基于CornYOLO的冠层内玉米果穗目标检测方法

高光甫[1], 王启磊[2], 宋丽雯[1], 冯海宽[3], 时 雷[1], 杨 浩[3], 刘 杨[1], 岳继博[1*]

(1. 河南农业大学 信息与管理科学学院,河南郑州 450002,中国; 2. 河南金苑种业有限公司,河南郑州 450002,中国; 3. 北京市农林科学院 信息技术研究中心/农业农村部农业遥感机理与定量遥感重点实验室,北京 100097,中国)

摘 要: **[目的/意义]** 玉米是主要粮食作物,果穗作为玉米关键表型性状,其形态、大小及颜色特征能够有效反映植株生长状态及潜在产量。传统的田间玉米果穗检测依赖人工,效率低且劳动强度大。随着密植栽培模式的推广,玉米冠层结构愈发密集,人工进入田间开展果穗测量不仅操作困难,还容易对植株造成机械损伤,进一步限制了数据的准确性与代表性。因此,亟需高效的自动化检测技术。**[方法]** 为实现复杂田间环境下玉米果穗的高效精准检测,提出一种基于改进YOLO11n(You Only Look Once 11)的CornYOLO模型。创新性地采用无人车搭载全景相机进行图像采集,构建了高质量的田间数据集,并在此基础上提出了3项核心模型改进:1)采用动态点空间注意力的跨阶段部分网络(Cross Stage Partial Network with Dynamic Pointwise Spatial Attention, C2PDA)以提升对遮挡目标的识别鲁棒性;2)引入特征优化模块(Feature Refinement Module, FRM)以增强多尺度目标检测能力;3)使用统一交并比(Unified Intersection Over Union, UIoU)损失函数以优化边界框回归精度。为作物田间表型高通量获取提供了一种从数据采集到智能识别的端到端解决方案。**[结果和讨论]** CornYOLO在复杂田间环境下表现出优异的检测性能,在验证集上mAP@50达到89.3%,相较于YOLO11n,F_1分数提升2.5个百分点。相较于其余基线模型,其mAP@50提升显著,最高达12.6个百分点。消融实验表明,C2PDA、FRM与UIoU这3个模块均对性能提升有积极贡献,其中C2PDA作用最为关键。**[结论]** CornYOLO模型能够高效精准地识别田间玉米果穗,为玉米育种表型分析和产量预测提供可靠的技术支持,推动玉米果穗信息提取的智能化发展。

关键词: 玉米果穗;无人车;YOLO11;全景相机;目标检测

中图分类号: TP181 **文献标识码:** A **文章编号:** SA202509005

引用格式:高光甫, 王启磊, 宋丽雯, 冯海宽, 时雷, 杨浩, 刘杨, 岳继博. 基于CornYOLO的冠层内玉米果穗目标检测方法[J]. 智慧农业(中英文), 2026, 8(1): 167-177. DOI: 10.12133/j.smartag.SA202509005

GAO Guangfu, WANG Qilei, SONG Liwen, FENG Haikuan, SHI Lei, YANG Hao, LIU Yang, YUE Jibo. Object Detection Method of Maize Ears Within Canopy Based on CornYOLO[J]. Smart Agriculture, 2026, 8(1): 167-177. DOI: 10.12133/j.smartag.SA202509005 (in Chinese with English abstract)

0 引 言

玉米是全球主要粮食作物之一,广泛应用于食品加工,具有重要的营养价值和经济意义[1,2]。培育高产品种并实施科学种植管理是提升玉米产量和质量的核心策略[3]。果穗作为玉米关键表型性状,其形态、大小及颜色特征能够有效反映植株生长状态及潜在产量[4,5]。因此,精准获取不同生育期田间玉米果穗图像,对于育种研究和产量预测具有重要价值[6]。目前,田间玉米果穗表型检测主要依赖人工操作,劳动强度高、效率低下,难以满足现代育种研究的高通量需求。当前,基于地面或者近地面无人机/无人车的作物表型分析技术正处于快速发展阶段,已成为智慧农业和精准育种的核心工具[7,8]。通过搭载多光谱、高光谱、热红外及可见光传感器,无人机能够高效获取作物冠层的高分辨率图像,实现对植株高度、叶面积指数、生物量、胁迫响应等关键表型性状的无损监测[9-11]。

结合人工智能算法,研究人员能够自动化提取和分析海量图像数据,显著提升表型识别的精度与

收稿日期:2025-09-02

基金项目:河南省自然科学基金(252300421839);国家自然科学基金项目(42101362, 42571462)

作者简介:高光甫,硕士研究生,研究方向为无人机图像处理。E-mail:gaoguangfu@stu.henau.edu.cn

*通信作者:岳继博,博士,副教授,研究方向为无人机遥感图像分析。E-mail:yuejibo@henau.edu.cn

copyright©2026 by the authors (登录www.smartag.net.cn免费获取电子版全文)

效率[12, 13]。目标检测技术在农业领域的应用正迅速拓展，通过高效识别与定位作物、杂草、病虫害及成熟果实，显著提升农业生产的智能化水平[14-16]。借助深度学习模型，无人机或地面设备可实时分析田间图像，精准监测作物生长状态、区分杂草与作物以实现精准施药，或自动识别果实成熟度引导机器人采收[17-19]。当前，目标检测技术作为计算机视觉的核心分支，已在玉米表型检测领域得到深入且广泛的应用，显著推动了作物监测与性状提取的自动化进程。例如，ALZADJALi等[20]提出一种基于YOLO（You Only Look Once）的检测模型，用于从无人机图像中提取玉米雄穗，实现了92%的检测准确率。该方法通过融合多光谱图像技术，有效支持了田间实时监测，但仍需应对云层遮挡与光照变化带来的干扰。QI等[21]则开发了MT-YOLO（Missed Tassels-YOLO）模型，专门用于识别杂交玉米中的遗漏雄穗，其平均精度达到93.1%，体现出优异的识别鲁棒性。FALAHAT和KARAMI[22]提出了一种改进的YOLOv5n网络，用于玉米穗的检测与计数，该模型通过结构优化显著提高了计算效率，能够在单位时间内处理更多图像，在保持高精度的同时展现出卓越的实时性能与吞吐量优势。为推进玉米田病害的精准防治，YADAV等[23]开发了一套基于YOLOv5模型的无人机智能喷药系统。该系统通过机载视觉设备实时捕捉田间影像，并利用轻量化的YOLOv5模型快速识别病害区域，进而驱动无人机实现针对性定点施药。这一方法有效实现了对玉米田病害的实时监测与防控，为精准农业中的病害管理提供了可行的技术路径。目前，玉米表型研究已在冠层性状检测方面取得了不错的成果。冠层内部的结构、果穗及籽粒等性状，因其蕴含更为直接和关键的表型信息，在产量评估与病虫害识别等应用中具有重要研究价值。例如，KHAKI等[24]开发了一种基于卷积神经网络（Convolutional Neural Networks, CNN）的玉米籽粒检测网络。该模型基于包含不同光照条件的采摘后玉米数据集进行训练，能够有效应对非标准光照环境，实现玉米雌穗图像的自动籽粒定位与计数。SPRAGUE等[25]为量化玉米果穗的生长角度，基于YOLOv8目标检测架构，提出了一种先进的果穗朝向估计模型。该方法通过计算机视觉技术实现了对果穗空间姿态的自动化解译，有效替代了传统人工测量。赵仲文等[26]基于生产线采集的新鲜玉米数据集，开发了用于检测采摘后玉米果穗优劣的SS-YOLOv8（SPDConv SimAM-YOLOv8）模型，该模型兼顾了检测精度与轻量化需求，为鲜食玉米的自动化分选提供了可行方法。为评估机械脱粒后的玉米质量，FU等[27]基于RGB图像，利用图像处理与分析方法对完整籽粒与损伤部分进行区分，为该环节的质量控制提供了技术依据。同样，CHEN等[28]利用改进的YOLOv7模型，对采摘后的玉米果穗进行检测，实现了对缺粒和苞叶残留等质量问题的有效识别。值得注意的是，上述关于玉米冠层内果穗识别研究，多依赖于实验室环境下采集的采摘后果穗图像，然而，在真实田间场景中，玉米植株种植密集、冠层层次复杂，果穗通常位于植株中下部，常受到叶片和茎秆的严重遮挡，导致基于无人机或手持设备的图像采集难以获取完整、清晰的果穗信息。叶片与茎秆的密集遮蔽不仅削弱了果穗的可见区域，导致视觉特征提取不完整，也进一步加大了复杂场景下稳健检测的难度。这也导致当前目标检测算法在果穗识别中普遍存在识别精度有限、误检与漏检率高等问题。

综上，当前图像检测技术在玉米果穗检测场景下，仍面临以下问题。一是由于玉米植株种植密集、冠层结构复杂，采用人工或无人机平台均难以获取高质量的冠层内部图像。二是现有检测模型在复杂田间环境下识别精度仍有待进一步提升。三是目前研究多聚焦于实验室环境下果穗检测，而针对大田环境下果穗检测研究相对稀缺。

针对上述问题，本研究旨在开发一种适用于大田复杂环境的玉米果穗检测方案。通过搭载全景相机的地面无人车，系统采集试验田不同小区的玉米图像数据。在算法层面，本研究基于YOLO11n模型提出一种改进的CornYOLO检测框架，以提升田间自然场景下果穗检测的精度与效率，以期为玉米育种表型解析与产量预估提供可靠的技术支持。

1 数据与方法

1.1 研究区与现场试验设计

研究区位于河南省荥阳市（34°36′～34°59′N，113°7′～113°30′E），地处中纬度地带，属暖温带大陆性季风气候，年平均气温14.3 ℃，平均年降水量约为608.8 mm。研究区试验田含80组玉米育种品系，每个小区的尺寸为2.5 m × 5 m，小区共种植4行玉米。于2023年夏季玉米的6个关键生育期（P_1～P_6）采集田间图像。采集时期如下：P_1（8月11日），P_2（8月18日），P_3（9月1日），P_4（9月6

日），P_5（9月14日），P_6（9月21日）。

实验过程中使用大疆机甲大师S1（中国深圳市大疆创新科技有限公司）搭载全景相机（Insta360 X3）来捕获试验田不同小区的玉米图像，其中全景相机搭载1个4 800万像素的传感器。采集时，为保障相机与玉米果穗处于相近高度，全景相机通过延长杆固定于地面无人车上（距地约1.5 m）。在每1小区的中央位置停放无人车进行图像捕获，确保成像质量和完整覆盖。研究共采集480张图片，其中每个小区在每个时期对应1张图片，基于无人车和全景相机的玉米果穗图像采集方案如图1所示，其中图1a为无人车，图1b为数据采集使用的全景相机，图1c为搭载全景相机的无人车，图1d为使用实验设备拍摄的图像。



a. 大疆机甲大师S1　b. 全景相机　c. 搭载了全景相机的无人车　d. 实际拍摄的田间图片

图1　基于无人车和全景相机的玉米果穗图像采集方案

Fig. 1　Corn ear image acquisition scheme based on unmanned vehicle and panoramic camera

1.2　数据标注和预处理

对原始数据进行随机裁剪，设定最大重合率不超过35%，最终获得1 152张像素尺寸为4 480×4 480的图像。部分数据如图2所示，其中，图2a为未成熟玉米田间图像，图2b为成熟玉米田间图像。从图2的样本可见，复杂的田间环境给检测带来了巨大困难：果穗不仅被茎叶严重遮挡，而且其颜色与背景趋于一致，这使得准确识别的难度大大增加。之后使用LabelImg工具对图像进行标注，生成1 152张标注样本。实验采用动态数据增强策略，通过在每次训练迭代中对图像应用随机变换显著增加模型接触的视觉数据多样性。该策略有效提升模型的泛化能力，并有助于抑制过拟合现象。动态数据增强参数配置如表1所示。

数据标注后将数据集按照4∶1∶1的比例划分为训练集、测试集、验证集，具体包括768张训练图像、192张测试图像和192张验证图像。验证集用于评估模型的泛化性和准确性，而测试集用于评估模型的稳健性。



a. 未成熟玉米　b. 成熟玉米

图2　部分田间玉米图像展示

Fig. 2　Partial field corn image data display

表1　CornYOLO动态数据增强参数配置

Table 1　CornYOLO dynamic data augmentation parameter configuration

变量	数值
Fliplr	0.500
hsv_h	0.015
hsv_s	0.700
hsv_v	0.400
Translate	0.100
Scale	0.500

1.3　CornYOLO模型

YOLO11n是KHANAM和HUSSAIN[29]基于YOLOv8提出的改进模型，包含5个不同版本：YOLO11n、YOLO11s、YOLO11m、YOLO11l和YOLO11x。各版本因特征提取模块和卷积核结构设计的差异，在参数量与计算复杂度上存在显著不同。尽管YOLO11在多数场景表现优异，但在复杂

田间环境下仍面临挑战，例如对图像中尺寸差异显著的玉米目标易产生漏检，且在玉米果穗受茎叶遮挡时检测性能大幅下降。综合考虑训练资源消耗与模型性能要求，本研究选择以YOLO11n为基础进行改进，旨在提升其对复杂田间玉米果穗的检测能力。CornYOLO相对于YOLO11主要有以下改进：

（1）提出动态点空间注意力的跨阶段部分网络（Cross Stage Partial Network with Dynamic Pointwise Spatial Attention, C2PDA），用于替换逐点空间注意力的跨阶段部分网络（Cross Stage Partial with Pointwise Spatial Attention， C2PSA）。该模块通过动态通道加权和位置感知建模提升对遮挡目标的识别性能。

（2）引入特征优化模块（Feature Rrefinement Module, FRM），以替换YOLO11n中的空间金字塔池化-快速（Spatial Pyramid Pooling-Fast， SPPF）模块。该模块通过双向池化和空间-通道协同注意力机制增强多尺度目标检测能力，优化大尺度玉米果穗的空间定位。

（3）采用统一交并比（Unified Intersection Over Union, UIoU）损失函数，通过动态权重分配优化低质量检测框，提升模型在密集目标场景下的定位精度和收敛稳定性。

基于上述改进构建的CornYOLO模型，其整体网络结构如图3所示。

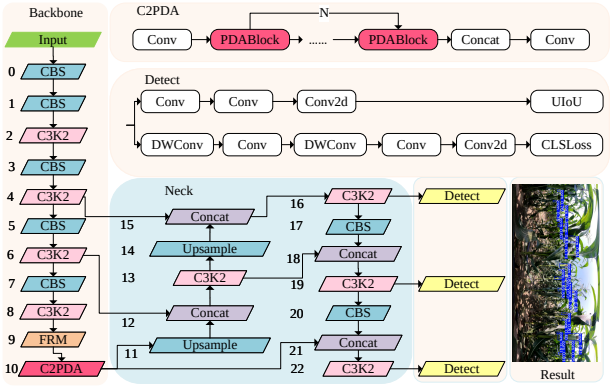


图3 CornYOLO模型架构

Fig. 3 Model architecture of CornYOLO

1.3.1 C2PDA改进模块

C2PDA模块通过协同整合动态通道加权机制与位置感知建模，增强特征表达的空间判别能力和通道敏感性，从而显著提升模型在茎叶遮挡、多尺度目标等复杂农田场景下的玉米果穗识别性能。其核心部分动态点空间注意力（Dynamic Pointwise Spatial Attention, PDA）模块结构如图4所示。

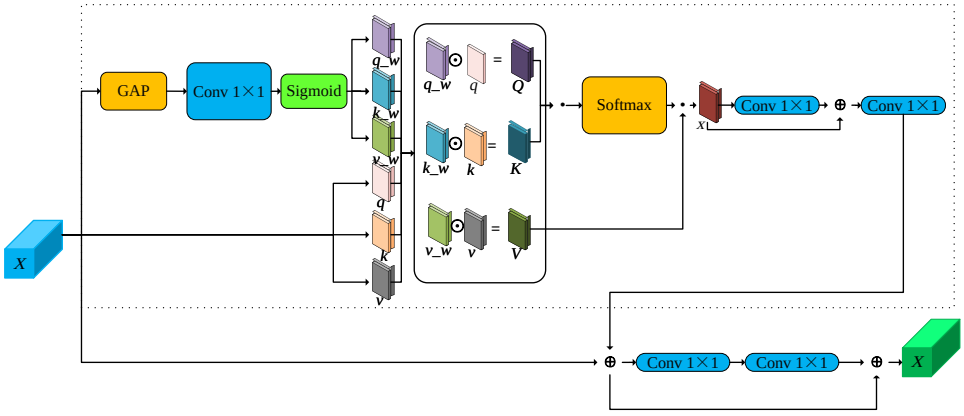


图4 PDA模块架构

Fig. 4 Module architecture of PDA

PDA模块基于多头注意力框架融入通道感知的动态权重生成和卷积位置编码机制，对于维度为$X \in R^{B \times C \times H \times W}$的输入特征图，模块通过通道注意力机制生成动态权重张量$\boldsymbol{Y} \in R^{B \times 3 \times H_{num} \times 1 \times 1}$，如公式（1）所示，并将其解耦为$W_q$，$W_k$，$W_v$这3组权重，如公式（2）所示，分别对多头注意力的$Q$、$K$和$V$特征进行通道级重加权，如公式（3）～公式（5）所示。

$$Y_{b,k,1,1} = \sigma\left(\sum_{C=1}^{C} W_{k,c} \times \left(\frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W} X_{b,c,i,j}\right) + b_k\right) \quad (1)$$

$$\boldsymbol{Q}_w, \boldsymbol{K}_w, \boldsymbol{V}_w = \mathrm{split}\left(Y_{b,k,1,1}\right) \quad (2)$$

$$\boldsymbol{Q} = \boldsymbol{Q} \odot \boldsymbol{Q}_w \quad (3)$$

$$\boldsymbol{K} = \boldsymbol{K} \odot \boldsymbol{K}_w \quad (4)$$

$$\boldsymbol{V} = \boldsymbol{V} \odot \boldsymbol{V}_w \quad (5)$$

式中：b为当前处理的样本在批次中的位置；k为当前计算的输出通道位置；c为当前求和的输入

通道位置；i为特征图高度方向索引；j为特征图宽度方向索引；X为输入特征；$W_{k,c}$为权重参数；b_k为偏置参数；$\sigma()$为Sigmoid函数；$\boldsymbol{Q}_w$为查询向量的通道注意力权重；$\boldsymbol{K}_w$为键向量的通道注意力权重；$\boldsymbol{V}_w$为值向量的通道注意力权重；$\boldsymbol{Q}$为查询向量；$\boldsymbol{K}$为键向量；$\boldsymbol{V}$为值向量。

同时，原始位置编码机制被重构优化，以使其在复杂田间环境中能表现出更优异的性能。为确保位置信息与语义表征处于统一特征空间，在本研究中，位置编码信息被添加到经过注意力聚合后的特征，而非原始特征。

1.3.2　FRM特征优化模块

YOLO11通过SPPF块实现多尺度目标检测。SPPF借助单一尺寸的池化核进行递归池化操作，有效提升模型对不同尺寸目标的检测性能。然而，其串行处理机制可能导致不同尺度特征间的平衡性不足，在特定场景下更倾向于关注较小目标特征，从而可能影响大尺度玉米果穗检测的精度。为优化模型在玉米田间对不同尺度玉米果穗的精确识别能力，本研究采用FRM模块[30]替代原始的SPPF模块。

FRM是一种基于方向性特征分解与自适应注意力机制的特征优化模块。它通过水平与垂直双向池化捕获空间结构信息，并结合空间-通道协同注意力实现特征动态校准。模块采用异构卷积生成通道权重图，并利用多尺度空间池化提取方向敏感特征；随后通过双向特征融合与像素级权重分配，显著增强目标区域的特征响应强度。其结构如图5所示。在复杂农业场景的玉米果穗检测任务中，FRM能有效提升多尺度目标的鉴别能力，尤其通过抑制茎叶背景干扰和优化大尺度玉米果穗的空间定位特征，克服传统串行池化结构的尺度感知偏差问题，实现检测精度与泛化性的协同提升。

1.3.3　UIoU回归损失函数

损失函数用于衡量模型预测与真实值之间的偏差程度。经典的YOLO11模型采用CIoU（Complete Iou）损失函数，该方法易受低质量预测框干扰，尤其是当预测框与真实框共中心点或长宽比相同时，会丧失梯度校准能力，导致其在密集目标场景下存在局限性。基于此，本研究采用UIoU作为替代方案。

UIoU[31]是一种创新的边界框回归损失函数，它关注不同质量预测框之间的权重分配，通过IoU值本身或其单调函数来自适应地调整每个预测框损

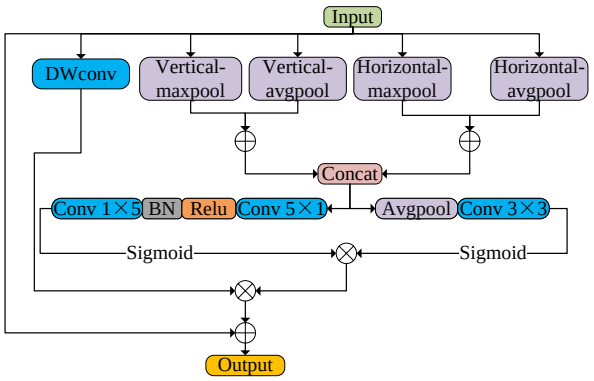


图5　FRM模块架构

Fig. 5　Module architecture of FRM

失项的权重，赋予较低质量预测框更高的权重，促使其优先优化，而较高质量预测框的权重相对降低以避免过度调整。这种分配策略使得模型在训练过程中能够更有效地聚焦于更具提升空间的困难样本，提升模型在各种复杂场景下的定位精度和收敛稳定性。UIoU的计算公式如公式（6）和公式（7）所示。

$$\mathrm{IoU}=\frac{B_{\mathrm{gt}}\cap B_{\mathrm{pred}}}{B_{\mathrm{gt}}\cup B_{\mathrm{pred}}} \tag{6}$$

$$\mathrm{UIoU}=1-\mathrm{IoU}+\lambda\frac{d^2\left(B_{\mathrm{gt}},B_{\mathrm{pred}}\right)}{L^2}+v\times\alpha_v+\beta(\Delta\omega+\Delta\mathrm{h}) \tag{7}$$

式中：B_{gt}为真实边界框的面积，像素的平方；B_{pred}为预测边界框的面积，像素的平方；$d^2\left(B_{\mathrm{gt}},B_{\mathrm{pred}}\right)$为这两个边界框的中心之间的欧几里得距离；$L$为覆盖两个边界框的最小封闭框的对角线长度；$\lambda$为中心距离权重系数，调节预测框与真实框中心点偏离的惩罚强度；$v\times\alpha_v$为长宽比一致性分量及其关联的校正因子；$\beta(\Delta\omega+\Delta\mathrm{h})$，为尺寸差异调制系数。

1.4　基准模型和评价指标

为全面评估CornYOLO模型的性能，本研究选取当前主流或集成最新检测技术的基准模型进行对比实验，确保基准方法代表领域前沿水平。在YOLO系列中，本研究选择YOLOv3-tiny[32]、YOLOv5n[33]、YOLOv8n[34]、YOLOv10n[35]和YOLO13n[36]，这些模型代表该系列不同的开发阶段，并且为当前主流或最新技术。其他技术方面选择RT-DETR（Real-Time Detection Transformer）模型为额外的基线模型。

本研究使用装配有Python 3.10.18，Pytorch

2.5.1 环境的笔记本进行模型的构建和改进，并在 Windows 11 操作环境下进行训练和验证。电脑搭载 Intel i7-13650HX 的 CPU，内存大小为 24 GB，GPU 为 NVIDIA GeForce RTX 4070，显存大小为 8 GB。所有实验均采用统一的超参设置，批量大小（Batch_size）为 8，学习率（Learning Rate）为 0.02，优化器为 AdamW，训练轮次（Epochs）为 200。

研究采用四项评估指标对模型性能进行定量分析，即平均精度均值（Mean Average Precision, mAP），精确率（Precision, P），召回率（Recall, R），以及 F_1 分数（F_1-Score）。这些指标通过以下数学定义实现量化评估，如公式（5）～公式（8）所示。

$$\mathrm{mAP}=\frac{1}{N}\sum_{i=1}^{N}\int_0^1 p_i(r)\,dr \tag{8}$$

$$P=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{9}$$

$$R=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{10}$$

$$F_1=\frac{2\cdot \mathrm{P}\cdot \mathrm{R}}{\mathrm{P}+\mathrm{R}} \tag{11}$$

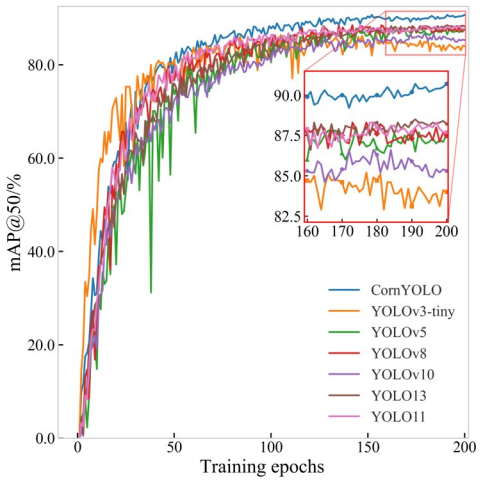
式中：N 为类别数；$p_i(r)$ 为类别 r 的精度-召回率曲线；TP 为正确检测为阳性样品的数量；FP 为错误检测为阳性样品的数量；FN 为错误检测为阴性样品的数量。

2 结果与分析

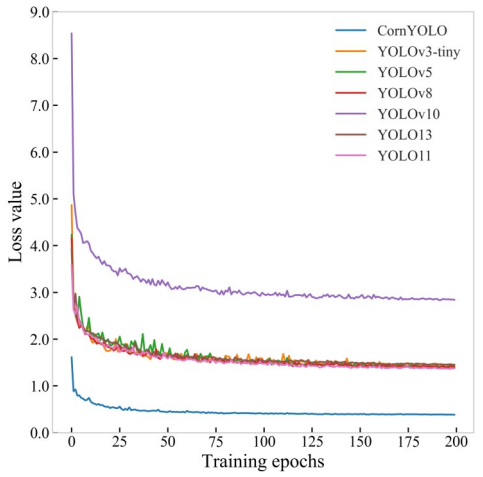
2.1 CornYOLO 与不同模型对比

图 6a 和图 6b 展示了模型训练过程中 mAP@50 和 box_loss 变化曲线。从图 6b 中可以看出，所有模型均在训练到 150 轮时 box_loss 趋于稳定，说明 200 轮训练足以满足训练需求。从图 6a 可以看出，CornYOLO 在训练阶段明显优于其他模型。此外可以看出，YOLOv3-tiny 表现相对较差，mAP@50 值小于 85.0%，YOLO13n、YOLO11n、YOLOv8n，YOLOv5n 达到 85.0% 以上的 mAP@50 值。特别的，只有 CornYOLO 达到 90% 以上的 mAP@50 值。

表 2 展示出不同基准模型的玉米果穗目标检测结果对比。从表 2 可以看出，YOLO11n 在验证集上的表现优于其他模型。YOLO11n 在复杂玉米田场景下的检测能力略优于其他版本，而本研究提出的 CornYOLO 模型进一步实现了显著提升，相较于 YOLO11n、YOLOv3-tiny、YOLOv5n、YOLOv8n、YOLOv10n、YOLO13n 和 RT-DETR，CornYOLO



a. mAP@50 曲线变化



b. box_loss 变化曲线

图 6 训练过程中不同模型测试集 mAP@50 和 box_loss 变化曲线

Fig. 6 Variation curves of mAP@50 and box_loss on the test set for different models during training

的 mAP@50 分别提升 1.1、2.2、1.9、1.8、5.7、2.4 和 12.6 个百分点，F_1 分数分别提升 2.5、2.6、3.3、2.7、5.8、3.2 和 12.6 个百分点。这证实 CornYOLO 能更精准地识别复杂田间环境中的玉米目标，并在漏检与误检之间取得更好的平衡。

2.2 C2PDA 模块有效性测试

为验证提出的 C2PDA 是否有效，将其与当前主流的改进方法进行对比，结果如表 3 所示。对原本的 C2PSA 模块进行改进后，模型性能均有所提升，说明原本的 C2PSA 模块在复杂田间环境中的特征提取能力存在不足。C2PDA 模块在验证集上表现最佳，相较原始 C2PSA 模块，mAP@50 提升

表2　玉米果穗检测研究中不同模型的检测结果

Table 2　Detection experimental results of different models in maize ear detection research

模型	mAP@50/%	*R*/%	*P*/%	F_1
YOLOv3-tiny	87.1	77.5	87.7	82.3
YOLOv5n	87.4	80.2*	82.9	81.6
YOLOv8n	87.5	76.2	88.1	82.2
YOLOv10n	83.6	74.4	84.3	79.1
YOLO11n	88.2	78.3	86.9	82.4
YOLO13n	86.9	76.2	88.1	81.7
RT-DETR	76.7	71.7	72.9	72.3
CornYOLO	89.3*	79.5	91.0*	84.9*

注：*为当前列最高值。

0.5个百分点；与Mona（Multi-cognitive Visual Adapter）、CGLU（Convolutional Gated Linear Unit）和EMA（Efficient Multi-scale Attention）模块相比，C2PDA的mAP@50提升幅度为0.1个百分点～1.2个百分点。说明C2PDA模块有效地提升了模型在复杂田间环境的检测能力。

表3　CornYOLO使用不同模块的实验结果

Table 3　Experimental results of cornYOLO with different modules

模型	mAP@50/%	*R*/%	*P*/%	F_1
C2PSA	88.2	78.3	86.9	82.4
C2PSA-CGLU	88.3	78.8	89.3*	83.8*
C2PSA-EMA	87.5	79.4	85.8	82.5
C2PSA-Mona	88.6	82.1*	84.2	83.1
C2PDA	88.7*	81.5	84.2	82.9

注：*为当前列最高值。

2.3　不同损失函数的实验结果比较

为探究不同损失函数对模型性能的影响，本研究对比了CIoU、UIoU、PIoU（Powerful Iou）、Wise-IoU及SIoU（Scylla Iou）的表现，实验结果如图7所示。结果显示，采用UIoU作为损失函数时，模型取得最优性能，其mAP@50值较基准提升0.6个百分点。相比之下，其余损失函数的提升效果有限，PIoU和Wise-IoU的mAP@50值相较于CIoU分别下降0.8和0.5个百分点，而SIoU的性能则与CIoU相当。该结果表明，UIoU的动态权重调整机制能更好地优化困难样本的学习过程，从而更适用于本研究的复杂农田场景目标检测任务。

2.4　消融实验对比

为评估CornYOLO添加模块的有效性，本研究采用控制变量法进行消融实验，实验结果如表4所示。实验表明，在基础YOLO11n模型上引入各改进模块均有效提升田间玉米果穗的检测性能。其中，C2PDA模块的贡献最为显著，mAP@50提升0.5个百分点，F_1得分提升0.5个百分点。随后，为进一步增强多尺度特征提取能力而引入FRM模块，成功优化对不同尺度目标的检测效果，F_1得分提升1.5个百分点。然而，模块叠加产生低质量检测框，原有损失函数对此类框的优化效率不足，导致更改后模型的mAP@50没有变化。为解决低质量框问题并进一步提升性能，本研究引入UIoU损失函数。UIoU通过动态权重分配机制，赋予低质量检测框更高权重进行优先优化，同时适当降低高质量预测框的权重以避免过度调整，从而有效提升模型精度。最终构建的CornYOLO模型性能优异，相较于原始YOLO11n，F_1得分提升2.5个百分点，mAP@50提升1.1个百分点。实验结果充分表明，CornYOLO在复杂田间环境下对玉米果穗的检测能力较YOLO11n模型实现有效提升。

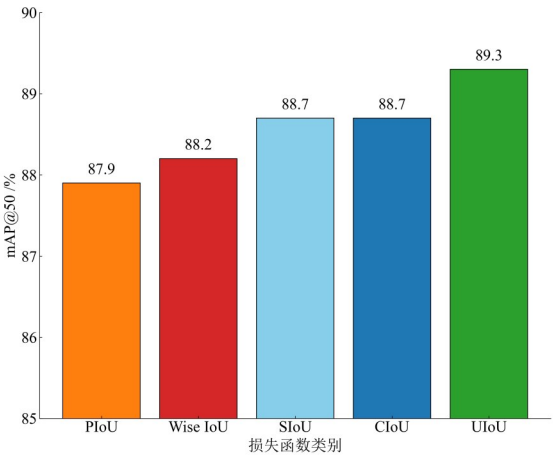


图7　CornYOLO使用不同损失函数时模型mAP@50对比图

Fig. 7　CornYOLO comparison of model mAP@50 with different loss functions

图8展示了添加不同模块后的预测热力图，该可视化结果直观揭示了各模块对检测效果的改进。其中，图8a为验证图像的关键区域放大图，图8b为只添加了C2PDA模块的CornYOLO检测结果热力图，图8c为添加C2PDA和FRM模块的CornYOLO检测结果热力图，图8d为完整的CornYOLO检测结果热力图。

2.5　检测效果可视化分析

为更加直观地展示CornYOLO在复杂田间环境中的识别能力，使用Grad-CAM（Gradient-Weight-

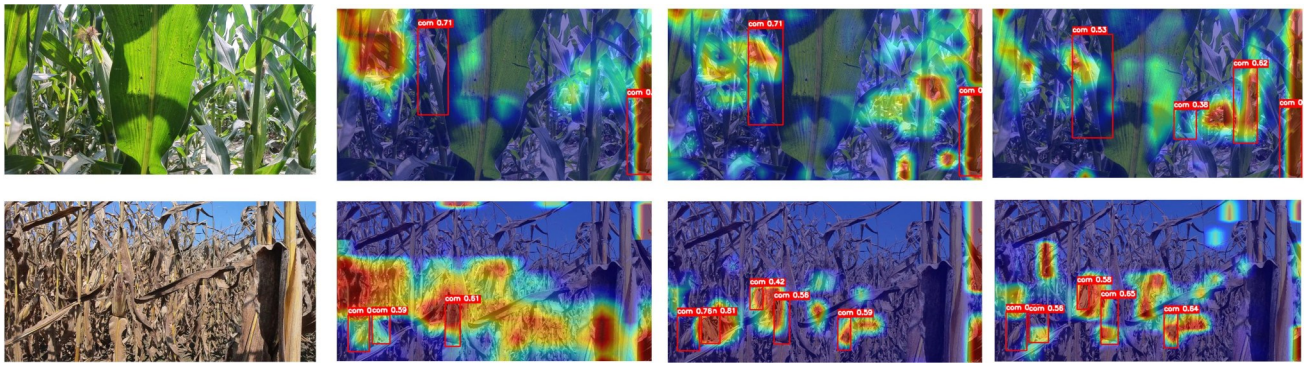
表4 CornYOLO消融实验结果

Table 4 CornYOLO ablation study results

模型				结果			
YOLO11n	C2PDA	FRM	UIoU	mAP@50/%	*R*/%	*P*/%	F_1
√	×	×	×	88.2	78.3	86.9	82.4
√	√	×	×	88.7	81.5*	84.2	82.9
√	×	√	×	87.3	78.2	86.1	82.0
√	×	×	√	87.8	79.2	87.9	83.3
√	√	√	×	88.7	80.4	88.8	84.4
√	√	√	√	89.3*	79.5	91.0*	84.9*

注：*为当前列最高值。√为添加该模块，×为未添加该模块。

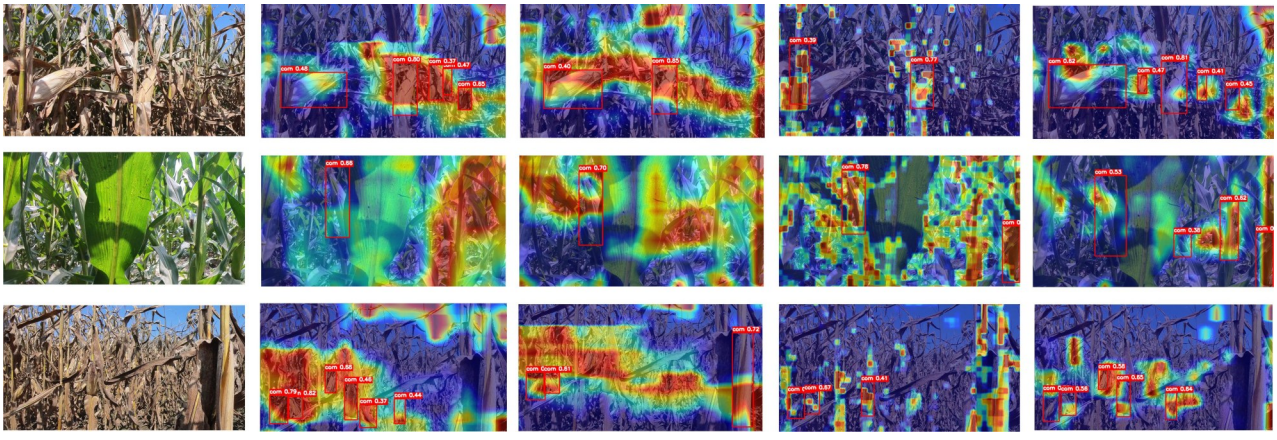
ed Class Activation Mapping）对YOLOv8n、YOLO11n、YOLO13n和CornYOLO的检测结果进行对比，结果如图9b～图9e所示，其中图9a为验证图像的关键区域放大图。可以看出，YOLOv8n、YOLO11n和YOLO13n在处理复杂图像时，其注意力机制易发生分散，尤其当叶片、茎秆与果穗颜色相近时，模型会错误地将注意力集中于背景叶片或茎部，导致误识别或漏检。相比之下，CornYOLO能够更准确地聚焦于果穗区域，即使在果穗被茎部遮挡的情况下，其注意力仍集中于果穗可见部分，而非整个茎秆。这种高度集中的注意力机制是其实现更优检测性能的关键原因。



a. 区域放大图 b. C2PDA c. C2PDA+FRM d. CornYOLO

图8 CornYOLO不同模块效果对比图

Fig. 8 CornYOLO comparison of different modules



a. 区域放大图 b.YOLOv8n c.YOLO11n d.YOLO13n e.CornYOLO

图9 CornYOLO检测效果对比图

Fig. 9 CornYOLO performance comparison

3 讨 论

玉米果穗表型参数是评估玉米产量与品质的关键，其精确提取对精准农业和遗传育种至关重要。然而，在密集的田间环境中实现这些参数的高通量获取，仍是一个待解决的技术挑战[37]。传统田间果穗检测主要依赖人工目视评估，存在效率低下、劳动强度大等局限，难以满足现代农业高通量表型分析的需求。本研究提出一种基于无人车搭载全景相机采集田间图像，并结合CornYOLO模型实现果穗精准识别的自动化方案，为玉米育种提供高效的技术支持[38]。CornYOLO通过集成C2PDA和FRM模块，显著提升模型在复杂田间环境中对遮挡和多

尺度目标的检测能力。C2PDA模块通过动态通道加权和位置感知建模增强特征表示的鲁棒性，有效应对茎叶遮挡问题；FRM模块通过双向池化和空间-通道协同注意力机制优化多尺度果穗的空间定位精度。此外，引入UIoU损失函数通过动态权重分配机制，优先优化低质量检测框，进一步提升模型的定位精度和收敛稳定性。实验结果表明，CornYOLO在验证集上F_1分数提升2.5个百分点，mAP@50达89.3%，精确率和召回率分别提高4.1和1.2个百分点。这些改进使CornYOLO能够高效、精准地提取田间玉米果穗影像，为表型精准评估和产量预测提供坚实的技术基础，有力推动农业生产的智能化和精准化发展。

尽管CornYOLO在田间玉米果穗检测中表现出色，但仍存在一定局限性。首先，在田间路况复杂或不平整的环境下，无人车采集的图像质量可能下降，导致模型出现误检或漏检，影响检测性能。其次，考虑玉米品种和生长环境的多样性，当前数据集的覆盖范围有限，模型的泛化能力需进一步验证和优化。最后，需要针对复杂田间环境，开发适应性更强的无人车数据采集方案，优化相机稳定性和图像质量，减少环境干扰对检测精度的影响。此外，需要系统性收集涵盖不同玉米品种、地理区域和生长环境的数据，构建更具代表性的数据集，以增强模型的泛化能力和鲁棒性。还需要进一步探索轻量化模块和高效注意力机制，降低计算复杂度，提升模型在边缘设备上的部署能力，以适应实际农业场景的需求。

4 结 论

本研究基于YOLO11n模型提出一种改进的CornYOLO网络，专为田间复杂环境下的玉米果穗检测。通过引入FRM模块增强多尺度目标鉴别能力、C2PDA模块提升遮挡场景下的检测性能，以及UIoU损失函数优化边界框回归精度，CornYOLO显著提升田间果穗检测的效率和准确性。实验结果表明，CornYOLO在验证集上mAP@50达89.3%，F_1分数提高2.5个百分点，精确率和召回率分别提升4.1和1.2个百分点，CornYOLO展现出其在复杂田间环境中优秀的检测性能和泛化能力。本研究提出的CornYOLO模型为田间玉米果穗的精准识别提供一种高效、可靠的解决方案，为玉米育种表型分析和产量预测提供重要技术支撑，显著推动农业智能化和精准化的发展。

利益冲突声明：本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献：

[1] AI Y F, JANE J L. Macronutrients in corn and human nutrition [J]. Comprehensive Reviews in Food Science and Food Safety, 2016, 15(3): 581-598.
[2] AKHTAR M S, ZAFAR Z, NAWAZ R, et al. Unlocking plant secrets: A systematic review of 3D imaging in plant phenotyping techniques[J]. Computers and Electronics in Agriculture, 2024, 222: 109033.
[3] DARRAH L L, MCMULLEN M D, ZUBER M S. Breeding, genetics and seed corn production[M]// Corn: AACC International Press, 2019: 19-41.
[4] MIRBOD O, CHOI D, HEINEMANN P H, et al. On-tree apple fruit size estimation using stereo vision with deep learning-based occlusion handling[J]. Biosystems Engineering, 2023, 226: 27-42.
[5] WANG L L, ZHAO Y J, XIONG Z J, et al. Fast and precise detection of *Litchi* fruits for yield estimation based on the improved YOLOv5 model[J]. Frontiers in Plant Science, 2022, 13: 965425.
[6] SARKAR S, OSORIO LEYTON J M, NOA-YARASCA E, et al. Integrating remote sensing and soil features for enhanced machine learning-based corn yield prediction in the southern US[J]. Sensors, 2025, 25(2): 543.
[7] 余兴娇, 樊凯, 霍雪飞, 等. 基于无人机影像多特征融合的夏玉米LAI动态估计[J]. 农业工程学报, 2025, 41(4): 124-134.
YU X J, FAN K, HUO X F, et al. Dynamic estimation of LAI in summer maize based on multi-feature fusion of UAV images[J]. Transactions of the Chinese Society of Agricultural Engineering, 2025, 41(4): 124-134.
[8] 张晓东, 蔡宗耀, 胡炼, 等. 基于多维成像特征+UGV的设施蔬菜表型参数检测方法[J]. 农业机械学报, 2025, 56(6): 509-517.
ZHANG X D, CAI Z Y, HU L, et al. Detection method of phenotypic parameters of protected vegetables based on multi-dimensional imaging features +UGV[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(6): 509-517.
[9] ZHANG S X, YUE J B, WANG X Y, et al. Segmentation and fractional coverage estimation of soil, illuminated vegetation, and shaded vegetation in corn canopy images using CCSNet and UAV remote sensing[J]. Agriculture, 2025, 15(12): 1309.
[10] YUE J B, YANG G J, LI C C, et al. Estimation of winter wheat above-ground biomass using unmanned aerial vehicle-based snapshot hyperspectral sensor and crop height improved models[J]. Remote Sensing, 2017, 9(7): 708.
[11] YUE J B, WANG J, ZHANG Z Y, et al. Estimating crop leaf area index and chlorophyll content using a deep learning-based hyperspectral analysis method[J]. Computers and Electronics in Agriculture, 2024, 227: 109653.
[12] YUE J B, YANG H, FENG H K, et al. Hyperspectral-to-image transform and CNN transfer learning enhancing soybean LCC estimation[J]. Computers and Electronics in Agriculture, 2023, 211: 108011.
[13] 岳继博, 冷梦蝶, 田庆久, 等. 叶片多理化参数的高光谱遥感与深度学习估算[J]. 光谱学与光谱分析, 2024, 44(10): 2873-2883.
YUE J B, LENG M D, TIAN Q J, et al. Estimation of leaf physical and chemical parameters based on hyperspectral remote sensing and deep learning technologies[J]. Spectroscopy and Spectral Analysis, 2024, 44(10): 2873-2883.
[14] WANG N, FU S W, RAO Q, et al. Insect-YOLO: A new method of crop insect detection[J]. Computers and Electronics in Agriculture, 2025, 232: 110085.
[15] FAN X P, SUN T, CHAI X J, et al. YOLO-WDNet: A lightweight and accurate model for weeds detection in cotton field[J]. Computers and Electronics in Agriculture, 2024, 225: 109317.
[16] 任锐, 孙海霞, 张淑娟, 等. 基于改进YOLOv8n的不同栽培模式下玉露香梨轻量化检测[J]. 农业工程学报, 2025, 41(5): 145-155.

REN R, SUN H X, ZHANG S J, et al. Lightweight detection method for 'Yuluxiang' pear under different cultivation modes based on improved YOLOv8n[J]. Transactions of the Chinese Society of Agricultural Engineering, 2025, 41(5): 145-155.
[17] DANG F Y, CHEN D, LU Y Z, et al. YOLOWeeds: A novel benchmark of YOLO object detectors for multi-class weed detection in cotton production systems[J]. Computers and Electronics in Agriculture, 2023, 205: 107655.
[18] 闫彬, 樊攀, 王美茸, 等. 基于改进YOLOv5m的采摘机器人苹果采摘方式实时识别[J]. 农业机械学报, 2022, 53(9): 28-38, 59.
YAN B, FAN P, WANG M R, et al. Real-time identification of apple picking mode of picking robot based on improved YOLOv5m [J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(9): 28-38, 59.
[19] JING R, NIU Q L, TIAN Y Y, et al. Sunflower-YOLO: Detection of sunflower capitula in UAV remote sensing images[J]. European Journal of Agronomy, 2024, 160: 127332.
[20] ALZADJALI A, ALALI M H, VEERANAMPALAYAM SIVAKUMAR A N, et al. Maize tassel detection from UAV imagery using deep learning[J]. Frontiers in Robotics and AI, 2021, 8: 600410.
[21] QI J T, DING C C, ZHANG R R, et al. UAS-based MT-YOLO model for detecting missed tassels in hybrid maize detasseling[J]. Plant Methods, 2025, 21(1): 21.
[22] FALAHAT S, KARAMI A. Maize tassel detection and counting using a YOLOv5-based model[J]. Multimedia Tools and Applications, 2023, 82(13): 19521-19538.
[23] YADAV P K, THOMASSON J A, HARDIN R, et al. AI-driven computer vision detection of cotton in corn fields using UAS remote sensing data and spot-spray application[J]. Remote Sensing, 2024, 16(15): 2754.
[24] KHAKI S, PHAM H, HAN Y, et al. Convolutional neural networks for image-based corn kernel detection and counting[J]. Sensors, 2020, 20(9): 2721.
[25] SPRAGUE N, EVANS J, MARDIKES M. Corn ear detection and orientation estimation using deep learning[EB/OL]. arXiv: 2412.14954, 2024.
[26] 赵仲文, 张永立, 韩镇宇, 等. 基于改进的SS-YOLOv8轻量化鲜食玉米果穗优劣检测模型[J]. 农业工程学报, 2025, 41(11): 183-192.
ZHAO Z W, ZHANG Y L, HAN Z Y, et al. Improved SS-YOLOv8 lightweight ear detection model for fresh corn[J]. Transactions of the Chinese Society of Agricultural Engineering, 2025, 41 (11): 183-192.
[27] FU J, YUAN H K, ZHAO R Q, et al. Peeling damage recognition method for corn ear harvest using RGB image[J]. Applied Sciences, 2020, 10(10): 3371.
[28] CHEN J, LONG D, YANG S. Research on corn ears defect detection algorithm based on improved YOLOv7[J]. Academic Journal of Engineering and Technology Science, 2024, 7(3): 39-47.
[29] KHANAM R, HUSSAIN M. YOLOv11: An overview of the key architectural enhancements[EB/OL]. arXiv: 2410.17725, 2024.
[30] WEI J F, NI L Y, LUO L, et al. GFS-YOLO11: A maturity detection model for multi-variety tomato[J]. Agronomy, 2024, 14(11): 2644.
[31] LUO X J, CAI Z H, SHAO B, et al. Unified-IoU: For high-quality object detection[EB/OL]. arXiv: 2408.06636, 2024.
[32] REDMON J, FARHADI A. YOLOv3: An incremental improvement[EB/OL]. arXiv: 1804.02767, 2018.
[33] XIANG W T, WU D C, WANG J. Enhancing stem localization in precision agriculture: A two-stage approach combining YOLOv5 with EffiStemNet[J]. Computers and Electronics in Agriculture, 2025, 231: 109914.
[34] XU J S, YANG S Y, LIANG Q, et al. Transillumination imaging for detection of stress cracks in maize kernels using modified YOLOv8 after pruning and knowledge distillation[J]. Computers and Electronics in Agriculture, 2025, 231: 109959.
[35] WANG A, CHEN H, LIU L H, et al. YOLOv10: Real-time end-to-end object detection[EB/OL]. arXiv: 2405.14458, 2024.
[36] LEI M Q, LI S Q, WU Y H, et al. YOLOv13: Real-time object detection with hypergraph-enhanced adaptive visual perception[EB/OL]. arXiv: 2506.17733, 2025.
[37] RESENDE E L, BRUZI A T, SILVA CARDOSO EDA, et al. High-throughput phenotyping: Application in maize breeding[J]. AgriEngineering, 2024, 6(2): 1078-1092.
[38] WARMAN C, SULLIVAN C M, PREECE J, et al. A cost-effective maize ear phenotyping platform enables rapid categorization and quantification of kernels[J]. The Plant Journal, 2021, 106(2): 566-579.

Object Detection Method of Maize Ears Within Canopy Based on CornYOLO

GAO Guangfu[1], WANG Qilei[2], SONG Liwen[1], FENG Haikuan[3], SHI Lei[1], YANG Hao[3], LIU Yang[1], YUE Jibo[1*]

(1. *College of Information and Management Science*, *Henan Agricultural University*, *Zhengzhou* 450002, *China*; 2. *Henan Jinyuan Seed Industry Co. , Ltd. , Zhengzhou* 450002, *China*; 3. *Key Laboratory of Quantitative Remote Sensing in Agriculture*, *Ministry of Agriculture and Rural Affairs/ Information Technology Research Center*, *Beijing Academy of Agriculture and Forestry Sciences*, *Beijing* 100097, *China*)

Abstract: [Objective] As a major grain crop, maize plays a critical role in global food security. The ears of maize serves as a key phenotypic trait, providing essential information on the plant's physiological and agronomic status. Its morphological characteristics, size, and color effectively reflect the plant's growth status and potential yield. Therefore, accurately acquiring images of maize ears in the field across different growth stages is crucial for breeding research and yield prediction. Traditional field detection of maize ears relies heavily on manual labor, which is not only inefficient and labor-intensive but also struggles to meet the high-throughput demands of modern precision breeding programs. There is an urgent need for efficient, automated detection technologies that can operate reliably under real-world field conditions. To address the requirement for efficient acquisition of maize ears phenotypic traits in field breeding work, the objective of this research is to develop a robust object detection solution suitable for large-scale field environments. An improved CornYOLO model based on the YOLO11n (You Only Look Once) architecture was designed to enhance the detection accura-

cy and efficiency of maize ears in complex field environments. **[Methods]** Image data were acquired using an unmanned ground vehicle (UGV) equipped with a high-resolution panoramic camera, which traversed multiple experimental plots under varying lighting and growth conditions. A dataset containing 1 152 annotated samples was constructed, covering diverse ear morphologies and occlusion scenarios. Dynamic data augmentation techniques were applied during training to enhance the model's generalization capability. Three key enhancements were introduced to the YOLO11n detection framework. First, a cross stage partial network with dynamic pointwise spatial attention (C2PDA) module was designed to replace the cross stage partial with pointwise spatial attention (C2PSA) module in the YOLO11 backbone network. This module enhanced spatial discriminability and channel sensitivity in feature representation through the collaborative integration of a dynamic channel weighting mechanism and position-aware modeling. It significantly improves the model's performance in identifying maize ears under challenging field conditions such as occlusion of stems and leaves and multi-scale target distribution. Second, the spatial pyramid pooling-fast (SPPF) module in the original model was replaced with an feature refinement module (FRM) to optimize multi-scale feature fusion. The FRM functions via directional feature decomposition and an adaptive attention mechanism. It captures fine-grained spatial structural information through horizontal and vertical bidirectional pooling and combines spatial-channel cooperative attention for dynamic feature calibration, thereby improving recognition accuracy across varying ear sizes and complex backgrounds. Finally, the unified intersection over union (UIoU) loss function was introduced to optimize bounding box regression accuracy. UIoU is an innovative loss function that emphasizes weight allocation among prediction boxes of different qualities. It adaptively adjusted the weight of each prediction box's loss term based on the IoU value or its monotonic function, assigning higher weights to lower-quality predictions to prioritize their optimization, while reducing weights for high-quality boxes to prevent over-optimization. **[Results and Discussions]** Experimental results demonstrate that CornYOLO achieved a mAP@50 of 89.3% on the validation set, with the F_1-Score increasing by 2.5 percentage points. Compared to widely used lightweight models including YOLOv3-tiny, YOLOv5n, YOLOv8n, YOLOv10n, real-time detection transformer (RT-DETR) and YOLO13n, CornYOLO showed significantly superior detection performance in complex field environments, with mAP@50 improvements of 2.2, 1.9, 1.8, 5.7, 12.6 and 2.4 percentage points, respectively. These results fully validate that CornYOLO can efficiently and accurately extract maize ear images under field conditions, providing a technical foundation for precise phenotypic evaluation and yield prediction. Furthermore, ablation studies were conducted: Introducing the C2PDA module improved the model's mAP@50 by 0.5 percentage points and the F_1-Score by 0.5 percentage points. However, after incorporating the FRM module, which successfully enhanced multi-scale detection performance and increased the F_1-Score by 1.5 percentage points, the integration of these two modules resulted in the generation of a small number of low-quality detection boxes. The original loss function was inefficient in optimizing such boxes, resulting in no improvement in mAP@50 after the modification. To address this issue, the UIoU loss function was introduced. By dynamically adjusting weight assignments based on prediction quality, it significantly improved the regression performance for low-quality detection boxes, thereby enhancing the localization accuracy and convergence stability of the model in dense target scenarios. The final CornYOLO model exhibited excellent overall performance: Compared to the original YOLO11n, the F_1-Score increased by 2.5 percentage points and mAP@50 improved by 1.1 percentage points. The experimental results fully demonstrate that CornYOLO effectively enhances the detection capability for maize ears in complex field environments compared to the baseline YOLO11n model. **[Conclusions]** The CornYOLO model proposed in this study incorporates three key components: C2PDA, FRM, and UIoU, which enhances model convergence and localization performance in dense and occluded scenes, enables the model to effectively and precisely identify maize ears under practical conditions, thereby providing reliable technical support for phenotypic analysis and yield prediction in maize breeding. Future work will focus on extending the model to other crop types and further optimizing inference efficiency for real-time deployment on mobile platforms.

Key words: maize ears; unmanned ground vehicle; YOLO11; panoramic camera; object detection

Foundation items: Natural Science Foundation of Henan (252300421839); National Natural Science Foundation (42101362, 42571462)

First author: GAO Guangfu, E-mail: gaoguangfu@stu.henau.edu.cn

***Corresponding author:** YUE Jibo, E-mail: yuejibo@henau.edu.cn