

基于 DD-MA-YOLOv11 的设施番茄花穗检测与 花朵状态识别方法

高雅楠^{1,2,3}, 柳平增^{1,2,3*}, 张宇轩^{1,2,3}, 朱珂^{1,2,3},
张艳^{1,2,3}, 于群^{1,2,3}, 温孚江¹

(1. 山东农业大学信息科学与工程学院, 泰安 271018; 2. 农业农村部黄淮海智慧农业技术重点实验室, 泰安 271018; 3. 山东农业大学农业大数据研究中心, 泰安 271018)

摘要: 针对设施番茄花果期管理中花穗与花朵目标尺度小、背景复杂、遮挡严重, 且单阶段检测方法难以同时兼顾花穗精准定位与花朵状态精细识别的问题, 该研究提出了一种基于 DD-MA-YOLOv11 的级联识别网络。该网络以整株番茄图像作为输入, 第一级通过引入 DLKA 注意力机制与 Dynamic Head 动态检测头增强 YOLOv11 的小目标检测能力, 实现花穗区域的准确定位; 第二级针对复杂背景下花朵状态差异细微、易受遮挡干扰的识别任务, 以 MobileNetV4 作为主干网络, 并结合 ATFL 焦点损失函数, 实现花朵状态的高精度识别。试验结果表明, 第一级 DD-YOLOv11 模型在花穗检测任务中, 精确率、召回率、mAP@0.5 及 F1 分数分别为 95.27%、90.35%、94.19%、92.73%, 较原模型提升 4.02、5.25、8.49 和 4.66 个百分点。第二级 MA-YOLOv11 模型在花朵状态识别任务中, 精确率、召回率、mAP@0.5 及 F1 分数分别较原模型提升了 5.24、2.97、5.31 和 4.07 个百分点, 整体级联模型在单输入条件下的处理效率为 38.4 帧/s, 满足实时处理要求。该模型实现了花穗检测与花朵状态识别, 可为番茄的授粉、摘心等农事操作提供依据。

关键词: 设施番茄; 花穗检测; 花朵状态识别; 级联目标检测; YOLOv11

doi: 10.11975/j.issn.1002-6819.202509254

中图分类号: S126

文献标志码: A

文章编号: 1002-6819(2026)-12-0227-12

高雅楠, 柳平增, 张宇轩, 等. 基于 DD-MA-YOLOv11 的设施番茄花穗检测与花朵状态识别方法[J]. 农业工程学报, 2026, 42(12): 227-238. doi: 10.11975/j.issn.1002-6819.202509254 <http://www.tcsae.org>

GAO Yanan, LIU Pingzeng, ZHANG Yuxuan, et al. Method for detecting tomato flower clusters and recognizing flower status based on DD-MA-YOLOv11[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2026, 42(12): 227-238. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202509254 <http://www.tcsae.org>

0 引言

番茄是世界上种植面积第二大蔬菜, 也是中国种植的主要蔬菜种类之一^[1], 其产量和出口量均位居世界前列^[2]。番茄富含维生素、番茄红素、类胡萝卜素、类黄酮等多种营养成分, 具有极高的营养价值和保健功效^[3-4]。在番茄生产过程中, 不同生育时期的栽培管理要求各不相同。其中, 花果期是番茄产量与品质形成的关键阶段^[5], 在此阶段通过授粉、摘心等措施可提高授粉质量、调控植株营养生长, 从而促进果实健全发育与均匀成熟, 提升产量和品质^[6-9]。然而, 在实际生产中, 花穗数量、花朵开放状态及其空间分布等关键表型信息具有明显的时序性和空间差异性, 目前主要依赖人工经验判断, 难以实现连续、客观的精准获取, 制约了设施番茄生产向高效化、智能化发展的需求。因此, 实现对设施番茄花穗及花朵发育状态的快速、准确识别, 是支撑授粉、摘

心等关键农事操作自动化与智能决策的重要前提。

近年来, 基于深度学习的农业视觉感知技术在作物目标检测领域得到广泛应用, 相关研究主要集中在果实检测与成熟度识别任务中, 通过引入多尺度特征融合、注意力机制及网络结构优化等方法提升复杂环境下的检测性能^[10-12]。此外, 在植株结构表型提取方面, 部分研究结合图像处理与深度学习方法对作物结构信息进行分析, 但主要集中于茎叶结构及果实相关特征建模^[13-15]。相较而言, 针对番茄花穗及花朵发育状态的研究仍相对不足。在花朵相关研究方面, 现有工作主要集中于单花检测及花期分类任务。如余贤海等^[16]提出 TFDC-Net 算法, 基于改进 YOLOv5s 模型实现番茄花朵检测, 并结合分类网络完成花期与姿态识别。赵春江等^[17]基于改进 YOLOv3 与多尺度特征融合结构实现番茄花期识别。然而, 在实际设施生产环境中, 番茄花穗通常呈簇状分布, 花朵之间存在密集排列与遮挡现象, 同时不同发育阶段花朵在颜色、纹理及形态特征上差异较小, 使得模型在复杂场景下容易出现定位偏差与类别混淆问题。

作物视觉感知方法通常可分为基于二维图像与基于三维信息的方法两类。基于三维信息的方法依赖 RGB-D 或多视角重建技术, 可获取较为完整的空间结构信息, 并已应用于花朵检测与授粉目标定位等任务中^[18-19]。然而, 该类方法对硬件条件与数据处理流程要求较高, 难

收稿日期: 2025-09-27 修订日期: 2026-05-10

基金项目: 山东省重点研发计划项目(2022TZXD0030)、中央引导地方科技发展专项资金项目(YDZX2022073)

作者简介: 高雅楠, 研究方向为农业大数据技术与工程。

Email: gaoyan2022@163.com

*通信作者: 柳平增, 博士, 博士生导师, 教授, 研究方向为农业物联网与大数据、智慧农业等。Email: pzliu@sdau.edu.cn

以在设施农业场景中实现长期稳定应用。相比之下,基于二维图像的方法具有成本低、部署灵活等优势,成为当前农业视觉感知的主流方向。在二维图像检测方法中,卷积神经网络(CNN)已成为核心技术框架,其中目标检测方法主要包括两阶段与单阶段两类。两阶段方法(如Faster R-CNN、Mask R-CNN)在精度方面表现较优,但推理效率较低^[20-21];单阶段方法(如SSD及YOLO系列)在保证较高精度的同时具有更好的实时性,因此在农业场景中应用更为广泛^[22]。相关研究表明,YOLO系列模型在复杂场景目标检测任务中具有较好的适应性,例如尚钰莹等^[23]基于YOLOv5实现苹果花朵检测,吕志远等^[24]结合ViT与YOLO构建级联检测框架以提升复杂环境下检测性能。HAO等^[25]基于YOLOv11构建番茄成熟度检测模型,验证了其在复杂场景下的有效性。

尽管上述研究在各自任务中取得了一定进展,但现有基于YOLO的改进方法大多针对单一目标检测任务设计,其网络结构通常更适用于目标尺度变化相对有限的场景。当花穗小目标检测与花朵状态细粒度识别同时存在于同一整株图像时,单一检测框架往往难以同时兼顾全局感受野建模与局部细粒度特征表达,容易导致花穗定位不稳定或花朵状态判别精度受限。

针对上述问题,本文面向设施番茄花果期农事管理需求,将设施番茄花果期感知任务划分为花穗检测与花朵状态识别两个子任务,提出一种“先空间定位、后精细识别”的两阶段级联感知策略,第一阶段输出花穗空间区域,第二阶段在该区域内对花朵进行状态识别,从而有效缓解整株图像中尺度差异大、背景干扰强对识别性能的影响。具体而言,本文以YOLOv11为基础构建两阶段级联识别模型,第一阶段引入DLKA注意力机制与Dynamic Head动态检测头,实现番茄花穗的小目标检测,为后续精细识别提供稳定的花穗区域;第二阶段引入MobileNetV4主干网络与ATFL焦点损失函数,在花穗区域内,针对复杂遮挡且形态差异细微等特点,开展花朵状态识别。通过两阶段级联策略,在保证整体检测效率的同时,提升设施环境下花穗定位的稳定性和花朵状态识别的准确性,以期为设施番茄授粉、摘心等自动化农事操作提供可靠的视觉感知基础。

1 材料与方法

1.1 数据获取与预处理

1.1.1 数据采集

本研究使用的番茄图像数据集采集于山东省菏泽市巨野县某农业科技有限公司园区4号温室(115.87°E, 35.34°N),温室东西长度120 m,南北跨度12 m,试验番茄品种为双飞8号。数据采集期间,在温室顶部和侧墙分别布设一台RGB摄像机,支持侧视、俯视多视角采集,系统设定每日8:00至17:00定时定点自动采集图像数据并实时上传至云服务器。为进一步丰富数据集,采用手机辅助采集方式,以获取在不同光照条件、拍摄角度及背景环境下的图像。

1.1.2 数据预处理与数据集构建

对采集的原始图像进行质量筛选,剔除模糊及质量不佳样本,最终获得番茄花穗图像2400张。采用Labellmg工具对花穗目标进行标注。按照8:1:1比例将数据划分为训练集、验证集与测试集。在此基础上,对数据进

行离线数据增强,以提升模型泛化能力。增强方法包括随机高斯噪声、随机亮度调整及随机水平翻转等。经过增强后,训练集样本扩展至3600张,数据集划分结果如表1所示。

表1 花穗数据集划分

Table 1 Division of the tomato flower cluster dataset

数据集 Dataset	番茄花穗图像数量 Number of tomato flower cluster images
训练集 Training set	2 880
验证集 Validation set	360
测试集 Testing set	360

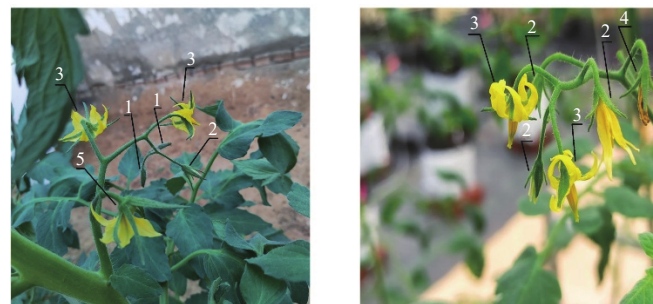
第二阶段花朵状态识别数据集基于第一阶段花穗检测模型的输出结果构建。利用已训练完成的花穗检测模型对原始图像进行处理,利用训练完成的花穗检测模型对原始图像进行处理,根据模型输出的花穗检测框裁剪获得对应的花穗区域图像,以减少人工精确裁剪引入的理想化偏差,提升模型在实际应用中的鲁棒性。

经文献查阅、专家咨询等,将番茄花朵状态分为花蕾期、花裂期、全开期、谢花期、初果期五种类别,通过识别花裂期和全开期的花朵判断授粉需求,具体花朵状态分类如表2所示,花朵状态各类别示意图如图1所示。

表2 花朵状态分类

Table 2 Classification of tomato flower states

花朵状态 Flower state	分类依据 Classification criteria
花蕾期 Flower bud stage	花朵尚未开放,花瓣完全包裹在萼片内,形态呈椭圆或球形,颜色偏浅绿或黄绿色
花裂期 Cracked bud stage	花瓣开始从萼片中伸出,花朵部分开放,花瓣基部可见,颜色由浅黄逐渐加深
全开期 Full bloom stage	花瓣完全展开,花形平展,雌蕊和雄蕊清晰可见,是授粉和受精的最佳时期
谢花期 Withered flower stage	花瓣开始萎蔫、卷曲或脱落,颜色逐渐褪去,雄蕊花粉减少或消失
初果期 Initial fruit stage	花瓣完全脱落,花萼包裹着幼果,果实直径较小且呈绿色



1. 花蕾期 2. 花裂期 3. 全开期 4. 谢花期 5. 初果期
1. Flower bud stage 2. Cracked bud stage 3. Full bloom stage 4. Withered flower stage 5. Initial fruit stage

图1 花朵状态各类别示意图

Fig.1 Schematic diagram of different flower state categories

在获得花穗区域图像后,使用Labellmg标注工具对花朵状态数据集进行人工标注,按照8:1:1的比例将番茄花朵状态图像划分为训练集、验证集和测试集。为增强数据集多样性,提高模型对不同光照、姿态及遮挡条件下花朵的识别能力,在数据集构建阶段采用随机旋转、镜像翻转、尺度缩放、光照扰动、Mixup及Cutout等离线数据增强方法对原始样本进行扩充。经数据增强后,共获得4800张花朵状态图像。以番茄花朵作为目标对象,共标注23258个标签,各标签分布如表3所示。

表 3 花朵状态各状态标签分布
Table 3 Distribution of labels for different flower states

花朵状态 Flower state	花朵状态各状态标签数量 Number of samples per flower state category
花蕾期 Flower bud stage	4 823
花裂期 Cracked bud stage	5 176
全开期 Full bloom stage	6 394
谢花期 Withered flower stage	3 274
初果期 Initial fruit stage	3 591

1.2 YOLOv11 模型分析

番茄花穗检测与花朵状态识别是设施番茄智能管理的重要基础。其中，花穗目标在整株图像中占比较小，且常受到叶片遮挡和复杂背景干扰；不同花朵发育阶段之间存在一定相似性，对模型的目标检测能力和特征表达能力提出了较高要求。同时，为满足设施农业智能装备实时作业需求，模型还需兼顾检测精度、推理速度和部署效率。

YOLO 系列算法作为典型的单阶段目标检测算法，具有检测速度快、部署方便和精度较高等优点，已广泛应用于农业目标检测领域。YOLOv11^[26] 在 YOLO 系列模型基础上进一步优化了特征提取与特征融合结构，能够在保证实时性的同时获得较好的检测性能。因此，本文选择 YOLOv11 作为两阶段级联检测框架的基础模型。

1.3 目标检测网络构建

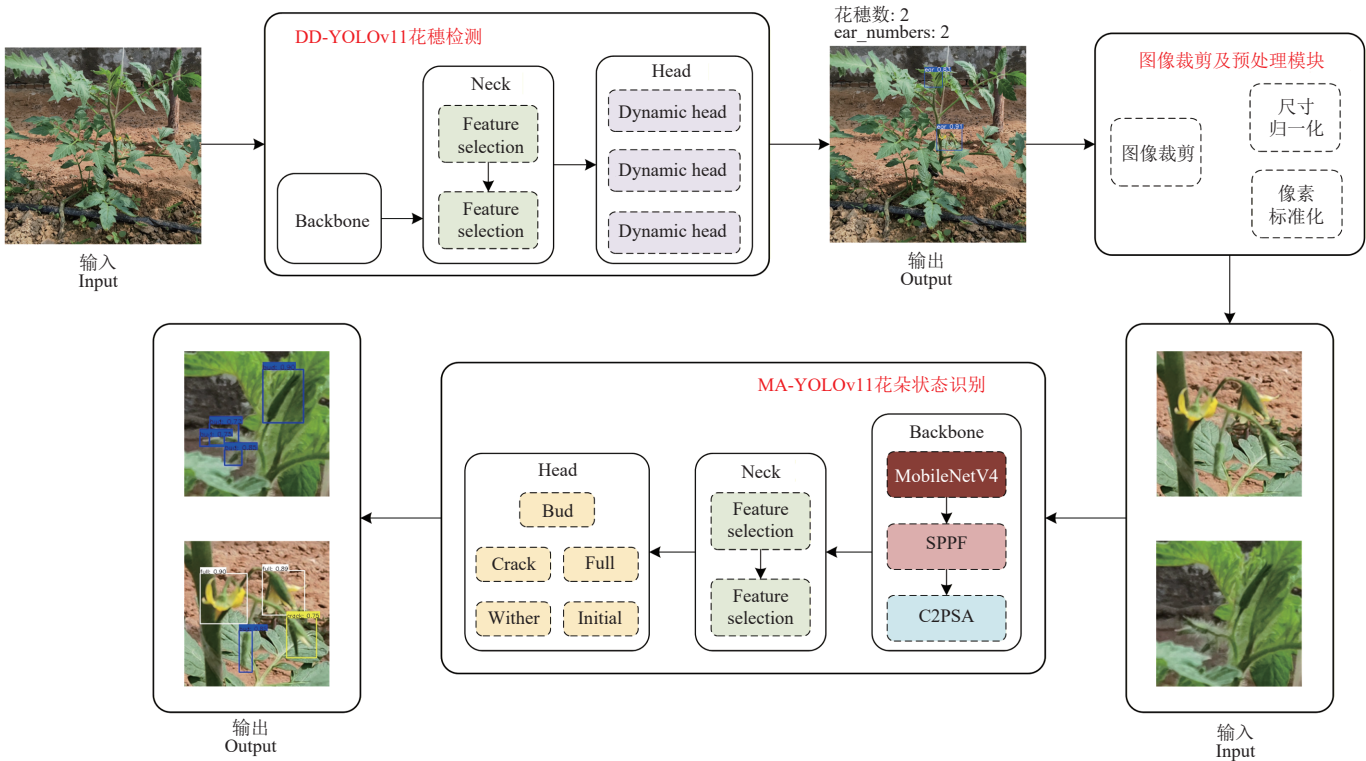
1.3.1 模型整体结构

在设施番茄花果期管理中，授粉与摘心等操作依赖花穗和花朵状态的准确识别。然而，花穗体积小且密集，花朵易被叶片遮挡，且不同发育阶段的形态差异细微，

同时温室环境存在复杂背景干扰，增加了花穗目标检测和花朵状态识别的难度，难以满足实际农事操作需求。

为解决上述问题，本研究基于 YOLOv11 的高精度与实时性能，提出了级联识别模型 DD-MA-YOLOv11。该模型通过级联两个改进的 YOLOv11 模型，分阶段完成花穗空间定位与花朵状态精细识别。第一阶段构建 DD-YOLOv11 花穗检测模型，在 YOLOv11 中引入 DLKA 注意力机制与 Dynamic Head 动态检测头。针对整株图像中花穗目标尺度较小、分布密集及存在叶片遮挡等问题，DLKA 通过大卷积核与可变形卷积增强对花穗及其邻近结构的特征建模能力，Dynamic Head 利用尺度与空间感知机制，提高模型对不同尺度花穗及复杂遮挡区域的检测稳定性。第二阶段构建 MA-YOLOv11 花朵状态识别模型，以第一阶段输出的花穗区域为输入，在局部区域内开展花朵状态识别。针对花朵状态间差异细微、局部遮挡以及类别分布不均衡等问题，模型采用轻量化主干网络 MobileNetV4，以降低计算复杂度并提升推理效率；同时引入自适应阈值焦点损失（adaptive threshold focal loss, ATFL）函数，根据花朵状态类别的样本分布自适应分配权重，增强对样本数量较少及易混淆状态的识别能力。

整体流程为首先利用 DD-YOLOv11 对整株番茄图像进行花穗检测，根据检测结果裁剪并预处理花穗区域，随后将其输入 MA-YOLOv11 进行花朵状态识别，最终输出识别结果，整体流程如图 2 所示。



注：Feature selection 为特征选择；Dynamic head 为动态检测头模块；SPPF 为快速空间金字塔池化模块；C2PSA 为跨阶段注意力模块；MobileNetV4 为轻量化移动端网络；Bud 为花蕾期；Crack 为花裂期；Full 为全开期；Wither 为谢花期；Initial 为初果期。下同。
Note: Feature selection represents feature selection; Dynamic head represents dynamic detection head module; SPPF represents spatial pyramid pooling fast module; C2PSA represents cross-stage attention module; MobileNetV4 represents lightweight mobile network; Bud represents flower bud stage; Crack represents cracked bud stage; Full represents full bloom stage; Wither represents withered flower stage; Initial represents initial fruit stage. The same below.

图 2 整体流程图
Fig.2 Overall flow diagram

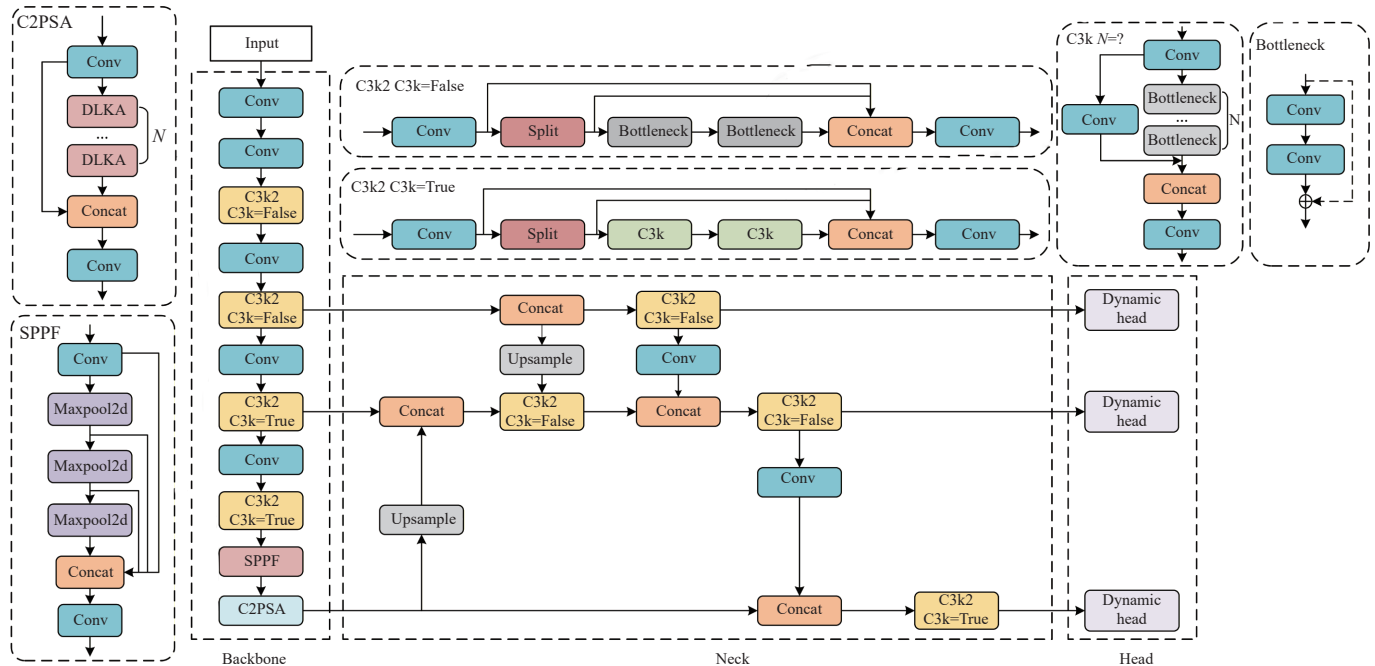
1.3.2 DD-YOLOv11 花穗检测模型

在番茄生长过程中,花穗的检测对于摘心操作具有重要指导意义。然而,花穗在番茄整株图像中通常尺寸较小、像素占比低且特征有限,属于典型的小目标检测问题,识别难度较大。针对这一问题,本研究提出了 DD-YOLOv11 花穗检测模型。该模型在 YOLOv11 主干网络的 C2PSA 模块中引入 DLKA 注意力机制,以增强模型对小目标的关注能力,并结合 Dynamic head 提升检

测头的特征表征与自适应能力,在保持计算效率的基础上显著提高花穗的检测精度。DD-YOLOv11 花穗检测模型的整体结构如图 3 所示。

1) DLKA 注意力机制

可变形大卷积核注意力机制 DLKA^[27] 是一种简化注意力机制,通过大卷积核提取上下文关系,并借助可变形卷积使采样网格自适应变形,从而提升模型对多维数据模式的适应能力。DLKA 模块架构如图 4 所示。



注: Conv 为卷积模块; C3k2 为复合卷积模块; Concat 为特征连接模块; DLKA 为可变形大卷积核; Upsample 为上采样模块; Split 为特征通道分离操作; Bottleneck 为基础残差单元; C3k 为复合卷积模块; MaxPool2d 为二维最大池化模块。N 表示对应子模块的堆叠重复次数。下同。
Note: Conv is a convolution module; C3k2 is a composite convolution module; Concat is a feature concatenation module; DLKA is a deformable large-kernel convolution module; Upsample is an upsampling module; Split is a feature channel splitting operation; Bottleneck is a basic residual unit; C3k is a composite convolution module; MaxPool2d is a two-dimensional max-pooling module. N represents the stacking repetition times of the corresponding submodules. The same below.

图 3 DD-YOLOv11 花穗检测模型整体结构

Fig.3 Overall architecture of the DD-YOLOv11 model for tomato flower cluster detection

大卷积核具备捕获上下文信息的能力,可在一定程度上模拟自注意力机制的大感受野行为。该模块通过结合深度可分离卷积 (DW Conv) 与深度可分离膨胀卷积 (DW-D Conv), 构建大卷积核结构, 在扩大感受野的同时有效控制参数规模与计算复杂度, 提升模型运行效率。在此基础上, DLKA 机制引入可变形卷积^[28], 该机制依据目标尺度自适应调整采样位置。针对番茄花穗尺寸差异明显、结构形态多样的特点, 可变形卷积能够根据目标特征动态调整采样位置, 更精准地捕获花穗关键区域信息, 从而增强模型对不同尺度花穗的表征能力, 并有效抑制复杂背景带来的干扰。

本研究在 YOLOv11 主干网络中的 C2PSA 模块中引入 DLKA 模块, 替代原有的注意力机制。通过在深层特征层引入 DLKA, 网络能够在较大感受野范围内建模花穗目标与周围植株结构之间的上下文关系, 从而增强深层语义特征中与花穗相关的判别性表达。经 DLKA 增强后的深层语义特征通过 FPN 结构向浅层逐级传递, 并与高分辨率特征进行多尺度融合, 使检测头在保持空间定位精度的同时获得更具判别性的语义约束, 提升复杂设施环境下花穗小目标在背景干扰条件下的检测稳定性与

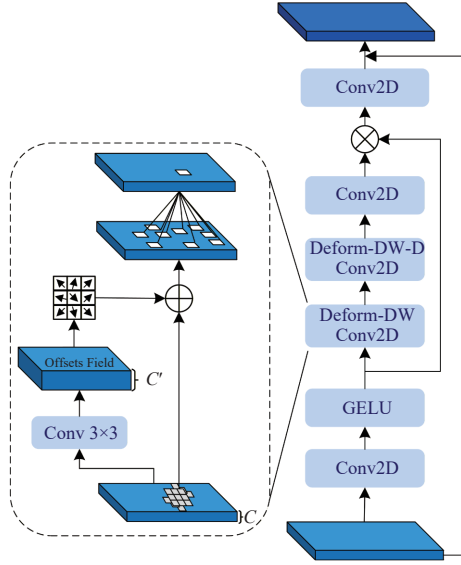
鲁棒性。

2) Dynamic Head

Dynamic Head 是一种基于动态特征聚合的新型检测头架构^[29], 通过集成空间尺度、空间位置信息以及任务感知机制, 利用多头自注意力增强模型的特征表达能力, 从而动态聚焦于关键区域与细微特征。适用于检测复杂背景中的小目标或被遮挡对象, 能够有效提升检测的准确性与鲁棒性。

Dynamic head 网络架构如图 5 所示。尺度感知注意力是 Dynamic head 的重要组成部分。在花穗检测任务中, 不同大小的花穗在图像中具有不同的语义重要性。该机制能够依据不同尺度的语义重要性, 在特征层次自适应地融合多尺度信息, 增强模型对花穗尺度变化的感知能力, 进而提升检测头在处理多尺度目标的性能。空间感知注意力主要作用于空间维度 (高度×广度), 专注于识别具有判别性的空间区域。首先利用可变形卷积对注意力学习过程进行稀疏化处理, 使卷积采样点能够根据花穗的具体形态和位置自适应调整, 精准提取花穗特征。随后, 通过对同一空间位置跨不同特征层的信息进行自适应聚合, 融合多级别特征, 使特征响应更集中于花穗

的关键空间区域。该方法有助于在复杂的番茄植株背景下准确区分花穗与茎叶等干扰物，从而提高检测精度。

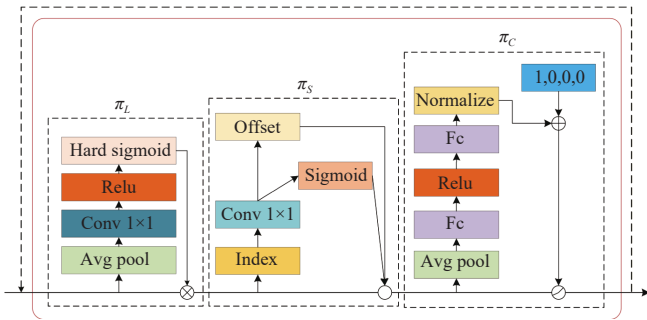


注：Conv2D 为标准二维卷积模块；Conv3×3 为 3×3 尺寸卷积模块；Offsets Field 为偏移场特征图；Deform-DW Conv2D 为可变形深度卷积模块；Deform-DW-D Conv2D 为可变形空洞深度卷积模块；GELU 为高斯误差线性激活函数；⊕ 为逐元素相加融合操作；⊗ 为逐元素相乘加权操作；C 为输入特征通道数；C' 为偏移场输出通道数。下同。

Note: Conv2D is a standard two-dimensional convolution module; Conv 3×3 is a 3×3 size convolution module; Offsets Field is the offset field feature map; Deform-DW Conv2D is a deformable depthwise convolution module; Deform-DW-D Conv2D is a deformable dilated depthwise convolution module; GELU is the Gaussian error linear activation function; ⊕ denotes element-wise addition fusion operation; ⊗ denotes element-wise multiplication weighting operation; C represents the number of input feature channels; C' represents the number of output channels of the offset field. The same below.

图 4 DLKA 注意力机制模块架构

Fig.4 Architecture of the DLKA attention mechanism module



注： π_L 为尺度感知注意力子分支； π_S 为空间感知注意力子分支； π_C 为任务感知注意力子分支。Avg pool 为平均池化；Index 为索引特征；Offset 为特征偏移量；fc 为全连接层；normalize 为归一化操作。

Note: π_L is a scale-aware attention sub-branch; π_S is a spatial-aware attention sub-branch; π_C is a task-aware attention sub-branch. Avg pool denotes average pooling; Index denotes index features; Offset denotes feature offset; fc denotes fully connected layer; normalize denotes normalization operation.

图 5 Dynamic head 网络架构

Fig.5 Architecture of the Dynamic head

在本研究中，将 Dynamic head 结构引入 YOLOv11 的检测头中，将多维注意力机制集成于原有架构。多种注意力机制在不同维度协同工作、互为补充，增强了检测头的特征表征能力，提升了模型对番茄花穗的检测性能。

1.3.3 图像裁剪及预处理模块

为实现两阶段检测网络的有效衔接，本研究在花穗检测模块与花朵状态识别模块之间设计了图像裁剪及预

处理模块。第一阶段 DD-YOLOv11 模型以 960×960 像素图像作为输入进行花穗检测。该模块的主要作用是将第一阶段花穗检测模型输出的检测框转换为适用于第二阶段模型输入的候选区域图像。

具体而言，根据第一阶段模型输出的花穗检测框坐标，从原始图像中裁剪出对应的花穗区域。为满足第二阶段 MA-YOLOv11 模型的输入要求，首先采用等比例缩放与边缘填充相结合的方式，对裁剪后的花穗区域图像进行尺寸归一化处理，统一调整为 640×640 像素；随后对图像进行像素值标准化处理，将 RGB 图像像素值由 [0,255] 线性映射至 [0,1] 区间，以减小不同图像间亮度差异对模型训练和推理的影响。

经预处理后的花穗区域图像被输入至第二阶段花朵状态识别模型，以实现花朵状态识别。通过该模块的设置，第一阶段的花穗检测结果有效传递至第二阶段花朵状态识别模型，保证了两阶段方法的整体闭环性和特征表达的完整性，为后续的花朵状态识别提供了可靠的数据基础。

1.3.4 MA-YOLOv11 花朵状态识别模型

设施环境中，番茄花朵易受叶片及其他花朵遮挡，导致关键特征信息难以完整获取。同时，生产管理中花朵状态的实时监测具有重要需求，尤其是对花裂期、全开期等适宜授粉的关键阶段的及时识别，可有效把握最佳授粉时机，提升授粉成功率与果实品质。为解决上述问题，本研究提出了 MA-YOLOv11 花朵状态识别模型，该模型在 YOLOv11 的基础上采用轻量化主干网络 MobileNetV4，以降低计算复杂度并提升推理效率；同时引入自适应阈值焦点损失（adaptive threshold focal loss, ATFL）函数，以改善类别不平衡条件下的识别效果，提升模型在复杂遮挡环境下的精度与鲁棒性。MA-YOLOv11 花朵状态识别模型整体结构如图 6 所示。

1) MobileNetV4

MobileNetV4 引入通用反向瓶颈（universal inverted bottleneck, UIB）搜索块和 Mobile MQA 注意力模块创新设计，UIB 模块结构示意图如图 7 所示。

UIB 结构在倒瓶颈块中引入了两个可选的深度卷积，旨在提升特征提取的灵活性和效率^[30]。其中，第一个深度卷积位于扩展层之前，负责对输入特征进行初步处理，从而更好地提取低层次特征，即便花朵存在部分遮挡，仍能从输入特征中提取出花朵的边缘、纹理等低层次特征。第二个深度卷积位于扩展层和投影层之间，主要用于增强模型对深层次特征的提取能力。深层次特征能够融合更多上下文信息，在花朵被遮挡的情况下，可通过周围未遮挡区域的信息及其语义关联，推断出花朵的整体状态。MobileNetV4 在网络设计中兼顾模型性能与计算开销，通过优化网络结构减少参数数量和计算复杂度；并通过结合粗粒度和细粒度的神经架构搜索（neural architecture search, NAS）策略，对网络结构进行自动化搜索与优化。

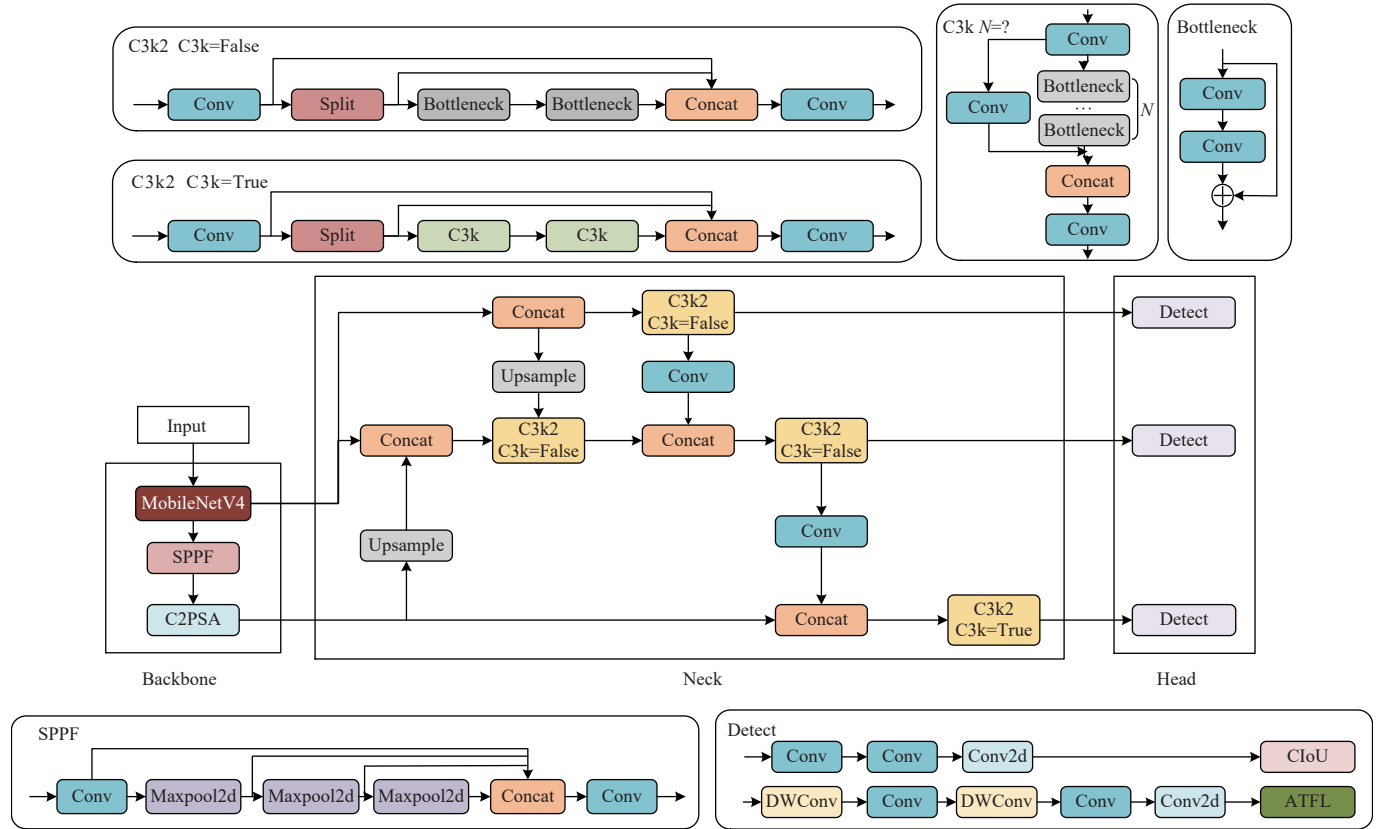
本研究采用 MobileNetV4 替换 YOLOv11 的主干网络，旨在提升推理速度，增强在遮挡和复杂背景下的特征提取能力，从而能够有效提升模型的鲁棒性与实时检测性能，为花朵状态识别任务提供了可靠支持。

2) ATFL 焦点损失函数

自适应阈值焦点损失（adaptive threshold focal loss, ATFL）是一种动态调整损失权重的损失函数^[31]，旨在缓

解类别分布差异带来的影响, 并提高模型对难分类样本的关注, 从而提高检测性能。与阈值焦点损失 (threshold focal loss, TFL) 采用固定阈值和调制参数不同, ATFL 通过引入自适应权重分配和动态阈值调节机制, 实现损失权重的动态调整。由表 3 可知, 花朵状态

数据集中不同类别标签数量存在一定差异, 同时花蕾期、花裂期和全开期等相邻发育阶段在形态特征上具有较高相似性, 容易产生分类混淆。因此, 本文引入 ATFL 损失函数, 以增强模型对少样本类别和困难样本的学习能力, 提高花朵状态识别性能。



注: Detect 是检测头模块; DWConv 为深度可分离卷积模块; Conv2d 为二维标准卷积模块; CIoU 为完整 IoU 损失函数; ATFL 为自适应焦点损失函数。
Note: Detect is a detection head module; DWConv is a depthwise separable convolution module; Conv2d is a two-dimensional standard convolution module; CIoU is the complete IoU loss function; ATFL is the adaptive focal loss function.

图 6 MA-YOLOv11 花朵状态识别模型整体结构

Fig.6 Overall architecture of the MA-YOLOv11 flower state recognition model

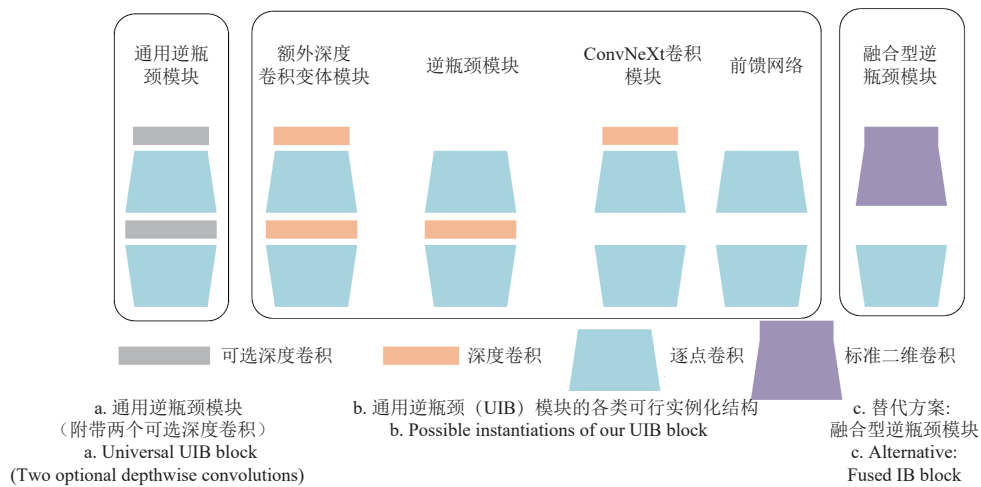


图 7 通用逆瓶颈 (UIB) 模块结构示意图

Fig.7 Structure diagram of the universal inverted bottleneck (UIB) block

基于原始 Focal Loss, ATFL 通过引入自适应权重分配和动态阈值调节机制, 根据样本预测结果动态调整损失权重, 提高模型对少样本类别和困难样本的学习能力。

ATFL 通过自适应阈值机制为各样本计算最优权重, 使损失函数更精确地反映样本重要性; 同时利用聚焦机制抑制易分类样本的损失贡献, 加强对难分类样本的关

注。通过优化损失计算过程，加快了关键特征的学习速度，进一步提高了训练效率与实时检测性能。

阈值焦点损失 (TFL) 通过降低易分类样本的损失权重，同时增加难分类样本的损失权重，减轻易分类样本对训练过程的影响。TFL 计算如下：

$$\text{TFL} = \begin{cases} -(\lambda - P_i)^\eta \log(P_i), & P_i \leq 0.5 \\ -(1 - P_i)^\gamma \log(P_i), & P_i > 0.5 \end{cases} \quad (1)$$

式中 P_i 为模型对目标类别的预测概率； γ 为高置信易分类样本的自适应调制系数； η 和 λ ($\lambda > 1$) 为超参数。针对不同数据集与模型结构，通过多次试验调整超参数以取得最佳性能。

为提升自适应性，对 η 和 γ 进行了修改。通常，高预测概率对应较低的信息增益，而低概率则意味着较高的信息增益。因此，自适应调制因子 γ 定义为

$$\gamma = -\ln(\hat{p}_c) \quad (2)$$

式中 $\hat{p}_c$ 为类别自适应预测概率。

在网络训练后期，过高的期望概率会降低难分类样本的比例，导致模型对困难样本的学习不足。因此， η 定义为

$$\eta = -\ln(P_i) \quad (3)$$

本文在 YOLOv11 中引入 ATFL 替代传统交叉熵损失，通过动态调整不同样本的损失权重，旨在提高模型对少样本类别和困难样本的学习能力，增强对易混淆花朵状态的判别能力，从而提升花朵状态识别性能。

1.4 试验环境及评价指标

1.4.1 试验环境与参数

试验基于 Pytorch1.13.0+cu117 深度学习框架开展，试验运行环境为 Windows 11 操作系统，硬件平台为 Intel Core i7-13700F CPU 与 GeForce RTX 4070 GPU，开发语言为 Python 3.8。

在模型训练过程中，结合前期研究基础与多次试验调优，以平衡训练效率与模型性能。最终训练策略采用 SGD 优化器，初始学习率设置为 0.001，动量系数为 0.937，权重衰减系数为 0.005，训练轮数 (epochs) 设置为 300，批量大小 (batch size) 为 16。

在不同任务的输入设置方面，花穗检测任务的输入图像尺寸统一调整为 960×960 像素，花朵状态识别任务的输入图像尺寸设置为 640×640 像素。

1.4.2 评价指标

为全面评估所提出的两阶段模型的检测与识别性能，选取了准确率 (precision, P)、召回率 (recall, R)、平均精度均值 (mean average precision, mAP)、推理速度 (frames per second, FPS) 以及 F1 分数 (F1-score, F1) 作为评价指标。各指标计算式如下：

$$P = \frac{T_p}{T_p + F_p} \times 100\% \quad (4)$$

$$R = \frac{T_p}{T_p + F_n} \times 100\% \quad (5)$$

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i \quad (6)$$

$$\text{FPS} = \frac{K}{T} \quad (7)$$

$$\text{F1} = \frac{2 \times P \times R}{P + R} \times 100\% \quad (8)$$

式中 T_p 指在检测过程中被正确识别为目标类别的样本数量； F_p 指在检测过程中被错误识别为目标类别，但实际上并不属于该类别的样本数量； F_n 指在检测过程中未能识别为目标类别，但实际上属于该类别的样本数量； N 为检测类别的数目； AP_i 是第 i 个类别的平均精度； K 为处理图像数量， T 为处理时间，s。

2 结果与分析

2.1 消融试验

2.1.1 花穗检测模型消融试验结果

为评估花穗检测模型中对 YOLOv11 采用的两种改进策略的有效性，进行了消融试验，结果如表 4 所示。

表 4 DD-YOLOv11 模型消融试验结果

Table 4 Ablation study results of the DD-YOLOv11 model

DLKA	Dynamic head	精确率 Precision $P/\%$	召回率 Recall $R/\%$	平均精度 均值 mAP@ $_{0.5}/\%$	F1 分数 F1/%	推理速度 FPS/ (帧·s ⁻¹)
-	-	91.25	85.10	85.70	88.07	56.4
√	-	93.48	87.21	90.92	90.24	55.3
-	√	92.63	88.35	91.34	90.44	57.1
√	√	95.27	90.35	94.19	92.73	59.5

注：-表示不使用该模块，√表示使用该模块。下同。

Note: - means not using this module, and √ means using this module. The same below.

由表 4 可知，YOLOv11 模型的精确率、召回率、平均精度均值和 F1 分数分别为 91.25%、85.10%、85.70% 和 88.07%，整体性能存在一定局限。引入 DLKA 模块后，精确率提升至 93.48%，召回率提升至 87.21%，平均精度均值提升至 90.92%，F1 分数提升至 90.24%，结果表明，DLKA 模块在一定程度上扩大了特征感受野范围，从而增强了模型对小目标花穗特征的表达能力，使检测性能得到改善。引入 Dynamic head 模块后，精确率提升至 92.63%，召回率提升至 88.35%，平均精度均值提升至 91.34%，F1 分数提升至 90.44%。其中召回率提升幅度比较显著，表明动态检测头通过增强分类与定位特征的表达能力，有助于降低花穗的漏检率。同时引入 DLKA 与 Dynamic head 后，模型的精确率达到 95.27%，召回率提升至 90.35%，mAP@ $_{0.5}$ 提升至 94.19%，F1 分数达 92.73%，相较基准模型，改进模型的精确率、召回率、mAP@ $_{0.5}$ 和 F1 分数分别提升 4.02、5.25、8.49 和 4.66 个百分点，说明两者具有良好的互补性。整体而言，两种改进策略联合应用能够有效提升花穗检测性能。

2.1.2 花朵状态识别模型消融试验结果

为评估各模块在花朵状态识别任务中的贡献，本研究以 YOLOv11 为基准模型，通过逐步引入关键模块进行消融试验，分析其对各项性能指标的影响，结果如表 5 所示。

由表可知，YOLOv11 模型在花朵状态识别任务中的精确率、召回率、平均精度均值和 F1 分数分别为 91.02%、88.77%、90.27% 和 89.88%，各类别 AP 显示，花蕾期和谢花期识别性能相对较低，表明模型在复杂背景与遮挡场景下仍存在一定提升空间。

表 5 MA-YOLOv11 模型消融试验结果
Table 5 Ablation study results of the MA-YOLOv11 model

MobileNetV4	ATFL	精确率 Precision P/%	召回率 Recall R/%	AP/%					平均精度均值 mAP@ _{0.5} /%	F1 分数 F1/%	推理速度 FPS/ (帧·s ⁻¹)
				花蕾期 Bud	花裂期 Cack	全开期 Full	谢花期 Witherr	初果期 Initial			
-	-	91.02	88.77	84.50	92.83	96.21	87.58	90.31	90.27	89.88	60.5
√	-	93.36	89.45	88.12	93.05	96.42	92.14	91.16	92.18	91.36	61.8
-	√	92.58	90.13	85.74	93.81	97.05	93.61	95.14	93.07	91.34	60.4
√	√	96.26	91.74	91.61	97.37	98.93	94.05	95.94	95.58	93.95	61.1

在保持整体网络结构不变的情况下，将主干网络替换为 MobileNetV4 后，模型的精确率、召回率、平均精度均值和 F1 分数分别提升至 93.36%、89.45%、92.18% 和 91.36%。各花朵状态的 AP 均有不同程度提升，尤其是花蕾期与初果期的识别效果改善较为明显。表明 MobileNetV4 通过深度可分离卷积与轻量化设计，在降低冗余计算的同时增强了对全局语义信息的建模能力，有助于提升模型在复杂环境下的特征提取效率与判别稳定性。

在基准模型中引入 ATFL 模块后，精确率和召回率分别提升至 92.58% 和 90.13%，平均精度均值达到 93.07%。从各类别 AP 分析可见，与基准模型相比，花裂期、全开期及谢花期的识别性能提升显著，表明 ATFL 损失函数通过自适应调制因子与阈值机制，增强模型对难分类样本的关注，从优化目标函数层面提升花朵状态的判别能力。

当同时引入 MobileNetV4 与 ATFL 模块后，模型在精确率、召回率、平均精度均值和 F1 分数指标上分别达到 96.26%、91.74%、95.58% 和 93.95%，较 YOLOv11 模型提升了 5.24、2.97、5.31 和 4.07 个百分点，各花朵状态的 AP 值均有所提升。结果表明，MobileNetV4 在全局特征提取与计算效率方面的优势，与 ATFL 模块在局部细粒度特征增强和多尺度信息建模方面形成了良好的互补关系。两者协同作用，使模型在保持高效推理能力的同时，提升了对复杂花朵状态的识别精度与鲁棒性。

2.2 不同模型检测性能对比

2.2.1 花穗检测模型结果分析

为验证本文所提出的 DD-YOLOv11 花穗检测模型的有效性，选取 SSD、Faster R-CNN、RT-DETR-R34、YOLOv5、YOLOv8、YOLOv11、YOLOv12 和 YOLOv13 等主流目标检测算法进行对比分析，试验结果如表 6 所示。

由表 6 可知，SSD 与 Faster R-CNN 整体性能较弱。SSD 的精确率和召回率仅为 72.84% 和 65.37%，平均精度均值不足 70%，精度明显不足。Faster R-CNN 的精

率虽达到 83.92%，但召回率和检测速度均明显不足。RT-DETR-R34 在检测精度和推理速度方面具有一定优势，精确率和平均精度均值分别达到 88.73% 和 86.87%，但召回能力受限，召回率仅为 76.48%。

YOLO 系列模型在检测精度和实时性之间表现出较好的平衡性。其中 YOLOv5 的召回率为 88.30%，F1 分数达到 86.47%，具有较好的召回能力，但精确率相对偏低。YOLOv8 在精确率和平均精度均值指标上分别提升至 90.98% 和 89.09%，但召回率下降至 78.37%，对部分花穗目标存在漏检。YOLOv11 在精确率、召回率和 F1 分数分别达到 91.25%、85.10% 和 88.07%，但平均精度略低。YOLOv12 与 YOLOv13 在精确率、召回率及检测速度方面与 YOLOv11 表现接近，但在花穗检测场景中，整体性能有限。

表 6 DD-YOLOv11 模型试验结果对比
Table 6 Comparison of experimental results of the DD-YOLOv11 model and other models

模型 Models	精确率 Precision P/%	召回率 Recall R/%	平均精度 均值 mAP@ _{0.5} /%	F1 分数 F1/%	推理速度 FPS/ (帧·s ⁻¹)
SSD	72.84	65.37	68.12	68.89	43.6
Faster R-CNN	83.92	76.25	78.46	79.91	19.2
RT-DETR-R34	88.73	76.48	86.87	82.19	44.3
YOLOv5	84.70	88.30	87.10	86.47	52.6
YOLOv8	90.98	78.37	89.09	84.21	55.1
YOLOv11	91.25	85.10	85.70	88.07	56.4
YOLOv12	90.78	83.46	87.92	86.96	56.1
YOLOv13	90.96	84.29	85.07	87.32	55.9
DD-YOLOv11	95.27	90.35	94.19	92.73	59.5

本研究改进后的 DD-YOLOv11 花穗检测模型的精确率、召回率、平均精度均值及 F1 分数分别达到 95.27%、90.35%、94.19% 和 92.73%，同时推理速度提升至 59.5 FPS。DLKA 注意力机制增强了花穗关键特征的表达能力，Dynamic head 优化了多尺度特征融合过程，从而提高了小目标花穗的检测性能。

2.2.2 花朵状态识别结果分析

为评估 MA-YOLOv11 花朵状态识别模型的性能，选取多个目标检测模型进行对比分析，试验结果如表 7 所示。

表 7 MA-YOLOv11 模型试验结果对比
Table 7 Comparison of experimental results of the MA-YOLOv11 model and other models

模型 Models	精确率 Precision P/%	召回率 Recall R/%	AP/%					平均精度均值 mAP@ _{0.5} /%	F1 分数 F1/%	推理速度 FPS/ (帧·s ⁻¹)
			花蕾期 Bud	花裂期 Crack	全开期 Full	谢花期 Witherr	初果期 Initial			
SSD	74.36	66.42	56.31	77.85	88.03	58.69	63.97	68.97	70.17	44.5
Faster R-CNN	82.95	75.80	68.11	83.41	92.84	72.77	79.57	79.34	79.20	21.7
RT-DETR-R34	87.91	84.26	80.15	87.62	93.58	82.97	86.83	86.23	86.05	46.7
YOLOv5	85.50	82.30	75.21	87.73	93.76	80.04	84.26	84.20	83.88	56.2
YOLOv8	89.10	86.50	82.84	90.12	95.11	85.93	89.75	88.75	87.80	58.7
YOLOv11	91.02	88.77	84.50	92.83	96.21	87.58	90.31	90.27	89.88	60.5
YOLOv12	90.47	87.29	82.86	91.37	95.15	85.82	88.94	88.83	88.85	59.7
YOLOv13	89.86	86.43	82.11	90.59	94.62	84.73	87.86	87.98	88.12	58.9
MA-YOLOv11	96.26	91.74	91.61	97.37	98.93	94.05	95.94	95.58	93.95	61.1

由表 7 可知, SSD 和 Faster R-CNN 模型的平均精度均值分别为 68.97% 和 79.34%, 尤其在花蕾期、谢花期的识别效果较差, 不适用于花朵状态识别; RT-DETR-R34 整体性能较为均衡, 其精确率、召回率和平均精度均值分别为 87.91%、84.26% 和 86.23%, 但在花蕾期和谢花期的 AP 分别为 80.15% 和 82.97%, 对花朵状态的区分能力有限。

YOLOv5 和 YOLOv8 性能有所提升, 平均精度均值分别达到 84.20% 和 88.75%, 但对关键状态花朵的识别精度仍然有限。YOLOv11 在各花朵状态识别上表现较为均衡, 精确率、召回率和平均精度均值分别达到 91.02%、88.77% 和 90.27%, 整体表现稳定且实时, 但识别精度仍有提升空间。YOLOv12 与 YOLOv13 的整体性能与 YOLOv11 接近, 未在花朵状态识别任务中体现出明显优势。其中, YOLOv12 的平均精度均值为 88.83%, YOLOv13 为 87.98%, 精度性能不高。

改进后的 MA-YOLOv11 模型通过引入 MobileNetV4 主干网络和 ATFL 损失函数, 提升了特征提取能力及关键状态花朵的检测性能。试验结果显示, MA-YOLOv11 的精确率、召回率、F1 分数和平均精度均值分别达到 96.26%、91.74%、93.95% 和 95.58%, 均高于其他对比模型; 其中, 花蕾期、花裂期、全开期、谢花期和初果期的 AP 分别达到 91.61%、97.37%、98.93%、94.05% 和 95.94%, 在 5 种花朵状态识别任务中均表现出较高的检测精度。同时, 模型推理速度达到 61.1 帧/s, 为所有对比模型中的最高水平。结果表明, MA-YOLOv11 在保证实时性的同时, 显著提升了不同花朵状态的检测精度和整体识别性能, 尤其增强了花裂期和全开期等关键授粉阶段的识别能力, 为番茄精准授粉作业提供了可靠的技术支撑。

2.2.3 两阶段检测分析

为验证两阶段级联检测流程在花朵状态识别任务中的整体性能, 本文对比分析了单阶段 MA-YOLOv11 整图花朵状态识别方法与两阶段级联检测流程在最终花朵状态输出上的检测效果。单阶段方法直接对整幅图像进行花朵状态识别, 而两阶段方法在识别前增加了花穗检测与区域裁剪过程。结果如表 8 所示。

表 8 单阶段与两阶段级联检测流程性能对比
Table 8 Performance comparison between single-stage and two-stage detection pipelines

检测流程 Detection pipeline	精确率 Precision P/%	召回率 Recall R/%	F1 分数 F1/%	平均精度均值 mAP@0.5/%	推理速度 FPS/ (帧·s ⁻¹)
单阶段 Single-stage	84.62	81.35	82.95	83.78	45.8
两阶段 Two-stage	92.18	86.47	89.24	90.83	38.4

从表 8 可见, 两阶段级联检测流程相比单阶段整图识别在花朵状态识别任务中表现更优。两阶段方法的精确率、召回率、F1 分数和平均精度均值分别达到 92.18%、86.47%、89.24% 和 90.83%, 较单阶段分别提升 7.56、5.12、6.29 和 7.05 个百分点。在推理效率方面, 在设施农业连续监测与农事辅助作业中, 视觉感知系统通常在 10~20 帧/s 可满足实时感知与决策需求。两阶段流程的平均推理速度为 38.4 帧/s, 相比单阶段的 45.8 帧/s

略有下降, 主要由于增加了花穗检测与裁剪步骤, 但仍满足实时处理需求。综合来看, 两阶段流程在提升识别精度的同时保持较高帧率, 适用于花朵状态识别任务。

为分析不同模型对检测任务的适应性, 本文将 DD-YOLOv11 和 MA-YOLOv11 分别应用于花穗检测和花朵状态识别任务, 并对其性能进行比较, 结果如表 9 所示。

表 9 不同模型在花穗与花朵检测任务上的跨任务性能对比
Table 9 Cross-task performance comparison of different models

模型 Models	检测对象 Detection target	精确率 Precision P/%	召回率 Recall R/%	平均精度均值 mAP@0.5/%	F1 分数 F1/%
DD-YOLOv11	花穗	95.27	90.35	94.19	92.73
	花朵	78.63	72.47	74.12	75.47
MA-YOLOv11	花穗	82.31	79.12	80.24	80.70
	花朵	96.26	91.74	95.58	93.95

如表 9 所示, DD-YOLOv11 在花朵检测任务上的精确率、召回率、平均精度均值和 F1 分数分别为 78.63%、72.47%、74.12% 和 75.47%, 低于其在花穗检测任务中的性能。主要因为 DD-YOLOv11 引入的 DLKA 注意力机制和 Dynamic Head 主要针对多尺度花穗特征进行优化, 对花朵的状态特征表征不足, 导致跨任务性能下降。另一方面, MA-YOLOv11 在花穗检测任务上的精确率、召回率、平均精度均值和 F1 分数为 82.31%、79.12%、80.24% 和 80.70%, 低于其在花朵检测任务的指标。这是由于 MA-YOLOv11 采用轻量化主干 MobileNetV4 并结合 ATFL 损失函数优化花朵状态特征, 其对花穗的空间与形态特征捕捉能力有限。上述结果表明, 两个模型在非本任务上的性能均有所下降, 验证了针对不同任务特点进行模型设计的必要性。

2.3 模型结果可视化分析

2.3.1 花穗检测结果可视化分析

由图 8 的可视化结果可以看出, 基准模型 YOLOv11 能够基本定位花穗目标, 但在检测框边界拟合与目标完整性方面仍存在一定不足。相比之下, 本文提出的 DD-YOLOv11 模型在花穗边界的划分上更加精准, 检测结果更为完整。在开花坐果期, 图像中仅包含单个花穗, DD-YOLOv11 能够保持较高的检测准确性; 在结果期, 随着花穗数量增多, 模型需要同时区分多个小尺度目标, DD-YOLOv11 仍能够保持较高的检测完整性与置信度, 体现出良好的检测能力。

2.3.2 花朵状态识别结果可视化分析

图 9 的可视化结果表明, YOLOv11 在多数情况下可检测到花朵目标, 但仍存在一定程度的误检与漏检。在样品 1 中, YOLOv11 将花穗中部的枝条结构误判为花蕾; 在样品 2 中, YOLOv11 将右下方处于初果期的花朵错误分类为全开期; 在样品 3 中, YOLOv11 存在花蕾目标漏检的情况。相比之下, MA-YOLOv11 在上述场景中未出现相应的误检或漏检问题, 能够对不同花期的花朵实现稳定、准确的识别。同时, 在检测框的边界拟合上更加精准, 目标贴合度更高, 置信度分布整体高于 YOLOv11。

为直观分析不同模型在花朵状态识别任务中的分类性能差异, 本文引入混淆矩阵对 YOLOv11 与 MA-YOLOv11 在各花期状态上的识别与混淆情况进行对比分析, 如图 10 所示。

由 YOLOv11 的混淆矩阵可知, 模型虽能较好识别

各花期类别,但在相邻生育阶段仍存在一定混淆,主要表现为花蕾期与花裂期、以及谢花期与初果期之间的误判,表明模型对细粒度花期差异的区分能力仍有不足。相比之下,MA-YOLOv11 的混淆矩阵呈现出更加集中的对角线分布特征。各花期类别的正确识别样本数量均有所提升,且非对角线元素显著减少,表明模型对不同花期状态的判别更加准确。其中,花蕾期和花裂期的混淆数量降低;同时,谢花期与初果期之间的混淆亦得到有效缓解,模型对花朵生长后期状态的区分能力增强。

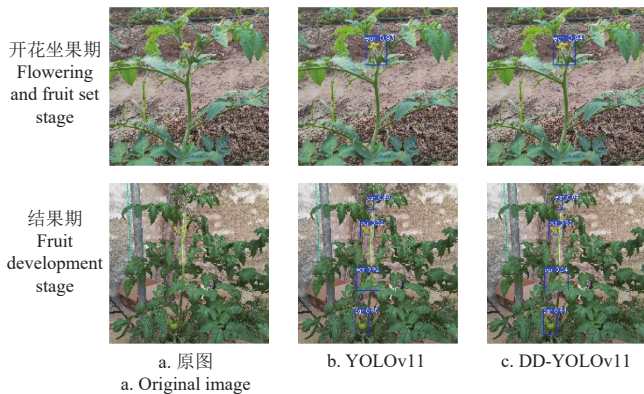


图 8 DD-YOLOv11 模型可视化结果对比

Fig.8 Comparison of visualization results of the DD-YOLOv11 model

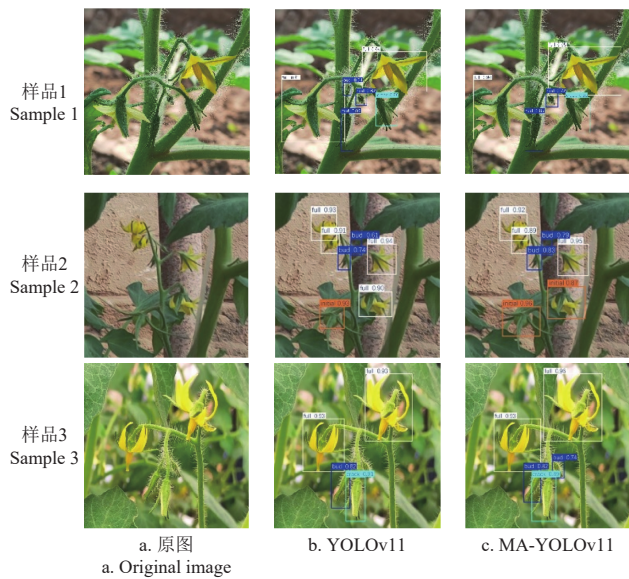


图 9 MA-YOLOv11 模型可视化结果对比

Fig.9 Comparison of visualization results of the MA-YOLOv11 model

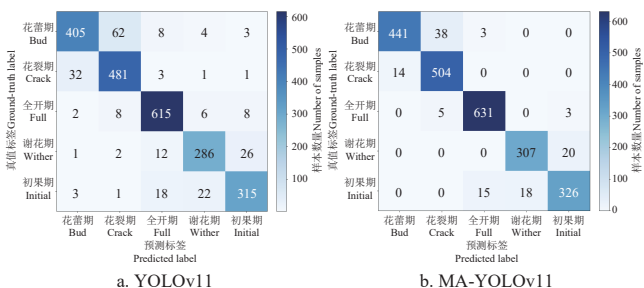


图 10 不同模型混淆矩阵

Fig.10 Confusion matrices of different models

总体来看, DD-MA-YOLOv11 在各花期状态上的分类稳定性优于 YOLOv11。

3 结论

本文针对设施环境下番茄花穗目标尺度小、花朵状态识别易受遮挡影响等问题,提出了 DD-MA-YOLOv11 两阶段级联识别模型,主要结论如下:

1) 提出了 DD-YOLOv11 花穗检测模型。通过引入 DLKA 注意力机制和 Dynamic head 模块,增强了模型对小尺度花穗目标的特征提取能力,花穗检测任务的召回率和平均精度均值分别提高 5.25 和 8.49 个百分点,有效提升了复杂设施环境下花穗目标的检测性能,为后续花朵状态识别提供了可靠的目标定位结果。

2) 提出了 MA-YOLOv11 花朵状态识别模型。相比基准 YOLOv11 模型,精确率、召回率、平均精度均值和 F1 分数分别提高 5.24、2.97、5.31 和 4.07 个百分点,提高了番茄花朵细粒度状态识别精度,增强了不同花朵状态之间的判别能力。

3) 构建了 DD-MA-YOLOv11 两阶段级联识别框架,实现了花穗定位、目标区域裁剪与花朵状态识别的协同处理。相比单阶段识别方案,两阶段级联策略能够兼顾花穗检测精度与花朵状态识别性能,模型平均推理速度达到 38.4 帧/s,在保证识别精度的同时满足设施农业连续监测与在线作业的实时性要求,为设施番茄花期智能监测提供了一种有效的技术方案。

尽管本文提出的方法在典型设施环境下取得了较好的检测与识别效果,但在花朵高度密集、重叠严重场景下,对个体花朵的区分能力仍有进一步提升空间,且模型在极端光照变化条件下的鲁棒性仍需进一步增强。未来将重点研究实例分割技术在重叠花朵分离中的应用,并结合多光谱、多模态信息融合以及迁移学习和领域自适应等方法,提高模型在复杂环境及不同作物场景中的泛化能力,为设施农业智能感知与精准管理提供更加可靠的技术支撑。

[参考文献]

- [1] 常心怡, 吉光鹏, 牛岭磊, 等. 地力差减法在南疆温室大棚番茄推荐施肥中的应用研究[J]. 中国瓜菜, 2023, 36(9): 60-65.
CHANG Xinyi, JI Guangpeng, NIU Linglei, et al. Study on the application of differential ground strength reduction method in the recommended fertilization of tomatoes in greenhouse in Southern Xinjiang[J]. China Cucurbits and Vegetables, 2023, 36(9): 60-65. (in Chinese with English abstract)
- [2] 顾文娟, 刘浩状, 魏金, 等. 基于 FPBW-YOLO v8 的复杂场景下番茄果实识别方法[J]. 农业机械学报, 2025, 56(8): 467-478.
GU Wenjuan, LIU Haozhuang, WEI Jin, et al. Tomato Fruit Recognition in Complex Scenes Based on FPBW-YOLO v8[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(8): 467-478. (in Chinese with English abstract)
- [3] 刘中良, 高俊杰, 张艳艳, 等. 不同有机基质配方对设施番茄产量及品质的影响[J]. 上海交通大学学报(农业科学版), 2019, 37(3): 34-38.
LIU Zhongliang, GAO Junjie, ZHANG Yanyan, et al. Effects of organic substrates on yield and quality of tomato in solar greenhouse[J]. Journal of Shanghai Jiao Tong University

- (Agricultural Science Edition), 2019, 37(3): 34-38. (in Chinese with English abstract)
- [4] 牛艳, 王晓静, 陈翔, 等. 中国番茄产业发展的现状问题和对策及宁夏番茄产业发展成效[J]. 黑龙江农业科学, 2022(12): 70-74.
NIU Yan, WANG Xiaojing, CHEN Xiang, et al. Current Situation, Problems and Countermeasures of Tomato Industry Development in China and Achievements of Tomato Industry Development in Ningxia[J]. Heilongjiang Agricultural Sciences, 2022(12): 70-74. (in Chinese with English abstract)
 - [5] 曾华. 设施番茄优质高产水肥管理技术[J]. 现代农村科技, 2023 (5): 70, 64.
 - [6] 孙艳军, 徐刚, 高文瑞, 等. 不同授粉方式对日光温室番茄产量、品质及效益的影响[J]. 中国蔬菜, 2017(6): 38-41.
SUN Yanjun, XU Gang, GAO Wenrui, et al. Effects of Different Pollination Methods on Yield, Quality and Benefit of Tomato in Solar Greenhouse[J]. China Vegetables, 2017(6): 38-41. (in Chinese with English abstract)
 - [7] 王锐竹, 王丽, 田猛, 等. 熊蜂授粉对设施番茄产量、品质及效益的影响[J]. 天津农林科技, 2018(2): 5-6.
 - [8] 魏代国. 设施番茄花期提高坐果率的关键措施[J]. 农业工程技术, 2024, 44(18): 60-61.
 - [9] OHTA K, IKEDA D. Effects of pinching treatment on harvest term and plant growth in processing tomato[J]. Canadian Journal of Plant Science, 2017, 97(1): 92-98.
 - [10] 任晶秋, 万恩晗, 单蜜, 等. 基于改进 YOLO v8n 网络的番茄成熟度实时检测算法[J]. 农业机械学报, 2025, 56(3): 374-382, 450.
REN Jingqiu, WAN Enhuan, SHAN Mi, et al. Improved YOLO v8n Network for Real-time Detection of Tomato Maturity[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(3): 374-382, 450. (in Chinese with English abstract)
 - [11] 刘芳, 刘玉坤, 林森, 等. 基于改进型 YOLO 的复杂环境下番茄果实快速识别方法[J]. 农业机械学报, 2020, 51(6): 229-237.
LIU Fang, LIU Yukun, LIN Sen, et al. Fast Recognition Method for Tomatoes under Complex Environments Based on Improved YOLO[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(6): 229-237. (in Chinese with English abstract)
 - [12] GAO G, SHUAI C, WANG S, et al. Using improved YOLO V5s to recognize tomatoes in a continuous working environment[J]. Signal, Image and Video Processing, 2024, 18: 4019-4028.
 - [13] 彭程, 李帅, 苗艳龙, 等. 基于三维点云的番茄植株茎叶分割与表型特征提取[J]. 农业工程学报, 2022, 38(9): 187-194.
PENG Cheng, LI Shuai, MIAO Yanlong, et al. Stem-leaf segmentation and phenotypic trait extraction of tomatoes using three-dimensional point cloud[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(9): 187-194. (in Chinese with English abstract)
 - [14] ZHANG L, HUANG Z, YANG Z, et al. Tomato stem and leaf segmentation and phenotype parameter extraction based on improved red billed blue magpie optimization algorithm[J]. Agriculture, 2025, 15: 180.
 - [15] WANG Y, LIU Q, YANG J, et al. A method for tomato plant stem and leaf segmentation and phenotypic extraction based on Skeleton extraction and supervoxel clustering[J]. Agronomy, 2024, 14: 198.
 - [16] 余贤海, 孔德义, 谢晓轩, 等. 基于深度学习的番茄授粉机器人目标识别与检测[J]. 农业工程学报, 2022, 38(24): 129-137.
YU Xianhai, KONG Deyi, XIE Xiaoxuan, et al. Deep learning-based target recognition and detection for tomato pollination robots[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(24): 129-137. (in Chinese with English abstract)
 - [17] 赵春江, 文朝武, 林森, 等. 基于级联卷积神经网络的番茄花期识别检测方法[J]. 农业工程学报, 2020, 36(24): 143-152.
ZHAO Chunjiang, WEN Chaowu, LIN Sen, et al. Tomato flowering stage recognition and detection method based on cascaded convolutional neural networks[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(24): 143-152. (in Chinese with English abstract)
 - [18] 文朝武, 龙洁花, 张宇, 等. 基于 3D 视觉的番茄授粉花朵定位方法[J]. 农业机械学报, 2022, 53(8): 320-328.
WEN Chaowu, LONG Jiehua, ZHANG Yu, et al. Positioning method of tomato pollination flowers based on 3D vision[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(8): 320-328. (in Chinese with English abstract)
 - [19] LIU S Y, ZHANG X M, WANG X Y, et al. Tomato flower pollination features recognition based on binocular gray value-deformation coupled template matching[J]. Computers and Electronics in Agriculture, 2023, 214: 108345.
 - [20] 邓颖, 吴华瑞, 朱华吉. 基于实例分割的柑橘花朵识别及花量统计[J]. 农业工程学报, 2020, 36(7): 200-207.
DENG Ying, WU Huarui, ZHU Huaji. Recognition and counting of citrus flowers based on instance segmentation[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(7): 200-207. (in Chinese with English abstract)
 - [21] 李翠明, 杨柯, 申涛, 等. 基于改进 Faster R-CNN 的苹果采摘视觉定位与检测方法[J]. 农业机械学报, 2024, 55(01): 47-54.
LI Cuiming, YANG Ke, SHEN Tao, et al. Vision detection method for picking robots based on improved Faster R-CNN[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(01): 47-54. (in Chinese with English abstract)
 - [22] 杨碧云, 叶丽萍, 田珺, 等. 番茄采摘机器人关键技术研究进展、挑战与展望[J]. 吉林农业大学学报, 2025, 47(4): 741-752.
YANG Biyun, YE Liping, TIAN Jun, et al. Research progress, challenges, and prospects of key technologies of tomato harvesting robots[J]. Journal of Jilin Agricultural University, 2025, 47(4): 741-752. (in Chinese with English abstract)
 - [23] 尚钰莹, 张倩如, 宋怀波. 基于 YOLOv5s 的深度学习在自然场景苹果花朵检测中的应用[J]. 农业工程学报, 2022, 38(9): 222-229.
SHANG Yuying, ZHANG Qianru, SONG Huaibo. Application of deep learning using YOLOv5s to apple flower detection in natural scenes[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(9): 222-229. (in Chinese with English abstract)
 - [24] 吕志远, 张付杰, 魏晓明, 等. 采用组合增强的 YOLOX-ViT 协同识别温室室内番茄花果[J]. 农业工程学报, 2023, 39(4): 124-134.
LYU Zhiyuan, ZHANG Fujie, WEI Xiaoming, et al. Synergistic recognition of tomato flowers and fruits in

- greenhouse using combination enhancement of YOLOX-ViT[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2023, 39(4): 124-134. (in Chinese with English abstract)
- [25] HAO Y, RAO L, FU X, et al. Tomato ripening detection in complex environments based on improved BiAttFPN fusion and YOLOv11-SLBA modeling[J]. *Agriculture*, 2025, 15: 1310.
- [26] KHANAM R, HUSSAIN M. YOLOv11: An overview of the key architectural enhancements[EB/OL]. (2024-10-23)[2025-05-26]. <https://arxiv.org/abs/2410.17725>.
- [27] AZAD R, ARIMOND R, AGHDAM E K P, et al. Beyond self-attention: deformable large kernel attention for medical image segmentation[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA, 2024: 1287-1297.
- [28] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks[C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE Computer Society, 2017: 764-773.
- [29] DAI X, CHEN Y, XIAO B, et al. Dynamic head: Unifying object detection heads with attentions[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: IEEE Computer Society, 2021: 7373-7382.
- [30] 马鹏伟, 周杰. 基于改进 YOLOv7 的复杂环境下的葡萄成熟度检测[J]. *农业工程学报*, 2025, 41(3): 171-178.
- MA Pengwei, ZHOU Jie. Detecting grape ripeness in complex environments using improved YOLOv7[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2025, 41(3): 171-178. (in Chinese with English abstract)
- [31] ZHAO M, DUAN Y, GAO T, et al. A lightweight model for small-target pig eye detection in automated estrus recognition[J]. *Animals*, 2025, 15: 1127.

Method for detecting tomato flower clusters and recognizing flower status based on DD-MA-YOLOv11

GAO Yanan^{1,2,3}, LIU Pingzeng^{1,2,3*}, ZHANG Yuxuan^{1,2,3}, ZHU Ke^{1,2,3},
ZHANG Yan^{1,2,3}, YU Qun^{1,2,3}, WEN Fujiang¹

(1. School of Information Science and Engineering, Shandong Agricultural University, Tai'an 271018, China; 2. Key Laboratory of Huang-Huai-Hai Smart Agricultural Technology, Ministry of Agriculture and Rural Affairs, Tai'an 271018; 3. Agricultural Big-Data Research Center, Shandong Agricultural University, Tai'an 271018, China)

Abstract: Accurate perception of tomato inflorescences and flower states can often be required to support key operations in greenhouse tomato production, such as pollination and topping at the flowering and fruiting stage. However, inflorescences and flowers are characterized by small target size, dense spatial distribution, complex backgrounds, and frequent occlusion by leaves and stems in practical greenhouse environments. Single-stage detection models cannot simultaneously realize stable inflorescence localization and high-precision flower state recognition when operating on whole-plant images. In this study, a two-stage cascaded framework of visual perception was developed to detect tomato inflorescence and flower states using an improved YOLO version 11 network. A “spatial localization followed by fine-grained recognition” strategy was adopted to decompose the overall perception task into two sequential subtasks. In the first stage, an inflorescence detection model was constructed to enhance the baseline YOLO version eleven network with a Deformable Large Kernel Attention mechanism and a Dynamic Head detection structure. Deformable Large Kernel Attention Mechanism also employed large convolutional kernels with deformable convolution to capture long-range contextual information between inflorescences and adjacent peduncles, while morphological variations were also considered at different growth stages and plant structures. The dynamic head module incorporated scale- and spatial-aware feature modeling, thereby enabling the detector to robustly handle inflorescences of varying sizes for the complex regions, where inflorescences and leaves overlapped. The output precise spatial regions corresponded to inflorescences, which served as reliable regions of interest for subsequent analysis. In the second stage, a flower state recognition model was designed to operate exclusively within the inflorescence regions in the first stage. A lightweight backbone network, MobileNetV4, was adopted to reduce computational complexity for inference efficiency, while preserving feature representation. An Adaptive Task-aligned Focal Loss function was introduced to balance sample distribution among different flower developmental states. This loss function dynamically adjusted category weights, according to classification difficulty and sample frequency, thereby enhancing recognition performance for the minority and easily confused flower states under occlusion and cluttered backgrounds. Experiments were conducted on a greenhouse tomato image dataset at multiple growth stages and complex environments. In the inflorescence task, the first-stage model achieved substantial improvements in performance, compared with the baseline network, with the precision, recall, mean average precision at an intersection-over-union threshold of 0.5, and F1-score increasing by 4.02, 5.25, 8.49, and 4.66 percentage point, respectively. These results demonstrated that the attention and detection head enhancements significantly improved small-target detection stability in complex scenes. In the flower state recognition task, the second-stage model further improved precision, recall, mean average precision at the same threshold, and F1-score by 5.24, 2.97, 5.31, and 4.07 percentage point, respectively, indicating stronger identification for fine-grained flower state classification. The cascaded framework achieved an average processing speed of 38.4 frames per second under a single-input condition, fully meeting the real-time requirements of continuous greenhouse monitoring and online agricultural operations. The cascaded framework effectively balanced detection accuracy and computational efficiency to decouple spatial localization from fine-grained recognition. The reliable inflorescence and flower state recognition from whole-plant images can provide a practical visual perception for pollination, topping, and intelligent operations in greenhouse tomato production.

Keywords: greenhouse tomato; inflorescence detection; flower state recognition; cascaded object detection; YOLOv11