

改进 YOLOv5su 模型检测桃树缩叶病

姚凌云, 周俊峰, 李丽

(1. 西南大学工程技术学院, 重庆 400715; 2. 重庆市丘陵山区农业装备重点实验室, 重庆 400715)

摘要: 为实现自然环境下桃树缩叶病的检测, 该研究提出了一种基于 YOLOv5su 的桃树缩叶病识别改进模型 DLL-YOLOv5su。首先, 针对桃树缩叶病目标特征变化较大的问题, 在骨干网络最后一层 C3 模块中加入可变形自注意力模块 (deformable attention, DA), 使模型更加关注目标区域, 降低背景对模型的影响, 提高模型在复杂背景下的拟合能力。其次在 SPPF (fast spatial pyramid pooling) 模块中引入 LSKA (large separable kernel attention) 结构, 大核卷积增大了模型的感受野, 使模型能够关注更多信息。最后, 提出了 LAWD (lightweight adaptive weighted downsampling) 模块, 使用轻量化的下采样结构替换卷积模块, 减少计算开销。在桃树缩叶病数据集上进行试验, 结果显示, DLL-YOLOv5su 模型权重大小为 17.6 MB, 检测速度为 83 帧/s。识别准确率 P 、召回率 R 和平均精度均值 mAP_{50} 分别达到了 80.7%、73.1% 和 80.4%, 相较于原始 YOLOv5su 分别提高了 4.2、2.4 和 4.3 个百分点。与 YOLOv3-tiny、Faster R-CNN、YOLOv7 和 YOLOv8 相比 mAP_{50} 分别高出了 28.5、11.8、2.1 和 4.1 个百分点。改进模型识别精度高, 误检、漏检率低, 检测速度满足实时检测的要求, 可以为桃树缩叶病的实时监测和预警提供参考。

关键词: 图像处理; 病害; 缩叶病; 目标检测; YOLOv5su; 可变形自注意力; 大核卷积; 轻量化

doi: 10.11975/j.issn.1002-6819.202403104

中图分类号: S24

文献标志码: A

文章编号: 1002-6819(2024)-14-0109-09

姚凌云, 周俊峰, 李丽. 改进 YOLOv5su 模型检测桃树缩叶病[J]. 农业工程学报, 2024, 40(14): 109-117. doi: 10.11975/j.issn.1002-6819.202403104 <http://www.tcsae.org>

YAO Lingyun, ZHOU Junfeng, LI Li. Improved YOLOv5su model for detecting peach leaf curl disease[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(14): 109-117. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.202403104 <http://www.tcsae.org>

0 引言

中国是全球最大的桃生产国, 2021 年中国桃子种植面积和产量分别占全球的 54.83% 和 64.08%, 桃产业总产值近千亿元^[1]。然而, 桃树缩叶病是桃树的主要病害之一, 其主要发生在嫩叶, 严重时危害嫩枝、花朵和果实, 造成桃树减产, 甚至影响桃树次年产量。桃树缩叶病的快速和准确的识别, 对防治措施具有重要意义。目前桃树缩叶病一般是果农依靠经验识别, 这种方式不仅识别准确率和效率低, 还会受到果农主观因素影响。因此, 桃树缩叶病发病情况的智能检测是桃产业急需实现的目标。

近年来随着计算机硬件技术的发展, 基于深度学习的目标检测方法被广泛应用于农业领域的识别任务中^[2-4]。王磊磊等^[5]在 YOLOv5 的模型中添加注意力模块, 并改进损失函数构建了 OMM-YOLO 模型用于平菇目标检测, 改进模型对各个成熟度的平菇检测精度均有提高。WANG 等^[6]提出了一种基于两阶段移动视觉的级联害虫检测方法, 该方法将作物分类模型作为预训练模型训练害虫检

测, 通过预先对作物种类进行区分再进行害虫识别解决物种间图像数据不平衡的问题。LI 等^[7]提出了一种基于改进 YOLOv5 的蔬菜病害检测方法, 该方法采用多尺度特征融合策略提高了特征提取能力, 减少了由于复杂背景造成的漏检和误检。兰玉彬等^[8]在 YOLOv5s 模型中加入 CA (coordinate attention, CA) 注意力减少小目标特征丢失的情况, 并引入 GhostNet 网络中的 Ghost 模块轻量化网络, 实现了对自然场景下生姜叶片病虫害识别准确率的提升。SU 等^[9]在 YOLOv5 模型中加入 SE (squeeze-and-excitation, SE) 解决了传统目标检测模型无法有效筛选出芸豆褐斑病的关键特征的问题。XIAO 等^[10]利用增强的 YOLOv7 和边缘计算对荔枝病害进行实时轻量级检测, 相比原始 YOLOv7 模型检测速度更快, 模型参数更少。HOU 等^[11]在 Faster-RCNN 网络中引入特征金字塔结构和 ISResNet (inception Squeeze-and-Excitation ResNet) 构建了一种深度学习模型, 该模型在检测苹果叶片病害方面具有较高的准确性和通用性。李颀^[12]采用基于回归的轻量级检测算法 RFBNet (reception field block network, RFBNet) 检测桃树的病虫害, 其方法识别速度快, 对桃树缩叶病的识别精度达到了 76.58%。YANG 等^[13]提出了一种基于 YOLOv8s 深度学习算法并结合 LW-Swin Transformer 模块的 LS-YOLOv8s 的草莓成熟度检测和分级模型, 增加了 Swin Transformer 模块,

收稿日期: 2024-03-15 修订日期: 2024-05-31

基金项目: 国家自然科学基金项目 (No. 52175121)

作者简介: 姚凌云, 教授, 博士, 博士生导师, 研究方向: 农业机器人动力学及智能检测。Email: lingyunyao@swu.edu.cn

利用多头自注意力机制提高模型的泛化能力。ZHANG 等^[14]提出了一种基于深度学习的苹果病害检测算法,设计了一种 Bole 卷积模块 (Bole convolution module, BCM) 减少苹果叶病图像特征提取过程中的冗余特征信息,提出了一种交叉注意力模块 (cross-attention module, CAM) 降低图像背景信息对模型的干扰。上述研究在农作物病害检测等方面取得了较好的效果,但仍存在如下问题:

多数研究仅在单一叶片上进行,对实际场景中出现的遮挡、光线等影响因素考虑不足;多数研究仅对病害进行分类,并未区分病害发展情况;针对桃树缩叶病识别的研究较少。

针对以上问题,本文以果园中桃树为目标,采集并搭建桃树缩叶病目标检测数据集。以 YOLOv5su 为基础模型,通过引入可变形自注意力机制、可分离大核卷积注意力和轻量化权重自适应下采样模块提高模型对特征的拟合能力,提出了一种基于改进 YOLOv5su 的桃树缩叶病检测模型 DLL-YOLOv5su,为实现桃树缩叶病发生的实时监测提供一种有效方法。

1 图像采集与数据集搭建

1.1 图像采集

本文的桃树缩叶病图像采集地点为中国重庆市农业科学院桃树种植基地,以“紫金红”油桃树为主要研究对象。采集时间为 2023 年 4—6 月,采集设备与方式分别为数码相机 Canon EOS M50 m2 拍摄照片和手机 SAMSUNG SM-G9910 拍摄视频,设备距离目标 20~80 cm,拍摄不同时段和视角下的桃树缩叶病图像,包含从桃树缩叶病发生早期到后期的各阶段图像。拍摄得到 291 张分辨率为 6 000×4 000 像素的桃树图像和 5 分 38 秒视频,视频帧率为 60 帧/s,分辨率为 3 840×2 160 像素。数码相机采集到的图像分辨率高,且包含的内容多,故对 291 张高分辨率图像进行裁剪,筛选出包含缩叶病目标的图像 1 358 张。对视频进行抽帧和筛选,剔除过于模糊和不包含目标的背景图片得到 162 张桃树缩叶病图像。合并两个图像集得到 1 520 张桃树缩叶病图像。

1.2 数据集搭建

采用轻量级图像注释工具“LabelImg”将采集到的桃树缩叶病图像按照 YOLO 格式标注,并以 8:2:1 的比例随机划分为训练集 1 105 张、验证集 277 张和测试集 138 张。由于数据集均于同一场景中拍摄,为避免过拟合,提高模型鲁棒性和泛化能力,采用添加高斯噪声、改变亮度、Cutout、旋转、翻转、平移等图像处理方法随机组合对训练集进行数据增广,扩充后的训练集包含 6 630 张图片。

如图 1 所示,在桃树缩叶病数据集中,每一张图片均包含病害目标,且病害分布随机,包含多种光照强度与方向。其中早期缩叶病标签数一共为 19 419,中期缩叶病标签数为 11 172,后期缩叶病标签数为 10 069。

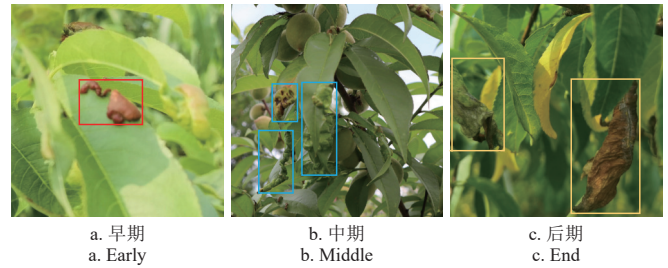


图 1 桃树缩叶病不同发生时期示例

Fig.1 Examples of peach leaf curl disease in different periods

2 桃树缩叶病识别算法与改进

2.1 YOLOv5su 网络

YOLO 是目前流行的目标检测和图像分割模型之一,由华盛顿大学的 Joseph Redmon 和 Ali Farhadi 开发,因其高速度和高精度的优点而迅速流行起来。YOLOv5 主要由输入 (Input)、骨干 (Backbone)、颈部 (Neck) 和头部 (Head) 组成。

YOLOv5u^[15]在 YOLOv5 模型的基础架构上,集成了解耦头部结构,将分类任务与回归任务解耦。同时采用了无需预定义锚框匹配真实框的 Anchor-free 策略避免预定义锚框尺寸不合理的问题。YOLOv5u 根据网络深度和宽度由小到大分为 YOLOv5nu、YOLOv5su、YOLOv5mu、YOLOv5lu、YOLOv5xu 五种不同结构。

2.2 改进 YOLOv5su

本研究的试验图像皆取自室外自然环境中的桃树缩叶病发病区域。为了减轻光照和复杂背景环境等因素的不利影响,提高目标检测模型的鲁棒性,对 YOLOv5su 算法进行了改进,构建了 DLL-YOLOv5su。具体改进措施如下:

1) 在骨干网络最后一层 C3 模块的 Bottleneck 结构中加入 Version Transformer^[16]的可变形自注意 (DA) 模块^[17]组成新的 C3-DA 模块。

2) 在快速空间金字塔池化模块 (SPPF) 中加入了 Version Attention Network^[18]中的可分离大核卷积注意力模块^[19]构成 SPPF-LSKA 模块。

3) 在感受野注意力卷积 (receptive-field attention convolution, RFACnv)^[20]的基础上提出了一种轻量化权重自适应下采样模块 (lightweight adaptive weighted downsampling, LAWD) 替换部分初始网络结构中的 CBS 模块。

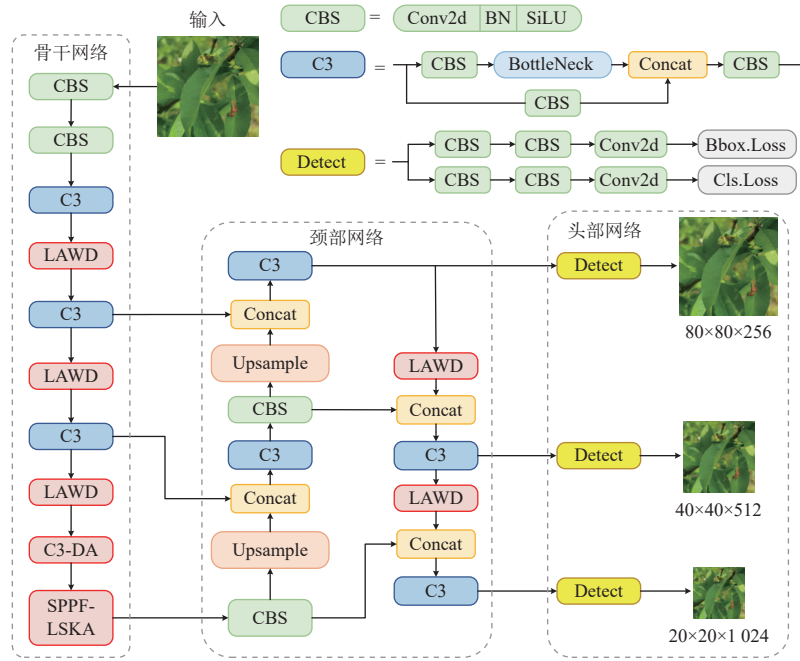
改进后的 DLL-YOLOv5su 模型结构如图 2 所示。

2.2.1 可变形自注意力机制

自注意力最早在自然语言处理模型 Transformer^[21]中被提出。视觉转换器 (version transformer, ViT) 将自注意力机制用于视觉领域,构造了多头注意力模块 (multi-head self-attention)。自注意力有三个重要元素: 查询 (query)、键 (key) 和值 (value)。其输出式如下:

$$\text{Attention}(q, k, v) = \text{Softmax}(qk^T / \sqrt{d_k})v \quad (1)$$

式中 q 、 k 和 v 分别表示查询向量、键向量和值向量, d_k 表 k 的维度, Softmax 表示归一化。

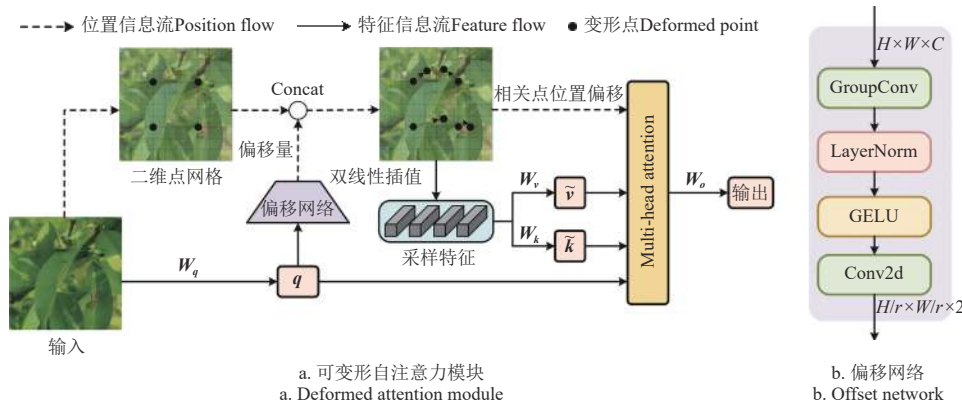


注：Upsample 为上采样模块，Concat 为从通道维度进行特征图拼接，Conv2d 为卷积层，BN 表示批量归一化层，SiLU 为激活函数，Bbox.Loss 为定位损失，Cls.Loss 为分类损失。
 Note: Upsample is the upsampling module, Concat is the feature map concatenation from the channel dimension, Conv2d is the convolutional layer, BN represents the batch normalization layer, SiLU is the activation function, Bbox.Loss is the regression loss, and Cls.Loss is the classification loss.

图 2 DLL-YOLOv5su 模型结构图
 Fig.2 DLL-YOLOv5su model structure diagram

自注意力可以捕获全局信息，并建立特征通道和目标之间的长期依赖关系。然而简单地扩大感受野，会导致内存和计算资源的浪费，并且特征会受到背景区域影响^[17]。为了有效地建模和捕获全局语义信息，在主干网络最后一层 C3 模块中加入可变形自注意力模块强化缩

叶病目标信息，抑制背景信息。可变形自注意力使用多组可变形的采样点确定特征图的重要区域，并以此重要区域为依据对桃树缩叶病不同特征之间的关系进行建模。这种灵活的方案使变形自注意力模块能够专注于相关区域并捕获更多信息特征。可变形自注意力结构如图 3 所示。



注： q 表示查询向量， \tilde{k} 和 \tilde{v} 分别表示变形的键和值向量， W_q 、 W_k 、 W_v 和 W_o 分别表示查询向量、键向量、值向量和输出向量的投影矩阵， z 表示输出特征图， H 、 W 和 C 分别表示图像高度、宽度和通道数， r 为下采样缩减比例，GroupConv 为组卷积，LayerNorm 为层归一化，GELU 为激活函数。
 Note: q is the query vector, \tilde{k} and \tilde{v} represent the deformed key and value vectors, \tilde{x} represents the feature map sampling result, W_q , W_k , W_v and W_o denote the projection matrices of the query vector, key vector, value vector and output vector respectively, z is the output feature map, H , W and C represent the image height, width, and number of channels, r is the downsampling reduction ratio, GroupConv is a group convolution, LayerNorm is a layer normalization, and GELU is an activation function.

图 3 变形注意力模块与偏移网络结构图
 Fig.3 Deformable attention module and offset network structure diagram

如图 3a 所示，变形注意力模块首先对输入特征图 x 生成一个均匀的二维点网格 p ，网格中的点作为采样点。将特征图线性投影到查询向量，然后馈送到偏移子网络 θ_{offset} 以生成偏移量 Δp ，使用此偏移量与采样点相加得到

变形点。在变形点处利用双线性插值函数进行采样得到采样特征 \tilde{x} 。其式为：

$$q = xW_q, \tilde{k} = \tilde{x}W_k, \tilde{v} = \tilde{x}W_v \quad (2)$$

$$\Delta p = \theta_{\text{offset}}(q) \quad (3)$$

$$\tilde{\mathbf{x}} = \phi(\mathbf{x}; p + \Delta p) \quad (4)$$

$$\phi(\mathbf{z}; (p_x, p_y)) = \sum_{(r_x, r_y)} g(p_x, r_x) g(p_y, r_y) \mathbf{z}[r_x, r_y, :] \quad (5)$$

式中 \mathbf{W}_q 、 \mathbf{W}_k 、 \mathbf{W}_v 分别为查询向量、键向量和值向量的投影矩阵, $g(a, b) = \max(0, 1 - |a - b|)$ (r_x, r_y) 索引 $\mathbf{z} \in \mathbb{R}^{H \times W \times C}$ 上的所有位置。由于 $g(a, b)$ 仅在最接近 (p_x, p_y) 的 4 个积分点上非零, 可以简化式 (5) 为

$$\mathbf{z}^{(m)} = \sigma \left(\frac{\mathbf{q}^{(m)} \tilde{\mathbf{k}}^{(m)T}}{\sqrt{d}} + \phi(\widehat{\mathbf{B}}; \mathbf{R}) \right) \tilde{\mathbf{v}}^{(m)} \quad (6)$$

$$m = 1, \dots, M. \quad (7)$$

$$\mathbf{z} = \text{Concat}(\mathbf{z}^{(1)}) \mathbf{W}_o \quad (7)$$

式中 $\mathbf{z}^{(m)}$ 表示第 m 个头部输出向量, $\sigma(\cdot)$ 表示 Softmax 函数, M 表示多头注意力头部数量, $d = C/M$ 表示每个头部的维度, $\mathbf{q}^{(m)}$ 、 $\tilde{\mathbf{k}}^{(m)}$ 、 $\tilde{\mathbf{v}}^{(m)} \in \mathbb{R}^{N \times d}$ 分别表示查询、键和值向量的对应头的元素, \mathbf{W}_o 表示输出向量投影矩阵。 $\phi(\widehat{\mathbf{B}}; \mathbf{R}) \in \mathbb{R}^{H \times W \times H \times W}$ 对应于 Swin Transformer^[22] 中的位置嵌入。将每个头部的特征连接在一起并通过 \mathbf{W}_o 投影得到输出 \mathbf{z} 。将变形注意力模块加入到 C3 模块中组成 C3-Deformable attention 模块和 Bottleneck-DA 模块结构, 如图 4 所示。

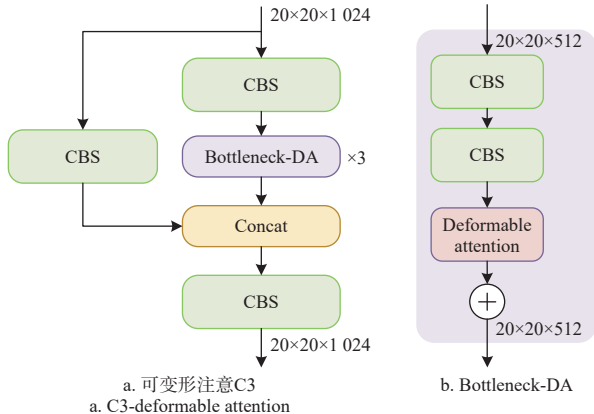


图 4 C3-DA 模块结构图

Fig.4 C3-DA module structure diagram

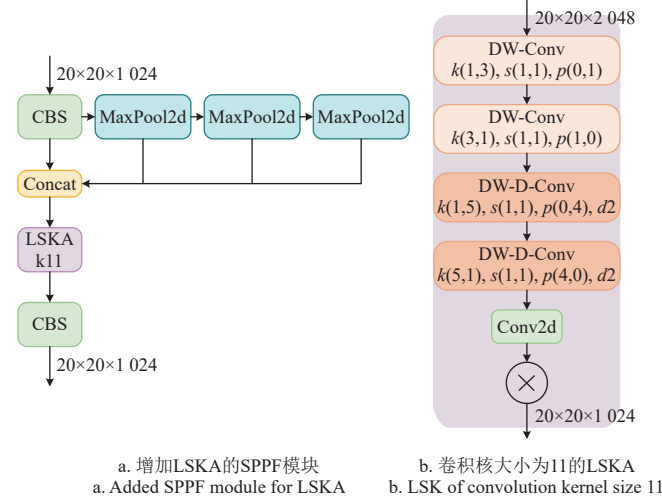
2.2.2 可分离大核卷积注意力机制

大卷积核注意力 (large kernel attention, LKA) 模块^[18] 在 VAN 中发挥了卓越的性能, 使得 VAN 在一系列基于视觉的任务中的表现超过了 ViT 和 CNN 等模型。LKA 吸收了卷积和自注意力的优点, 包括局部结构信息、长程依赖和适应性, 同时避免了通道维度上忽略适应性等缺点。大核卷积模块包括 3 个部分: 深度卷积 (depth-wise convolution, DWC)、深度扩张卷积 (depth-wise dilation convolution, DWDC) 和通道卷积 (1×1 convolution)。LKA 模块可以表示为

$$\text{LKA}(\mathbf{x}) = C^{1 \times 1}(\text{DWD_C}(\text{DW_C}(\mathbf{x}))) \quad (8)$$

式中 \mathbf{x} 是输入特征图, $\text{LKA}(\mathbf{x})$ 是注意力输出特征图, $C^{1 \times 1}$ 表示 1×1 卷积, $\text{DW_C}(\cdot)$ 表示深度卷积, $\text{DWD_C}(\cdot)$ 表示深度扩张卷积。

LSKA 将 LKA 中的深度卷积层的二维卷积核分解为级联的水平和垂直一维卷积核。与标准的 LKA 相比, LSKA 可以提供相当的性能, 并且计算复杂度和内存占用更低。为了提高 SPPF 的特征信息获取能力, 在 SPPF 层中引入可分离大卷积核注意力, 构成 SPPF-LSKA 层, 其模型如图 5 所示。



a. 增加LSKA的SPPF模块

a. Added SPPF module for LSKA

b. 卷积核大小为11的LSKA

b. LSKA of convolution kernel size 11

注: MaxPool2d 为最大池化层, DW-Conv 为深度卷积, DW-D-Conv 为深度扩张卷积, k 表示卷积核, s 表示步长, p 表示像素填充, $d2$ 表示卷积核元素间隔为 1。

Note: MaxPool2d is the maximum pooled layer, DW-Conv is the deep-wise convolution, DW-D-Conv is the deep-wise dilation convolution, k represents the convolution kernel, s represents the step size, p represents the pixel filling, and $d2$ represents the convolution kernel element interval of 1.

图 5 SPPF-LSKA 模块结构图

Fig.5 SPPF-LSKA module structure diagram

2.2.3 轻量化权重自适应下采样

标准卷积运算提取特征图是通过卷积核与相同尺寸的感受域相乘然后求和得到。每个感受域相同空间位置处的特征共享相同的卷积核参数。因此, 标准卷积操作没有考虑不同位置所包含的信息差异, 这在一定程度上限制了卷积神经网络的性能。ZHANG 等^[23] 观察到组卷积可以减少模型的参数和计算开销, 但是组内信息之间交互不足会影响网络性能; HUANG 等^[24] 通过重用特征进行融合来改进特征信息, 以解决网络梯度消失的问题; DAI 等^[25] 提出了可变形卷积, 通过学习偏移来改变卷积核的采样位置, 进一步提高了卷积神经网络的性能。ZHANG 等^[20] 提出了感受野注意力卷积 (receptive-field attention convolution, RFACConv), 该方法动态地确定每个特征在感受野中的权重。

本文为减少标准卷积的卷积核参数共享导致的空间信息损失与轻量化模型, 在 RFACConv 的基础上提出轻量化权重自适应下采样 (lightweight adaptive weight downsampling, LAWD) 替换模型中的部分卷积模块。其结构如图 6 所示。

LAWD 上半部分为权重路线, 首先采用的平均池化聚合每个感受域信息, 再通过 1×1 卷积交互不同通道间的空间信息, 提高网络性能^[26-28]。对得到的权重矩阵进行维度转换, 这样在既不损失权重信息的前提下可以有

效减少网络的参数量。

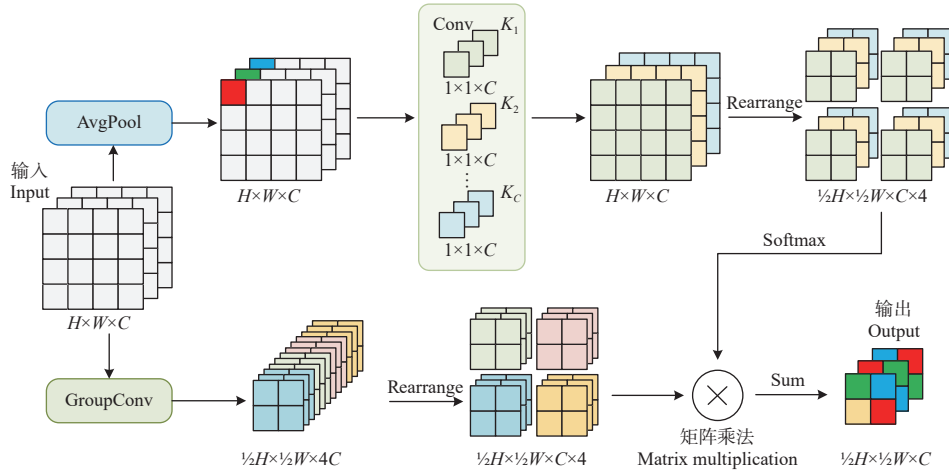
下半部分为特征提取路线, 首先通过组卷积, 对输入特征图进行切片操作, 将输入特征图的 W 、 H 维度信息保存到通道中, 再进行卷积操作得到的二倍下采样特征图可以很好地保留原始图像信息。相较于普通卷积, 采用分组大小为 16 的分组卷积可以将参数量减少到普通卷积的 $1/16$ 。再对特征图进行维度变换, 并与权重相乘。最后对第四个维度进行求和得到最终下采样的特征图。LAWD 可以通过廉价的操作为不同的感受域生成权重矩

阵, 避免了卷积操作的卷积核参数共享问题, 并且减少了模型参数量。LAWD 计算过程可以表示为

$$\text{LAWD}(\mathbf{x}) = \text{Sum}(\text{Softmax}(\mathbf{C}^{1 \times 1}(\text{Avgpool}(\mathbf{x}))) \times \text{SiLU}(\text{BN}(g^{3 \times 3}(\mathbf{x})))) \quad (9)$$

式中 \mathbf{x} 表示输入特征图, $g^{3 \times 3}$ 表示 3×3 组卷积。

改进后的 DLL-YOLOv5su 桃树缩叶病检测整体网络结构中, LAWD 模块替换了骨干网络后三层 CBS 模块和颈部网络的两层下采样卷积。



注: AvgPool 为平均池化层, Rearrange 为张量维度调整, $K_1 \sim K_c$ 表示第 1~ C 的卷积核, Sum 表示求元素和。

Note: Avgpool is the average pooling layer, Rearrange is the tensor dimension adjustment, $K_1 \sim K_c$ represents the 1st to C convolution kernel, and Sum is the element sum.

图 6 轻量级权重自适应下采样结构图

Fig.6 Lightweight adaptive weight downsampling structure diagram

2.3 试验环境与评价指标

2.3.1 试验环境设置

本文研究中针对桃树缩叶病数据集的训练和测试的试验平台环境为: 操作系统为 64 位 Windows11, 处理器型号为 12 th Gen Inter (R) Core (TM) i5-12400F 2.5 GHz, 显卡型号为 Nvidia GeForce RTX 4060Ti (8G), 内存为 32GB (3 200 MHz), 深度学习框架为 PyTorch 2.0, CUDA 版本为 11.7, 编程平台为 PyCharm, 编程语言为 Python3.11。

本文所有试验均在相同环境下进行, 选择 YOLOv5su 作为桃树缩叶病检测的原始模型, 为避免过高的学习率导致训练精度震荡, 模型收敛困难, 超参数文件选择 low 版本。训练采用的批量大小设置为 8, 初始学习率为 0.01, 动量设置为 0.937, 训练图片尺寸为 640×640 像

素, 训练轮数设置为 300 轮。

2.3.2 评价指标

本文主要以准确率 (precision, P)、召回率 (recall, R)、平均检测精度 (average precision, AP), 平均检测精度均值 (mean average precision, mAP)、每秒帧数 (frames per second, FPS) 和权重大小 (MB) 作为评价指标。

3 结果与分析

3.1 基线模型选择

YOLOv5su 是在 YOLOv5s 基础上提出的, 为了选择合适的基线模型, 本文对部分版本进行了对比试验, 试验结果如表 1 所示。

表 1 基线模型对比结果

Table 1 Baseline model comparison results

模型 Models	准确率 Precision P /%	召回率 Recall R /%	平均精度均值 Mean average precision/%		权重大小 Model size/MB	帧速率 Frame per second FPS/(帧·s ⁻¹)	平均精度 Average precision AP/%		
			mAP ₅₀	mAP _{50~95}			早期	中期	后期
							Early	Mid	End
YOLOv5nu	75.3	68.9	73.8	44.3	7.6	105.7	75.1	76.2	70.1
YOLOv5su	76.5	70.7	76.1	46.8	18.5	119.7	78.3	75.3	74.6
YOLOv5 mu	77.1	72.4	78.6	48.0	50.5	78.0	79.9	80.2	75.6
YOLOv5s	74.4	69.6	73.0	44.8	14.5	83.0	78.0	70.1	70.7

注: mAP₅₀ 为 IoU 阈值为 0.5 时的平均精度均值, mAP_{50~95} 为 IoU 阈值为 0.5 到 0.95, 步长为 0.05 时 mAP 的平均值。

Note: mAP₅₀ is the average precision when the intersection over union threshold is 0.5, mAP_{50~95} is the average value of mAP for IoU thresholds of 0.5 to 0.95 in steps of 0.05.

由表1结果可以看出, YOLOv5su模型检测效果优于YOLOv5nu, 模型权重大小为18.5 MB, 远小于YOLOv5mu的50.5 MB。对比YOLOv5s, YOLOv5su在准确率、召回率和平均检测精度方面均优, 模型增加了4 MB。因此, 引入无锚框的分体式头部结构有效地提升了模型性能。综合以上因素, 本文选择YOLOv5su作为基线模型。

3.2 消融试验结果

为了验证本文改进和提出的方法对于桃树缩叶病检测是否有效, 以及评估各个方法所达到的效果, 设计了消融试验。试验结果如表2所示, 模型改进训练过程参

数如图7所示。

由消融试验结果可以看出各个改进方法的效果: 改进C3-DA模块之后, 模型相较于初始模型召回率提高了3.5个百分点, 平均精度均值 mAP_{50} 提高了2.3个百分点, 但是模型权重增大了0.6 MB; 引入LSKA模块之后, 模型在准确率和平均精度 mAP_{50} 方面分别提高了2.5和2.3个百分点, 模型权重增加了2.2 MB; 使用LAWD替换部分CBS模块之后的模型召回率降低了1个百分点, 但准确率和平均精度均值分别上升了0.4和1.7个百分点, 模型大小减少了18.9个百分点。

表2 消融试验结果
Table 2 Ablation test results

模型 Models	准确率 Precision P/%	召回率 Recall R/%	平均精度均值 Mean average precision/%		权重大小 Model size/MB	帧速率 Frame per second FPS/(帧·s ⁻¹)	平均精度 Average precision AP/%		
			mAP ₅₀	mAP _{50~95}			早期 Early	中期 Mid	后期 End
YOLOv5su	76.5	70.7	76.1	46.8	18.5	119.7	78.3	75.3	74.6
YOLOv5su+C3-DA	76.7	74.2	78.4	48.3	19.1	96.1	79.9	80.4	75.1
YOLOv5su+LSKA	79.0	71.0	78.4	49.0	20.7	114.8	80.4	79.2	75.7
YOLOv5su+LAWD	76.9	69.7	77.8	47.8	15.0	96.5	79.1	78.5	75.9
DLL-YOLOv5su	80.7	73.1	80.4	50.4	17.6	83.0	80.6	81.5	79.1

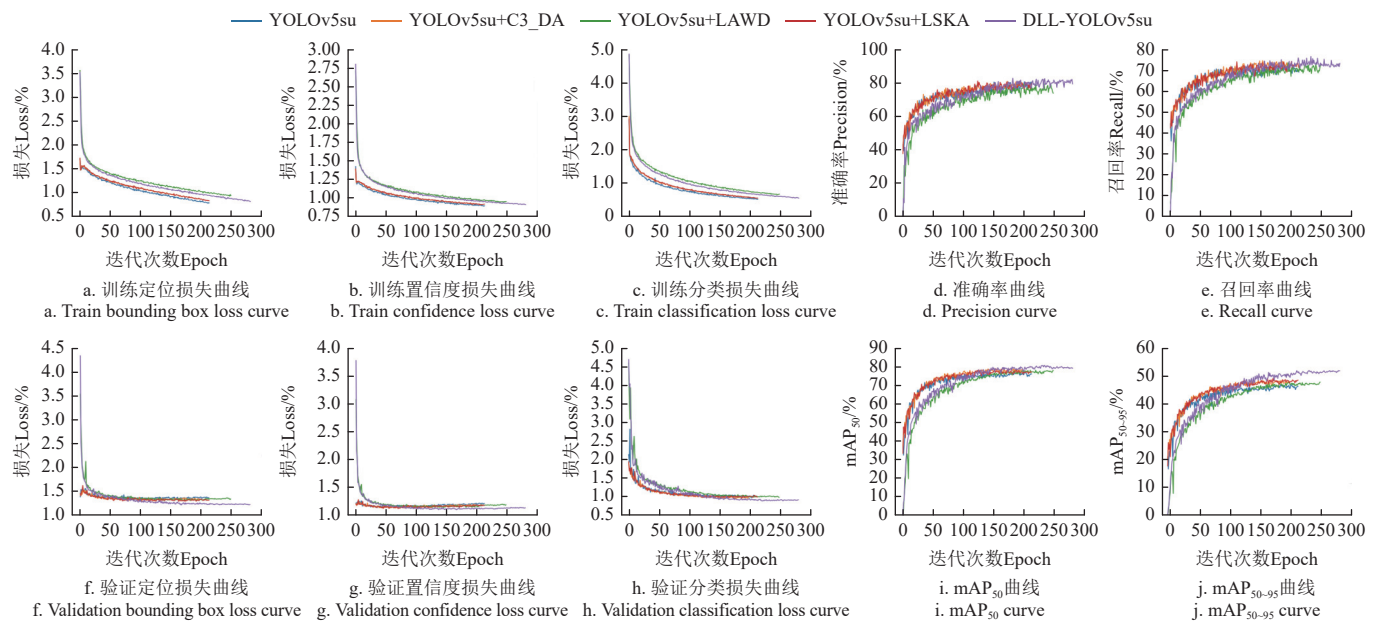


图7 改进模型性能对比曲线

Fig.7 Performance comparison curve of the improved model

由图7可以看出, 改进后的DLL-YOLOv5su模型收敛速度稍慢, 但在训练200轮左右时效果已经超越初始版本的YOLOv5su。最终本研究提出的DLL-YOLOv5su桃树缩叶病检测模型准确率、召回率和平均精度均值分别达到了80.7%、73.1%和80.4%。相比原YOLOv5su模型分别提升了4.2、2.4和4.3个百分点。检测速度为83.0帧/s, 可以满足桃树缩叶病实时检测的要求。

3.3 目标检测模型对比试验

为验证本文提出模型的有效性, 将其他主流目标检测模型与改进后的模型进行对比试验。所有模型均在相同试验环境下, 对相同数据集进行试验, 试验结果如表3所示。

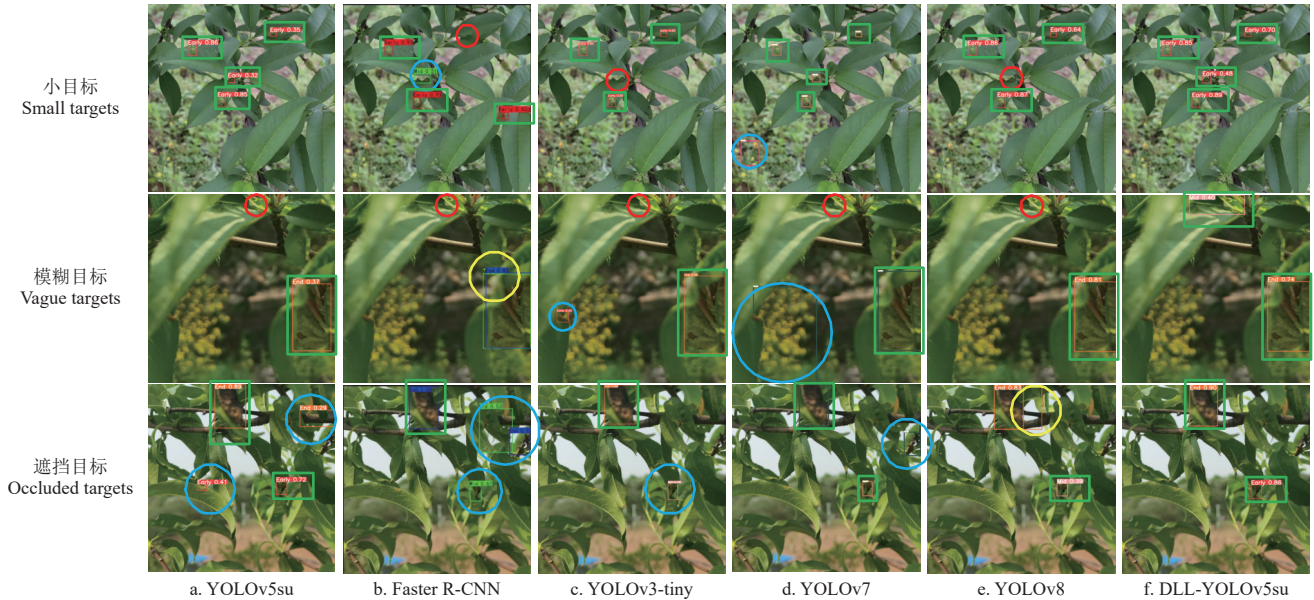
由表3可以看出, YOLOv5su原始模型对比YOLOv7和YOLOv8, 在桃树缩叶病识别平均精度均值接近的情况下, 权重最小, 证明了本文选择其作为基线模型的可行性。在对原始模型进行针对桃树缩叶病特征识别进行改进之后, 改进的DLL-YOLOv5su模型平均精度均值 mAP_{50} 相较于Faster R-CNN、YOLOv3-tiny、YOLOv7和YOLOv8分别高出28.5、11.8、2.1和4.1个百分点。召回率为73.1%, 仅低于YOLOv7模型1.5个百分点, 而准确率80.7%高于YOLOv7, 因为对与目标检测任务, 召回率与准确率难以同时兼顾。同时改进的DLL-YOLOv5su模型权重为17.6MB仅为YOLOv7的23.5%。

表 3 对比试验结果
Table 3 Contrast test results

模型 Models	准确率 Precision P/%	召回率 Recall R/%	mAP ₅₀ /%	权重大小 Model size/MB	平均精度 Average precision AP/%		
					早期 Early	中期 Mid	后期 End
YOLOv5su	76.5	70.7	76.1	18.5	78.3	75.3	74.6
Faster R-CNN	-	-	51.9	110.9	31.5	62.2	62.1
YOLOv3-tiny	73.9	67.2	68.6	17.5	69.3	66.2	70.4
YOLOv7	80.1	74.6	78.3	74.8	82.4	78.5	74.0
YOLOv8s	80.0	70.6	76.3	22.5	78.0	77.9	73.1
DLL-YOLOv5su	80.7	73.1	80.4	17.6	80.6	81.5	79.1

为进一步验证 DLL-YOLOv5su 对桃树缩叶病检测的效果, 随机选择测试集图像对模型测试, 测试结果如图 8 所示。从图中可以看出, 对于桃树缩叶病早期的小目标检测任务, 改进的 DLL-YOLOv5su 模型检测精度更高。Faster R-CNN 和 YOLOv7 均出现了不同程度的误检;

对于模糊目标, 仅有改进的模型检测到了两个正确的目标, 其余模型均仅检出一个目标或出现误检; 对于遮挡的目标识别, YOLOv5su 模型将桃树枝干误识别为缩叶病目标, 而改进的 DLL-YOLOv5su 模型识别准确率高, 且误检率低。



注: 绿色方框处为正确检出, 红色圆圈处为漏检, 蓝色圆圈处为误检, 黄色圆圈处为重复检出。
Note: The green boxes are correctly checked out, the red circle is missed, the blue circle is false, and the yellow circle is double-checked.

图 8 不同环境下各模型对桃树缩叶病识别效果

Fig.8 Recognition effect of each model on peach leaf curl disease in different environments

试验结果表明, 本文提出的 DLL-YOLOv5su 模型相较于主流目标检测模型, 对不同发生时期的桃树缩叶病具有更好的识别效果; 相较于文献 [12] 中的桃树病害检测模型, 对桃树缩叶病的检测精度高出了 3.82 个百分点, 且能够反映桃树缩叶病的发生情况。DLL-YOLOv5su 模型能够为桃树缩叶病的针对性措施提供技术支持。

4 结论

本文通过对自然环境下采集的桃树缩叶病图像进行增强处理, 搭建了桃树缩叶病数据集, 提出了一种基于改进的 YOLOv5su 网络的目标检测模型 DLL-YOLOv5su 以提高桃树缩叶病在自然环境中的检测准确率。研究结果如下:

1) 本文以 YOLOv5su 为基础网络, 在其骨干网络的中引入可变形自注意力机制和可分离大内核注意模块,

并在感受野卷积的基础上提出了权重自适应下采样模块替换网络中部分卷积模块, 实现了在提高识别准确率和平均精度的前提下, 轻量化模型。改进的 DLL-YOLOv5su 模型的准确率、召回率和平均精度均值 mAP₅₀ 分别达到了 80.7%、73.1% 和 80.4%, 相比初始模型分别提高了 4.2、2.4 和 4.3 个百分点。改进模型在桃树缩叶病 3 个发生时期的检测效果均高于初始模型, 说明了改进方法的有效性。

2) 通过对比试验与 Faster R-CNN、YOLOv3-tiny、YOLOv7、YOLOv8s 等主流目标检测模型进行比较, 改进的 DLL-YOLOv5su 模型的准确率最高, 平均精度均值 mAP₅₀ 分别高出 28.5、11.8、2.1、4.1 个百分点。与桃树病害检测模型相比, DLL-YOLOv5su 精度也更高, 进一步展示了 DLL-YOLOv5su 桃树缩叶病检测模型的优势。DLL-YOLOv5su 模型能够识别自然环境下桃树缩叶病发生的 3 个阶段, 对早期目标的识别效果提升可以更

好地达到预警的效果, 对于模糊和遮挡目标也具备一定推理能力。

[参 考 文 献]

- [1] 徐磊, 陈超. 中国桃产业经济分析与发展趋势[J]. 果树学报, 2023, 40(1): 133-143.
XU Lei, CHEN Chao. Economic situation and development countermeasures of Chinese peach industry[J]. Journal of Fruit Science, 2023, 40(1): 133-143. (in Chinese with English abstract)
- [2] 贾少鹏, 高红菊, 杭潇. 基于深度学习的农作物病虫害图像识别技术研究进展[J]. 农业机械学报, 2019, 50(S1): 313-317.
JIA Shaopeng, GAO Hongju, HANG Xiao. Research progress on image recognition technology of crop pests and diseases based on deep learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(S1): 313-317. (in Chinese with English abstract)
- [3] 刘诗怡, 胡滨, 赵春. 基于改进 YOLOv7 的黄瓜叶片病虫害检测与识别[J]. 农业工程学报, 2023, 39(15): 163-171.
LIU Shiyi, HU Bin, ZHAO Chun. Detection and identification of cucumber leaf diseases based improved YOLOv7[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(15): 163-171. (in Chinese with English abstract)
- [4] LIU J, WANG X. Plant diseases and pests detection based on deep learning: A review[J]. Plant Methods, 2021, 17: 22.
- [5] 王磊磊, 王斌, 李东晓, 等. 基于改进 YOLOv5 的菇房平菇目标检测与分类研究[J]. 农业工程学报, 2023, 39(17): 163-171.
WANG Leilei, WANG Bin, LI Dongxiao, et al. Object detection and classification of pleurotus ostreatus using improved YOLOv5[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2023, 39(17): 163-171. (in Chinese with English abstract)
- [6] WANG F, WANG R, XIE C, et al. Fusing multi-scale context-aware information representation for automatic in-field pest detection and recognition[J]. Computers and Electronics in Agriculture, 2020, 169: 105222.
- [7] LI J, QIAO Y, LIU S, et al. An improved YOLOv5-based vegetable disease detection method[J]. Computers and Electronics in Agriculture, 2022, 202: 107345.
- [8] 兰玉彬, 孙斌书, 张乐春, 等. 基于改进 YOLOv5s 的自然场景下生姜叶片病虫害识别[J]. 农业工程学报, 2024, 40(1): 218-224.
LAN Yubin, SUN Binshu, ZHANG Lechun, et al. Identifying diseases and pests in ginger leaf under natural scenes using improved YOLOv5s[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2024, 40(1): 218-224. (in Chinese with English abstract)
- [9] SU Y, LI H, WANG X, et al. Improvement of the YOLOv5 model in the optimization of the brown spot disease recognition algorithm of kidney bean[J]. Plants, 2023, 12(21): 3765.
- [10] XIAO J, KANG G, WANG L, et al. Real-time lightweight detection of lychee diseases with enhanced YOLOv7 and edge computing[J]. Agronomy, 2023, 13(12): 2866.
- [11] HOU J, YANG Chen, HE Y, et al. Detecting diseases in apple tree leaves using FPN-ISResNet-Faster RCNN[J]. European Journal of Remote Sensing, 2023, 56(1): 2186955.
- [12] 李硕. 桃树病害和害虫图像检测系统的研究与实现[D]. 泰安, 山东农业大学, 2021.
LI Qi. Research and Implementation of Image Detection System for Peach Diseases and Pests[D]. Tai'an, Shandong Agricultural University, 2021. (in Chinese with English abstract)
- [13] YANG S, WANG W, GAO S, et al. Strawberry ripeness detection based on YOLOv8 algorithm fused with LW-Swin Transformer[J]. Computers and Electronics in Agriculture, 2023, 215: 108360.
- [14] ZHANG Y, ZHOU G, CHEN A, et al. A precise apple leaf diseases detection using BCTNet under unconstrained environments[J]. Computers and Electronics in Agriculture, 2023, 212: 108132.
- [15] JOCHER G. Ultralytics YOLOv8 Docs/YOLOv5[EB/OL]. (2023-11-12)[2024-01-11] <https://docs.ultralytics.com/models/yolov5>
- [16] ALEXEY D, LUCAS B, ALEXANDER K, et al. An image is worth16x16words:Transformersforimagerecognitionatscale[C]// International Conference on Learning Representations (ICLR). Vienna, Austria, 2021.
- [17] XIA Z, PAN X, SONG S, et al. Vision transformer with deformable attention[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, Louisiana: IEEE, 2022: 4784-4793.
- [18] GUO M, LU C, LIU Z, et al. Visual attention network[J]. Computation Visual Media, 2023, 9(4): 733-752.
- [19] LAU K W, PO L M, REHMAN Y A U. Large separable kernel attention: Rethinking the large kernel attention design in CNN[J]. Expert Systems with Applications, 2023, 236: 121352.
- [20] ZHANG X, LIU C, YANG S, et al. RFACConv: Innovating spatial attention and standard convolutional operation[EB/OL]. (2023-04-06)[2024-03-28]. <https://arxiv.org/abs/2304.03198v6>
- [21] VASWANI A, NOAM M S, NIKI P, et al. Attention is All you Need [EB/OL]. (2017-06-12) [2023-08-02], <https://arxiv.org/abs/1706.03762>.
- [22] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, Canada: IEEE, 2021: 9992-10002.
- [23] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City: IEEE, 2018: 6848-6856.
- [24] HUNAG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI: IEEE, 2017: 4700-4708.
- [25] DAI J, QI H, XIONG Y, et al. Deformable convolutional networks[C]// Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy: IEEE, 2017:

- 764-773.
- [26] BELLO I, ZOPH B, VASWANI A, et al. Attention augmented convolutional networks[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, South Korea: IEEE, 2019: 3286-3295.
- [27] VASWANI A, RAMACHANDRAN P, SRINIVAS A, et al. Scaling local self-attention for parameter efficient visual backbones[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN: IEEE, 2021: 12894-12904.
- [28] SRINIVAS A, LIN T, PARMAR N, et al. Bottleneck transformers for visual recognition[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN: IEEE, 2021: 16519-16529.

Improved YOLOv5su model for detecting peach leaf curl disease

YAO Lingyun , ZHOU Junfeng , LI Li

(1. College of Engineering and Technology, Southwest University, Chongqing 400715, China;

2. Chongqing Key Laboratory of Agricultural Equipment for Hilly and Mountainous Regions, Chongqing 400715, China)

Abstract: Peach trees are susceptible to leaf curl disease in humid regions in spring, leading to a reduction in peach production. Current monitoring cannot timely and accurately identify the leaf curl disease in large-scale orchards. Therefore, intelligent detection can be expected for the peach tree diseases using deep learning, in order to minimize the pesticide application. However, most research has relied on individual-picked leaves in disease identification. It is still challenging to apply to disease detection in real-world scenarios. The primary research can also emphasize to differentiation among various disease types. Yet it is lacking in the accurate occurrence of leaf disease. It is also difficult to identify the peach leaf curl, due to the complex background of disease images and the small size of the early target. In this study, a DLL-YOLOv5su model was introduced to detect peach leaf curl disease in natural environments. Firstly, a deformable attention module was integrated into the backbone network of YOLOv5su. The C3-DA module was formed to recognize the targets with significant size variations, in order to improve the feature fitting of the model. Additionally, a large separable kernel attention (LSKA) module was incorporated into the spatial pyramid pooling fusion (SPPF) layer to expand the receptive field of the network. The sensitivity to large targets was enhanced without complexity. Lastly, a lightweight adaptive weighted downsampling (LAWD) module was proposed on the receptive field attention convolution (RFACConv), in order to replace the convolutional modules within the network for feature extraction. Both lightweight and accuracy were achieved concurrently. The dataset was also built for the peach leaf curl disease, including the early, middle, and late stages. 1520 images were then collected from the peach orchard of Chongqing Academy of Agricultural Sciences, China. Among them, the number of pictures in the training set increased from 1105 to 6630 after data augmentation. Subsequently, ablation experiments were conducted to validate the improvement under the same device environment. Each improvement was observed. The recall rate increased by 3.5 percentage points under the C3-DA module, compared with the initial model, while the mean average precision (mAP_{50}) increased by 2.3 percentage points; The accuracy and average precision (mAP_{50}) increased by 2.5 and 2.3 percentage points, respectively, after the LSKA module was introduced; The model size decreased by 18.9 percentage points, after the LAWD was used to replace some CBS modules. The ablation experiments revealed that the improved DLL-YOLOv5su model was achieved with a detection accuracy of 80.7%, a recall rate of 73.1%, and an average precision mean average precision (mAP_{50}) of 80.4% on the test set of peach leaf curl disease. Compared with the initial model, these metrics increased by 4.2, 2.4, and 4.3 percentage points, respectively. The frame rate (FPS) of 83 frames per second fully met the requirements of real-time detection for peach leaf curl. A comparison experiment was designed to compare the detection performance of DLL-YOLOv5su with the current mainstream model of target detection. The performance of the DLL-YOLOv5su model outperformed the most mainstream models, including YOLOv3-tiny, YOLOv7, Faster R-CNN, and YOLOv8, with higher accuracy and precision. The recall rate was only 1.5 percentage points lower than YOLOv7, but the model size was only 23.5% of YOLOv7's. In summary, the improved DLL-YOLOv5su model was achieved in the real-time and accurate detection of peach leaf curl occurrences, thereby enhancing the efficiency of smart peach orchards. Resource allocation was also optimized to facilitate the precise application of pesticides, particularly for the high crop yields and healthy peach trees.

Keywords: image processing; diseases; leaf curl disease; object detection; YOLOv5su; deformable attention; large kernel convolution; light weight