU G,KOKKINOS I,et al.DeepLab:semantic image segmentation with deep convolutional nets,atrous convolution,and fully connected CRFs[J].IEEE transactions on pattern analysis and machine intelligence,2018,40(4):834-848.

[29] 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]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 17-24,2023,Vancouver,BC,Canada.Vancouver:IEEE,2023:7464-7475.

[30] YAN J H,ZHOU Z,ZHOU D J,et al.Underwater object detection algorithm based on attention mechanism and cross-stage partial fast spatial pyramidal pooling[J/OL].Frontiers in marine science,2022,9:1056300[2025-09-25].https://doi.org/10.3389/fmars.2022.1056300.

[31] ZHENG Z H,WANG P,REN D W,et al.Enhancing geometric factors in model learning and inference for object detection and instance segmentation[J].IEEE transactions on cybernet-

ics,2022,52(8):8574-8586.

[32] MA S L, XU Y. MPDIoU: a loss for efficient and accurate bounding box regression [PP/OL]. arXiv, 2023: 2307.07662 [2025-09-25].https://arxiv.org/abs/2307.07662.

[33] GALE T, ELSEN E, HOOKER S. The state of sparsity in deep neural networks [PP/OL]. arXiv, 2019: 1902.09574 [2025-09-25].https://arxiv.org/abs/1902.09574.

[34] GALE T, ZAHARIA M, YOUNG C, et al. Sparse GPU kernels for deep learning [C]//SC20: International Conference for High Performance Computing, Networking, Storage and Analysis. November 9-19, 2020, Atlanta, GA, USA. Atlanta: IEEE, 2020: 1-14.

[35] LEE J, PARK S, MO S, et al. Layer-adaptive sparsity for the magnitude-based pruning [PP/OL]. arXiv, 2020: 2010.07611 [2025-09-25].https://arxiv.org/abs/2010.07611.

[36] ROMERO A, BALLAS N, KAHOU S E, et al. FitNets: hints for thin deep nets[PP/OL].arXiv, 2014: 1412.6550[2025-09-25].https://arxiv.org/abs/1412.6550.

[37] SHU C Y, LIU Y F, GAO J F, et al. Channel-wise knowledge distillation for dense prediction[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). October 10-17, 2021, Montreal, QC, Canada. Montreal: IEEE, 2021: 5291-5300.

[38] TONG Z J, CHEN Y H, XU Z W, et al. Wise-IoU: bounding box regression loss with dynamic focusing mechanism [PP/OL]. arXiv, 2023: 2301.10051 [2025-09-25]. https://arxiv.org/abs/2301.10051.

[39] ZHANG H, ZHANG S J. Focaler-IoU: more focused intersection over union loss[PP/OL].arXiv, 2024: 2401.10525[2025-09-25].https://arxiv.org/abs/2401.10525.

[40] YUE K, DENG J, ZHOU F. Matching guided distillation[PP/OL].arXiv, 2020: 2008.09958[2025-09-25]. https://doi.org/10.48550/arXiv.2008.09958.

[41] WEI Y, PAN X, QIN H, et al. Quantization mimic: towards very tiny CNN for object detection[C]//European Conference on Computer Vision (ECCV) 2018. September 8-14, 2018, Munich, Germany. Cham: Springer, 2018: 267-283.

# A pruning and knowledge distillation-based YOLOv8 lightweight method for detecting leaf diseases in apple

ZHANG Shuaiping, SHI Lei, ZHENG Guang, WANG Huixin, YIN Fei

*College of Information & Management Sciences, Henan Agricultural University, Zhengzhou 450046, China*

**Abstract** A lightweight, real-time model of detecting leaf diseases in apple, named as SMPD-YOLO, was proposed based on YOLOv8n to solve the problems in existing methods of detecting leaf diseases in apple in terms of accuracy, real-time performance, and robustness under complex and noisy environments. The spatial pyramid pooling fast cross stage partial CSP (SPPFCSPC) module was embedded into the backbone network to enhance the capability of feature fusion. The minimum point distance-IoU (MPDIoU) was introduced as the bounding box regression loss function to improve the accuracy and convergence speed of the model. The model volume and floating point operations were further reduced via layer adaptive magnitude-based pruning (LAMP). A channel-wise knowledge distillation (CWD) strategy was used to boost detection performance. The results showed that the improved SMPD-YOLO model had a mean average precision (mAP@0.5) of 90.20% and a frame rate of 133.3 frames per second (FPS), while the weight and FLOPs of model was 5.0 MB and $7.3\times10^{9}$ s$^{-1}$, respectively. The improved model maintained excellent robustness under complex and noisy environments including strong illumination, weak light, and blurred images. It is indicated that the SMPD-YOLO model combines high accuracy, lightweight design, and real-time performance, enabling the efficient detection of leaf diseases on resource-constrained equipment.

**Keywords** apple; detection of leaf diseases; YOLOv8; lightweight; adaptive magnitude-based pruning; knowledge distillation

(责任编辑:葛晓霞)