

doi:10.6041/j.issn.1000-1298.2026.01.012

基于 GA - DETR 的自然场景谷类作物穗部检测方法

王佳诗¹ 崔晨曦¹ 杜奥博¹ 石祥龙¹ 刘进宝¹ 邓雪寒¹
杨万能² 宋 鹏² 段凌凤² 翟瑞芳^{1,2}

(1. 华中农业大学信息学院, 武汉 430070; 2. 华中农业大学作物遗传改良全国重点实验室, 武汉 430070)

摘要: 全球三大谷类作物(水稻、小麦、玉米)穗部的检测,是精准农业与谷类作物表型分析中的基础任务。然而,在复杂田间环境中,穗部分布密集、尺度差异显著且存在大量小目标严重影响检测精度。为解决这些问题,本研究提出一种基于 RT - DETR 架构的门控注意力 DETR (GA - DETR),并引入 3 个创新组件:针对谷物穗部细节特征,设计门控机制 C2F (GMC2F) 模块,通过动态通道加权与跨阶段局部特征融合,提升骨干网络的特征判别能力。为解决谷物穗部形态差异导致的尺度不匹配问题,提出注意力上采样尺度序列特征融合 (AUSFF) 模块,借助 3D 卷积强化多尺度特征依赖。针对无人机图像中小目标检测难题,提出 FPIoU 损失函数,结合目标尺寸自适应加权与难度感知分层策略,优化对难检测样本的处理能力。在水稻穗数据集 (RiceR)、小麦穗数据集 (GWHD) 和无人机玉米雄穗数据集 (MTC - UAV) 上,GA - DETR 性能优于基准模型 RT - DETR 及其他 5 种主流检测模型,mAP@0.5 分别达到 92.8%、91.7% 和 91.3%,对于水稻穗数据集 (RiceR) 模型内存占用量减少 32.5%,浮点运算量降低 14.4%。在 GWHD 数据集的穗部计数任务中,该框架性能超过 5 种主流框架,平均绝对误差 (MAE) 为 5.650,均方根误差 (RMSE) 为 7.383。GA - DETR 有效平衡了检测精度与效率,为谷物穗部通用检测框架提供了跨物种特征建模范式,且兼容小麦、水稻、玉米等多种谷类作物,以及地面相机、无人机等不同采集平台的数据,可支持谷物自动化高通量田间表型监测,进一步推动精准农业发展。

关键词: 谷类作物; 穗部检测; 密集目标; 目标检测; 无人机; GA - DETR

中图分类号: TP301.6 文献标识码: A 文章编号: 1000-1298(2026)01-0125-15 OSID: 

Method for Spike Detection of Cereal Crops in Natural Scenes Based on GA - DETR

WANG Jiashi¹ CUI Chenxi¹ DU Aobo¹ SHI Xianglong¹ LIU Jinbao¹ DENG Xuehan¹
YANG Wanneng² SONG Peng² DUAN Lingfeng² ZHAI Ruifang^{1,2}

(1. College of Informatics, Huazhong Agricultural University, Wuhan 430070, China
2. National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China)

Abstract: The detection of inflorescences from the world's three major cereal crops (rice panicles, wheat spikes, and maize tassels) is a fundamental task in precision farming and cereal crop phenotyping. However, accurate detection remains challenging due to dense distributions, significant scale variations, and small-target in complex environments, which substantially compromise the precision of detection. To tackle these issues, gated attention - DETR (GA - DETR), an architecture based on RT - DETR was proposed, which introduced three novel components: aiming at the delicate tip features of cereal inflorescences, a gated mechanism C2F (GMC2F) module was proposed to enhance backbone feature discrimination through dynamic channel weighting and cross-stage local feature integration. To address the scale mismatch caused by differences in the shapes of cereal inflorescences, an attention upsample scale sequence feature fusion (AUSFF) module was proposed, which enhanced the robustness of multi-scale dependency modeling through 3D convolutions. For difficult small target in UAV images, a FPIoU loss function was proposed, which combined target-size adaptive weighting and difficulty-aware

收稿日期: 2025 - 09 - 25 修回日期: 2025 - 10 - 19
基金项目: 农业农村部科技项目和国家自然科学基金项目(32170411)
作者简介: 王佳诗(2002—),女,硕士生,主要从事农业信息工程研究,E-mail: jiashi_wang@webmail.hzau.edu.cn
通信作者: 翟瑞芳(1979—),女,副教授,博士,主要从事农业信息工程研究,E-mail: rfzhai@mail.hzau.edu.cn

stratification to optimize the performance on hard samples. GA-DETR performed better than the baseline RT-DETR and five mainstream detection models on the RiceR dataset, GWHD dataset, and MTC-UAV dataset, including rice panicles, wheat spikes, and maize tassels, achieving mAP@0.5 of 92.8%, 91.7%, and 91.3%, respectively, while for RiceR dataset reducing model parameters by 32.5% and floating-point computational load by 14.4%. The proposed framework surpassed five state-of-the-art frameworks in inflorescence counting on GWHD dataset, achieving an MAE of 5.650 and an RMSE of 7.383. It effectively balanced accuracy and efficiency, providing a cross-species feature modeling paradigm for the universal detection framework of cereal inflorescence morphology. Compatible with diverse cereal crops (e.g., wheat, rice, maize) and data from different acquisition platforms (ground cameras, UAVs), it supported automated high-throughput field phenotyping monitoring of cereals, further advancing precision agriculture.

Key words: cereal crop; spike detection; dense targets; object detection; UAV; GA-DETR

0 引言

谷类作物的穗部作为作物产量形成的核心器官,其产量直接关乎粮食安全战略^[1-2]。谷类作物的穗数是衡量作物生产力的核心指标,其精准检测与计数是实现科学施肥、合理灌溉及产量预测的基础^[3-4]。然而,传统人工计数方法效率低且易受主观因素干扰,难以满足现代大规模农业生产需求。早期基于图像处理的研究多依赖颜色特征提取、纹理分析等传统算法,如 COINTAULT 等^[5]提出的混合特征半自动计数方法,但该类方法过度依赖人工设计的特征,阻碍了其捕捉田间条件下谷类作物穗部复杂特征的能力。

在谷物穗部检测领域,主流方法主要分为基于卷积神经网络(CNN)的方法与基于 Transformer 的方法。其中,基于 CNN 的方法(如 Faster R-CNN、YOLO 系列)凭借成熟的特征提取机制与高效的推理性能,成为应用最广泛的技术路线之一。Faster R-CNN、Mask R-CNN 等 R-CNN 系列多阶段框架,在复杂田间背景下展现出较强的鲁棒性^[6]。后续研究通过自适应多尺度窗口策略^[7]、基于区域建议网络(RPN)优化多尺度特征融合^[8]等改进,进一步提升了其处理尺度变化的能力。但该类方法固有依赖区域建议生成与特征复用的特性,导致效率较低:RPN 的滑动窗口机制需为每幅图像生成数千个候选区域,且每个候选区域需重复提取特征,计算成本高,无法满足谷物穗部检测的实时性需求。与之相比,单阶段的 YOLO 模型无需生成区域建议,直接预测目标类别与位置,可实现快速推理,适用于田间实时监测^[9-10];轻量化设计进一步提升了其在边缘设备上的适用性^[11-12]。但 YOLO 的单阶段架构在小目标检测中存在固有缺陷:固定网格单元与锚框的设计,使其难以精准捕捉小型谷物穗部的精细特征,在密植场景下易出现漏检现象。

现有谷物穗部检测方法面临共性挑战:密植导

致的穗部重叠增加目标区分难度并引发漏检,模型参数量大导致计算成本高^[13-15]。为解决这些问题,众多研究从模型结构优化入手:通过改进特征金字塔网络(FPN)实现特征融合^[16-18]、优化骨干网络设计^[16-17],以提升特征提取能力;轻量化设计方面,WheatLFANet 通过高效卷积与网络剪枝实现实时检测^[12],KHAKI 等^[16]采用 MobileNetV2 作为骨干网络;BHAGAT 等^[18]通过混合深度可分离卷积减少参数量。但这些方法往往会牺牲特征表征能力,在面对谷物穗部复杂形态变化时精度下降。注意力机制虽能改善遮挡问题,但过度强调局部区域而忽略全局上下文,导致高密度分布场景下权重分配不佳^[19-20];Focal Loss 等损失函数可缓解类别不平衡,但存在超参数敏感问题,且与 Soft-NMS 结合时,在密集簇中易产生高误检率^[21]。

随着 Transformer 架构在计算机视觉领域的兴起,其建模全局依赖关系的独特优势为解决谷物穗部检测的复杂挑战提供了新思路。基于 Transformer 的架构其核心优势在于通过自注意力机制捕捉长距离依赖关系并整合全局上下文信息^[22]。与依赖局部特征聚合的 FPN 或卷积块注意力模块(CBAM)不同,Transformer 可动态捕捉所有空间位置的交互关系,适用于存在尺度变化与密集遮挡的复杂农业环境。在谷物穗部检测中,这种全局建模能力可通过考虑穗部的相对空间关系与整体分布,更好地区分重叠穗部。2020 年,CARION 等^[23]提出 DETR 无锚框、无非极大值抑制(NMS)的端到端集合预测,但计算成本高且小目标性能差,限制了其应用;而 RT-DETR(Real-time detector)的出现为谷物穗部检测带来新的契机,RT-DETR 通过混合编码器提升效率,但在谷物穗部检测的复杂背景下,仍存在多尺度特征融合不足的问题^[24],无法充分捕捉密集簇与形态多变目标的边缘及空间位置信息,导致小型穗部漏检与重叠区域定位误差。

现有研究大多针对单一作物类别,且模型往往

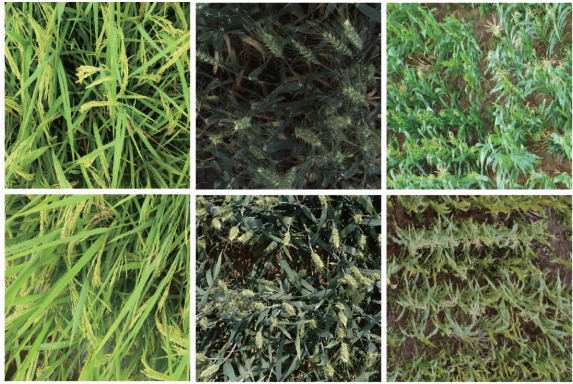
依赖于特定图像采集设备,缺乏能够同时适用于多种谷物穗部(如水稻、小麦、玉米)并可兼容不同拍摄设备(如无人机与地面相机)的通用检测框架。不同作物穗部在形态、分布和背景干扰等方面存在显著差异:小麦与水稻穗细小且易重叠,玉米雄穗体积大、形态多样,均增加了检测难度^[25-26]。

针对这一现状,本研究提出一种新型 GA - DETR 检测网络,通过动态特征调制、多尺度注意力融合与自适应损失机制,有效提升在多作物、多设备图像中的泛化能力与检测精度。该模型能够同时支持从无人机和地面相机获取的图像输入,实现不同田间场景下水稻穗、小麦穗和玉米雄穗等多种谷物的高效、精准检测,为大规模谷物表型分析及智慧农业应用提供可靠技术支撑。

1 材料和方法

1.1 数据集

水稻、小麦、玉米作为全球三大谷类作物及核心主食,具有高度代表性,因此本研究聚焦这 3 种作物的穗部检测。研究所用的 3 个数据集采集方式存在差异:水稻、小麦穗通过地面相机采集,玉米雄穗通过无人机采集,各数据集对应的典型样本如图 1 所示。



(a) 水稻穗 (b) 小麦穗 (c) 玉米雄穗

图 1 谷类作物穗部数据集代表性样本
Fig. 1 Representative samples from cereal crop inflorescence datasets

第 1 个数据集源自 TAN 等^[27]的研究,本研究中将其统一命名为 RiceR 数据集,专门用于水稻穗部检测。该数据集包含 5 416 幅高分辨率图像,涵盖水稻拔节期、抽穗期、灌浆期 3 个关键生育阶段,分为 3 789 幅训练图像与 1 627 幅测试图像。为满足模型训练过程中对超参数调整与过拟合监测的需求,本研究采用“训练集内部分拆”策略,从 3 789 幅训练图像中按比例 8:2 随机划分验证集,最终确定训练集 3 031 幅、验证集 758 幅、测试集 1 627 幅。RiceR 数据集可实现高分辨率的近距离表型分析,能捕捉水稻穗的精细细节,侧重评估模型在处理密

植与复杂背景干扰时的性能。

第 2 个数据集是广泛应用于小麦穗检测的全球小麦穗检测(GWHD)系列数据集^[28]。GWHD_2020 是首个标准化基准数据集,数据通过多种地面表型平台与相机采集共 4 700 幅 RGB 图像。GWHD_2021 新增来自 5 个国家的 1 722 幅图像,表型多样性更高且噪声更少。本研究采用 GWHD_2021 数据集,并将其统一命名为 GWHD 数据集,该数据集包含训练图像 3 655 幅、验证图像 1 476 幅、测试图像 1 381 幅^[29]。该数据集适用于大规模检测场景,覆盖广泛的地理区域与作物品种。

第 3 个数据集是 MTC - UAV 数据集,专门用于无人机图像中的玉米雄穗检测,包含 400 多个玉米品种^[30]。该数据集有 306 幅高分辨率图像(分辨率为 5 472 像素 × 3 648 像素),图像由无人机在高度 12.5 m 处采集,因品种与生育阶段不同,玉米雄穗尺度存在差异,且存在叶片、茎秆遮挡现象。本研究将原始无人机图像裁剪后,得到 1 691 幅 640 像素 × 640 像素大小的图像,从 1 352 幅训练图像中按比例 8:2 随机划分验证集,最终确定训练集 1 082 幅、验证集 270 幅、测试集 339 幅。

上述数据集涵盖了田间场景中的主要干扰因素,包括密植、尺度变化、小目标等,且图像通过地面相机与无人机采集,能够全面评估模型在不同作物、不同场景下谷物穗部检测的通用性与效率。

1.2 GA - DETR 网络结构

如图 2 所示,GA - DETR 以原始 RT - DETR 架构为基础框架,通过 3 项关键改进提升其在谷类作物穗部检测中的性能:将骨干网络中的所有 C2F 模块替换为本文提出的 GMC2F 模块,该模块整合门控分组卷积,在保证效率的同时提升特征表征能力;在特征融合阶段,增加 AUSFF 模块以有效结合深层特征图的高层语义信息与浅层特征图的细节信息;在损失计算阶段引入 FPIoU 损失,优化模型对小目标与遮挡目标的学习过程。

1.2.1 GMC2F

为解决谷类作物穗部分布密集的问题,本文提出一种改进的 CSPDarknet 模块——GMC2F,其核心创新点在于将门控分组卷积(gConv)^[31]与 C2F 结构整合,通过动态特征调制与高效计算,提升模型对多尺度作物穗部的表征能力。该模块借鉴 gMLP 与 GLU 门控机制的设计思路,旨在克服传统 CNN 在密集目标检测中存在的特征冗余与非线性建模能力有限的问题。由图 3 可知,GMC2F 模块的具体结构如下:设输入特征图为 $x \in \mathbf{R}^{H \times W \times C}$,其中 H 、 W 、 C 分别表示特征图的空间高度、宽度与通道维度。首先通

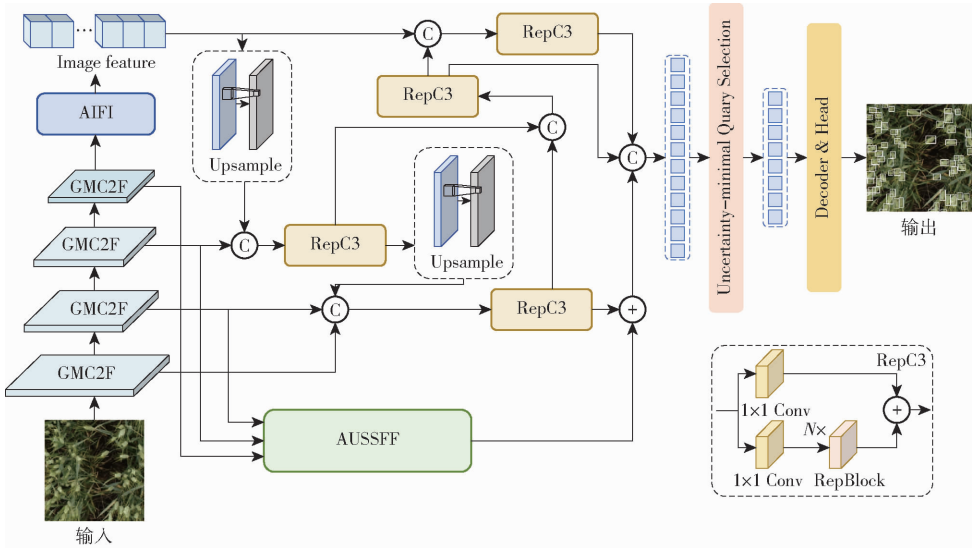


图 2 GA-DETR 结构图

Fig. 2 Diagram of GA-DETR structure

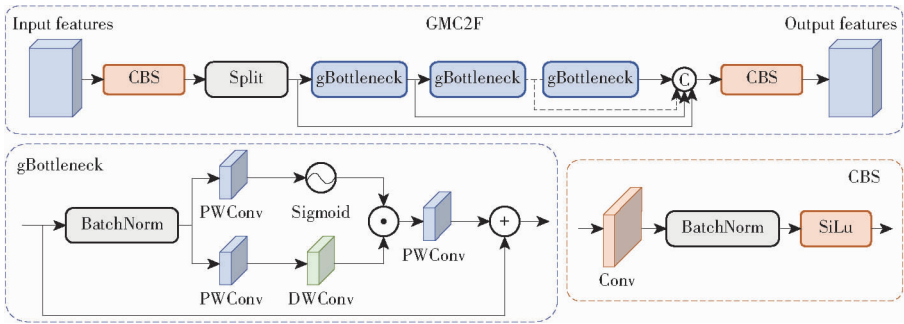


图 3 GMC2F 模块结构图

Fig. 3 Diagram of GMC2F module structure

过批归一化稳定特征分布,归一化后的特征 $\hat{x}$ 通过两条并行路径处理:门控分支与特征变换分支。在门控分支中,先通过逐点卷积(PWConv)压缩通道数,再进行线性投影,即

$$x_1 = \text{Sigmoid}(\text{PWConv}(\hat{x})) \quad (1)$$

式中 PWConv—— 1×1 卷积核,将通道数从 C 压缩至 C/r (压缩比 $r=4$)

Sigmoid——激活函数,可抑制非关键通道(如背景),同时增强谷物穗部特征

x_1 ——门控权重

特征变换分支采用“逐点卷积(PWConv) + 深度可分离卷积(DWConv)”的级联结构^[32],即

$$x_2 = \text{DWConv}(\text{PWConv}(\hat{x})) \quad (2)$$

式中 DWConv—— 3×3 核深度可分离卷积

x_2 ——特征变换分支的输出特征

经过上述处理后,门控权重 x_1 与局部特征 x_2 通过逐元素乘法得到调制后的特征表征,即

$$\hat{y} = x_1 \odot x_2 \quad (3)$$

式中 $\odot$ ——逐元素乘法

$\hat{y}$ ——调制后的特征表征

通过门控分数对空间特征进行动态加权,强化诸如重叠谷物穗部交接处的重要性特征。最终输出通过残差连接整合,加法残差结构可在反向传播过程中保留原始特征完整性,缓解梯度消失问题,即

$$y = x + \text{PWConv}(\hat{y}) \quad (4)$$

式中 x ——输入特征 y ——输出特征

该设计通过门控机制实现非线性动态特征调制,缓解了传统 ReLU/GELU 激活函数伴随的梯度消失问题。与注意力机制相比,GMC2F 借助门控卷积实现高效特征调制,在应用于谷物穗部检测任务时展现出显著优势:门控分支与特征变换分支并行,可在动态调制输入特征的同时实现轻量化计算。具体而言,门控分支生成通道注意力权重,有效抑制土壤、叶片等背景的冗余信息;特征变换分支引入深度可分离卷积,在削减计算开销的同时,完整保存了穗部细微的纹理特征。两条分支通过乘法融合,并结合跳跃连接,不仅增强了梯度流动,还提升了模型对判别性特征的响应能力。

1.2.2 AUSFF

为解决 RT-DETR 架构中金字塔特征相关性利用不足的关键问题(传统融合方法如简单拼接或

逐元素求和无法保留跨尺度依赖关系), 本研究提出一种新型多尺度特征融合模块 AUSSFF, 结构如图 4 所示。该模块通过 1×1 卷积统一浅层特征 P3 (含小目标关键细节)、中层特征 P4 与深层特征 P5 (富含高维语义信息) 的通道维度, 随后引入创新的注意力上采样机制。该机制通过双路径结构生成高质量特征: 一条路径采用转置卷积学习自适应上采样核, 另一条路径通过双线性插值保持特征平滑性, 并借助 1×1 卷积增强特征表达。对两条路径的特征进行拼接处理后, 通过由全局平均池化、 1×1 卷积及 HardSigmoid 激活构成的通道注意力机制, 对特征重要性展开加权, 最后以 1×1 卷积完成特征融合, 即

$$X_1 = \text{ConvTrans}(X) \oplus (\text{Upsample}(X)) \quad (5)$$

$$X_2 = X_1 \odot \sigma(\text{Conv}(\text{GAP}(X))) \quad (6)$$

$$\text{AttnUp}(X) = \text{DWConv}_{1 \times 1}(X_2) \quad (7)$$

式中 X_1 ——双路径特征相加结果
 X_2 —— X_1 经通道注意力加权
 X ——输入特征图
 ConvTrans —— 2×2 核、步长为 2 的转置卷积
 Upsample ——2 倍双线性插值
 GAP ——全局平均池化
 AttnUp ——注意力上采样模块
 σ ——HardSigmoid 激活函数
 $\odot$ ——注意力加权的哈达玛积
 $\text{DWConv}_{1 \times 1}$ —— 1×1 深度可分离卷积

针对中层特征 P4, 通过单阶段级联注意力上采样模块实现 2 倍分辨率提升, 即

$$P_4^{up} = \text{AttnUp}(P_4') \quad (8)$$

式中 P_4' ——特征 P4 经过卷积处理后得到的中间特征

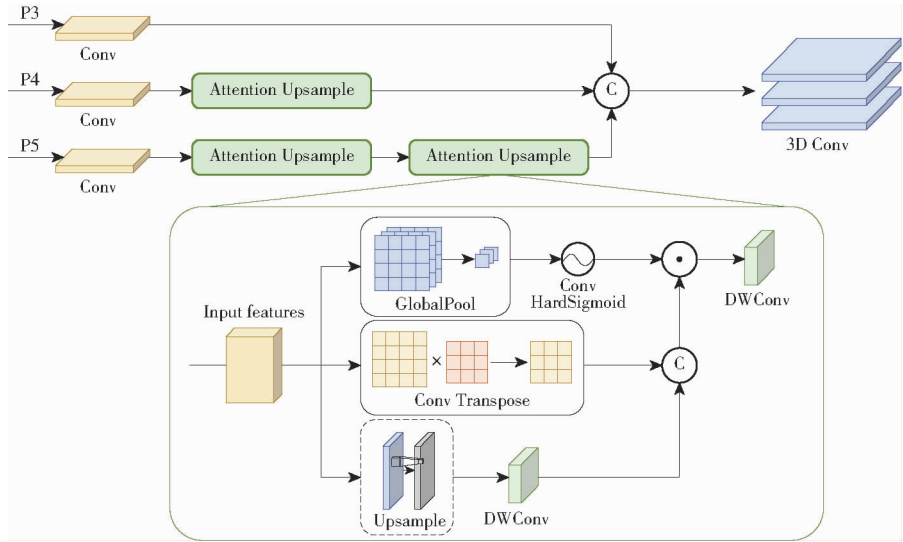


图 4 AUSSFF 模块结构图

Fig. 4 Diagram of AUSSFF module structure

P_4^{up} ——特征 P_4' 经上采样后的特征, 由 P_4' 通过单阶段级联注意力上采样模块后得到

针对深层特征 P5, 通过两阶段级联注意力上采样实现 4 倍渐进式上采样, 以保留特征完整性并减少细粒度细节损失, 即

$$P_5^{up} = \text{AttnUp}(\text{AttnUp}(P_5')) \quad (9)$$

式中 P_5' ——特征 P5 经过卷积处理后得到的中间特征

P_5^{up} ——特征 P_5' 经上采样后的特征, 由 P_5' 通过两阶段级联注意力上采样模块后得到

将所有特征统一至 P3 分辨率后, 模块基于尺度空间理论构建四维特征体, 即

$$F_{4d} = \text{Concat}(P_i \oplus \text{dim}_t) \quad (10)$$

$$(P_i \in (P_3', P_4^{up}, P_5^{up}))$$

式中 $\oplus \text{dim}_t$ ——通过特征复制扩展时间轴

F_{4d} ——四维特征体, $F_{4d} \in \mathbf{R}^{H \times W \times T \times C}$

Concat ——特征拼接

P_i ——参与四维特征体构建的多尺度特征集合, 具体包括 P_3' (特征 P3 经预处理后的特征)、 P_4^{up} 、 P_5^{up}

通过 3D 卷积进行跨尺度特征提取, 随后通过时间最大池化降维, 最终生成分辨率提升的融合特征^[33], 即

$$F_c = \text{Conv}_{3D}(F_{4d}) \quad (11)$$

式中 Conv_{3D} —— $1 \times 1 \times 1$ 核 3D 卷积

F_c ——融合特征

为应对谷物穗部目标的空间变化, AUSSFF 模块结合单阶段与两阶段级联注意力上采样操作, 协同利用转置卷积、双线性插值与通道注意力, 在增强

语义表达的同时保留谷物穗部边缘细节。融合特征经 3D 卷积与池化处理后,可有效捕捉跨尺度依赖关系,特别适用于谷物穗部检测任务,能够精准识别密集簇中的重叠目标。

1.2.3 FPIoU

在目标检测任务中,边界框回归的精准度对目标定位的准确程度起着直接决定性作用,其中,损失函数的设计是制约边框回归性能的核心要素。现有研究多聚焦于挖掘边框间几何关系(如 IoU、GIoU)来优化回归效果,却普遍忽视了田间场景中固有的样本难度分布问题——易样本(尺寸正常、视野清晰、光照适宜的目标)在训练中占比高,而难样本(遮挡或像素小于 50 的穗部)在训练中获得关注不足。这种不平衡导致模型偏向简单样本,降低了密集种植场景中的定位精度。针对这一问题,本研究提出一种创新的边框回归自适应损失函数 FPIoU,通过动态调整难易样本,关注分配优化回归性能。FPIoU 融合两项关键技术:①借鉴 Focaler-IoU 所采用的线性区间映射思想,对 IoU 损失进行重新设计,利用区间参数的动态调整,实现对不同样本的特异性聚焦^[34]。②针对传统 IoU 类损失固有的锚框膨胀问题,引入 PIoU 的尺寸自适应惩罚机制^[35],通过构建惩罚因子 P ,依据预测框与目标框的尺寸偏差动态调整惩罚权重,在规避锚框不合理膨胀的同时,进一步增强回归稳定性。惩罚因子的形式化定义为

$$P = \frac{\frac{dw_1}{w_{gt}} + \frac{dw_2}{w_{gt}} + \frac{dh_1}{h_{gt}} + \frac{dh_2}{h_{gt}}}{4} \tag{12}$$

式中 dw_1 、 dw_2 、 dh_1 、 dh_2 ——预测框与目标框对应边的绝对距离

w_{gt} 、 h_{gt} ——目标框的宽度与高度

为进一步优化损失函数,引入质量自适应惩罚项,基于锚框质量动态调整:该惩罚项(公式(12))通过对低质量锚框分配较低权重,缓解茎秆、叶片等模糊区域导致的边界偏移问题。此外,重构的 IoU 损失融入难度敏感重映射机制,在指定区间 $[d,u]$ 内对 IoU 值进行线性变换,并通过惩罚函数 $f(x)$ 强化难样本关注,即

$$f(x) = 1 - e^{-x^2} \tag{13}$$

$$FPIoU = \begin{cases} 0 & (IoU < d) \\ \frac{IoU - d}{u - d} - f(P) & (d \leq IoU \leq u) \\ 1 & (IoU > u) \end{cases} \tag{14}$$

$$Loss_{FPIoU} = 1 - \frac{IoU - d}{u - d} + f(P) \tag{15}$$

式中 $FPIoU$ ——重构后的 FPIoU

$Loss_{FPIoU}$ ——重构后的 Loss

IoU ——原始交并比

d 、 u 取值范围为 $[0,1]$,通过系统调整 d 、 u 值,可优化 FPIoU 对特定回归样本的精准聚焦能力。

针对无人机图像中小而隐蔽穗部检测的难题,FPIoU 损失通过线性区间映射重新定义传统 IoU 损失,实现易、难样本间的动态注意力调整;同时,尺寸自适应惩罚因子 P 基于真实框差异调整惩罚权重,有效避免锚框不合理膨胀。该机制可强化对密集、遮挡、小尺寸穗部等难样本的学习,同时提升边界框回归精度。

1.3 试验设置

谷类作物穗部检测模型的训练与验证的硬件配置如下:AMD EPYC 7642 CPU(2.30 GHz)、90 GB 内存、NVIDIA RTX 3090 GPU(24 GB 显存);软件环境为 CUDA 11.8、PyTorch 2.0.0、Python 3.8。模型训练共进行 300 个训练轮次,最终通过多个数据集验证模型性能。详细超参数设置如表 1 所示。

表 1 试验训练参数

Tab.1 Experimental training parameters

参数	数值/类型
训练轮次	300
批量大小	16
图像尺寸/(像素×像素)	640×640
初始学习率	0.000 1
优化器	AdamW

1.4 评价指标

本文采用客观指标进行评估:精确率(P)、召回率(R)、交并比(IoU)阈值为 0.5 时的平均精度(mAP@0.5)、IoU 阈值在 0.5~0.95 区间的平均精度(mAP@0.5:0.95)用于评估目标检测准确性,浮点运算量衡量计算复杂度,模型内存占用量评估内存需求。

2 试验结果与分析

2.1 模型训练

模型 GA-DETR 在 3 个数据集上的训练性能指标变化如图 5 所示。精确率曲线(图 5a)显示,模型在 GWHD 数据集上稳定性最佳,精确率达 90.6%,其次为 RiceR 数据集(88.8%)、MTC-UAV 数据集(87.8%),体现出较强的穗部目标区分能力;召回率方面(图 5b),RiceR 数据集以 88.4% 领先,GWHD 与 MTC-UAV 数据集分别达到 86.8%、84.6%;mAP@0.5 指标(图 5c)中,3 个作物数据集均超过 90%,其中 RiceR 数据集达 92.8%,表明穗

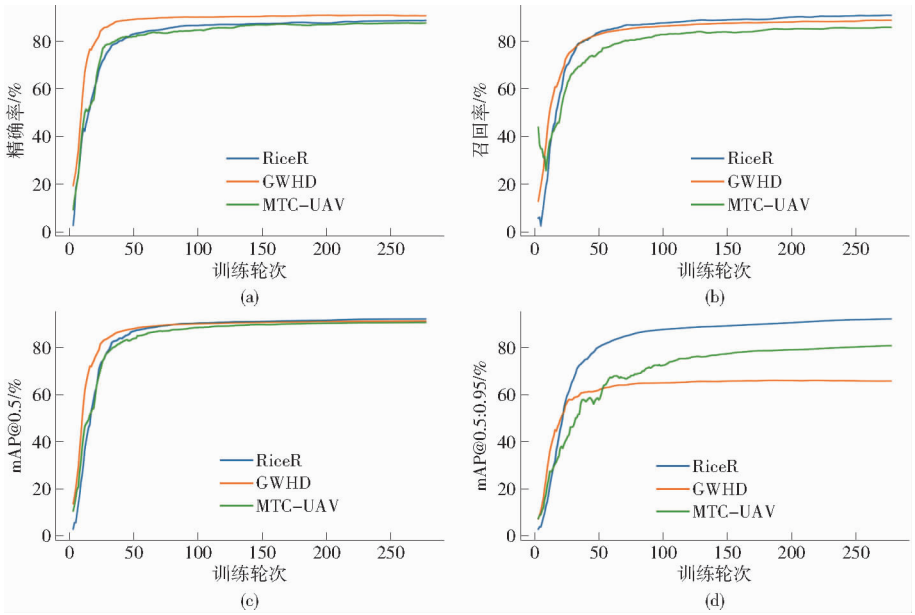


图 5 模型在 3 个数据集上的训练性能指标

Fig. 5 Training performance metrics of model on three datasets

部边界框定位精度较高。mAP@ 0.5 : 0.95 指标 (图 5d) 中,RiceR 数据集和 MTC - UAV 数据集达到 60% 以上、GWHD 数据集低于 50%。这种差异反映出模型对不同谷类作物穗部的泛化能力存在细微差别,在 RiceR 与 MTC - UAV 数据集的尺度与遮挡问题处理上表现尤为突出。总体而言,模型在不同作物与不同设备成像场景下保持了一致的有效性,随着训练轮次增加,模型对多种谷类作物穗部特征的

学习能力不断增强,最终趋于稳定。

2.2 模型对比

为全面评估模型在谷类作物穗部检测的性能,本研究将 GA - DETR 与 5 种目标检测算法 (YOLO v5m^[36], YOLO v8m, YOLO v10m^[37], YOLO 11m^[38], RT - DETR - R18) 进行了对比试验。试验在 3 个数据集上进行:RiceR 数据集、GWHD 数据集和 MTC - UAV 数据集 (表 2)。

表 2 在所有数据集上 GA-DETR 与其他检测算法的性能对比结果							
Tab. 2 Comparison results between GA-DETR and other detection algorithms on all datasets							
数据集	模型	P/%	R/%	mAP@ 0. 5/%	mAP@ 0. 5 : 0. 95/%	模型内存占用量/MB	浮点运算量
RiceR	YOLO v5m	81. 8	84. 1	88. 8	64. 2	50. 5	6. 40 × 10 ¹⁰
	YOLO v8m	81. 2	82. 0	86. 6	62. 6	52. 1	7. 87 × 10 ¹⁰
	YOLO v10m	82. 0	83. 5	88. 9	64. 8	33. 5	6. 34 × 10 ¹⁰
	YOLO 11m	82. 1	84. 3	89. 2	65. 2	40. 5	6. 76 × 10 ¹⁰
	RT - DETR - R18	85. 3	84. 5	90. 0	65. 0	40. 5	5. 69 × 10 ¹⁰
	GA - DETR	88. 8	88. 4	92. 8	69. 5	27. 3	4. 87 × 10 ¹⁰
GWHD	YOLO v5m	90. 8	85. 2	90. 9	48. 6	50. 5	6. 40 × 10 ¹⁰
	YOLO v8m	91. 2	85. 0	91. 0	48. 4	52. 1	7. 87 × 10 ¹⁰
	YOLO v10m	94. 1	82. 5	89. 6	54. 4	33. 5	6. 34 × 10 ¹⁰
	YOLO 11m	93. 2	81. 2	89. 1	54. 5	40. 5	6. 76 × 10 ¹⁰
	RT - DETR - R18	90. 5	85. 6	90. 3	48. 7	40. 5	5. 69 × 10 ¹⁰
	GA - DETR	90. 6	86. 8	91. 8	49. 0	27. 3	4. 87 × 10 ¹⁰
MTC - UAV	YOLO v5m	87. 1	85. 4	90. 1	55. 8	50. 5	6. 40 × 10 ¹⁰
	YOLO v8m	87. 9	81. 0	85. 5	56. 0	52. 1	7. 87 × 10 ¹⁰
	YOLO v10m	87. 4	83. 6	89. 7	49. 3	33. 5	6. 34 × 10 ¹⁰
	YOLO 11m	84. 5	83. 8	89. 0	56. 2	40. 5	6. 76 × 10 ¹⁰
	RT - DETR - R18	86. 5	84. 2	90. 9	60. 0	40. 5	5. 69 × 10 ¹⁰
	GA - DETR	87. 8	84. 6	91. 3	60. 8	27. 3	4. 87 × 10 ¹⁰

在 RiceR 数据集上,GA - DETR 在所有指标上表现优异,实现了 88. 8% 的精确率 (P)、88. 4% 的召回率 (R)、92. 8% 的 mAP @ 0. 5 和 69. 5% 的 mAP@ 0. 5 : 0. 95。对比 RT - DETR - R18,GA - DETR

模型内存占用量减少了 32.5% (减少至 27.3 MB), 浮点运算量降低了 14.4% (减小至 4.87×10^{10}), 同时 mAP@0.5 和 mAP@0.5:0.95 分别提升 2.8、4.5 个百分点。精确率和召回率分别提高 3.5、3.9 个百分点,减少叶/茎误检与密集穗部漏检。对比 YOLO 模型,GA-DETR 的 mAP@0.5、mAP@0.5:0.95 比 YOLO 11m 分别高 3.6、4.3 个百分点,召回率比 YOLO v5m 和 YOLO v8m 高 4.3、6.4 个百分点,表明 GA-DETR 在叶干扰和复杂背景噪声场景中更有效。

在 GWHD 数据集上,GA-DETR 实现了 91.8% 的 mAP@0.5,高出其他模型 0.8~2.7 个百分点。GA-DETR 的召回率 (86.8%),超过 YOLO v5m (85.2%)、YOLO v8m (85.0%)、RT-DETR-R18 (85.6%),显著减少了漏检。尽管 YOLO v10m (54.5%)和 YOLO 11m(54.4%)的 mAP@0.5:0.95 较高,但依赖分类置信度阈值提高和轻量级架构,本质上牺牲了召回率(仅 82.5%和 81.2%),导致杂草遮挡场景检测的完整性受损。GA-DETR 的精确率 (90.6%)不及 YOLO v10m(94.1%),因为其基于 Transformer 的架构会优先考虑全局上下文建模以增强召回率。虽然这种设计增强了模型在野外场景中的检测能力,但它有时会对小型、密集的目标误检。

在 MTC-UAV 数据集上,GA-DETR 在所有模型中保持领先水平。与 RT-DETR-R18 相比,精确率提高 1.3 个百分点,召回率提高 0.4 个百分点, mAP@0.5 提高 0.4 个百分点, mAP@0.5:0.95 提高 0.8 个百分点,表明 GA-DETR 适配尺度变化的无人机玉米雄穗图像。与 YOLO 系列相比,召回率比 YOLO v8m 高 3.6 个百分点,精确率比 YOLO 11m 高 3.3 个百分点,说明其在无人机图像中效果稳定。该数据集的性能提升较小,推断是因数据集规模较小(约 2 000 幅)以及裁剪后图像的分辨率降低。

在 3 类谷类作物穗部检测中,GA-DETR 在检测精确率、召回率和计算效率方面表现良好,证明其不同作物、成像方式、环境复杂度场景下的鲁棒性与适应性较好。

为验证多数据集融合训练对 GA-DETR 模型通用能力的效果,本研究新增 3 个数据集融合训练方案:合并 RiceR 水稻、GWHD 小麦、MTC-UAV 玉米数据集以训练通用模型权重,并在 3 类作物独立测试集上评估性能(表 3)。

在 RiceR 水稻测试集,通用模型精确率较独立训练下降 4.1 个百分点、召回率下降 4.2 个百分点,因为 RiceR 数据集中的水稻存在密植程度高、穗部

表 3 单数据集训练与 3 个数据集融合训练模型性能对比
Tab.3 Performance comparison of models trained with single dataset training and three-dataset fusion training %

训练方式	数据集	P	R	mAP@0.5	mAP@0.5:0.95
独立训练	RiceR	88.8	88.4	92.8	69.5
	GWHD	90.6	86.8	91.8	49.0
	MTC-UAV	87.8	84.6	91.3	60.8
融合训练	RiceR	84.7	84.2	89.8	65.5
	GWHD	90.5	86.5	91.2	49.0
	MTC-UAV	85.5	83.7	90.0	57.4

与叶片特征差异小的检测难点,与小麦的粗壮穗部、高空视角的玉米分散雄穗差异显著,融合训练中模型对水稻特征的学习深度被稀释,故降幅相对更大;虽有此降幅,但精确率和召回率仍保持在 84% 左右;而在 GWHD 小麦测试集中的精确率较独立训练仅下降 0.1 个百分点、召回率下降 0.3 个百分点、mAP@0.5:0.95 与独立训练持平,性能近乎一致,能适配小麦多地域样本的复杂分布;在 MTC-UAV 玉米测试集,通用模型精确率较独立训练下降 2.3 个百分点、召回率下降 0.9 个百分点, mAP@0.5 仍保持在 90%,可精准捕捉无人机高空视角下玉米雄穗的小目标特征。

值得注意的是,通用模型以极小性能损耗打破“单一权重仅适配单一作物”的局限,3 类作物 mAP@0.5 最大降幅仅 3.0 个百分点 (RiceR 数据集),却实现水稻、小麦、玉米的全覆盖检测,且无需切换权重,为谷类作物穗部检测规模化应用提供支撑,并验证了融合训练提升模型通用性的有效性。

2.3 消融试验

本研究基于 RT-DETR 基线模型,引入 GMC2F、AUSFF 和 FPIoU 损失函数进行消融试验(表 4)。基线模型 RT-DETR 实现了 85.3% 的精确率,84.5% 的召回率,90.0% 的 mAP@0.5 和 65.0% 的 mAP@0.5:0.95。仅添加 GMC2F 将精确率提升至 85.4%,同时召回率降至 83.7%,mAP@0.5 和 mAP@0.5:0.95 分别下降 0.4 个百分点和 1.5 个百分点,表明其特征过滤机制以牺牲小目标换高置信度目标优先。单独引入 AUSFF 显著提升所有指标,精确率提高 1.8 个百分点至 87.1%,召回率提高 1.3 个百分点至 85.8%,mAP@0.5 提高 1.2 个百分点至 91.2%,mAP@0.5:0.95 提高 1.3 个百分点至 66.3%,因其多尺度融合保留了细粒度特征;FPIoU 损失函数效果最突出,mAP@0.5 提高到 91.5%,mAP@0.5:0.95 提高至 68.2%,虽召回率保持在基线(84.5%),但其通过动态聚焦难易样本

表 4 在 RiceR 数据集上每个模块组合的消融试验

Tab. 4 Ablation experiments for each module combination on RiceR dataset							%
RT - DETR	GMC2F	AUSFF	FPIoU	P	R	mAP@0.5	mAP@0.95
√				85.3	84.5	90.0	65.0
√	√			85.4	83.7	89.6	63.5
√		√		87.1	85.8	91.2	66.3
√			√	88.5	84.5	91.5	68.2
√	√	√		87.0	85.5	89.6	65.3
√	√		√	86.8	83.9	91.5	67.1
√		√	√	87.4	87.3	91.7	67.7
√	√	√	√	88.8	88.4	92.8	69.5

注：“√”表示添加此模块。

优化边框回归精度。

当 GMC2F 与 AUSFF 联合作用时,模型精确率较基线提升 1.7 个百分点至 87.0%,召回率较基线提升 1.0 个百分点至 85.5%,这表明 GMC2F 的动态通道加权能筛选出更具判别性的作物穗部特征,为 AUSFF 的多尺度融合提供优质输入,有效推动精确率与召回率提升;而当 GMC2F 与 FPIoU 组合

时,模型精确率较基线显著提升 1.5 个百分点至 86.8%,mAP@0.5:0.95 较基线提高 2.1 个百分点至 67.1%,召回率较基线下降 0.6 个百分点至 83.9%,说明 GMC2F 强化关键特征后,FPIoU 的难度分层损失能高效优化边界框,但小幅影响召回率;当联合使用 AUSFF 与 FPIoU 时,召回率提升至 87.3%,mAP@0.5 提升至 91.7%,虽 mAP@0.5:0.95 降至 67.7%,验证了多尺度上下文能够实现挑战性目标的精确定位。最终完整模型实现了最佳性能,精确率为 88.8%,召回率为 88.4%,mAP@0.5 为 92.8%,mAP@0.5:0.95 为 69.5%,分别比基线提高了 3.5、3.9、2.8、4.5 个百分点。这些结果不仅直观体现性能提升,更表明在复杂且不断变化的检测场景时,GMC2F 的特征选择、AUSFF 的多尺度感知和 FPIoU 损失的自适应回归可以协同工作,全面提高整体检测性能。

为明确 GMC2F、AUSFF、FPIoU 在多作物场景的适配性,本研究在 GWHD 小麦、MTC - UAV 玉米数据集开展单模块消融试验,通过逐一加入各模块并对比基线模型精确率、召回率、mAP@0.5,定位性能增益方向(表 5)。

表 5 在 GWHD 和 MTC - UAV 数据集上单模块的消融试验

Tab. 5 Ablation experiments of single modules on GWHD and MTC - UAV datasets									%
RT - DETR	GMC2F	AUSFF	FPIoU	GWHD			MTC - UAV		
				P	R	mAP@0.5	P	R	mAP@0.5
√				90.5	85.6	90.3	86.5	84.2	90.9
√	√			90.6	86.8	91.1	86.7	85.1	90.9
√		√		90.8	86.4	91.3	87.1	84.6	91.3
√			√	91.5	85.7	90.9	87.5	84.9	91.5
√	√	√	√	90.6	86.8	91.8	87.8	84.6	91.3

注：“√”表示添加此模块。

在 GWHD 小麦数据集上,GMC2F 将召回率提升 1.2 个百分点、mAP@0.5 提升 0.8 个百分点,通过强化小麦穗特征,有效改善小麦密植场景的目标召回能力;AUSFF 将召回率提升 0.8 个百分点、mAP@0.5 提升 1.0 个百分点,以多尺度融合优化小麦穗尺度差异的问题;FPIoU 使精确率提升 1.0 个百分点、mAP@0.5 提升 0.6 个百分点,其通过优化边界框回归,重点提升检测精确性。

在 MTC - UAV 玉米数据集上,GMC2F 使精确率微升 0.2 个百分点、召回率提升 0.9 个百分点且 mAP@0.5 保持不变,说明其在不影响核心精度前提下提升目标检出率;AUSFF 使精确率提升 0.6 个百分点、召回率提升 0.4 个百分点,验证多尺度感知对高空视角玉米穗分散且视角差异大问题的优化;FPIoU 使精确率提升 1.0 个百分点、召回率提升

0.7 个百分点、mAP@0.5 提升 0.6 个百分点,实现精确率与召回率的协同提升。

整体而言,3 个模块在两类数据集均呈现正向提升。其中 FPIoU 对精确率的提升最显著(均为 1.0 个百分点);AUSFF 对 mAP@0.5 的优化更突出;GMC2F 则在召回率提升上表现更优。

2.4 结果可视化

2.4.1 谷物作物检测可视化

定性评估包括 3 种主要谷类作物——水稻穗、小麦穗和玉米雄穗,通过将 GA - DETR 与 YOLO v5、YOLO v8、YOLO 11、RT - DETR 在 RiceR、GWHD、MTC - UAV 数据集的代表性田间场景中进行可视化对比分析。如图 6 所示(注:红色标记为误检,蓝色标记为重复检测),在水稻穗检测(图 6a)中,GA - DETR 能够准确捕捉到叶片杂乱

中的细长分支结构,而 YOLO v5/YOLO v8 往往将相邻的茎误分类为穗;对于小麦穗(图 6b),它有效地区分了密集排列且成熟度不同的穗,优于出现检测框重叠的 YOLO v8 和漏检小目标穗的 RT-DETR。在玉米雄穗场景(图 6c)中,RT-DETR 表

现出过度的边界框重叠,YOLO 模型则出现漏检,GA-DETR 展示了较优的适应性。这些跨作物的试验结果共同证实了 GA-DETR 超越作物特定形态差异的能力,为大规模、多谷物表型研究奠定了技术基础。

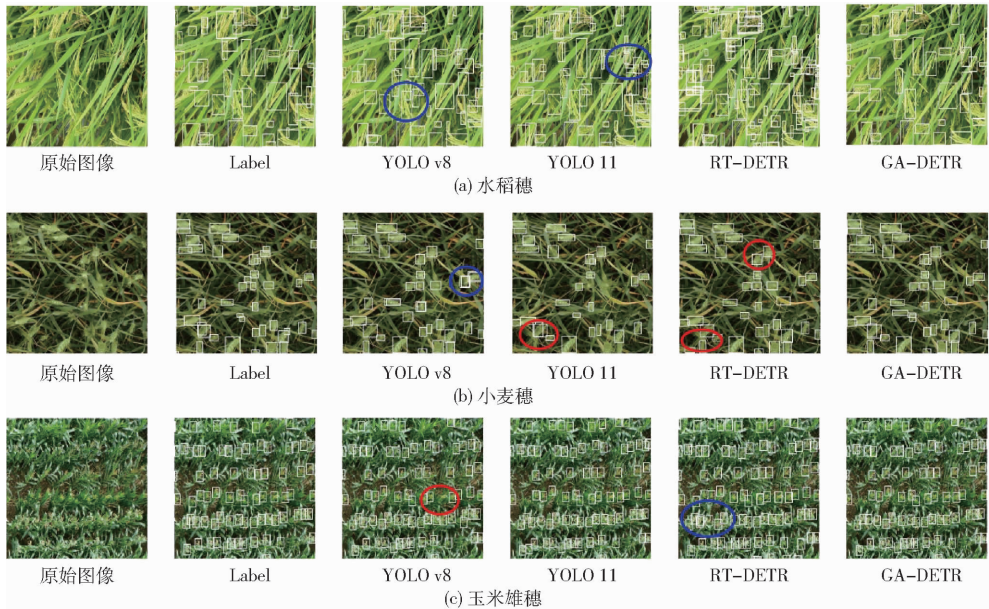


图 6 不同模型在所有数据集上的性能视觉比较

Fig. 6 Performance visual comparison of various detection models on all datasets

聚焦 GWHD 数据集(图 7),GA-DETR 的检测框始终与实际穗部的形状精确匹配。而 RT-DETR 显示出更广泛的覆盖范围,由于特征关注不足,在处理复杂形态或背景干扰时会出现模糊或者错误。在 高密度小麦穗场景中,如图 7a、7b 所示,GA-DETR 的优越边界检测能力能够更完整地捕获目标,而其他模型要么漏检被叶遮挡的穗,要么漏检重叠的穗;在光线过曝条件下,如图 7c 所示,YOLO v8 和 RT-

DETR 表现出显著的漏检,而 GA-DETR 可靠地识别出目标,显示出对光照变化的适应性。

2.4.2 复杂场景检测可视化

图 8 的可视化对比揭示了 GA-DETR 在极端场景下的技术优势,在图 8b 的密集遮挡场景中,RT-DETR 由于通道特征混淆,将叶片纹理错误地识别为穗,从而产生误检。而 GA-DETR 通过 GMC2F 的门控卷积的特征动态抑制了高频叶片噪

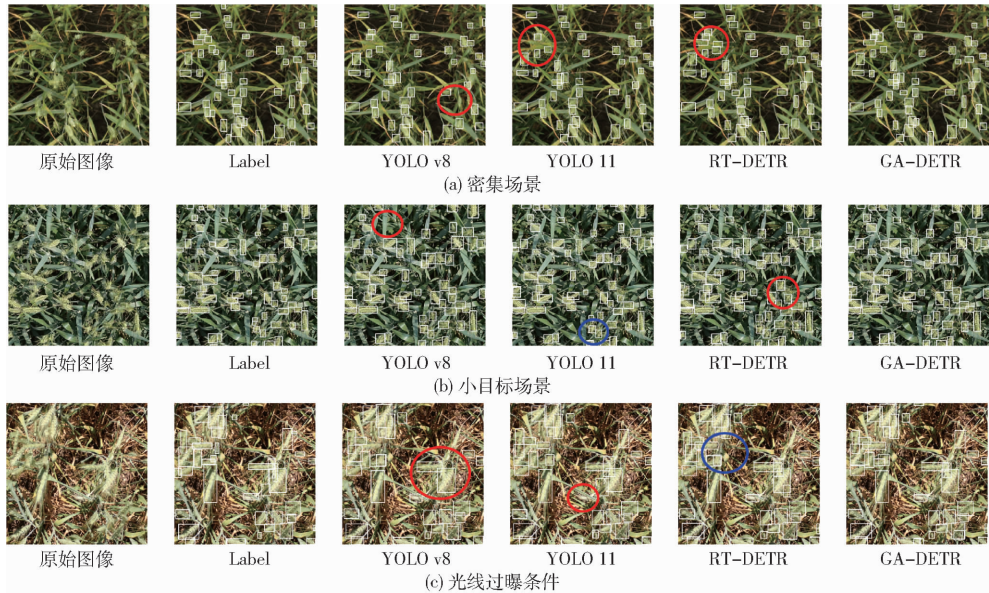


图 7 不同模型在 GWHD 数据集上的性能视觉比较

Fig. 7 Performance visual comparison of various detection models on GWHD dataset

声并增强了尖端曲率特征,实现了精确的穗-叶区分。如图 8b、8c 所示的重叠目标场景进一步揭示了 RT-DETR 的局限性,它将单个穗生成重复的边界框,而 GA-DETR 即使在涉及叶片遮挡和小穗尺寸的情况下,也能通过 FPIoU 的难度分级优化保持对单一目标的准确定位,为困难样本优先学习,并使用尺寸敏感性惩罚锚点偏差。这些比较结果说明了 GA-DETR 在处理复杂场景时较强的检测能力,尤其在区分相似纹理、克服遮挡和准确检测小目标的场景。

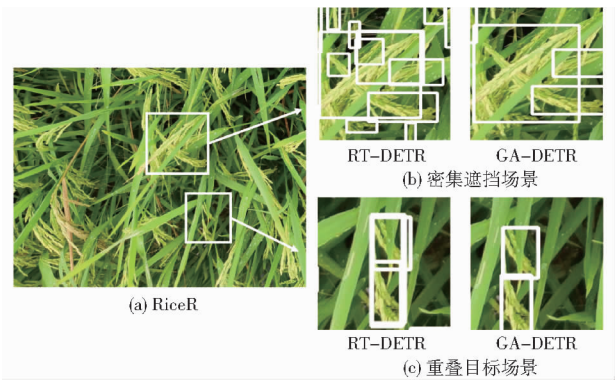


图 8 密集稻穗场景下的检测可视化 (RiceR 数据集)
Fig. 8 Detection visualization in dense rice panicle scenarios (RiceR dataset)

如图 9 所示,GA-DETR 在极端场景中表现出技术优势,在有阴影遮挡的环境下(图 9b),光强衰减导致的高频纹理丢失显著削弱了麦穗与背景的对比度,致使 RT-DETR 完全漏检麦穗目标。相比之下,GA-DETR 成功识别了低可见度的穗,这归功于 FPIoU 损失函数的双重机制设计:通过提升小目标的损失权重进行非线性重映射强化小目标梯度反馈;同步激活的尺寸敏感惩罚机制则基于目标与锚框的几何相似性约束边界框形变,双机制协同保障了弱光环境下的定位鲁棒性。对于图 9c 中的边缘目标,常规卷积的固定感受野在图像边界产生特征

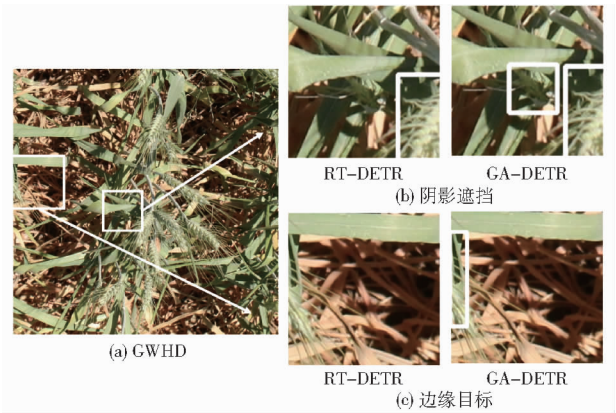


图 9 遮挡与边缘场景中的检测可视化 (GWHD 数据集)
Fig. 9 Detection visualization in occlusion and edge scenarios (GWHD dataset)

采样残缺,引发 RT-DETR 出现边缘漏检。GA-DETR 的完整检测则依托 AUSSFF 模块的多尺度融合架构:通过通道注意力引导的上采样在空间维度重建边缘梯度场,结合三维卷积对跨层级特征进行关联,有效增强边界穗粒感知能力。上述试验说明,GA-DETR 在两组不同的极端场景下都能实现超过 RT-DETR 的检测效果。

2.4.3 自采数据集检测可视化

本研究为验证三数据集融合训练后通用权重的跨场景适配性,选择 4 幅自采水稻无人机视角图像(飞行高度 5 m,涵盖 2 幅抽穗期、2 幅灌浆期生育阶段,均未参与前期融合训练),采用“无额外训练”模式直接加载该通用权重进行检测推理。从图 10 的直观效果来看,该融合权重在水稻 UAV 场景下展现出与生育期相关的检测差异:①图 10a 为抽穗期,水稻穗部虽存在一定密集度,但融合权重仍能实现较好的检测效果,检测框与穗部实际轮廓重合度较高,误检数量单幅图像小于等于 3 个,整体能准确捕捉核心穗部目标。②图 10b 为灌浆期,受穗部更密集(穗间重叠度提升)、灌浆期穗部形态与训练数据中特征差异略大的影响,检测效果较抽穗期有所下降,虽然核心穗部仍能有效识别,但是出现少量检测框偏移及漏检。本研究设计的多作物通用模型,其设计目标侧重于跨作物场景的适配性而非单一作物密集场景的极致精度,因此在水稻 UAV 密集场景下的表现已满足通用检测需求,进一步证明该融合权重在未参与训练的水稻 UAV 新场景中具备实用的泛化能力,为模型在田间多生育期水稻 UAV 检测中的应用提供了可视化支撑。

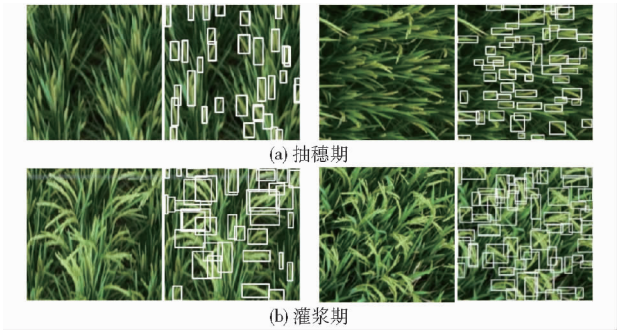


图 10 通用权重在高空视角水稻场景的检测可视化
Fig. 10 Detection visualization of general-purpose weights in high-altitude rice scenario

2.4.4 Grad-CAM 可视化结果

为了直观展示 GA-DETR 的优势,使用 Grad-CAM 进行注意力可视化。结果显示,GA-DETR 的热力图在小麦穗区域表现出高度集中和精确的响应(图 11a、11b),热力峰值与真实分布紧密契合,有效抑制背景干扰,极大提升定位精度。相反,YOLO v8

生成的热力图常出现大范围弥散错误的点,而 RT-DETR 虽然覆盖区域较广但缺乏精细聚焦。在重叠、密集稻穗的极端场景下(图 11c、11d),GA-DETR 展示了卓越的目标分离和特征区分能力,而 YOLO v8 的热力图模糊且无序,这表明在处理复杂空间关系时存在困难;RT-DETR 亦因细粒度特征捕捉不足导致注意力分布杂乱。

在无人机图像中检测玉米雄穗是一项相当困难的任务,因为图像上的雄穗通常尺寸较小。对于图 11e、11f 中的小目标场景下,GA-DETR 的热力图精确聚焦于雄穗区域,即使在密集的叶片或茎周围,

重叠雄穗之间也有明显的边界,这表明它具有很强的捕捉无人机图像中玉米雄穗特征的能力。相比之下, YOLO v8 和 YOLO 11 的热力图过度扩散到背景区域。

2.4.5 谷物穗部计数性能比较

谷物穗部计数对于产量估计和表型分析至关重要,因为它直接反映了作物生长状况和生产力。为了验证框架在穗部计数任务中的有效性,从 GWHD、RiceR、MTC-UAV 3 个数据集中分别随机选取 200 幅测试图像进行评估。结果显示,GA-DETR 在不同作物和成像环境下均实现了优异的性能。在 GWHD 数据集上,GA-DETR 的决定系数达

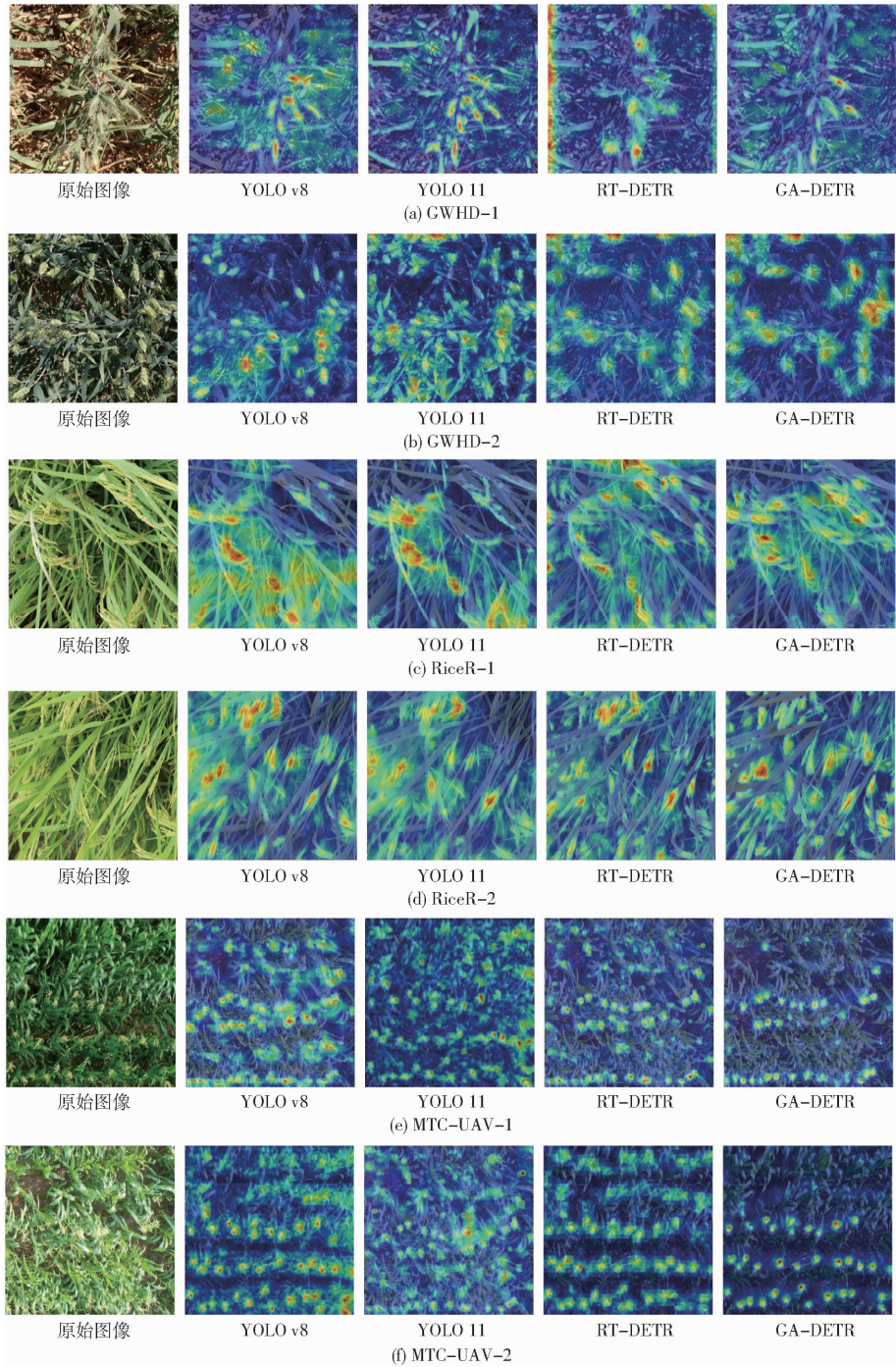


图 11 Grad-CAM 注意力可视化对比

Fig. 11 Comparison of Grad-CAM attention visualization

0.908(图 12), 优于 RT - DETR 的 0.896。该模型表现出卓越的预测精度, 具体如下: ①手动计数和预测计数之间有很强的线性相关性。②平均绝对误差

(MAE) 为 5.650。③均方根误差 (RMSE) 为 7.383。这些指标共同表明, 在小麦穗检测任务中, 模型拟合较好且误差较小。

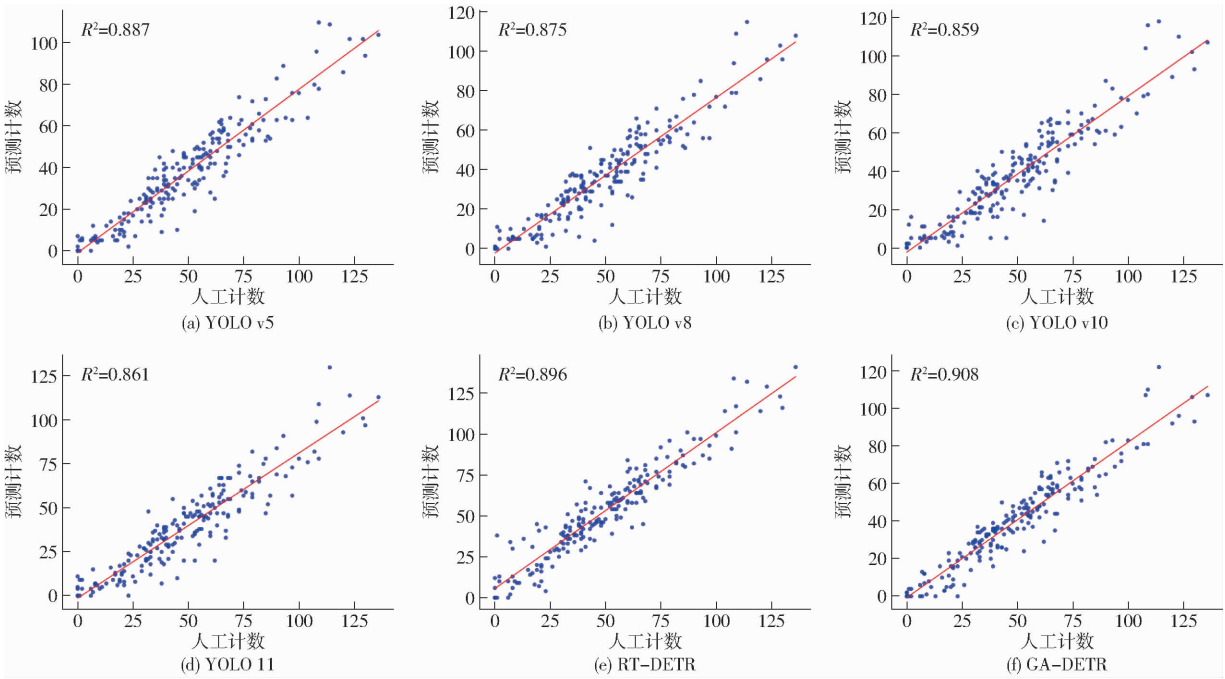


图 12 模型在小麦穗计数相关性上的比较 (GWHD 数据集)

Fig. 12 Comparison of models on wheat spike count prediction correlation (GWHD dataset)

在 RiceR 水稻数据集上的测试中, GA - DETR 同样表现优异, 其 R^2 达到 0.889, MAE 和 RMSE 分别为 3.665 和 5.396, 优于其他对比模型 (表 6), 显示出模型对密集穗型作物仍具有良好的适应性和准确性。

表 6 计数指标比较

Tab. 6 Comparison of counting metrics

数据集	模型	MAE	RMSE	R^2
GWHD	YOLO v5m	6.097	7.925	0.887
	YOLO v8m	6.361	8.393	0.875
	YOLO v10m	6.999	9.293	0.859
	YOLO 11m	7.109	9.340	0.861
	RT - DETR	7.008	9.137	0.896
	GA - DETR	5.650	7.383	0.908
RiceR	YOLO v5m	3.863	6.189	0.839
	YOLO v8m	3.832	5.551	0.887
	YOLO v10m	4.628	5.918	0.768
	YOLO 11m	4.005	5.814	0.870
	RT - DETR	4.747	6.187	0.878
	GA - DETR	3.665	5.396	0.889
MTC - UAV	YOLO v5m	5.597	7.168	0.853
	YOLO v8m	2.667	3.494	0.955
	YOLO v10m	3.699	4.754	0.918
	YOLO 11m	4.160	5.216	0.901
	RT - DETR	3.597	4.582	0.927
	GA - DETR	3.323	4.207	0.936

在 MTC - UAV 无人机采集的玉米穗计数任务中, YOLO v8m 表现最佳, 其 MAE (2.667)、RMSE (3.494) 和 R^2 (0.955) 均优于其他所有模型。GA - DETR 在该数据集上的表现为: MAE 为 3.323、RMSE 为 4.207、 R^2 为 0.936, 优于 YOLO v5m、YOLO v10m, 但与 YOLO v8m 存在差距, 这表明针对无人机高空成像的小目标场景, 模型仍有优化空间。

总体而言, GA - DETR 在 GWHD 和 RiceR 数据集上展现了较优的综合性能, 特别是在传统地面采集图像和密集穗型作物上表现出色。尽管在 MTC - UAV 数据集上未取得最优结果, 但其性能仍稳定居于前列, 展现了模型的泛化能力。

上述结果证明了 GA - DETR 在复杂田间环境下面临稻穗重叠、密集分布等挑战时, 凭借其高效的特征提取与注意力机制, 能够实现高精度的穗部计数, 为谷物穗部产量的自动化精准估算提供了可靠的技术支撑。

3 结论

(1) 本研究提出了一种用于谷类作物穗部检测的新型模型——GA - DETR, 该模型可有效应对穗部重叠、尺度变化等复杂场景下的检测挑战, 适用于小麦、水稻、玉米等作物的穗部检测任务。

(2) 针对密集穗部目标检测难题, 设计了门控

动态特征调制模块(GMC2F),该模块将门控卷积机制与跨阶段局部特征结构相结合,有效增强模型对密集区域的判别特征提取能力,提高了目标在复杂分布下的识别精度。

(3)为缓解花穗重叠对检测精度的影响,构建了注意力上采样尺度序列特征融合模块(AUSSFF),通过双路径注意力上采样与3D卷积策略强化多尺度特征整合,显著提升了模型对重叠穗部的感知与定位鲁棒性。

(4)面向无人机图像中小目标检测需求,提出了FPIoU自适应损失函数,结合难度分级聚焦策略与尺寸敏感的惩罚机制,优化了不同复杂度样本的

回归权重分配,有效提升小尺度穗部的识别与定位性能。

(5)在多种作物公开数据集(RiceR、GWHD和MTC-UAV)上的试验表明,GA-DETR在mAP@0.5、mAP@0.5:0.95等关键指标上均优于当前主流检测模型。对于RiceR数据集,与基线模型RT-DETR相比,GA-DETR在提升检测精度的同时,模型内存占用量减少32.5%,浮点运算量降低14.4%。本研究构建了一个适用于多种谷物穗部的通用检测框架GA-DETR,不仅克服了光照多变、背景复杂等环境挑战,也为农业图像处理中的跨物种特征建模提供了技术支撑。

参 考 文 献

[1] KHAN M I R, PALAKOLANU S R, CHOPRA P, et al. Improving drought tolerance in rice: ensuring food security through multi-dimensional approaches[J]. *Physiologia Plantarum*, 2020, 172(2): 645-668.

[2] TYCZEWSKA A, WOŹNIAK E, GRACZ J, et al. Towards food security: current state and future prospects of agrobiotechnology[J]. *Trends in Biotechnology*, 2018, 36(12): 1219-1229.

[3] XUE H, XU X, ZHU Q, et al. Rice yield and quality estimation coupling hierarchical linear model with remote sensing[J]. *Computers and Electronics in Agriculture*, 2024, 218: 108731.

[4] 张远琴,肖德琴,陈焕坤,等.基于改进Faster R-CNN的水稻稻穗检测方法[J].*农业机械学报*, 2021,52(8):231-240. ZHANG Yuanqin, XIAO Deqin, CHEN Huankun, et al. Rice panicle detection method based on improved Faster R-CNN[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021,52(8):231-240. (in Chinese)

[5] COINTAULT F, GUERIN D, GUILLEMIN J P, et al. In-field Triticum aestivum ear counting using colour-texture image analysis[J]. *New Zealand Journal of Crop and Horticultural Science*, 2008, 36(2): 117-130.

[6] HASAN M M, CHOPIN J P, LAGA H, et al. Detection and analysis of wheat spikes using convolutional neural networks[J]. *Plant Methods*, 2018, 14(1): 100.

[7] XU C, JIANG H, YUEN P, et al. MHW-PD: a robust rice panicles counting algorithm based on deep learning and multi-scale hybrid window[J]. *Computers and Electronics in Agriculture*, 2020, 173: 105375.

[8] LI L, HASSAN M A, YANG S, et al. Development of image-based wheat spike counter through a Faster R-CNN algorithm and application for genetic studies[J]. *Crop Phenotyping Studies with Application to Crop Monitoring*, 2022, 10(5): 1303-1311.

[9] ZHANG S, YANG Y, TU L, et al. Comparison of YOLO-based sorghum spike identification detection models and monitoring at the flowering stage[J]. *Plant Methods*, 2025, 21(1): 20.

[10] 韩科立,王振坤,余永峰,等.基于改进YOLO 11n模型的棉花田间复杂环境障碍物检测方法[J].*农业机械学报*, 2025, 56(5):111-120. HAN Keli, WANG Zhenkun, YU Yongfeng, et al. Obstacle detection method for complex cotton field environments based on improved YOLO 11n model[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2025,56(5):111-120. (in Chinese)

[11] LI R, SUN X, YANG K, et al. A lightweight wheat ear counting model in UAV images based on improved YOLOv8[J]. *Frontiers in Plant Science*, 2025, 16:1536017.

[12] YE J, YU Z, WANG Y, et al. WheatLFANet: in-field detection and counting of wheat heads with high-real-time global regression network[J]. *Plant Methods*, 2023, 19(1): 103.

[13] CHEN R, LU H, WANG Y, et al. High-throughput UAV-based rice panicle detection and genetic mapping of heading-date-related traits[J]. *Frontiers in Plant Science*, 2024, 15:1327507.

[14] LIANG Y, LI H, WU H, et al. A rotated rice spike detection model and a crop yield estimation application based on UAV images[J]. *Computers and Electronics in Agriculture*, 2024, 224: 109188.

[15] 姜海燕,徐灿,陈尧,等.基于田间图像的局部遮挡小尺寸稻穗检测和计数方法[J].*农业机械学报*, 2020,51(9):152-162. JIANG Haiyan, XU Can, CHEN Yao, et al. Detecting and counting method for small-sized and occluded rice panicles based on in-field images[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2020,51(9):152-162. (in Chinese)

[16] KHAKI S, SAFAEI N, PHAM H, et al. WheatNet: a lightweight convolutional neural network for high-throughput image-based wheat head detection and counting[J]. *Neurocomputing*, 2022, 489: 78-89.

[17] TENG Z, CHEN J, WANG J, et al. Panicle-Cloud: an open and ai-powered cloud computing platform for quantifying rice panicles from drone-collected imagery to enable the classification of yield production in rice[J]. *Plant Phenomics*, 2023, 5:

0105.

[18] BHAGAT S, KOKARE M, HASWANI V, et al. WheatNet-Lite: a novel light weight network for wheat head detection[C]//International Conference on Computer Vision, 2021.

[19] WEN C, WU J, CHEN H, et al. Wheat spike detection and counting in the field based on spikeretinanet[J]. Frontiers in Plant Science, 2022, 13: 821717.

[20] 于俊伟, 陈威威, 郭园森, 等. 基于改进 Oriented R-CNN 的旋转框麦穗检测与计数模型[J]. 农业工程学报, 2024, 40(6): 248-257.

YU Junwei, CHEN Weiwei, GUO Yuansen, et al. Improved Oriented R-CNN-based model for oriented wheat ears detection and counting[J]. Transactions of the CSAE, 2024, 40(6): 248-257. (in Chinese)

[21] 鲍文霞, 苏彪彪, 胡根生, 等. 基于 FE-P2Pnet 的无人机小麦图像麦穗计数方法[J]. 农业机械学报, 2024, 55(4): 155-164, 289.

BAO Wenxia, SU Biaobiao, HU Gensheng, et al. Method for counting wheat ears in uav images based on FE-P2Pnet[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(4): 155-164, 289. (in Chinese)

[22] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, arXiv:1706.03762.

[23] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//European Conference on Computer Vision, 2020: 213-229.

[24] ZHAO Y, LV W, XU S, et al. DETRs beat YOLOs on real-time object detection[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024: 16965-16974.

[25] 曹英丽, 刘亚帝, 马殿荣, 等. 基于最优子集选择的水稻穗无人机图像分割方法[J]. 农业机械学报, 2020, 51(8): 171-177, 188.

CAO Yingli, LIU Yadi, MA Dianrong, et al. Best subset selection based rice panicle segmentation from UAV image[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(8): 171-177, 188. (in Chinese)

[26] 陈鹏, 王兴蓉, 章军, 等. 基于改进 YOLOv9m 的多品种玉米雄穗检测方法[J]. 农业工程学报, 2025, 41(17): 122-131.

CHEN Peng, WANG Xingrong, ZHANG Jun, et al. Multi-varieties maize tassels detection method based on improved YOLOv9m[J]. Transactions of the CSAE, 2025, 41(17): 122-131. (in Chinese)

[27] TAN S, LU H, YU J, et al. In-field rice panicles detection and growth stages recognition based on RiceRes2Net[J]. Computers and Electronics in Agriculture, 2023, 206: 107704.

[28] DAVID E, MADEC S, SADEGHI-TEHRAN P, et al. Global wheat head detection (GWHD) dataset: a large and diverse dataset of high-resolution rgb-labelled images to develop and benchmark wheat head detection methods[J]. Plant Phenomics, 2020, 2020: 3521852.

[29] DAVID E, SEROUART M, SMITH D, et al. Global wheat head detection 2021: an improved dataset for benchmarking wheat head detection methods[J]. Plant Phenomics, 2021, 2021: 9846158.

[30] LU H, LIU L, LI Y N, et al. TasselNetV3: explainable plant counting with guided upsampling and background suppression[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-15.

[31] SONG Y, ZHOU Y, QIAN H, et al. Rethinking performance gains in image dehazing networks[J]. Computer Vision and Pattern Recognition, 2022, arxiv preprint arxiv:2209.11448.

[32] CHOLLET F. Xception: deep learning with depthwise separable convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1251-1258.

[33] KANG M, TING C M, TING F F, et al. ASF-YOLO: a novel yolo model with attentional scale sequence fusion for cell instance segmentation[J]. Image and Vision Computing, 2024, 147: 105057.

[34] ZHANG H, ZHANG S. Focaler-IoU: more focused intersection over union loss[J]. Computer Vision and Pattern Recognition, 2024, arXiv preprint arXiv: 2401.10525.

[35] LIU C, WANG K, LI Q, et al. Powerful-IoU: more straightforward and faster bounding box regression loss with a nonmonotonic focusing mechanism[J]. Neural Networks, 2024, 170: 276-284.

[36] KHANAM R, HUSSAIN M. What is YOLOv5: a deep look into the internal features of the popular object detector[J]. Computer Vision and Pattern Recognition, 2024, arXiv preprint arXiv: 2407.20892.

[37] WANG A, CHEN H, LIU L, et al. Yolov10: real-time end-to-end object detection[J]. Advances in Neural Information Processing Systems, 2024, 37: 107984-108011.

[38] JOCHER G, QIU J. Ultralytics YOLO 11 (Version 11.0.0)[CP]. <https://github.com/ultralytics/ultralytics>.