doi:10.6041/j.issn.1000-1298.2024.06.023

基于 YOLO v7 - ST - ASFF 的复杂果园环境下苹果成熟度检测方法

苗荣慧[1,2]　李港澳[2]　黄宗宝[1]　李志伟[1,2]　杜慧玲[3]

(1. 山西农业大学信息科学与工程学院, 太谷 030801;
2. 山西农业大学农业工程学院, 太谷 030801; 3. 山西农业大学基础部, 太谷 030801)

摘要: 针对复杂果园环境下目标检测算法参数量大、鲁棒性差等问题,本文提出一种改进的 YOLO v7 网络模型用于苹果成熟度(未成熟、半成熟、成熟)检测。以 YOLO v7 为基线网络,在特征提取结构中引入窗口多头自注意力机制(Swin transformer,ST),极大地降低网络参数量与计算量;为提高模型对远景图像中小目标的检测能力,在特征融合结构中引入自适应空间特征融合(Adaptively spatial feature fusion,ASFF)模块优化 Head 部分,有效利用图像的浅层特征和深层特征,加强特征尺度不变性;采用 WIoU(Wise intersection over union)代替原始 CIoU(Complete intersection over union)损失函数,在提高检测准确率的同时加快模型收敛速度。试验结果表明,本文改进的 YOLO v7 - ST - ASFF 模型在苹果图像测试集上的检测速度和准确率均有显著提高,不同成熟度检测精确率、召回率和平均精度均值可达 92.5%、84.2% 和 93.6%,均优于 Faster R - CNN、SSD、YOLO v3、YOLO v5、YOLO v7 以及 YOLO v8 目标检测模型;针对多目标、单目标、顺光、逆光、远景、近景以及套袋、未套袋苹果目标的检测效果都较好;本文网络模型内存占用量为 53.4 MB,模型平均检测时间(Average detection time,ADT)为 45 ms,均优于其他目标检测模型。改进的 YOLO v7 - ST - ASFF 模型能够满足复杂果园环境下苹果目标的检测,可为果蔬机器人自动化采摘提供技术支撑。

关键词: 苹果; 成熟度检测; 复杂果园环境; YOLO v7; 窗口多头自注意力机制; ASFF

中图分类号: TP391　　**文献标识码**: A　　**文章编号**: 1000-1298(2024)06-0219-10　　**OSID**:

Maturity Detection of Apple in Complex Orchard Environment Based on YOLO v7 - ST - ASFF

MIAO Ronghui[1,2]　LI Gang'ao[2]　HUANG Zongbao[1]　LI Zhiwei[1,2]　DU Huiling[3]

(1. *College of Information Science and Engineering*, *Shanxi Agricultural University*, *Taigu* 030801, *China*
2. *College of Agricultural Engineering*, *Shanxi Agricultural University*, *Taigu* 030801, *China*
3. *Department of Basic Science*, *Shanxi Agricultural University*, *Taigu* 030801, *China*)

Abstract: In response to large number of parameters and poor robustness of object detection algorithms in complex orchard environment, an improved YOLO v7 network for apple maturity (immature, semimature, mature) detection was proposed. With YOLO v7 as the baseline network, a window multi-head self-attention mechanism (Swin transformer, ST) was adopted into the feature extraction structure to greatly reduce the parameters and computational complexity. In order to improve the ability of the model for detecting small targets in distant images, adaptively spatial feature fusion (ASFF) module was adopted into the feature fusion structure to optimize the Head part, effectively utilizing shallow and deep features and enhancing the performance of the feature scale invariance. Wise intersection over union (WIoU) was used to replace the original complete intersection over union (CIoU) loss function, thus accelerating the convergence speed and detection accuracy. The experimental results showed that the

收稿日期: 2024 - 03 - 08　修回日期: 2024 - 04 - 02

基金项目: 财政部和农业农村部:国家现代农业产业技术体系建设项目(CARS - 06 - 14.5 - A21)、中央引导地方科技发展资金项目(YDZJSX20231A042)、山西省谷子现代农业产业技术体系建设项目(2023CYJSTX04 - 04)、山西省重点研发重大项目(2022ZDYF119)和山西省基础研究计划项目(202203021212428)

作者简介: 苗荣慧(1990—),女,讲师,博士生,主要从事图像分析与机器视觉研究,E-mail: ronghui092@163.com

通信作者: 李志伟(1969—),男,教授,博士生导师,主要从事精细农业装备及计算机控制研究,E-mail: lizhiweitong@163.com

improved YOLO v7 – ST – ASFF model had significantly improved the detection speed and accuracy on the test set of the apple images. The average detection precision, recall, mean average precision (mAP) for different maturity levels can reach 92.5%, 84.2% and 93.6%, all of which were better than that of Faster R – CNN, SSD, YOLO v3, YOLO v5, YOLO v7 and YOLO v8 object detection models. The detection effects were good for multi, single, front-light, backlight, distant and close targets, as well as bagged and unpacked targets. The size of the model was 53.4 MB, and the ADT was 45 ms, which was also better than that of other models. The improved YOLO v7 – ST – ASFF model can meet the detection of apple targets in complex orchard environment, providing effective exploration for automated fruit and vegetable picking by robots.

Key words: apple; maturity detection; complex orchard environment; YOLO v7; ST; ASFF

0 引言

我国农业生产不断向规模化、集约化、精准化方向发展,对农业装备智能化、自动化需求也在不断增加[1]。苹果是我国产量最大的水果,其自动化采摘技术尤其是视觉识别技术更是当前研究的热点[2]。快速准确实现苹果目标检测是自动化采摘的关键[3]。复杂果园环境下的苹果目标检测受诸多因素的影响:枝叶对果实产生的遮挡;不同天气条件下复杂光照环境的影响;果实之间的相互重叠遮挡;采摘机器人本身算力有限;检测算法运算量大、参数量高等。因此,随着采摘成本不断增加、农业劳动力紧缺,实现复杂果园环境下苹果目标的智能化、自动化检测具有非常重要的实际意义。

苹果成熟度检测本质上属于目标检测范畴[4-5]。目标检测方法主要分为基于机器学习和深度学习。近年来,基于自动特征提取的深度学习目标检测方法已广泛应用于苹果目标检测。武星等[6]提出一种轻量化的 YOLO v3 卷积神经网络模型用于苹果目标检测,该方法在检测速度和准确率方面均有显著提高,其检测速度为 116.96 f/s,F1 值为 94.57%,平均精度均值为 94.69%。闫彬等[7]针对苹果的采摘方式提出一种基于改进的 YOLO v5m 苹果识别方法,将苹果采摘方式分为可直接采摘、迂回采摘和不可采摘 3 类,其识别结果均优于原 YOLO v5m、YOLO v3 和 EfficientDet – D0。LIU 等[8]提出一种基于 YOLO v5s 的苹果目标检测方法,通过引入空间到深度的卷积模块避免信息丢失,并结合上下文增强模块丰富特征金字塔网络信息,该方法验证了 YOLO v5s 模型的稳健性。以上研究虽然能够实现苹果目标的检测,但大多只针对成熟期苹果目标进行检测,而实际果园环境复杂、多种品种类型水果同时种植、成熟度不同,对不同成熟度苹果目标的检测能够直接决定果实的运输和贮藏方式,对果实价格及风味也会造成较大影响。同时,面对非结构化的苹果园,受自然光线、天气、绿色果实、采集角度、样本数量等影响,以上算法精度、鲁棒性及适用性相较于传统机器学习算法虽有大幅提升,但在检测远距离及遮挡目标时仍存在漏检和置信度偏低的问题,难以满足果园测产和自动采摘的实时作业需求,平衡检测速度、检测精度和模型计算复杂度仍是当前算法研究的重点。

YOLO v7[9]网络模型与其前身 YOLO v5 相比,通过改进骨干网络和特征融合方法等,进一步提升了检测精度;使用空间金字塔池化(Spatial pyramid pooling, SPP) + 路径聚合网络(Path aggregation network, PANet)实现多尺度特征融合,在保持较高检测精度的同时提高了检测速度。因此,本文针对复杂果园环境不同成熟度苹果进行检测,并对模型进行轻量化设计,增强小目标及远景目标的检测性能,提高模型对不同场景的适应性。以 YOLO v7 为基础网络实现复杂果园环境下苹果目标的检测,改进原有 YOLO v7 模型,将窗口多头自注意力机制引入原始骨干网络,增强模型对不同环境的泛化能力;在特征融合网络引入 ASFF 模块实现特征空间权重的自适应调整和融合,以提升对远景小目标的识别效果。

1 材料与方法

1.1 图像采集

本研究试验数据集采集于山西省晋中市太谷区白城村试验果园(北纬 37°21′,东经 112°30′),研究对象为富士苹果。考虑到机械臂在采摘过程中需要适应复杂的果园环境,采用 OPPO Reno5 型智能手机选取多种角度分别采集晴天、阴雨天等条件下的苹果图像,采集时间为 2023 年 7—9 月,采集距离为 0.5 ~ 1.5 m。共采集苹果图像 1 225 幅,图像尺寸为 3 468 像素 × 4 624 像素,格式为 JPEG。复杂果园环境包括顺光、逆光、阴雨天、单果、多果、近景、远景、果实重叠、枝叶遮挡等情况,部分样本图像如图 1 所示。

1.2 数据增强

复杂果园环境下采集的图像样本数量分布如表 1



图 1　复杂果园环境下苹果样本图像

Fig. 1　Sample images of apple in complex orchard environment

所示。由表 1 可知，由于存在光照与天气等不确定因素，获取的图像样本分布不均衡，解决样本不均衡问题可以消除模型对不同类别的偏向性，降低模型过拟合风险[10]。为了提高苹果目标检测模型的鲁棒性和泛化性，需对图像样本进行数据扩增[11]。本文采用垂直翻转、水平翻转、伽马变换、自适应直方图变换、随机亮度等方法增强图像对比度，提高模型获取目标信息的能力；考虑到图像获取过程中由于枝叶抖动、相机晃动等情况导致的运动模糊情况，对图像进行运动模糊处理。经过数据扩增之后每种环境下图像数量均在 1 200 幅左右，总样本图像为 3 015 幅，具体样本数如表 1 所示。将数据集按照比例 7∶2∶1划分为训练集(2 110 幅)、验证集(603 幅)和测试集(302 幅)。

表 1　复杂果园环境下苹果图像样本数量分布

Tab. 1　Distribution of apple image samples in complex orchard environment

样本类别	环境											
	单果	多果	顺光	逆光	夜间	雨后	遮挡	无遮挡	套袋	未套袋	近景	远景
原始样本	78	1 147	632	593	122	258	982	243	226	999	947	278
扩充样本	1 135	1 258	1 085	1 205	1 218	1 156	1 089	1 123	1 034	1 187	1 205	1 203

1.3　数据集搭建

采用 Labelme 标注软件对不同成熟度苹果数据集进行人工标注，标注框为苹果最小外接矩形，尽可能减少背景对目标检测的干扰。根据果实成熟情况，将苹果目标分别标注为未成熟(Immature)、半成熟(Semimature)及成熟(Mature)3 种类别。具体标注规则为：图像中枝叶遮挡严重的苹果目标不予标注，颜色为绿色的目标标注为未成熟，着色面积低于 80% 的目标标注为半成熟，着色面积高于 80% 至全红的目标标注为成熟。标注后产生的标注信息文件为 json 类型文件，需将数据格式转换为 YOLO 模型所需 txt 文件类型。

1.4　苹果成熟度检测模型建立

1.4.1　改进 YOLO v7 网络模型构建

YOLO v7 网络结构由输入端(Input)、骨干网络(Backbone)和检测头(Head)3 部分组成[12]，该模型提出了重参数化卷积(RepConv)、高效聚合网络(Efficient layer aggregation network，ELAN)、动态标签分配等策略，使其在检测精度和效率上取得较好的平衡[13]。检测头部分引入 Rep 模块与 SPPCSPC 模块，Rep 模块分为训练模块和推理模块，训练模型时将不同卷积层以及批归一化层合并为一个卷积模块，在网络推理时对训练模块中的参数进行重参数化操作，在保证模型性能的条件下加速了网络推理[14]。

YOLO v7 网络中对特征提取仅采用卷积操作，卷积层通常只会建模邻域像素之间的关系，难以捕捉长距离的依赖关系，导致小目标特征易丢失，出现远景果实漏检问题。而自注意力机制更加关注每个像素特征之间的关系，可以对卷积操作所得特征进行补充，使得模型能够更好地捕捉长距离的依赖关系[15－17]，提高对远景果实的检测能力。YOLO v7 采用多尺度特征图思想实现特征融合，以获取丰富的语义信息，然而，YOLO 系列的单阶段目标检测模型在特征融合过程中，不同层级之间的特征存在信息不一致性问题。本研究为复杂果园环境下的苹果成熟度检测，因此，以 YOLO v7 为基础模型，采用窗口多头自注意力机制改进原始骨干网络，充分提取复杂果园环境下目标的上下文信息，在提高特征表达能力的同时降低模型计算量，进而达到模型轻量化的目的；为了解决跨尺度融合问题，在检测头部分引入 ASFF 结构，该结构通过自适应地融合来自不同层级特征图的信息，可增强模型对不同尺度目标的检测能力，有效解决漏检或误检问题。改进后的 YOLO v7－ST－ASFF 苹果成熟度检测模型结构如图 2 所示。

1.4.2　Swin transformer 模块

融合不同尺度的特征是提高目标检测网络识别性能的一个重要手段[18]。低层特征图的分辨率较高，但语义性较低，所含噪声也较多；高层特征图具

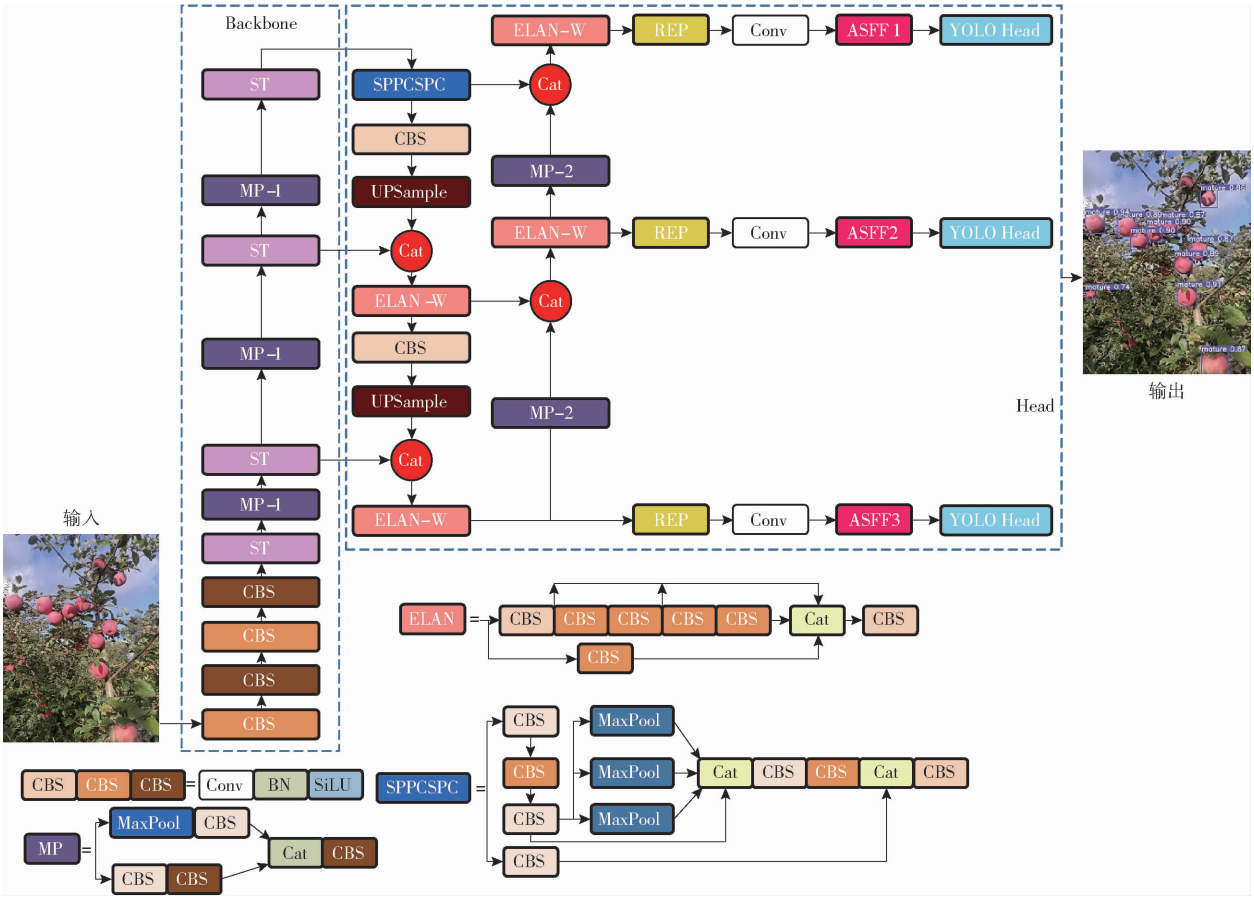


图 2 基于 YOLO v7－ST－ASFF 的复杂果园环境下苹果成熟度检测模型结构

Fig. 2 Maturity detection model structure of apple in complex orchard environment based on YOLO v7－ST－ASFF

有较强的语义信息,但其分辨率较低,对图像细节的感知能力也较差。因此,对高层与低层特征进行有效融合,是改善模型检测性能的关键。

Transformer[19] 模块通过引入自注意力机制,使模型能够全局感知输入的上下文信息,从而提高模型的表达能力和推理能力。近年来,Transformer 在目标检测、语义分割等任务中都表现出优异的性能。Transformer 从自然语言处理应用到计算机视觉领域,需面临视觉实体变化大、图像分辨率高等问题。而 Swin transformer[20] 引入 2 个关键概念——分层特征图和滑动窗口注意力,可有效解决以上问题。Swin transformer 模块整体架构如图 3 所示,它由图像分块层(Patch partition)、线性嵌入层(Linear embedding)和 4 个 Layer 组成,每个 Layer 由多个图块拼接层(Patch merging)和 Swin transformer block 模块构成。

在 ResNet 等卷积神经网络中,特征图下采样大多采用卷积操作完成,而 Swin transformer 模块中的下采样则采用无卷积 Patch merging 技术实现,其中“Patch”是特征图中的最小单位。假设输入图像的高度、宽度、通道数分别为 H、W、3,通过 Patch partition 层可将每 4×4 相邻的像素分块为一个 Patch,并沿通道方向展开,使得图像维度变为 $(H/4)\times(W/4)\times 48$;然后,通过 Linear embedding 层对每个像素的通道数据做线性变换,使图像维度变成 $(H/4)\times(W/4)\times C$,其中 C 为投影维度。通

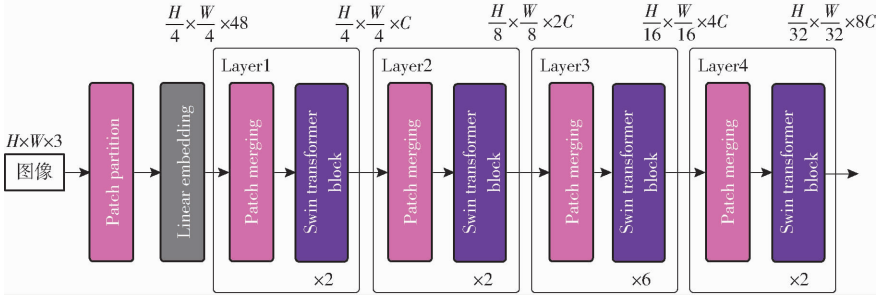


图 3 Swin transformer 模块结构

Fig. 3 Structure of Swin transformer module

过每个 Patch merging 后，输出特征图高和宽会减半，通道数翻倍，其中 Layer2、Layer3、Layer4 的输出特征图分别为（$H/8$）×（$W/8$）×$2C$、（$H/16$）×（$W/16$）×$4C$ 和（$H/32$）×（$W/32$）×$8C$。随着网络层次的加深，节点的感受野也在不断扩大。

Swin transformer 以 Transformer 结构为基础，将标准多头自注意力模块（Multi-head self-attention，MSA）替换为窗口多头自注意力模块（Window multi-head self-attention，W－MSA）和滑动窗口多头注意力模块（Shift window multi-head self-attention，SW－MSA）。图 4 为 Swin transformer block 结构。该模块由 2 个子单元组成，每个子单元由一个归一化层（Norm）、一个注意力模块，后接另一个归一化层和一个多层感知机（Multilayer perceptron，MLP）层组成；第 1 个子单元使用 W－MSA 模块，第 2 个子单元使用 SW－MSA 模块。

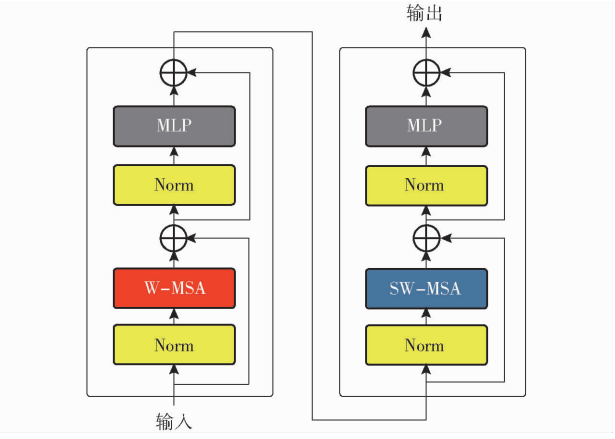


图 4　Swin transformer block 结构

Fig. 4　Structure of Swin transformer block

1.4.3　ASFF 模块

传统的 YOLO v7 模型在特征金字塔网络（Feature pyramid networks，FPN）的基础上结合了自顶向下和自底向上的特征融合功能，能够自适应地调整各尺度特征的空间融合权重。但多尺度融合会造成不同层级之间的特征信息冲突，即不一致性。不同特征尺度之间的不一致性是基于 FPN 的单阶段检测器的主要限制。

ASFF 模块能够对冲突信息进行空间过滤以抑制不一致性，可改善特征的尺度不变性[21]。因此，本文引入 ASFF 模块来抑制不同层级之间特征的信息冲突，提高 PANet 的融合效果，进而提高目标检测能力。ASFF 模块结构如图 5 所示。

1.5　损失函数

YOLO v7 的损失函数由分类损失、回归框损失和目标检测损失 3 部分构成，其中回归框损失为 CIoU。在目标检测中，框回归（Bounding box regression，BBR）是决定目标定位性能的关键步骤。YOLO v7 中的 CIoU 未考虑难易样本的平衡问题，导致评估结果可能存在偏差。因此，本研究使用 WIoU[22] 损失函数代替 CIoU，该损失函数能够有效解决质量较好和质量较差样本间的 BBR 平衡问题，提高模型的收敛速度和检测准确性。

1.6　试验环境配置与网络参数设置

试验所用操作系统为 Windows 10，CPU 型号为 Intel（R）Core（TM）i7－12700F CPU@2.10 GHz，GPU 型号为 NVIDIA GeForce RTX 3080，运行内存为 32 GB，1 TB 机械硬盘，编程语言为 Python 3.9，使用的深度学习框架为 Pytorch 1.13.0，GPU 加速库为 CUDA 11.7、CUDNN 8.4.1。训练时初始学习率为 0.001，批处理量为 16，动量为 0.9，迭代次数为 100。网络在前 20 轮训练时，损失值快速下降，训练 80 轮之后损失值基本趋于稳定。因此，本研究将训练 100 轮结束后输出的模型确定为面向苹果采摘机器人的目标检测模型。

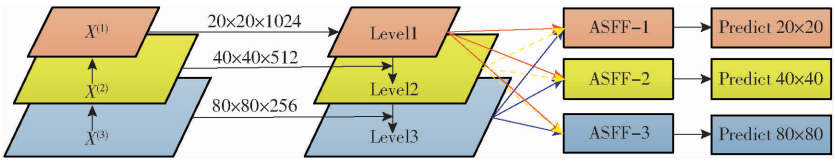


图 5　ASFF 模块结构

Fig. 5　Structure of ASFF module

1.7　模型评价指标

对苹果成熟度的检测需要考虑检测精度和速度。在模型检测精度方面，选用精确率（Precision，P）、召回率（Recall，R）和平均精度（Average precision，AP）作为评价指标；在模型检测性能方面，选取平均精度均值（mAP）、平均检测时间（ADT）和模型内存占用量作为评价指标。

2　试验结果与分析

2.1　YOLO v7－ST－ASFF 模型试验结果

为验证 YOLO v7－ST－ASFF 模型的性能，对测试集中的 302 幅苹果图像进行评价，表 2 为本文模型在不同成熟度级别的检测结果，图 6h 为改进模型的部分检测结果图。从表2可知，本文模型针对未

表 2 YOLO v7 – ST – ASFF 模型检测结果

Tab. 2 Detection results of YOLO v7 – ST – ASFF model

%

类别	指标		
	精确率	召回率	平均精度均值
未成熟	93.4	83.7	92.7
半成熟	89.9	87.9	93.5
成熟	94.2	81.0	94.6
平均值	92.5	84.2	93.6

成熟、半成熟和成熟苹果目标检测的平均精度均值分别为 92.7%、93.5% 和 94.6%，不同成熟度目标检测的精确率、召回率和平均精度均值可达 92.5%、84.2% 和 93.6%，表明检测效果较好。从图 6h 中可知，本文模型针对多目标、单目标、顺光、逆光、远景、近景以及套袋、未套袋环境下的苹果目标检测效果都较好，能够满足复杂环境下苹果目标的检测。

2.2 YOLO v7 – ST – ASFF 模型消融试验结果

本研究设置了 5 组消融试验，分别为 ASFF 作为 YOLO v7 检测头、YOLO v7 替换骨干网络为 Transformer 模块、YOLO v7 替换骨干网络为 Swin transformer 模块、YOLO v7 模型和本文 YOLO v7 – ST – ASFF 模型，所有模型采用的损失函数均为 WIoU。从定量角度分析 5 组网络性能，并对测试集中的苹果图像进行客观评价，结果如表 3 所示。从表 3 可知，本文提出的 YOLO v7 – ST – ASFF 模型平均精度均值均高于其他 4 组模型，且模型内存占用量有一定的降低。

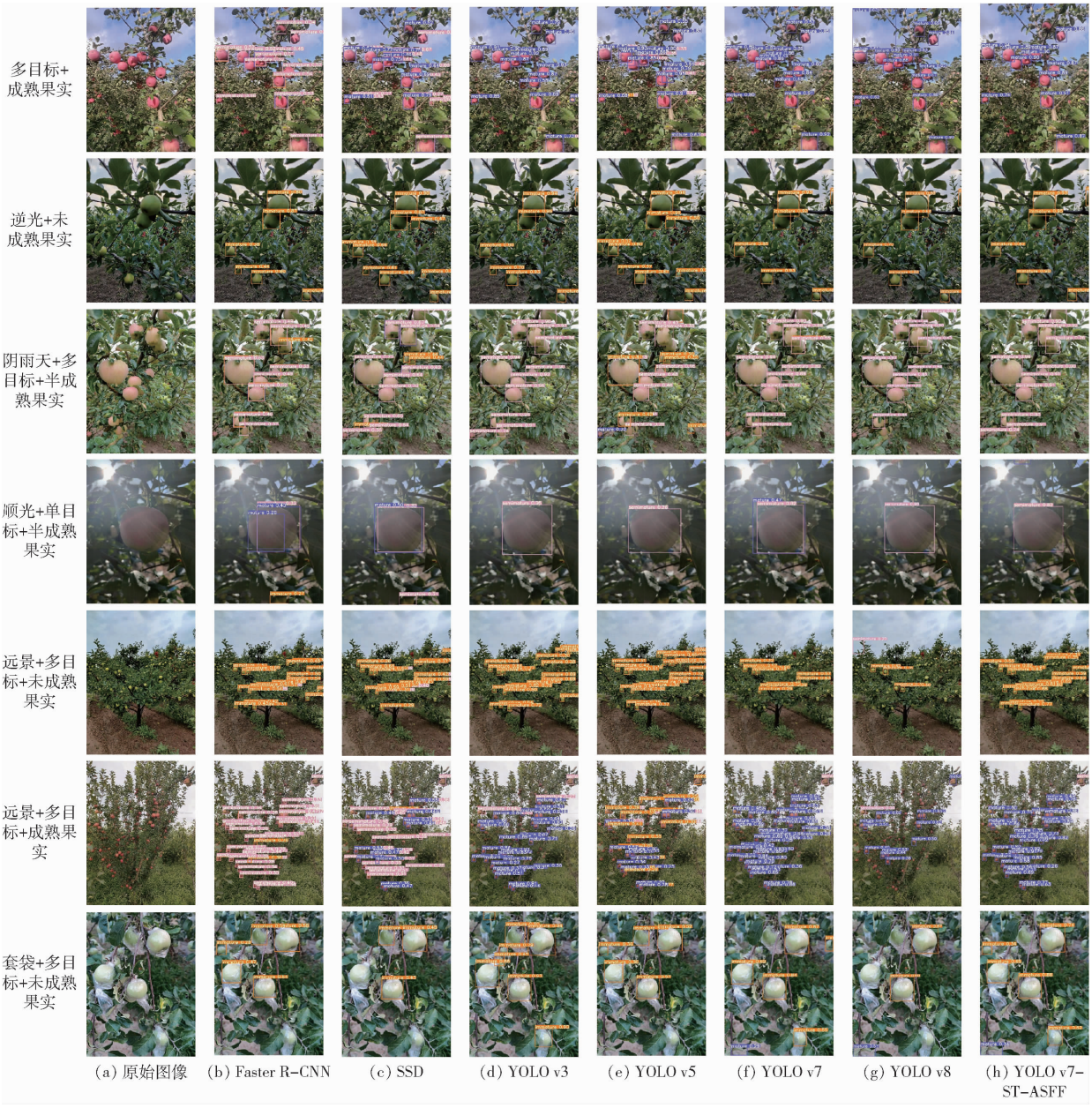


图 6 不同目标检测模型对比

Fig. 6 Comparison results of different object detection models

表 3　YOLO v7－ST－ASFF 模型消融试验结果

Tab. 3　Ablation results of YOLO v7－ST－ASFF model

模型	精确率/%	召回率/%	平均精度均值/%	模型内存占用量/MB
YOLO v7－ASFF	62.5	71.2	73.5	65.3
YOLO v7－Transformer	80.7	82.5	83.6	64.0
YOLO v7－Swin transformer	86.5	83.7	89.8	71.3
YOLO v7	88.8	83.5	91.7	64.0
YOLO v7－ST－ASFF	92.5	84.2	93.6	53.4

2.3　不同目标检测模型对比

2.3.1　不同目标检测模型定性与定量对比

为定性评价 YOLO v7－ST－ASFF 模型的检测结果，本文将改进模型与 Faster R－CNN[23]、SSD[24]、YOLO v3[25]、YOLO v5[26]、YOLO v7 以及 YOLO v8[27]目标检测模型对测试集中的苹果图像进行算法对比，结果如图 6 所示。

从图 6b 可知，使用 Faster R－CNN 模型，近景、远景多目标成熟果实检测效果都较差，多数目标被检测为半成熟级别；逆光未成熟果实检测效果较好，能够检测出小目标；阴雨天多目标半成熟果实、套袋未成熟果实检测效果尚可，但检测准确率较低；顺光单目标检测效果较差，存在重复检测问题；远景多目标未成熟检测效果一般，存在漏检情况。从图 6c 可知，使用 SSD 模型，近景、远景多目标成熟果实检测准确率较低，多数目标被检测为半成熟级别；逆光未成熟果实检测效果较好，能够检测出小目标；阴雨天多目标半成熟果实、套袋未成熟果实检测准确率较低；顺光单目标检测效果较差，存在重复检测问题；远景多目标未成熟检测效果较差，存在漏检情况。从图 6d 可知，使用 YOLO v3 模型，多目标成熟果实、阴雨天多目标半成熟果实、逆光未成熟果实、顺光单目标果实检测准确率都较高；远景多目标未成熟检测效果较差，存在漏检情况；远景多目标成熟果实检测效果尚可，少数目标被检测为半成熟级别；套袋未成熟果实检测效果尚可。从图 6e 可知，使用 YOLO v5 模型，多目标成熟果实检测效果较差，存在重复检测问题；逆光未成熟果实检测效果较差，存在误检情况，部分检测结果将叶片检测为果实；阴雨天多目标半成熟果实、远景多目标成熟果实检测效果较差，存在检测级别误检情况；顺光单目标果实检测准确率较低；远景多目标未成熟果实、套袋未成熟果实检测效果尚可，但检测准确率较低。从图 6f 可知，使用 YOLO v7 模型，多目标成熟果实、阴雨天多目标半成熟果实、远景多目标成熟果实、套袋未成熟果实检测效果都较好，检测准确率较高；逆光未成熟果实检测效果较好，能够对拍摄不完整目标进行检测；顺光单目标果实检测效果较差，存在重复检测问题；远景多目标未成熟检测效果较差，由于远景未成熟目标与叶片杂草颜色差异较小，只有部分目标能够检测出，且检测准确率较低。从图 6g 可知，使用 YOLO v8 模型，多目标成熟果实、阴雨天多目标半成熟果实、顺光单目标果实检测准确率较高；逆光未成熟果实检测效果尚可；远景多目标未成熟检测效果较差，只有极少部分目标能够检测出，且检测准确率较低；远景多目标成熟果实、套袋未成熟果实检测效果较差，检测准确率较低。从图 6h 可知，使用 YOLO v7－ST－ASFF 模型，多目标成熟果实、逆光未成熟果实、阴雨天气多目标半成熟果实、顺光单目标果实、远景多目标成熟果实检测效果均较好，检测准确率较高；远景多目标未成熟检测效果尚可，存在漏检问题；套袋未成熟果实检测效果一般，存在漏检情况。

为定量比较改进模型的性能，本文将改进模型与 Faster R－CNN、SSD、YOLO v3、YOLO v5、YOLO v7、YOLO v8 模型在测试集上进行性能比较，表 4 为不同检测模型在测试集中的性能结果。从表 4 可知，与其他模型相比，本文改进模型精确率、召回率及平均精度均值均最优，分别为 92.5%、84.2% 和 93.6%。通过使用 ST 模块改进骨干网络能够减少模型内存占用量，本文模型为 53.4 MB，ADT 为 45 ms，均优于其他模型。图 7 为不同检测模型在训练集上的总损失值对比。从图 7 可知，不同模型训练轮次大多在 80 后趋于稳定，收敛到固定值；Faster R－CNN 模型随着训练轮次的增加，损失值波动较大，其他模型整体变化趋势差距较小；与其他模型相比，本文提出的 YOLO v7－ST－ASFF 模型收敛速度最快。

表 4　不同检测模型在测试集中的性能

Tab. 4　Performance of different detection models in test sets

模型	精确率/%	召回率/%	平均精度均值/%	ADT/ms	模型内存占用量/MB
Faster R－CNN	55.8	76.6	69.6	82	65.3
SSD	57.5	74.6	72.6	67	57.5
YOLO v3	73.7	78.3	78.1	78	71.3
YOLO v5	75.5	78.6	79.6	56	55.8
YOLO v7	88.8	83.5	91.7	89	64.0
YOLO v8	89.8	83.4	90.8	53	58.3
YOLO v7－ST－ASFF	92.5	84.2	93.6	45	53.4

2.3.2　不同目标检测模型不同成熟度对比

表 5 为不同检测模型在测试集中不同成熟度的检测结果。从表 5 可知，与其他算法相比，本文改进

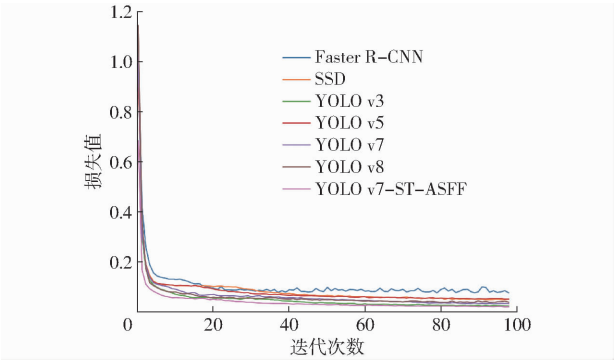


图 7 不同目标检测模型训练损失值对比

Fig. 7 Training loss results of different detection models

的 YOLO v7 - ST - ASFF 算法针对未成熟、半成熟和成熟苹果目标检测的平均精度均值均最高。Faster R - CNN 模型针对半成熟苹果目标检测结果最差，针对未成熟苹果目标检测结果良好。SSD 模型针对不同成熟度苹果目标检测结果均优于 Faster R - CNN 模型。YOLO v3 和 YOLO v5 模型针对不同成熟度苹果目标检测结果均不太理想。YOLO v7 和 YOLO v8 模型针对不同成熟度苹果目标的检测结果均较好，其平均精度均值可达 90% 以上。

表 5 不同检测模型在测试集中不同成熟度检测结果

Tab. 5 Performance of different detection models in different maturity in test sets %

模型	平均精度			平均精度均值
	未成熟	半成熟	成熟	
Faster R - CNN	86.0	54.9	67.8	69.6
SSD	74.1	66.6	77.2	72.6
YOLO v3	76.7	75.5	82.1	78.1
YOLO v5	82.9	74.4	83.6	79.6
YOLO v7	92.8	89.1	93.2	91.7
YOLO v8	90.2	89.8	92.4	90.8
YOLO v7 - ST - ASFF	92.7	93.5	94.6	93.6

2.4 不同回归框损失函数对比

原始 YOLO v7 模型采用 CIoU 作为预测框和真实框之间的损失函数，即回归框损失。CIoU 将边界框的长宽比作为惩罚项加入到损失函数中，一定程度上可加快预测框的回归收敛过程，但当收敛到预测框和真实框的长和宽呈线性比例时，惩罚项则恒为零，存在不合理性。因此，对比分析 CIoU、EIoU（Efficient intersection over union）、GIoU（Generalized intersection over union）、SIoU（Smooth intersection over union）和 WIoU 损失函数对模型性能的影响。不同损失函数在训练集上的性能及损失值如表 6 和图 8 所示。从表 6 可知，原始 CIoU 损失函数的精确率、召回率和平均精度均值都最低，WIoU 平均精度均值为 93.6%，均优于其他损失函数。从图 8 可以看出，EIoU 损失值最大，均高于 CIoU、GIoU、SIoU 和 WIoU；GIoU 和 SIoU 损失曲线下降趋势相似，且均高于 CIoU 和 WIoU；CIoU 损失曲线在经过一次上升下降之后，在迭代次数为 30 左右时出现明显的上升再下降波动，表明该模型存在一定过拟合现象，无法满足模型收敛性要求；WIoU 拥有更小的收敛损失，且收敛速度较快，与其他损失函数相比，从性能分析和收敛速度上该损失函数都最优。

表 6 不同损失函数在训练集中的性能

Tab. 6 Performance of different loss functions in training sets

损失函数	精确率/%	召回率/%	平均精度均值/%	模型内存占用量/MB
CIoU	79.4	78.2	81.8	53.4
EIoU	80.4	89.3	89.9	53.4
GIoU	89.2	84.9	91.9	53.4
SIoU	91.7	85.1	92.2	53.4
WIoU	92.5	84.2	93.6	53.4

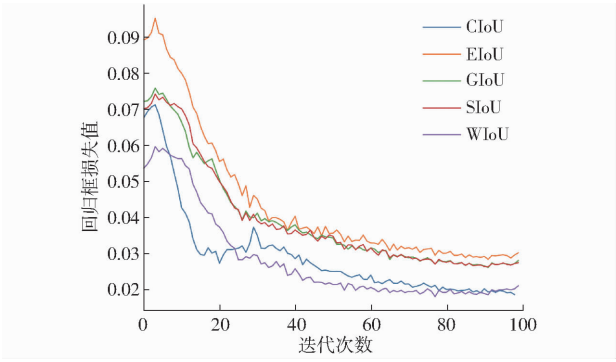


图 8 不同损失函数在训练集上的损失值对比

Fig. 8 Training loss curves of different loss functions

2.5 热力图

目标检测热力图是图像中信息对于预测结果的贡献排名，分数越高（颜色越红）的地方表示在输入图像中该区域对网络的响应越高、贡献越大。本研究采用 Grad - CAM（Gradient-weighted class activation mapping）方法对改进 YOLO v7 - ST - ASFF 模型进行热力图可视化，热力图中像素值表示该像素区域对于检测结果的重要性。图 9 为部分复

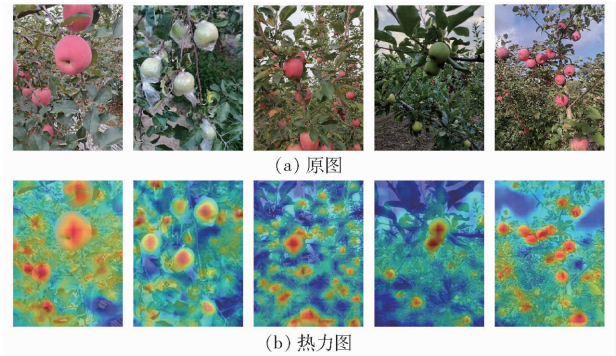


图 9 YOLO v7 - ST - ASFF 模型检测热力图

Fig. 9 Heat maps of YOLO v7 - ST - ASFF model

杂果园环境下苹果图像检测热力图，从图 9 中可看出，本文 YOLO v7 - ST - ASFF 模型的检测热力图更加贴近真实果实区域，能够注意到不同环境情况下的苹果目标。

3　结论

（1）针对复杂果园环境下目标检测网络对不同成熟度苹果检测精度不高问题，改进原有 YOLO v7 模型，将 ST 模块引入原始骨干网络，有效获取长距离依赖关系，提高特征表达能力以适应复杂的果园环境。引入 ASFF 改进特征融合网络实现特征空间权重的自适应调整和融合，模型在小尺度、远景图像下的目标检测性能提升明显。

（2）本文模型对未成熟、半成熟和成熟苹果目标检测的平均精度均值分别为 92.7%、93.5% 和 94.6%，所有成熟度级别检测的平均精确率、召回率和平均精度均值可达 92.5%、84.2% 和 93.6%。针对多目标、单目标、顺光、逆光、远景、近景以及套袋、未套袋检测效果都较好，能够满足复杂环境下苹果目标检测。

（3）设置了 5 组消融试验验证网络性能，分别为 ASFF 作为 YOLO v7 检测头、YOLO v7 替换骨干网络为 Transformer 模块、YOLO v7 替换骨干网络为 Swin transformer 模块，原始 YOLO v7 模型和本文 YOLO v7 - ST - ASFF 模型。从定量的角度分析 5 组网络性能，试验结果表明，本文改进的 YOLO v7 - ST - ASFF 模型平均精度均高于其他 4 组模型，且模型内存占用量有一定的降低。

（4）与 Faster R - CNN、SSD、YOLO v3、YOLO v5、YOLO v7 及 YOLO v8 模型对测试集中的苹果图像进行对比，改进的 YOLO v7 - ST - ASFF 模型针对多目标成熟果实、逆光未成熟果实、阴雨天气情况下多目标半成熟果实、顺光单目标果实检测效果都较优，且检测准确率较高。本文模型与其他模型相比，精确率、召回率及平均精度均值也最优。通过使用 ST 模块改进骨干网络能够减少模型内存占用量，降低对硬件的要求，提升模型可移植性。在模型训练过程中，改进 YOLO v7 - ST - ASFF 模型收敛速度最快。

（5）由于远景未成熟目标与叶片颜色相似程度较高，本文模型针对远景未成熟多目标果实、套袋未成熟果实检测效果一般，存在漏检情况，影响了检测结果。

参　考　文　献

[1] 龙燕，高研，张广犇. 基于改进 HRNet 的单幅图像苹果果树深度估计方法[J]. 农业工程学报，2022，38(23)：122 - 129.
LONG Yan，GAO Yan，ZHANG Guangben. Depth estimation of apple tree in single image using improved HRNet[J]. Transactions of the CSAE，2022，38(23)：122 - 129.（in Chinese）

[2] 曾晏林，贺壹婷，蔺瑶，等. 基于 BCE - YOLO v5 的苹果叶部病害检测方法[J]. 江苏农业科学，2023，51(15)：155 - 163.
ZENG Yanlin，HE Yiting，LIN Yao，et al. A detection method for apple leaf diseases based on BCE - YOLO v5[J]. Jiangsu Agricultural Sciences，2023，51(15)：155 - 163.（in Chinese）

[3] ZHANG C，ZHANG H，ZHANG C. Multi-objective detection and tracking algorithm for unmanned street view based on improved YOLO v5 and DeepSORT under complex scenes[C]//2022 IEEE 4th International Conference on Civil Aviation Safety and Information Technology（ICCASIT），2022.

[4] YAN B，FAN P，LEI X，et al. A real-time apple targets detection method for picking robot based on improved YOLO v5[J]. Remote Sensing，2021，13(9)：1619.

[5] TIAN Yunong，YANG Guodong，WANG Zhe，et al. Apple detection during different growth stages in orchards using the improved YOLO - V3 model[J]. Computers and Electronics in Agriculture，2019，157：417 - 426.

[6] 武星，齐泽宇，王龙军，等. 基于轻量化 YOLO v3 卷积神经网络的苹果检测方法[J]. 农业机械学报，2020，51(8)：17 - 25.
WU Xing，QI Zeyu，WANG Longjun，et al. Apple detection method based on light YOLO v3 convolutional neural network[J]. Transactions of the Chinese Society for Agricultural Machinery，2020，51(8)：17 - 25.（in Chinese）

[7] 闫彬，樊攀，王美茸，等. 基于改进 YOLO v5m 的采摘机器人苹果采摘方式实时识别[J]. 农业机械学报，2022，53(9)：28 - 38，59.
YAN Bin，FAN Pan，WANG Meirong，et al. Real-time apple picking pattern recognition for picking robot based on improved YOLO v5m[J]. Transactions of the Chinese Society for Agricultural Machinery，2022，53(9)：28 - 38，59.（in Chinese）

[8] LIU J，WANG C，XING J. YOLO v5 - ACS：improved model for apple detection and positioning in apple forests in complex scenes[J]. Forests，2023，14(12)：2304.

[9] ZHANG X，HUANG D. Research on UAV ground target detection based on improved YOLO v7[C]//2023 3rd International Conference on Computer，Control and Robotics（ICCCR），2023：28 - 32.

[10] LI Lili，YANG Hua，WANG Bin. Cerasus humilis cultivars identification with small-sample and unbalanced dataset based on EfficientNet - B0 + Ranger networks[J]. INMATEH-Agricultural Engineering，2023，69(1)：225 - 236.

[11] YU G，SHAO B. Garbage classification and detection based on improved YOLO v7 network[C]//2023 International Conference on Pattern Recognition，Machine Vision and Intelligent Algorithms（PRMVIA），2023：103 - 107.

[12] 于春涛,李金阳,石文强,等. 基于改进 YOLO v7 的杂交大豆苗期胚轴颜色检测模型[J]. 中国农业大学学报, 2024, 29(2): 11 - 22.
YU Chuntao, LI Jinyang, SHI Wenqiang, et al. Color detection model of hybrid soybean hypocotyl based on an improved YOLO v7 object detection model[J]. Journal of China Agricultural University, 2024, 29(2): 11 - 22. (in Chinese)

[13] 李非凡,沈新锋,赵和平,等. 基于计算机人工智能 YOLO v7 算法的物体表面缺陷检测方法[C]//第十七届全国信号和智能信息处理与应用学术会议论文集,2023:9.
LI Feifan, SHEN Xinfeng, ZHAO Heping, et al. Object surface defect detection method based on computer artificial intelligence YOLO v7 algorithm[C]// Proceedings of the 17th National Conference on Signal and Intelligent Information Processing and Application, 2023: 9. (in Chinese)

[14] 邹振涛,李泽平. 改进 YOLO v7 的航拍图像目标检测[J/OL]. 计算机工程与应用: 1 - 12[2023 - 12 - 27]. http://kns.cnki.net/kcms/detail/11.2127.TP.20231016.1108.016.html.
ZOU Zhentao, LI Zeping. Improved YOLO v7 for UAV image object detection[J/OL]. Computer Engineering and Applications: 1 - 12[2023 - 12 - 27]. http://kns.cnki.net/kcms/detail/11.2127.TP.20231016.1108.016.html. (in Chinese)

[15] WANG S, WU D, ZHENG X. TBC - YOLO v7: a refined YOLO v7-based algorithm for tea bud grading detection[J]. Frontiers in Plant Science, 2023, 14: 1223410.

[16] MENG X, LI C, LI J, et al. YOLO v7 - MA: improved YOLO v7-based wheat head detection and counting[J]. Remote Sensing, 2023, 15(15): 3770.

[17] HE Y, SU B, YAN J, et al. Research on underwater object detection of improved YOLO v7 model based on attention mechanism: the underwater detection module YOLO v7 - C[C]// Proceedings of the 2022 4th International Conference on Robotics, Intelligent Control and Artificial Intelligence, 2022: 302 - 307.

[18] HASSAN M, HUDA S, HASSAN M, et al. Early detection of cardiovascular autonomic neuropathy: a multi-class classification model based on feature selection and deep learning feature fusion[J]. Information Fusion, 2022, 77: 70 - 80.

[19] STRUDEL R, GARCIA R, LAPTEV I, et al. Segmenter: transformer for semantic segmentation[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 7262 - 7272.

[20] 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, 2021: 10012 - 10022.

[21] 耿艳利,林彦伯,付艳芳,等. 基于 Transformer 与自适应空间特征融合的群猪目标检测算法研究[J]. 东北农业大学学报 ,2023, 54(1): 88 - 96.
GENG Yanli, LIN Yanbo, FU Yanfang, et al. Study on pigs target detection algorithm based on Transformer and adaptively spatial feature fusion[J]. Journal of Northeast Agricultural University, 2023, 54(1): 88 - 96. (in Chinese)

[22] 齐向明,柴蕊,高一萌. 重构 SPPCSPC 与优化下采样的小目标检测算法[J]. 计算机工程与应用, 2023, 59(20): 158 - 166.
QI Xiangming, CHAI Rui, GAO Yimeng. Algorithm of reconstructed SPPCSPC and optimized downsampling for small object detection[J]. Computer Engineering and Applications, 2023, 59(20): 158 - 166. (in Chinese)

[23] KALIDINDA L D, VIJAYABASKAR V. Plant disease detection using Faster RCNN networks[C]// 2022 International Conference on Computing, Communication and Power Technology (IC3P), 2022: 260 - 263.

[24] 李振波,李萌,吴宇峰,等. 基于优化 SSD 算法的冰鲜鲳鱼新鲜度评估方法研究[J]. 农业机械学报, 2021, 52(增刊): 472 - 481.
LI Zhenbo, LI Meng, WU Yufeng, et al. Evaluation method of iced pomfret freshness based on improved SSD[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(Supp.): 472 - 481. (in Chinese)

[25] WANG K, LIU M. YOLO v3-MT: a YOLO v3 using multi-target tracking for vehicle visual detection[J]. Applied Intelligence, 2022, 52(2): 2070 - 2091.

[26] REN J, WANG Z, ZHANG Y, et al. YOLO v5-R: lightweight real-time detection based on improved YOLO v5[J]. Journal of Electronic Imaging, 2022, 31(3): 033033.

[27] ZHOU Y, ZHU W, HE Y, et al. YOLO v8-based spatial target part recognition[C]//2023 IEEE 3rd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA). IEEE, 2023: 1684 - 1687.