doi:10.6041/j.issn.1000-1298.2025.11.007

基于 VSC - DETR 的田间环境苜蓿种荚成熟度判别方法

于镇伟[1]　田富洋[1,2]　王　强[3]　闫银发[1]　张一诺[2]　宋占华[1]

(1. 山东农业大学机械与电子工程学院, 泰安 271018;
2. 智能农业机器人山东省高等学校未来产业工程研究中心, 泰安 271018;
3. 中国农业机械化科学研究院呼和浩特分院有限公司, 呼和浩特 010010)

摘要: 针对确定苜蓿种子适收期时存在种荚识别效率低、种荚成熟度判断模型泛化能力差等问题,提出了一种苜蓿种荚成熟度判别模型(VSC - DETR),可精准确定田间苜蓿种子成熟度等级。在 RT - DETR(Real-time detection Transformer)模型基础上,利用 VanillaNet 网络替代 RT - DETR 主干网络,提升模型在复杂自然环境下识别苜蓿种荚的计算效率和泛化能力;设计了新型卷积神经网络模块以增强检测苜蓿种荚小目标的能力,将尺度特征交互模块中的多头注意力替换为卷积加性标记混合器模块,优化模型对全局信息的理解;利用多元分支模块改进了 RepC3(Re-parameterized cross stage partial network with 3 convolutions),将下采样过程中的标准卷积替换为线性可变形卷积,提升了苜蓿种荚成熟度特征提取的准确性和鲁棒性。试验结果表明,VSC - DETR 的 mAP@0.5 达 94.23%,平均检测时间为 88.3 ms。相较于 RT - DETR,精确率、召回率和 mAP@0.5 分别提升 7.65、14.2、11.36 个百分点,且平均检测时间减少 20.4 ms。该研究结果能够快速、准确地识别不同成熟度的苜蓿种荚,可为苜蓿种子适收期的确定提供技术支撑。

关键词: 苜蓿种荚; 成熟度; 小目标识别; 田间环境; VSC - DETR

中图分类号: S572; TP391.4　　**文献标识码**: A　　**文章编号**: 1000-1298(2025)11-0084-09　　**OSID**:

Method for Determining Maturity of Alfalfa Seed Pods in Field Based on VSC - DETR

YU Zhenwei[1]　TIAN Fuyang[1,2]　WANG Qiang[3]　YAN Yinfa[1]　ZHANG Yinuo[2]　SONG Zhanhua[1]

(1. *College of Mechanical and Electronic Engineering*, *Shandong Agricultural University*, *Taian* 271018, *China*
2. *Shandong Higher Education Institution Future Industry Engineering Research Center of Intelligent Agricultural Robots*, *Taian* 271018, *China*
3. *Hohhot Branch of Chinese Academy of Agricultural Mechanization Sciences Co., Ltd.*, *Hohhot* 010010, *China*)

Abstract: Uneven maturity of alfalfa seeds hinders to determination of the optimal harvesting time. There are several algorithms used to assess the maturity level of alfalfa seed; however, existing algorithms are inefficient under natural conditions due to low automation and poor generalization capabilities. An efficient alfalfa seed pod maturity identification model (VSC - DETR) that can accurately determine the optimal harvest time was proposed. The proposed real-time detection Transformer (RT - DETR) model enhanced computational efficiency and generalization ability by replacing the main network of RT - DETR with a VanillaNet network, improving the model's ability to identify seed pods in complex natural environments. In addition, the SPDConv CSP - OmmiKernel (SPOM) module was designed to enhance the model's detection capability for small alfalfa seed pod targets. By replacing the multi-head attention in the scale feature interaction module (Attention-based cross-scale feature interaction, AIFI) with the convolutional additive token mixer (CATM) module, the model's understanding of global information was optimized. Additionally, the reparameterized cross-stage partial network with three convolutional layers (RepC3) was optimized by introducing the diversity branch block (DBB) module, and standard convolutions were replaced with linear deformable convolutions during down sampling to improve

收稿日期: 2025 - 08 - 26　修回日期: 2025 - 08 - 29
基金项目: 国家重点研发计划项目(2022YFD2001902)
作者简介: 于镇伟(1992—),男,副教授,博士,主要从事智能农机装备研究,E-mail: zhenweiyu615@126.com
通信作者: 宋占华(1984—),男,教授,博士,主要从事智能农机装备研究,E-mail: songzh@sdau.edu.cn

accuracy. The experimental results showed that the mAP@0.5 of VSC－DETR reached 94.23%, with average detection time of 88.3 ms. Compared with RT－DETR, the precision, recall, and mAP@0.5 were improved by 7.65, 14.2, and 11.36 percentage points, respectively, while the average detection time was decreased by 20.4 ms. This research achievement enabled the quick and accurate identification of alfalfa seed pods of different maturities, providing strong support for selecting high-quality seeds, thereby improving seed germination rates and production efficiency.

Key words: alfalfa seed pods; maturity; small target recognition; field environment; VSC－DETR

0　引言

苜蓿作为优质牧草,是重要的饲料作物,但其种子产量及质量与美国、加拿大等发达国家相比还存在较大差距[1-3]。苜蓿种子适收期的准确判断是影响苜蓿种子产量、品质及后续利用的关键环节。在黄熟期末至完熟初期收获,可最大限度减少田间损失和加工损耗,同时保证种子具有高发芽率、强活力及优良的商品性状,为苜蓿种子生产的经济效益和后续种植的建植成功提供关键保障[4-5]。然而,当前苜蓿种子适收期仍然依靠人工局部观察进行判断。这种方法不仅费时费力,还会受到大田种子成熟不均匀和缺乏客观量化标准影响,导致判断结果不准确。因此,基于苜蓿种荚成熟度判断模型建立客观的种子适收期评价体系,对提高种子产量、品质,保持饲草产业稳定发展具有重要意义[6]。

计算机视觉与深度学习技术在农业目标识别领域展现出巨大潜力,这为苜蓿种荚识别及成熟度判断提供了新的研究方向[7-9]。以卷积神经网络为代表的深度学习算法被广泛应用于多种作物的成熟度判别[10-13]。已有相关学者开展作物成熟评价中成熟度识别判断及分类相关研究[14-15],同时也取得了一些成果。黄健熙等[16]基于归一化植被指数对小麦成熟情况进行预测,能较好地实现小麦全程生长发育及成熟度动态检测。DIGMAN 等[17]利用近红外反射光谱仪和支持向量机预测青豌豆成熟度。石文强[18]基于光谱反射率、影像信息的颜色和纹理特征,建立了大豆成熟度识别模型,模型识别率为 100%。TIAN 等[19]基于无人机航拍图像和改进的轻量级 MobileViT 对苜蓿种荚成熟度智能判别方法进行研究,判断苜蓿种荚的青熟期、黄熟期、完熟期,判断精度为 92.23%,然而田间环境对其成熟度判别模型的轻量化方法和判别准确率都提出了更高要求。

针对上述问题,本文提出一种苜蓿种荚识别及成熟度判别模型(VSC－DETR),基于 RT－DETR 网络增添新型卷积神经网络模块,提高对苜蓿种荚小目标的敏感程度;并将其多头注意力机制更换为卷积加性标记混合器,提升模型对细节特征的识别能力;同时,将原骨干网络更换为 VanillaNet 模块,增强模型泛化能力,以实现轻量化。最后开展田间试验,测试模型准确度和鲁棒性。

1　材料与方法

1.1　数据集构建

1.1.1　图像采集

于 2024 年 7 月 5—25 日(每日 09:00—11:00、15:00—18:00 采集 2 次)在内蒙古自治区鄂尔多斯市鄂托克旗乌兰镇赛乌素草籽场(39°0′34″N,107°17′01″E)进行数据采集。苜蓿品种为中苜 3 号,标准化种植模式,种植面积 33 hm^2。采用自主设计的自走式图像采集系统获取苜蓿种荚图像,如图 1 所示。该装置主要由相机、存储与处理单元、驱动装置、电源和支架等部分组成。为满足专用苜蓿种子收获机需求,相机的安装高度为 3.0～3.5 m,拍摄角度为 45°俯拍。为确保图像数据的准确性和可靠性,采用五点取样法采取图像。每个区域均长 15 m、宽 2 m,并将每个区域划分成 5 个单元格(3 m×2 m)。

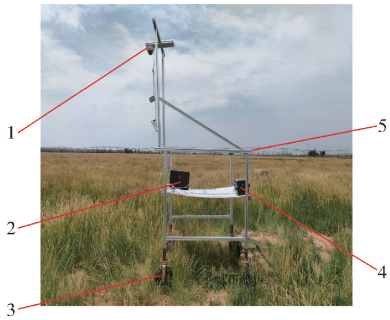


图 1　图像采集装置

Fig. 1　Image acquisition device

1. 相机　2. 存储与处理单元　3. 驱动装置　4. 电源　5. 支架

1.1.2　数据集制备

苜蓿种荚生长期颜色转变如图 2 所示。在青熟期,苜蓿种荚的 RGB 值范围为 0～100(R)、120～255(G)、0～100(B);在黄熟期,种荚开始转变成灰色,其 RGB 值范围为 100～150(R)、100～150(G)、100～150(B);在完熟期,种荚颜色转为黑色,其 RGB 值范围为 0～50(R)、0～50(G)、0～50(B)。

为增强模型泛化能力,采用平均哈希算法对图像进行预处理,去除相似度超过 95% 的图像,得到 1 354 幅苜蓿种荚图像(4 000 像素×3 000 像素)。



(a) 绿色 (b) 灰色 (c) 黑色

图 2 不同成熟期的苜蓿种荚

Fig. 2 Alfalfa seed pods at different stages of maturity

使用 LabelImg 工具进行标注,并选择 Mixup 算法增强图像,最终构成苜蓿种荚数据集(5 416 幅图像、108 320 个标注框)。按照比例 8∶1∶1将其划分为训练集、测试集和验证集。

1.2 基于 VSC - DETR 的苜蓿种荚成熟度识别模型

本研究提出一种改进 RT - DETR 的苜蓿种荚识别及成熟度判别模型 VSC - DETR,该模型主要由主干网络、高效混合编码器和带有辅助预测头的 Transformer 解码器组成。使用 VanillaNet 主干网络代替 RT - DETR 主干网络,提高模型的计算效率和泛化能力;设计 SPOM 模块(SPDConv CSP - OmmiKernel)提升小目标检测能力,并通过将 AIFI 模块中的多头注意力替换为 CATM(Convolutional additive token mixer)模块,增强模型对全局信息的理解。此外,为提升模型对细微特征的感知,利用 DBB 模块对 RepC3 进行改进,并将下采样过程中的标准卷积替换为线性可变形卷积 LDConv,优化特征提取的准确性和鲁棒性;将解码器中 IoU 感知查询模块的损失函数 GIoU 替换为 Ins - IoU,以提高目标边界形状一致性的精确度。VSC - DETR 模型结构如图 3 所示。

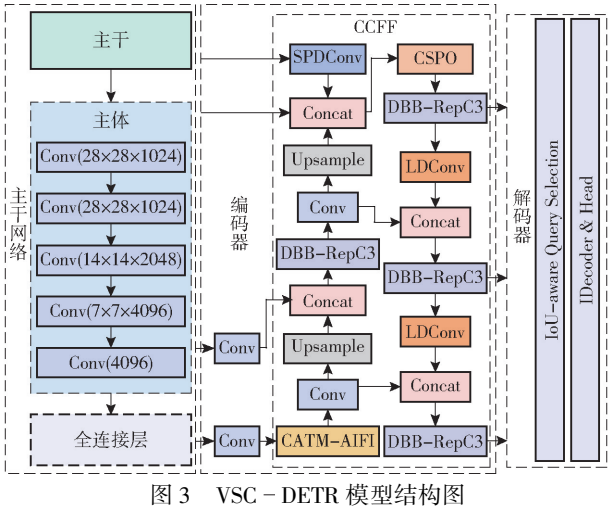


图 3 VSC - DETR 模型结构图

Fig. 3 Model structure diagram of VSC - DETR

1.2.1 VanillaNet 主干架构

为减少模型参数量和计算量,将 RT - DETR 主干网络更换为一种极简的神经网络架构 VanillaNet[20]。使用步幅为 2 的最大池化层来减小尺寸和特征图,采用平均池化层计算特定区域内的平均值来减少特征图尺寸,在每个 1×1 卷积层后添加激活函数,增加模型的非线性能力,提高种荚检测实时性。

1.2.2 SPOM 模型

为增强模型对小目标(种荚)特征的提取能力,设计一种针对小目标的特征金字塔结构。SPOM 模型对特征信息的提取过程如图 4 所示。对 P2 层增加 160×160 的特征图输出,融合 OmmiKernel 算法、CSP 思想和 SPD 模块降低模型复杂度,确保算法能适应不同尺寸和形态的目标,也能更好地处理种荚图像的噪声或遮挡,提高种荚检测鲁棒性。

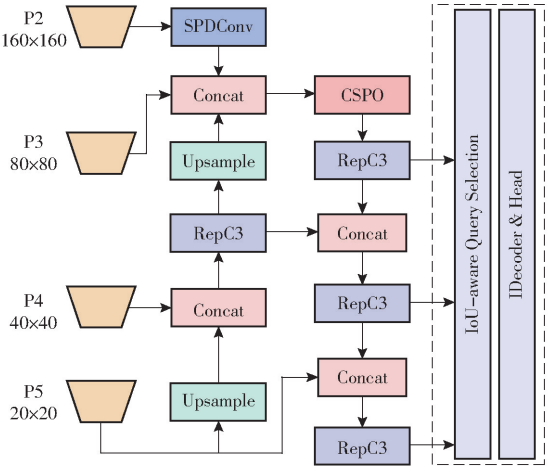


图 4 SPOM 模型结构图

Fig. 4 SPOM model structure diagram

1.2.3 CATM - AIFI

为增强苜蓿种荚细粒度特征的捕捉和全局建模能力,将 AIFI 中的多头注意力机制(Multi-head attention)更换为 CATM(Convolutional additive token mixer)模块,提出轻量高效的内尺度特征交互模块 CATM - AIFI,提高对密集种荚的检测精度。

1.2.4 下采样算子

改进下采样中标准卷积,替换为线性可变形卷积 LDConv[21],可有效地融合目标不同尺度和位置信息,使特征图保留更多有用的空间和语义信息。同时,其空间变形机制可使感受野在空间上自适应调整,从而捕捉种荚的复杂形状和结构。

1.2.5 Ins - IoU

VSC - DETR 模型的损失函数为分类损失[22]与边界框回归损失[23]之和。采用 Inner - Shape - IoU(Ins - IoU)损失函数衡量边界框回归损失。根据不同的检测器和检测任务进行自我调整,提高模型泛化能力。

2 结果与分析

2.1 模型评价指标

为验证 VSC - DETR 模型性能,本研究选择精

确率、召回率和平均精度均值(mean Average Precision,mAP)作为模型性能的评价指标,mAP@0.5 为在单一交并比阈值为 0.5 时计算得到 mAP,mAP@[0.5:0.95]为把 IoU 阈值从 0.5 开始,以 0.05 为步长递增到 0.95,在每个阈值下分别计算一次 mAP 并取平均值;选择参数量、模型内存占用量和浮点运算数作为模型复杂程度的评价指标;以平均检测时间作为模型推理速度的评价指标。

2.2　训练平台与参数配置

该训练平台搭载 64 位 Ubuntu 22.10 操作系统,Python 3.9.19、Pytorch 1.13.1、torchvision 0.14.1 和 CUDA 11.7 搭建了深度学习环境,批次大小设置为 4,迭代次数设置为 300。

2.3　不同主干模型对比

为分析不同主干模型对苜蓿种荚识别的影响,以 VCS－DETR 为基准网络,采用不同主干模型进行特征提取,基于各评价指标验证模型性能。选取的主干模型有:RepViT[24]、StarNet[25] 和 MobileNetV4[26] 等。使用不同主干模型的性能如表 1 所示。

表 1　不同主干模型性能

Tab.1　Performance of different backbone models

主干模型	精确率/%	召回率/%	mAP@0.5/%	mAP@[0.5:0.95]/%	平均检测时间/ms	模型内存占用量/MB
RepViT	70.97	56.90	62.67	37.87	92.5	46.1
StarNet	74.41	62.81	70.39	44.63	89.8	43.1
MobileNetV4	85.52	65.21	72.12	46.68	89.1	41.8
SwinTransformer	55.51	45.24	47.63	29.65	95.9	92.0
EfficientFormerV2	94.33	83.54	93.40	64.38	96.6	94.1
VanillaNet	95.46	91.12	94.23	68.91	88.3	74.2

采用 VanillaNet 主干模型的精确率、召回率、mAP@0.5 和 mAP@[0.5:0.95]均最高,分别为 95.46%、91.12%、94.23% 和 68.91%。单幅图像检测时间为 88.3 ms,是所有模型中最短的。这表明 VanillaNet 模型的特征提取能力较强、实时性较好,适用于检测效率要求较高的应用场景。

使用 RepViT、StarNet 和 MobileNetV4 3 种主干的模型内存占用量分别为 46.1、43.1、41.8 MB,相较于 VanillaNet 主干网络,分别降低了 28.1、31.1、32.4 MB。然而,与采用 VanillaNet 主干网络的模型相比,其余主干网络在性能和速度上仍存在一定差距,其中性能最好的是 MobileNetV4,其精确率、召回率和 mAP@0.5 分别为 85.52%、65.21%、72.12%,仍远低于 VanillaNet 主干模型。这表明尽管这 3 种主干在减小模型尺寸方面表现出色,但在检测精度方面表现不足,未能有效捕捉到识别目标的特征信息。

融合 SwinTransformer[27] 主干的模型在性能上表现最差,精确率、召回率和 mAP@0.5 分别为 55.51%、45.24%、47.63%。使用 EfficientFormerV2[28] 的模型在精确率和 mAP@0.5 上表现较好,但召回率和 mAP@[0.5:0.95]分别比 VanillaNet 主干模型低 7.58、4.53 个百分点,模型内存占用量和平均检测时间也最差。这说明使用 EfficientFormerV2 作为主干会产生漏检现象,并伴随较高的计算成本。

边界框回归损失用于衡量预测框与真实框之间的差异。损失曲线能反映模型的训练状态和改进合理性,使用不同主干模型的边界框损失曲线如图 5 所示。由图 5 可知随着训练轮次增加,各模型的定位置损失呈下降趋势。达到 280～300 轮时,损失曲线保持平稳,达到收敛状态。与其他模型相比,VanillaNet 主干模型的损失曲线较平稳,这表明在训练过程中稳定性较强,参数的迭代更平滑,未出现过拟合现象。

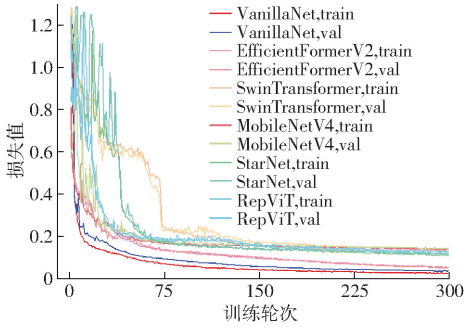


图 5　不同主干网络损失曲线

Fig.5　Loss curves using different backbone networks

综上所述,在检测精度方面,VanillaNet 主干模型表现最好,mAP@0.5 达到 94.23%;在模型内存占用量方面,MobileNetV4 主干模型最小;在检测速度方面,VanillaNet 主干处理单幅图像的时间最少。因此,对于苜蓿种荚的识别任务,VanillaNet 作为主干网络为最佳选择。

2.4　不同 AIFI 的模型对比

为探究不同 AIFI 对苜蓿种荚识别的影响,对 AIFI 进行不同改进并验证模型性能。训练结果如表 2 所示。

表 2 不同改进 AIFI 模型训练结果

Tab. 2 Training results by using different AIFI models

模型	精确率/%	召回率/%	mAP@0.5/%	mAP@[0.5:0.95]/%	参数量	浮点运算数
DAT - AIFI	93.23	86.71	91.46	66.85	2.784×10^7	1.903×10^{11}
EAA - AIFI	85.74	77.60	82.92	52.97	2.783×10^7	1.903×10^{11}
CGA - AIFI	84.68	80.82	81.23	47.78	2.767×10^7	1.901×10^{11}
HiLo - AIFI	92.14	84.31	88.61	61.61	2.780×10^7	1.902×10^{11}
CATM - AIFI	95.46	91.12	94.23	68.91	2.791×10^7	1.902×10^{11}

由表 2 可知,在精确率方面,使用 DAT - AIFI、HiLo - AIFI 和 CTAM - AIFI 的性能较突出,分别为 93.23%、92.14% 和 95.46%,表明这 3 种 AIFI 对苜蓿种荚的特征信息更加敏感,能够有效捕捉种荚的细节特征,降低模型误检率。

在召回率方面,与 CATM - AIFI 模型相比,采用 DAT - AIFI、EAA - AIFI、CGA - AIFI 和 HiLo - AIFI 模型分别下降 4.41、13.52、10.3、6.81 个百分点。这表明通过 CATM 模块改进 AIFI 能提高苜蓿种荚检出率,有效减少漏检。

在 mAP@0.5 方面,使用 DAT - AIFI 和 CATM - AIFI 的模型表现最好,分别为 91.46% 和 94.23%,这表明采用这 2 种模块可以显著提升模型的检测精度,能更准确地识别和定位苜蓿种荚,从而具有更高的应用价值。这 2 种模型的 mAP@[0.5:0.95]在所有模型中性能最为突出,分别为 66.85% 和 68.91%,这说明融合 DAT - AIFI 和 CATM - AIFI 模块的模型可以在更精确位置上拟合目标边界,进一步提升了检测准确性和可靠性。

综上所述,相较于其他 AIFI 模型,采用 CATM - AIFI 改进的 VSC - DETR 模型性能提升更为显著,且没有额外增加参数量和浮点运算数,因此该模型在苜蓿种荚识别的效果上更加突出。根据图 6 中损失函数曲线可以看出,融合 CATM - AIFI 的模型在训练中可以高效且稳定地收敛至较低损失值。与其他模型相比,鲁棒性更强,在田间环境中的适应能力更强。

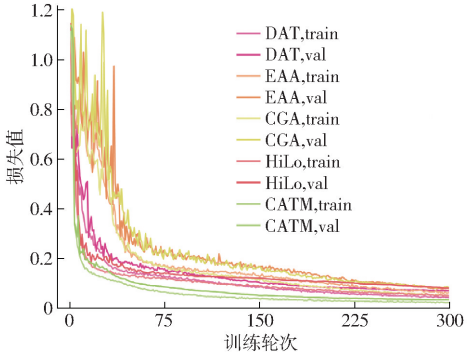


图 6 不同 AIFI 模块损失曲线

Fig. 6 Loss curves for different AIFI modules

2.5 不同 RepC3 模型对比

为探究不同模块改进 RepC3 对苜蓿种荚识别的影响,融合不同 RepC3 模型进行比较,各模型性能如表 3 所示。

表 3 不同 RepC3 模型训练结果

Tab. 3 Training results of models by using different RepC3s

模型	精确率/%	召回率/%	mAP@0.5/%	mAP@[0.5:0.95]/%	参数量	浮点运算数
DWR - RepC3	93.91	86.76	91.94	64.77	2.791×10^7	1.902×10^{11}
DGC - RepC3	92.57	83.54	89.28	61.06	2.657×10^7	1.835×10^{11}
DRB - RepC3	91.36	84.17	88.59	60.92	2.618×10^7	1.815×10^{11}
RET - RepC3	93.54	84.56	90.47	59.24	2.656×10^7	1.834×10^{11}
DBB - RepC3	95.46	91.12	94.23	68.91	2.791×10^7	1.902×10^{11}

融合 DWR - RepC3 模块的模型在精确率、mAP@0.5 和mAP@[0.5:0.95]上表现良好,分别为 93.91%、91.94%、64.77%。而召回率表现较差,仅有 86.76%。这说明该模型的多尺度特征融合能力受限,未能充分提取苜蓿种荚的特征信息,出现漏检现象。与其他模型相比,使用 DGC - RepC3 和 DRB - RepC3 改进模型的性能较差,这说明此改进方式不适用识别苜蓿种荚这类小目标。与融合 DWR - RepC3 模块的模型相比,采用 RET - RepC3 改进的模型性能差距较小,精确率、召回率、mAP@0.5 和 mAP@[0.5:0.95]分别降低 0.3、2.2、1.47、5.53 个百分点。基于 DBB - RepC3 改进模型的性能指标表现最好,特别在召回率方面,达到 91.12%,显著优于其他模型。这说明该模块能增强捕捉苜蓿种荚细粒度特征能力,有效改善漏检状况。

综合考虑模型的性能和复杂度,基于 DBB - RepC3 改进模型在参数量和浮点运算数变化较小情况下,达到最高检测精度。因此,对于苜蓿种荚成熟

度识别，采用 DBB－RepC3 的改进方法最适宜。

2.6　与现有目标检测模型对比

为探究不同目标检测模型对苜蓿种荚识别的影响，本研究在试验条件和数据集相同的情况下，根据精确率、召回率、mAP@0.5、浮点运算数、参数量和模型内存占用量等评价指标进行比较。除 RT－DETR 外，选取的目标检测网络为 RetinaNet、Faster R－CNN、YOLO v8。各模型训练结果如表 4 所示。

表 4　VSC－DETR 与现有目标检测模型训练结果

Tab. 4　Training results of VSC－DETR and existing object detection models

模型	精确率/%	召回率/%	mAP@0.5/%	参数量	平均检测时间/ms	模型内存占用量/MB
RetinaNet	67.32	52.07	55.74	3.637×10^7	92.55	145.8
Faster R－CNN		68.04	56.32		280.47	310.9
YOLO v8	85.97	74.51	80.53	1.114×10^7	126.84	24.6
RT－DETR－R50	89.67	78.38	84.93	4.196×10^7	64.00	86.1
RT－DETR－R18	87.81	76.92	82.87	1.988×10^7	108.7	40.5
VSC－DETR	95.46	91.12	94.23	2.791×10^7	88.3	74.2

由表 4 可知，RT－DETR－R18 模型的召回率较低，仅为 76.92%，VSC－DETR 模型与其相比，精确率、召回率和 mAP@0.5 都得到明显提升，分别提升 7.65、14.2、11.36 个百分点。在模型复杂程度方面，RT－DETR－R18 的模型内存占用量和参数量分别为 40.5 MB 和 1.988×10^7，均低于 VSC－DETR 模型。在检测效率方面，RT－DETR－R18 模型的平均检测时间为 108.7 ms，比 VSC－DETR 模型检测所需时间长。这表明虽然 RT－DETR－R18 在模型复杂度上表现较为轻量，但其在检测精度和速度上仍存在提升空间。

RetinaNet 在召回率上表现最差，仅为 52.07%，相比其他模型会产生更多的漏检。与 VSC－DETR 相比，精确率、召回率和 mAP@0.5 分别降低 28.14、39.05、38.49 个百分点，且模型内存占用量增加近 2 倍。这说明 RetinaNet 并不适合苜蓿种荚这类小目标的检测任务。

Faster R－CNN 模型的平均检测时间和内存占用量均为最高，分别为 280.47 ms 和 310.9 MB，其检测速度和效率可能会受到影响。VSC－DETR 模型与之相比，在召回率和 mAP@0.5 分别提升 23.08、37.91 个百分点的条件下，模型内存占用量降低约 76.13%，平均检测时间减少 192.17 ms。这表明 Faster R－CNN 在模型轻量化和检测效率方面存在明显瓶颈，制约了其在实时性部署中的应用表现。

RT－DETR－R50 精确率为 89.67%，mAP@0.5 为 84.93%，仅次于 VSC－DETR 模型，但 RT－DETR－R50 的参数量比 VSC－DETR 高出 1 倍多。VSC－DETR 与之相比，精确率提升 5.79 个百分点，召回率提升 12.74 个百分点，模型内存占用量降低 11.9 MB。同时，其较大的参数量和模型内存占用量限制了它在资源受限环境中的应用。

YOLO v8 在参数量和模型内存占用量上表现最好，分别为 1.114×10^7 和 24.6 MB。但在性能方面，VSC－DETR 与之相比，精确率、召回率和 mAP@0.5 分别提高 9.49、16.61、13.7 个百分点，而平均检测时间减少 38.54 ms。这说明 YOLO v8 具有较小的模型规模和较低的计算资源需求，适合在计算能力有限的边缘设备上进行部署，但在精度和效率上仍需进一步优化。

综上所述，本研究所提出的 VSC－DETR 模型在性能上具有最好的表现，但在召回率、模型内存占用量和平均检测时间上仍存在提升空间，因此后续对 VCS－DETR 进行剪枝蒸馏操作。各模型的损失曲线如图 7 所示，与其他模型相比，VSC－DETR 模型能够以最快的速度收敛至最低损失值，说明该模型在预测目标的位置和尺寸方面表现出更高的准确性，对数据的拟合能力更强。

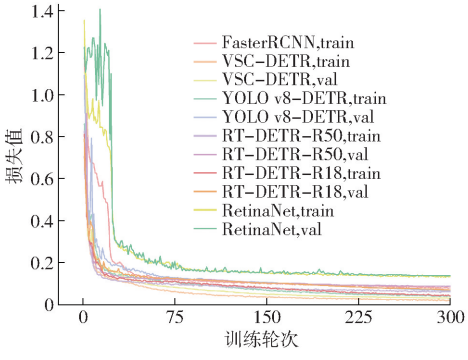


图 7　模型损失曲线

Fig. 7　Loss curves of models

2.7　消融试验

为验证不同改进方法对模型整体性能的影响，以 RT－DETR 为基准模型，设计了消融试验，通过使用不同模块检验改进方法的有效性，并选用相同的苜蓿种荚数据集进行训练，试验结果如表 5 所示。

表 5 消融试验结果

Tab. 5 Ablation test results %

RT－DETR	VanillaNet	SPOM	CATM－AIFI	LDConv	DBB－RepC3	Ins－IoU	精确率	召回率	mAP@0.5
√							87.81	76.92	82.87
√	√						90.49	80.24	87.15
√		√					94.98	85.58	91.18
√			√				89.74	80.29	86.69
√				√			87.32	77.64	83.34
√					√		88.07	80.25	85.41
√	√	√					94.25	88.18	92.51
√	√	√	√				95.27	89.63	93.16
√	√	√	√	√			95.26	90.17	93.61
√	√	√	√	√	√		95.14	90.89	93.86
√	√	√	√	√	√	√	95.46	91.12	94.23

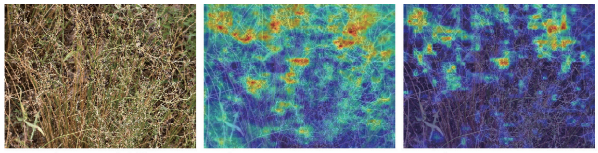
注:"√"表示添加此模块。

由表 5 可以看出,在基准网络上添加 SPOM 模块后,模型的识别效果得到有效改进,精确率、召回率和 mAP@0.5 分别提升 7.17、8.66、8.31 个百分点,这说明使用该模块后能有效提高模型对小目标的特征感知能力,帮助网络更加准确地定位和识别苜蓿种荚,但融合 SPOM 模块后的召回率仅为 85.58%,可能会导致漏检情况的发生。将 LDConv 模块加入基准网络中,精确率下降 0.49 个百分点,召回率、mAP@0.5 分别提高 0.72、0.47 个百分点,说明单独添加该模块可能会出现误检情况。以 RT－DETR 为基准网络,单独增加 CATM－AIFI 和 DBB－RepC3 模块,精确率分别提高 1.93、0.26 个百分点,召回率提高 3.37、3.33 个百分点,mAP@0.5 分别提升 3.82、2.54 个百分点。在 RT－DETR 的基础上,将主干模型替换为 VanillaNet,并添加 SPOM 模块,精确率、召回率和 mAP@0.5 分别提高 6.44、11.26、9.64 个百分点,这表明同时添加这 2 种模块,能够增强模型在复杂场景中的特征提取能力;在上述条件下,添加 CATM－AIFI、DBB－RepC3 和 LDConv 模块,与 RT－DETR 模型相比较,精确率、召回率、mAP@0.5 分别提升 7.33、13.97、10.99 个百分点。

2.8 热力图可视化

根据大田复杂环境下苜蓿种荚特性,选择 HiResCAM 对模型的识别效果进行可视化展示。改进前后模型的热力图对比结果如图 8 所示。

由图 8b 可以看出,改进前模型在识别苜蓿种荚时,不仅关注待检测目标,还聚焦于背景和其他目标,图 8c 则展示了基于 RT－DETR 改进后的 VSC－DETR 模型的热力图可视化结果,模型改进后降低了对背景和非检测目标的重视程度,能够更加准确地聚焦于苜蓿种荚的关键特征,对于样本有正向影响的像素点,激活区域更广,响应强度更大,充分解释了模型的检测机制。



(a) RGB图像 (b) RT-DETR-R18 (c) VSC-DETR

图 8 改进前后模型热力图

Fig. 8 Heat maps of before and after improvement of model

3 结论

(1)提出的苜蓿种荚识别及成熟度判别模型 VSC－DETR,在 RT－DETR 网络的基础上增添 SPOM 模块,提高了其对苜蓿种荚这类小目标的敏感程度;对 AIFI 模块进行改进,将其多头注意力机制更换为 CATM 模块;将原骨干网络更换为极简的神经网络架构 VanillaNet;在编码器中引入 LDConv 和 DBB－RepC3 模块;将原损失函数替换为 Ins－IoU。

(2)改进后模型精确率和召回率分别为 95.46%和 91.12%,mAP@0.5 达到 94.23%,平均检测时间为 88.3 ms。相较于原始模型和现有研究,各项评价指标均得到提升,可为苜蓿种子适收期的确定提供技术支持。

(3)通过 HiResCAM 对 VSC－DETR 模型进行热力图可视化分析,该模型能够更有效地捕捉到重要的特征或区域,且对不同输入的响应更加精确,进而提高了模型的可解释性和检测精度。

参 考 文 献

[1] 潘烤鑫, 张庆, 尤泳, 等. 小粒径苜蓿种子机械梳脱增流式收获机设计与试验[J]. 农业机械学报, 2025, 56(2): 342－356.

PAN Kaoxin, ZHANG Qing, YOU Yong, et al. Design and experiment of flow-enhanced mechanical combining harvester for small-seeded alfalfa[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(2): 342－356. (in Chinese)

[2] 刘伟, 李凤鸣, 李伟, 等. 牧草种子收获机研究现状与展望[J]. 农业工程, 2021, 11(7): 17－20.
LIU Wei, LI Fengming, LI Wei, et al. Research status and prospect of forage grass seed harvester[J]. Agricultural Engineering, 2021, 11(7): 17－20. (in Chinese)

[3] 李涛, 王振华, 翟改霞, 等. 苜蓿草植株离散元模型建立与参数标定[J]. 农业机械学报, 2024, 55(增刊2): 187－199.
LI Tao, WANG Zhenhua, ZHAI Gaixia, et al. Establishment and parameter calibration of discrete element model of alfalfa plant[J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(Supp.2): 187－199. (in Chinese)

[4] ZHANG X, WANG H, WANG F, et al. Parameter calibration of discrete element model for alfalfa seeds based on EDEM simulation experiments[J]. International Journal of Agricultural and Biological Engineering, 2024, 17(3): 33－38.

[5] 王树栓, 石凤翎. 野生黄花苜蓿种子生长发育动态分析[J]. 植物生理学报, 2025, 61(7): 962－972.
WANG Shushuan, SHI Fengxiang. Study on the growth and development dynamics of wild *Medicago falcata* seeds[J]. Plant Physiology Journal, 2025, 61(7): 962－972. (in Chinese)

[6] 王德成, 贺长彬, 武红剑, 等. 苜蓿生产全程机械化技术研究现状与发展分析[J]. 农业机械学报, 2017, 48(8): 1－25.
WANG Decheng, HE Changbin, WU Hongjian, et al. Review of alfalfa full-mechanized production technology[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(8): 1－25. (in Chinese)

[7] YANG S, JIA Z, YI K, et al. Rapid prediction and visualization of safe moisture content in alfalfa seeds based on multispectral imaging technology[J]. Industrial Crops and Products, 2024, 222: 119448.

[8] WANG X, ZHANG H, SONG R, et al. Non-destructive identification of naturally aged alfalfa seeds via multispectral imaging analysis[J]. Sensors, 2021,17: 5804.

[9] ZHANG L, ZHANG Q, WU J, et al. Moisture detection of single corn seed based on hyperspectral imaging and deep learning[J]. Infrared Physics and Technology, 2022,125: 104279.

[10] 刘艺豪, 周名扬, 陈度, 等. 基于深度学习和视觉显著性算法的自然环境中烟叶成熟度判别[J]. 农业机械学报, 2025, 56(7): 502－512.
LIU Yihao, ZHOU Mingyang, CHEN Du, et al. Maturity discrimination of tobacco leaves in natural environments based on deep learning and visual saliency algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(7): 502－512. (in Chinese)

[11] 吴尘萱, 栗君谊, 韩亚芬, 等. 面向植物合成生物学的植物工厂生产关键技术与装备研究进展及应用展望[J]. 农业机械学报, 2025, 56(7): 300－315.
WU Chenxuan, LI Junyi, HAN Yafen, et al. Research progress and application prospects of key technologies and equipment for plant factory production in plant synthetic biology[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(7): 300－315. (in Chinese)

[12] FAISAL M, ALSULAIMAN M, ARAFAH M, et al. IHDS: intelligent harvesting decision system for date fruit based on maturity stage using deep learning and computer vision[J]. IEEE Access, 2020, 8: 167985－167997.

[13] THORIQ A I, ZUHRI M H, PURWANTO P, et al. Classification of banana maturity levels based on skin image with HSI color space transformation features using the KNN method[J]. Journal of Development Research, 2022, 6(1): 11－15.

[14] 胡心雨, 叶惠源, 向东, 等. 基于无人机多光谱技术的田间烟叶成熟度判别研究[J]. 智慧农业导刊, 2025, 5(8): 1－5.
HU Xinyu, YE Huiyuan, XIANG Dong, et al. Maturity determination of tobacco leaves in the field based on UAV multi-spectral technology[J]. Journal of Smart Agriculture, 2025, 5(8): 1－5. (in Chinese)

[15] WEI X, WU L, GE D, et al. Prediction of the maturity of greenhouse grapes based on imaging technology[J]. Plant Phenomics, 2022, 2022: 9753427.

[16] 黄健熙, 牛文豪, 马鸿元, 等. 基于时间序列 HJ－1 A/B 卫星数据的冬小麦成熟期预测[J]. 农业机械学报, 2016, 47(11): 278－284.
HUANG Jianxi, NIU Wenhao, MA Hongyuan, et al. Prediction of maturity data for winter wheat based on time series of HJ－1 A/B CCD images[J]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(11): 278－284. (in Chinese)

[17] DIGMAN M F, RUNGE W M. The utility of a near-infrared spectrometer to predict the maturity of green peas (*Pisum sativum*)[J]. Computers and Electronics in Agriculture, 2022, 193: 106643.

[18] 石文强 . 大豆成熟度智能识别与分区研究[D]. 大庆: 黑龙江八一农垦大学, 2022.
SHI Wenqiang. Study on intelligent recognition and zoning of soybean maturity[D]. Daqing: Heilongjiang Bayi Agricultural University, 2022. (in Chinese)

[19] TIAN F, ZHANG Y, SOOMRO S A, et al. SSOD－MViT: a novel model for recognizing alfalfa seed pod maturity based on semi-supervised learning[J]. Computers and Electronics in Agriculture, 2025, 236:110439.

[20] TAO Z, HE Y, LIN S, et al. SnakeNet:an adaptive network for small object and complex background for insulator surface

defect detection[J]. Computers and Electrical Engineering, 2024, 117: 109259.

[21] WANG F, JIANG X, HAN Y, et al. YOLO-LSDI: an enhanced algorithm for steel surface defect detection using a YOLOv11 network[J]. Electronics, 2025, 14(13): 2576.

[22] 廖彬, 张陶, 于炯, 等. NLGAE: 一种基于改进网络结构及损失函数的图自编码器节点分类模型[J]. 计算机科学, 2024, 51(10): 234-246.
LIAO Bin, ZHANG Tao, YU Jiong, et al. NLGAE: a graph autoencoder model based on improved network structure and loss function for node classification task[J]. Computer Science, 2024, 51(10): 234-246. (in Chinese)

[23] 朱雨, 郝晓丽, 牛保宁, 等. RetinaNet 目标检测算法改进[J]. 计算机工程与设计, 2024, 45(11): 3271-3278.
ZHU Yu, HAO Xiaoli, NIU Baoning, et al. Improvement of RetinaNet object detection[J]. Computer Engineering and Design, 2024, 45(11): 3271-3278. (in Chinese)

[24] WANG S, JIANG H, YANG J, et al. Lightweight tomato ripeness detection algorithm based on the improved RT-DETR[J]. Frontiers in Plant Science, 2024, 15: 1415297.

[25] MA L, ZHAO L, WANG Z, et al. Detection and counting of small target apples under complicated environments by using improved YOLOv7-tiny[J]. Agronomy, 2024, 13(5): 1419.

[26] QIN J, ZHOU H, YI H, et al. YOLOv7-GCM: a detection algorithm for creek waste based on improved YOLOv7 model[J]. Pattern Analysis and Applications, 2024, 27(4): 116.

[27] 季长清, 曹思雨, 李艳志, 等. 基于 SwinTransformer 的去雾算法在森林消防中的应用[J]. 消防科学与技术, 2025, 44(6): 839-845.
JI Changqing, CAO Siyu, LI Yanzhi, et al. Application of dehazing methods based on SwimTransformer in forest fire[J]. Fire Science and Technology, 2025, 44(6): 839-845. (in Chinese)

[28] LI X, CAI M, TAN X, et al. An efficient transformer network for detecting multi-scale chicken in complex free-range farming environments via improved RT-DETR[J]. Computers and Electronics in Agriculture, 2024, 224: 109160.