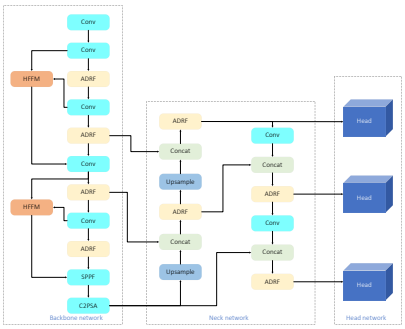
OEE | 光电工程 Opto-Electronic Engineering

Article

2026 年 6 月, 第 53 卷, 第 6 期

MSAF-YOLO：基于 YOLOv11 的玉米病虫害检测算法



李志明[1*]，贾改秀[2]

[1]河西学院信息技术与传媒学院，甘肃 张掖 734000；
[2]张掖市农业科学研究院，甘肃 张掖 734000

摘要：针对玉米病虫害检测任务中复杂背景以及小目标检测场景所存在的检测精度低以及漏检误检的问题，本文提出了一种改进型的 YOLOv11 模型：MSAF-YOLO。引入类残差反馈路径与通道注意力，设计了异构特征融合模块（heterogeneous feature fusion module，HFFM）模块，实现了跨层特征校准与增强，通过反馈连接将深层语义特征注入浅层特征中，解决了浅层网络语义信息不足，小目标特征在采样中丢失的问题。结合空间自适应权重与多分支空洞卷积，设计了 ADRF 模块，实现了像素级感受野调节，能够实现动态调整卷积层的感受野大小。解决了传统固定感受野卷积神经网络难以适应不同环境下目标检测的问题。实验结果表明，MSAF-YOLO 模型相较于原 YOLOv11 模型，在降低参数量的同时，mAP50 和 mAP50:95 在自建数据集 CORN 上分别提升了 3.2% 和 2.7%，在通用数据集 diseases 上分别提升了 3.7% 和 3.1%，证明了 MSAF-YOLO 模型的有效性。

关键词：玉米病虫害检测；YOLOv11；多尺度特征；卷积神经网络

The English title and abstract appear at the end of the paper.

DOI: 10.12086/oee.2026.250353 | CSTR: 32245.14.oee.2026.250353 | 中图分类号：TP391.4 | 文献标识码：A

李志明，贾改秀. MSAF-YOLO：基于 YOLOv11 的玉米病虫害检测算法 [J]. 光电工程，2026，**53**(6): 250353

Li Z M, Jia G X. MSAF-YOLO: A corn pest and disease detection algorithm based on YOLOv11[J]. *Opto-Electron Eng*, 2026, **53**(6): 250353

1 引　言

玉米作为全球重要的粮食作物，在农业生产、工业生产以及经济等领域具有关键作用[1]。据统计，近年来我国玉米种植面积和产量在谷物中占 40% 以上，且呈现逐步增加态势，故其稳定的生产对于保障我国粮食安全、促进经济发展具有深远意义。但是在玉米的生长过程中，病虫害的侵袭制约着其产量和质量的提升，目前玉米主要面临着黏虫、大斑病、锈病、茎腐病等病虫害的威胁，从而导致玉米生长受阻、产量降低[2]。传统的玉米病虫害检测方法主要是依赖于人工观察和判断，种植者需定期深入田间检查玉米植株，该方法的检测效率极低，且对于大面积的玉米种植区域，需耗费大量的人力和时间成本，此外人工检测无法实时监测，存在漏检和误判的情况[3]。因此设计一种自动化的、高效的玉米病虫害检测方法迫在眉睫。

玉米生长过程中常受多种病虫害威胁，其中玉米螟虫害、叶斑病、锈病是影响产量与品质的主要因素。这些病虫害在图像中呈现显著差异。

1) 玉米螟虫害：体长为 5~10 mm，在田间图像中仅

收稿日期：2025-11-17；修回日期：2026-02-12；录用日期：2026-02-28；网络出版日期：2026-06-25

基金项目：甘肃省自然科学基金项目 (23JRRG0012, 23JRRG0036)；甘肃省高校教师创新基金项目 (2025B-193)；教育部产学合作协同育人项目 (231005634170926)；甘肃省 2025 年高等教育教学改革研究项目 (GJJGB295)；教育部供需对接就业育人项目 (2024011199167)

*通信作者：李志明，zhangyelzm@163.com

版权所有©2026 中国科学院光电技术研究所

占数十像素，颜色与叶片相近，且常藏匿于叶背或叶鞘处，属于典型小目标。

2) 叶斑病：早期病斑呈小椭圆形淡黄色斑点 (直径 2~5 mm)，易与水滴、光照斑驳混淆，后期病斑扩大为长梭形灰褐色斑块，但边缘模糊，与健康组织过渡区域难界定。

3) 锈病：病斑呈橙色至褐色粉状小点 (直径 1~3 mm)，密集分布于叶面，常形成小目标密集区域，且颜色与土壤背景相似。

上述玉米病虫害的视觉特性给检测任务带来的挑战主要有小目标检测困难、早期病斑或幼虫在图像中占比极小、特征微弱、易在卷积下采样过程中丢失。复杂背景干扰，田间图像中存在叶片重叠、光照不均、泥土污渍、杂草等干扰，易导致误检。

近年来随着计算机视觉技术的不断发展，深度学习方法在目标检测领域有着广泛的应用，并取得了显著进展。YOLO (You only look once) 系列算法作为目标检测领域的典型算法，具有高效的检测速度和优异的性能[4]。目前，YOLOv11 算法[5] 作为该系列算法较新的迭代版本，在训练策略、网络架构，以及检测精度等方面做了进一步的优化和改进，具有较好的目标检测能力，可为玉米病虫害检测任务提供基本的技术支持。但现阶段玉米病虫害检测领域仍存在一些难题，主要有以下两点：

1) 复杂环境干扰：田间环境存在复杂背景、光照变化、病虫害形态多样及遮挡[6]，易导致模型在复杂环境下难以精准定位并识别目标。

2) 小目标检测困难：早期或小型的玉米病虫害在图像中占比小[7]，特征不明显，易导致漏检和错检。

针对上述问题，本文基于多尺度特征协同优化与自适应感受野卷积神经网络[8]，设计了一种 MSAF-Net (multi-scale adaptive feedback network) 的改进主干网络，利用 MSAF-Net 对 YOLOv11 模型的主干网络进行重构，提出了一种改进的 YOLOv11 模型：MSAF-YOLO 模型。本文在该模型中主要设计了 HFFM 分层特征反馈模块 (hierarchical feature feedback module) 和 ADRF 自适应动态感受野模块 (adaptive dynamic receptive field module)。

1) HFFM 分层特征反馈模块：引入类残差反馈路径与通道注意力，实现跨层特征校准与增强，通过反馈连接将深层语义特征注入浅层特征中。解决了浅层网络语义信息不足，小目标特征在采样中丢失的问题。

2) ADRF 自适应动态感受野模块：通过空间自适应权重与多分支空洞卷积，实现像素级感受野调节，能够实现动态调整卷积层的感受野大小。解决了传统固定感受野卷积神经网络难以适应不同环境下目标检测的问题。

2 相关工作

在 2015 年，Redmon 等[9] 提出了 YOLO 模型，为目标检测技术领域奠定了基础。初代 YOLO 之后，相继出现了 YOLOv2[10] 和 YOLOv3[11] 模型，通过引入锚框、多尺度特征融合等机制，提升了模型对目标的检测能力和精度。YOLOv4[12] 在 YOLOv3 的基础上，集成了多种高效的训练与模块化技术，在保持高精度的前提下降低计算量和参数量。近年来，YOLO 系列算法得到了进一步的发展，随着 YOLOv8[13] 到 YOLOv13[14] 的迭代版本推出，YOLO 算法实现了轻量化的同时，达到更高的精度上限。YOLO 系列算法也逐步应用到多个领域，如交通[15]、农业[16]、工程[17] 等领域。

目前，随着深度学习方法的不断发展，近年来许多学者都基于该方法在玉米病虫害检测领域中做了相关的研究。Maginga 等[18] 提出了一种利用小波变换与混合 CNN-LSTM 模型进行玉米早期病害识别，结合不同物理特征的数据，提升了模型的鲁棒性，但其计算量增加，在数据稀缺场景下 (如特定病害样本少)，模型易出现过拟合。Wang 等[19] 结合高光谱成像技术，设计了一种混合的 1D-CNN-BiLSTM 模型，降低了模型训练的参数量，使模型更轻量化，但在数据采集过程较为复杂，且模型的泛化能力差。Liu 等[20] 基于 DETR 模型，采用 Transformer 架构的端到端目标检测模型，提出了一种 SMSERT-DETR 模型，大幅减少了计算量和参数量，且具有较好的泛化能力，但对于小目标检测困难。承等[21] 基于 YOLOv5n 模型通过引入注意力机制、优化特征网络等策略，解决了检测效率低的问题，但在复杂背景下，如不同光照、角度等，检测效果较差。施等[22] 基于 YOLOv8 模型，结合边缘计算技术解决了现有方法计算量大、参数量多的问题，同时保持了较优的精度。

为应对田间环境中目标尺度多变、背景复杂的问题，自适应感受野机制逐渐被引入农业目标检测中。例如，Chen 等[23] 提出 ASPP 模块，通过多个不同空洞率的并行空洞卷积捕获多尺度上下文信息。在后续许多农业病害检测研究 (如苹果病害、黄瓜病害) 中，将其引入到目标检测网络 (如 Faster R-CNN，YOLO)，以提升对病斑

多尺度特征的感知能力，但对小目标特征保留不足。Tan 等[24] 提出加权双向特征金字塔网络 (BiFPN)，通过可学习的权重快速、高效地进行多尺度特征融合。该方法被广泛用于改进 YOLO 系列 (如 YOLOv7 的改进)，用于柑橘、番茄等作物病虫害检测，提升了不同尺度目标的检测精度，但缺乏显式的反馈路径。Dai 等[25] 提出可变形卷积 (deformable convolution) 和可变形 RoI 池化，通过额外的偏移量预测，使卷积核能自适应地聚焦于不规则或尺度不一的目标区域。在 YOLOv9 等模型的改进中，被用于玉米、小麦病害检测，以增强模型对不规则病斑和复杂背景下目标的鲁棒性。但该方法未考虑空间对齐问题，缺乏对反馈特征的动态校准，在复杂的田间环境，某些病虫害与植株颜色相近，该方法存在漏检和误检的问题。

针对目前所存在的问题，本文设计了 HFFM 分层特征反馈模块和 ADRF 自适应动态感受野模块，并集成 HFFM 与 ADRF，重构 YOLOv11 的主干网络，构建了一个协同工作的多尺度自适应特征提取系统。

3 MSAF-YOLO 模型

为了解决玉米病虫害检测中出现的小目标漏检、误检以及复杂环境下检测效果低的问题，本文提出了 MSAF-YOLO 模型，以提升对小目标的特征提取能力，增强复杂环境下模型的鲁棒性。MSAF-YOLO 模型的架构图如图 1 所示，主要设计了 HFFM 模块和 ADRF 模块，并集成这两个模块得到 MSAF-Net 主干网络。具体集成逻辑如下：

1) 特征图维度匹配策略

保持 YOLOv11 原有的多尺度输出结构，所有模块的输出特征图均与输入尺度保持一致，确保与后续颈部网络 (Neck network) 的维度兼容。

2) HFFM 和 ADRF 模块嵌入位置

在主干网络中设置两个 HFFM 模块：第一个 HFFM

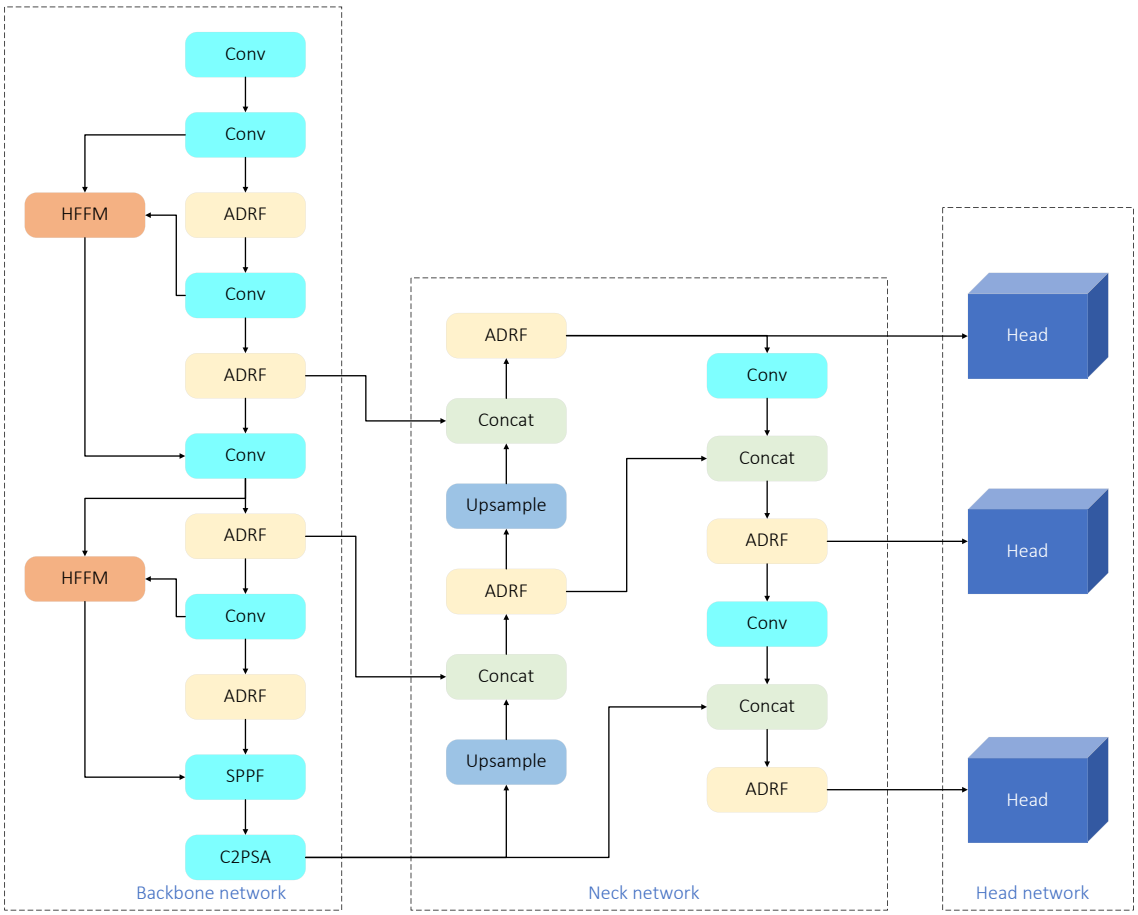


图 1 MSAF-YOLO 模型

Fig. 1 MSAF-YOLO model

嵌入在第二层至第四层路径，输入为第二层和第四层 Conv 中的特征，输出增强后的浅层特征。第二个 HFFM 嵌入在第六层到第八层路径，输入为第六层和第八层 Conv 中的特征，输出增强后的浅层特征。同时，用 ADRF 模块替换原主干网络和颈部网络中所有层级的 C3K2 模块。

3) 参数迁移与初始化策略

为保持训练稳定性，采用以下参数迁移策略：ADRF 替换 C3K2 时，保留原 C3K2 中卷积层的通道数、卷积核大小等配置，仅结构调整。HFFM 模块的卷积层采用 He 正态分布初始化，注意力层采用 Xavier 初始化。

HFFM 模块的核心思想是弥补浅层特征语义信息不足，通过利用深层特征的上下文信息来指导浅层特征，构建一条从深层到浅层的反馈路径，将富含语义的上下文信息精准地传递到负责目标定位的浅层特征，ADRF 模块通过图像局部特征自适应地融合不同感受野的特征，使模型的主干网络在面对复杂背景、尺度多变的小目标时，能够根据输入内容，动态调整每个空间位置的感受野，增强了特征分辨能力和鲁棒性。

本文使用 ADRF 模块替换 YOLOv11 模型原来主干网络以及颈部网络中 C3K2 标准卷积块，在 Backbone 和 Neck 的中间阶段引入 ADRF 模块，可使得模型在提取特征中具备动态自适应感受野的能力，在 Backbone 中从深层到浅层的连接处插入 HFFM 模块，可将深层的特征反馈给浅层，确保浅层特征富含足够的语义信息。由此得到改进后 YOLOv11 模型的主干网络：MSAF-Net。

3.1 HFFM 模块

在 YOLO 系列模型中，基于 CNN 的 FPN 网络及其变体通过自上而下的路径将深层语义传递至浅层，这种单向传递存在一个缺陷：当深层特征本身由于多次向下采样而丢失了小目标的特征信息时，其提供给浅层的特征信息是不完整的或存在噪声的。

在视觉感知系统中，高层语义信息可通过反馈路径调节低层特征响应，从而增强对细节特征的敏感性。借鉴这一思想，本文设计了 HFFM 模块，构建了一条从深层到浅层的显式反馈路径，旨在将富含全局语义信息的深层特征与保留空间细节的浅层特征进行自适应融合，以提升小目标与细微结构的特征表达能力。其与现有方法的主要区别如下：

1) 与残差网络 (ResNet) 的区别：传统残差连接旨在缓解梯度消失，其跳跃连接为恒等映射或线性变换，缺乏对特征通道的语义筛选能力。HFFM 则引入通道注意力机制，对反馈特征进行动态通道重校准，抑制无关噪声，增强目标相关特征。

2) 与特征金字塔 (FPN) 的区别：FPN 采用单向自上而下路径传递语义信息，缺乏从浅层到深层的反向指导。HFFM 则构建了双向特征交互机制，通过反馈路径实现语义信息从深层向浅层的注入，同时保留浅层细节，形成互补增强。

3) 与注意力机制 (如 ECA、CBAM) 的区别：现有注意力机制多作用于单层特征内部，而 HFFM 将注意力机制嵌入到跨层特征融合过程中，实现对反馈特征的通道级重要性评估，提升融合质量。

本文设计的 HFFM 模块的创新在于融合了反馈机制和通道注意力，深层特征虽然会丢失空间特征，但其包含强语义上下文信息。该模块旨在建立一条显式的反馈路径，将这种全局语义上下文特征精准地传递至浅层特征中，从而在早期阶段增强小目标的特征响应。HFFM 模块的结构图如图 2 所示，其详细工作流程如下述步骤所示。

步骤 1：特征对齐

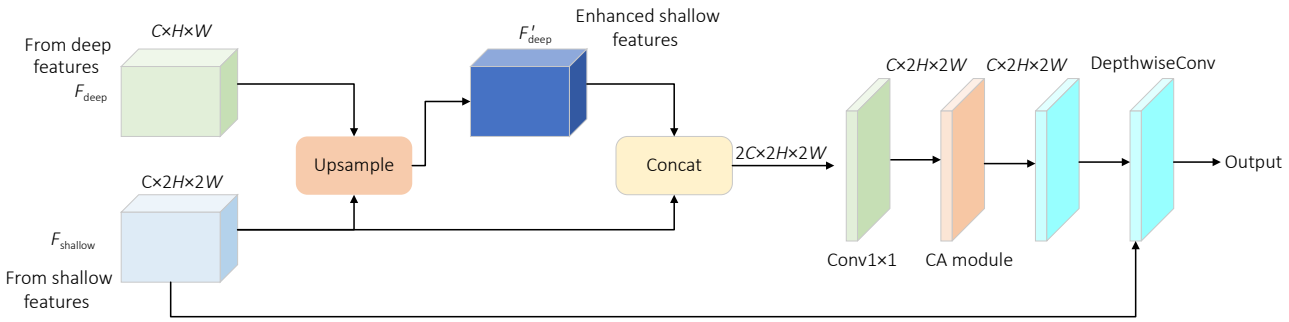


图 2 HFFM 模块结构图

Fig. 2 Structure of HFFM module

设深层特征$F_{\text{deep}} \in \mathbb{R}^{C\times H\times W}$和浅层特征$F_{\text{shallow}} \in \mathbb{R}^{C\times 2H\times 2W}$。首先对深层特征通过双线性插值进行2倍上采样至浅层特征的空间尺寸，得到空间尺寸与F_{shallow}相同的F'_{deep}，如式(1)所示：

$$F'_{\text{deep}} = Upsample_{2\times}(F_{\text{deep}}) . \tag{1}$$

该步骤的目的是使两个来源不同的特征图在空间维度上对齐，以便进行后续的融合操作。

步骤2：特征拼接与初步融合

对上采样后的深层特征F'_{deep}与原始的浅层特征F_{shallow}在通道维度上进行拼接(Concat)，如式(2)所示：

$$F_{\text{concat}} = Concat(F_{\text{shallow}}, F'_{\text{deep}}) , \tag{2}$$

其中：$F_{\text{concat}} \in \mathbb{R}^{2C\times 2H\times 2W}$，该步骤将来自浅层特征的细节信息和来自深层特征的语义信息进行初步融合，拼接后的特征通道数增加，为了控制计算复杂度，本文使用一个1×1的卷积进行降维，融合跨通道信息，如式(3)：

$$F_{\text{fused}} = Conv_{1\times1}(F_{\text{concat}}) . \tag{3}$$

步骤3：通道注意力校准

本文引入一个通道注意力子模块(CA Module)对F_{fused}进行加权，如图3所示。该模块同时进行全局平均池化(global average pooling, GAP)和全局最大池化(global max pooling，GMP)，生成两个不同的通道描述符，如式(4)、(5)所示：

$$GAP_{\text{c}} = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} F_{\text{fused}}(c,i,j) , \tag{4}$$

$$GMP_{\text{c}} = \max_{i,\ j} F_{\text{fused}}(c,\ i,\ j) . \tag{5}$$

对上述两个描述符进行拼接后，通过一个MLP(multilayer perceptron)，其主要由两个1×1卷积构成，中间包含ReLU激活和非线性瓶颈，最后通过Sigmoid函数(σ)生成通道权重向量$\boldsymbol{\alpha}$，如式(6)所示:

$$\boldsymbol{\alpha} = \sigma(\boldsymbol{W}_2\cdot\delta(\boldsymbol{W}_1\cdot Concat(GAP_{\text{c}}, GMP_{\text{c}}))) , \tag{6}$$

其中：σ表示Sigmoid激活函数；δ表示ReLU函数；$\boldsymbol{W}_1 \in \mathbb{R}^{(C/r)\times 2C}$和$\boldsymbol{W}_2 \in \mathbb{R}^{C\times(C/r)}$是$1\times1$卷积的权重；$r$是缩减比例。$\boldsymbol{\alpha}$源于融合了深层语义和浅层细节的$F_{\text{fused}}$，编码了每个通道对于识别小目标特征的重要性。故能够更准确地识别出哪些特征通道与小目标相关。最后，将$\boldsymbol{\alpha}$与F_{fused}进行相乘，得到校准后的特征，如式(7)所示。

$$F_{\text{calibrated}} = F_{\text{fused}} \odot \boldsymbol{\alpha} . \tag{7}$$

步骤4：融合与残差连接

校准后的特征$F_{\text{calibrated}}$通过一个3×3的深度可分离卷积，以平滑特征融合空间信息，且能够保持较低的计算量。最后与原始的F_{shallow}进行残差连接，如式(8)所示。

$$F_{\text{enhanced}} = DepthwiseConv_{3\times3}(F_{\text{calibrated}}) + F_{\text{shallow}} . \tag{8}$$

本文利用深度可分离卷积可在进一步融合空间上下文的同时，极大减少参数量和计算量。最后的残差连接可确保梯度能够直接回传，缓解训练过程中的梯度消失问题，并保留原始浅层特征中的信息。

3.2 ADRF模块

在YOLOv11模型的传统卷积神经网络架构中，C3K2标准卷积块每一层的感受野大小是固定的，主要由卷积核大小、层数和下采样率决定，这种固定的感受野在复杂场景的目标检测中存在明显不足。

1) 小目标：在玉米早期病斑中，需要小感受野来捕捉其局部特征，避免受到周围背景特征的干扰。

2) 大目标或密集目标：需要大感受野来理解全局特征和整体形态。

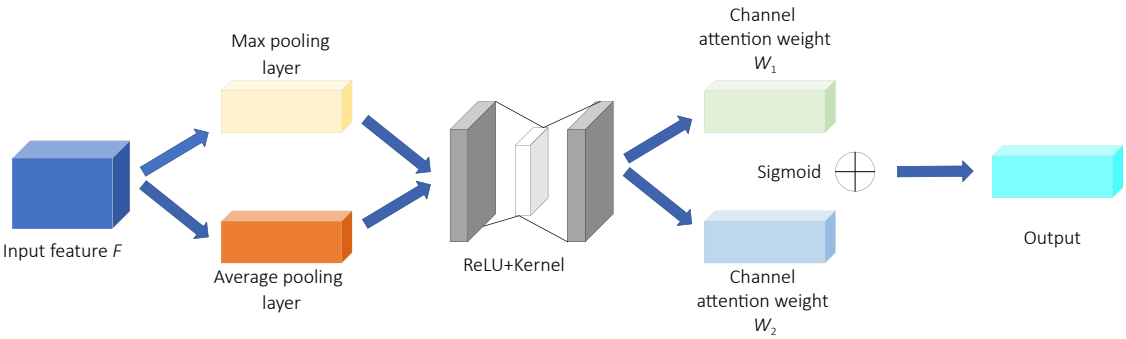


图3 通道注意力子模块

Fig. 3 Channel attention submodule

3) 复杂背景干扰：模型的网络架构需要自适应地关注目标特征，而非某些背景等噪声信息。

本文设计的 ADRF 模块核心创新在于，其通过并行多个不同的感受野卷积分支，引入一个轻量的空间权重生成网络 (SWN)，为每个像素位置动态地计算出一组最优的分支融合权重，根据输入特征，在前向传播过程中动态地、自适应地调整每个空间位置的特征提取感受野。

ADRF 模块与 ASPP (atrous spatial pyramid pooling) 的区别主要在于：ASPP 采用多个并行空洞卷积分支，但各分支权重固定，无法根据输入内容动态调整，而 ADRF 则引入空间权重生成网络 (SWN)，实现像素级动态权重分配，增强了对局部特征的适应性。

ADRF 模块与动态卷积 (dynamic convolution) 的区别：动态卷积通过注意力机制生成卷积核权重，计算开销大，难以部署于轻量平台。ADRF 保持各分支卷积核不变，仅通过 SWN 生成空间权重进行输出融合，在保持轻量化的前提下实现了感受野的自适应调节。

ADRF 模块的架构图如图 4 所示，其详细的工作流程如下述步骤所示。

步骤 1：多分支空洞卷积

本文在 ADRF 模块中并行设置了 3 个卷积分支，每个分支专门捕捉不同范围的上下文信息，其目的旨在覆盖从局部细节到全局上下文的特征尺度，空洞率的设置是基于对玉米病斑尺寸分布的统计分析：设边长为 l，小病斑 (l<16 pixel) 适用小感受野，占比约为 35%，中病斑 (16 pixel≤l≤ 48 pixel) 适用中等感受野，占比约 50%，大病斑 (l>48 pixel) 或密集区域需大感受野，占比约 15%，故空洞率设置为$d=[1,3,5]$。分支 1 的感受野最小，用于捕捉局部细节特征，如病斑的细微纹理、边缘等，如式 (9)。分支 2 的感受野中等，用于感知局部区域的上下文特征信息，如几个小病斑组成的簇，如公式 (10)。分支 3 的感受野最大，负责捕捉全局的上下文信息，如病斑在整个叶片上的分布模式，如式 (11)：

$$F_1 = DepthwiseConv_{3\times3}^{(d=1)}(F_{\text{input}}), \tag{9}$$

$$F_2 = DepthwiseConv_{3\times3}^{(d=2)}(F_{\text{input}}), \tag{10}$$

$$F_3 = DepthwiseConv_{3\times3}^{(d=3)}(F_{\text{input}}). \tag{11}$$

使用深度可分离卷积确保了在增加感受野的同时，参数量的增加是线性的，而不是幂增长，保持了模块的轻量化。使用空洞卷积可再不进行下采样，即不丢失分辨率的前提下扩大感受野。

步骤 2：空间权重生成网络—SWN

该子网络的设计是 ADRF 模块实现动态自适应的关键，根据输入特征F_{input}，即原始输入经过 3 个卷积分支所得到的特征 (F_1, F_2, F_3)，为每个像素点在各个分支上生成一组权重，SWN 的结构图如图 5 所示。

对输入F_{input}同时进行全局平均池化和全局最大池化，得到两个不同的全局上下文向量，如式 (12)、(13) 所示：

$$GAP_{\text{c}} = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}F_{\text{input}}(c,i,j), \tag{12}$$

$$GMP_{\text{c}} = \max_{i,j}F_{\text{input}}(c,i,j). \tag{13}$$

对上述两个不同的全局上下文向量进行拼接，通过第一个1×1卷积进行降维和非线性激活，通过第二个1×1卷积将特征映射到 3 个通道，对应 3 个不同的分支，然后利用广播机制将其扩展为与输入特征相同的空间尺寸$H\times W$，如式 (14)。在 SE-Net、ECA-Net 等经典注意力机制中，缩减比例通常设定在$r=[4,16]$之间，以平衡信息保留与计算效率。根据信息瓶颈理论，适度的降维可以过滤噪声、增强特征的表征能力，但过度降维会

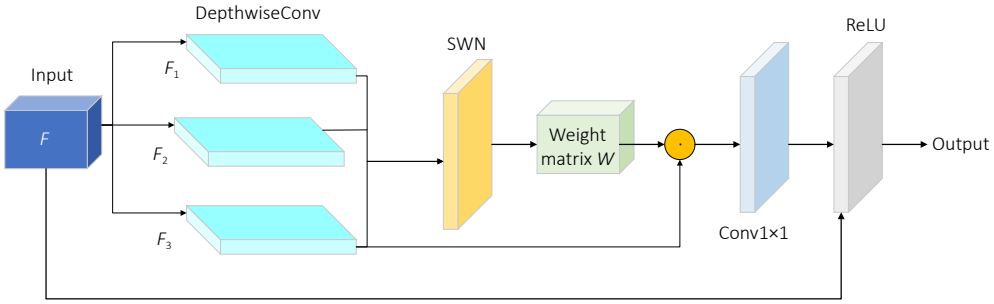


图 4 ADRF 模块结构图

Fig. 4 Structure of ADRF module

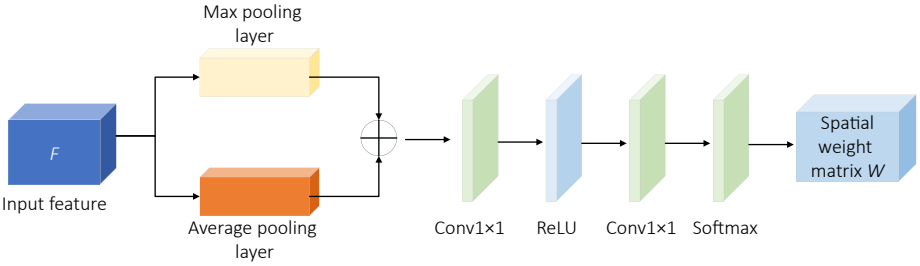


图 5 SWN 网络结构图

Fig. 5 Structure of SWN network

导致关键信息丢失。通过梯度传播分析，当$r=4$时，SWN 的梯度方差保持在合理范围内，既避免梯度消失，又防止梯度爆炸。

$$W' = \boldsymbol{W}_2 \cdot ReLU(W_1 \cdot Concat(GAP, GMP)), \tag{14}$$

其中：$\boldsymbol{W}_1 \in \mathbb{R}^{(C/4)\times 2C}$，$\boldsymbol{W}_2 \in \mathbb{R}^{3\times(C/4)}$。最后对每个空间位置$(i,j)$上的 3 个分支权重进行 Softmax 归一化，如式 (15)，最终可得到三个空间权重矩阵$\boldsymbol{W}_1$，$\boldsymbol{W}_2$，$\boldsymbol{W}_3$：

$$\boldsymbol{W}_k(i,j) = \frac{\exp(\boldsymbol{W}'_k(i,j))}{\sum_{m=1}^{3}\exp(\boldsymbol{W}'_m(i,j))}. \tag{15}$$

步骤 3：动态融合

对于每个分支的输出特征矩阵$\boldsymbol{F}_k$和其对应的空间权重矩阵$\boldsymbol{W}_k$，进行逐元素相乘，即利用权重矩阵对特征矩阵进行调整，如式 (16)：

$$F_{\text{fused}} = \sum_{k=1}^{3} \boldsymbol{W}_k \odot \boldsymbol{F}_k. \tag{16}$$

对于图像中的一个像素点(i,j)，如果其位于一个小而独立的病斑中心，SWN 会给分支 1 (小感受野) 分配更高的权重，因为其局部特征更重要。若其位于一个大面积病斑的内部，SWN 会给分支 3 (大感受野) 分配更高的权重，以利用其全局形态特征。若其位于复杂背景区域，SWN 会学习到一个平衡的权重，将目标与背景进行分离。

步骤 4：残差输出

融合后的特征F_{fused}通过一个1×1的点卷积，将多个分支融合后的特征信息在各个通道间进行传递和整合，如式 (17) 所示：

$$F_{\text{transformed}} = Conv_{1\times1}(F_{\text{fused}}). \tag{17}$$

最后，将变换后的特征与原始输入F_{input}相加，并通过 ReLU 激活函数确保梯度流动，保留原始特征的完整性，防止在优化过程中丢失已有特征，如式 (18) 所示：

$$F_{\text{out}} = ReLU(F_{\text{transformed}} + F_{\text{input}}). \tag{18}$$

4 实验与结果分析

4.1 数据集介绍

为满足玉米病虫害检测的小目标检测以及复杂背景下的检测，本文自制了 CORN 数据集。主要制作过程如下。

步骤 1：图像数据采集

为丰富数据集的样本多样性，本文利用网络爬虫技术从专业的农业科普网站、植物保护论坛等网页 (主要包括：PlantVillage 公开数据集中的玉米病害子集、AI Challenger 2018 农作物病害检测竞赛数据、农业农村部植物保护研究所公开的病虫害图像库)，系统爬取了一批玉米病虫害图像。

步骤 2：图像筛选与预处理

为确保数据质量，我们对图像进行了筛选：分辨率不低于 600 pixel×600 pixel，保证病斑细节可见；病斑区域无严重模糊；背景复杂度兼顾田间真实性与标注可行性。通过筛选后，共保留 6200 张图像，并统一缩放至 1024 pixel×1024 pixel，同时采用直方图均衡化进行颜色增强，以提升模型对光照变化的鲁棒性。

步骤 3：病虫害类别定义与标注

根据我国玉米生长过程中的主要病虫害因素，本文主要对常见的玉米枯叶病、叶斑病、锈病进行矩形边界框标注，其标注类别主要为健康玉米 (healthy)、枯叶病 (corn_blight)、叶斑病 (gray_leaf_spoot)、锈病 (corn_common_rust)、玉米害虫 (plaga)。每种病害类别 (枯叶病、叶斑病、锈病、害虫) 的代表性切片图像如图 6 所示，对于同一图像中出现的不同尺寸的多个目标，以及

部分被遮挡的目标，均进行独立、完整的标注。本文使用 Labellmg 工具进行标注工作，标注示例图如图 7 所示。

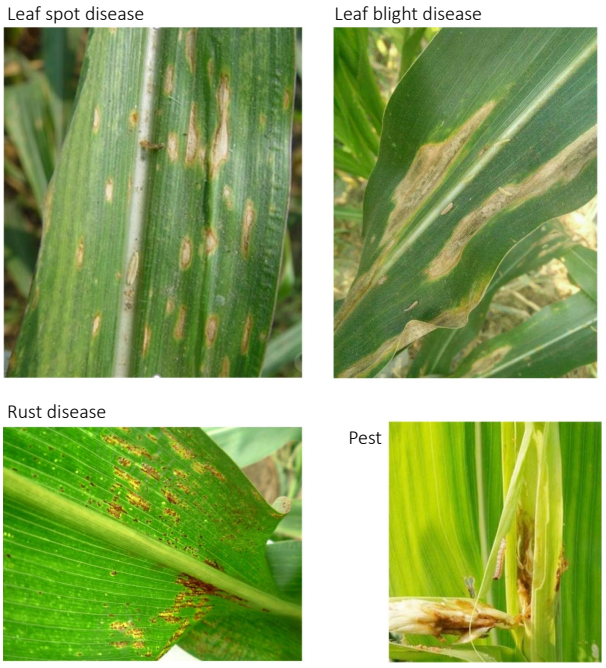


图 6 病害类别切片图

Fig. 6 Disease category slice diagram

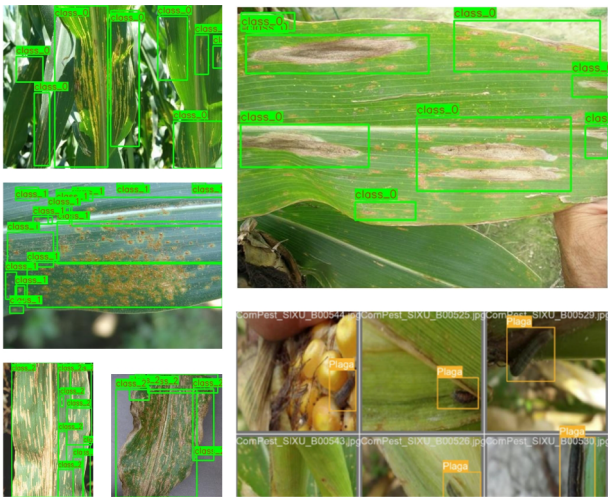


图 7 数据标注示例图

Fig. 7 Data Annotation example diagram

步骤 4：数据集划分

为对模型进行有效的训练和评估，本文将上述图像按 7:2:1 的比例随机划分为训练集 (4340 张图像)、验证集 (1240 张图像) 和测试集 (620 张图像)。具体的样本分布统计信息如表 1 所示。

表 1 CORN 数据集样本分布表

Table 1 CORN dataset sample distribution table

Category	Images	Bounding boxes
Healthy	1480	1620
Corn_blight	1350	2100
Gray_leaf_spoot	1250	1850
Corn_commom_rust	1400	1950
Plaga	1150	1500

为了进一步验证本文模型的泛化能力，本文除了在自制玉米病虫害检测数据集上进行实验，还在通用数据集 diseases 上进行对比实验和消融实验。Diseases 数据集共 4572 张图像，其类别主要分为病害玉米、虫害玉米、健康玉米，共包含 27000 个边界框。该数据集小目标占比较高，适用于验证本文模型的小目标检测能力。

4.2 实验环境与评价指标

本文的所有实验均在 Win 10 系统下进行，GPU 型号为 GeForce RTX 4060Ti (16 G)，训练框架为 Pytorch，CUDA11.3、Python3.9。实验的超参数配置如表 2 所示。

表 2 实验超参数设置表

Table 2 Experimental hyperparameter setting table

超参数名称	参数量	参数说明
训练轮数	200	总训练迭代次数，确保模型充分收敛
Batch_Size	8	每批次输入图像数，兼顾显存限制与训练稳定性
学习率	0.01	初始学习率，采用自适应调整
Momentum	0.937	优化器动量参数，加速收敛并平滑梯度更新
Image_size	640	输入图像尺寸，保持与YOLO系列标准设置一致

本文实验采用的评价指标主要有平均精度均值 (mAP) 和模型的参数量 (params)。

在目标检测领域，*AP* 是最核心的精度指标，其综合考虑了不同置信度阈值下精确率 (precision) 和召回率 (recall) 的性能，其计算过程首先对某一类别的所有预测框按置信度从高到低排序，在不同置信度阈值下，计算出一系列 (*P*,*R*) 点对，绘制 *P*-*R* 曲线，*P*-*R* 曲线下的面积即为该类别的 *AP* 值。*AP* 值的计算如式 (19) 所示。mAP是多个类别平均精度 (AP) 的平均值，是衡量模型整体检测性能的黄金标准，其计算如式 (20) 所示。

$$AP = \int_0^1 P(R)\mathrm{d}R\,, \tag{19}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i\,, \tag{20}$$

其中：$P(R)$为某个类别的 P-R 曲线。根据交并比 (IoU) 的不同取值，将 mAP 指标分为 mAP@0.5 和 mAP@0.5：0.95，mAP@0. 5 是 IoU 阈值为 0.5 时的 mAP 值，mAP@0.5：0.95 是指 IoU 阈值从 0.5 到 0.95 的平均 mAP。

4.3 HFFM 模块分析

为验证 HFFM 模块插入原模型的主干网络中的有效性和泛化性，本文分别在 YOLO 系列中具有不同任务功能的模型进行对比，包括目标检测 (YOLOv10n、YOLOv12)、实例分割 (YOLOv12 -seg) 和图像分类 (YOLOv12-cls)，虽然这些模型的任务目标不同，但其主干网络均基于 CNN 特征提取架构，因此可作为验证 HFFM 模块泛化性的合理基线。以及在较为先进的 DETR 系列算法 (RF-DETR[26]，RT-DETR[27]) 的主干网络中从深层到浅层的连接处插入 HFFM 模块，在 CORN 数据集上进行实验对比，实验结果如表 3 所示。实验结果表明，将 HFFM 模块作用在模型的主干网络中，mAP@0.5和mAP@0.5: 0.95均得到了显著提升，验证了 HFFM 模块的有效性，同时 HFFM 模块作用在不同系列的算法中均表现出较好的性能提升，证明了该模块具有较强的泛化性。

在 YOLOv11 模型上加入 HFFM 模块，模型改进前后的训练损失及验证损失对比如图 8 所示。由图 8 分析可得，加入 HFFM 模块后，模型在训练和验证阶段的 box_loss、cls_loss、dfl_loss 均有显著的下降，这表明模型成功地在训练数据中提升了精确地定位物体 (box_loss 下降)、准确地区分物体类别 (cls_loss 下降) 以及优化其边界框预测分布 (dfl_loss 下降) 的能力，并且在验证集上得到了同步体现，证明了模型具有良好的泛化能力。

表 3　在不同模型中加入 HFFM 模块的对比实验

Table 3　Comparative experiments incorporating the HFFM module into different models

Model	Before improvement		After improvement (+HFFM)	
	mAP@0.5	mAP@0.5：0.95	mAP@0.5	mAP@0.5：0.95
YOLOv10n	0.795	0.643	**0.819 (+2.4%)**	**0.659 (+1.6%)**
YOLOv12	0.817	0.653	**0.836 (+1.9%)**	**0.668 (+1.5%)**
YOLOv12-seg	0.827	0.652	**0.839 (+1.2%)**	**0.658 (+0.6%)**
YOLOv12-cls	0.829	0.659	**0.838 (+0.9%)**	**0.666 (+0.7%)**
RT-DETR	0.821	0.654	**0.836 (+1.5%)**	**0.664 (+1.0%)**
RF-DETR	0.832	0.665	**0.840 (+0.8%)**	**0.669 (+0.4%)**
YOLOv11	0.825	0.662	**0.846 (+2.1%)**	**0.681 (+1.9%)**

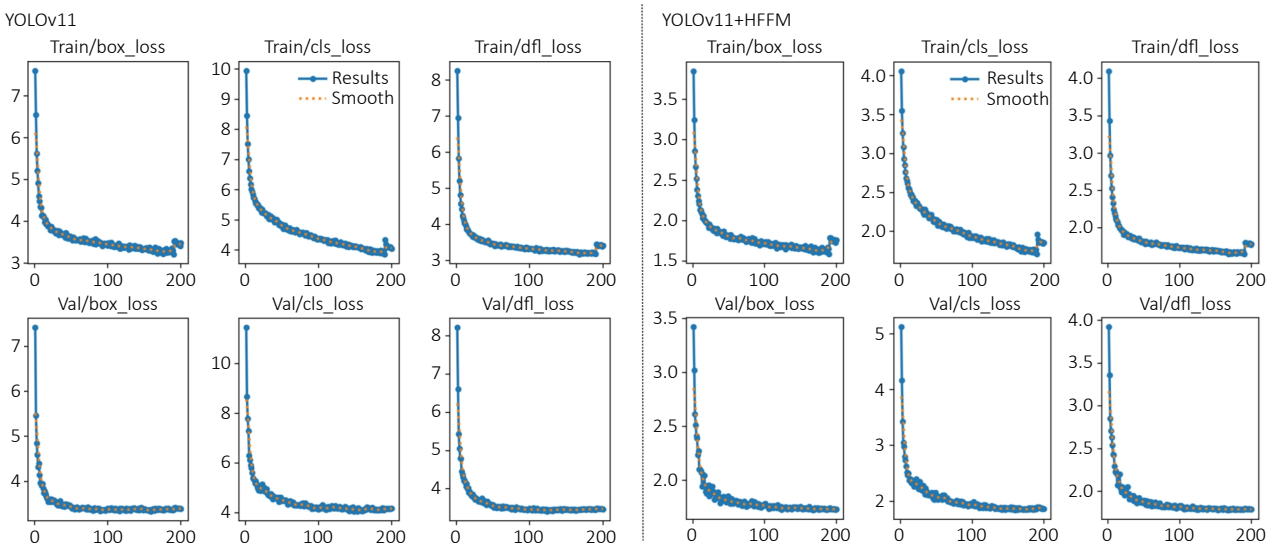


图 8　加入 HFFM 模块后的模型损失对比

Fig. 8　Comparison of model loss after integrating the HFFM module

为验证该模块中引入通道注意力校准的优势，本文在 HFFM 模块中分别使用空间注意力 (spatial attention，SA)、混合域注意力 (convolutional block attention module，CBAM)、坐标注意力 (coordinate attention，CA)、多头注意力 (multi-head attention, MHA) 来替换通道注意力，在 CORN 数据集上进行实验对比，如表 4 所示。由表 4 分析可得，相较于其他几种注意力，在该模块中引入通道注意力机制，在保持低参数量的同时，提升了 mAP@0.5和mAP@0.5：0.95值。

4.4 ADRF 模块分析

为验证 ADRF 模块的有效性，将其与原模型中的 C3K2 标准卷积模块，在 CORN 数据集上进行实验对比，如表 5 所示。实验表明，用 ADRF 模块替换 C3K2 模块后，mAP@0.5和mAP@0.5：0.95分别提升了 1.2% 和 0.7%，且下降了 5% 的参数量。C3K2 模块与 ADRF 模块对模型在训练和验证阶段的 box_loss、cls_loss、dfl_loss 的影响如图 9 所示。

表 4　不同注意力之间 mAP 值对比

Table 4　Comparison of mAP between different attentions

Module	Params/M	mAP@0.5	mAP@0.5：0.95
+SA	2.95	0.829	0.653
+CBAM	3.06	0.842	0.679
+CA	2.87	0.834	0.662
+MHA	2.81	0.836	0.668
Ours	**2.72**	**0.846**	**0.681**

由图 9 分析可得，本文设计的 ADRF 模块通过空间自适应权重与多分支空洞卷积，有效增强了模型的特征提取、空间定位和类别判别能力。通过 box_loss、cls_loss 和 dfl_loss 三项关键指标的显著的下降表明，ADRF 模块相比原模型中 C3K2 模块，能够让模型收敛更快、定位更准、泛化能力更强，更适用于解决复杂的玉米病虫害检测任务。

在 ADRF 模块中，本文引入了一个空间权重生成网络 SWN，为验证该网络在 ADRF 模块的优势，本文针对 ADRF 模块中有无 SWN 网络，在 CORN 数据集上进行了对比实验分析，如表 6 所示。数据表明，引入 SWN 网络后，mAP@0.5和mAP@0.5：0.95分别提升了 1.3% 和 0.8%，且参数量基本保持不变。

为验证 ADRF 模块在小目标、复杂背景下以及有遮挡的情况下的检测精度，本文按目标尺度、背景复杂度、遮挡程度对 CORN 数据集中的图片进行分类，并设计了

表 5　ADRF 和 C3K2 对比

Table 5　Comparison of ADRF and C3K2

Module	mAP@0.5	mAP@0.5：0.95	Params/M
+C3K2	0.825	0.662	2.79
+ADRF	**0.837 (+1.2%)**	**0.669 (+0.7%)**	**2.64 (−5%)**

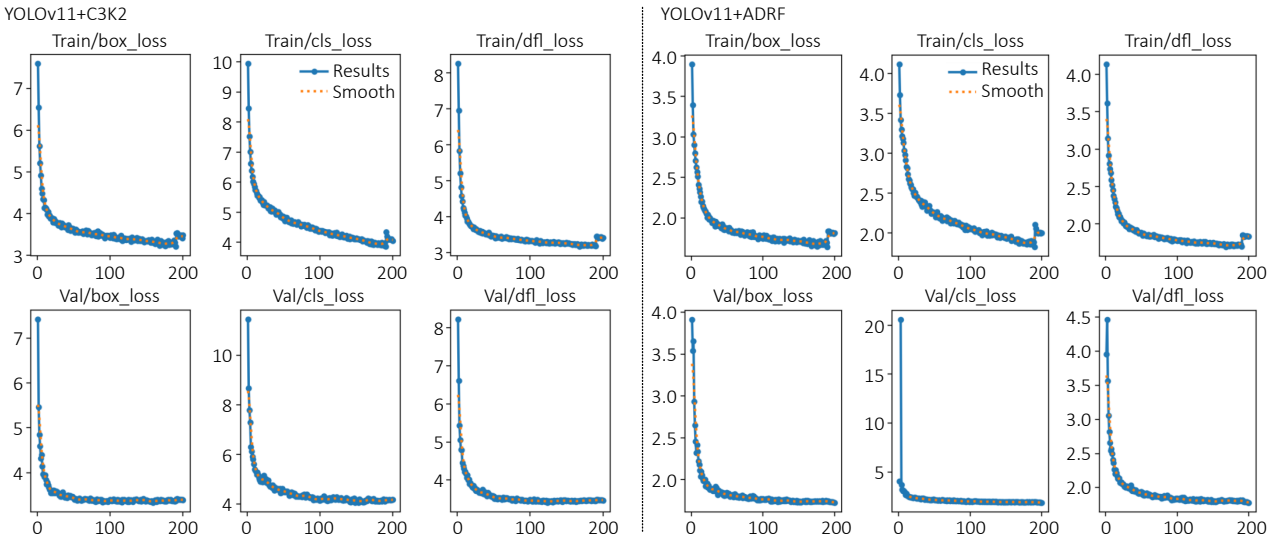


图 9　ADRF 模块与 C3K2 模块对模型性能的影响对比

Fig. 9　Comparison of the Impact of ADRF and C3K2 modules on model performance

对比实验，根据标注框面积占图像面积的比例分为三组：小目标 (面积占比<1%)、中目标 (1%≤面积占比≤10%)、大目标 (面积占比>10%)；根据图像背景杂乱程度分为两组：简单背景 (背景单一，无干扰物)、复杂背景 (多叶片、杂草、阴影等干扰)；根据目标被遮挡比例分为三组：无遮挡 (遮挡面积<10%)、部分遮挡 (10%≤遮挡面积≤50%)、严重遮挡 (遮挡面积>50%)。实验结果如表 7~9 所示。由表 7~9 分析可得：ADRF 模块对小目标、复杂背景、遮挡场景的提升最为显著，说明该模块能够通过动态融合多尺度特征，增强对局部可见区域的感知能力，从而缓解遮挡等不利条件带来的信息损失问题。

4.5 消融实验分析

为了验证 ADRF 模块和 HFFM 模块对模型改进的有效性，本文以 YOLOv11 模型作为基线模型分别在 diseases 数据集和 CORN 数据集上进行消融实验，如表 10 所示。

模型 1 为 YOLOv11 基线模型；模型 2 为 YOLOv11 的网络架构的基础上，插入 HFFM 模块，在 CORN 数据集上，较基线模型mAP@0.5和mAP@0.5: 0.95分别提升了 2.1% 和 1.9%，在 diseases 数据集上，较基线模型mAP@0.5和mAP@0.5: 0.95分别提升了 1.9% 和 1.8%；模型 3 为在 YOLOv11 的网络架构的基础上，将 C3K2 模块替换为 ADRF 模块，在 CORN 数据集上，较基线模型mAP@0.5和 mAP@0.5: 0.95分别提升了 1.2% 和 0.7%，在 diseases 数据集上，较基线模型mAP@0.5和mAP@0.5: 0.95分别提升了 1.3% 和 0.6%；模型 4 为本文模型，在 CORN 数据集上，较基线模型mAP@0.5和mAP@0.5: 0.95分别提升了 3.2% 和 2.7%，在 diseases 数据集上，较基线模型mAP@0.5和mAP@0.5: 0.95分别

表 6 SWN 网络有效性对比

Table 6 SWN network effectiveness comparison

Module	mAP@0.5	mAP@0.5：0.95	Params/M
ADRF (−SWN)	0.824	0.661	2.63
ADRF (+SWN)	**0.837 (+1.3%)**	**0.669 (+0.8%)**	**2.64**

表 7 针对不同目标尺度的模块验证

Table 7 Module verification for different object scales

Target scale	−ADRF module		+ADRF module	
	mAP@0.5	mAP@0.5：0.95	mAP@0.5	mAP@0.5：0.95
Small target	0.782	0.621	0.810 (+2.8%)	0.648 (+2.7%)
Medium target	0.836	0.674	0.845 (+0.9%)	0.680 (+0.6%)
Big target	0.855	0.691	0.862 (+0.7%)	0.695 (+0.4%)

表 8 针对不同背景复杂度的模块验证

Table 8 Module verification under different background complexities

Complexity of background	−ADRF module		+ADRF module	
	mAP@0.5	mAP@0.5：0.95	mAP@0.5	mAP@0.5：0.95
Simple	0.860	0.698	0.865 (+0.5%)	0.702 (+0.4%)
Complex	0.805	0.648	0.842 (+3.7%)	0.673 (+2.5%)

表 9 针对不同遮挡程度的模块验证

Table 9 Module verification under different occlusion levels

Degree of obstruction	−ADRF module		+ADRF module	
	mAP@0.5	mAP@0.5：0.95	mAP@0.5	mAP@0.5：0.95
Unobstructed	0.852	0.690	0.855 (+0.3%)	0.692 (+0.2%)
Partially obstructed	0.810	0.660	0.842 (+3.2%)	0.678 (+1.8%)
Severely obstructed	0.785	0.629	0.816 (+3.1%)	0.651 (+2.2%)

表 10 消融实验对比

Table 10 Comparison of ablation experiments

Model	HFFM	ADRF	CORN		Diseases		Params/M
			mAP@0.5	mAP@0.5：0.95	mAP@0.5	mAP@0.5：0.95	
1			0.825	0.662	0.772	0.597	2.79
2	√		0.846	0.681	0.791	0.615	2.72
3		√	0.837	0.669	0.785	0.603	2.64
4	√	√	**0.857**	**0.689**	**0.809**	**0.628**	**2.58**

提升了 3.7% 和 3.1%。由消融实验分析可得，ADRF 模块和 HFFM 模块在玉米病虫害领域检测的有效性，改进后的 MSAF-YOLO 模型可实现玉米病虫害的高精度检测，同时可降低模型的参数量。

改进后的模型 (MSAF-YOLO) 与 YOLOv11 模型在 CORN 数据集和 diseases 数据集上的损失曲线对比图如图 10 和图 11 所示。

由图 10 和图 11 分析可得，改进后的 MSAF-YOLO 模型在整体性能上显著优于基准模型 YOLOv11。具体分析如下。

1) 边界框回归损失分析：在两个数据集中，MSAF-YOLO 模型的 box_loss 曲线从训练初期开始就展现出更低的损失值，且收敛到一个更低的稳定平台。box_loss 衡量的是预测边界框与真实边界框之间的重合度误差 (如 CIoU)。更低的 box_loss 意味着模型能够更精确地定位图像中的目标物体，预测框的位置和大小更接近真实情况。MSAF-YOLO 模型在此方面上的显著优势，表明该模型有效增强了模型对目标空间位置和几何特征的感知能力。模型能够更准确地聚焦于目标本身，减少背景干扰，从而实现了更优的定位精度。

2) 分类损失分析：MSAF-YOLO 的 cls_loss 曲线在两大数据集上也始终低于 YOLOv11，并且收敛速度更快，最终稳定值更低。cls_loss 衡量的是模型对目标物体分类的正确性。更低的 cls_loss 则代表模型能更准确地将检测到的目标划分到正确的类别中 (例如区分不同种类的病害)。实验证明改进后的模型有效提升了模型的特征判别能力。多尺度特征融合机制帮助模型捕获了更多有助于区分细微类别差异的上下文信息，而注意力机制则进一步强化了对关键判别性特征的提取，抑制了无关特征，使得模型在复杂背景下也能做出更可靠的分类判断。

3) 分布焦点损失分析：在 dfl_loss 上，MSAF-YOLO 同样展现出了一致的优越性，其损失曲线明显低于 YOLOv11 模型。dfl_loss 是 YOLO 系列中用于优化边界

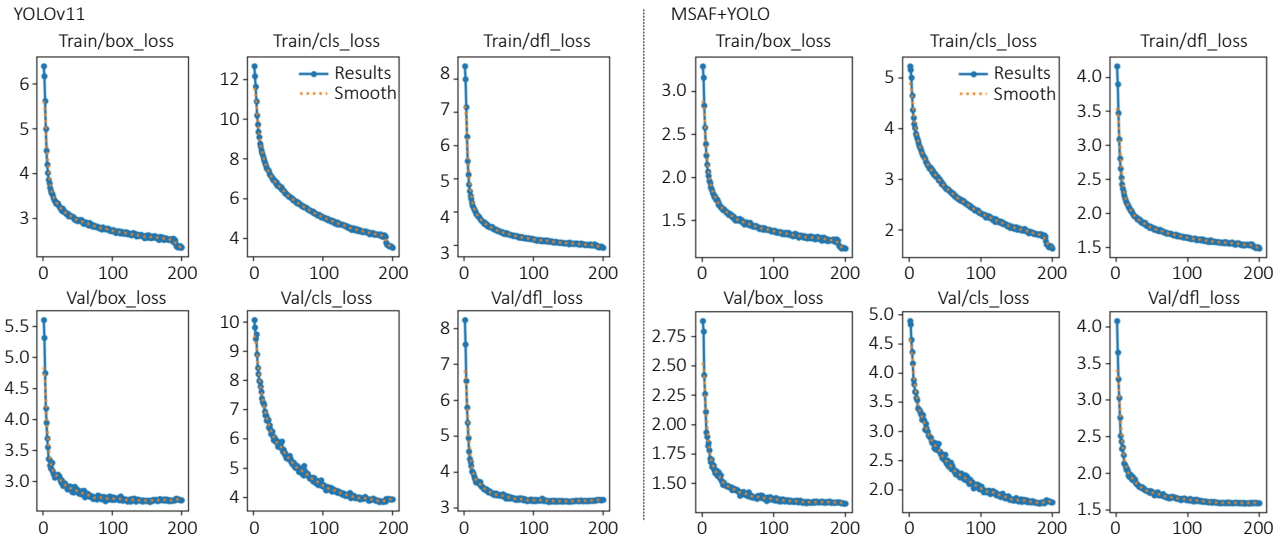


图 10 MSAF-YOLO 与 YOLOv11 在 CORN 数据集上的损失曲线对比

Fig. 10 Comparison of loss curves between MSAF-YOLO and YOLOv11 on the CORN dataset

框位置回归分布的一种损失。它不直接预测一个绝对坐标值，而是预测坐标值的一个分布，通过让分布更尖锐、更准确地聚焦到真实值附近来提升回归精度。MSAF-YOLO 模型更低的 dfl_loss 说明，模型对于边界框坐标位置的预测置信度更高、更集中。这与 box_loss 的降低是相互印证的。表明改进后的模型不仅定位更准，模型的回归预测更加精准。

4.6 对比实验分析

为进一步验证本文模型的优势，在 diseases 数据集和 CORN 数据集上分别与 YOLOv8n-ghost[28]、YOLOv10n、YOLOv11、YOLOv12、YOLOv12-seg、YOLOv12-cls、YOLOv13、RF-DETR、RT-DETR 等目前主流目标检测算法进行对比实验。其中，尽管 YOLOv12-seg 和 YOLOv12-cls 的主要任务并非检测，但其检测分支仍基于相似的主干网络设计，因此可作为有意义的对比基线，以体现 MSAF-YOLO 在检测任务上的专门优化效果，实验结果如表 11 所示。由表 11 分析可得，在 CORN 数据集上，本文模型相较于以上几种算法，mAP@0.5最低提升 2.4%，最高提升 6.2%，mAP@0.5: 0.95最低提升 2.4%，最高提升 4.6%；在 diseases 数据集上，mAP@0.5最低提升 1.3%，最高提

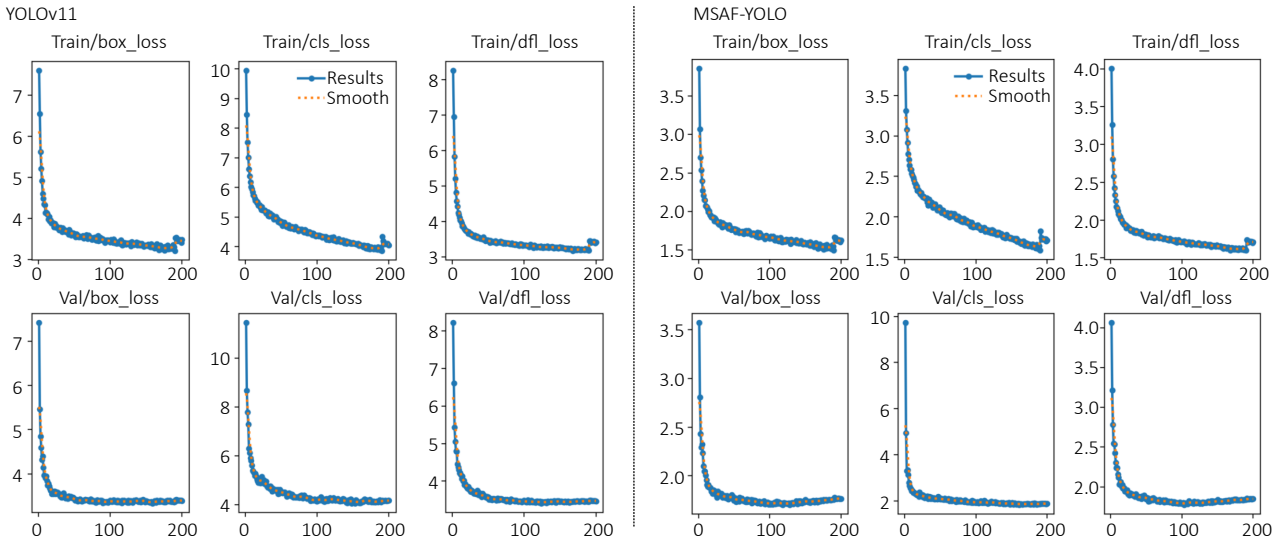


图 11 MSAF-YOLO 与 YOLOv11 在 diseases 数据集上的损失曲线对比

Fig. 11 Comparison of loss curves between MSAF-YOLO and YOLOv11 on the diseases dataset

表 11 与其他模型对比

Table 11 Comparison with other models

Model	CORN		Diseases		Params/M	GFLOPs
	mAP@0.5	mAP@0.5：0.95	mAP@0.5	mAP@0.5：0.95		
YOLOv11	0.825	0.662	0.772	0.597	2.79	6.7
YOLOv10n	0.795	0.643	0.745	0.552	2.80	6.8
YOLOv8n-ghost	0.815	0.649	0.768	0.571	2.13	6.4
YOLOv12	0.817	0.653	0.783	0.589	2.56	6.9
YOLOv12-seg	0.827	0.652	0.789	0.592	2.45	7.9
YOLOv12-cls	0.829	0.659	0.784	0.587	2.62	8.2
YOLOv8-SPPF-L	0.833	0.664	0.796	0.615	2.72	6.8
YOLOv13	0.819	0.653	0.785	0.583	2.68	6.3
RT-DETR	0.821	0.654	0.788	0.596	5.41	14.7
RF-DETR	0.832	0.665	0.792	0.613	8.52	15.2
Ours	**0.857**	**0.689**	**0.809**	**0.628**	**2.58**	**6.2**

升 6.4%， mAP@0.5: 0.95最低提升 1.3%，最高提升 7.6%。且能保持较低参数量和计算量。

4.7 实际检测效果展示

本节主要对模型在玉米病虫害领域上的检测进行直观展示，对比本文模型和 YOLOv11 模型在 CORN 数据集中测试数据上的检测效果，如图 12 所示。在 CORN 数据集上的直观对比显示，本文模型相较 YOLOv11 实现了更精准的病虫害识别。YOLOv11 模型对玉米叶片上小尺寸病斑存在明显漏检，且对密集目标的边界定位较为模糊；而本文模型显著提升了对玉米病斑等细微特征的捕捉能力，不仅减少了漏检现象，更提升了模型的检测效果，显著优于 YOLOv11 模型。

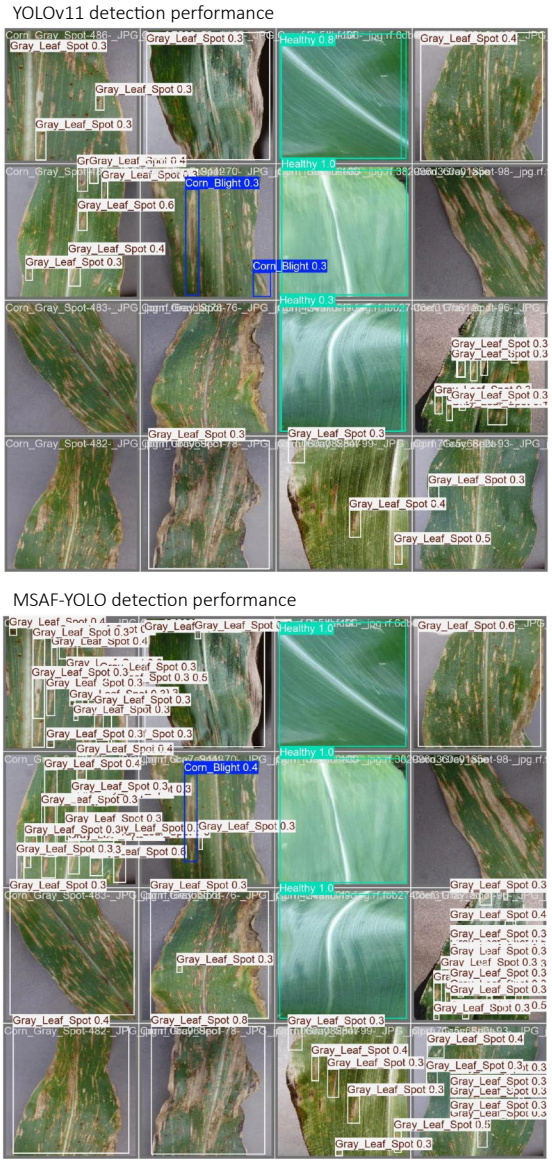


图 12 YOLOv11 和 MSAF-YOLO 的检测效果对比

Fig. 12 Comparison of detection performance between YOLOv11 and MSAF-YOLO

为增强本文模型在实际应用中的有效性，本文使用训练后的 MSAF-YOLO 模型针对模糊、倾斜、遮挡、多目标重叠等实际场景的图像进行检测效果测试，如图 13 所示。由图 13 可以看出，MSAF-YOLO 可以对模糊特征、倾斜特征、遮挡特征以及多特征重叠的病虫害叶片进行精准的识别和检测，证明了本文所提出的方法对田间复杂干扰因素具有较好的适应能力，验证了模型对于田间图像检测的鲁棒性。

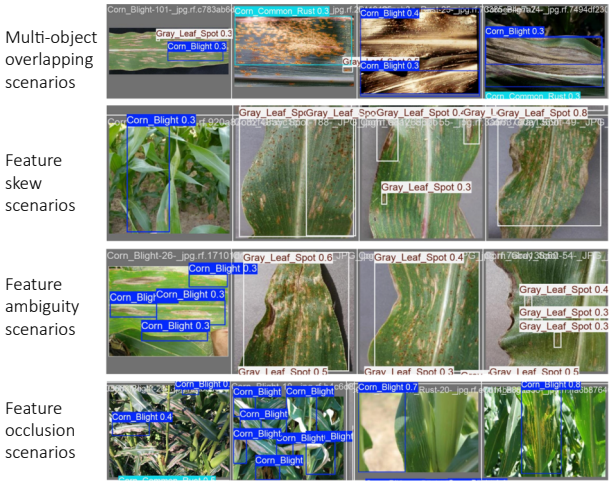


图 13 针对不同场景 MSAF-YOLO 的检测效果

Fig. 13 Detection performance of MSAF-YOLO across different scenarios

5 结 语

本文主要针对玉米病虫害检测领域中的小目标检测以及复杂背景干扰的问题，设计了 HFFM 分层特征反馈模块和 ADRF 自适应动态感受野模块。HFFM 模块不同于简单的特征拼接，而是通过通道注意力机制实现了特征增强，并且利用残差结构使得模型易于训练，降低了计算量和参数量，同时可增强对小目标检测的性能。ADRF 模块不受固定感受野的限制，能够根据图像特征动态调整，对于复杂背景、多尺度目标共存等场景，能够有效地提取出目标特征。相较于原 YOLOv11 模型，本文模型在 CORN 数据集上，mAP@0.5和mAP@0.5: 0.95分别提升了 3.2% 和 2.7%，在 diseasea 数据集上，mAP@0.5和mAP@0.5: 0.95分别提升了 3.7% 和 3.1%，

通过实验证明了本文模型的有效性。

在玉米病虫害检测领域上，未来仍有很大的研究空间，可探索融合多模态数据，避免单一的视觉信息在玉米病虫害早期识别和检测上的局限性，可利用多模态技术构建一个更全面的作物健康检测系统。

作者贡献

李志明：研究资料的收集、分析，实验的设计与完成，论文写作及修改；贾改秀：论文修改。

利益冲突

所有作者声明无利益冲突

参考文献

1. Wang Q H, Liu Y K, Zheng Q, et al. SMC-YOLO: a high-precision maize insect pest-detection method[J]. *Agronomy*, 2025, **15**(1): 195.
2. Pfordt A, Paulus S. A review on detection and differentiation of maize diseases and pests by imaging sensors[J]. *J Plant Dis Prot*, 2025, **132**(1): 40.
3. Gülmez B. Advancements in maize disease detection: a comprehensive review of convolutional neural networks[J]. *Comput Biol Med*, 2024, **183**: 109222.
4. Zheng H W, Wang F, Gao J B. DES-YOLO: a more accurate object detection method[J]. *Opto-Electron Eng*, 2024, **51**(11): 240212.
郑华伟, 王飞, 高建邦. DES-YOLO: 一种更精确的目标检测方法[J]. 光电工程, 2024, **51**(11): 240212.
5. Hao S N, Gao E J, Ji Z L, et al. BCS_YOLO: research on corn leaf disease and pest detection based on YOLOv11n[J]. *Appl Sci*, 2025, **15**(15): 8231.
6. Shi J, Lin S S, Luo J G, et al. Study on a detection method for crop diseases and insect pests based on YOLO v5s improved model[J]. *Jiangsu Agric Sci*, 2023, **51**(24): 175−183.
施杰, 林双双, 罗建刚, 等. 基于 YOLO v5s 改进模型的玉米作物病虫害检测方法[J]. 江苏农业科学, 2023, **51**(24): 175−183.
7. Zhang Y W, Wang X X, Fan X F. Research progress on corn and tomato diseases and pests detection technology based on deep learning[J]. *Jiangsu Agric Sci*, 2024, **52**(10): 10−20.
张友为, 王鑫鑫, 范晓飞. 基于深度学习的玉米和番茄病虫害检测技术研究进展[J]. 江苏农业科学, 2024, **52**(10): 10−20.
8. Zhu S P, Zhang Y, Tang M J, et al. Ripe fruit detection model of passion fruit based on improved YOLO v10n[J]. *Trans Chin Soc Agric Mach*, 2025, **56**(10): 549−557,595.
祝诗平, 张越, 唐茂杰, 等. 基于改进 YOLO v10n 的百香果成熟果实检测方法[J]. 农业机械学报, 2025, **56**(10): 549−557,595.
9. Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016: 779–788. https://doi.org/10.1109/CVPR.2016.91.
10. Guo L, Liao Y, Yao H Y, et al. An electrical insulator defects detection method combined human receptive field model[J]. *J Control Sci Eng*, 2018, (1): 2371825.
11. Zheng J H, Wang D Y, Geng Z X. Real-time detection of safety hazards in coal mines utilizing an enhanced YOLOv3 algorithm[J]. *Trait Signal*, 2023, **40**(4): 1565−1572.
12. Hsiao Y T, Sheu J S, Ma H. Efficient object detection and intelligent information display using YOLOv4-tiny[J]. *Adv Technol Innovation*, 2024, **9**(1): 42−49.
13. Wang S M, Xu H Y, Zhu X Z, et al. Lightweight small object detection algorithm for aerial photography based on improved YOLOv8n: PECS-YOLO[J]. *Comput Eng*, 2025, **51**(9): 280−293.
王舒梦, 徐慧英, 朱信忠, 等. 基于改进 YOLOv8n 的航拍轻量化小目标检测算法: PECS-YOLO[J]. 计算机工程, 2025, **51**(9): 280−293.
14. Feng Z, Liu F H. Balancing feature symmetry: IFEM-YOLOv13 for robust underwater object detection under degradation[J]. *Symmetry*, 2025, **17**(9): 1531.
15. Pan H, Guan S P, Zhao X Y. LVD-YOLO: an efficient lightweight vehicle detection model for intelligent transportation systems[J]. *Image Vision Comput*, 2024, **151**: 105276.
16. Wei M, Zhan W. YOLO_MRC: a fast and lightweight model for real-time detection and individual counting of Tephritidae pests[J]. *Ecol Inf*, 2024, **79**: 102445.
17. Dong D H, Li J Z, Liu H Y, et al. EA-YOLO: an efficient and accurate UAV image object detection algorithm[J]. *IEEJ Trans Electr Electron Eng*, 2025, **20**(1): 61−68.
18. Maginga T J, Masabo E, Bakunzibake P, et al. Using wavelet transform and hybrid CNN – LSTM models on VOC & ultrasound IoT sensor data for non-visual maize disease detection[J]. *Heliyon*, 2024, **10**(4): e26647.
19. Wang Z L, Fan S X, An T, et al. Detection of insect-damaged maize seed using hyperspectral imaging and hybrid 1D-CNN-BiLSTM model[J]. *Infrared Phys Technol*, 2024, **137**: 105208.
20. Liu Y A, Yan F, Li S Y, et al. SMSERT-DETR: spatial multi-scale efficient real-time detector for maize disease detection[J]. *J Supercomput*, 2025, **81**(8): 880.
21. Cheng D Y, Zhao W, He W D, et al. Identification method of crop diseases and insect pests based on improved YOLOv5n model[J]. *Jiangsu J Agric Sci*, 2024, **40**(11): 2021−2031.
承达瑜, 赵伟, 何伟德, 等. 基于改进 YOLOv5n 模型的农作物病虫害识别方法[J]. 江苏农业学报, 2024, **40**(11): 2021−2031.
22. Shi J, Xiong K X, Li Z, et al. Maize pest and disease detection system based on lightweight improved YOLOv8 model and edge computing[J]. *Jiangsu J Agric Sci*, 2025, **41**(2): 313−322.
施杰, 熊凯祥, 李志, 等. 基于轻量化改进 YOLOv8 模型和边缘计算的玉米病虫害检测系统[J]. 江苏农业学报, 2025, **41**(2): 313−322.
23. 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 Trans Pattern Anal Mach Intell*, 2018, **40**(4): 834−848.
24. Tan M X, Pang R M, Le Q V. EfficientDet: scalable and efficient object detection[C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 2020: 10778–10787. https://doi.org/10.1109/CVPR42600.2020.01079.
25. Dai J F, Qi H Z, Xiong Y W, et al. Deformable convolutional networks[C]//*Proceedings of 2017 IEEE International Conference on Computer Vision*, Venice, Italy, 2017: 764–773. https://doi.org/10.1109/ICCV.2017.89.
26. Dahiya N, Prakash D, Kundu S, et al. Optimised RFO tuned RF-DETR model for precision urine microscopy for renal and systemic disease diagnosis[J]. *Sci Rep*, 2025, **15**(1): 25842.
27. Liu C, Zhang Y L, Shen J F, et al. Improved RT-DETR for infrared ship detection based on multi-attention and feature fusion[J]. *J Mar*

Sci Eng, 2024, **12**(12): 2130.

28. Xiong E J, Zhang R F, Liu Y H, et al. Ghost-YOLOv8 detection algorithm for traffic signs[J]. *Comput Eng Appl*, 2023, **59**(20): 200−207.
熊恩杰, 张荣芬, 刘宇红, 等. 面向交通标志的 Ghost-YOLOv8 检测算法[J]. 计算机工程与应用, 2023, **59**(20): 200−207.

作者简介



【通信作者】李志明 (1980-)，男，硕士，副教授，主要研究方向为计算机视觉、计算机网络应用。
E-mail：zhangyelzm@163.com



贾改秀 (1986-)，女，硕士，副研究员，主要研究方向为制种玉米连作障碍。
E-mail：415173205@qq.com

MSAF-YOLO: A corn pest and disease detection algorithm based on YOLOv11

Li Zhiming[1*], Jia Gaixiu[2]

[1] School of Information Technology and Media, Hexi College, Zhangye, Gansu 734000, China;
[2] Zhangye Academy of Agricultural Sciences, Zhangye, Gansu 734000, China

Abstract

Objective Corn, as one of the world's most vital staple crops, plays an indispensable role in agricultural production, industrial applications, and economic development. In China, corn planting area and yield currently account for over 40% of total grain output, a proportion that continues to grow steadily, underscoring its strategic significance for national food security and economic stability. However, throughout the growth cycle, corn is persistently threatened by various pests and diseases, including corn borer, leaf spot, rust, stalk rot, and armyworm, which severely restrict both yield and quality. Traditional pest and disease detection methods rely heavily on manual field inspection—a labor-intensive, time-consuming, and inefficient approach that struggles to meet the demands of large-scale cultivation. Moreover, manual detection is inherently subjective and discontinuous, often resulting in missed diagnoses and misjudgments. Consequently, developing an automated, efficient, and accurate detection method for corn pests and diseases has become an urgent necessity.

In recent years, deep learning-based object detection technologies have achieved remarkable success across various domains, with the YOLO (You Only Look Once) series of algorithms gaining particular prominence due to their exceptional balance of detection speed and accuracy. YOLOv11, the latest iteration in this series, incorporates further optimizations in training strategies, network architecture, and detection precision, providing a solid technical foundation for agricultural pest identification tasks. Nevertheless, when applied to corn pest detection in real-world field scenarios, mainstream models including YOLOv11 continue to exhibit critical limitations: 1) Complex environmental interference—the intricate backgrounds, variable lighting conditions, diverse pest morphologies, and frequent occlusions characteristic of field environments substantially impair the model's ability to accurately localize and recognize targets; 2) Small-target detection difficulty—early-stage or physically small pests and disease lesions occupy extremely limited pixel regions in images and display indistinct features, leading to elevated rates of missed detections and false negatives.

Methods To address these challenges, this paper proposes an improved YOLOv11 model designated as MSAF-YOLO (Multi-Scale Adaptive Feedback YOLO). The core innovation lies in reconstructing the backbone network using a novel MSAF-Net (Multi-Scale Adaptive Feedback Network) architecture, which synergistically integrates two specially designed modules: the Hierarchical Feature Feedback Module (HFFM) and the Adaptive Dynamic Receptive Field Module (ADRF).

Foundation item: Gansu Provincial Natural Science Foundation (23JRRG0012, 23JRRG0036), Gansu Provincial Higher Education Faculty Innovation Fund (2025B-193), Ministry of Education Industry-Academia Collaboration Project for Collaborative Talent Development (231005634170926), Gansu Province 2025 Higher Education Teaching Reform Research Project (GJJGB295), and Ministry of Education Supply-Demand Matching Employment Education Project (2024011199167)

*Correspondence: Li Z M, E-mail: zhangyelzm@163.com

The Hierarchical Feature Feedback Module addresses the fundamental problem of semantic information deficiency in shallow feature layers and the progressive loss of small-target features during repeated downsampling operations. Conventional feedforward convolutional networks propagate information unidirectionally from shallow to deep layers; while deep layers acquire rich semantic representations, shallow layers remain dominated by fine-grained spatial details but lack high-level semantic context. HFFM introduces residual-style feedback pathways coupled with channel attention mechanisms to establish cross-layer feature calibration and enhancement. Specifically, deep feature maps containing abstracted semantic information are selectively fed back and fused with corresponding shallow feature maps through learnable gating mechanisms. The channel attention component adaptively recalibrates feature responses, emphasizing informative channels while suppressing less useful ones. This feedback-driven feature refinement significantly enhances the model's capacity to extract and preserve discriminative characteristics of small pests and incipient disease symptoms, effectively mitigating the information attenuation typically observed across deep convolutional architectures.

The Adaptive Dynamic Receptive Field Module confronts the inherent inflexibility of standard convolutional neural networks, which employ fixed geometric receptive fields that remain constant regardless of input content. Such rigidity proves particularly disadvantageous when processing agricultural images containing objects at vastly different scales—from millimeter-scale rust pustules to centimeter-scale expanded leaf spot lesions—under heterogeneous background conditions. ADRF achieves pixel-level receptive field adaptation through an elegantly designed integration of spatially adaptive weights and multi-branch dilated convolutions. For each spatial location in the feature map, the module dynamically computes context-dependent weights that govern the relative contributions of parallel convolutional branches employing different dilation rates. This mechanism enables the network to expand its receptive field when processing large contextual regions while maintaining fine resolution for small-target localization, all modulated by the local structural characteristics of the input features. The resulting adaptive receptive field substantially improves model robustness and generalization across diverse pest and disease categories, growth stages, and environmental conditions.

Results and Discussions The proposed MSAF-YOLO model was rigorously evaluated on two datasets: a self-constructed corn pest and disease dataset (CORN) encompassing representative categories including corn borer, leaf spot, and rust under various field conditions, and the publicly available general plant disease dataset (diseases) to assess cross-domain generalization capability. Experimental results demonstrate substantial and consistent improvements across multiple evaluation metrics. On the CORN dataset, MSAF-YOLO achieved mean Average Precision at 50% IoU (mAP50) of 85.7% and mAP50:95 of 68.9%, representing significant gains of 3.2 and 2.7 percentage points respectively over the baseline YOLOv11 model. On the diseases dataset, mAP50 and mAP50:95 reached 80.9% and 62.8%, improving by 3.7% and 3.1% respectively. Notably, these accuracy enhancements were accomplished alongside parameter optimization, demonstrating that MSAF-YOLO achieves a favorable trade-off between detection precision and computational efficiency.

Conclusions In conclusion, this research makes three principal contributions: 1) it systematically identifies and characterizes the specific challenges confronting corn pest and disease detection in complex field environments, particularly the coupled difficulties of small-target detection and background interference; 2) it proposes two novel modules—HFFM and ADRF—that respectively address semantic feature deficiency and fixed receptive field limitations through innovative feedback mechanisms and dynamic convolution strategies; 3) it delivers a practically deployable model that achieves state-of-the-art detection performance on corn pest and disease tasks while maintaining architectural efficiency suitable for real-time agricultural applications. The proposed MSAF-YOLO framework not only advances the methodological frontier of object detection in agricultural contexts but also offers a viable technical pathway toward intelligent, automated crop health monitoring systems.

Keywords: corn pest and disease detection; YOLOv11; multi-scale features; convolutional neural network

Li Z M, Jia G X. MSAF-YOLO: A corn pest and disease detection algorithm based on YOLOv11[J]. *Opto-Electron Eng*, 2026, **53**(6): 250353; DOI: 10.12086/oee.2026.250353

Opto-Electronic Journals Group
www.oejournal.org

扫描二维码，获取 PDF 全文