DOI: 10.13718/j.cnki.xdzk.2025.07.019
刘愚成，梁新成，李法霖，等. 基于改进 YOLOv8 的柠檬果实识别方法[J]. 西南大学学报(自然科学版)，2025，47(7)：219-230.

基于改进 YOLOv8 的柠檬果实识别方法

刘愚成，　梁新成，　李法霖，　张峰菱，　李云伍

西南大学 工程技术学院，重庆 400715

摘要：为了解决人工采摘柠檬果实成本高、效率低下等问题，实现在复杂环境下对柠檬果实的快速精准识别，提出一种基于改进 YOLOv8 模型的柠檬果实识别方法。首先，在主干网络中引入 SPDConv 模块来提升模型对低分辨率图像和小目标的检测精度；然后，加入 EMA 注意力机制来有效提取被遮挡柠檬的特征；最后，将 CIoU 边界框损失函数替换为 Wise-IoU 以降低对高质量锚框的依赖并增强模型的泛化能力。在自建数据集上进行测试，结果表明：YOLOv8-SEW 模型的精确率、召回率、平均精度均值分别为 94.5%、85.7%及 92.4%，与改进前相比分别提升 1.0%、4.2%和 2.9%。单幅图像的检测时间为 44.8 ms，可以实现柠檬的快速准确识别，为实现自动化柠檬果实采摘提供技术基础。

关　键　词：柠檬识别；YOLOv8；SPD 卷积模块；Wise-IoU 损失函数；注意力机制

中图分类号：TP391；S23　　**文献标识码**：A　　开放科学(资源服务)标识码(OSID)：

文 章 编 号：1673-9868(2025)07-0219-12

A Lemon Fruit Recognition Method Based on Improved YOLOv8

LIU Yucheng，　LIANG Xincheng，　LI Falin，
ZHANG Fengling，　LI Yunwu

College of Engineering and Technology，Southwest University，Chongqing 400715，China

Abstract: In order to address the challenges of high costs and low efficiency of manual picking lemon fruits, and achieve swift and precise identification of lemon fruits in intricate environments, a lemon fruit recognition method based on the improved YOLOv8 model was established. Firstly, the SPDConv module was introduced into the backbone network to enhance the accuracy of model's detection for low-resolution images and small targets. Then, the EMA attention mechanism was added to effectively extract the features of obscured fruits. Finally, the CIoU bounding box loss function was replaced with Wise-IoU to re-

收稿日期：2024-03-12
基金项目：国家自然科学基金项目(52405294)；贵州省科技计划项目(黔科合支撑〔2021〕一般 171)。
作者简介：刘愚成，硕士研究生，主要从事智慧农业的研究。
通信作者：李云伍，博士，教授。

duce the dependence on high-quality anchor boxes and improve the generalization ability of the model. Tested on a self-constructed dataset, the YOLOv8-SEW model exhibited precision, recall and mean average precision values of 94.5%, 85.7% and 92.4% separately. Compared with before improvement, the precision, recall and mean average precision of the model was increased by 1.0%, 4.2% and 2.9%, respectively. The detection time for a single image was 44.8 ms, enabling rapid and accurate identification of lemon fruits, thus providing a technological foundation for automatic harvesting lemon fruits.

Key words: recognition of lemon fruits; YOLOv8; SPD convolution module; Wise-IoU loss function; attention mechanism

丘陵山区的柠檬因其皮厚气香、出汁率高而受到市场的广泛欢迎，在农业和食品行业中具有重要的经济价值和市场需求[1-2]。而农业机械化帮助实现规模化种植与采摘是农业现代化的重要组成部分，且是乡村振兴这一国家战略的重要手段[3]。传统的柠檬果实检测方法通常依靠人工视觉判断，效率低下且容易受主观因素影响。因此，开发高效精确的自动化柠檬目标检测模型对于提高采摘效率、降低人工成本具有重要意义[4]。

当前，基于卷积神经网络的目标检测算法在果实检测领域取得显著进展[5]。深度学习目标检测算法分为两类：一类是基于候选区域的双阶段目标检测算法[6]，主要代表算法有RCNN(Regions with Convolutional Neural Network)和Faster RCNN[7-9]等；另一类是基于回归的单阶段目标检测算法，主要代表算法有SSD(Single Shot Detector)[10]和YOLO(You Only Look Once)[11-14]等。相比其他目标检测算法，YOLO具有检测速度快、流程简单等优势。这种算法可以同时完成物体识别和位置定位的任务，而且不需要额外的候选区域提取过程，从而大大减少了计算量和内存占用。由于其高效、准确和易于实现的特点，YOLO系列算法被广泛应用于许多领域，例如果实采摘、自动驾驶、医疗影像分析等。

对于柠檬果实识别，文献[15]提出了改进YOLOv3的柠檬果实检测方法，精确率达到96.28%，但未以平均精度均值做参考，将相同网络用于葡萄数据集后出现大幅度精确率下降，模型依赖高质量锚框且泛化能力较弱；而对于其他果实识别，文献[16]在不均匀的环境条件下，用深度卷积神经网络对西红柿进行检测后精度达到98.3%，但YOLOv3训练后的权重过大，不利于在移动设备上部署；文献[17]在YOLOv4的基础上增强了特征提取并加深了网络结构，虽然检测樱桃果实的平均精度均值比改进前提高了0.15%，但检测精度仍然不佳；文献[18]在YOLOv5中加入CBAM注意力机制和Hardswish激活函数后提高了对咖啡瑕疵豆的精度均值，但光照对检测精度影响很大；文献[19]将Light-YOLO用于芒果果实的检测，在颈部网络中引入残差结构并加入注意力机制，从而提高模型的检测能力，但该模型在检测被严重遮挡的芒果时表现不佳。2023年初，Ultralytics公司发布了具有更快推理速度和更高精度的YOLOv8系列，且训练和调整也更加容易，现已成为最受欢迎的目标检测算法之一，在果实识别等领域得到了广泛应用[20]。文献[21]提出了一种改进MHSA-YOLOv8的模型来检测番茄果实的成熟度，在测试场景中的平均精度均值达到86.4%，但由于只引入了MHSA注意力机制，在精度和召回率方面都有一定的提升空间；文献[22]提出了改进YOLOv8的板栗果实识别方法，引入了加权双向特征金字塔网络，更改了边界框损失函数，加快了模型的收敛速度，且使模型的召回率和平均精度均值分别提升了1.5%和1.8%；文献[23]基于YOLOv8引入针对小目标的空间到深度的卷积，提升了检测草莓果实的精度，但由于自然场景中的柠檬果实与树叶色彩相近，且容易被遮挡，因此精确检测复杂自然背景中的柠檬是一个技术难点。

为改进柠檬果实的识别精度，提出一种基于改进YOLOv8的目标检测算法，在主干网络中加入针对小目标的SPDConv(Space-to-Depth Convolution)模块[24]和EMA(Efficient Multi-Scale Attention)高效的多尺度注意力模块[25]，精确识别复杂自然环境中的柠檬，并将损失函数更改为Wise-IoU(Wise Intersection over Union) Loss[26]，以期降低训练的损失值并提高模型的收敛速度。

1　数据样本采集和预处理

1.1　数据样本采集

以潼南柠檬为研究对象，于重庆市潼南区柏梓镇柠檬种植基地进行数据采集。在柠檬果实即将成熟的季节，在专业摄影软件里设置好快门速度、ISO 感光度和白平衡等参数后使用手机摄像头进行拍摄，分别采集不同成熟度、多拍摄角度(平拍、仰拍和俯拍)和多光照角度(顺光和逆光)的柠檬图像，以保证识别的准确度和鲁棒性。采集图像分辨率为 4032×3024，颜色为 sRGB，格式为 jpg。去除图像重复、抖动造成的图像模糊和完全不存在的柠檬图像后，得到 890 幅柠檬图像进行后续的训练，如图 1 所示。

a. 顺光　b. 逆光　c. 侧光　d. 被枝干遮挡　e. 被树叶遮挡　f. 果实重叠

图 1　成熟柠檬图像

1.2　数据样本预处理和图像增强

在复杂情况下，柠檬图像的颜色会因为不同天气和不同角度的光照而产生变化。这种变化导致采集到的柠檬果实图像之间的颜色差异较大。此外，有的柠檬会被遮挡，有的柠檬距摄像头较远，这些情况都会使得果实的形状特征难以被准确提取。为了确保数据的准确性，使用标注软件 LabelImg 对经过筛选的柠檬图像进行人工标注，得到包含柠檬果实宽、高和中心坐标的 xml 文件。为了适应 YOLO 模型所需的标注文件类型(txt 格式)，使用 Python 编程将之前的 xml 格式标注文件转换成 txt 格式标注文件。这样就得到了可以在 YOLO 模型中使用的数据集。YOLO 数据的格式如图 2 所示。

class	$\overline{x_c}$	$\overline{y_c}$	$\overline{w}$	$\overline{h}$

图 2　YOLO 数据格式

由于光照和天气等因素的不确定性，视觉信息对模型的训练和应用有着显著的影响。为了增加训练集的多样性，更好地捕捉柠檬的特征，对训练集图像进行了数据增强。其目的是使生成的数据更接近真实数据的分布，从而提高检测模型的准确性。此外，数据增强还可以大大减少机器学习中的泛化误差，使模型学习到更加具有鲁棒性的特征。为了避免训练过拟合和样本不均衡，利用 Python 编程对训练集中的图像进行了随机加噪声、改变亮度、裁剪、镜像和旋转的数据增强处理，以适应不同光照和天气条件下的柠檬图像，并提高其在实际场景中的测试表现，数据增强的具体方法如图 3 所示。通过以上方法，最终的样本图像扩增至 4 450 幅，其中将训练集、验证集、测试集按 8∶1∶1 的比例划分，得到训练集 3 560 幅，验证集 445 幅，测试集 445 幅。

a. 原图　b. 旋转　c. 镜像　d. 裁剪　e. 改变亮度

图 3　数据增强

2 改进的YOLOv8-SEW网络模型

2.1 YOLOv8-SEW网络结构

目前，以YOLO系列为代表的基于回归的单阶段算法相比于双阶段算法能够在保持较高的检测精确率的同时实现较快的目标识别速度，非常适合应用于需要实时处理复杂自然环境下目标检测的场景[27]。YOLOv8由主干网络(Backbone)、颈部网络(Neck)和检测头网络(Head)3部分组成，其网络结构如图4所示，组成主干网络和颈部网络的子模块如图5所示，其中：CBS模块由Conv2d(卷积层)+BatchNorm2d(归一化)+SiLU(激活函数)组成；C2f卷积模块通过更多的分支跨层连接，丰富了模型的梯度流，使网络能够学习到更丰富的特征；SPPF模块采用了串行+并行的MaxPool2d(最大池化)处理；节点C表示拼接，节点S表示分割；s表示步长，k表示卷积核大小，n表示有n个模块。经过数据增强后的图片输入到网络后进行预处理，将RGB图片的尺寸缩放至网络设定的640×640×3，随后传入到主干网络中。

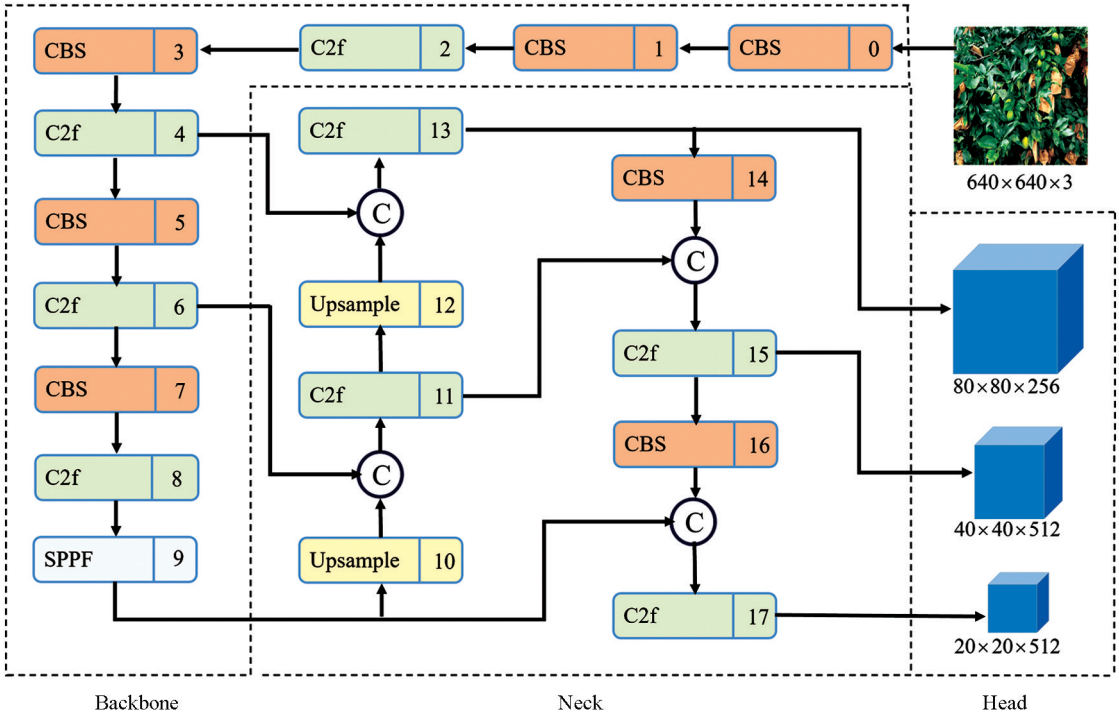


图4 YOLOv8网络结构

YOLOv8-SEW模型结构如图6所示，以YOLOv8系列中参数最小且检测速度最快的YOLOv8n为基线网络进行3项改进。

1) 主干网络中加入了SPDConv卷积模块，达到细粒化的学习特征，避免被遮挡的柠檬果实特征丢失。

2) 引入EMA多尺度注意力模块来保留特征信息同时降低计算成本，确保空间语义特征在每个特征组中均匀分布，从而提高识别的准确性和鲁棒性。

3) 将损失函数替换为Wise-IoU，摆脱模型对高质量锚框的依赖。

2.2 SPDConv(空间到深度的非跨行卷积)

目前卷积神经网络(CNN)架构中存在一些常见的设计缺陷，即使用跨行卷积或者池化层，这种设计会导致细粒度信息的丢失并降低特征学习有效性。为了解决此问题，引入一种新的CNN模块—SPDConv。SPDConv由一个空间到深度层后接一个非跨行卷积层组成，适用于大多数CNN架构，可以替代原来的CNN架构所使用的跨行卷积或者池化层，在图像分辨率低和目标很小的复杂任务中也具有良好的性能。

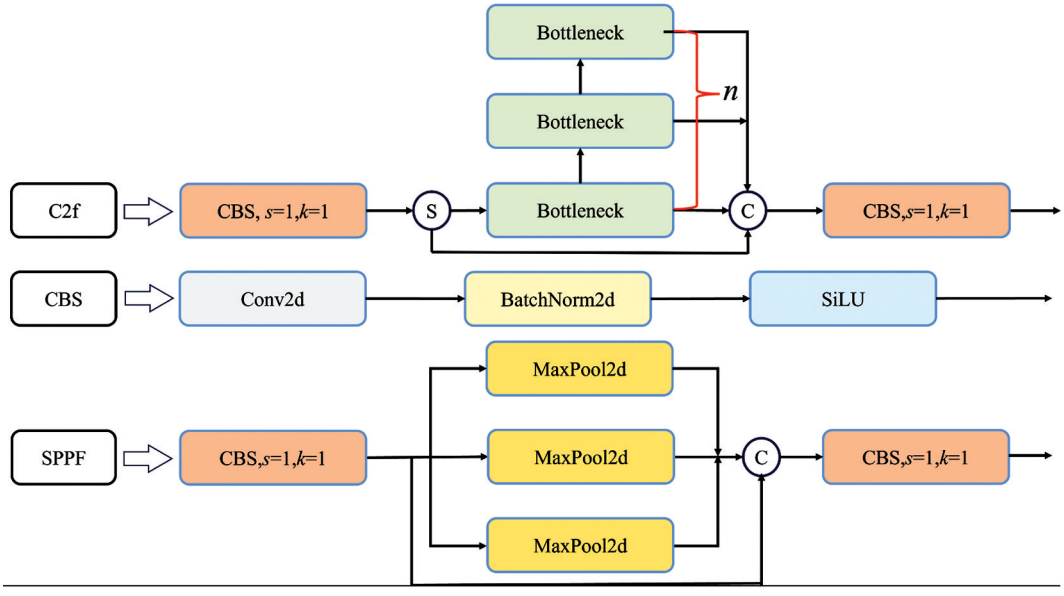


图 5　各模块结构图

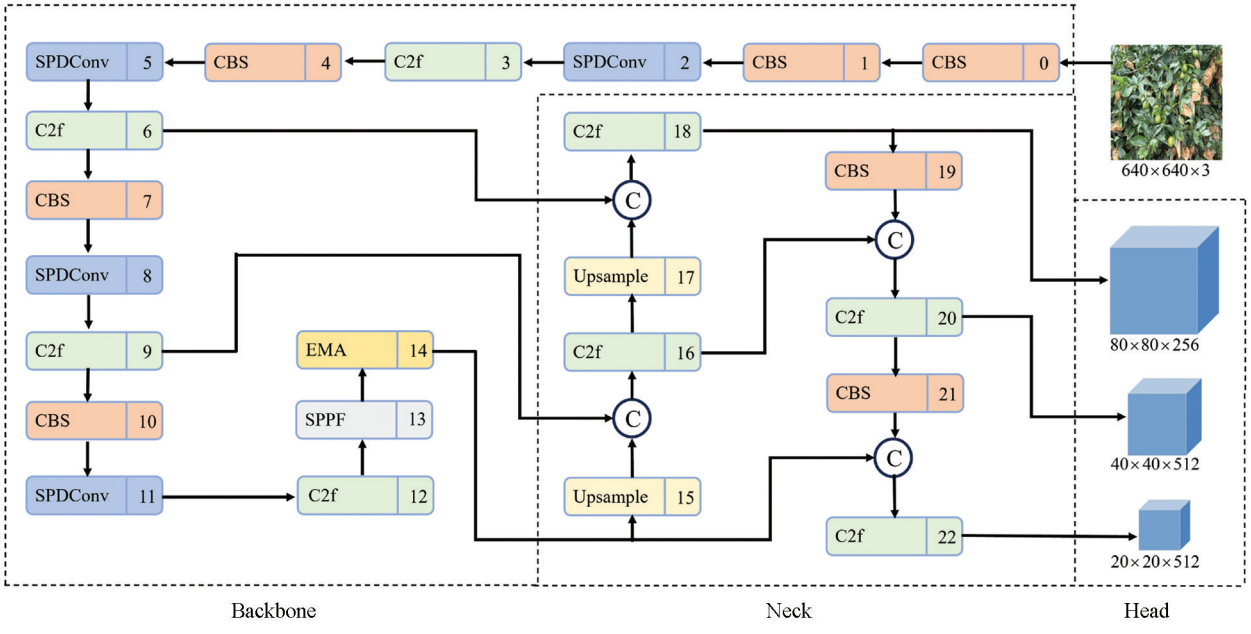


图 6　YOLOv8-SEW 网络结构

原始目标检测模型的设计在处理小目标和低分辨率图像时会出现细粒度信息丢失的问题，导致无法充分学习小目标的特征。为更好地捕捉小目标的细节信息，并提高模型对小目标的检测能力，在 YOLOv8 中引入 SPDConv 卷积模块，以实现有效的特征学习。根据 SPDConv 模块内部原理，对于 CNN 内部的空间到深度层的特征图进行下采样操作。将任意大小的中间特征图 $X(S,S,C_1)$，通过指定 l_{scale}（下采样参数）来切分出一系列子特征图：

$$f_{0,0}=X[0:S:l_{\text{scale}},0:S:l_{\text{scale}}]$$

$$f_{1,0}=X[1:S:l_{\text{scale}},0:S:l_{\text{scale}}]$$

$$\vdots$$

$$f_{l_{\text{scale}}-1,l_{\text{scale}}-1}=X[l_{\text{scale}}-1:S:l_{\text{scale}},l_{\text{scale}}-1:S:l_{\text{scale}}]$$

通常子特征图 $f_{x,y}$ 由原始的特征图 $X(i,j)$ 经过下采样操作得到，以图 7 的 $l_{\text{scale}}=2$ 为例，对特征图

X 进行2倍的下采样，得到4个大小为$\frac{S}{2}\times\frac{S}{2}\times C_1$的子特征图 $f_{0,0}$、$f_{1,0}$、$f_{0,1}$、$f_{1,1}$，然后沿着通道维度将这4个子特征图拼接起来，最终得到的特征图 $X'\left(\frac{S}{l_{\text{scale}}},\frac{S}{l_{\text{scale}}},l^2_{\text{scale}}C_1\right)$ 较原始特征图 X 在空间维度减少了一个 l_{scale}，而通道维度增加了一个 l^2_{scale}。在 $C_2<l^2_{\text{scale}}C_1$ 的条件下添加一个步长为1的非跨行卷积来尽可能多地保留特征信息，其中节点 C 表示拼接。

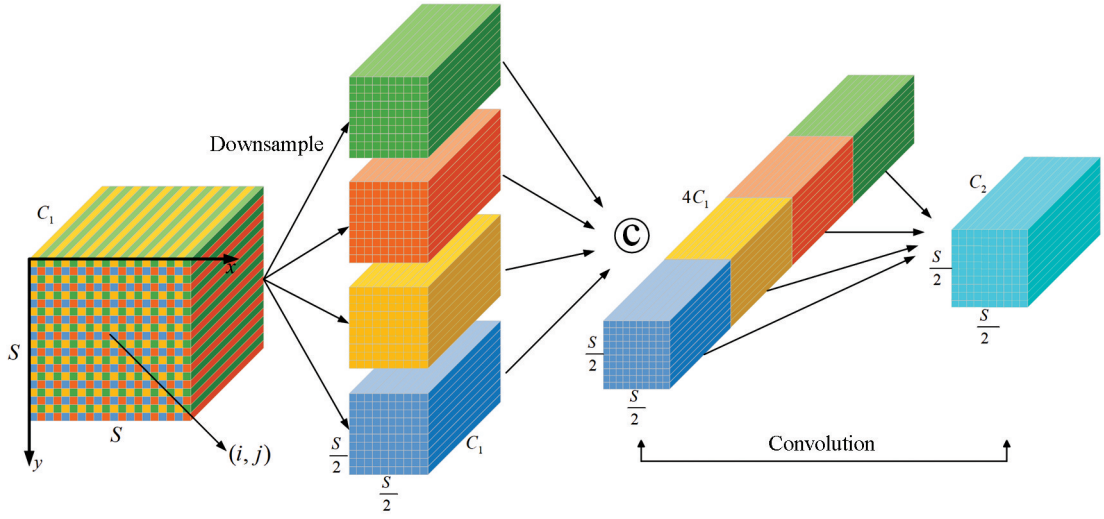


图7 SPDConv卷积模块结构图

2.3 EMA注意力机制

在各种目标检测任务中，注意力机制[28-31]发挥了重要作用，注意力机制的两种类型(通道注意力机制[32]和空间注意力机制[33])可显著增加可识别的特征表示。作为通道注意力机制的代表，SE(Squeeze-and-Excitation)[34]专注于跨维度交互作用，用于提取通道注意力；CBAM(Convolution Block Attention Module)[35]将跨维度注意力权重整合到输入特征中去，用特征图中空间维度和通道维度之间的关系建立了跨空间和通道的信息；SGE(Spatial Group-Wise Enhance)[36]为改进不同语义子特征表示的空间分布，将通道维度分组为多个子特征。

由上述注意力机制看出，通道降维和分组对特征提取有显著效果，但会降低检测效率和增加延迟。基于以上启发，文献[25]在分组结构的基础上，提出了一种不降维的高效多尺度注意力机制EMA，用多尺度并行子网络来建立短期和长期的依赖关系。该机制将部分通道维度转为多个分组的子特征，以避免使用通用卷积进行降维操作，同时通过跨空间学习方法融合两个并行子特征图。EMA的通道分组和多尺度并行网络更有利于提升检测性能和降低计算开销。EMA注意力机制结构如图8所示，其中：$C//G$ 表示将通道 C 划分为 G 个子特征图。

在图8的结构图中可看出EMA使用了特征分组和多尺度结构，有利于在特征学习中建立有效的短期和长期依赖关系来获得更好的表现。在特征分组中：EMA对于任意的输入特征图 $\boldsymbol{X}\in\mathbb{R}^{C\times H\times W}$，令 $\boldsymbol{X}=[\boldsymbol{X}_0,\boldsymbol{X}_i,\cdots,\boldsymbol{X}_{G-1}]$，取 $G\leqslant C$，沿着通道维度方向将通道 C 划分为 G 个子特征图来学习注意力的权重，并用权重加强每个子特征的表示。在并行子网上：EMA通过3条线路来提取注意力权重并赋值给分组特征图，并在通道方向上对跨通道信息交互进行建模，获取通道之间的依赖关系。在跨空间学习中：提出了一种不同空间维度方向的跨空间信息聚合方法，实现更丰富的特征融合，引入两条1×1支路和一条3×3支路的张量输出，再利用(X，Y)Avg Pool(二维全局平均池化)对1×1支路的输出进行全局空间信息编码，将最小支路的输出直接转化为对应的维度形状，即 $\boldsymbol{R}_1^{1\times C//G}\times\boldsymbol{R}_3^{C//G\times H\times W}$，其中 $\boldsymbol{R}$ 代表输出的特征图。同样在3×3支路对全局空间信息进行编码，最后将3条支路的输出特征图计算为两个空间注意力权重值

的集合，然后用 Sigmoid 函数来获得像素级的特征信息。二维全局平均池化的操作公式如下：

$$z_c = \frac{1}{H \times W} \sum_{j}^{H} \sum_{i}^{W} x_c(i, j) \tag{1}$$

式中：c 是输入通道的个数；H、W 是输入特征的空间维度；z_c 是二维全局平均池化的符号；x_c 是第 c 通道的输入特征。

在 YOLOv8 的颈部网络之前插入一个 EMA 注意力机制，用 EMA 注意力机制将主干网络所输出的特征结合通道信息和上下文信息来区分不同尺度下被遮挡的柠檬果实，并将特征信息传递到特征融合网络中，以提升模型的识别性能。

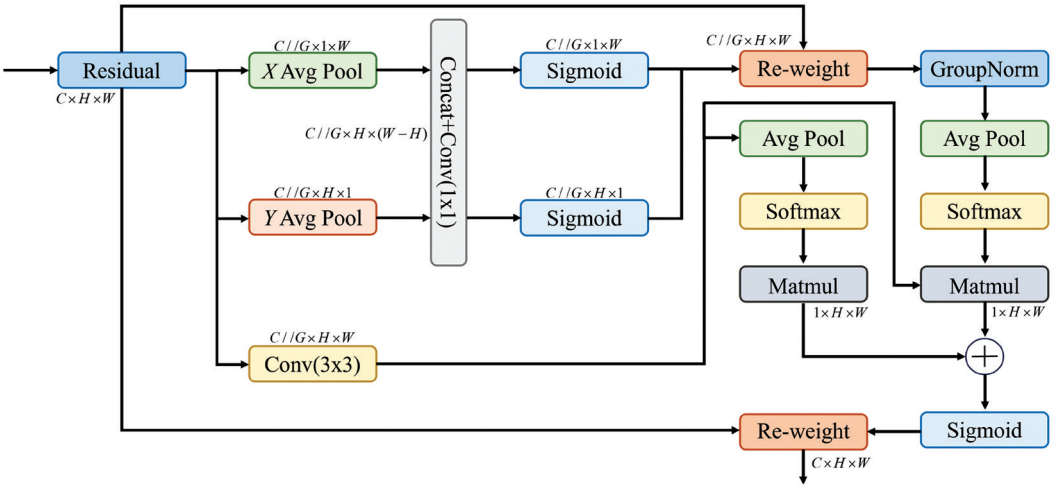


图 8　EMA 注意力机制结构图

2.4　Wise-IoU(WIoU)损失函数

良好定义的边界框回归损失函数能显著提高目标检测的性能，但现有大部分工作都假设训练样本是高质量的，并在此基础上加强损失函数的拟合能力，但这种方法对于一些低质量的样本，则会起反向作用。YOLOv8 模型采用的网络预测边框坐标损失函数是 CIoU(Complete Intersection over Union) Loss，虽然 CIoU Loss 损失函数综合考虑了重叠面积、中心距离和纵横比等影响因素，但当预测框与真实框的高宽比呈线性关系时，CIoU Loss 的惩罚项可能会降至 0。因此无论高质量还是低质量的锚框，都可能对回归损失产生负面影响。

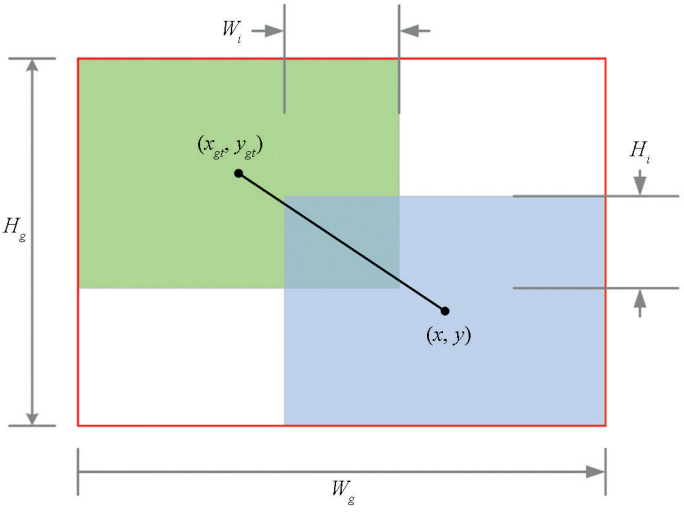


图 9　真实框和预测框交并集面积

在目标检测中，IoU 表示交并比，通常被用来评估模型的预测结果与真实标注间的匹配程度(图 9)。本文将 CIoU 替换成具有动态聚焦机制的边界框回归损失函数 Wise-IoU(WIoU)，其损失函数见式(2)：

$$L_{\mathrm{WIoU}} = rR_{\mathrm{WIoU}}L_{\mathrm{IoU}},\ R_{\mathrm{WIoU}} \in [1, e),\ L_{\mathrm{IoU}} \in [0, 1] \tag{2}$$

式(2)中：L_{IoU}、L_{WIoU} 是以 IoU 和 WioU 为损失的 Loss 函数，R_{WIoU} 是真实框和预测框中心点之间的归一化距离，

$$R_{\mathrm{WIoU}} = \exp\left(\frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g^2 + H_g^2)}\right) \tag{3}$$

$$\beta=\frac{L_{\mathrm{IoU}}^{*}}{L_{\mathrm{IoU}}}\in[0,+\infty) \tag{4}$$

$$r=\frac{\beta}{\delta\alpha^{\beta-\delta}} \tag{5}$$

其中：(x,y)、(x_{gt},y_{gt})是真实框和预测框的中心坐标；W_g、H_g 是边界框的大小；r 是梯度收益因子；α 和δ 表示超参数；β 是锚框的离群度；L_{IoU}^{*} 是 LIoU 的单调聚焦系数。r 通过减少高质量样本对损失值的影响，根据边界框的梯度动态调整增益，并在训练后期降低低质量锚框所产生的有害梯度，使模型更专注于普通质量的锚框，从而提升模型在目标定位上的表现能力。WIoU 使用动态非单调评估锚框质量的机制，降低对高质量锚框的依赖，提高模型的泛化性能和定位对象的能力。

3 实验结果与分析

3.1 实验平台及参数设置

实验平台的硬件配置 CPU 为 Inter i9 13900F，内存为 32 GB，显卡为 NVIDIA GeForce RTX4090，显存为 24 GB。软件配置为 Windows 11 操作系统，采用 Python 3.10 作为编程语言，使用 Pytorch 的深度学习框架和 CUDA 11.8 和 Cudnn 8.9.7 的深度学习加速库。训练过程的参数设置如表 1 所示。

表 1 参数设置

参数	设置	参数	设置
训练周期/次	100	初始学习率	0.01
每批次图片数量/张	64	动量因子	0.937
输入图像尺寸/像素	640×640	权重衰减	0.000 5
优化器	SGD	数据增强	1.0
早停等待轮次/次	50		

3.2 评价指标

以精确率(Precision，P)、召回率(Recall，R)、$F1$ 值、平均精度均值(mean Average Precision，mAP)和权重(Weights)5 个常用指标作为评价模型性能的指标。精确率用于评估模型预测的准确性，召回率用于评估模型预测的完整性，$F1$ 得分是综合考虑精确率和召回率的指标，平均精度均值用于计算不同阈值的平均精确度，权重大小通常是衡量模型复杂度和存储需求的重要指标之一，具体计算公式如下：

$$P=\frac{T_{\mathrm{P}}}{T_{\mathrm{P}}+F_{\mathrm{P}}} \tag{6}$$

$$R=\frac{T_{\mathrm{P}}}{T_{\mathrm{P}}+F_{\mathrm{N}}} \tag{7}$$

$$F1=\frac{2\times P\times R}{P+R} \tag{8}$$

$$AP=\int_{0}^{1}P(R)\mathrm{d}R \tag{9}$$

$$mAP=\frac{1}{N}\sum_{k=0}^{N}AP(k) \tag{10}$$

其中：T_{P} 是判断为正类的正类数；F_{N} 是判断为负类的正类数；F_{P} 是判断为正类的负类数；N 表示参与计算的类别数量。

3.3 消融实验

为验证改进方法的有效性，以上述评价指标为参考，使用相同的数据集、实验平台和训练参数，通过

消融实验对每个改进点进行排列组合实验来评估对最初算法的优化效果。改进方法分别为 EMA、SPDConv 和 WioU，实验方法如下：① 在原始 YOLOv8 的基础上，评估单独加入各改进方法对原始算法的提升效果；② 评估对 3 个改进点进行两两组合后的实验效果；③ 加入 3 个改进方法后寻找最优组合。由表 2 可知，与原始算法相比，3 种改进方法均有提升目标检测的效果，其中加入 SPDConv 后对模型的提升效果最为明显，精确率、召回率、$F1$ 值和平均精度均值分别提升了 1.3%、0.1%、0.6%和 1.5%。在权重稍有增加的情况下，经过 3 项改进方法后的 YOLOv8-SEW 模型，精确率、召回率、$F1$ 值和平均精度均值分别提升了 1.0%、4.2%、2.8%和 2.9%。

表 2　消融实验

序号	EMA	SPDConv	WIoU	精确率/%	召回率/%	$F1$/%	平均精度均值/%	权重/MB
1	—	—	—	93.5	81.5	87.1	89.5	6.0
2	√	—	—	93.4	81.1	86.8	90.4	6.0
3	—	√	—	94.8	81.6	87.7	91.0	6.4
4	—	—	√	92.2	83.2	87.5	90.7	6.0
5	√	√	—	94.2	82.9	88.2	91.7	6.4
6	√	—	√	92.9	84.1	88.3	91.0	6.0
7	—	√	√	94.3	84.5	89.1	92.1	6.4
8	√	√	√	94.5	85.7	89.9	92.4	6.4

注：√表示使用该模块；—表示未使用该模块。

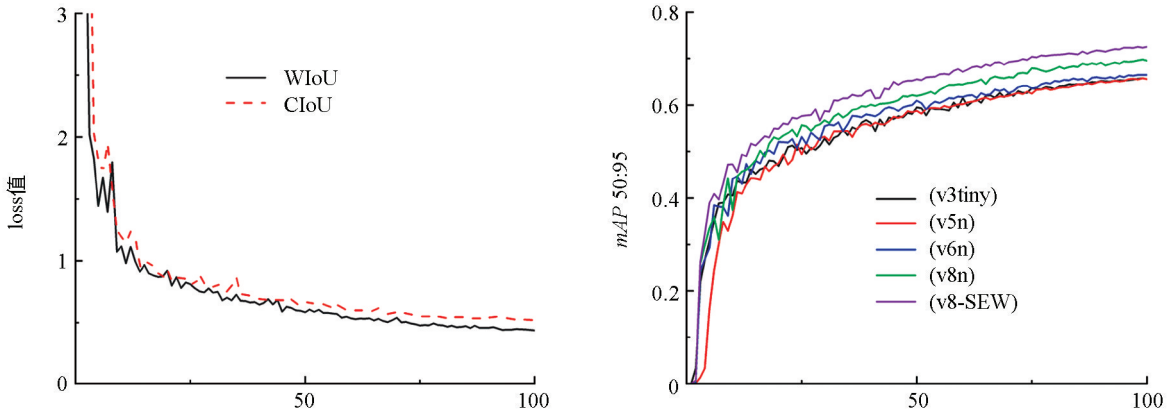
3.4　对比实验

为保证本研究的算法适合使用在各种移动设备上，选择单阶段目标检测算法各个版本参数量最少的模型进行柠檬识别对比实验。在相同的数据集上将本研究的 YOLOv8-SEW 网络与 YOLOv3tiny、YOLOv5n、YOLOv6n 和 YOLOv8n 进行训练与测试。由表 3 可知，与其他主流算法相比，YOLOv8-SEW 具有更好的精确率、召回率、$F1$ 值和平均精度均值。与改进前的 YOLOv8n 相比，在精确率、召回率和平均精度均值上也有 1.0%、4.2%和 2.9%的提升。

表 3　主流算法对比实验

方法	精确率/%	召回率/%	$F1$/%	平均精度均值/%	权重/MB	检测时间/(帧·s^{-1})
YOLOv3tiny	93.4	80.4	86.4	88.2	23.2	26.6
YOLOv5n	93.1	81.1	86.7	89.1	3.7	22.9
YOLOv6n	92.5	77.1	84.1	86.8	8.3	31.9
YOLOv8n	93.5	81.5	87.1	89.5	6.0	28.4
YOLOv8-SEW	94.5	85.7	89.9	92.4	6.4	22.3

由图 10 可以看出，将损失函数由 CIoU 替换成 WIoU 后加快了模型的收敛速度并且损失值更小；mAP 50∶95 曲线图可以看出 YOLOv8-SEW 网络在不同 IoU 阈值下的平均精度均值优于所有基线网络。为更为直观地表现改进前后模型的检测性能，对部分图片进行测试，统计结果见表 4，实验效果图见图 11。由表 4 可知改进后漏检率降低 5.4%，误检率降低 4.9%。由图 11 可以看出改进后模型的效果优于改进前。尽管在加入针对小目标的卷积模块和注意力机制后，平均检测时间从 28.4 帧/s 降低至 22.3 帧/s，但在实际应用中，目前的帧率仍能满足移动设备的要求。因此，在保证这些要求的前提下，各项精确率和召回率得到了提升，综合性能表现最佳。



a. 损失值曲线　　b. *mAP* 50:95曲线

图10　损失值和 *mAP* 50：95 的曲线图

表4　改进前后检测数量统计

模型	正确数目/个	漏检数目/个	误检数目/个
YOLOv8n	165	20	14
YOLOv8-SEW	175	10	5

图11　改进前后算法测试对比

4　结论

为实现在复杂背景中对柠檬果实的快速精确识别，提出了基于 YOLOv8 的单阶段目标检测算法 YOLOv8-SEW。首先在主干网络中加入 EMA 注意力机制，将所输出通道信息和上下文信息相结合以加强不同尺度下被遮挡的柠檬果实的特征学习；其次引入 SPDConv 卷积模块用于处理小目标和低分辨率图像细粒度信息丢失的问题；最后将损失函数替换为 WIoU，降低模型对高质量锚框依赖的同时提高泛化性能和定位对象的能力，并且梯度下降速度和收敛后损失值均优于改进前。

对 3 个改进点进行排列组合的消融实验和对比实验，结果表明 3 个改进方法对模型的性能均有提升效果，并且优于所有的基线网络模型。在自建柠檬数据集上的实验表明，改进后的 YOLOv8-SEW 模型的精

确率、召回率和平均精度均值分别达到 94.5%、85.7%、92.4%，与 YOLOv8 相比分别增长 1.0%、4.2%、2.9%。

在实际检测场景中随着果树与相机的距离拉远，柠檬果实在图片中占据的像素更小，后续可以加入小目标检测层来提高对小目标果实的检测性能。YOLO 模型只能获取图像的二维坐标信息，后续可通过双目相机获取深度信息进行坐标融合得到果实的三维坐标，为自动化采摘机器提供定位方法。

参考文献：

[1] 潼南区委宣传部. 潼南：小柠檬“走出去”开拓国内国际大市场 [J]. 重庆与世界，2023(12)：56-59.

[2] 代小红，钟光跃，曹树梅，等. 安岳柠檬品牌培育现状及建议 [J]. 四川农业科技，2023(1)：119-121.

[3] 周冰. 农业机械化助力乡村振兴中的影响与作用 [J]. 当代农机，2024(2)：39-40，42.

[4] 郑太雄，江明哲，冯明驰. 基于视觉的采摘机器人目标识别与定位方法研究综述 [J]. 仪器仪表学报，2021，42(9)：28-51.

[5] LECUN Y，BENGIO Y，HINTON G. Deep Learning [J]. Nature，2015，521(7553)：436-444.

[6] ZHANG X H，WANG H P，XU C A，et al. A Lightweight Feature Optimizing Network for Ship Detection in SAR Image [J]. IEEE Access，2019，7：141662-141678.

[7] GIRSHICK R，DONAHUE J，DARRELL T，et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation [C] //2014 IEEE Conference on Computer Vision and Pattern Recognition. New York：IEEE，2014：580-587.

[8] GIRSHICK R. Fast R-CNN [C] //2015 IEEE International Conference on Computer Vision (ICCV). New York：IEEE，2015：1440-1448.

[9] REN S Q，HE K M，GIRSHICK R，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] LIU W，ANGUELOV D，ERHAN D，et al. SSD：Single Shot MultiBox Detector [C] // Computer Vision – ECCV 2016. Cham：Springer International Publishing，2016：21-37.

[11] REDMON J，FARHADI A. YOLO9000：Better，Faster，Stronger [C] //2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York：IEEE，2017：6517-6525.

[12] REDMON J，FARHADI A. YOLOv3：An Incremental Improvement [EB/OL]. (2018-04-08) [2024-03-14]. https：//arxiv.org/abs/1804.02767.

[13] BOCHKOVSKIY A，WANG C Y，LIAO H M. YOLOv4：Optimal Speed and Accuracy of Object Detection [EB/OL]. (2020-04-23) [2024-03-14]. https：//arxiv.org/abs/2004.10934v1.

[14] 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). New York：IEEE，2016：779-788.

[15] LI G J，HUANG X J，AI J Y，et al. Lemon-YOLO：An Efficient Object Detection Method for Lemons in the Natural Environment [J]. IET Image Processing，2021，15(9)：1998-2009.

[16] LAWAL M O. Tomato Detection Based on Modified YOLOv3 Framework [J]. Scientific Reports，2021，11(1)：1447.

[17] GAI R L，CHEN N，YUAN H. A Detection Algorithm for Cherry Fruits Based on the Improved YOLO-V4 Model [J]. Neural Computing and Applications，2023，35(19)：13895-13906.

[18] 张成尧，张艳诚，张宇乾，等. 基于 YOLOv5 的咖啡瑕疵豆检测方法 [J]. 食品与机械，2023，39(2)：50-56，175.

[19] ZHONG Z Y，YUN L J，CHENG F Y，et al. Light-YOLO：A Lightweight and Efficient YOLO-Based Deep Learning Model for Mango Detection [J]. Agriculture，2024，14(1)：140.

[20] RUIZ-PONCE P，ORTIZ-PEREZ D，GARCIA-RODRIGUEZ J，et al. POSEIDON：a Data Augmentation Tool for Small Object Detection Datasets in Maritime Environments [J]. Sensors，2023，23(7)：3691.

[21] LI P，ZHENG J S，LI P Y，et al. Tomato Maturity Detection and Counting Model Based on MHSA-YOLOv8 [J]. Sensors，2023，23(15)：6701.

[22] 李茂，肖洋轶，宗望远，等. 基于改进YOLOv8模型的轻量化板栗果实识别方法[J]. 农业工程学报，2024，40(2)：1-9.

[23] LUO Q，WU C B，WU G J，et al. A Small Target Strawberry Recognition Method Based on Improved YOLOv8n Model[J]. IEEE Access，2024，12：14987-14995.

[24] SUNKARA R，LUO T. No More Strided Convolutions or Pooling：A New CNN Building Block for Low-Resolution Images and Small Objects[C] // Machine Learning and Knowledge Discovery in Databases. Cham：Springer Nature Switzerland，2023：443-459.

[25] OUYANG D L，HE S，ZHANG G Z，et al. Efficient Multi-Scale Attention Module with Cross-Spatial Learning[C] // ICASSP 2023 - 2023 IEEE International Conference on Acoustics，Speech and Signal Processing (ICASSP). New York：IEEE，2023：1-5.

[26] TONG Z，CHEN Y，XU Z，et al. Wise-IoU：Bounding Box Regression Loss with Dynamic Focusing Mechanism[EB/OL]. (2020-01-24)[2024-03-19]. https：//arxiv. org/abs/2301. 10051.

[27] 赵德安，吴任迪，刘晓洋，等. 基于YOLO深度卷积神经网络的复杂背景下机器人采摘苹果定位[J]. 农业工程学报，2019，35(3)：164-173.

[28] BRAUWERS G，FRASINCAR F. A General Survey on Attention Mechanisms in Deep Learning[J]. IEEE Transactions on Knowledge and Data Engineering，2023，35(4)：3279-3298.

[29] GUO M H，XU T X，LIU J J，et al. Attention Mechanisms in Computer Vision：A Survey[J]. Computational Visual Media，2022，8(3)：331-368.

[30] NIU Z Y，ZHONG G Q，YU H. A Review on the Attention Mechanism of Deep Learning[J]. Neurocomputing，2021，452：48-62.

[31] LAI Q X，KHAN S，NIE Y W，et al. Understanding More about Human and Machine Attention in Deep Neural Networks[J]. IEEE Transactions on Multimedia，2020，23：2086-2099.

[32] WAN D H，LU R S，SHEN S Y，et al. Mixed Local Channel Attention for Object Detection[J]. Engineering Applications of Artificial Intelligence，2023，123：106442.

[33] YANG H，YUAN C F，ZHANG L，et al. STA-CNN：Convolutional Spatial-Temporal Attention Learning for Action Recognition[J]. IEEE Transactions on Image Processing，2020，29：5783-5793.

[34] JIN X，XIE Y P，WEI X S，et al. Delving Deep into Spatial Pooling for Squeeze-and-Excitation Networks[J]. Pattern Recognition，2022，121：108159.

[35] 赵茂程，邹涛，齐亮，等. 基于MobileViT-CBAM的枇杷表面缺陷检测方法[J]. 农业机械学报，2024，7(2)：1-10.

[36] LI X，HU X L，YANG J. Spatial Group-Wise Enhance：Improving Semantic Feature Learning in Convolutional Networks[EB/OL]. (2020-04-23)[2024-06-14]. https：//arxiv. org/abs/1905. 09646v2.

责任编辑 张枸
柳剑