doi:10.6041/j.issn.1000-1298.2026.13.015

基于 DGA－DETR 模型的芦笋选择性采收精准识别方法研究

夏先飞[1,2]　吕俊潼[1,2]　韦树谷[3]　宫庆硕[1,4]　黄　玲[3]　王申莹[1,4]　梁明旭[1,5]

（1.农业农村部南京农业机械化研究所，南京 210014；2.苏州中农院华东农业科技中心，苏州 215331；
3.四川省农业科学院经济作物研究所，成都 610300；4.农业农村部现代农业装备重点实验室，南京 210014；
5.南京工程学院机械工程学院，南京 211167）

摘要：针对芦笋机器人采收作业过程中，芦笋植株分布密集、遮挡严重导致感知精度不足、漏检率高、重叠实例粘连及跨域泛化能力弱等问题，对芦笋细长几何形态特性和密集遮挡的空间特征进行深入分析，提出了一种融合动态门控注意力与边缘感知下采样机制的实例分割模型 DGA－DETR（Dynamic gated attention-detection Transformer）。通过设计动态门控注意力（DGA）模块，利用实例语义驱动的动态门控机制实现跨空间与跨尺度的特征精准聚焦，从机理层面有效抑制密集遮挡下背景噪声干扰；构建了边缘感知上下采样机制，以强化边界解耦能力，提升模型对芦笋细长形态及遮挡交界区域的分割建模精度。为验证分割模型的有效性，构建了不同光照条件及遮挡程度的机收芦笋数据集用于模型训练，涵盖像素、场景、语义多层次域偏移，并搭建跨域数据集进行测试。实验结果表明，所提 DGA－DETR 模型在测试集上的掩码精确率、掩码召回率、Mask mAP@0.5 和 Mask mAP@0.5:0.95 分别达到 91.60%、67.24%、90.98% 和 59.13%，较基线模型分别提升 5.33、8.43、6.56、9.51 个百分点，显著降低了密集遮挡条件下芦笋实例漏分割与掩码不连续现象。跨域测试结果表明，该方法在遮挡与复杂背景条件下芦笋的精准识别表现出良好鲁棒性与泛化能力，研究结果可为其他嫩茎类作物在复杂环境下机器人采收作业提供技术参考。

关键词：采收机器人；芦笋；实例分割；遮挡识别；动态门控注意力；DGA－DETR 模型

中图分类号：S126　　**文献标识码**：A　　**文章编号**：1000-1298(2026)13-0187-13　　**OSID**：

Precise Identification Method for Selective Harvesting of Asparagus Based on DGA－DETR Model

Xia Xianfei[1,2]　Lü Juntong[1,2]　Wei Shugu[3]　Gong Qingshuo[1,4]　Huang Ling[3]　Wang Shenying[1,4]　Liang Mingxu[1,5]

(1. *Nanjing Institute of Agricultural Mechanization*, *Ministry of Agriculture and Rural Affairs*, *Nanjing* 210014, *China*
2. *East China Agricultural Science and Technology Center*, *Suzhou Zhongnong Institute*, *Suzhou* 215331, *China*
3. *Industrial Crop Institute*, *Sichuan Academy of Agricultural Sciences*, *Chengdu* 610300, *China*
4. *Key Laboratory of Modern Agricultural Equipment*, *Ministry of Agriculture and Rural Affairs*, *Nanjing* 210014, *China*
5. *School of Mechanical Engineering*, *Nanjing Institute of Technology*, *Nanjing* 211167, *China*)

Abstract: Aiming at the technical challenges in the robotic harvesting process of asparagus, such as insufficient perception accuracy, high miss rate, adhesion of overlapping instances, and weak cross-domain generalization caused by dense plant distribution and severe occlusion, an in-depth analysis of the slender geometric morphology and spatial features of dense occlusion in asparagus was conducted. An instance segmentation model, dynamic gated attention-detection transformer (DGA－DETR), which integrated dynamic gated attention and edge awareness, was proposed. By designing a dynamic gated attention (DGA) module, an instance-semantic-driven dynamic gating mechanism was utilized to achieve precise feature focusing across space and scales, effectively suppressing background noise interference under dense occlusion at the mechanistic level. Additionally, an edge-aware up-and-down sampling mechanism was constructed to strengthen boundary decoupling capabilities, improving the model's

收稿日期：2026－03－10　修回日期：2026－03－22

基金项目：苏州中农院华东农业科技中心科技创新项目（ECS－KY－N－2025006）、南京市重大科技专项（前沿技术）项目（202512115）和四川省农业科学院科技成果中试熟化与示范转化项目（2025ZSSFJD13）

作者简介：夏先飞（1987—），男，研究员，博士，主要从事农业机器人及收获机械化技术研究，E-mail：xiaxianfei@caas.cn

segmentation modeling accuracy for the slender morphology of asparagus and occlusion boundary regions. To verify the effectiveness of the segmentation model, an asparagus harvesting dataset covering different lighting conditions and occlusion levels was constructed for model training, encompassing multi-level domain shifts in pixels, scenes, and semantics, and a cross-domain dataset was established for testing. Experimental results demonstrated that the proposed DGA - DETR model achieved Mask precision, Mask recall, Mask mAP@0.5, and Mask mAP@0.5:0.95 of 91.60%, 67.24%, 90.98%, and 59.13%, respectively, representing improvements of 5.33, 8.43, 6.56, and 9.51 percentage points over the baseline model, and significantly reducing missed instances and mask discontinuities under dense occlusion conditions. Cross-domain test results indicated that this method exhibited excellent robustness and generalization capabilities for the precise identification of asparagus under occlusion and complex backgrounds. The research results can provide a technical reference for the robotic harvesting of other tender-stem crops in complex environments.

Key words: harvesting robot; asparagus; instance segmentation; occlusion identification; dynamic gated attention; DGA - DETR model

0 引言

芦笋作为我国重要的高附加值出口蔬菜,产业规模居世界首位,但其采收环节劳动力成本占比超30%,研制芦笋采收机器人已成为产业转型升级的迫切需求。芦笋嫩茎纤细且具有典型的密集簇生模式,田间作业环境存在严重的物理遮挡与复杂空间拓扑关系[1],导致现有视觉模型在像素级边缘感知、个体实例解耦及跨场景泛化等方面性能显著下降,难以满足采收机器人末端执行器精准定位与避障的作业需求[2]。

当前,国内外学者在农作物目标识别与分割领域已开展了大量研究,在农作物识别[3-4]、病虫害检测[5]等任务中取得了广泛应用。目标识别算法主要有两大范式,以 Faster R - CNN 等为典型的两阶段目标检测算法先通过区域生成提取候选区域,再利用卷积神经网络完成样本分类,检测精度优势显著,但检测速度慢,难以满足实时性采收要求[6]。以 YOLO 系列[7-9]、SSD[10]等算法为代表的单阶段目标检测算法可直接通过网络推理输出目标位置及其判别结果,平衡了速度与精度,更适配农业机器人实时作业的需求。为了获取更精细的形态特征[11],研究者引入了语义分割和实例分割技术。语义分割算法通过对图像每个像素分配语义标签,实现了对田间场景的整体结构解析,代表模型有 FCN、U - Net[12]、DeepLab[13]系列。然而,语义分割在处理密集群体时存在本质局限,它只能完成区域级语义区分,无法区分同类个体。当多株作物相互接触或遮挡时,语义分割会将相邻的茎秆判定为同一个连通区域,导致机器人无法提取单株芦笋的中心轴线。实例分割技术[14-15]通过融合目标检测的个体分类能力与语义分割的像素建模能力,实现了对同类中不同实例的独立掩码生成,代表模型有 YOLACT、SOLO 系列、Mask R - CNN[16]。然而,传统的基于卷积神经网络的实例分割模型多采用先检测后分割的流程作业,在处理芦笋这类目标极度密集且相互重叠的场景时,容易受限于候选框质量,导致重叠个体的特征提取出现混淆。此外,卷积算子的局部感受野特性使其难以捕捉细长茎秆在被遮挡后的全局空间联系,难以从整体上维持掩码的连贯性。

针对传统算法在复杂关联建模上的不足,近年来基于 Transformer 的端到端检测框架性能显著[17],通过全局注意力机制实现目标与查询的直接匹配,为密集遮挡场景下的实例解析提供了新范式。其中,RF - DETR[18]作为该系列的先进代表,融合神经架构搜索与多尺度特征融合策略,在精度与效率间取得良好平衡。然而,针对芦笋这类细长形态且处于高密度遮挡的农业目标,RF - DETR 仍存在明显局限。首先,其自注意力机制对所有查询进行无差别交互,致使密集区域的注意力资源易被土壤、杂草等背景噪声稀释,引发特征混淆;其次,特征融合策略与损失函数对遮挡边界敏感性及细长结构连续性的建模细粒度不足,在处理芦笋遮挡部分时,常因空间特征丢失导致分割掩码出现掩码不连续或边缘轮廓失真。

针对上述问题,提出一种面向密集遮挡场景的芦笋实例分割模型 DGA - DETR(Dynamic gated attention-detection Transformer)。基于 RF - DETR 基线框架,主要开展两项针对性改进:在解码阶段引入动态门控注意力(Dynamic gated attention,DGA)模块,通过由实例查询语义驱动的门控调制增强前景相关响应并抑制背景噪声;在分割头中设计边缘感知上下采样机制,通过条件融合与边界增强提升细长结构边界保持能力及重叠区域分离能力。为全面评估所提模型在复杂环境下的性能,构建包含像素、场景和语义三层次域偏移的芦笋图像数据集,并在

此基础上开展跨域测试实验，以系统分析模型在密集遮挡及复杂背景条件下的分割性能。

1　数据集采集与预处理

1.1　数据采集

本研究于 2025 年 8 月至 10 月在四川省农业科学院简阳芦笋基地开展芦笋视觉训练数据采集，采集过程全面覆盖芦笋生长环境的多样性。如图 1 所示，在不同光照、芦笋生长阶段、形态及背景复杂度下，采集芦笋静态图像与动态视频序列，获取芦笋及其生长环境的视觉数据。所有原始视觉数据均进行标准化预处理，视频数据通过关键帧提取技术转化为静态图像序列，作为模型训练基础数据。



图 1　跨域多场景数据集图例

Fig. 1　Cross-domain multi-scenario dataset

本次共采集原始图像 500 幅，其中包含使用相机拍摄的静态图像 420 幅，以及从动态视频序列中通过关键帧提取技术获得的静态图像 80 幅。采用 Labelme 工具对每幅图像中的芦笋实例进行精细化地多边形标注，严格遵循“每株独立、边界清晰、不重复、不漏标”的原则。所有标注结果经双人交叉验证，确保实例分割标注的准确性与一致性。采用分层随机抽样方式按照比例 8∶1∶1将数据集划分为训练集、验证集和测试集。

1.2　数据增强

针对农业场景下深度学习模型易因样本不足导致的过拟合问题，研究仅对训练集样本实施多维度数据增强策略提升模型的鲁棒性与泛化能力，将训练集样本规模扩充至 2 400 幅的具体方法包括：①几何空间增强：几何变换类增强实现样本的平移、旋转、缩放与翻转变换。②光学特征增强：通过像素级增强完成图像亮度、对比度、饱和度的自适应调整。③干扰注入：引入高斯噪声添加、色彩空间转换及随机裁剪等方式，丰富数据集的样本特征分布，实现数据集的有效扩充，数据增强效果如图 2 所示。

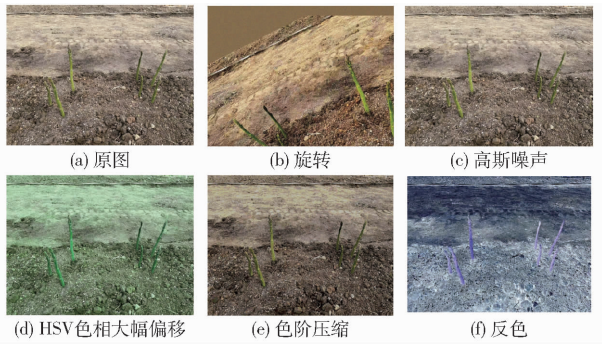


图 2　数据增强效果图

Fig. 2　Data augmentation results

1.3　跨域验证集构建

为验证 DGA－DETR 在未知复杂环境下的泛化性能，本研究系统性地构建了与训练集严格独立的芦笋跨域验证子集，如图 3 所示。

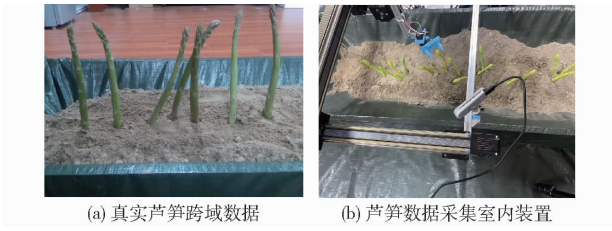


图 3　仅用于测试的跨域数据集

Fig. 3　Cross-domain dataset for testing purposes only

该子集通过在受控室内环境中还原真实田间生境，模拟密集种植下的极端工况，其核心在于集成 3 类相互交织的域偏移，构成对田间采收复杂性的系统化表征[19]：①像素偏移：利用不同感光特性的成像设备，在动态光照波动下进行采集，模拟自然光温差导致的色彩风格偏移及传感器噪声，旨在检验模型对像素级表观细节的鲁棒性。②空间偏移：通过调整相机俯仰角及水平视角，采集从单株清晰可见到多株高度重叠、交织的图像序列。这种空间拓扑结构的变化，重点评估模型在复杂遮挡下对个体实例的区分能力。③目标语义偏移：涵盖芦笋不同发育阶段的形态差异以及自然生长的尺度变异。该子集不参与模型的参数迭代与训练过程，仅用于验证模型的跨域泛化性能。

2　芦笋实例分割模型构建

2.1　DGA－DETR 基础骨干网络

针对芦笋在田间环境下呈现的密集簇生分布及植株间频繁重叠形成的复杂遮挡特性，本文构建了以 RF－DETR 为基础架构的 DGA－DETR 模型，RF－DETR基线模型本身已采用 DINOv2[20] 作为视觉骨干网络，本文并未对其骨干结构进行额外替换或重设计。基于此，在保持基线骨干网络与主体分割框架一致的前提下，重点针对密集遮挡场景下的特征筛选与边界恢复问题，引入 DGA 模块和边缘感知上下采样机制，以提升模型对芦笋细长结构、遮挡交界区域及重叠实例的表征能力。

如图 4 所示，将输入图像划分为图像块，并送入 Transformer 块堆叠而成的编码层中，提取具有丰富语义信息的多层次特征。在骨干网络末端，引入多

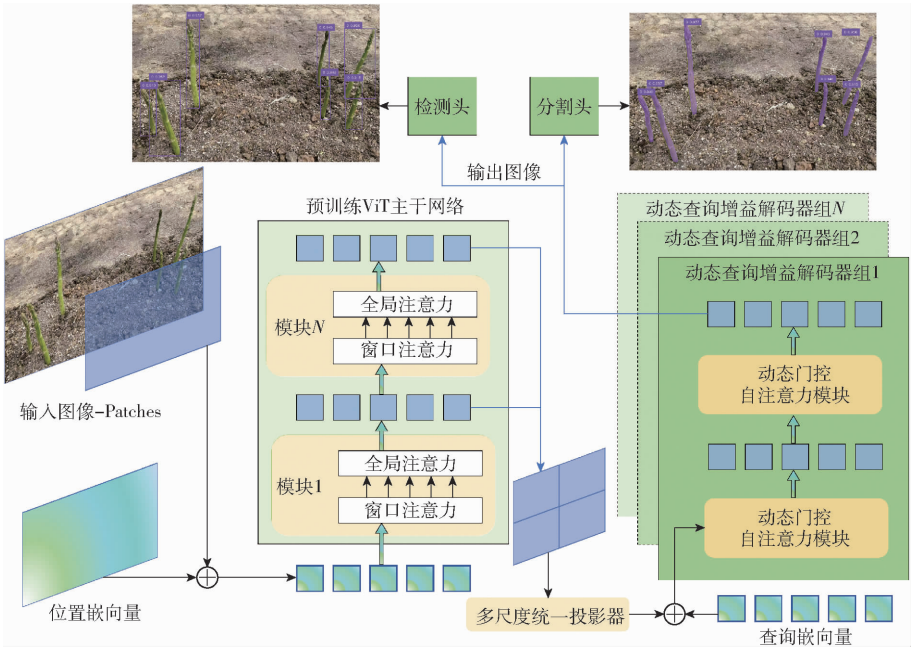


图 4 DGA－DETR 网络模型图

Fig. 4 DGA－DETR network model diagram

尺度统一投影器对来自不同层次的特征进行整合与降维,形成结构化的视觉先验。并将 DGA 模块嵌入于解码器中的“查询特征”交互过程中,该模块能够自适应地调节特征间的关系权重,从而在保持特征分辨率的同时增强全局上下文建模能力。

2.2 动态门控注意力模块

在非结构化农田环境下,RF－DETR 的原始自注意力机制将所有查询特征视为同质集合,进行无差别全局交互。这种无偏向性的全局交互模式极易受背景噪声与遮挡物干扰,芦笋目标具有显著的高长径比特征,且受茎叶遮挡及光影干扰,这导致目标查询特征被高熵的背景噪声稀释,特征判别边界趋于模糊,使得关键目标查询与背景查询的权重分布出现均值化倾向,降低了模型在密集工况下的识别精度。

受大语言模型门控注意力研究的启发[21],本研究设计 DGA 模块,以缓解密集遮挡条件下实例建模被背景噪声稀释的问题。该模块由实例查询特征自适应生成门控信号,并根据查询的实例语义状态动态变化,对注意力响应进行条件调制。由于解码器中的实例查询在迭代匹配过程中逐步聚合目标的形状线索、空间位置及局部上下文信息,因此得到实例语义驱动的门控向量可以抑制与背景干扰或邻近目标混叠相关的无效响应。基于此,DGA 并非对查询特征实施静态重加权,而是在实例语义约束下对注意力响应实施动态增益控制。其门控向量 $\boldsymbol{g}$ 表示为

$$\boldsymbol{g}=\sigma(\boldsymbol{W}_g\boldsymbol{Q}+\boldsymbol{b}_g)\quad(\boldsymbol{Q}\in\mathbf{R}^{N\times d}) \tag{1}$$

式中 $\boldsymbol{Q}$——解码器当前层的实例查询特征

N——特征点数量

d——特征维度

$\boldsymbol{W}_g$、$\boldsymbol{b}_g$——门控中可学习的权重矩阵与偏置项,$\boldsymbol{W}_g\in\mathbf{R}^{d\times d}$,$\boldsymbol{b}_g\in\mathbf{R}^d$

σ——Sigmoid 激活函数,将门控响应映射至稳定区间,用以表征不同特征点对目标的贡献程度

为防止在大规模并行计算中出现注意力过载或梯度爆炸,引入重缩放函数 $R(\boldsymbol{g},\tau)$,其可描述为

$$R(\boldsymbol{g},\tau)=1+\tau\tanh(\boldsymbol{g}) \tag{2}$$

式中,$\tanh(\boldsymbol{g})$ 函数将门控增益限制在范围内,参数 τ 控制了增益波动的最大振幅。

该设计确保了特征激活既能获得必要的增强,又不会脱离合理的数值区间,从而维持了深度神经网络在训练过程中的鲁棒性。

最终,DGA 模块将重缩放后的增益应用于原始注意力权重,并采用残差结构输出,即

$$\mathrm{DGA}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{Attn}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})\odot R(\boldsymbol{g},\tau) \tag{3}$$

其中

$$\mathrm{Attn}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V} \tag{4}$$

式中,$\boldsymbol{K}$ 为键向量,代表图像中所有候选区域的特征,通过 $\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}$ 计算两者的相似度,确定模型应将注意力分配给哪些区域;d_k 代表键向量的维度,$\sqrt{d_k}$为缩放因子,用于防止数值过大导致梯度消失;$\boldsymbol{V}$ 为值向量,包含了实际特征信息,经相似度加权后输出;$\odot$表示逐元素乘法,这意味着门控信号 $R(\boldsymbol{g},\tau)$ 会精准地对每一个注意力响应点进行按需分配增益;$\mathrm{Attn}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})$表示基础注意力算子。

该设计确保了门控增益被限制在合理的数值范围内，既增强了前景目标的特征激活，又维持了优化过程的数值稳定性。在网络作用路径上，DGA 模块嵌入于解码器内部的查询特征交互阶段，并不替代基础注意力计算，而是在注意力响应输出后由实例语义驱动的动态门控向量对响应进行实例级调制。经门控调制后的特征进一步参与查询更新与后续掩码预测，从而增强与当前芦笋实例一致的前景响应，抑制背景土壤、杂草及邻近植株遮挡所引起的无效响应。

与已有注意力重加权或特征门控方法相比，DGA 的主要差异体现在门控信号来源、作用位置及任务适配目标 3 个方面。现有方法通常依据全局池化特征或卷积特征图生成相对统一的通道权重或空间权重，其调制模式更多依赖整体统计信息。相比之下 DGA 的门控向量直接由解码器中的实例查询生成，因此能够随实例匹配过程中语义状态的变化而动态调整。在作用位置上，传统门控机制大多施加于主干网络或特征融合阶段的特征图，主要针对特征值进行增强或抑制，而 DGA 作用于解码器中的注意力响应，对实例查询与图像特征交互后的结果实施实例级调制。进一步地，考虑到本文面向的是单类别、细长形态且密集遮挡显著的芦笋实例分割任务，DGA 的设计重点并不在于提升一般意义上的全局表征能力，而在于强化与当前目标实例一致的前景响应，抑制邻近植株及复杂背景引起的特征混叠，从而提升遮挡场景下的实例区分能力与掩码连续性。因此，DGA 的核心并非对中间特征图进行统一重标定，而是针对实例查询与图像特征交互后的响应结果实施动态、实例级的语义调制。

2.3　边缘感知上下采样机制

不同于常规上采样仅在低分辨率粗掩码生成后通过直接插值恢复空间尺度，本文提出的边缘感知上下采样机制将实例语义约束、边界响应增强和高分辨率轮廓修正整合到统一的掩码生成流程中。在中间特征层中，空间特征与实例查询的条件融合使掩码响应在生成阶段即具备实例区分能力，从而减弱相邻芦笋在重叠区域中的响应混叠。在此基础上，结合掩码响应的局部变化构建边界注意力，对潜在边界区域进行强化，以维持细长茎秆轮廓的连续性。随后，在高分辨率恢复阶段引入边缘一致性校正，借助浅层边缘信息补偿双线性插值引起的轮廓平滑和边界偏移。因而，该机制的关键改进并不在于插值算子的替换，而在于通过语义约束、边界增强与轮廓修正的协同作用，提高掩码边界质量以及相邻重叠实例的分离效果。相较于常规上采样仅恢复空间尺度，本文方法在掩码生成阶段引入实例区分约束，并在边界增强与高分辨率校正阶段，提高遮挡边界与重叠区域的分割准确度。实现对芦笋掩码 M 的精确生成，即

$$M = \mathrm{Up}(\sigma(\boldsymbol{\Phi}_s \oplus \boldsymbol{\Phi}_q)) \tag{5}$$

式中，$\oplus$ 表示特征元素级交互操作，用以实现语义边界的对齐；σ 为非线性激活函数，用于增强特征的选择性表达；Up 是双线性插值算法，代表最终的上采样算子，旨在将掩码响应图恢复至原图分辨率；$\boldsymbol{\Phi}_s$ 为下采样增强后的空间特征，保留了芦笋茎秆轮廓与局部边缘几何信息；$\boldsymbol{\Phi}_q$ 为解码器输出的实例感知查询特征，承载了当前目标实例的语义与结构信息。二者的条件交互并非单纯的特征叠加，而是利用实例查询对空间响应进行目标相关筛选，保留属于当前实例的响应，而抑制与相邻重叠个体或背景区域相关的响应。由此，该掩码并不是普通意义上的低分辨率掩码生成，而是一个在实例语义约束下完成局部结构解耦的过程，这为后续上采样阶段保持边界清晰性和分离重叠区域奠定了基础。

为进一步增强模型在密集遮挡条件下的边界表达能力，在式(5)的掩码生成过程中引入轻量级边界注意力模块。该模块利用掩码响应图的局部变化特性，对空间位置进行自适应加权，使模型在生成实例掩码时更加关注目标边界区域。将式(5)中激活后的掩码响应表示为

$$P = \sigma(\boldsymbol{\Phi}_s \odot \boldsymbol{\Phi}_q) \tag{6}$$

其中，P 为低分辨率空间上的掩码响应。由于目标边界通常对应响应变化较为剧烈的位置，因此根据响应图的局部变化程度构建边界注意力权重，即

$$A = 1 + |\nabla P| \tag{7}$$

其中，A 为边界注意力权重，$|\nabla P|$ 表示 P 的空间变化幅度，用于表示潜在边界位置。随后将边界注意力权重作用于空间特征，实现对边界区域的轻量增强，即

$$\boldsymbol{\Phi}_s \leftarrow A \odot \boldsymbol{\Phi}_s \tag{8}$$

在密集遮挡场景中，相邻芦笋之间的真实分界通常表现为局部响应由一个实例向另一实例快速过渡的区域，这类区域在掩码响应图上同样对应较大的空间变化幅度。因此该轻量级边界注意力模块无需额外复杂网络结构，仅通过强化掩码响应的空间变化信息建模就可实现边界增强，使模型在遮挡区域能够更准确地恢复细长结构的连续轮廓，提高交界区域的对比度。

推理阶段，本研究仍采用双线性插值将低分辨率掩码响应恢复至原始图像尺寸，但该插值步骤仅承担尺度重建功能，为避免插值平滑导致的轮廓偏

移与交界模糊,在上采样后进一步引入边缘约束项以修正遮挡边界的模糊问题,提取上采样后掩码的边缘特征与骨干网络浅层特征的边缘信息进行余弦相似度对齐,对相似度低于阈值的边缘像素进行微调,降低遮挡边界与真实目标轮廓的偏差。

3 模型训练与评价指标

3.1 实验环境与参数设置

实验在 Linux x86_64 系统下进行,硬件平台为 Intel Xeon Platinum 8173M(16 核)CPU 与 NVIDIA GeForce RTX 3090 GPU,内存为 48 GB,显存为 23.57 GB。软件环境包括 PyTorch 2.8.0、Python 3.10.12、CUDA 12.8 及 cuDNN 9.10.2。训练策略继承 RF-DETR 的默认优化范式并结合本任务特点进行适配,采用带预热的余弦退火学习率调度以改善早期收敛的稳定性,并提升中后期的精细收敛能力。具体训练参数配置如表 1 所示。实例分割训练采用 RF-DETR 的集合预测与二分图匹配范式,启用指数滑动平均模型以获得更稳健的评估结果。

表 1 训练配置

Tab.1 Training configuration

参数	数值/型号
输入图像尺寸/(像素×像素)	384×384
优化器	AdamW
训练轮数	60
基础学习率	0.000 1
固定随机种子数	42
门控增益系数 alpha_s	1.0
门控增益系数 alpha_d	0.9
增益缩放系数 τ	0.2

针对提出的 DGA 改进,其关键超参数包括门控增益系数 alpha_s、alpha_d 以及增益缩放系数 τ。为提高方法的可复现性,本文对门控增益系数 alpha_s 与 alpha_d 进行了二维网格搜索实验。以 Mask mAP@0.5 与 Mask mAP@0.5:0.95 为实验评估指标,Mask mAP@0.5 表示在 IoU=0.5 的条件下,计算掩码的平均精度均值,Mask mAP@0.5:0.95 表示在 IoU 从 0.5 到 0.95(步长 0.05)多个阈值下计算 AP,然后取平均值。

如表 2 所示,当 alpha_s 取 1.0、alpha_d 取 0.9 时,模型在 Mask mAP@0.5 指标上取得最高值 84.9%,同时保持较高的 Mask mAP@0.5:0.95,因此选择该组合作为最终配置。当 alpha_s 或 alpha_d 进一步增大时,注意力增益过强会引入背景噪声,从而导致分割性能下降。

在确定 alpha_s 与 alpha_d 后,进一步对动态门控模块中的增益缩放系数 τ 进行单变量敏感性实验,设置 alpha_s=1.0、alpha_d=0.9,当 τ=0.2 时,模型的 Mask mAP@0.5:0.95 达到最高值 57.9%。当 τ 继续增大时,注意力权重波动增强,导致模型性能下降,见表 3。因此最终选取 0.2 作为默认参数。

表 2 alpha_s/alpha_d-二维网格搜索实验结果

Tab.2 alpha_s/alpha_d-two-dimensional grid search experiment %

alpha_s	alpha_d	Mask mAP@0.5	Mask mAP@0.5:0.95
1.0	0.8	82.0	53.1
1.0	0.9	84.3	56.8
1.0	1.0	84.9	51.8
1.1	0.8	84.3	54.0
1.1	0.9	78.9	47.4
1.1	1.0	81.4	50.0
1.2	0.8	83.1	52.4
1.2	0.9	84.4	50.7
1.2	1.0	81.1	39.4

表 3 τ-敏感性实验结果

Tab.3 τ-sensitivity experiment %

τ	Mask mAP@0.5	Mask mAP@0.5:0.95
0.1	79.7	55.2
0.2	81.8	57.9
0.3	78.0	52.1
0.4	80.8	57.2

3.2 评价指标

本研究的核心任务为复杂遮挡场景下的芦笋实例分割。考虑到实例分割结果同时依赖目标实例定位与像素级掩码预测,本文在评价体系中同时报告边界框层面的 Box 指标与掩码层面的 Mask 指标。其中,Box 指标用于表征模型对目标实例的定位、匹配及召回能力,Mask 指标用于衡量实例轮廓覆盖、边界贴合及重叠目标分离效果。

对检测性能的评价,采用精确率(Precision,P)与召回率(Recall,R)刻画预测正确性与覆盖完整性,其中精确率用于衡量预测为正的边界框中真正例所占比例,召回率用于衡量真实目标被成功检出的比例。

为全面评估模型在不同置信度阈值下的综合性能,采用平均精度(Average precision, AP)和平均精度均值(mean Average Precision, mAP)作为核心评价指标。

为进一步呈现模型性能,重点关注 mAP@0.5 与 mAP@0.5:0.95 两类指标。mAP@0.5 更侧重反映模型在相对宽松阈值下的实例匹配性能;mAP@0.5:0.95,更体现模型在严格定位条件下的

综合性能。

对于实例分割任务，采用与检测一致的评价框架但将交并比从边界框扩展到像素级掩码，分别计算 Mask 层面的精确率与召回率，输出 Mask mAP@0.5 与 Mask mAP@0.5:0.95，衡量模型在遮挡背景复杂条件下对实例轮廓覆盖与边界贴合的能力。

4　实验结果与分析

4.1　消融实验

为定量分析各改进模块对模型性能的独立贡献及协同作用，在统一实验设置下设计消融实验。以 RF－DETR 为基线，在不改变骨干网络与训练策略的前提下，分别引入边缘感知上下采样机制与动态门控注意力模块，并进一步构建二者联合模型。通过对不同模型配置在相同测试集上的性能进行对比，系统评估各模块在提升实例区分能力、边界建模能力及遮挡场景适应性方面的作用。

表 4 所示消融实验均建立在统一的 RF－DETR 基线框架之上，且各组实验均固定采用原始 DINOv2 骨干网络与一致的训练配置，因此表中性能差异可归因于本文新增的 DGA 模块与边缘感知上下采样机制，而非骨干网络替换所带来的收益。

表 4　消融实验结果

Tab.4　Ablation test results　　%

RF－DETR	上下采样	DGA 模块	边框精确率 P_{Box}	边框召回率 R_{Box}	Box mAP@0.5	Box mAP@0.5:0.95	掩码精确率 P_{Mask}	掩码召回率 R_{Mask}	Mask mAP@0.5	Mask mAP@0.5:0.95
√	×	×	95.58	88.35	92.05	69.30	86.27	58.81	84.42	49.62
√	√	×	94.24	86.21	91.60	69.55	90.92	66.68	87.76	58.45
√	×	√	94.62	89.49	93.20	72.26	91.90	67.01	89.69	59.52
√	√	√	95.47	88.79	93.30	71.43	91.60	67.24	90.98	59.13

注：√表示采用该改进方法，×表示不采用该改进方法。

基线模型掩码召回率仅为 58.81%，表明在遮挡场景下存在较为严重的漏检问题。引入边缘感知上下采样机制后，掩码召回率提升至 66.68%，说明该机制能够有效缓解遮挡交界区域的响应混叠问题，并增强细长目标边界的连续恢复能力。进一步结合 DGA 模块后，掩码精确率与召回率分别达到 91.60% 与 67.24%，表明改进的注意力机制能够有效抑制背景干扰，提升模型在复杂遮挡场景下的实例分割性能。

DGA 模块显著增强了实例响应判别能力，仅启用该模块时，边框召回率达 89.49%，掩码精确率达 91.90%，说明通过实例语义驱动的动态门控机制对查询特征进行动态加权后，模型能够有效抑制背景噪声干扰，提升目标特征的判别性，从而获得较高的掩码预测精度。当 DGA 模块与边缘感知上下采样机制协同工作时，模型在边框精确率、边框 mAP@0.5 及掩码召回率等指标上进一步提升，其中 P_{Box} 由 94.62% 提升至 95.47%，Mask mAP@0.5 从 89.69% 提升到 90.98%。这表明边缘感知上下采样机制有助于增强模型对细长茎秆轮廓及遮挡交界区域的细节恢复能力，使模型能够输出更加完整的目标掩码。

需要说明的是，与仅使用 DGA 模块相比，完整模型的掩码精确率 P_{Mask} 由 91.90% 小幅下降至 91.60%。结合模块设计目标分析，这一现象反映了模型在前景判别抑制噪声与边界细节恢复增强完整性之间存在权衡。DGA 模块更倾向于生成边界误检较少的掩码结果，而边缘感知上下采样机制在强化细长结构和遮挡边界恢复的同时，在掩码边界处进行优化，进而导致精确率指标小幅回落。

综合来看，完整模型在掩码指标上均取得显著提升。其中，在密集遮挡测试集下，模型的 Mask mAP@0.5 达到 90.98%，较基线模型提升 6.56 个百分点；Mask mAP@0.5:0.95 提升 9.51 个百分点，表明所提方法在提升分割精细度及遮挡场景适应能力方面具有明显优势。DGA 模块与边缘感知上下采样机制在功能上具有互补性，前者侧重提升目标查询的判别能力，后者侧重增强边界与细节建模能力。芦笋实例分割任务具有细长结构、密集分布及遮挡特点，完整模型 DGA－DETR 在保持较高掩码精度的同时，改善了目标轮廓恢复能力和掩码覆盖能力，因此更符合复杂田间场景下对分割完整性与识别稳定性的需求。

4.2　主流实例分割模型对比实验

为更加全面评估模型 DGA－DETR 在芦笋实例分割任务中的性能表现，并明确其在现有主流算法体系中的相对定位，在相同数据集与统一实验环境下，引入主流实例分割模型 Mask R－CNN、YOLACT、YOLO 11n－seg 和 YOLO 26n－seg 进行对比实验。所有模型均在相同数据集上进行训练与验证，并在统一评价指标体系下进行评估。模型对比结果如表 5 所示。

表 5 主流实例分割模型对比实验结果

Tab. 5 Comparison experiments results of mainstream instance segmentation models %

模型	边框精确率 P_{Box}	边框召回率 R_{Box}	Box mAP@0.5	Box mAP@0.5:0.95	掩码精确率 P_{Mask}	掩码召回率 R_{Mask}	Mask mAP@0.5	Mask mAP@0.5:0.95
YOLACT	51.81	62.90	82.53	47.61	43.35	52.20	80.60	43.35
Mask R-CNN	68.82	73.32	86.40	66.33	59.60	63.97	86.36	59.20
YOLO 11n-seg	90.10	86.80	90.60	71.40	90.40	67.60	90.12	52.70
YOLO 26n-seg	84.05	85.90	89.60	72.00	86.20	65.40	90.50	54.70
DGA-DETR	95.47	88.79	93.30	71.43	91.60	67.24	90.98	59.13

需要强调的是,本文的核心任务为芦笋实例分割,因此下述分析以掩码指标为主,检测指标仅作为实例定位质量的辅助评价。作为辅助定位指标,DGA-DETR 在 Box mAP@0.5 上达到 93.30%,表明模型能够为后续掩码生成提供较准确的实例匹配与定位先验。更关键的是,在核心实例分割指标上,DGA-DETR 取得了 90.98% 的 Mask mAP@0.5,优于其他模型,说明其在遮挡与重叠区域具有更优的轮廓覆盖能力与边界贴合能力,优于其他模型。在这关键指标上的突破,意味着 DGA-DETR 生成的分割掩码边缘更加精细、贴合目标轮廓,能够更好地应对芦笋生长环境中常见的遮挡与重叠挑战。

值得注意的是,尽管 Mask R-CNN 作为两阶段方法的代表,Mask mAP@0.5:0.95 为 59.20%,表现出较强的竞争力,但 DGA-DETR 也取得了 59.13% 的优异表现。在其他指标上,DGA-DETR 领先 Mask R-CNN。

综合来看,所提 DGA-DETR 模型在测试集上的掩码精确率、掩码召回率、Mask mAP@0.5 和 Mask mAP@0.5:0.95 分别达到 91.60%、67.24%、90.98% 和 59.13%,较基线模型分别提升 5.33、8.43、6.56、9.51 个百分点,表明该方法在分割完整性、边界精度及复杂遮挡场景下的实例识别能力方面均取得了显著提升。

4.3 模型性能定量分析

4.3.1 多任务损失演化与训练收敛稳定性

如图 5a 所示,RF-DETR 的 Mask mAP@0.5 在训练初期迅速收敛至 92%,但 Mask mAP@0.5:0.95 持续波动并在后期回落。这表明模型虽能快速定位目标,但对遮挡边界的精细分割能力易受背景噪声干扰,导致轮廓拟合不稳定。从损失曲线上看,总损失稳步下降,其中分类损失占主导且下降明显,而涉及掩码分割的损失项下降较为缓慢。这进一步印证,复杂遮挡背景下的像素级分割任务更具挑战性,模型的分割损失难以快速收敛。

图 5b 所示的 DGA-DETR 训练过程中,模型在早期 Mask mAP@0.5 快速上升之后,趋于平稳,波动

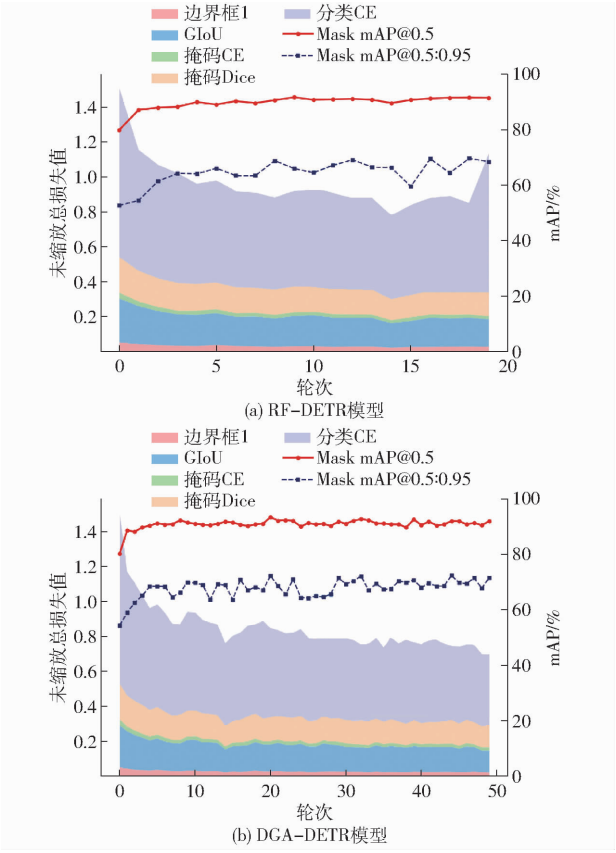


图 5 损失精度时序变化曲线

Fig. 5 Loss accuracy timing curves

范围小于 RF-DETR。并且 Mask mAP@0.5:0.95 曲线未出现 RF-DETR 中的回落现象。相较于 RF-DETR,DGA-DETR 的掩码损失收敛更稳定,且 Mask mAP@0.5:0.95 无回落现象,表明其对遮挡边界具有更鲁棒的拟合能力。分类损失虽仍占主导,但其下降过程更为平滑;掩码相关损失在整个训练中后期都保持低位且波动较小。表明 DGA-DETR 在各任务间实现了更优的优化平衡,不易因单一分支的梯度占优导致模型整体性能波动。

综合对比图 5a、5b 可知,与 RF-DETR 相比,DGA-DETR 的性能平台实现整体提升。DGA-DETR 在高阈值指标上的波动大幅减小,最低回落显著收窄。结合芦笋实例分割中细长目标易断裂、遮挡边界难辨别的任务特点,使得模型在实际应用中能够更稳定地输出精细、完整的分割轮廓。

4.3.2　损失指标相关性与瓶颈归因

在图 6a 所示的 RF－DETR 相关性分析中，模型在芦笋遮挡场景下的性能表现与各项损失呈现出明显的负相关，其中验证集 Mask mAP@0.5 与掩码损失、边界框损失的相关系数分别达到 －0.90 和 －0.89，模型整体性能的提升很大程度上取决于掩码质量和定位精度是否同步优化。通过分析 Mask mAP@0.5:0.95 指标，发现其与掩码损失的相关系数为 －0.82，明显高于分类损失的 －0.62，高 IoU 阈值下，模型对细长芦笋的实例掩码预测精度受限于掩码边界质量。枝叶遮挡导致的轮廓不连续会进一步放大该误差。此外，召回率与掩码损失、边界框损失的相关系数分别为 －0.63 和 －0.60，而与分类损失的相关系数只有 －0.42，进一步证实在遮挡导致的漏检中，起主导作用的是几何定位失败或掩码匹配不准，而非单纯的前景与背景判别失误。

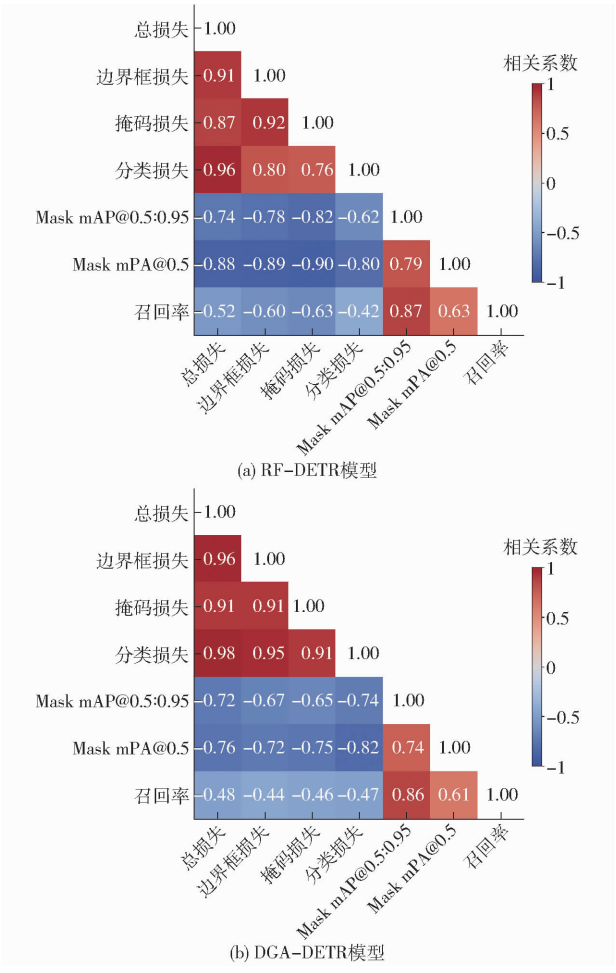


图 6　损失相关性热力图

Fig. 6　Loss correlation heatmap

图 6b 中 DGA－DETR 的相关性出现明显不同。尽管 Mask mAP@0.5 与各项损失呈现负相关，但其与分类损失的相关系数升至 －0.82，超过了与掩码损失的 －0.75 和边界框损失的 －0.72。表明构建 DGA 机制后，模型对遮挡芦笋的语义判别能力成为影响整体性能的主要因素，而掩码质量对模型性能的影响有所减弱。对于 Mask mAP@0.5:0.95 指标，DGA－DETR 与掩码损失的相关系数只有 －0.65，相比于 RF－DETR 的 －0.82 显著下降（负号代表负相关，不代表大小），说明模型在高阈值下的精细分割能力对掩码误差的敏感度降低，其性能上限更多由整体表征质量与实例可见性判别能力共同决定。与此同时，召回率与各项损失的相关系数分布更为均衡，均维持在 －0.44 ~ －0.47 之间，尤其是与分类损失的相关系数从 －0.47 提升至 －0.63，表明在有效抑制背景干扰后，漏检主要源于遮挡导致的目标语义可见性不足，而非掩码断裂或定位偏移。

综上所述，RF－DETR 在遮挡芦笋实例分割中的主要限制来自掩码质量与定位精度，而 DGA－DETR 通过动态门控增强有效查询、抑制冗余信息，缓解了掩码瓶颈，将性能制约因素部分转移至分类表征与任务一致性上。模型误差从明显的边界形状错误转向隐蔽的语义可见性不足，体现出所提模型的优越性。

4.4　模型实例分割定性分析

4.4.1　掩码可视化

在同域实验设定下，通过 3 组典型场景验证了遮挡程度与背景复杂度对芦笋实例分割模型性能的影响。图 7 展示了同域低分辨率、密集遮挡以及复杂重叠场景下的基线模型与改进模型的掩码可视化对比图。在图 7a 低分辨率场景中，DGA－DETR 对边缘芦笋的平均置信度 0.87 高于 RF－DETR 的 0.73，且标准差更小，表明其稳定性更好。在图 7b 密集遮挡场景下，基线模型漏检率达 23.5%，而 DGA－DETR 将漏检率降至 8.7%，同时掩码 IoU 提升 12.6 个百分点。图 7c 复杂重叠场景中，当遮挡率大于 60% 时，DGA－DETR 仍保持 76.3% 的有效实例识别率，较基线模型 52.1% 的有效实例识别率

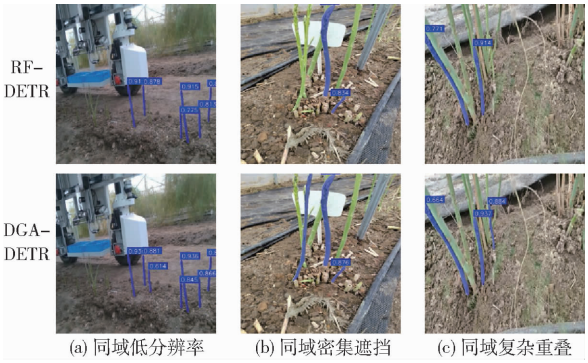
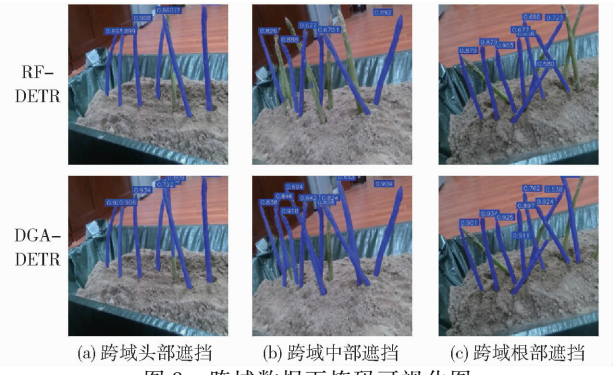


图 7　同域数据下掩码可视化图

Fig. 7　Mask visualization diagrams for data in the same domain

显著提升。实验表明,同域中密集遮挡与复杂重叠场景显著增加识别难度,DGA - DETR 在缓解遮挡与抑制噪声方面表现更优。

在跨域场景下对芦笋实例分割算法开展性能测试,如图 8 所示,设置跨域头部、中部及根部遮挡 3 种典型干扰场景以模拟田间复杂采收环境。如图 8a 所示,在跨域头部遮挡场景中,两类模型的单株分割精度均表现较好,而 DGA - DETR 对边缘目标的置信度波动更小,特征锚定稳定性优于 RF - DETR。如图 8b、8c 所示,针对跨域中部和根部遮挡情况,DGA - DETR 对遮挡目标的置信度高于 RF - DETR,并且漏检情况优于 RF - DETR 模型。通过跨域场景下实验分析,DGA - DETR 不仅在置信度稳定性上更优,对于边缘目标、密集遮挡目标的漏检率低于 RF - DETR,验证了其在跨域场景下的跨域适应能力。



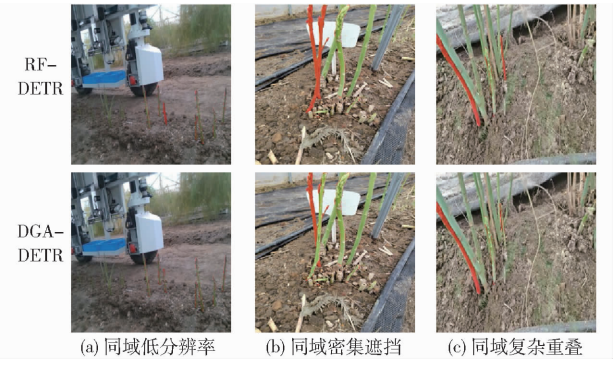
(a) 跨域头部遮挡 (b) 跨域中部遮挡 (c) 跨域根部遮挡

图 8 跨域数据下掩码可视化图

Fig. 8 Masked visualization charts for cross-domain data

4.4.2 缺失掩码区域可视化

图 9 展示了同域场景下的芦笋实例分割缺失掩码区域可视化对比结果,其中红色区域代表模型在实例分割过程中未能正确生成的目标掩码像素范围。两类模型在同域场景虽然都有不同程度的缺失掩码,但不论是同域低分辨率、密集遮挡还是复杂重叠环境下,RF - DETR 模型的红色区域明显多于 DGA - DETR 模型,表明 DGA - DETR 模型有效减少了漏检现象,且掩码缺失范围显著小于 RF - DETR 模型,在同域场景下的分割完整性与稳定性更优,对实际芦笋采摘场景的适配性更强。

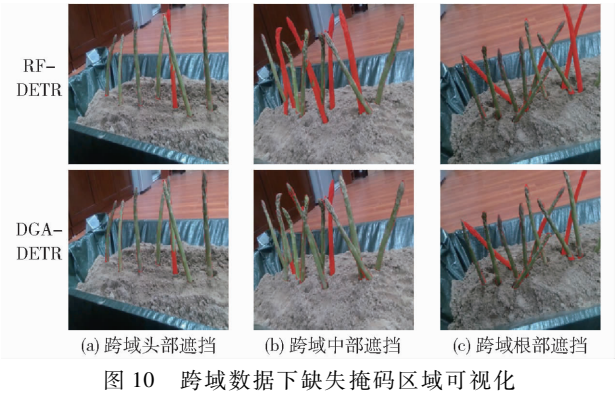


(a) 同域低分辨率 (b) 同域密集遮挡 (c) 同域复杂重叠

图 9 同域数据下缺失掩码区域可视化

Fig. 9 Visualization of missing mask regions in intra-domain data

图 10 展示了跨域场景下芦笋头部、中部以及根部实例分割的缺失掩码区域可视化结果。综合两类模型在跨域场景下的掩码缺失情况分析可知,RF - DETR 部分茎秆的头部出现掩码缺失,甚至存在整株芦笋完全漏检的情况,然而 DGA - DETR 模型丢失范围较小,无明显漏检。针对跨域中部及根部遮挡场景下的分析可知,RF - DETR 模型出现明显大面积丢失,甚至出现多株漏检现象。而 DGA - DETR 模型掩码缺失范围与漏检现象均明显得到改善。尽管受跨域场景特征分布偏移、芦笋茎秆细长等技术难点限制,DGA - DETR 通过改进的特征关联与背景抑制机制,仍有效提升了跨域遮挡场景下的分割稳定性。



(a) 跨域头部遮挡 (b) 跨域中部遮挡 (c) 跨域根部遮挡

图 10 跨域数据下缺失掩码区域可视化

Fig. 10 Visualization of missing mask regions in cross-domain data

4.4.3 注意力特征热力图

同域场景下的 3 组注意力热力图显示,两种模型在低分辨率、密集遮挡与复杂重叠场景下对芦笋目标的注意力响应存在显著差异。图 11a 显示,RF - DETR 在背景土壤区域出现注意力扩散,而 DGA - DETR 的注意力热力图聚焦性更强,对目标的特征锚定精度更高。图 11b 显示,DGA - DETR 模型对于被遮挡茎秆的红色区域更集中、颜色更深,直观表明 DGA - DETR 模型可以维持更好的轮廓连续性。图 11c 显示,在复杂重叠环境下,DGA - DETR 对重叠部位的芦笋特征仍能形成清晰的热力响应边界,有效区分相邻重叠的芦笋植株,而 RF - DETR 在重叠核心区的热力响应出现轻微融合。上述结果验证了 DGA - DETR 在同域场景下的有效性,图 11 也表明其在特征聚焦、遮挡抑制及重叠实例区分方面均有优化。

跨域场景下的芦笋实例分割面临着场景迁移带来的特征分布偏移问题,图 12 直观呈现了两类模型

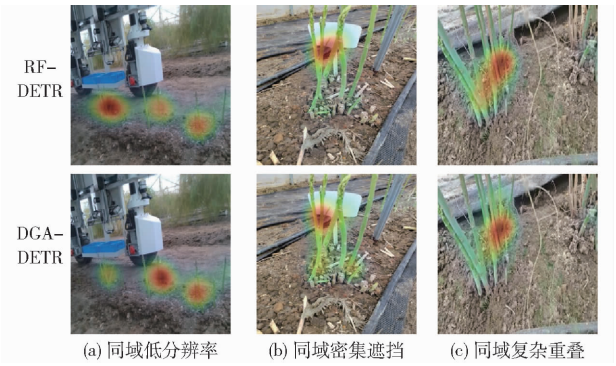


图 11　同域数据下注意力特征热力图

Fig. 11　Attention feature heatmaps for intra-domain data

在芦笋头部、中部及根部受遮挡情况下的响应特性。如图 12a 所示，虽然两类模型的热力响应均集中在芦笋茎秆头部，并且形成了连续且强烈的特征锚定，但 DGA－DETR 的热力响应范围更加集中，不曾出现扩散的情况。图 12b 显示，DGA－DETR 模型对植株交织导致的中部遮挡仍能维持一定的注意力强度。图 12c 显示，当出现芦笋根部遮挡情况，DGA－DETR 的热力响应范围更宽，并且特征覆盖更全面，而 RF－DETR 的响应只集中于头部和中部区域，对根部的关注度较低。图 12 揭示了模型在跨域遮挡下的注意力分配规律，并通过对比凸显了 DGA－DETR 在遮挡场景下的特征感知优势，为芦笋采摘机器人算法的跨域鲁棒性优化提供可量化的视觉依据。

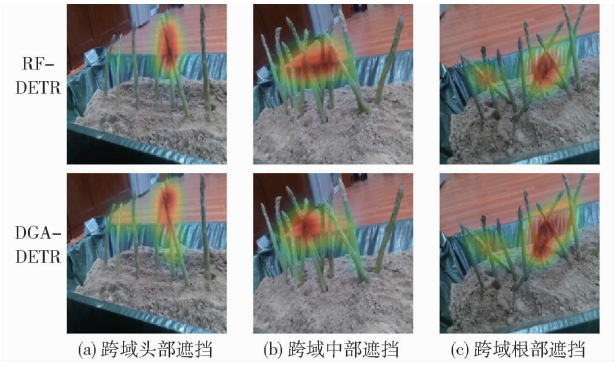


图 12　跨域数据下注意力特征热力图

Fig. 12　Attention feature heatmaps for cross-domain data

在此基础上，进一步分析模型在极端工况下的性能瓶颈。在遮挡率大于 80% 的极端密集遮挡场景中，芦笋茎秆有效可见区域的严重缺失导致目标全局几何特征与局部纹理特征大幅退化，是模型性能下降的核心诱因。该场景下，即使本文设计的边缘感知上下采样机制可强化边界特征提取，仍难以通过碎片化的局部区域还原茎秆完整的空间拓扑关系，易出现掩码断裂与实例漏检，这与本文遮挡率大于 60% 场景下的实验规律一致，同时，DGA 模块虽可有效抑制背景噪声，但对多实例的特征解耦能力仍存在局限，是极端遮挡场景下模型性能提升的核心瓶颈。在低光照、强逆光的极端光照场景中，图像信噪比降低导致芦笋与背景的边界特征模糊，局部光学不稳定引发的特征丢失进一步加剧了前景-背景的特征混淆，是模型精度退化的主要原因。针对上述极限工况，后续可通过引入芦笋茎秆形态先验约束、极端样本对抗数据增强、时序帧间特征融合等方式，进一步提升模型的极端环境适应能力。

4.5　实际工程效率评估与应用挑战

为验证 DGA－DETR 模型在芦笋自动采收机器人端的工程应用潜力，本研究对其计算复杂度和推理效率进行了量化评估。实验环境基于 NVIDIA GeForce RTX 3090 GPU，输入图像尺寸统一调整为 384 像素 ×384 像素。统计指标涵盖模型参数量、推理速度、推理延迟时间以及显存占用量，对比结果如表 6 所示。

表 6　模型复杂度与推理效率对比

Tab. 6　Comparison of model complexity and inference efficiency

模型	参数量	推理速度/(f·s^{-1})	推理延迟时间/ms	显存占用量/MB
RF－DETR	3.569×10^7	31.68	31.57	342.37
DGA－DETR	3.577×10^7	30.99	32.27	342.57
绝对变化量	8×10^4	－0.69	0.70	0.20
相对变化率/%	0.22	－2.18	2.22	0.06

实验数据表明，相较于基线模型 RF－DETR，DGA－DETR 在引入 DGA 模块后，参数量仅由 3.569×10^7 微增至 3.577×10^7，增幅仅为 0.22%；显存占用量几乎持平，增幅仅为 0.06%。在关键的实时性指标方面，DGA－DETR 的推理速度维持在 30.99 f/s，虽然较基准模型 31.68 f/s 有 －2.18% 微小波动，但推理延迟时间增加量严格控制在 0.70 ms 以内。

综上所述，DGA 模块以极微小的计算开销显著提升了感知精度。鉴于该模型的推理速度远超采收机器人实时作业所需的阈值，DGA－DETR 在嵌入式边缘计算平台或移动端设备上具有极高的部署可行性与实际应用价值。

但在实际应用中仍面临一些挑战，如像素掩码向三维作业空间的映射、田间动态作业的鲁棒性与嵌入式边缘部署的实时性等挑战。本模型输出的高精度实例分割掩码是定位的核心。在实际集成中，需利用相机标定参数，将像素坐标系下的掩码质心或生长轴线，结合深度感知信息映射至机器人基坐标系。芦笋细长且易损的生物特性要求投影变换过程具备极高的空间一致性，以确保末端执行的精准

切入。

实地作业中,采收车的振动以及田间光照的剧烈变化可能导致图像运动模糊或特征淹没。虽然 DGA 模块通过动态门控增强了特征抗噪能力,但在系统层面仍建议引入时序滤波或多帧融合策略,以平滑单帧分割产生的跳变,确保机械臂运动轨迹的平顺性。

针对芦笋采收机器人嵌入式平台的边缘部署需求,模型轻量化是本算法后续工程化落地的核心研究方向。本文模型复杂度评估结果表明,DGA - DETR 相较于基线模型仅增加 0.22% 的参数量与 0.06% 的显存占用量,且推理速度满足采收作业的实时性要求,具备良好的轻量化部署基础。后续将通过结构化剪枝与通道稀疏化训练方式压缩模型参数量,同时可以构建知识蒸馏框架,以本文 DGA - DETR 为教师模型、轻量型分割模型为学生模型,实现密集遮挡场景下特征表征能力的迁移,在保障模型核心精度指标的前提下,实现推理速度与模型参数量的平衡。

5 结论

(1)针对芦笋机器人采收中因植株细长形态、密集遮挡导致的掩码断裂与边缘失真等难题,本文提出 DGA - DETR 芦笋实例分割算法,设计 DGA 模块在解码阶段对实例查询相关的注意力响应进行动态门控调制,有效增强了前景实例表征并抑制了复杂背景干扰;构建边缘感知上下采样机制强化了对细长茎秆及遮挡边界的细粒度建模,显著提升了分割精度。

(2)通过实验对比分析,确定了 DGA - DETR 算法的最优配置参数及其在复杂工况下的性能界限。实验结果表明,在密集遮挡测试集下,模型对芦笋实例的掩码预测 mAP@0.5 达到 90.98%,较基准模型提升了 6.56 个百分点;反映更高分割精细度的 mAP@0.5:0.95 指标提升了 9.51 个百分点。这一显著提升表明了模型在复杂空间拓扑关系下的性能界限突破,为实现芦笋个体的精准解耦与边界建模提供了支撑。

(3)跨域泛化性验证与消融实验表明,DGA - DETR 在像素、空间及语义三层次域偏移下均保持了稳定的感知能力。可视化分析证实,其注意力响应分布更契合芦笋生物形态特征,边界表征具有极强的连贯性。研究成果可为复杂田间环境下农业机器人的视觉导航、避障及精准采收提供关键技术支撑。

参 考 文 献

[1] 汪小旦,李为民,王琳,等. 基于改进 YOLACT + + 的成熟芦笋检测-判别-定位方法[J]. 农业机械学报, 2023, 54 (7): 259 - 271.
Wang Xiaochan, Li Weimin, Wang Lin, et al. Method of detection-discrimination-localization for mature asparagus based on improved YOLACT + + algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(7): 259 - 271. (in Chinese)

[2] 张萍. 白芦笋选择性采收机器人关键技术研究[D]. 泰安: 山东农业大学, 2024.
Zhang Ping. Research on key technology of selective harvesting robot for white asparagus[D]. Tai'an: Shandong Agricultural University, 2024. (in Chinese)

[3] 吴擎,韦润轩,周乐,等. 基于改进 YOLOv5s 的轻量化鲜茶叶识别方法[J]. 智能化农业装备学报(中英文), 2025, 6(1): 1 - 14.
Wu Qing, Wei Runxuan, Zhou Le, et al. Lightweight fresh tea leaf recognition method based on improved YOLOv5s[J]. Journal of Intelligent Agricultural Mechanization, 2025, 6(1): 1 - 14. (in Chinese)

[4] 刘莫尘, 褚镇源, 崔明诗, 等. 基于改进 YOLO v8 - Pose 的红熟期草莓识别和果柄检测[J]. 农业机械学报, 2023, 54 (增刊 2): 244 - 251.
Liu Mochen, Chu Zhenyuan, Cui Mingshi, et al. Red ripe strawberry recognition and stem detection based on improved YOLO v8 - Pose[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(S2): 244 - 251. (in Chinese)

[5] 张林鎧, 巴音塔娜, 曾庆松. 基于 StyleGAN2 - ADA 和改进 YOLO v7 的葡萄叶片早期病害检测方法[J]. 农业机械学报, 2024, 55(1): 241 - 252.
Zhang Linxuan, Ba Yintana, Zeng Qingsong. Grape disease detection method based on StyleGAN2 - ADA and improved YOLO v7[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(1): 241 - 252. (in Chinese)

[6] Cui J, Zhang X, Zhang J, et al. Weed identification in soybean seedling stage based on UAV images and Faster R - CNN[J]. Computers and Electronics in Agriculture, 2024, 227: 109533.

[7] Xu D W, Huang X P, Zhao Z H, et al. Pha - YOLO: a multi-view Phalaenopsis flower detection and counting method based on improved YOLO11 with dynamic reference selection and adaptive thresholding[J]. Computers and Electronics in Agriculture, 2026, 243: 111446.

[8] Shen Q, Zhang X, Shen M, et al. Multi-scale adaptive YOLO for instance segmentation of grape pedicels[J]. Computers and Electronics in Agriculture, 2025, 229: 109712.

[9] 黄华,张昊,胡晓林,等. 基于改进 YOLO v5 的复杂环境下花椒簇识别与定位方法[J]. 农业机械学报, 2024, 55(3):

243－251.
Huang Hua, Zhang Hao, Hu Xiaolin, et al. Recognition and localization method for peper clusters in complex environments based on improved YOLO v5[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(3): 243－251. (in Chinese)
[10] 张立杰,周舒骅,李娜.基于改进 SSD 卷积神经网络的苹果定位与分级方法[J]. 农业机械学报,2023,54(6):223－232.
Zhang Lijie, Zhou Shuhua, Li Na. Positioning and grading method of apples based on improved SSD convolutional neural network[J]. Transactions of the Chinese Society for Agricultural Machinery,2023,54(6):223－232. (in Chinese)
[11] 翟肇裕,张梓涵,徐焕良,等. YOLO 算法在动植物表型研究中应用综述[J]. 农业机械学报, 2024, 55(11): 1－20.
Zhai Zhaoyu, Zhang Zihan, Xu Huanliang, et al. Review of applying YOLO family algorithms to analyze animal and plant phenotype[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(11): 1－20. (in Chinese)
[12] Jiang K L, Hu W J, Han P. Multi-task segmentation of oilseed rape and weeds in agricultural fields: a hybrid approach combining enhanced UNet and unsupervised clustering[J]. Computers and Electronics in Agriculture,2025,238:110827.
[13] Ding Y, Yang W, Zhang J. An improved DeepLabV3 + based approach for disease spot segmentation on apple leaves[J]. Computers and Electronics in Agriculture,2025,231:110041.
[14] 张勤,庞月生,李彬. 基于实例分割的番茄串视觉定位与采摘姿态估算方法[J]. 农业机械学报, 2023, 54 (10): 205－215.
Zhang Qin, Pang Yuesheng, Li Bin. Visual positioning and picking pose estimation of tomato clusters based on instance segmentation[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(10): 205－215. (in Chinese)
[15] Li Z, Wang Z, Zhang H, et al. Light－AdaPSPNet: a novel lightweight adaptive weighted semantic segmentation network based on PSPNet for extracting *Spartina alterniflora* from dual-polarimetric SAR images[J]. Computers and Electronics in Agriculture,2026,242:111338.
[16] Meng K, Xu K, Cattani P, et al. Camouflaged cotton bollworm instance segmentation based on PVT and Mask R－CNN[J]. Computers and Electronics in Agriculture,2024,226:109450.
[17] Li Z, Shen Y, Tang J, et al. IMLL－DETR: an intelligent model for detecting multi-scale litchi leaf diseases and pests in complex agricultural environments[J]. Expert Systems with Applications,2025,273:126816.
[18] Zhao Z, Chen S, Ge Y, et al. RT－DETR－Tomato: tomato target detection algorithm based on improved RT－DETR for agricultural ssafety production[J]. Applied Sciences,2024,14(14):6287.
[19] Zhou H, Jiang F, Lu H. SSDA－YOLO: semi-supervised domain adaptive YOLO for cross-domain object detection[J]. Computer Vision and Image Understanding, 2023, 229: 103649.
[20] Oquab M, Darcet T, Moutakanni T, et al. Dinov2: learning robust visual features without supervision[PP/OL]. V1. arXiv (2023－04－14)[2026－03－08]. https://arXiv.org/abs/2304.07193.
[21] Qiu Z, Wang Z, Zheng B, et al. Gated attention for large language models: non-linearity, sparsity, and attention-ssink-free [PP/OL]. V1. arXiv (2025－05－10)[2026－03－08]. https://arXiv.org/abs/2505.06708.

(上接第 165 页)

[13] 刘文政, 周雪健, 平凤娇, 等. 基于可见-近红外光谱的鲜食葡萄成熟品质关键指标检测[J]. 农业机械学报, 2024, 55(2): 372－383.
Liu Wenzheng, Zhou Xuejian, Ping Fengjiao, et al. Detection of key maturity quality indicators of grapes based on visible-near infrared spectroscopy[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(2): 372－383. (in Chinese)
[14] Zou H, Lan T, Jiang Y, et al. Research on rapid detection methods of tea pigments content during rolling of black tea based on machine vision technology[J]. Foods, 2024, 13(23): 3718.
[15] Xu Z, Han Y, Zhao D, et al. Research progress on quality detection of livestock and poultry meat based on machine vision, hyperspectral and multi-source information fusion technologies[J]. Foods, 2024, 13(3): 445.
[16] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 770－778.
[17] Tan M, Le Q V. EfficientNet: rethinking model scaling for convolutional neural networks[C]// Proceedings of the 36th International Conference on Machine Learning. Cambridge, U.S.A.: PMLR, 2019: 6105－6114.
[18] Liu Z, Lin Y, Cao Y, et al. Swin Transformer: hierarchical vision transformer using shifted windows[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2021: 10012－10022.
[19] Jiao J, Tang Y M, Lin K Y, et al. DilateFormer: multi-scale dilated transformer for visual recognition[J]. IEEE Transactions on Multimedia, 2023,25:8906－8919.
[20] Hatamizadeh A, Kautz J. MambaVision: a hybrid Mamba-Transformer Vision backbone[C]// Proceedings of the Computer Vision and Pattern Recognition Conference. Piscataway: IEEE, 2025: 25261－25270.