文章编号：1008-1542（2026）00-0000-00

基于 FPB-DETR 的苹果成熟度检测算法

薛 婷，王震洲，孟志永，张秀清，杨 琳，邓 标

（河北科技大学信息科学与工程学院，河北石家庄 050018）

摘 要：针对当前苹果成熟度检测在大规模、光照和遮挡等情况时检测精度与效率较低的情况，提出一种基于 RT-DETR 改进的 FPB-DETR 检测模型。首先，在主干网络中引入频率自适应膨胀卷积（frequency-adaptive dilated convolution，FADC）模块，通过解决有效感受野与特征带宽的矛盾以及突破固定膨胀率的限制，精准聚焦苹果表面细微颜色渐变区、未成熟斑点及纹理条纹；其次，设计 Pola-AIFI（polaformer-attention-based intra-scale feature interaction）模块，解决负值忽略与信息熵过大的问题，抑制目标苹果在不同环境下的干扰；最后，在多尺度融合阶段引入双向特征金字塔网络（bi-directional feature pyramid network，BIFPN）结构，优化特征融合效率与关键信息聚焦能力，抑制成熟度特征传递中的歧义干扰。结果表明：本研究所提 FPB-DETR 模型的精确度、召回率和平均精度均值分别为 92.5%、92.7%和 96.8%，相比于原模型分别提升 2%、1.7%和 1.8%，均优于 Faster R-CNN、YOLOv5、YOLO v8、YOLO v11 以及 YOLO v12 目标检测模型，显著提升了模型检测能力；模型平均检测时间为 31 ms，满足对苹果成熟度实时检测的需求。本研究在特征提取、注意力机制与多尺度融合三方面实现创新，为智能采摘机器人更好的发展打下基础。

关键词：计算机神经网络；目标检测；成熟度；多尺度融合；RT-DETR

中图分类号：TP391.41 **文献标志码**：A

Apple maturity detection algorithm based on FPB-DETR

XUE Ting，WANG Zhenzhou，MENG Zhiyong，ZHANG Xiuqing，YANG Lin，DENG Biao

(School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang, Hebei 050018, China)

Abstract：To address the low accuracy and efficiency of apple maturity detection under large-scale, lighting, and occlusion conditions, an improved FPB-DETR detection model based on RT-DETR was proposed. Firstly, a frequency-adaptive dilated convolution (FADC) module was introduced into the backbone network to precisely focus on subtle color gradients, immature spots, and texture stripes on apple surfaces by resolving the conflict between effective receptive field and feature bandwidth, as well as overcoming the limitations of fixed dilation rates. Secondly, a polarity attention mechanism (polaformer-attention-based intra-scale feature interaction, Pola-AIFI) module was designed to mitigate the issues of negative value neglect and excessive information entropy, suppressing interference from target apples under varying environmental conditions. Finally, a bidirectional feature pyramid network (bi-directional feature pyramid network, BIFPN) structure was introduced during the multi-scale fusion stage to optimize feature fusion efficiency and key information focusing capability, reducing ambiguity interference in maturity feature transmission. The results show that the precision, recall rate and average accuracy of the FPB-DETR model proposed in this study are 92.5%, 92.7% and 96.8%, respectively, which increases by 2%, 1.7% and 1.8%, respectively compared with the original model, and are superior to those of Faster R-CNN, YOLOv5, YOLO v8, YOLO v11

收稿日期：2025-07-22；修回日期：2025-10-20；责任编辑：冯民
基金项目：国家自然科学基金(62441401)
第一作者简介：薛婷(2001—)，女，河北张家口人，硕士研究生，主要从事深度学习、图像处理方面的研究。
通信作者：孟志永，副教授。E-mail：26325940@qq.com

and YOLO v12 object detection models, significantly enhancing the detection capability of the model; The average detection time of the model is 31 ms, which meets the real-time detection requirements for apple maturity. This study demonstrates the effectiveness of innovations in three aspects: feature extraction, attention mechanism and multi-scale fusion, providing reference for the optimization design of intelligent harvesting robots.
Keywords：computer neural networks; object detection; maturity; multi-scale fusion; RT-DETR

中国作为全球最大的苹果生产国，2024 年产量高达 3734 万吨，然而出口量仅为 98 万吨，不足总产量的 3%，这一数据与中国苹果生产大国的地位极不匹配[1]。究其原因，主要在于适宜出口的优质果率偏低。因此，提升苹果品质已成为增强我国苹果产业国际竞争力的关键环节。在苹果产业链中，成熟度检测直接影响果实品质与商品价值，传统依赖人工判断的方式存在采摘周期长、标准不统一等问题，难以在最佳采收期内高效处理大规模果园，导致品质把控不足与资源浪费。现有采摘技术精度有限，误检率高，制约了优质果的筛选效果。

近些年来，基于深度学习的水果成熟度检测在农业领域得到了迅速发展，目前主流算法可分为基于 CNN（convolutional neural networks）和基于 Transformer[2]两大流派。CNN 检测算法包含一阶段与二阶段方法：前者以 YOLO（you only look once）[3-6]系列、SSD（single shot multibox detector）[7]、RetinaNet（retinanet: focal loss for dense object detection）[8]为代表，无需候选区域，直接输出目标类别与位置；后者如 R-CNN 系列[9]，需先生成候选区域再精细调整。在具体应用中，任梦真等[10]改进了 YOLOv8n，通过轻量化主干网络、引入 GhostNet 与改进注意力机制，在复杂场景下提升了检测精度；苗荣慧等[11]则优化 YOLOv7，引入多头自注意力机制，实现了果园环境中苹果成熟度的实时检测。然而，CNN 方法在处理小目标、遮挡或颜色变化时存在局限，且依赖锚框与 NMS（non-maximum suppression）等手工组件[12-13]。Transformer 类方法以 DETR（detection transformer）[14]和 RT-DETR 为代表，通过自注意力机制构建全局特征关联[15]，无需锚框与 NMS，流程更简洁。例如，LI 等[16]基于 DETR 提出半监督方法，有效监测严重遮挡下的作物生长，单一尺度特征图难以捕捉小目标细节，这在检测苹果成熟度这种颜色变化的细微细节时会出现的偏差；董适等[12]改进 RT-DETR，引入 RepVGG（re-parameterize vgg style convolutional networks）模块和级联分组注意力机制，使检测精确率提升至 95.8%，但有负值忽略与信息熵过大的问题。尽管如此，DETR 类方法仍存在计算复杂度高、小目标细节捕捉能力弱、信息熵过大等问题。因此，如何在保持实时性的同时进一步提升小目标识别精度，仍是当前研究的关键挑战。

基于此，本文研究基于 RT-DETR 提出改进的 FPB-DETR 检测模型，该模型通过三方面创新提升性能：首先，在主干网络中引入频率自适应膨胀卷积（frequency-adaptive dilated convolution，FADC）模块[17]，以增强对苹果表面细微特征的感知能力；其次，构建 Pola-AIFI（polaformer-attention-based intra-scale feature interaction）模块，优化特征交互过程并抑制环境干扰；最后，在多尺度特征融合阶段采用双向特征金字塔网络（bi-directional feature pyramid network，BIFPN）结构[18]，强化特征金字塔的跨尺度信息整合效率。最终，通过系统的实验验证 FPB-DETR 模型的综合性能。

1 FPB-DETR 网络结构

1.1 总体结构

RT-DETR 是基于 Transformer 改进的实时端到端目标检测器，由主干网络（backbone）[19]、高效混合编码器（hybrid encoder）[20-21]和带有辅助预测头的 Transformer 解码器三部分组成。针对苹果在复杂果园环境中面临的光照变化、枝叶遮挡及果实重叠等问题，确保检测系统满足实时性要求，解决 RT-DETR 模型在目标检测性能上的不足。本文以 RT-DETR-R18 为基准架构，提出了 FPB-DETR 检测模型。FPB-DETR 的整体网络架构如图 1 所示，其核心改进体现在三个模块的集成：FADC 模块通过自适应膨胀率、自适应核与频率选择三个组件，解决了传统卷积在固定膨胀率下无法同时捕获大小感受野特征的问题；Pola-AIFI 模块通过极性感知机制分离正负交互分量，避免负值信息丢失，提升注意力计算的准确性；BIFPN 则通过双向路径与加权融合机制，优化多尺度特征的信息流，减少特征传递中的歧义。该模型通过系统性结构优化，在保持端到端检测效率的基础上，显著提升了在复杂场景下的鲁棒性与精度。

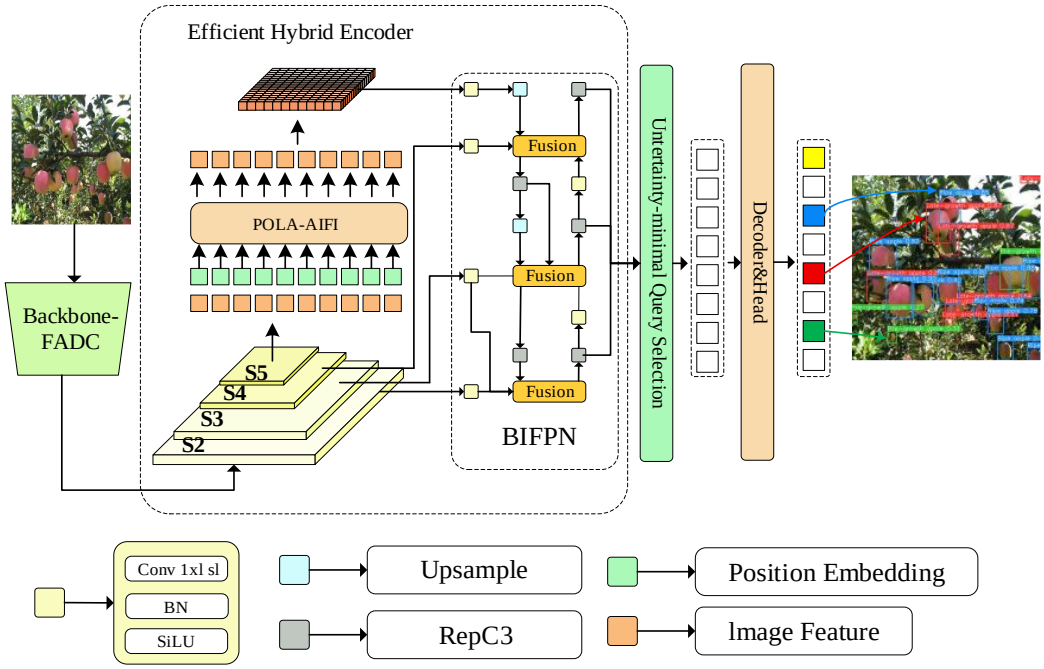


图 1　FPB-DETR 网络架构图

Fig. 1 FPB-DETR network architecture diagram

1.1.1　FADC 模块

RT-DETR 模型主干网络由多个 Conv（convolution）和 BasicBlock 组成，称之为 ResNet（residual network）。而 BasicBlock 使用 3×3 的卷积，采用固定膨胀率，对于需要小感受野的果柄微小结痂与需要大感受野的整果颜色渐变无法同时捕获。普通卷积在处理所有频率分量时使用同一标准，导致在光照强烈的情况下，反光区域被误判为成熟。为了解决以上问题，本文在 BasicBlock-s5 层引入 FADC 模块，使主干网络具备频域感知能力，更好的提高检测能力。FADC 结构如图 2 所示。

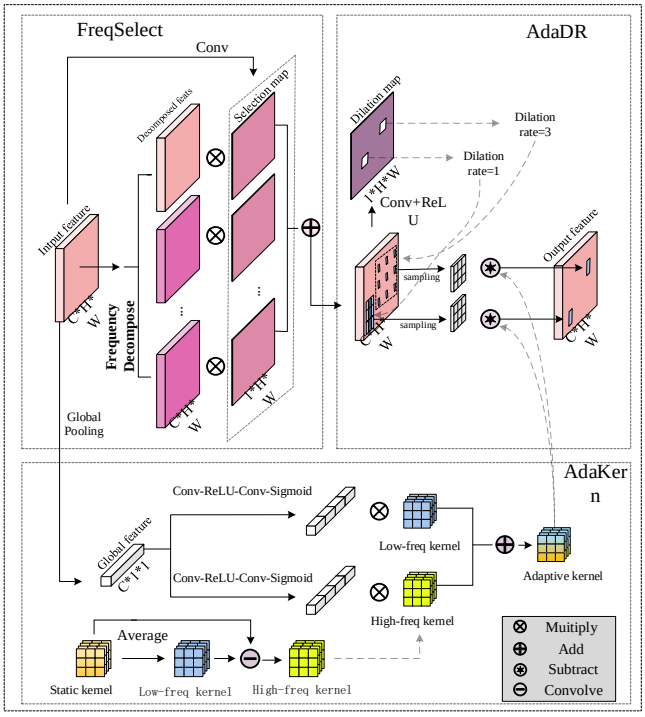


图 2　FADC 结构图

Fig.2 FADC structual diagram

如图 2 所示，FADC 模块融合了 3 个核心组件，即自适应膨胀率（adaptive dilation rate，AdaDR）、自适应核（AdaKern）和频率选择（FreqSelect）。自适应膨胀率公式如下：

$$Y(p)=\sum_{i=1}^{K\times K}W_i X(p+\Delta p_i\times\hat{D}(p))\ ,\tag{1}$$

式中：$Y(p)$ 为输出特征图中位置 p 的像素值；K 表示卷积核大小；W_i 表示卷积核第 i 个位置的权重参数；$X(p+\Delta p_i)$ 表示输入特征图中以 p 位置为参照、偏移量为 Δp_i 的像素值；$\hat{D}(p)$ 可以通过带有参数 θ 的卷积层进行预测，公式如下：

$$\theta=\max_{\theta}\left(\sum_{p\in \mathrm{HP}^-}\hat{D}(p)-\sum_{p\in \mathrm{HP}^+}\hat{D}(p)\right)\tag{2}$$

其中：HP^+ 表示具有最高高频功率像素；HP^- 表示具有最低高频功率像素。自适应空洞率（AdaDR）根据局部频率成分动态调整空间上的空洞率，平衡感受野和有效带宽。

FADC 中的自适应核（AdaKern）是一种动态生成卷积核参数的先进卷积方法，通过内容感知机制实现卷积核的样本级定制化，动态调整高频和低频成分。其权重可以分解如下：

$$W^{'}=\lambda_l\bar{W}+\lambda_h\hat{W}\ ,\tag{3}$$

式中：$W^{'}$ 为自适应生成的新卷积核权重；λ_l，λ_h 分别为低频分量和高频分量的动态权重；$\bar{W}$ 为低频分量；$\hat{W}$ 为高频残差分量。

FADC 的第 3 个核心组件频率选择（FreqSelect）是一种基于频域分析的特征优化机制，通过动态抑制噪声频段并增强目标特征频段，提升模型在复杂场景下的鲁棒性。具体来说，FreqSelect 通过在傅里叶域中应用不同的掩模将特征分解为不同的频带:

$$X_b=F^{-1}(M_b X_F)\tag{4}$$

式中：X_b 为原始输入特征图 X 中，仅包含某个特定频带 b 的成分所对应的空域特征图；X_F 为输入特征图 X 经过离散傅里叶变换后得到的频域表示；M_b 为一个用于提取相应频率的二进制掩码:

$$M_b(u,v)=\begin{cases}1 & \text{if}\ \phi_b\le\max(|u|,|v|)<\phi_{b+1}\\ 0 & \text{otherwise}\end{cases}\tag{5}$$

其式中：ϕ_b、ϕ_{b+1} 分别为 B+1 个预定义的频率阈值 $\{0,\phi_1,\phi_2,\cdots,\phi_{B-1},12\}$。FreqSelect 在空间上动态地对不同频段的频率分量进行加权。如式（6）所示：

$$\hat{X}(i,j)=\sum_{b=0}^{B-1}A_b(i,j)X_b(i,j)\ ,\tag{6}$$

式中：$\hat{X}(i,j)$ 为频率选择后学习到的频率平衡特征；$A_b(i,j)$ 为第 b 个频带的选择图；$X_b(i,j)$ 是使用离散傅里叶变换（DFT）将特征图 $X\in \mathrm{R}^{H\times W}$ 变换到频域。

FADC 模块通过 3 个方面提升苹果成熟度检测的精度和效率：一是自适应膨胀率根据区域频率特性动态调整，对高频的苹果边界用小膨胀率保留细节，对低频的苹果果肉区域用大膨胀率扩大感受野，平衡有效带宽与感知范围；二是自适应核分解卷积权重为高低频分量并动态调整比例，增强高频响应以捕捉表面斑点等关键细节；三是频率选择模块平衡特征中高低频分量，抑制过高频率以合理扩大感受野，同时减少计算量，兼顾精度与效率。

1.1.2 Pola-AIFI 模块

AIFI（attention-based intra-scale feature interaction）模块是 RT-DETR 的核心模块。通过 Transformer 的自注意力机制，实现同一尺度特征图内部的高效全局信息交互与融合。而其全局自注意力易受背景噪声影响，导致成熟度特征混淆。设计 Pola-AIFI，提升了树荫区域颜色识别稳定性。结构图如图 3 所示。

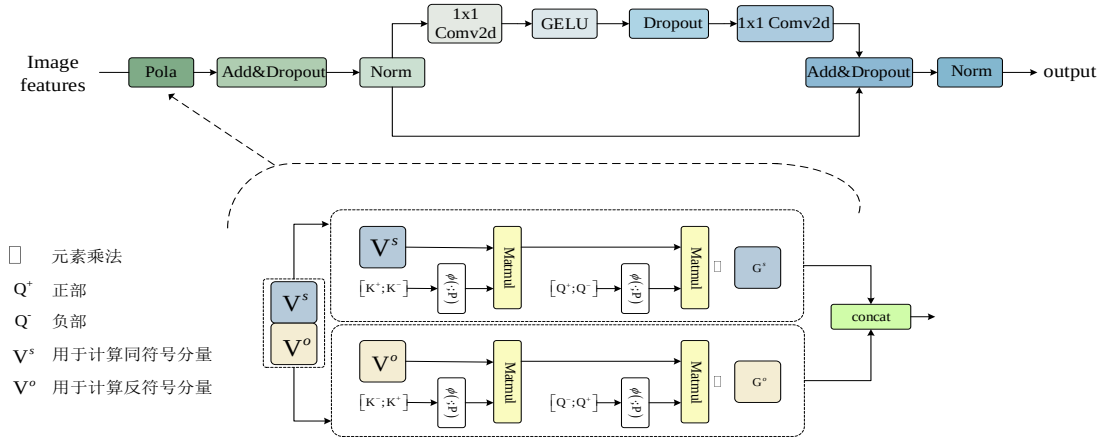


图 3　Pola-AIFI 结构图

Fig.3 Pola-AIFI structual diagram

图 3 该框架根据查询键对的极性，将其划分为两个独立的处理流：同符号分量和反符号分量。每个分量的缩放输出分别通过可学习的符号感知矩阵（$\boldsymbol{G^s}$ 和 $\boldsymbol{G^o}$）进行控制。此外，采用了具有可学习指数 p 的通道级幂函数来学习缩放过程，以此模拟 softmax 的锐度特性。

Pola-AIFI 模块通过纳入被忽略的负交互作用来解决先前线性注意力模型的局限性。先前线性注意力模型公式如式（7）所示：

$$\boldsymbol{o}_t = \frac{\sum_{i=1}^{N} \phi(\boldsymbol{q}_t)\phi(\boldsymbol{k}_i)^{\intercal}\boldsymbol{v}_i}{\sum_{j=1}^{N} \phi(\boldsymbol{q}_t)\phi(\boldsymbol{k}_j)^{\intercal}} = \frac{\phi(\boldsymbol{q}_t)\sum_{i=1}^{N} \phi(\boldsymbol{k}_i)\ \boldsymbol{v}_i}{\phi(\boldsymbol{q}_t)\sum_{j=1}^{N} \phi(\boldsymbol{k}_j)} \quad , \tag{7}$$

式中：$\boldsymbol{o}_t$ 为第 t 个位置的输出向量；$\boldsymbol{q}$ 表示查询矩阵，每一行对应一个查询向量；$\boldsymbol{k}$ 表示键矩阵，同样每一行对应一个键向量；$\boldsymbol{v}$ 是值矩阵，表示实际要被聚合的信息内容。而极性感知注意力背后的核心思想是为了解决现有线性注意力机制的局限性，这些机制通常会丢弃来自负成分的有价值信息。

Pola-AIFI 在处理负成分时，极性感知注意力将 query 和 key 向量分解为它们的正部和负部。这种分解允许机制分别考虑正相似度和负相似度对注意力权重的影响。对于查询向量 $\boldsymbol{q}$ 和键向量 $\boldsymbol{k}$，可以将它们分解为

$$\boldsymbol{q} = \boldsymbol{q}^{+} - \boldsymbol{q}^{-}, \boldsymbol{k} = \boldsymbol{k}^{+} - \boldsymbol{k}^{-} \quad , \tag{8}$$

式中：$\boldsymbol{q}_i^{+} = \max(\boldsymbol{q}_i, 0)$ 和 $\boldsymbol{q}_i^{-} = \max(-\boldsymbol{q}_i, 0)$ 分别代表 $\boldsymbol{q}$ 的正部和负部，对于 $\boldsymbol{k}$，上述规则同样成立。

将这些分解代入 $\boldsymbol{q}$ 和 $\boldsymbol{k}$ 的内积中，可以得到：

$$\langle \boldsymbol{q}, \boldsymbol{k} \rangle = \langle \boldsymbol{q}^{+}, \boldsymbol{k}^{+} \rangle + \langle \boldsymbol{q}^{-}, \boldsymbol{k}^{-} \rangle - \langle \boldsymbol{q}^{+}, \boldsymbol{k}^{-} \rangle - \langle \boldsymbol{q}^{-}, \boldsymbol{k}^{+} \rangle \tag{9}$$

前 2 项捕捉了同号成分之间的相似性，而后两项则代表了异号成分之间的相互作用。之前的线性注意力方法，如基于 ReLU 的特征映射，通过将负成分映射到零来消除它们，这在近似 $\boldsymbol{q}$，$\boldsymbol{k}$ 点积时会导致显著的信息丢失。

对此，极性感知注意力机制根据 $\boldsymbol{q}$、$\boldsymbol{k}$ 的极性将它们分开，并独立计算它们之间的相互作用。注意力权重的计算方式如下：

$$\mathrm{SM}(\boldsymbol{q}, \boldsymbol{k}^{\intercal}) = \exp(\boldsymbol{q}\boldsymbol{k}\ \ ) \approx (\phi(\boldsymbol{q}^{+})\phi(\boldsymbol{k}^{+})^{\intercal} + \phi(\boldsymbol{q}^{-})\phi(\boldsymbol{k}^{-})\ \ ) - (\phi(\boldsymbol{q}^{+})\phi(\boldsymbol{k}^{-})^{\intercal} + \phi(\boldsymbol{q}^{-})\phi(\boldsymbol{k}^{+})\ \ ) \tag{10}$$

Pola-AIFI 根据极性明确地将 $\boldsymbol{q}$，$\boldsymbol{k}$ 对分开，处理在内积计算过程中维度的同号和异号交互作用。这些交互作用在两个流中处理，从而能够更准确地重建原始的 softmax 注意力权重。为了避免不必要的复杂性，沿着通道维度拆分 $\boldsymbol{v}$ 向量，在不引入额外可学习参数的情况下处理这两种类型的交互作用。然后，将输出进行拼接，并通过一个可学习的符号感知矩阵进行缩放，以确保准确重建 $\boldsymbol{q}$，$\boldsymbol{k}$ 关系，如式（11）所示。

$$\boldsymbol{o}_t = \left[\frac{\phi\left(\left[\boldsymbol{q}_t^{+}; \boldsymbol{q}_t^{-}\right]\right)\sum_{i=1}^{N} \phi\left(\left[\boldsymbol{k}_i^{+}; \boldsymbol{k}_i^{-};\right]\right)^{\intercal}\boldsymbol{v}_i^{s}}{\phi\left(\left[\boldsymbol{q}_t^{+}; \boldsymbol{q}_t^{-}\right]\right)\sum_{j=1}^{N} \phi\left(\left[\boldsymbol{k}_i^{+}; \boldsymbol{k}_i^{-};\right]\right)^{\intercal}} \odot \boldsymbol{G^s}; \frac{\phi\left(\left[\boldsymbol{q}_t^{+}; \boldsymbol{q}_t^{-}\right]\right)\sum_{i=1}^{N} \phi\left(\left[\boldsymbol{k}_i^{-}; \boldsymbol{k}_i^{+};\right]\right)\ \boldsymbol{v}_i^{o}}{\phi\left(\left[\boldsymbol{q}_t^{+}; \boldsymbol{q}_t^{-}\right]\right)\sum_{j=1}^{N} \phi\left(\left[\boldsymbol{k}_i^{-}; \boldsymbol{k}_i^{+};\right]\right)} \odot \boldsymbol{G^o} \right] \quad , \tag{11}$$

式中：$\boldsymbol{q}_t^{+}$ 和 $\boldsymbol{q}_t^{-}$ 分别表示查询向量 $\boldsymbol{q}_t$ 中的正分量和负分量；$\boldsymbol{k}_i^{+}$ 和 $\boldsymbol{k}_i^{-}$ 分别表示键向量 $\boldsymbol{k}_i$ 中的正分量和负分

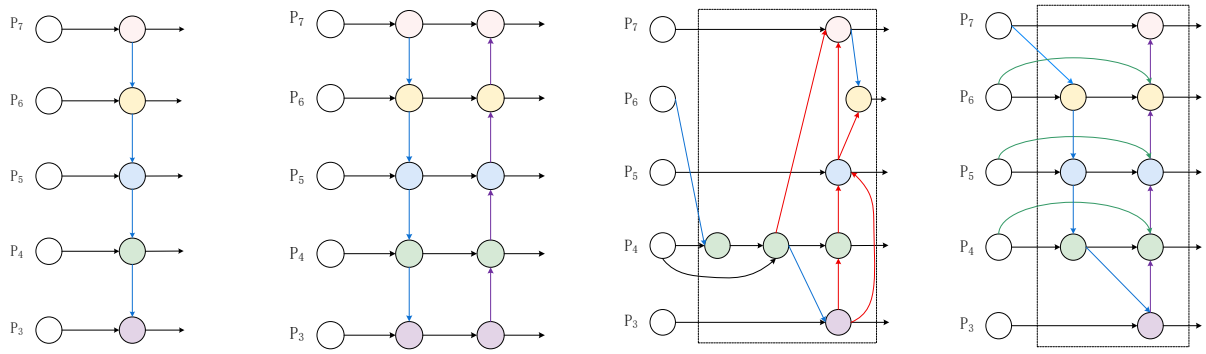
量；$\boldsymbol{v}_i^s$ 和 $\boldsymbol{v}_i^o$ 分别表示同符号和反符号交互对应的值向量；□ 表示逐元素乘法；$\boldsymbol{G}^s$ 和 $\boldsymbol{G}^o$ 是可学习的符号感知矩阵。

Pola-AIFI 模块通过极性感知计算增强特征交互、可学习幂函数缩放聚焦关键特征、低秩补偿卷积缓解高层退化，提升了检测精度；同时以线性注意力机制将计算复杂度从 $O(N^2)$降为 O(N)，提高了效率。通过设计 Pola-AIFI 能精准捕捉苹果表皮因成熟度变化而引发的细微物理特性改变（如蜡质层分布、微观粗糙度），这些特性直接影响光的偏振态，提供了超越颜色信息的、与成熟度高度相关的关键特征。提升了苹果成熟度判别的精度、鲁棒性以及对细微成熟度差异的分辨能力，为精准、稳定的自动化果园检测提供了强有力的技术支撑。

1.1.3 BIFPN 模块

RT-DETR 模型使用跨尺度特征融合模块（cross-scale feature fusion module，CCFM）进行特征融合。然而，其依赖简单的拼接或加和操作，容易引发特征信息丢失或退化，难以充分捕捉金字塔结构中不同层级特征图之间复杂的依赖关系与互补信息。这在苹果成熟度检测中尤为明显，因为苹果成熟度的判断依赖于多尺度特征的融合，而 CCFM 的简单融合方式难以有效捕捉细节特征，所以改用 BIFPN 模块。

BIFPN 通过引入双向连接与加权特征融合机制，能更高效地融合不同尺度特征，进而提升模型在多尺度场景下的泛化能力。这种双向融合机制不仅延续了传统 FPN（feature pyramid networks）自顶向下的特征融合路径，还新增了自底向上的路径，促使低层特征与高层特征实现更深度的交互。其中，自顶向下路径可强化小目标的特征表征，有效提升小目标的检测精度；而自底向上路径则能增强大目标的特征表达，让模型对不同大小的苹果实现更精准的识别。特征网络设计如图 4 所示。



a)FPN　　b)PANet　　c)NAS-FPN　　d)BIFPN

图 4　特征网络设计

Fig.4 Feature Network Design

1） FPN 引入了自顶向下的路径来融合从 3 级到 7 级的多尺度特征(P3-P7)；

2） PANet（path aggregation network）在 FPN 上增加了一个额外的自下而上通路；

3）NAS-FPN（neural architecture search for feature pyramid network）采用神经结构搜索找到不规则的特征网络拓扑，然后重复应用同一块；

4）BIFPN 具有更好的准确性和效率权衡，主要通过双向跨尺度连接和加权特征融合两个过程，实现不同尺度特征的高效整合与优化。

改进的 RT-DETR 中，BIFPN 模块通过简化跨尺度连接结构、重复双向路径融合多尺度特征来提升检测精度；同时采用加权特征融合和深度可分离卷积，在减少计算量的同时优化特征权重分配，从而提高检测效率。

BIFPN 通过双向交叉连接机制，实现了特征金字塔各层级间的动态交互。不仅支持高层语义特征向底层传递（自上而下方向），同时允许底层细节特征向上反馈（自下而上方向），从而构建了多尺度特征的双向融合通道。双向融合机制使得低层特征能够更好地与高层特征交互，防止信息丢失，提升模型对不同尺度目标的检测能力。BIFPN 通过引入可学习的权重机制，动态调整不同尺度特征。在融合过程中，每一级特征的权重由训练过程自动学习，确保最佳的特征整合。不同级别的特征通过加权相加的方式进行融合，图 4 d)展示了 BIFPN 第 6 层特征融合的计算过程，其数学表达式见式（12）、（13）。

$$P_6^{\text{td}} = \text{Conv}\left(\frac{\omega_1 \cdot P_6^{\text{in}} + \omega_2 \cdot \text{Resize}(P_7^{\text{in}})}{\omega_1 + \omega_2 + \varepsilon}\right) \tag{12}$$

$$P_6^{\text{out}} = \text{Conv}\left(\frac{\omega_1' \cdot P_6^{\text{in}} + \omega_2' \cdot P_6^{\text{td}} + \omega_3' \cdot \text{Resize}(P_5^{\text{out}})}{\omega_1' + \omega_2' + \omega_3' + \varepsilon}\right) \tag{13}$$

式中：P_6^{in} 表示第 6 级输入特征；P_6^{td} 表示自上而下路径中第 6 层的中间特征；ω_1 和 ω_2 是可学习的权重可以是一个标量(每个特征)、一个向量(每个通道)或一个多维张量(每个像素)；P_6^{out} 对应自下而上路径中该层的最终输出特征。该架构采用统一的计算范式，各层级特征均遵循相同的融合机制进行构建。为了进一步提高效率，采用深度可分离卷积进行特征融合，并在每次卷积后添加批量归一化和激活。

2 实　验

2.1 数据集采集与建立

数据来源网络图像与果园实地拍摄图像，选取复杂果园中 706 张图像。其中网络图像有 521 张（来源于 roboflow 公开数据集 Apple 及自主收集的网络图片，均经过人工筛选验证以确保质量），果园实地拍摄图像使用佳能 R10 相机拍摄，采集于蔚县下宫村乡浮图村苹果园共 185 张（品种为富士苹果），涵盖晴天、阴天、傍晚等不同时间段，枝叶遮挡、果实重叠等不同遮挡程度，单果、多果等不同场景，采集距离为 0.5~2 m，分辨率为 640×640，格式为 jpg。通过果实外观形状，分为幼果、生长初期、生长后期、成熟苹果四类。幼果果实小，表面密被茸毛，果皮为深绿色；生长初期的苹果，茸毛减少，果皮转浅绿色；生长后期的苹果表面光滑，开始显现淡红晕；成熟的苹果表面光滑，颜色饱满呈红色。为了进一步增强数据多样性并缓解可能存在的类别不平衡，本文使用旋转（±30°）、颜色调整（亮度、对比度、饱和度变化）、随机裁剪等扩充策略针对性扩充数据集到 2136 张，其中幼果 512 张、生长初期 528 张、生长后期 536 张、成熟苹果 560 张，各类别样本数量相对均衡，并按照 8:2 的比例随机分为训练集（1698 张）、验证集（438 张）。用 LabelImg 数据标注工具对不同成熟度的苹果进行人工标注，保存为“.txt”格式。部分不同环境下样本图片如图 5 所示。

a)晴天　b)阴天　c)近景　d)远景　e)多果　f)遮挡

图 5　苹果数据集

Fig.5 Apple dataset

2.2 实验平台和参数

所有实验均在同一台配置如下的计算机上完成：Windows 11(64 位)，Intel Core i7-13620H CPU，NVIDIA GeForce RTX 4090 GPU，软件环境为 CUDA 11.8、PyTorch 2.1.0 和 Python 3.10.14。模型训练使用 640×640 的输入图像，批量大小设为 4，共进行 150 轮(Epoch)。优化器采用 AdamW，初始学习率设置为 0.000 1，权重衰减为 0.000 1，损失函数为 GIoU。

2.3 评估指标

本文采用精确率(precision，*P*)、均值平均精度(mAP@0.5、mAP@0.5:0.95)、召回率(recall，*R*)、计算复杂度(giga floating-point operations per second，GFLOPs)和每秒处理帧率（frames per second，FPS）作为评估指标，用来验证 FPB-DETR 模型检测苹果成熟度的性能。相关公式如下：

P 反映模型预测结果的可靠性，计算公式为

$$\text{P}=\frac{\text{TP}}{\text{TP+FP}}\quad , \tag{14}$$

式中：TP（true positive）为正确识别的成熟苹果数量；FP（false positive）为误检为成熟的非苹果目标（如枝叶干扰或未成熟果实）。该指标直接决定模型在实际应用中的误报率。

R 表征模型对成熟苹果的覆盖能力，定义为

$$\mathrm{R}=\frac{\mathrm{TP}}{\mathrm{TP+FN}} \quad , \tag{15}$$

式中：FN（false negative）指漏检的成熟苹果数量。在农业采收场景中，高召回率可有效减少果实遗漏导致的产量损失。

平均精度（AP）基于 *P-R* 曲线下面积计算单类别检测稳定性：

mAP@0.5（mean average precision@0.5）：采用 0.5IoU 阈值时的平均精度，评估基础定位能力；mAP@0.5:0.95：IoU 阈值从 0.5 到 0.95（步长 0.05）的 mAP 均值，反映模型在严格定位要求下的鲁棒性。

$$\mathrm{AP}=\int_0^1 p(r)\mathrm{d}r \quad , \tag{16}$$

$$\mathrm{mAP}=\frac{1}{N}\sum_{i=1}^{N}\mathrm{AP}_i \quad , \tag{17}$$

式中：N 为成熟度等级数，本研究中 N =4 对应幼果/生长前期苹果/生长后期苹果/成熟苹果 4 个等级。

FPS 为每秒更新的图像帧数，是衡量画面流畅度的关键指标。

$$\mathrm{FPS}=\frac{s}{T'} \quad , \tag{18}$$

式中：s 为样本计数；T' 为所需的处理时间。

2.4 消融实验

以 RT-DETR 为基准网络，通过逐步集成本文提出的模块构建对比模型组，验证 FPB-DETR 各模块的有效性。实验结果见表 1。

表 1 消融实验结果

Tab.1 Results of ablation experiment

模型	FADC	Pola	BIFPN	mAP@0.5 /%	mAP@0.5:0.95 /%	*R*/%	*P*/%	FPS/ (帧/s)	Params (M)	GFLOPS (G)
1	—	—	—	95.0	71.7	91.0	90.5	99	19.8	57.0
2	√	—	—	95.7	75.0	91.4	92.1	96.1	20.3	57.6
3	√	√	—	96.0	74.0	91.8	92.8	181.0	20.2	53.6
4	√	—	√	95.9	75.2	90.8	92.7	98.0	20.4	60.6
5	—	√	√	95.5	73.4	91.2	92.1	158.7	20.5	64.6
6	√	√	√	96.8	74.5	92.7	92.5	142.9	20.4	56.7

实验 2，在主干网络中引入 FADC 模块，解决了有效带宽与感受野的矛盾以及全局固定膨胀率的局限，提升了目标苹果在复杂果园中的检测精度。实验结果表明，该模型在苹果成熟度识别任务中取得显著提升：精确率提高 1.6 个百分点，mAP@0.5 提升 0.7 个百分点。实验 3，在 FADC 基础上引入 Pola-AIFI 模块，解决了负值忽略与信息熵过大的问题，抑制了目标苹果在不同环境下的干扰，检测精确率相较于原模型提升 2.3 个百分点。实验 4，在 FADC 基础上，于多尺度特征融合阶段引入 BIFPN 结构，提升了 RT-DETR 的多尺度特征表达质量，使得 mAP@0.5 在原模型基础上提升 0.9 个百分点。实验 5，同时引入 Pola-AIFI 模块和 BIFPN 结构后，模型精准度也得到提升。实验 3、4、5 的结果表明，FADC 与 Pola-AIFI 模块及 BIFPN 结构具有良好的兼容性。实验 6 在同时引入三个模块后，在综合检测性能上达到了最优水平：mAP@0.5 为 96.8%，召回率为 92.7%，精确率为 92.5%。在效率方面，该组合取得了优异的推理速度（142.9 FPS），显著高于基线模型（99 FPS）；其参数量（20.4 M）与模型大小（56.7MB）均控制在较低水平，与仅引入 FADC 的模型（实验 2）相近，展现出极佳的效率-精度均衡性。虽然在 mAP@0.5:0.95 这一严格定位精度指标上（74.5%）略低于仅使用 FADC（75.0%）或 FADC+BIFPN（75.2%）的配置，但从整体性能来看，该组合在分类精度、召回率、推理效率等方面均表现出更佳的均衡性与实用性。

2.5 热力图分析

为了更直观地显示 FPB-DETR 模型的优势，采用 Grad-CAM 方法生成热图与 RT-DETR 对比，结果如图 6 所示，其中红色部分越深说明关注度越高。图 a）显示，在晴天时，由于光线充足，图片中的各个部分清晰，两个模型均未出现漏检现象，可以看出，FPB-DETR 的热图更为清晰，红色部分显著，对苹果的关注度更高；图 b）显示，在阴天的时候，光线较差，图片较昏暗，两个模型精度均有所下降，RT-DETR 模型苹果的关注度低于 FPB-DETR 模型；图 c）显示，在近果的情况时，2 个模型都呈现出高精度；图 d）显示，在检测目标距离变远后，对于较难检测的苹果，RT-DETR 模型出现了局部漏检现象，而 FPB-DETR 模型未受到距离的限制，在远距离情况下仍保持高精度的检测；图 e）显示，在检测目标数量堆积，密度变大的情况下，模型的检测难度增加，RT-DETR 模型开始出现明显漏检现象，而 FPB-DETR 模型将检测目标全部检测，无一漏检；图 f 显示，在有树叶遮挡的情况下，RT-DETR 模型在处理被遮挡的苹果时出现漏检的现象，将其当作树叶处理，而 FPB-DETR 模型仍保持高精度，高关注度。RT-DETR 模型对苹果的关注程度不高，在有遮挡物时精度下降，而 FPB-DETR 在各种复杂环境下都呈现出优异的精度，验证了改进模型的可行之处。

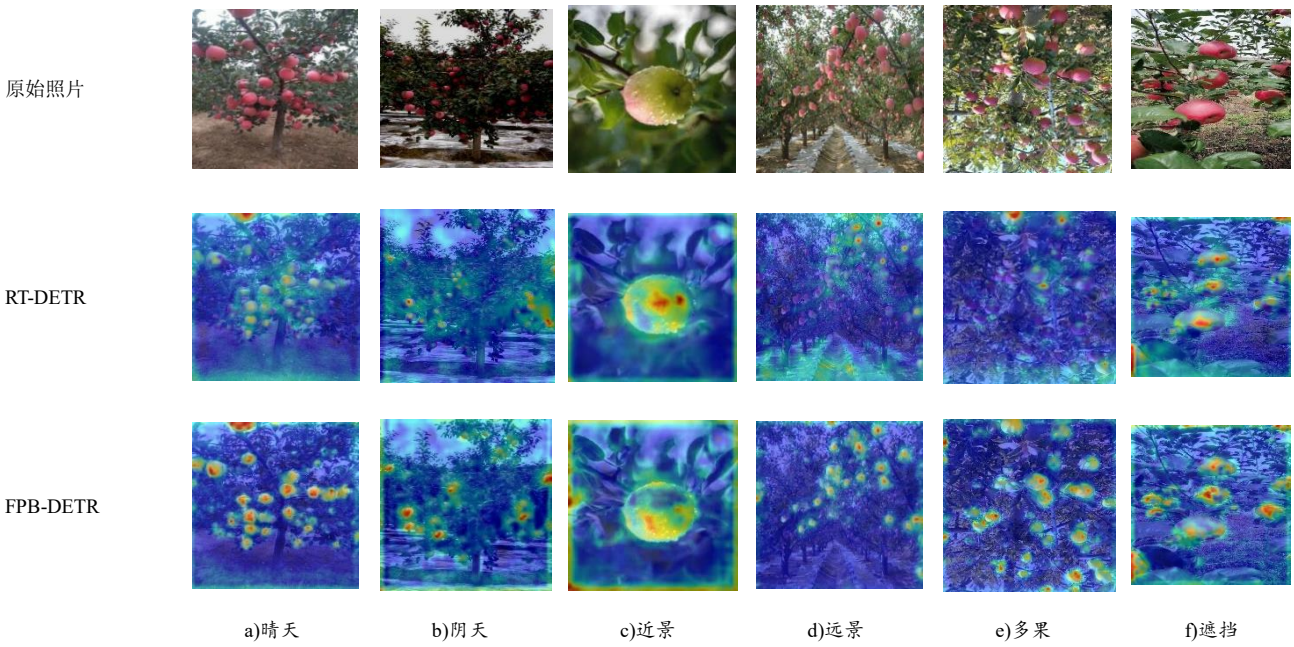


图 6 热力图

Fig.6 Heatmap

2.6 不同算法对比实验

如表 2 所示，综合对比实验表明，在统一数据集与实验下，FPB-DETR 在苹果成熟度检测任务中展现出全方位优势。

表 2 模型对比结果

Tab.2 Model comparison results

模型	mAP@0.5 (%)	mAP@0.5:0.95 (%)	R (%)	P (%)	FPS (帧/s)	Params (M)	GFLOPS (G)
FasterRCNN	87.0	61.0	89.0	68.8	78.5	42.0	78.3
YOLOv5m	86.1	60.9	80.5	80.4	117.4	25.0	64.0
YOLOv8m	88.5	52.8	80.3	86.1	128.0	25.9	78.8
YOLOv8-BIFPN-GLSA	86.2	60.7	78.8	81.8	130.2	25.6	80.4
YOLOv11m	91.2	65.1	85.9	85.8	140.1	20.1	68.0
YOLOv12m	88.0	61.1	78.9	85.9	136.7	20.2	67.5
RT-DETR	95.0	71.7	91.0	90.5	99.0	19.8	57.0
FPB-DETR	96.8	74.5	92.7	92.5	142.9	20.4	56.7

根据表 2 的数据，FPB-DETR 的 mAP@0.5 达到 96.8 %，mAP@0.5:0.95 为 74.5 %，*R* 为 92.7 %，*P* 为 92.5 %，在所有对比模型中均位列第一。该模型在保持高精度的同时实现了优异的推理速度，FPS 高达 142.9 帧/秒，超越 RT-DETR(99.0 FPS)及其他 YOLO 系列模型。在模型效率方面，尽管 FPB-DETR 的参数量为高于 YOLOv11m、YOLOv12m 和 RT-DETR，但低于 Faster R-CNN(42.0M)和 YOLOv5m(25.0 M)等模型。同时，其计算复杂度为 56.7 GFLOPS，低于多数对比模型，表明其具有优异的计算效率。综合整体性能，FPB-DETR 在精度、速度和模型复杂度之间达到了最佳平衡，具备良好的实用价值。

2.7 可视化分析

为了更直观地观察 FPB-DETR 的增强效果，检测结果如图 7 所示。图 7 a）行中，可以看出除 FPB-DETR 以外的其他检测图均有漏检现象，从图 7 b）—d）可以看出，全部 YOLO 模型和 RT-DETR 模型都出现漏检或误检现象，只有 FPB-DETR 模型没有漏检和误检，判断正确，做到了高精度的检测。

图 7 检测图

Fig.7 Detection diagram

3 结 语

针对复杂果园环境下的苹果成熟度检测需求，提出了一种改进的 RT-DETR 检测算法。通过引入 FADC 网络增强主干网络的特征提取能力，采用 Pola-AIFI 提升目标定位精度，并借助 BIFPN 增强多尺度特征融合，有效提升了模型对成熟苹果的识别性能。改进后的模型在精确率、召回率和平均精度上均有显著提升，检测速度也得到明显改善，为果园自动化监测提供了可行方案。本研究对果园巡检机器人、自动化精准采收装备等智能农业系统的开发具有参考价值。

本文算法在强反光场景下的适应性、跨品种泛化能力以及模型复杂度方面仍存在局限，未来拟在轻量化设计和多模态融合方面继续深入探索。

参考文献/References:

[1] 陈青,殷程凯,郭自良,等.苹果采摘机器人关键技术研究现状与发展趋势[J].农业工程学报,2023,39(4):1-15.

CHEN Qing,YIN Chengkai,GUO Ziliang,et al.Current status and future development of the key technologies for apple picking robots[J].Transactions of the Chinese Society of Agricultural Engineering,2023,39(4):1-15.

[2] 魏玉梅,江涛,白金燕.基于 Transformer 的遥感图像目标检测算法研究[J].计算机科学与应用,2024,14(4):105-114.

WEI Yumei,JIANG Tao,BAI Jinyan.Research on remote sensing image target detection algorithm based on Transformer[J].Computer Science and Application,2024,14(4):105-114.

[3] HUSSAIN M.YOLOv1 to v8:Unveiling each variant-A comprehensive review of YOLO[J].IEEE Access,2024,12:42816-42833.

[4] CHEN Chunling,ZHENG Ziyue,XU Tongyu,et al.YOLO-based UAV technology:A review of the research and its applications[J].Drones,2023.DOI:10.3390/drones7030190.

[5] DOSREIS D H,WELFER D,de SOUZA LEITE CUADROS M A,et al.Mobile robot navigation using an object recognition software with RGBD images and the YOLO algorithm[J].Applied Artificial Intelligence,2019,33(14):1290-1305.

[6] TAN Lu,HUANGFU Tianran,WU Liyao,et al.Comparison of RetinaNet,SSD,and YOLO v3 for real-time pill identification[J].BMC Medical Informatics and Decision Making,2021.DOI:10.1186/s12911-021-01691-8.

[7] NGUYEN N D,DO T, NGO T D,et al.An evaluation of deep learning methods for small object detection[J].Journal of Electrical and Computer Engineering,2020.DOI:10.1155/2020/3189691.

[8] VEZZALI E,BOLELLI F,SANTI S,et al.State-of-the-art review and benchmarking of barcode localization methods[J].Engineering Applications of Artificial Intelligence,2025.DOI:10.1016/j.engappai.2025.110259.

[9] REN Shaoqing,HE Kaiming,GIRSHICK R B,et al.Faster R-CNN:Towards real-time object detection with region proposal networks[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2017,39(6):1137-1149.

[10] 任梦真,罗雨竹,黄英来.改进 YOLOv8n 的蓝莓冠层果实成熟度识别方法[J/OL].哈尔滨理工大学学报(2025-05-22)[2025-09-13].https://link.cnki.net/urlid/23.1404.N.20250522.0916.004.

REN Mengzhen,LUO Yuzhu,HUANG Yinglai.Improved YOLOv8n detection method for fruit ripeness on blueberry canopies[J/OL].Journal of Harbin University of Science and Technology (2025-05-22)[2025-09-13].https://link.cnki.net/urlid/23.1404.N.20250522.0916.004.

[11]苗荣慧,李港澳,黄宗宝,等.基于 YOLO v7-ST-ASFF 的复杂果园环境下苹果成熟度检测方法[J].农业机械学报,2024,55(6):219-228.

MIAO Ronghui,LI Gang'ao,HUANG Zongbao,et al.Maturity detection of apple in complex orchard environment based on YOLO v7-ST-ASFF[J].Transactions of the Chinese Society for Agricultural Machinery,2024,55(6):219-228.

[12]董适,赵国瑞,苟豪,等.基于改进 RT-Detr 的黄瓜果实选择性采摘识别方法[J].农业工程学报,2025,41(1):212-220.

DONG Shi,ZHAO Guorui,GOU Hao,et al.Identifying cucumber fruits during selective picking using improved RT-Detr[J].Transactions of the Chinese Society of Agricultural Engineering,2025,41(1):212-220.

[13]LIN Ailiang,CHEN Bingzhi,XU Jiayu,et al.DS-TransUNet:Dualswin transformer U-Net for medical image segmentation[J].IEEE Transactions on Instrumentation and Measurement,2022,71:1-15.

[14]TIAN Yanfeng,AHMAD R B,ABDULLAH N A B.Accurate and efficient insulator maintenance:a DETR algorithm for drone imagery[J].PLoS One,2025.DOI:10.1371/journal.pone.0318225.

[15]廖律超,邹伟东,杨佳龙,等.基于注意力机制和微分跟踪器的宽度学习系统[J].深圳大学学报(理工版),2024,41(5):583-593.

LIAO Lyuchao,ZOU Weidong,YANG Jialong,et al.Broad learning system based on attention mechanism and tracking differentiator[J].Journal of Shenzhen University(Science & Engineering),2024,41(5):583-593.

[16]LI Haoyan,SHI Fanhuai.A DETR-like detector-based semi-supervised object detection method for *Brassicachinensis* growth monitoring[J].Computers and Electronics in Agriculture,2024.DOI:10.1016/j.compag.2024.108788.

[17]CHEN Linwei,GU Lin,ZHENG Dezhi.Frequency-adaptive dilated convolution for semantic segmentation[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).Seattle:IEEE,2024:3414-3425.

[18]TAN Mingxing,PANG Ruoming,LE Q V.EfficientDet:Scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).Seattle:IEEE,2020:10778-10787.

[19]XIAO Yi,WANG Xinqing,ZHANG Peng,et al.Object detection based on faster R-CNN algorithm with skip pooling and fusion of contextual information[J].Sensors,2020.DOI:10.3390/s20195490.

[20]TANG Siyuan,BAO Qiangqiang,JI Qingyu,et al.Improvement of RT-DETR model for ground glass pulmonary nodule detection[J].PLoS One,2025.DOI:10.1371/journal.pone.0317114.

[21]XIN Diye,LI Tianqi.Revolutionizing tomato disease detection in complex environments[J].Frontiers in Plant Science,2024.DOI:10.3389/fpls.2024.1409544.