基于改进RT-DETR的棉花成熟度检测算法

史绮梦，汪 军*，徐晓峰，张玮逸

（安徽工程大学计算机与信息学院，安徽芜湖 241000，中国）

摘 要：［**目的/意义**］为满足棉花不同生育阶段对水肥需求差异化调控的需求，实现棉花生长过程的精细化管理，本研究提出一种高精确率、轻量化的棉花成熟度视觉检测算法，为田间管理提供技术支持。［**方法**］基于实时检测Transformer模型，构建棉花成熟度检测模型（Cotton Maturity-Detection Transformer, CM-DETR）。首先，在主干网络中引入重参数化分组卷积空间增强轻量注意力网络模块，融合渐进式卷积与可重参数化卷积，平衡模型轻量化与特征表达能力；其次，构建深度鲁棒特征下采样模块，通过多尺度特征融合增强小目标及细微成熟特征感知；最后，采用聚焦式完整交并比损失函数，提高样本不平衡条件下的定位精度。［**结果和讨论**］CM-DETR模型在平均精度阈值为0.5时达到了80.8%、平均精度在阈值为0.5到0.95时达到51.1%，相比于基准模型分别提升3.7和1.8个百分点，同时模型参数量和计算量分别下降了31.7%和22.8%。［**结论**］研究表明，CM-DETR兼顾高精度与轻量化特性，可为棉花生产过程中的水肥精准调控与最佳采收时机判定提供可靠的技术支持。

关键词：棉花成熟度；RT-DETR；目标检测；RGCSPELAN；DRFD；Focaler-CIoU损失函数

中图分类号：TP391 **文献标识码：**A **文章编号：**SA202512013

引用格式：史绮梦, 汪军, 徐晓峰, 张玮逸. 基于改进RT-DETR的棉花成熟度检测算法[J]. 智慧农业(中英文), 2026, 8(4): 100-113. DOI：10.12133/j.smartag.SA202512013

SHI Qimeng, WANG Jun, XU Xiaofeng, ZHANG Weiyi. Cotton Maturity Detection Algorithm Based on Improved RT-DETR[J]. Smart Agriculture, 2026, 8(4): 100-113. DOI：10.12133/j.smartag.SA202512013 (in Chinese with English abstract)

0 引 言

棉花作为重要的经济作物之一，其在纺织、农业经济中占据着不可替代的地位[1]。棉花的产量与品质高度依赖于生长阶段的精细化管理，不同阶段对水、肥的需求迥异[2-4]。传统棉花成熟度监测主要依赖人工田间巡视与经验判断，此方法不仅劳动强度大、效率低下，且易因主观误差导致管理决策失当。随着农业规模化生产的推进和劳动力成本的攀升，传统模式已无法满足现代精准农业的需求[5-7]，因此，研发一种高效、准确的自动化棉花成熟度检测技术是非常重要的。

近年来，深度学习技术在图像识别和模式分类等领域展现出卓越性能，推动了卷积神经网络（Convolutional Neural Network, CNN）等模型的广泛应用[8]。随着深度学习技术的持续发展，CNN不仅在计算机视觉领域取得了重要突破，在农业图像感知领域也日益深入，尤其在农作物识别与害虫检测中展现出巨大的潜力[9, 10]。对于经济作物病害检测问题，国内外学者开展了研究探索。例如，XUE等[11]针对茶树叶片病虫害的检测任务，对YOLOv5（You Only Look Once）模型进行了改进，在一定程度上提升了检测精度，但在复杂背景的适应能力仍有限。KINDA等[12]在YOLOv8中引入实时检测Transformer模型（Real-Time Detection Transformer, RT-DETR）主干网络用于航空影像中的棉花病害检测，提升了检测效果，但模型的计算开销大。针对棉花果实相关任务，TAN等[13]提出了一种结合Transformer与多目标跟踪的棉铃计数方法，利用RT-DETR实现逐帧检测和边界框定位，有效避免重复计数，提升计数任务的准确性，但在

收稿日期：2025-12-12

基金项目：国家自然科学基金（62406004）；安徽高校协同创新项目（GXXT-2019-020）

作者简介：史绮梦，硕士研究生，研究方向为深度学习与目标检测，E-mail：2240932105@stu.ahpu.edu.cn

*通信作者：汪 军，教授，研究方向为智能感知、图像处理与模式识别，E-mail：wangjun@ahpu.edu.cn

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

复杂背景下对小目标特征的识别仍不足。YADAV等[14]和彭炫等[15]分别针对自生棉花植株和棉花顶芽检测对YOLOv5模型进行改进，在特定场景下提升了检测精度，但在复杂田间环境中的小目标识别及多阶段特征识别方面仍存在局限。VERMA等[16]提出了一种基于YOLOv8的改进算法进行棉花生长监测，部分阶段识别精度有所提升，但整体性能仍有进一步优化的空间。

目前，目标检测算法已从基于区域的卷积神经网络（Region-based Convolutional Neural Network, R-CNN）[17, 18]和YOLO[19, 20]系列为代表的经典模型，逐步发展为基于Transformer的新范式，其中RT-DETR[21, 22]作为一种典型的端到端检测模型，摒弃了传统CNN模型（如YOLO系列）所依赖的非极大值抑制（Non-Maximum Suppression, NMS）后处理机制，通过全局建模提升了对小目标和遮挡目标的检测鲁棒性，并减少人工阈值设置带来的不确定性。然而，在棉花成熟度检测任务中，RT-DETR的主干网络结构较复杂、模块间特征交互受限，导致模型对不同生育阶段棉花细微纹理差异和遮挡小目标的特征提取不足，进而影响在复杂田间环境下的检测稳定性和多阶段特征区分能力。

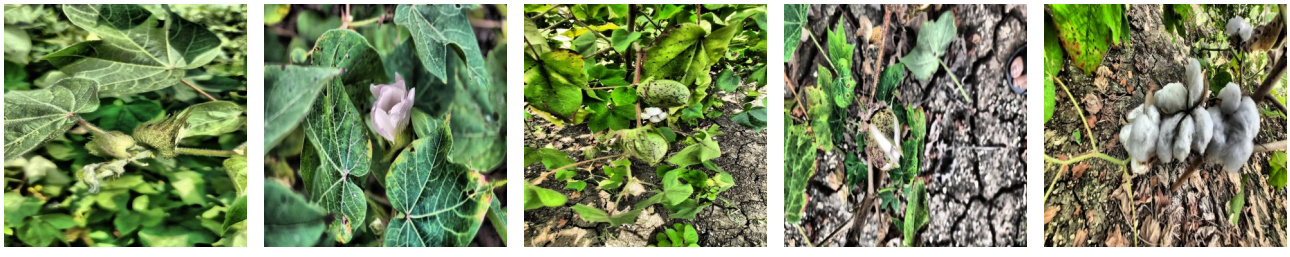
综上所述，尽管CNN和YOLO系列在检测精度和速度上取得显著优势，但在小目标、遮挡及多阶段特征识别上仍存在不足。RT-DETR虽在端到端检测上具有优势，但主干网络存在特征冗余、多尺度特征提取不足，以及模型稳定性不足的问题仍未解决。因此，本研究提出一种基于RT-DETR改进的棉花成熟度检测模型（Cotton Maturity-Detection Transformer, CM-DETR），主要贡献包括：1）针对主干网络特征冗余与计算复杂度较高的问题，引入重参数化分组卷积空间增强轻量注意力网络模块（Re-parameterized Group Convolution Spatial Enhancement Lightweight Attention Network, RGCS-PELAN），通过通道划分与深度可分离卷积降低模型参数量与计算复杂度，并结合RepConv可重参数化多分支结构在保证推理效率的同时增强特征表达能力，实现轻量化与检测性能的有效平衡；2）针对图像中低分辨率（目标边界模糊、纹理细节不足）及小目标特征（像素面积小于32×32的花蕾期及早铃期棉铃）丢失问题，采用DRFD下采样模块以增强多尺度特征融合能力，有效提升小目标与细微特征（边缘轮廓与内部纹理）的检测性能；3）为优化边界框预测的精度与稳定性，引入聚焦式完整交并比损失函数（Focaler Complete Intersection over Union, Focaler-CIoU），通过样本平衡、几何约束和动态聚焦机制，优化边界框回归精度并提升整体检测性能。最后，通过试验对所提出的CM-DETR模型的性能进行验证。

1　材料与方法

1.1　数据集获取

本研究采用的数据集主要来自公开数据集Maturity analysis，该数据集面向棉花成熟度分析任务构建，原始图像采集于实际棉花种植区域的自然农田环境，覆盖棉花生长周期内的多个生育阶段，拍摄条件为自然光照下的多视角场景，图像背景复杂，能够较真实地反映田间环境下棉花成熟度的视觉特征。

在数据预处理阶段，对原始数据进行了系统的清洗，去除了重复样本及标注不完整的图像，以保证数据质量。并在此基础上重新划分数据集比例，确保试验设置的合理性与可重复性。根据棉花生育期特征，将数据集划分为花蕾期、棉花花期、早铃棉期、裂棉铃期和成熟棉铃期5个类别，典型图像实例见图1所示。



a. 花蕾期　b. 棉花花期　c. 早铃棉期　d. 裂棉铃期　e. 成熟棉铃期

图1　不同阶段的棉花图像

Fig. 1　Images of cotton at different stages

1.2 数据集处理

Maturity analysis 数据集中的图像均统一缩放至640×640像素的固定尺寸，并以8∶1∶1的比例划分为训练集、验证集和测试集，其中包括1 861个训练样本、232个验证样本和234个测试样本。为了增强模型的泛化能力和鲁棒性，对训练集使用数据增强技术，包括随机旋转、缩放、颜色调整等方法进行扩充，经过数据清洗最终得到2 327张图像。棉花在不同阶段的成熟度类别数量如表1所示。

表1 棉花成熟度数据集

Table 1 The dataset of cotton maturity

棉花成熟度类别	图像数量/张
花蕾期	1 378
棉花花期	602
早铃棉期	3 667
裂棉铃期	400
成熟棉铃期	2 757

注:一张图片中可能包括多个类别,表中统计数量为类别的图片数量,不同于数据集图像数。

2 模型改进

2.1 基线模型RT-DETR

RT-DETR是一种基于Transformer架构的实时目标检测模型，主要由主干网络（Backbone）、高效混合编码器（Hybrid Encoder）和带辅助预测头的解码器（Decoder）三部分组成，该模型在保持较高检测精度的同时，提升了计算效率，适用于实时目标检测任务。模型中的高效混合编码器采用基于注意力的尺度内特征交互模块（Attention-based Intra-scale Feature Interaction, AIFI）和基于卷积的跨尺度特征融合模块（CNN-based Cross-scale Feature Fusion, CCFF），其中AIFI模块类似于Transformer的编码器，负责对S5特征进行自注意力编码，增强模型的特征表达能力；CCFF模块基于卷积神经网络结构，通过在特征融合路径中引入由多个卷积层构成的融合单元，显著提升多尺度特征的整合能力。解码器采用交并比（Intersection over Union, IoU）感知查询选择机制（IoU-aware Query Selection），该机制通过对编码器特征的联合潜在变量进行建模，优化初始查询质量，从而有效抑制冗余预测，提升目标定位的准确性。RT-DETR通过上述结构设计，在保证实时性的同时，实现了较高的检测性能，适用于农业作物监测（如棉花成熟度检测）等实际应用场景。

2.2 改进模型CM-DETR

本研究针对RT-DETR模型在棉花成熟度检测任务中存在的计算冗余、小目标特征易丢失，以及边界框回归精度不足等问题，提出了改进的棉花成熟度检测模型CM-DETR。

首先，将渐进式卷积和RGCSPELAN模块精巧结合构建一个轻量化的主干网络，并采用可重参数化卷积RepConv（Re-parameterizable Convolution）[23]替代标准残差块，以强化梯度流通和特征提取能力；其次，在颈部网络中引入下采样模块DRFD，该模块借助多路径结构与鲁棒特征提取策略，提升特征表达的稳定性，并且能在保留关键信息的同时抑制背景噪声干扰，从而增强模型在复杂背景和尺度变化场景下对小目标（如花蕾、棉铃等）的检测精度与鲁棒性；最后，为进一步提升对边界框预测的准确性，采用Focaler-CIoU损失函数，动态调整困难样本与简单样本的权重，有效提升模型对花蕾、棉铃等小尺度目标及密集重叠目标的定位能力。改进后的CM-DETR模型如图2所示。

2.2.1 RGCSPELAN模块

RT-DETR模型所依赖的ResNet等传统架构，虽然能有效提取层次化特征，但在棉花成熟度检测任务中仍存在多尺度特征提取不足、计算开销大、特征冗余等问题，尤其是在区分不同生长阶段棉花目标的细微特征时（尺寸相近但纹理细节有细微特征差异的目标）受限。针对上述问题，本研究引入RGCSPELAN模块构建轻量化主干网络，在降低模型参数量与计算复杂度的同时，提升对棉花不同生长阶段细微特征的判别性能。

RGCSPELAN是一种高效的特征提取模块，其核心在于两大高效设计：通道卷积划分策略与可重参数化卷积单元（RepConv）。首先，部分卷积机制通过通道级分组处理输入特征图$F \in R^{C \times H \times W}$，该模块将通道划分成两部分：一部分（比例为$\alpha$）送入计算量较大的深度可分离卷积（Depthwise Separable Convolution）以提取空间上下文信息；另一部分则通过轻量化的1×1卷积实现快速的信息传递与特征映射。最终将两部分特征融合，在保证特征表达能力的同时提升模型的运算效率，其计算过程可表示为公式（1）：

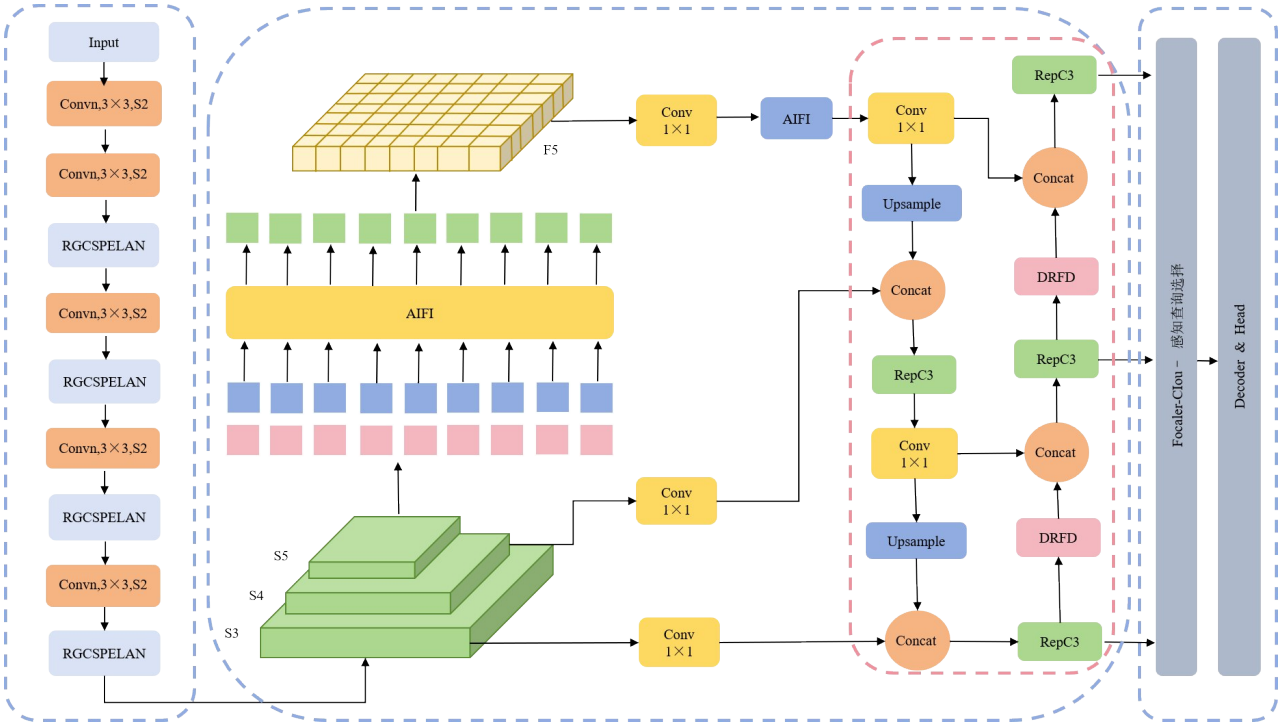


图2　CM-DETR网络结构示意图

Fig. 2　Schematic diagram of CM-DETR network structure

$$F_{out}=\mathrm{Concat}(\mathrm{DWConv}(F_{1:\alpha C}),\mathrm{Conv}_{1\times1}(F_{\alpha C:C}))\quad(1)$$

式中：DWConv（·）表示深度可分离卷积；$\mathrm{Conv}_{1\times1}$（·）为逐点卷积操作；Concat（·）表示通道拼接；α为通道划分比例参数，本试验中统一设置α=0.500，即将输入通道均匀划分为两部分，以在计算效率与特征表达能力之间取得平衡。

其次，为进一步提升特征提取能力，该模块摒弃了传统的残差瓶颈（Bottleneck）结构，转而采用RepConv作为核心特征提取单元，以增强特征辨识能力。RepConv在训练阶段采用多分支并行结构，融合3×3卷积、1×1卷积和恒等映射等操作，通过丰富的梯度路径增强模型的学习能力，从而有效提升模型对复杂特征的捕捉能力，其训练阶段的表达形式为公式（2）：

$$Y=\mathrm{Conv}_{3\times3}(F)+\mathrm{Conv}_{1\times1}(F)+F\quad(2)$$

式中：Y表示RepConv模块在训练阶段的输出特征图；$\mathrm{Conv}_{3\times3}$（·）表示3×3标准卷积操作，用于提取局部空间特征。

在推理阶段，通过结构重参数化技术将上述多分支结构等效融合为一个单一的标准卷积核，在保证性能的同时提升推理速度，其融合表示为公式（3）：

$$\hat{W}=W_{3\times3}+pad(W_{1\times1})+I\quad(3)$$

式中：$\hat{W}$表示结构重参数化后的等效3×3卷积核；$W_{3\times3}$表示3×3卷积分支的卷积核权重；$W_{1\times1}$表示1×1卷积分支的卷积核权重，pad（·）表示补零操作以统一核尺寸；I表示恒等映射。

RepConv是一种训练时多分支、推理时单分支的可重参数化卷积结构，是RGCSPELAN模块的核心。其结构在训练阶段包含三个并行分支，具体包括一个3×3卷积、一个1×1卷积以及一个恒等映射分支，其中，3×3卷积作为主分支用于提取局部空间特征，1×1卷积作为辅助分支起到通道压缩与信息补充的作用，恒等映射分支能够保留输入特征并促进梯度反向传播，从而提升模型的表达能力和训练稳定性。在推理阶段，通过结构重参数化技术将多分支结构等效融合为单一的3×3卷积核，既保持了训练阶段的性能优势，又显著提升了推理效率，其结构如图3所示。

在此基础上，为充分发挥RepConv的特征建模优势并将其有效融入整体网络结构中，本研究进一步构建了RGCSPELAN模块，其结构如图4所示。其中，图4a展示了RGCSPELAN模块的整体结构，图4b给出了模块内部的具体实现方式。该模块首先通过1×1卷积对输入特征图进行通道调整，随后在通道维度上按比例划分为两个分支：一条分支引入RepConv模块以实现多尺度特征提取，另一条

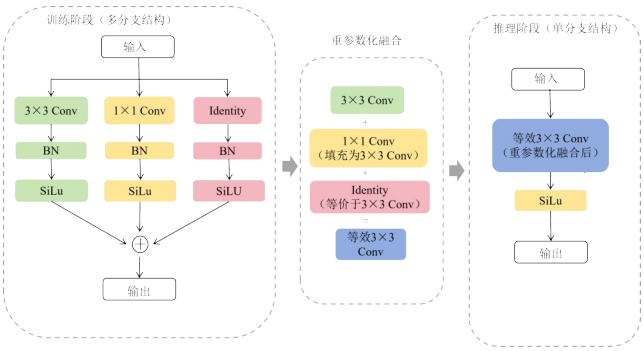


图 3 RepConv 模块结构图

Fig. 3 Structure diagram of RepConv module

分支通过多层 3×3 卷积逐级堆叠以增强语义表达能力。两条分支的输出特征经拼接操作进行融合，并通过 1×1 卷积完成特征整合与输出。在该结构中，3×3 卷积的堆叠数量为 n−1，用于适配不同的网络深度需求。当 n=3 时，RepConv 分支由两个连续的 3×3 卷积和一个 1×1 卷积组成，该配置在有效扩大感受野的同时增强了对局部细节特征的感知能力。综合检测性能与计算开销的平衡，本研究最终采用 n=3 作为 RGCSPELAN 模块的结构参数。

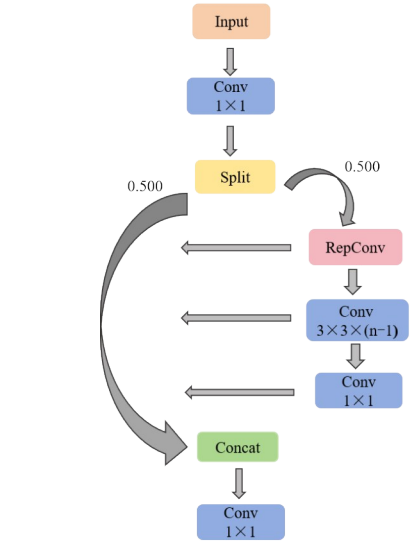
此外，RGCSPELAN 模块引入通道缩放因子与分支比例控制机制，通过通道扩展系数 e 与比例系数 scale 对隐藏及中间特征通道宽度进行约束，从而在网络构建阶段实现对模块参数量与计算复杂度的灵活调节。该设计使模块在保持结构一致性的前提下，可根据任务需求自适应调整模型规模，既满足轻量化部署需求，又兼顾高精度检测场景。

2.2.2 DRFD 下采样模块

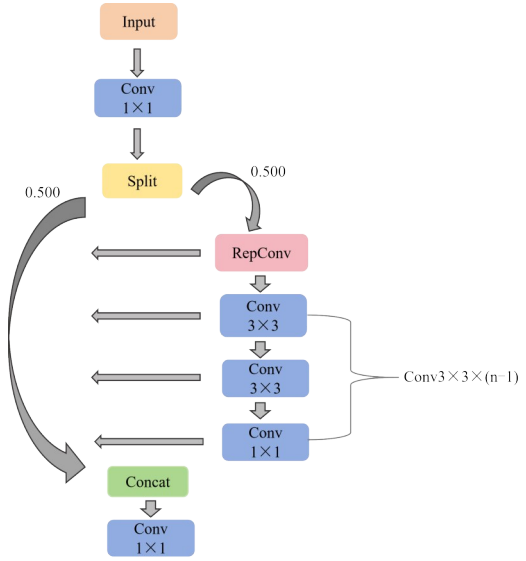
在农业监测中，棉花图像存在的低分辨率、小尺寸目标特征信息不足等问题，使传统计算机视觉方法在棉花生长状态识别中面临严峻挑战[24]。特别是在棉花成熟度检测任务中，不同生长阶段的形态特征差异细微，而传统卷积神经网络中的多次下采样操作易造成小目标特征的丢失，从而降低检测精度。

针对上述问题，本研究引入深层稳健特征下采样 DRFD[25] 模块，在下采样过程中通过多分支并行结构引入特征补偿机制，以增强多尺度特征表达，如图 5 所示。

DRFD 模块由 D_{conv}、D_{cut} 和 D_{max} 这 3 条并行下采样路径构成。设输入特征图为 Y，3 条路径分别对其进行下采样处理以获得特征表示。其中，D_{conv} 路径采用分组深度卷积（Depthwise Conv）实现下采样，并结合批归一化与高斯误差线性单元（Gauss-



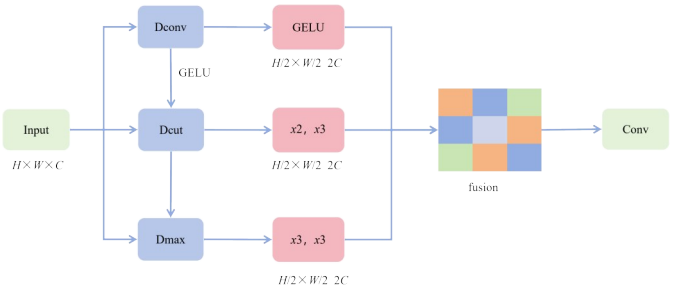
a. RGCSPELAN 模块整体结构



b. RGCSPELAN 模块内部结构

图 4 RGCSPELAN 模块的结构图

Fig. 4 RGCSPELAN module architecture



注：H、W、C 表示特征图的高度、宽度、通道数；D_{conv} 表示分组+深度可分离卷积下采样；D_{cut} 表示切块拼接下采样；D_{max} 表示分组卷积+最大池化下采样；GELU 表示高斯误差线性单元激活函数；fusion 表示特征融合操作。

图 5 DRFD 模块结构图

Fig. 5 Architecture of the DRFD module

ian Error Linear Unit，GELU）激活函数增强特征非线性表达，其输出通道数由C扩展至$2C$，计算过程表示为公式（4）：

$$x_1=\mathrm{GELU}(\mathrm{D}_{\mathrm{conv}}(Y)) \tag{4}$$

式中：$\mathrm{D}_{\mathrm{conv}}$（·）表示分组深度卷积下采样操作；$x_1$表示$\mathrm{D}_{\mathrm{conv}}$路径下采样后的特征图，空间尺寸减半，通道数翻倍。

$\mathrm{D}_{\mathrm{cut}}$路径通过切块拼接方式实现下采样，其空间结构如图6所示。具体而言，$\mathrm{D}_{\mathrm{cut}}$首先将输入特征图在空间维度上划分为多个子块，并按照预定义规则对不同位置子块进行重排与拼接，从而在降低空间分辨率的同时保留局部细节与位置信息，有利于小目标特征的保持，其输出同样扩展至$2C$通道，计算过程表示为公式（5）：

$$x_2=\mathrm{D}_{\mathrm{cut}}(Y) \tag{5}$$

式中：$\mathrm{D}_{\mathrm{cut}}$（·）表示切块重排下采样操作；$x_2$表示$\mathrm{D}_{\mathrm{cut}}$路径下采样后的特征图。

$\mathrm{D}_{\mathrm{max}}$路径则采用分组卷积结合最大池化的方式实现下采样，通过最大池化操作增强对显著响应区域的关注能力，同时借助分组卷积完成通道扩展，其输出为公式（6）：

$$x_3=\mathrm{D}_{\mathrm{max}}(Y) \tag{6}$$

式中：$\mathrm{D}_{\mathrm{max}}$（·）表示分组卷积与最大池化结合的下采样操作；$x_3$表示$\mathrm{D}_{\mathrm{max}}$路径下采样后的特征图。

最后，将3三条路径得到的特征x_1、x_2和x_3在通道维度进行拼接，使特征通道数扩展至$6C$，并通过一个步长为1的1×1卷积层进行特征融合与通道压缩，最终输出通道数为$2C$，整体融合过程可表示为公式（7）：

$$X=\mathrm{fusion}(x_1, x_2, x_3) \tag{7}$$

式中：fusion（·）表示通道维度拼接及1×1卷积融合操作，X为DRFD模块最终输出特征图。通过并行融合三种具有互补特性的下采样路径，DRFD模块在有效降低特征图分辨率的同时缓解了单一路径下采样导致的细节信息损失问题，增强了模型对不同尺度棉花目标及复杂农田背景的鲁棒性，从而提升棉花不同成熟检测阶段目标的检测精度与稳定性。

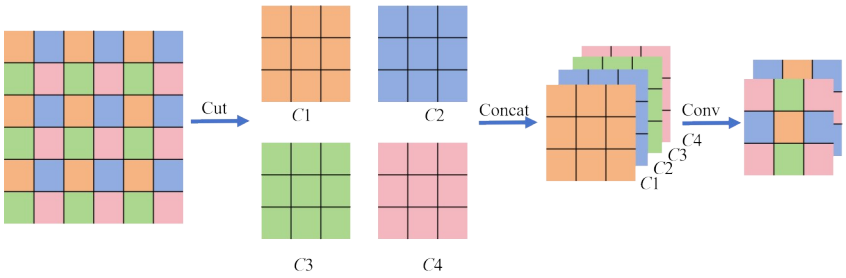


图6　$\mathrm{D}_{\mathrm{cut}}$模块的空间结构图

Fig. 6　Spatial structure diagram of the $\mathrm{D}_{\mathrm{cut}}$ module

2.2.3　Focaler-CIoU损失函数模块

边界框回归是目标检测中关键的组成部分，其性能直接影响模型对目标位置的预测精度。目前广泛采用的回归损失函数多基于IoU及其改进版本，如广义IoU（Generalized IoU, GIoU）[26]、距离IoU（Distance IoU, DIoU）[27]和完全IoU（Complete IoU, CIoU）[28]等。这些方法侧重于从几何层面建模预测框与真实框之间的关系。然而，这类方法普遍忽略了对训练样本的预测质量与回归难度的有效区分，从而在复杂生育阶段及小目标场景下难以兼顾精度与稳定性。

针对上述问题，本研究采用了一种改进的回归损失函数Focaler-CIoU [29]。该方法在CIoU基础上融合了焦点调制机制，有效地将样本难度信息嵌入到损失计算过程中，从而在保持几何约束建模能力的同时，实现对高质量预测样本的梯度强化，降低低质量样本对模型训练的负面影响。

Focaler-CIoU损失的形式定义如公式（8）和公式（9）所示。

$$L_{\mathrm{Focaler-IoU}}=1-\mathrm{IoU}^{\mathrm{Focaler}} \tag{8}$$

$$L_{\mathrm{Focaler-CIoU}}=L_{\mathrm{CIoU}}+\mathrm{IoU}-\mathrm{IoU}^{\mathrm{Focaler}} \tag{9}$$

式中：$L_{\mathrm{Focaler-IoU}}$表示引入焦点调制机制后的IoU损失；$L_{\mathrm{Focaler-CIoU}}$表示最终的Focaler-CIoU回归损失；$L_{\mathrm{CIoU}}$表示完全交并比损失函数；$\mathrm{IoU}^{\mathrm{Focaler}}$表示经过焦点调制后的IoU值，其表达式如公式（10）所示。

$$\mathrm{IoU}^{\mathrm{Focaler}}=\begin{cases}0, & \mathrm{IoU}<d\\ \dfrac{\mathrm{IoU}-d}{u-d}, & d\leqslant \mathrm{IoU}\leqslant u\\ 1, & \mathrm{IoU}>u\end{cases} \tag{10}$$

式中：d与u分别表示IoU调制函数的下限阈值和上限阈值。

通过设置阈值d与u，其中［d，u］∈［0，1］，并对其取值进行调节，可以使$IoU^{Focaler}$关注不同的回归样本，该函数能够自适应地调整不同质量样本对整体损失的影响程度。当IoU较低时，样本对损失的贡献被抑制；当IoU处于中间区间时，通过线性增长强化中间难度样本的回归权重；当IoU较高时，则给予最大梯度反馈。此机制能够有效缓解训练过程中由于样本不平衡带来的梯度主导问题，提升模型对复杂目标的回归适应能力。

从参数设定角度看，阈值参数d与u分别用于界定低质量样本的抑制区间和高质量样本的强化区间。针对棉花多生育期检测中小目标占比高、遮挡现象频繁的特点，在花蕾期与早铃期阶段，目标像素面积有限且易受叶片遮挡，预测框在训练初中期难以实现精确的边界回归。受小目标像素离散效应及IoU对边界偏移高度敏感的影响，部分已覆盖目标主体但边界尚未贴合的预测样本，其IoU主要集中分布于0.2～0.5的区间，而在裂铃期与成熟期等目标尺寸大、边界更清晰的阶段，中高质量样本的IoU多分布在0.6以上。若直接采用通用阈值设置，容易对训练初中期占比较高的中等质量样本产生过度抑制，从而影响模型的收敛稳定性与定位精度。

基于上述IoU分布规律，本研究将d的取值范围设定为0.2～0.5，以有效抑制低IoU噪声样本的梯度干扰并保留中等质量样本的学习信号；将u的取值范围设定为0.6～0.9，以在模型逐步收敛后强化中高IoU样本在裂铃期与成熟期等复杂生育阶段的回归反馈，从而在梯度抑制与增强之间取得合理平衡，并通过参数对比试验验证了该取值设定的有效性。

2.3 模型训练环境与评价指标

2.3.1 环境配置与训练参数设置

本研究试验环境是Python 3.8.20，深度学习框架是PyTorch 1.13.0，所有模型都在NVIDIA GeForce RTX 4090上进行了训练和测试。在训练阶段，本研究使用AdamW优化器进行优化，初始学习率设置为0.001，权重衰减为0.000 1，动量为0.9，训练过程中保持固定学习率，未采用额外衰减策略，基于多组预试验结果分析，固定学习率在棉花多生育阶段检测任务中表现出更加平稳的收敛特性；相比之下，引入学习率衰减易在训练中后期造成梯度过快减小，从而影响小目标及遮挡样本的持续优化效果。因此，本研究训练过程中采用固定学习率策略。输入图像统一缩放至640×640，批大小设置为16，训练轮次为150轮。训练期间同步在验证集上监控模型性能以观察收敛趋势，但未作为提前终止依据；考虑到模型损失在约120轮后趋于稳定，最终统一采用固定轮次结束训练，未引入早停机制，以保证各模型在统一条件下进行对比。

2.3.2 评价指标

本研究选取如下指标评估模型的检测性能：精确率（Precision, P）、召回率（Recall, R）、F_1得分（F_1-Score）、平均精度（Average Precision, AP）、平均精度均值（Mean Average Precision, mAP）、参数量（Params）、计算量（GFLOPs）。其中P、R计算方法如公式（11）和公式（12）所示。

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{11}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{12}$$

式中：TP表示真正例，是为正样本被正确识别为正样本的个数；FP表示假正例，是为负样本被错误识别为正样本的个数；FN表示假负例，是为正样本被错误识别为负样本的个数。

F_1（F_1-Score）是分类任务中常用的综合评价指标，其定义为P与R的调和平均数。具体计算如公式（13）所示。

$$F_1 = 2 \times \frac{P \times R}{P + R} \tag{13}$$

平均精度AP是由P和R计算得到的，其定义为某一类别对应的P-R曲线下的面积，用于计算不同类别的平均精确度，计算如公式（14）所示。

$$\mathrm{AP} = \int_0^1 P(R)\,d(R) \tag{14}$$

在此基础上，本研究采用mAP作为模型的评价指标，mAP表示对数据集中所有类别的AP取平均，其计算如公式（15）所示。

$$\mathrm{mAP} = \frac{\sum_{i=1}^{k} \mathrm{AP}_i}{k} \tag{15}$$

式中：k表示目标类别总数。

其中mAP50表示在IoU阈值为0.5时各类别的平均精度；mAP50～95表示IoU从0.5到0.95（步长为0.05）之间的平均精度。

3 结果与分析

3.1 消融实验

本研究为了进一步验证各改进模块对模型性能

的影响，在保持试验环境及超参数一致的条件下，以RT-DETR为基线模型开展消融实验，逐步引入RGCSPELAN模块、DRFD下采样模块及Focaler-CIoU损失函数，试验结果如表2所示。

表2 CM-DETR模型消融实验结果

Table 2 Ablation study results of the CM-DETR model

RT-DETR	RGCSPELAN	DRFD	Focaler-CIoU	*P*/%	*R*/%	F_1/%	mAP50/%	mAP50～95/%	参数量/M	计算量/GFLOPs	帧率/(帧/s)
√	×	×	×	79.5	74.5	76.9	77.1	49.3	19.9	57.0	77.3
√	√	×	×	81.4	75.3	78.2	78.2	50.3	13.8	44.5	83.8
√	×	√	×	83.3	74.6	78.7	77.8	49.3	19.6	56.5	75.9
√	×	×	√	80.2	76.8	78.5	78.9	49.8	19.9	57.0	76.4
√	√	√	×	82.2	76.6	79.3	78.8	49.4	13.6	44.0	83.3
√	√	×	√	79.8	77.7	78.7	79.7	50.1	13.8	44.5	87.6
√	×	√	√	79.7	74.9	77.2	78.3	49.9	19.6	56.5	74.6
√	√	√	√	82.1	77.4	79.7	80.8	51.1	13.6	44.0	83.6

注：√表示使用该模块；×表示未使用该模块。

由表2可知，各模块对模型性能产生不同程度的正向提升。首先，引入RGCSPELAN模块后，模型的*P*、*R*、mAP50和mAP50～95分别提升了1.9、0.8、1.1和1.0个百分点，同时模型的参数量和计算量分别下降了30.7%、21.9%。这表明RGCSPELAN通过通道划分与RepConv可重参数化结构，在显著降低模型复杂度的同时增强了局部纹理与细微特征的表达能力，从而有效提升对花蕾期与早铃期等小尺寸、纹理差异细微目标的判别精度。

在此基础上，进一步引入DRFD下采样模块后，模型的*P*、*R*、mAP50和mAP50～95分别提升了2.7、2.1、1.7、0.1个百分点，DRFD通过多路径并行下采样与特征补偿机制，在降低空间分辨率的同时有效保持局部结构信息，从而增强模型在遮挡及复杂背景场景下对棉花目标的特征表达能力。

在上述结构改进的基础上引入Focaler-CIoU损失函数，最终模型mAP50较基线提升3.7个百分点，该损失函数通过对回归样本质量进行自适应调制，强化中高质量样本的梯度反馈，抑制低质量样本的干扰，在棉花多生育阶段检测任务中有效提升了边界框回归的稳定性与定位精度。

综合三种改进后，将3种方法组合在一起的模型性能最优，相比于原模型RT-DETR的*P*、*R*、mAP50和mAP50～95分别提升了2.6、2.9、3.7和1.8个百分点，模型参数量和计算量分别下降了31.7%和22.8%。最终，CM-DETR在检测精度与计算效率方面均表现出明显优势，验证改进模块在棉花成熟度检测任务中的有效性。

3.2 不同主干网络的对比

为研究RGCSPELAN主干网络的合理性，选取多种具有代表性的主干网络，如上下文引导（Context Guided）网络[30]、上下文引导残差特征金字塔网络（Context-Guided Residual Feature Pyramid Network, CGRFPN）[31]等主干网络进行对比试验，所有的试验参数均保持一致，只改变模型的主干网络，以mAP50、mAP50～95、参数量和帧率（Frames Per Second, FPS）等作为评价指标，最终试验结果如表3所示。

表3 RT-DETR模型搭载不同主干网络的对比试验结果

Table 3 Results of RT-DETR comparison experiment with different backbone networks

模型	*P*/%	*R*/%	mAP50/%	mAP50～95/%	参数量/M	计算量/GFLOPs	帧率/(帧/s)
RT-DETR(baseline)	79.5	74.5	77.1	49.3	19.9	57.0	42.8
+RGCSPELAN	81.4	75.3	78.2	50.3	13.8	44.5	52.8
+EMBSFPN	81.6	72.7	76.6	49.2	17.9	48.6	33.9
+PACAPN	81.1	75.1	76.4	48.9	19.9	38.8	38.8
+Context Guided	80.1	73.7	77.1	48.7	16.5	47.6	43.7
+CGRFPN	81.7	75.2	76.0	48.9	19.2	48.2	41.0

由表3可知，引入不同主干网络后，基线模型的检测精度有所提升，并且基线模型的参数量都降低，其中，采用RGCSPELAN作为主干网络时，各项指标提升最为显著，改进后的RT-DETR的P、R、mAP50、mAP50～95和FPS分别达到81.4%、75.3%、78.2%、50.3%、52.8帧/s。试验结果表明，引入RGCSPELAN主干网络后，在模型参数量显著降低30.7%的同时，使模型P提升1.9个百分点，mAP50和mAP50～95分别提升了1.1和1.0个百分点，检测速度也提升了10帧/s，该主干网络在检测精度与推理速度方面均表现出显著优势。

综上所述，引入RGCSPELAN主干网络后，在提高棉花成熟度检测精度的同时也降低了模型的参数量，RGCSPELAN模块通过优化特征提取方式，能够有效捕捉棉花的细节特征，更适应于对棉花成熟度的检测任务。进一步考虑到RGCSPELAN模块中通道划分比例参数α决定了深度可分离卷积分支与轻量分支的通道分配，从而影响深度特征建模能力与计算复杂度的平衡。在确定RGCSPELAN为最优主干网络的基础上，对不同α取值进行对比试验，结果如表4所示。

表4 RGCSPELAN模块中不同通道划分比例α取值的对比

Table 4 Comparison of different channel partitioning ratio α values in the RGCSPELAN module

α取值	P/%	R/%	F_1/%	mAP50/%	mAP50～95/%
0.250	81.5	76.7	79.0	79.4	49.9
0.375	80.9	75.6	78.2	78.1	48.7
0.500	82.1	77.4	79.7	80.8	51.1
0.625	80.2	75.7	77.9	80.2	50.6
0.750	79.5	75.1	77.2	78.5	49.5

试验结果表明，当α=0.500时模型在各项指标上均达到最优，其中mAP50和mAP50～95分别达到80.8%和51.1%。这是因为α过小会限制深度可分离卷积分支对空间上下文与局部细节的提取能力；而α过大会增加高计算分支占比，削弱轻量分支的信息传递效率，从而导致特征冗余并降低推理速度。当α=0.500时，两类分支在特征表达能力与计算开销之间实现最佳平衡，使模型既能充分捕获棉花目标的细微纹理差异，又保持较高推理效率。因此，本研究最终选用RGCSPELAN作为主干网络，并将通道划分比例α固定为0.500，以兼顾计算效率和特征表达能力。

3.3 不同损失函数的对比

本研究在改进的CM-DETR模型中引入Focaler-CIoU作为边界回归损失函数，以提升模型的收敛速度与检测精度，为了验证其有效性，本研究以RT-DETR作为基线模型，对比了多种主流及专用损失函数下的性能表现，包括CIoU、DIoU、GIoU，以及面向小目标设计的专用损失函数NWD（Normalized Wasserstein Distance）、Slide Loss、Shape-IoU（Shape-Aware Intersection over Union）。Focaler-CIoU与上述损失函数的模型收敛情况及性能对比结果如图7所示。

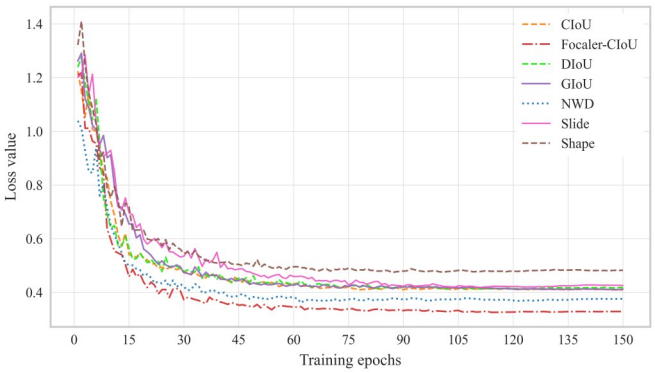


图7 棉花成熟度检测研究中不同损失函数的曲线对比图

Fig. 7 Comparison of loss function curves in cotton maturity detection research

试验结果表明，采用Focaler-CIoU损失函数后，模型在训练过程中表现出更快的收敛速度，且收敛后损失值也低于其他对比方法。在检测精度方面，Focaler-CIoU相比主流基础损失函数CIoU、DIoU和GIoU（其mAP50分别为79.4%、78.9%和78.4%），分别提升了1.4、1.9和2.4个百分点；与专为小目标设计的损失函数NWD、Slide Loss和Shape-IoU（其mAP50分别为78.0%、77.2%和78.2%）相比，分别提升了2.8、3.6和2.6个百分点。综上所述，Focaler-CIoU不仅有效加快了模型的收敛速度，还显著提升了检测精度，并且在小目标场景下表现出不错的综合性能。尽管Focaler-CIoU损失函数已展现出优势，但其性能仍受到阈值的影响，为了进一步优化Focaler-CIoU损失函数的性能并确定其阈值参数的最优取值，本研究进一步对阈值参数d与u的不同取值组合进行对比试验，结果如表5所示。

试验结果表明，Focaler-CIoU损失函数的性能受到阈值参数d与u取值的影响，由表5可知，在不同参数组合中，当取d = 0.50、u = 0.60时，模型

表5 Focaler-CIoU模块中不同阈值参数 d、u 取值的对比

Table 5 Comparison of different values of threshold parameters d and u in the Focaler-CIoU module

d取值	u取值	P/%	R/%	F_1/%	mAP50/%	mAP50～95/%
0.20	0.90	80.8	76.0	78.3	79.9	50.2
0.25	0.70	81.7	75.6	78.5	77.9	48.7
0.40	0.85	80.2	72.6	76.2	76.1	48.2
0.55	0.75	82.3	73.4	77.6	78.5	49.2
0.30	0.80	77.4	71.3	74.2	74.6	47.4
0.50	0.60	82.1	77.4	79.7	80.8	51.1

取得最优性能，其mAP50和mAP50～95分别达到80.8%、51.1%，同时保持82.1%的精确率和77.4%的召回率，表明该参数配置在精确率与泛化能力间达到了最佳平衡。相比之下，其他参数虽然在个别指标上表现良好（如d = 0.55、u = 0.75时精确率为82.3%），但综合性能不及该参数配置。本研究通过阈值参数对比试验，明确阈值参数对模型性能的调节机制，为Focaler-CIoU损失函数的阈值设置提供依据。

综上所述，在棉花的多生长周期检测任务中，Focaler-CIoU展现出良好的适应性。面对不同生长阶段目标尺度变化及遮挡重叠等复杂情况，该损失函数通过强化中高质量样本的优化，提高了模型对小目标及复杂场景的检测稳定性，从而为棉花成熟度识别与田间智能管理提供可靠的数据支持。

3.4 不同模型性能对比

为了进一步验证CM-DETR模型在棉花成熟度检测任务中的性能，本研究将其与多种先进检测模型进行对比试验，包括单阶段检测器：改进去噪锚框DINO（DETR with Improved deNoising anchOr boxes），YOLO系列的YOLOv5m、YOLOv8m、YOLOv10m，以及基于DETR架构的RT-DETR模型。所有模型均在相同训练参数配置下进行试验：统一使用AdamW优化器（初始学习率0.001，权重衰减0.0001，动量0.9），训练轮次150轮，批处理大小16，输入图像尺寸640×640，并采用一致的数据增强策略。对比试验结果如表6所示。

表6 棉花成熟度检测研究中不同模型的检测结果

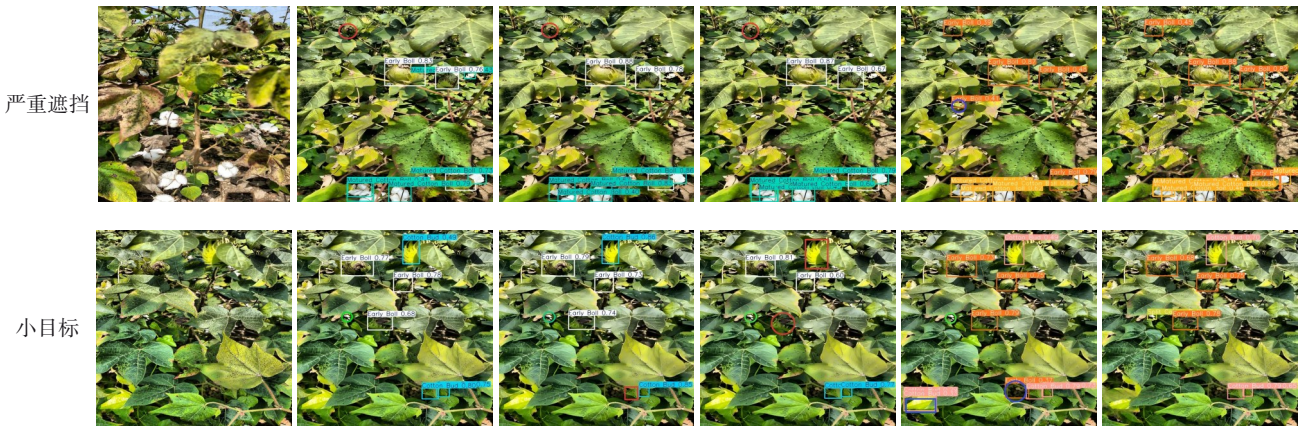
Table 6 Detection experimental results of different models in cotton maturity detection research

模型	P/%	R/%	F_1/%	mAP50/%	mAP50～95/%	参数量/M	计算量/GFLOPs
DINO	82.2	69.1	75.1	77.9	48.0	46.8	268.2
YOLOv5m	79.4	71.1	75.0	78.7	50.5	22.1	64.0
YOLOv8m	76.7	73.2	74.9	77.6	48.6	25.9	79.3
YOLOv10m	75 .1	73.7	74.3	78.0	50.0	15.3	58.9
RT-DETR	79.5	74.5	76.9	77.1	49.3	19.9	57.0
CM-DETR	82.1	77.4	79.7	80.8	51.1	13.6	44.0

由表6可知，改进后的CM-DETR模型的P、R、mAP50和mAP50～95分别为82.1%、77.4%、80.8%和51.1%，整体在所有模型中最高。与DINO模型相比，本研究模型的参数量和检测速度都明显优于DINO模型；改进的CM-DETR模型相比于YOLOv5m模型，P和R分别提高了2.7和6.3个百分点；相较于YOLOv8m、YOLOv10m，改进模型的精确率更高，对小目标的识别也更准确。与DINO、YOLOv5m、YOLOv8m、YOLOv10m和RT-DETR基准模型相比，CM-DETR模型的mAP50分别提高了2.9、2.1、3.2、2.8和3.7个百分点；mAP50～95分别提高了3.1、0.6、2.5、1.1和1.8个百分点。相较于以上的模型，CM-DETR模型的浮点计算量和参数量分别为44 GFLOPs和13.6 M，与DINO、YOLOv5m、YOLOv8m、YOLOv10m和RT-DETR模型相比，CM-DETR模型的计算量分别降低了83.6%、31.3%、44.5%、25.3%和22.8%；参数量分别降低70.9%、38.5%、47.5%、11.1%和31.7%。由此可知，CM-DETR模型具有较好的检测精度的同时降低了模型的计算量和参数量。

为了进一步验证CM-DETR模型在棉花成熟度检测任务中对复杂场景的检测性能，本研究选取DINO、YOLOv8m、YOLOv10m、基准模型RT-DETR，以及改进后的CM-DETR模型，对其在小目标与严重遮挡场景下的检测结果进行可视化对比分析，结果如图8所示。

从可视化结果看出，在遮挡严重的场景下，DINO、YOLOv8m、YOLOv10m均出现漏检情况，由于棉花叶片的严重遮挡以及绿色叶片与早棉铃的绿色花苞颜色相近，无法识别出早棉铃的重要特



a. 原图 b. DINO c. YOLOv8m d. YOLOv10m e. RT-DETR f. CM-DETR

注:红色圆圈标记的是漏检的早铃棉期,红色方框标记的是漏检的花蕾期,绿色圆圈标记的是漏检的裂棉铃期,蓝色圆圈标记的是误检的早铃棉期,蓝色方框标记的是误检的花蕾期,其余颜色的方框是正确类别。

图8 不同模型在棉花成熟度检测结果的可视化对比

Fig. 8 Visualization of detection results of different models in cotton maturity detection

征，造成了漏检，而RT-DETR出现误检情况，误将棉花的叶片识别成早棉铃。

在小目标检测方面，YOLOv8m、YOLOv10m对棉花蕾的检出率较低，早期的棉花蕾较小，颜色与棉花绿色叶片相近，造成了漏检。此外，图8中的裂棉铃期的目标也因尺寸较小，表观特征不明显，除了CM-DETR能够正确检测外，其余四种模型均未能识别。在RT-DETR模型检测过程中，出现误检情况，棉花叶片边缘具有锯齿状裂齿，花蕾期苞片（副萼）的裂齿同样呈锯齿状，棉花叶片与花蕾期的特征相似，从而导致误检的发生，误将棉花叶片识别成早棉铃，进一步说明模型在处理小目标检测时的挑战。

在本研究提出的改进的CM-DETR模型与其他5类模型的棉花成熟度可视化对比图中，发现各类算法或多或少存在误检和漏检的问题，有一些棉花叶片被错误识别为早棉铃和棉花花蕾，同时也有某些棉花蕾未被检测出来。经过改进后的算法可以适应不同的复杂场景，不仅满足棉花成熟度检测对高精度的要求，还有效降低棉花在复杂场景检测过程中出现漏检、误检的问题，验证了改进后的算法具有较好的检测效果。

3.5 棉花成熟度检测分析

3.5.1 棉花不同成熟度类别的性能对比

为探究模型对棉花不同成熟阶段的识别能力差异，并验证改进后模型CM-DETR在检测性能方面的有效性，本研究针对棉花的五个成熟度类别（花蕾期、棉花花期、早铃棉期、裂棉铃期和成熟棉铃期），分别进行了改进前后的AP值对比试验，试验结果如表7所示。

表7 棉花成熟度检测研究中不同成熟度类别的AP对比结果

Table 7 Comparison results of AP for different maturity categories in cotton maturity detection research

棉花成熟度类别	改进前RT-DETR平均精度		改进后CM-DETR平均精度	
	AP50/%	AP50～95/%	AP50/%	AP50～95/%
花蕾期	44.4	59.8	47.4	63.3
棉花花期	62.4	78.6	64.8	83.3
早铃棉期	62.2	80.5	62.4	81.8
裂棉铃期	61.8	78.3	70.0	86.7
成熟棉铃期	70.8	85.8	70.6	86.6

试验结果表明，CM-DETR模型在不同成熟期的检测精度都有所提升（除成熟棉铃期），其中裂铃棉期的AP50和AP50～95分别提升了8.2和8.4个百分点，棉花花期的AP50和AP50～95分别提升了2.4和4.7个百分点，这表明改进方法有效增强了模型对棉花成熟阶段特征的判别能力；花蕾期的AP50和AP50～95分别提升了3.0和3.5个百分点，表明改进方法缓解了小尺寸目标方面的漏检问题；成熟棉铃期改进后的AP50～95的提升说明Focaler-CIoU有效强化了高质量样本的回归权重，提升了边界框的定位精度，使得模型整体性能保持稳定。

综上所述，改进后的模型CM-DETR在棉花花期和裂铃棉期的检测精度方面提升显著，同时早铃棉期和花蕾期的检测精度也有一定的提升，成熟棉

铃期的整体性能保持稳定，验证了CM-DETR模型对棉花成熟阶段识别能力的有效性。

3.5.2　棉花成熟度检测效果热力图分析

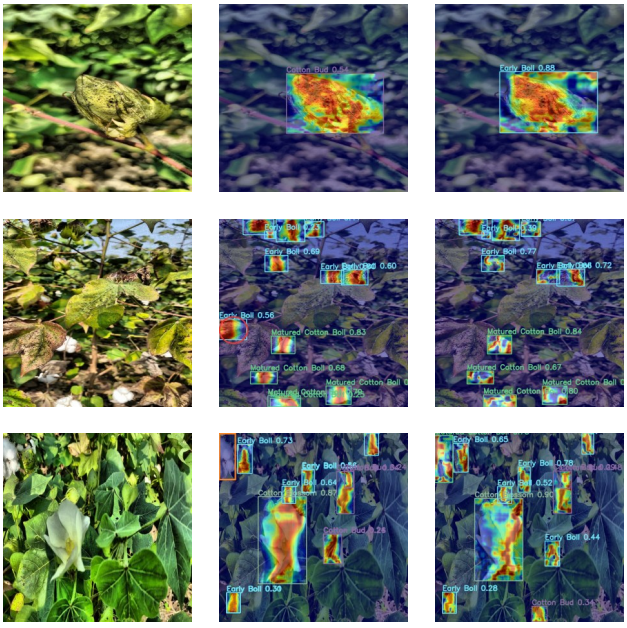
为更加直观地验证改进效果，本研究对棉花成熟度的检测效果进行分析。图9中展示了原始图、基准模型RT-DETR和改进的CM-DETR模型的检测效果对比，为直观且定量地验证模型改进效果，从检测精度、误检、漏检这三个方面对两种模型的性能进行分析。第一行热力图显示，两类模型均能正确识别出棉花的花蕾期，但CM-DETR的检测置信度达到0.88，较RT-DETR的0.54提升了62.9%，具有较高的识别精度。第二行热力图显示，在复杂背景的干扰下，CM-DETR模型精准检测出早铃棉期和裂铃棉期，其早铃棉检测的平均置信度为0.77，较基准模型（0.69）有所提升；而RT-DETR模型则误将棉花叶片识别成早铃棉，置信度为0.58。第三行结果显示，CM-DETR模型能够准确检测出早铃棉期、棉花花期和成熟棉铃期，并正确对应其类别，其中棉花花期的检测置信度达到了0.90，早铃棉期平均置信度为0.53，也高于基准模型的0.48，而RT-DETR模型在识别过程中未能检测到成熟棉铃棉期，出现漏检情况。

综上所述，CM-DETR在保持检测精度的同时，显著降低了误检和漏检率。相较于基准模型，改进后CM-DETR模型在检测精度与鲁棒性方面均表现出优势，表明该模型具有更强的特征提取能力，从而有效提升复杂场景下棉花目标检测的准确性与稳定性。

4　结　论

针对传统棉花成熟度检测依赖人工巡田，存在效率低、成本高，以及现有模型识别精度低和难以有效识别花蕾等小目标，也无法准确区分多阶段特征等问题。本研究提出了一种改进的棉花成熟度检测算法CM-DETR，主要结论如下。

1）为构建一个轻量且高效的主干网络，引入RGCSPELAN模块，该模块将渐进式卷积与可重参数化卷积巧妙结合，增强模型的特征表达能力和梯度流通效率，并且可以通过缩放因子控制RGCSPELAN的大小，使其可以兼顾小模型和大模型。接着采用下采样模块DRFD，通过多分支特征融合和特征补偿机制，有效减少下采样导致的小目标特征丢失问题，提高复杂背景下的检测稳定性。最后采用Focaler-CIoU损失函数，实现对不同难度样本的动态权重调整，提升模型的检测精度。



a. 原图　　b. RT-DETR　　c. CM-DETR

注：红色圆圈标记的是误检的早铃棉期，橙色方框标记的是漏检的成熟棉铃期，蓝色和紫色方框标记的是正确类别。

图9　改进前后棉花检测热力图分析

Fig. 9　Thermal map analysis of cotton detection before and after improvement

2）试验结果表明，改进后的CM-DETR模型相比RT-DETR模型精确率、召回率、mAP50和mAP50～95分别提升了2.6、2.9、3.7和1.8个百分点，同时浮点运算量和参数量分别降低了22.8%和31.7%，在实现更高精度的同时降低计算开销。

3）CM-DETR能够有效识别不同成熟度的棉花目标，参数量少、计算效率高，可为精准灌溉、科学施肥与采收决策提供数据支持，从而提升棉花产量与品质。

4）尽管模型在小目标和遮挡严重的复杂场景下表现良好，但在实际应用中仍存在局限：在不同光照条件、极端天气或杂乱背景下的检测稳定性可能受到影响；大规模农田部署时，计算资源和实时处理能力亦需考虑。未来研究可进一步探索模型在多光照、多环境条件下的鲁棒性及轻量化优化，以全面提升模型在实际农田环境中的泛化能力与实用性。

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] 李玉超, 张立杰, 乔心培. 棉花产业协同集聚对棉花绿色全要素生产率的影响[J]. 中国生态农业学报(中英文), 2025, 33(5): 985-998.
LI Y C, ZHANG L J, QIAO X P. The impact of cotton industry synergistic agglomeration on cotton green total factor productivity [J]. Chinese Journal of Eco-Agriculture, 2025, 33(5): 985-998.

[2] 范军亮, 白振涛, 李云霞, 等. 灌水施氮和缩节胺用量对南疆棉花产量品质和水肥利用效率的影响[J]. 农业工程学报, 2024, 40 (9): 68-78.
FAN J L, BAI Z T, LI Y X, et al. Effects of irrigation amount, nitrogen rate and mepiquat chloride dose on cotton yield, quality and water-fertilizer use efficiency in Southern Xinjiang of China [J]. Transactions of the Chinese Society of Agricultural Engineering, 2024, 40(9): 68-78.

[3] 吴华瑞, 李静晨, 杨雨森. 基于大语言模型的个性化作物水肥管理智能决策方法[J]. 智慧农业(中英文), 2025, 7(1): 11-19.
WU H R, LI J C, YANG Y S. Intelligent decision-making method for personalized vegetable crop water and fertilizer management based on large language models[J]. Smart Agriculture, 2025, 7 (1): 11-19.

[4] 吴琼, 刘保军, 李慧, 等. 不同基肥水平对棉苗生长和棉花产量的影响[J]. 新疆农业科学, 2020, 57(4): 740-745.
WU Q, LIU B J, LI H, et al. Effects of different basal fertilizer levels on cotton seedling growth and cotton yield[J]. Xinjiang Agricultural Sciences, 2020, 57(4): 740-745.

[5] 吴传云, 冯健, 陈传强, 等. 我国棉花产业现状与机械化发展情况分析[J]. 中国农机化学报, 2021, 42(5): 215-221.
WU C Y, FENG J, CHEN C Q, et al. Analysis of the status quo and mechanization development of cotton-producing industry in China[J]. Journal of Chinese Agricultural Mechanization, 2021, 42(5): 215-221.

[6] 辛明华, 王占彪, 韩迎春, 等. 新疆机采棉发展回顾、现状分析及措施建议[J]. 中国农业科技导报, 2021, 23(7): 11-20.
XIN M H, WANG Z B, HAN Y C, et al. Review, status and measures of Xinjiang machine-picked cotton[J]. Journal of Agricultural Science and Technology, 2021, 23(7): 11-20.

[7] SHAO L T, GONG J Q, FAN W Q, et al. Cost comparison between digital management and traditional management of cotton fields: Evidence from cotton fields in Xinjiang, China[J]. Agriculture, 2022, 12(8): 1105.

[8] 李瑞鑫, 曹悦, 司秀丽, 等. 基于改进YOLOv8n的樱桃成熟度检测研究[J]. 智能化农业装备学报(中英文), 2025, 6 (4): 40-49.
LI R X, CAO Y, SI X L, et al. Research on cherry ripeness detection based on improved YOLOv8n[J]. Journal of Intelligent Agricultural Mechanization, 2025, 6(4): 40-49.

[9] CHOUDHARY K, DECOST B, CHEN C, et al. Recent advances and applications of deep learning methods in materials science[J]. npj Computational Materials, 2022, 8: 59.

[10] 王东方, 汪军. 基于迁移学习和残差网络的农作物病害分类[J]. 农业工程学报, 2021, 37(4): 199-207.
WANG D F, WANG J. Crop disease classification with transfer learning and residual networks[J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(4): 199-207.

[11] XUE Z Y, XU R J, BAI D, et al. YOLO-tea: A tea disease detection model improved by YOLOv5[J]. Forests, 2023, 14(2): 415.

[12] KINDA Z, MALO S, BAYALA T R. Detection of cotton diseases by YOLOv8 on UAV images using the RT-DETR backbone[C]// Ambient Intelligence-Software and Applications-15th International Symposium on Ambient Intelligence. Cham, Germany: Springer, 2025: 3-13.

[13] TAN C J, LI C Y, SUN J, et al. Multi-object tracking for cotton boll counting in ground videos based on transformer[C]// American Society of Agricultural and Biological Engineers, ASABE Annual International Meeting, 2024. DOI: 10.13031/aim.202400619.

[14] YADAV P K, THOMASSON J A, SEARCY S W, et al. Assessing the performance of YOLOv5 algorithm for detecting volunteer cotton plants in corn fields at three different growth stages[J]. Artificial Intelligence in Agriculture, 2022, 6: 292-303.

[15] 彭炫, 周建平, 许燕, 等. 改进YOLOv5识别复杂环境下棉花顶芽[J]. 农业工程学报, 2023, 39(16): 191-197.
PENG X, ZHOU J P, XU Y, et al. Cotton top bud recognition method based on YOLOv5-CPP in complex environment[J]. Transactions of the Chinese Society of Agricultural Engineering, 2023, 39(16): 191-197.

[16] VERMA P, PAUL A, MACHAVARAM R, et al. Cotton growth stages detection using fine-tuned YOLOv8 deep learning model [C]// Proceedings of the 2024 8th International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence. New York, USA: ACM, 2024: 20-25.

[17] XIE X, CHENG G, WANG J B, et al. Oriented R-CNN for object detection[C]// 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, New Jersey, USA: IEEE, 2021: 3500-3509.

[18] BHARATI P, PRAMANIK A. Deep learning techniques-r-CNN to mask R-CNN: A survey[C]// Computational Intelligence in Pattern Recognition. Singapore, Germany: Springer, 2020: 657-668.

[19] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, New Jersey, USA: IEEE, 2016: 779-788.

[20] 于英杰, 吴坤澍, 冉朋鑫, 等. 基于改进YOLOv7的核桃仁分级研究与试验[J]. 智能化农业装备学报(中英文), 2025, 6 (4): 17-23.
YU Y J, WU K S, RAN P X, et al. Research and experiment on walnut kernel grading based on improved YOLOv7[J]. Journal of Intelligent Agricultural Mechanization, 2025, 6(4): 17-23.

[21] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]// Computer Vision-ECCV 2020. Cham, Germany: Springer, 2020: 213-229.

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

[23] SOUDY M, AFIFY Y, BADR N. RepConv: A novel architecture for image scene classification on Intel scenes dataset[J]. International Journal of Intelligent Computing and Information Sciences, 2022: 1-11.

[24] 吴建成, 郭荣佐, 成嘉伟, 等. 注意力特征融合的快速遥感图像目标检测算法[J]. 计算机工程与应用, 2024, 60(1): 207-216.
WU J C, GUO R Z, CHENG J W, et al. Fast remote sensing image object detection algorithm based on attention feature fusion [J]. Computer Engineering and Applications, 2024, 60(1): 207-216.

[25] LU W, CHEN S B, TANG J, et al. A robust feature downsampling module for remote-sensing visual tasks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 4404312.

[26] QIAN X L, ZHANG N N, WANG W. Smooth GIoU loss for oriented object detection in remote sensing images[J]. Remote Sensing, 2023, 15(5): 1259.

[27] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.

[28] DU S J, ZHANG B F, ZHANG P, et al. An improved bounding box regression loss function based on CIOU loss for multi-scale object detection[C]// 2021 IEEE 2nd International Conference on Pattern Recognition and Machine Learning (PRML). Piscataway, New Jersey, USA: IEEE, 2021: 92-98.

[29] ZHANG H, ZHANG S J. Focaler-IoU: More focused intersection over union loss[PP/OL]. arXiv: 2401.10525, 2024.

[30] WU T Y, TANG S, ZHANG R, et al. CGNet: A light-weight context guided network for semantic segmentation[J]. IEEE Transactions on Image Processing, 2021, 30: 1169-1179.

[31] WANG D R, PENG J M, LAN S, et al. CTDD-YOLO: A lightweight detection algorithm for tiny defects on tile surfaces[J]. Electronics, 2024, 13(19): 3931.

Cotton Maturity Detection Algorithm Based on Improved RT-DETR

SHI Qimeng, WANG Jun*, XU Xiaofeng, ZHANG Weiyi

（*School of Computer and Information Science*, *Anhui Polytechnic University*, *Wuhu* 241000, *China*）

Abstract: [Objective] Cotton maturity assessment is a vital task in precision agriculture, playing a crucial role in supporting timely irrigation, fertilization, and harvesting decisions. However, traditional monitoring approaches are time-consuming and labor-intensive, and current deep learning-based models often struggle to effectively recognize cotton bolls at varying maturity stages, especially in complex field environments with dense foliage, occlusion, and illumination changes. To address these challenges, a high-accuracy and lightweight computer vision model was proposed for cotton maturity detection. **[Methods]** An enhanced detection framework named cotton maturity-detection transformer (CM-DETR) was proposed, based on an improved real-time detection transformer (RT-DETR) architecture. CM-DETR incorporated three core architectural innovations that significantly improved both detection accuracy and computational efficiency. First, to construct a lightweight and efficient backbone, a novel feature extraction module named Re-parameterized Group Convolution Spatial Enhancement Lightweight Attention Network (RGCSPELAN) was introduced. This module integrated Progressive Convolution, which captured hierarchical and local contextual features, with Re-parameterized Convolution (RepConv), which reduced computational complexity during inference by transforming multi-branch structures into a single-path representation. The combination effectively enhanced the model's feature representation capabilities and gradient propagation while minimizing the number of parameters and FLOPs. Furthermore, RGCSPELAN was designed with a scalable architecture, allowing its computational capacity to be adjusted via a scaling factor. This ensured compatibility with both small and large models, facilitating flexible deployment across resource-constrained edge devices and high-performance systems alike. Second, to address the issue of small target feature loss, a new module termed Deep Robust Feature Downsampling (DRFD) was proposed. DRFD employed a multi-scale feature fusion strategy by integrating multiple downsampling branches (e.g., convolutional, cut-based, and max-pooling pathways). This design enabled the model to retain fine-grained spatial details while expanding its receptive field. Third, the original loss function in RT-DETR was replaced with Focaler-CIoU, and an adaptive regression optimization strategy integrating sample reweighting and geometric constraints was implemented to improve bounding box localization under complex conditions. **[Results and Discussions]** Experimental results demonstrated that CM-DETR achieved mAP50 and mAP50~95 scores of 80.8% and 51.1%, respectively, outperforming the baseline model by 3.7 and 1.8 percentage points. Meanwhile, CM-DETR reduced the parameter count and computational cost by 31.7% and 22.8%, respectively, indicating a favorable trade-off between detection accuracy and model efficiency. The incorporation of the DRFD module enhanced the model's sensitivity to small and subtly distinct features related to cotton maturity, improved robustness under diverse field conditions, and enabled more precise detection of cotton bolls at different growth stages. Moreover, the optimized regression strategy contributed to more stable bounding box prediction performance in scenarios involving occlusion, scale variation, and dense foliage. Overall, the proposed architectural improvements effectively strengthened feature representation capability while maintaining lightweight characteristics, thereby demonstrating practical applicability in real-time agricultural environments. **[Conclusions]** The proposed CM-DETR model provides an efficient and scalable solution for automated cotton maturity detection. By enhancing multi-stage feature recognition, improving small-target sensitivity, and reducing the demand on computational resources, CM-DETR serves as a reliable tool for intelligent decision-making in precision agriculture. Its practical deployment can support more accurate timing for irrigation, fertilization, and harvesting, thereby contributing to improved crop management and yield optimization.

Key words: cotton maturity; RT-DETR; object detection; RGCSPELAN; DRFD; Focaler-CIoU loss function

Foundation items: National Natural Science Foundation of China (62406004); Collaborative Innovation Project of Anhui Higher Education Institutions (GXXT-2019-020)

First author: SHI Qimeng, E-mail: 2240932105@stu.ahpu.edu.cn

***Corresponding author:** WANG Jun, E-mail: wangjun@ahpu.edu.cn