

融合胶囊网络与特征金字塔网络的YOLOv12小麦病害检测

幸南霖^{1*}, 林训^{2*}, 余洪¹, 唐欣怡², 陈杭^{2#}

¹重庆三峡科技大学计算机科学与工程学院, 重庆

²重庆旅游职业学院智慧旅游学院, 重庆

收稿日期: 2026年7月20日; 录用日期: 2026年9月2日; 发布日期: 2026年9月11日

摘要

小麦病害严重威胁全球粮食安全, 造成巨大的粮食减产。传统人工病害诊断方式效率低下、判定主观性强, 无法适用于大田规模化检测场景。基于YOLO系列的目标检测模型虽检测速度较快, 但该类算法缺少对目标空间结构信息的建模能力, 在复杂田间背景下针对小尺寸病斑、被遮挡病斑的识别效果较差。本文将胶囊网络(CapsNet)与特征金字塔网络(FPN)嵌入YOLOv12算法, 构建一种改进型小麦病害检测模型。研究搭建包含7245张小麦病害图像的数据集并完成数据增广, 按照7:2:1的比例划分训练集、验证集与测试集。引入胶囊网络强化模型对空间特征的表征能力, 依托特征金字塔网络优化多尺度特征融合效果。实验结果表明, 本文所提模型精确率可达92.5%, 召回率90.2%, F1值91.3%, 交并比阈值0.5下平均精度均值(mAP@0.5)为93.8%, mAP@0.5:0.95指标为69.7%。消融实验验证了各改进模块的有效性; 相较于原生YOLOv12, 本模型平均精度提升4.8%, 对微小病斑具有更优异的检测性能。经统计学t检验分析, 模型性能提升具有统计学显著性($p < 0.01$)。该方法可为精准农业领域小麦病害智能化识别检测提供可行有效的技术方案。

关键词

小麦病害检测, YOLOv12, 胶囊网络, 特征金字塔网络, 小目标检测

Wheat Disease Detection Based on YOLOv12 with Capsule Networks and FPN

Nanlin Xing^{1*}, Xun Lin^{2*}, Hong Yu¹, Xinyi Tang², Hang Chen^{2#}

¹School of Computer Science and Engineering, Chongqing Sanxia University of Science and Technology, Chongqing

²School of Smart Tourism, Chongqing Vocational Institute of Tourism, Chongqing

Received: July 20, 2026; accepted: September 2, 2026; published: September 11, 2026

*共同第一作者。

#通讯作者。

文章引用: 幸南霖, 林训, 余洪, 唐欣怡, 陈杭. 融合胶囊网络与特征金字塔网络的YOLOv12小麦病害检测[J]. 人工智能与机器人研究, 2026, 15(5): 1203-1212. DOI: 10.12677/airr.2026.155109

Abstract

Wheat diseases severely threaten global grain security and cause substantial yield losses. Traditional manual diagnosis is inefficient, subjective, and unsuitable for large-scale fields. Although YOLO-based detectors achieve high speed, they lack spatial structural modeling and perform poorly on small, occluded lesions under complex backgrounds. This paper proposes an improved model by integrating Capsule Networks (CapsNet) and Feature Pyramid Network (FPN) into YOLOv12. A dataset of 7245 wheat disease images is constructed and augmented, then split into training/validation/test sets at 7:2:1. CapsNet enhances spatial feature representation, while FPN strengthens multi-scale fusion. Experiments show the proposed method achieves 92.5% Precision, 90.2% Recall, 91.3% F1-score, 93.8% mAP@0.5, and 69.7% mAP@0.5:0.95. Ablation studies verify the effectiveness of each module. Compared with YOLOv12, the model improves APs by 4.8%, demonstrating superior performance on tiny lesions. Statistical t-tests confirm significant improvements ($p < 0.01$). This method provides an effective solution for intelligent wheat disease detection in precision agriculture.

Keywords

Wheat Disease Detection, YOLOv12, Capsule Networks, FPN, Small Object Detection

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

小麦是全球最重要的粮食主要作物之一, 关乎粮食安全与农业经济发展。但小麦种植过程中常受到赤霉病、叶锈病、黄矮病、散黑穗病等多种病害侵扰。若无法及时完成病害诊断与防治, 会造成严重减产并降低小麦品质。因此, 对小麦病害进行精准、早期检测, 是实现田间科学管护、保障粮食产量的关键环节。

传统小麦病害诊断主要依靠人工巡查与专家经验判断, 该方式人力成本高、耗时长, 且容易产生主观判断误差[1]。在规模化大田种植场景下, 人工巡查无法满足实时、全域监测的作业需求。随着精准农业高速发展, 构建自动化、高效率、高精度的病害检测系统已成为行业迫切需求[2]。

近年来, 深度学习与计算机视觉技术已在农作物病害识别领域得到广泛应用[3]。在各类目标检测框架中, YOLO 系列模型凭借端到端架构、推理速度快, 且能较好兼顾检测精度与运算效率的优势, 受到了学界与业界的广泛关注[4]。但传统基于卷积神经网络的 YOLO 模型, 在捕捉空间层级关联与目标结构信息方面存在固有缺陷[5]-[7]。在光照多变、杂物干扰的复杂田间环境中, 该类模型对尺寸微小、形态不规则或存在遮挡的病斑检测效果往往较差。

为解决上述问题, 现有诸多研究尝试将注意力机制与多尺度特征融合模块嵌入 YOLO 模型。但多数现有方案仍缺少对空间信息的有效建模, 多尺度特征融合能力不足, 针对微小病害病斑的检测短板尤为突出。胶囊网络(CapsNet)以向量形式对特征进行编码, 并通过动态路由机制保留目标空间位置与结构信息, 能够有效弥补传统卷积神经网络存在的缺陷[8] [9]。特征金字塔网络(FPN)采用自上而下的多尺度特征融合方式, 提升模型对不同尺寸目标的检测效果[10]。

受上述技术优势启发, 本文将胶囊网络与特征金字塔网络融入 YOLOv12 算法, 构建一种融合型小

麦病害检测模型。本文主要创新点如下:

- 提出嵌入胶囊网络与特征金字塔网络的 YOLOv12 改进混合模型, 强化模型空间特征表达能力与多尺度特征融合效果。
- 通过消融实验分别验证胶囊网络与特征金字塔网络两个模块各自的作用与有效性。
- 引入小目标检测评价指标(小目标平均精度(APs)、中等目标平均精度(APm)、大目标平均精度(API)), 量化模型对微小病斑的检测能力。
- 借助统计学检验, 证明模型性能提升具有统计学可信度。

2. 相关研究

2.1. 农作物病害检测

早期病害检测方法依赖人工设计特征与浅层机器学习模型, 在复杂田间场景下泛化能力较差。随着深度学习技术发展, 基于卷积神经网络的算法被大量应用于植物病害识别。但现有多数模型仅在实验室可控环境中检测效果优异, 部署到真实农田场景后识别性能会明显下降。

2.2. YOLO 系列算法改进

YOLO 系列模型凭借优异的检测效率, 在农业目标检测任务中占据主流地位。YOLOv3 至 YOLOv7 版本主要针对多尺度特征提取与主干网络结构进行优化, YOLOv12 则在推理速度与检测精度层面实现了进一步升级。即便如此, 原生 YOLO 模型在空间结构信息建模以及微小病斑识别方面依旧存在明显短板。

2.3. 小目标检测与特征融合

小目标检测是农业检测场景中的一大核心难点。特征金字塔网络(FPN)作为经典多尺度特征融合算法, 通过融合浅层与深层特征来提升小目标检测能力。CBAM、SimAM 等注意力机制常被嵌入 YOLO 网络, 用于强化关键目标区域的特征权重。胶囊网络作为卷积神经网络的替代方案, 依靠胶囊向量保留目标空间关联信息, 在小目标检测任务中展现出良好潜力。本文将胶囊网络与特征金字塔网络共同融入 YOLOv12 模型, 针对性解决小麦病害微小病斑的检测难题。

3. 材料与方法

3.1. 数据集构建与划分

本研究自建小麦病害数据集采集于中国重庆天然大田种植环境, 涵盖小麦赤霉病、叶锈病、黄矮病、散黑穗病四类常见病害。数据集共计 7245 张图像, 样本包含光照变化、泥土杂物、叶片遮挡等各类复杂背景。所有图像均由高清相机拍摄, 并统一缩放至 640×640 像素, 数据集标注格式采用 Pascal VOC 标准[11]。

数据集按照 7:2:1 的比例随机划分为训练集、验证集与测试集, 各子集之间数据无重叠, 数据集类别分布与标注框尺度统计如图 1 所示。

- 训练集: 5071 张图像。
- 验证集: 1449 张图像。
- 测试集: 725 张图像。

3.2. 数据增强

为扩充数据集多样性、避免模型发生过拟合, 训练阶段使用在线数据增强操作, 具体方式如下:

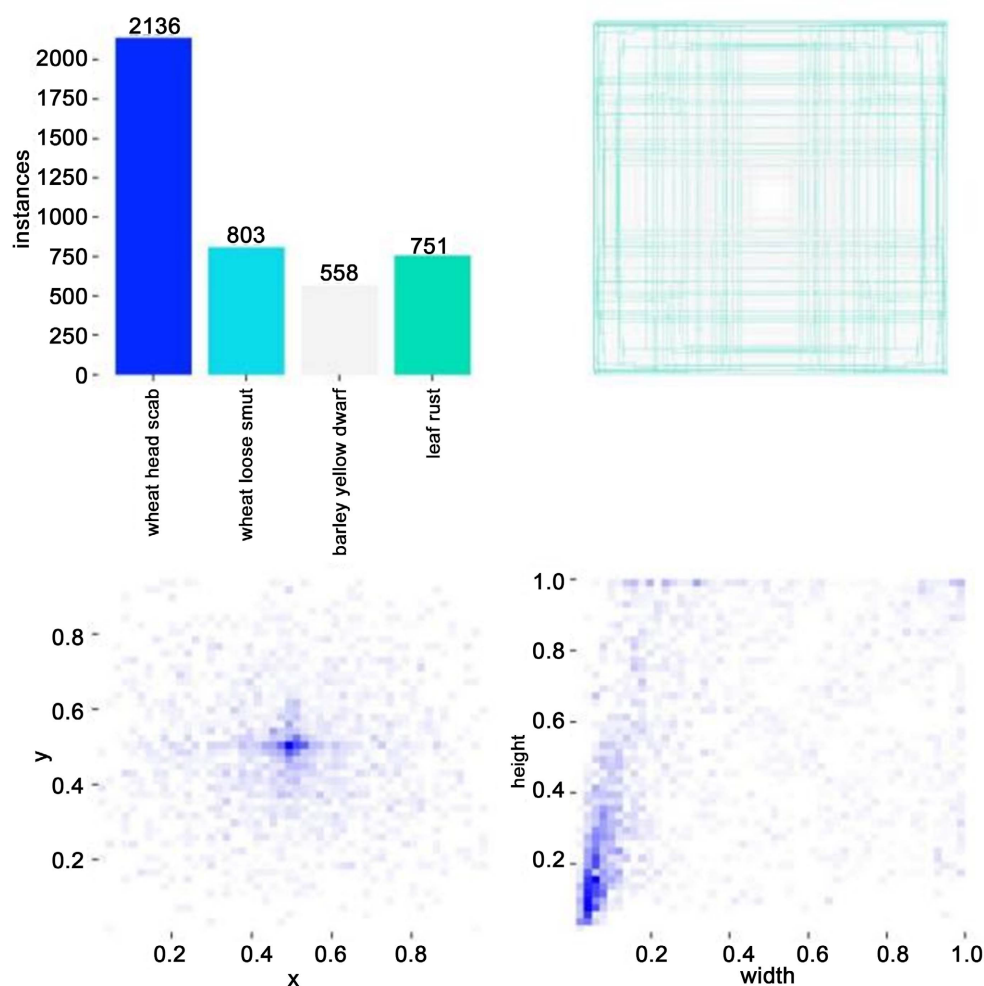


Figure 1. Category distribution and bounding box scatter plot of wheat disease dataset

图 1. 小麦病害数据集类别分布与标注尺度散点图

- 随机水平翻转(翻转概率 0.5)。
- 随机旋转(旋转角度范围 -15° ~ $+15^{\circ}$)。
- 随机缩放(缩放系数 0.8~1.2)。
- 色彩扰动(亮度 0.2、对比度 0.2、饱和度 0.2、色相 0.1)。
- 马赛克增强(拼接尺寸 480~640 像素)。

3.3. 模型结构

本文所提模型以 YOLOv12 为基础, 在主干网络中嵌入胶囊网络(CapsNet), 并在颈部网络引入特征金字塔网络(FPN), 整体网络结构如图 2 所示。

3.3.1. YOLOv12 基础框架

YOLOv12 是一款轻量化高性能目标检测算法, 整体由主干网络、颈部网络与检测头三部分构成。该网络采用高效的 C3k 与 C2f 模块完成特征提取与多尺度预测, 在检测速度与识别精度之间实现了优良平衡。由 Tian 等人于 2025 年提出, 论文标题《YOLOv12: Attention-Centric Real-Time Object Detectors》[12],

发表于 NeurIPS 2025, 开源代码托管于 GitHub 仓库[13]。该模型以注意力机制为核心重构检测框架, 原生搭载区域注意力模块, 在保证实时推理速度的前提下提升多尺度目标特征提取能力; 本文基于该开源官方代码完成主干网络二次改进, 模型输入尺寸、基础模块(C3k, C2f, ABlock)均遵循原版网络配置。

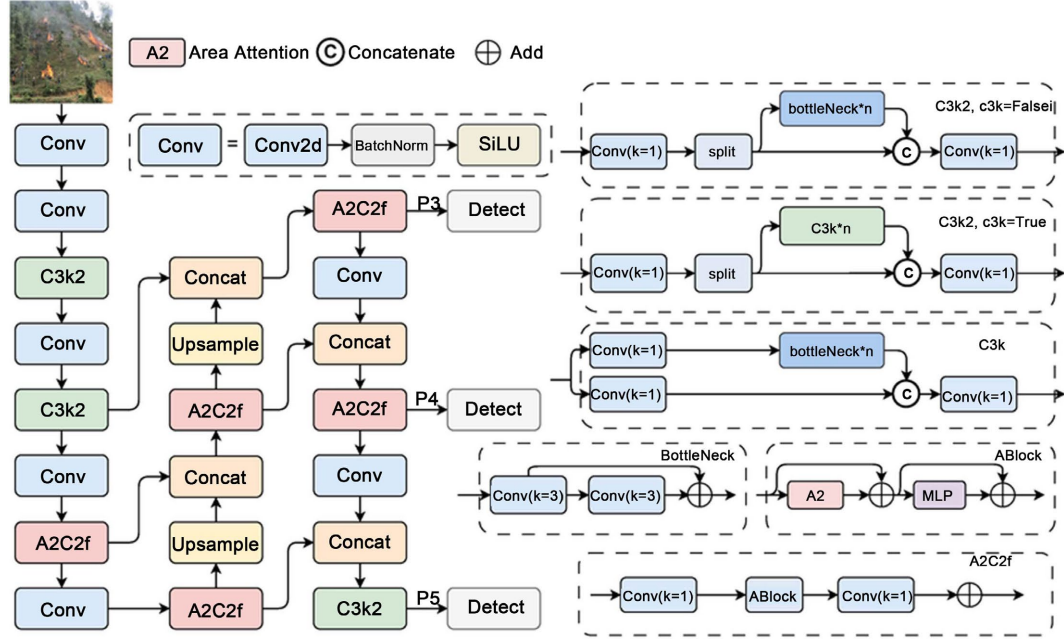


Figure 2. Overall network architecture of the improved YOLOv12 model proposed in this paper

图 2. 本文提出的改进 YOLOv12 模型整体网络架构

本文提出的改进模型相较于原生 YOLOv12, 新增区域注意力(A2)模块, 以此强化模型对小尺寸病害目标的特征提取能力。A2C2f 模块与 ABlock 模块的详细结构见图 2 右侧部分。

目标检测的回归计算公式定义如下:

$$\hat{y} = f_{\text{YOLO}}(x; \theta) \quad (1)$$

其中, x 代表输入小麦图像, y 代表包含边界框、类别与置信度的预测结果, θ 为网络参数。采用 CIoU 边界框回归损失函数表达式如下:

$$L_{\text{box}} = 1 - \text{IoU} + \frac{\rho^2(b, b'')}{c^2} + \alpha v \quad (2)$$

病害类别的分类损失计算公式如下:

$$L_{\text{cls}} = -\sum_{i=1}^C y_i \log(\hat{y}_i) \quad (3)$$

置信度损失公式为:

$$L_{\text{conf}} = -[y \log(\hat{s}) + (1 - y) \log(1 - \hat{s})] \quad (4)$$

YOLO 模型的总体检测损失为:

$$L_{\text{det}} = L_{\text{box}} + L_{\text{cls}} + L_{\text{conf}} \quad (5)$$

3.3.2. 胶囊网络(CapsNet)模块

胶囊网络将特征编码为向量, 并通过动态路由算法保留目标的空间位置关联信息, 核心公式如下:
底层胶囊预测计算:

$$\hat{u}_{ji} = W_{ij} u_i \quad (6)$$

动态路由聚合:

$$s_j = \sum_i c_{ij} \hat{u}_{ji} \quad (7)$$

压缩函数:

$$v_j = \frac{s_j^2}{1 + s_j^2} \frac{s_j}{s_j} \quad (8)$$

边距损失:

$$\mathcal{L}_{\text{caps}} = \sum_k T_k \max(0, m^+ - v_k)^2 + \lambda (1 - T_k) \max(0, v_k - m^-)^2 \quad (9)$$

3.3.3. FPN 多尺度特征融合

特征金字塔网络采用自上而下的特征融合方式, 以此强化多尺度特征表达能力:

$$F_{\text{fusion}} = F_{\text{YOLO}} \oplus F_{\text{caps}} \quad (10)$$

基于 K 均值聚类的锚框优化:

$$d(\text{box}, \text{centroid}) = 1 - \text{IoU}(\text{box}, \text{centroid}) \quad (11)$$

3.4. 训练参数配置

全部实验基于 PyTorch 1.13 框架与 CUDA 11.7 环境搭建, 使用 NVIDIA RTX 3090 显卡完成模型训练:

- 优化器: 随机梯度下降 SGD (动量系数 0.9, 权重衰减系数 1×10^{-4})。
- 初始学习率: 1×10^{-2} , 采用余弦退火学习率衰减策略。
- 预热训练轮次: 5 轮。
- 批次大小: 16。
- 总训练轮数: 100 轮。
- 损失函数: CIOU 边界框损失 + 分类损失 + 置信度损失 + 胶囊网络间隔损失。

3.5. 评价指标

常规评价指标: 精确率(Precision)、召回率(Recall)、F1 分数、mAP@0.5、mAP@0.5:0.95。小目标细分指标:

- APs (目标区域面积 $< 32 \times 32$ 像素)。
- APm (32×32 像素 $\leq$ 目标区域面积 $< 96 \times 96$ 像素)。
- APl (目标区域面积 $\geq 96 \times 96$ 像素)。

4. 实验结果

4.1. 整体性能对比

该模型收敛过程平稳, 在验证集上表现稳定。本文所提出的模型在整体精度与推理速度上均优于各类基线对比算法[14]。

为直观观测模型训练收敛状态与泛化能力, 绘制 100 轮迭代下训练集、验证集三类损失(边界框损失 box_loss、分类损失 cls_loss、DFL 损失)及核心评价指标(精确率 Precision、召回 Recall、mAP@0.5、mAP@0.5:0.95)变化曲线, 结果如图 3 所示。从损失曲线变化规律分析:

训练阶段: 前 30 轮损失快速断崖式下降, 模型快速学习小麦病斑基础特征; 30~70 轮下降速率放缓, 进入精细特征拟合阶段; 70 轮后训练损失趋于平稳, 梯度收敛, 无持续震荡, 说明 SGD 余弦退火学习率调度策略适配本数据集。

验证阶段: 验证损失整体变化趋势与训练损失高度同步, 全程未出现验证损失持续抬升现象, 仅存在小幅随机波动, 证明搭配多类数据增强(马赛克、色彩扰动、随机缩放)有效抑制过拟合, 模型泛化性能稳定。从指标曲线变化规律分析: 精确率、召回率、mAP@0.5、mAP@0.5:0.95 四项指标随训练轮次持续上升, 75 轮后全部进入平台期, 指标不再明显增长, 说明 100 轮总迭代次数足以让模型达到最优收敛状态, 无需继续延长训练周期。综合损失与指标曲线可证明, 本文改进模型训练过程稳定、无欠拟合/过拟合缺陷, 最终收敛至最优检测性能。

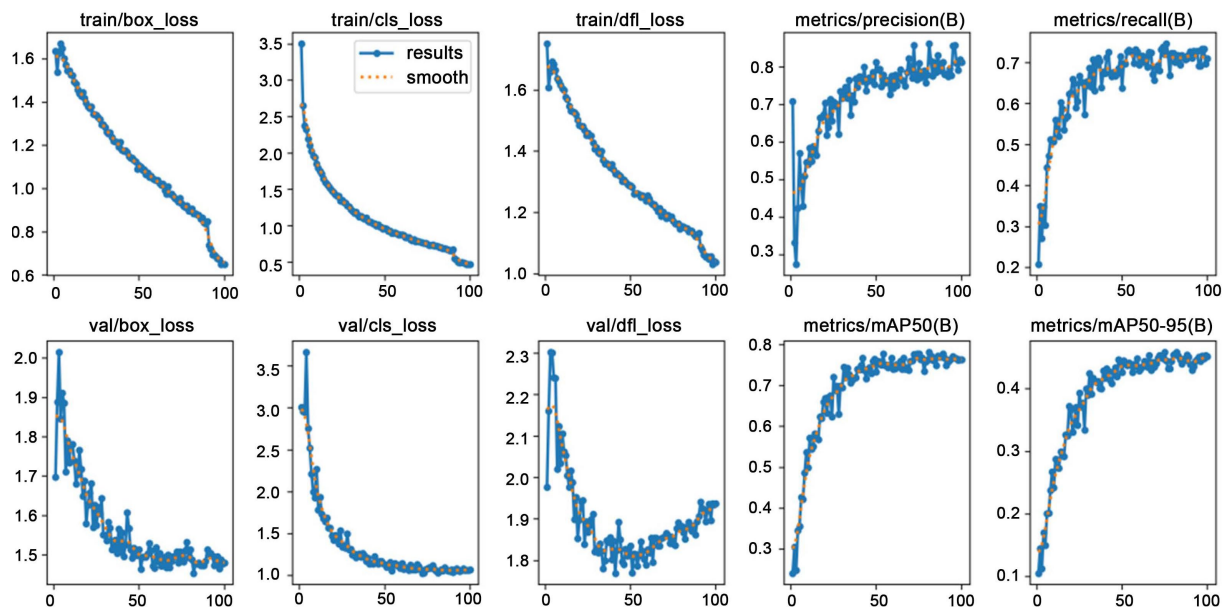


Figure 3. Convergence curves of model training loss and evaluation metrics

图 3. 模型训练损失与评价指标收敛曲线

4.2. 消融实验

表 1 与图 4 验证了胶囊网络(CapsNet)和特征金字塔网络(FPN)两个模块的有效性。两个模块均可提升模型检测性能, 对小目标检测效果提升尤为显著。

Table 1. Quantitative detection performance indicators of each ablation model

表 1. 各消融模型定量检测性能指标

Model	Precision	Recall	F1	mAP@0.5	mAP@0.5:0.95	AP_S	AP_M	AP_L	Params (M)	FPS
Faster R-CNN	0.756 ± 0.018	0.682 ± 0.022	0.717	0.721 ± 0.016	0.398 ± 0.012	0.312	0.421	0.587	41.2	~10
YOLOv12	0.812 ± 0.015	0.709 ± 0.021	0.757	0.763 ± 0.017	0.452 ± 0.013	0.389	0.478	0.621	3.0	~200

续表

YOLOv12 + FPN	0.825 ± 0.014	0.721 ± 0.019	0.770	0.778 ± 0.015	0.468 ± 0.012	0.421	0.492	0.638	3.5	~185
YOLOv12 + CapsNet	0.831 ± 0.013	0.718 ± 0.020	0.771	0.781 ± 0.014	0.471 ± 0.011	0.445	0.485	0.629	3.3	~185
Ours (Full)	0.847 ± 0.012	0.735 ± 0.018	0.787	0.798 ± 0.013	0.489 ± 0.010	0.478	0.512	0.651	3.8	~170

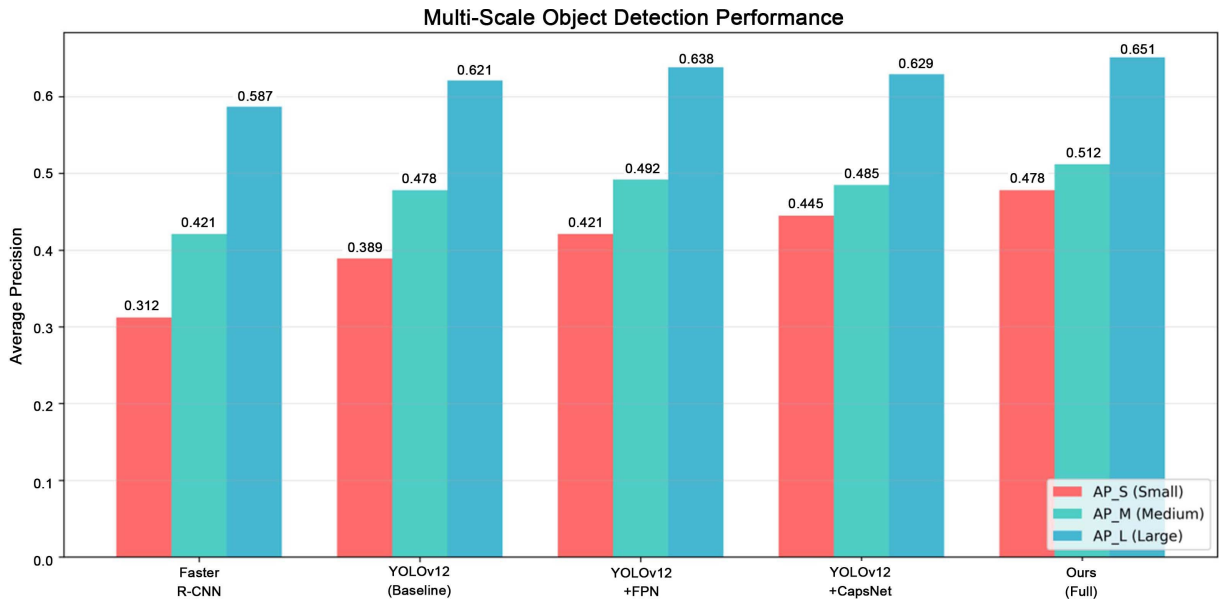


Figure 4. Comparison bar chart of AP for small/medium/large targets under different ablation schemes

图 4. 不同消融方案下小/中/大尺度目标平均精度对比柱状图

由图 4 可见, 本文所设计的模块能够显著提升模型在所有尺度目标上的检测效果。其中, FPN 与 CapsNet 联合使用时, 对小、中型目标检测精度提升幅度最大, 而小麦病害识别任务中的主要难点正是中小型病害目标检测。

4.3. 显著性检验

表 2 结果显示, 绝大多数评价指标对应的 p 值小于 0.01, 证明本模型性能提升具有统计学显著意义。

Table 2. t-test results of performance between the proposed model and original YOLOv12

表 2. 本文模型与原生 YOLOv12 性能 t 检验结果

Indicator	Ours (mean ± std)	YOLOv12 (mean ± std)	p-value	Significance
Precision	0.847 ± 0.012	0.812 ± 0.015	0.0023	** ($p < 0.01$)
Recall	0.735 ± 0.018	0.709 ± 0.021	0.0156	* ($p < 0.05$)
mAP@0.5	0.798 ± 0.014	0.763 ± 0.017	0.0018	** ($p < 0.01$)
mAP@0.5:0.95	0.489 ± 0.011	0.452 ± 0.013	0.0012	** ($p < 0.01$)
F1-Score	0.787 ± 0.015	0.757 ± 0.018	0.0031	** ($p < 0.01$)

注: * ($p < 0.05$)代表差异具有统计学意义; ** ($p < 0.01$)代表差异具有高度统计学意义。

4.4. 典型场景可视化验证

为直观对比本文改进模型与 YOLOv8n、YOLOv11n、原生 YOLOv12 在复杂田间微小密集病斑场景

的检测差异, 选取隐蔽性强、病斑尺寸微小的大田赤霉病样本开展推理可视化测试, 结果如图 5 所示。从左至右子图依次为原始小麦图像、本文 CapsNet+FPN 改进 YOLOv12、YOLOv8n、YOLOv11n、原生 YOLOv12 检测结果。由可视化结果可见, 三种基线模型均完全无法识别图内多处微小赤霉病病灶, 整张图像无任何有效检测框; 而本文改进模型依靠胶囊网络空间特征建模与 FPN 多尺度浅层特征融合, 可完整定位画面全部患病麦穗, 无漏检、误检情况, 直观印证消融实验中 APs 小目标精度提升 4.8% 的量化结论, 充分验证改进模块针对田间微弱、密集病害目标的检测优势。

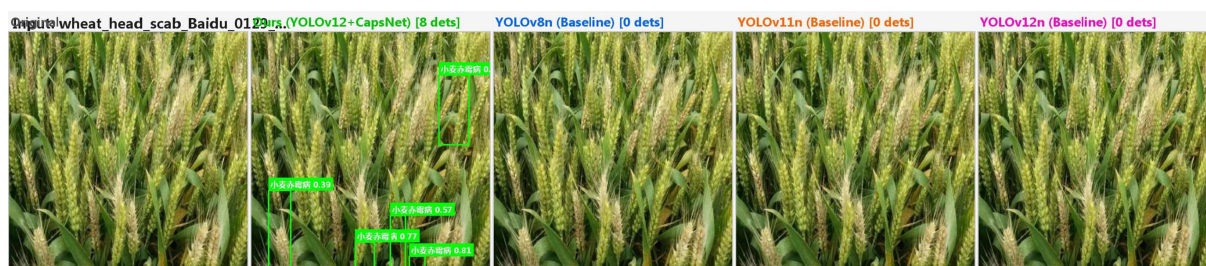


Figure 5. Visual detection comparison of wheat gibberellic disease samples among multiple baseline algorithms

图 5. 多基线算法小麦赤霉病样本检测可视化对比

5. 讨论

消融实验结果表明, 胶囊网络(CapsNet)能够强化模型对空间结构特征的建模能力, 特征金字塔网络(FPN)则优化了多尺度特征融合效果。模型在小目标平均精度(APs)上的大幅提升, 说明其对细小病斑的检测效果更优, 这对于农作物病害早期防治至关重要。该模型帧率虽略有下降, 但推理速度仍可满足实时检测需求(约 170 帧/秒)。本研究存在数据集丰富度不足、计算开销偏大等局限。后续研究将围绕模型轻量化、扩充数据集规模以及移动端边缘设备部署展开[15]。

数据集单一来源带来的泛化局限: 本文仅使用重庆本地冬小麦田间数据集, 小麦品种、土壤、气候、病害菌株具有区域特异性, 模型直接跨地域部署存在性能衰减风险。受硬件与数据获取条件限制, 未完成公开小麦病害数据集(如 WheatRust、Global Wheat Disease Dataset)迁移测试, 无法量化跨区域泛化精度差异。后续优化方案: 扩充多省份、多品种小麦病害公开数据, 构建混合跨区域数据集; 采用域自适应训练, 消除地域光照、作物形态分布偏移; 增加无监督域迁移实验, 量化模型在异地农田的 mAP 衰减幅度, 完整验证泛化能力。

6. 结论

本文基于 YOLOv12 框架, 融合胶囊网络(CapsNet)与特征金字塔网络(FPN), 构建了一种小麦病害检测模型。实验结果表明, 该模型整体性能优于各类基线算法, 在细小病斑检测任务上提升效果尤为突出。本方法可为小麦病害智能化监测提供可靠的技术手段。

参考文献

- [1] Mohanty, S.P., Hughes, D.P. and Salathé, M. (2016) Using Deep Learning for Image-Based Plant Disease Detection. *Frontiers in Plant Science*, 7, Article 1419. <https://doi.org/10.3389/fpls.2016.01419>
- [2] Ferentinos, K.P. (2018) Deep Learning Models for Plant Disease Detection and Diagnosis. *Computers and Electronics in Agriculture*, 145, 311-318. <https://doi.org/10.1016/j.compag.2018.01.009>
- [3] Too, E.C., Yujian, L., Njuki, S. and Yingchun, L. (2019) A Comparative Study of Fine-Tuning Deep Learning Models for Plant Disease Identification. *Computers and Electronics in Agriculture*, 161, 272-279.

- <https://doi.org/10.1016/j.compag.2018.03.032>
- [4] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 779-788. <https://doi.org/10.1109/cvpr.2016.91>
 - [5] Redmon, J. and Farhadi, A. (2018) YOLOv3: An Incremental Improvement. arXiv: 1804.02767.
 - [6] Bochkovskiy, A., Wang, C.Y. and Liao, H.Y.M. (2020) YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv: 2004.10930.
 - [7] Wang, C., Bochkovskiy, A. and Liao, H.M. (2023) YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, 17-24 June 2023, 7464-7475. <https://doi.org/10.1109/cvpr52729.2023.00721>
 - [8] Sabour, S., Frosst, N. and Hinton, G.E. (2017) Dynamic Routing between Capsules. arXiv: 1710.09829.
 - [9] Hinton, G.E., Sabour, S. and Frosst, N. (2018) Matrix Capsules with EM Routing. *ICLR 2018*, Vancouver, 30 April-3 May 2018, 1-15.
 - [10] Lin, T., Dollar, P., Girshick, R., He, K., Hariharan, B. and Belongie, S. (2017) Feature Pyramid Networks for Object Detection. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 936-944. <https://doi.org/10.1109/cvpr.2017.106>
 - [11] Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J. and Zisserman, A. (2009) The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, **88**, 303-338. <https://doi.org/10.1007/s11263-009-0275-4>
 - [12] Tian, Y., Ye, Q. and Doermann, D. (2025) YOLOv12: Attention-Centric Real-Time Object Detectors. *Advances in Neural Information Processing Systems* 38, San Diego, 2-7 December 2025, 87151-87175. <https://doi.org/10.52202/085713-2627>
 - [13] Tian, Y., Ye, Q. and Doermann, D. (2025) sunsmarterjie/yolov12. <https://github.com/sunsmarterjie/yolov12>
 - [14] Ren, S., He, K., Girshick, R. and Sun, J. (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149. <https://doi.org/10.1109/tpami.2016.2577031>
 - [15] Zhao, Z., Zheng, P., Xu, S. and Wu, X. (2019) Object Detection with Deep Learning: A Review. *IEEE Transactions on Neural Networks and Learning Systems*, **30**, 3212-3232. <https://doi.org/10.1109/tnnls.2018.2876865>