

基于改进YOLOF模型的田间农作物害虫检测方法

彭红星^{1,2} 徐慧明¹ 高宗梅³ 田兴国⁴ 邓倩婷⁵ 咸春龙⁶

(1. 华南农业大学数学与信息学院, 广州 510642; 2. 农业农村部华南热带智慧农业技术重点实验室, 广州 510642;
3. 华盛顿州立大学生物系统工程系, 普尔曼 WA99350; 4. 华南农业大学食品学院, 广州 510642;
5. 华南农业大学资产与实验室管理处, 广州 510642; 6. 华南农业大学经济管理学院, 广州 510642)

摘要: 田间害虫图像数据采集困难, 并且传统的检测模型大多使用复杂的特征金字塔 (Feature pyramid network, FPN) 结构提升精度, 这在一定程度上影响了检测的实时性。为此, 本研究通过设计诱虫灯装置构建害虫数据集 FieldPest5, 并且对无 FPN 结构的检测器 YOLOF 进行改进, 提出兼顾检测精度和效率的害虫检测模型 YOLOF_PD。首先, 增加 Cutout 数据增强方法缓解害虫图像中的遮挡问题, 并且使用 CIoU 损失函数获得更好的框回归位置; 其次, 在原有坐标注意力机制 (Coordinate attention, CA) 的全局平均池化 (Global average pooling, GAP) 路径中增加全局最大池化 (Global max pooling, GMP) 路径, 并且使用可学习参数自适应更新不同路径的权重, 提出自适应坐标注意力机制 (Adaptive coordinate attention, ACA), 增强模型的信息表征能力; 最后, 对 YOLOF 膨胀编码器中的 Projector 和 Residual 模块进行改进, 在 Projector 模块的 3×3 卷积后引入 ACA 注意力机制, 在 Residual 模块中融合 3×3 的深度可分离卷积和 1×1 的逐点卷积, 提出 Dilated_Dwise_ACA 编码器, 提高 YOLOF 对小尺度害虫的检测性能。实验结果表明: 改进后的 YOLOF_PD 模型在 FieldPest5 测试集上的平均精度均值 (Mean average precision, mAP) 为 93.7%, 较改进前提升 2.1 个百分点, 并且检测时图像传输速率为 42.4 f/s, 能够满足害虫快速检测的要求。对比 Cascade R-CNN、RetinaNet、ATSS 等模型, YOLOF_PD 模型在检测效果和检测速度方面均取得了良好性能。

关键词: 田间害虫检测; 诱虫灯; YOLOF; 自适应坐标注意力; 多尺度特征

中图分类号: TP391.4 文献标识码: A 文章编号: 1000-1298(2023)04-0285-10

OSID:



Insect Pest Detection of Field Crops Based on Improved YOLOF Model

PENG Hongxing^{1,2} XU Huiming¹ GAO Zongmei³ TIAN Xingguo⁴ DENG Qianting⁵ XIAN Chunlong⁶

(1. College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China

2. Key Laboratory of Smart Agricultural Technology in Tropical South China,
Ministry of Agriculture and Rural Affairs, Guangzhou 510642, China

3. Department of Biological Systems Engineering, Washington State University, Pullman WA99350, USA

4. College of Food Science, South China Agricultural University, Guangzhou 510642, China

5. Assets and Laboratory Management Office, South China Agricultural University, Guangzhou 510642, China

6. College of Economics and Management, South China Agricultural University, Guangzhou 510642, China)

Abstract: The wide distribution of pests in the field leads to difficulties in image data acquisition, and most of the traditional detection models use complex feature pyramid network (FPN) to enhance detection accuracy, which affects the real-time detection to some extent. To this end, the trap lamp device was designed to construct the pest dataset FieldPest5 and the detector YOLOF, which does not use the FPN structure, was improved to propose a pest detection model YOLOF_PD that balanced detection accuracy and efficiency. Firstly, the Cutout data augmentation method was added to alleviate the occlusion problem in the pest images, and the complete intersection over union (CIoU) loss function was used to obtain better box regression positions. Secondly, the adaptive coordinate attention (ACA) mechanism was proposed to enhance the information representation capability of the model. Specifically, the global maximum pooling (GMP) path was added to the global average pooling (GAP) path of the original

收稿日期: 2022-06-14 修回日期: 2022-07-06

基金项目: 国家自然科学基金项目 (61863011、32071912)、广东省乡村振兴战略专项项目 (2020KJ261)、广州市科技计划项目 (202002020016) 和广州市基础研究计划项目 (202102080337)

作者简介: 彭红星 (1976—), 男, 副教授, 博士, 主要从事机器视觉、智慧农业和农业机器人研究, E-mail: xyphx@scau.edu.cn

通信作者: 咸春龙 (1965—), 男, 教授, 博士生导师, 主要从事智慧农业、土地资源管理和教育经济与管理研究, E-mail: chlbian@scau.edu.cn

coordinate attention (CA) mechanism, and the weights of different paths were updated adaptively by using learnable parameters. Finally, the Dilated_Dwise_ACA encoder was proposed to improve the performance of YOLOF for small-scale pest detection. Improvements were made to the projector and residual modules in the dilated encoder. The ACA attention mechanism was introduced after the 3×3 convolution in the projector module, and in the Residual module 3×3 depth-separable convolution and 1×1 pointwise convolution were fused. The experimental results showed that the improved YOLOF_PD model mAP achieved 93.7% on the FieldPest5 test set, which was 2.1 percentage points higher than that of the model before improvement, and the detection speed was 42.4 f/s, which can meet the requirements of fast pest detection. Compared with Cascade R-CNN, RetinaNet and ATSS, YOLOF_PD achieved good performance in terms of detection effect and detection speed. The research result can lay a solid foundation for field pest data collection as well as real-time pest detection.

Key words: pest detection of field; insect trap lamp; YOLOF; adaptive coordinate attention; multiscale features

0 引言

害虫对农作物的质量和产量构成了严重威胁,造成农业产业的巨额损失^[1-2]。及时准确的害虫检测是有效制定害虫管理计划的前提条件。然而,害虫的种类繁多,并且部分害虫间的相似度较高,这使得害虫检测任务更加具有挑战性。

传统的害虫检测方法主要是依靠人工观察或者是专家上门指导,存在效率低、主观性强、时效性差等问题,很难得到广泛应用^[3]。为了减轻这种负担,设计一种自动检测害虫的方法至关重要。王林惠等^[4]使用 MoblieNet 作为害虫图像特征提取网络,并且使用 Faster R-CNN 对柑橘红蜘蛛和蚜虫进行检测,准确率分别达到 91.0% 和 89.0%,单帧图像平均处理时间为 286 ms。TENG 等^[5]提出了新型的害虫检测网络 MSR-RCNN,使用多尺度的超分辨率特征增强模块提高小尺寸、多尺度和高相似度害虫的检测性能,在自建的害虫数据集 LLPD-26 上实现了 67.4% 的精确率。LI 等^[6]使用粘性诱捕器捕获温室中的粉虱和蓟马,提出基于 R-CNN 的端到端模型 TPEst-RCNN,在验证集上的平均 F1 值和平均精度分别达到 94.4% 和 95.2%。WANG 等^[7]提出了 S-RPN 的网络,将注意力机制引入残差网络,并设计了一种新颖的自适应感兴趣区域选择方法,能够从特征金字塔 (Feature pyramid network, FPN) 的不同层次学习特征,在自建的 AgriPest21 数据集上平均召回率达到 89.0%,平均精度均值 (Mean average precision, mAP) 达到 78.7%。JIAO 等^[8]在 FPN 中引入自适应特征融合模块提取更丰富的害虫特征,开发自适应增强模块减少最高级别特征图的信息损失,建立了一个两阶段的基于区域的卷积神经网络 (Convolutional neural networks, CNN),在 AgriPest21 数据集上准确率达到 77.0%。

尽管上述研究具有良好的实验结果,但是基于深度学习的害虫检测需要大量的害虫样本进行监督学习^[9]。而田间害虫分布广泛,导致害虫图像数据收集困难。为此本研究拟设计一套诱虫灯害虫采集装置,收集包含水稻、大豆和青菜等田间农作物害虫的数据集 FieldPest5。此外,农业害虫检测大多建立在以 Faster R-CNN^[10] 为代表的两阶段检测器上,很难达到实时性的要求。在两阶段^[11] 和一阶段的优秀检测器中^[12],FPN^[13] 成为提升检测精度的重要组件,这不可避免地制约了检测速度的提升。以 RetinaNet^[14] 为基准,CHEN 等^[15] 提出了仅使用单级别特征图的一阶段目标检测器 YOLOF,在 COCO 数据集上得到了和使用 FPN 结构检测器相当的准确率,但是速度快了 2.5 倍。然而,YOLOF 中去除了 FPN 结构以提升检测速度,这对于小尺寸的害虫检测并不友好。为此,本研究在 YOLOF 检测器的基础上进行改进。针对害虫图像中存在的遮挡问题,使用 Cutout^[16] 数据增强方法进行缓解,并使用 CIoU (Complete intersection over union) 回归损失^[17] 代替 GIoU (Generalized intersection over union) 回归损失获得更准确的框回归位置。针对模型在检测小尺寸害虫时存在漏检等问题,拟对 YOLOF 中的核心结构膨胀编码器进行改进。首先,受视觉注意力系统的启发,将改进的坐标注意力机制^[18] 引入到膨胀编码器的 Projector 模块中,以获得更丰富的小物体细节特征,其次,使用深度卷积 (Depthwise convolution, DWConv) 和 1×1 的普通卷积 (Convolution, Conv) 重新设计膨胀编码器的 Residual 模块,更好地提取不同尺度的害虫特征。

1 实验数据集建立

1.1 诱虫灯图像采集设备

田间害虫一般分布较广,并且需要较多的人力

进行图像采集。为了更好地采集水稻、大豆、青菜等田间农作物的害虫图像,本研究以诱虫灯为基础,设计了一套自动化害虫采集设备,结构示意图如图 1 所示。

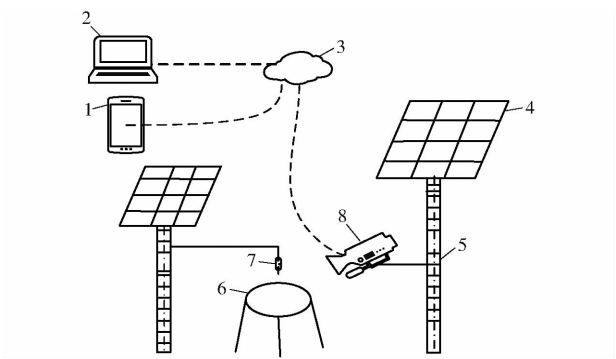


图 1 诱虫灯图像采集设备结构示意图

Fig. 1 Insect trap lamp image acquisition equipment structure

1. 手机 2. 计算机 3. 云服务器 4. 太阳能板 5. 立杆 6. 盛水容器 7. 诱虫灯 8. 太阳能摄像头

诱虫灯图像采集系统的整体工作流程如下:首先,诱虫灯通过特定的光波吸引田间害虫并将其诱杀,然后害虫落入盛有水的容器中,太阳能摄像头实时进行监控,并将数据传入到云服务器中进行保存,用户可以通过计算机或者手机等设备实时查看和筛选害虫图像,适合野外无电的恶劣环境,能够大幅度减少采集害虫图像的时间。

1.2 图像采集

该实验从 2021 年开始持续至今,在广东省广州、从化、惠州、梅州、韶关和阳山地区部署了 63 套诱虫灯图像采集设备。选择危害水稻、大豆以及青菜等田间农作物的草地贪夜蛾、斜纹夜蛾、小菜蛾、棉夜蛾和螟蛾作为目标害虫进行检测。每周更换 1 次容器中的水,经过筛选最终获得了包含 3 409 幅害虫图像的数据集 FieldPest5。数据集中每幅害虫图像的尺寸为 1 920 像素×1 080 像素,并且图像以 JPG 格式进行存储。捕获的害虫图像样本如图 2 所示。

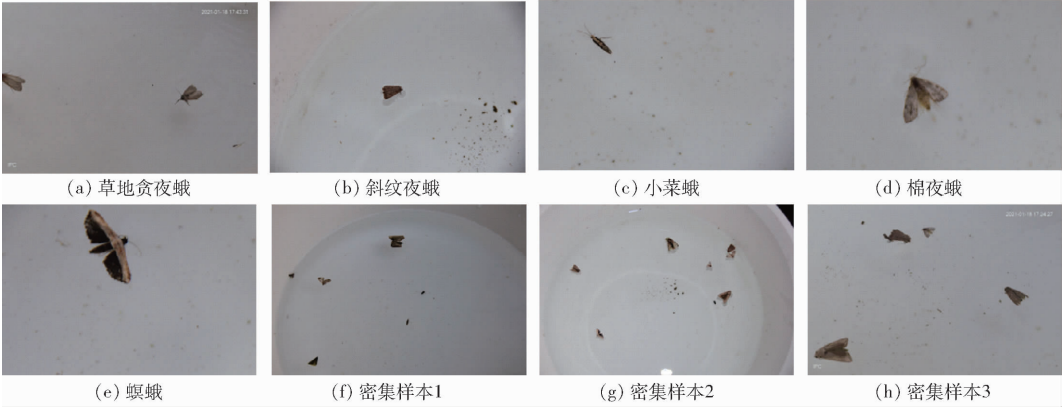


图 2 害虫图像样本

Fig. 2 Samples of pest images

从图 2 中可以看出,由于容器中的水会变浑浊,而且不同类别的害虫尺寸相差较大,其中螟蛾和棉夜蛾的尺寸较大,而小菜蛾的尺寸较小,这在一定程度上增加了害虫检测的难度。每类害虫危害的农作物类型和具体害虫目标数量如表 1 所示。

表 1 每类害虫的数量

Tab. 1 Number of pests in each category

害虫类别	危害农作物类型	目标数量
草地贪夜蛾	水稻等	2 725
斜纹夜蛾	水稻、大豆等	1 709
小菜蛾	青菜等	202
棉夜蛾	青菜等	380
螟蛾	水稻等	654
总计		5 670

从表 1 中可以看出,收集到的害虫图像中涉及的农作物类别有水稻、大豆和青菜等。草地贪夜蛾、斜纹夜蛾、螟蛾对水稻和大豆的危害较大,小菜蛾和

棉夜蛾对青菜的危害较大。诱虫灯设备对于草地贪夜蛾和斜纹夜蛾的吸引效果较好,捕获的图像数量较多,而小菜蛾的数量相对较少,最终获得了 5 670 个害虫目标。从总体上看 FieldPest5 数据集中存在一定的数据不平衡问题。

1.3 数据集划分和数据预处理

本文使用 LableImg 工具对害虫图像进行手动标注,保存为 PASCAL VOC^[19] 格式。将 FieldPest5 数据集按照 7: 1: 2 的比例划分为训练集、验证集和测试集。其中,训练集包含 2 386 幅图像,验证集包含 341 幅图像,测试集包含 682 幅图像。数据预处理主要包括图像标准化处理以及数据增强。图像标准化处理方式

$$O = \frac{I - M_{ean}}{S_{td}}$$

(1)

式中 O——归一化后的输出图像
I——输入图像

M_{ean} ——均值,取(103.53,116.28,123.675)
 S_{td} ——标准差,取(1.0,1.0,1.0)

考虑到 FieldPest5 数据集中包含的数据量较少,使用数据增强技术可以提高模型的泛化能力,减少过拟合。在目标检测任务中,目标需要事先进行框选,所以使用在线数据增强的方式更为适合。在原 YOLOF 检测器中使用了随机翻转、随机偏移和归一化等操作,从图 2 中可以看出,数据集中的部分数据存在遮挡情况。Cutout^[16]数据增强先随机选择一个固定大小的正方形区域(本文使用的正方形区域为 50 像素×50 像素),然后用 0 进行填充,从而将填充区域的图像信息遮挡,能够让卷积神经网络更好地利用图像的全局信息,而不是依赖于部分特定的视觉特征,这有利于提高模型的鲁棒性,缓解遮挡问题。使用 Cutout 数据增强的样本如图 3 所示。

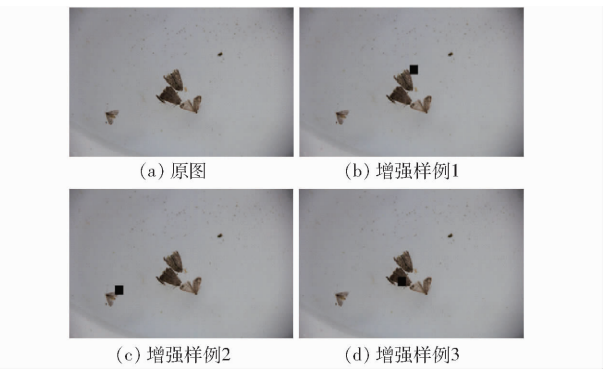


图 3 Cutout 数据增强样例
Fig. 3 Sample images of Cutout data augmentation

2 检测方法

2.1 YOLOF 检测器

FPN 为二阶段和一阶段的目标检测方法做出了巨大贡献,但是会降低检测器速度并且使得检测器的结构更加复杂。为此,YOLOF 通过引入膨胀编码器和均衡匹配策略作为替代方案,无需使用复杂的特征金字塔结构,只需要使用单层特征图,并且在 COCO 数据集上取得了良好的实验结果。单层特征图只能覆盖一个受限的感受野,为此 YOLOF 中使用膨胀编码器的结构增大感受野,如图 4 所示,其中 Dilated Conv 表示膨胀卷积; k 表示卷积核大小; $\times 4$ 表示堆叠次数为 4 次。

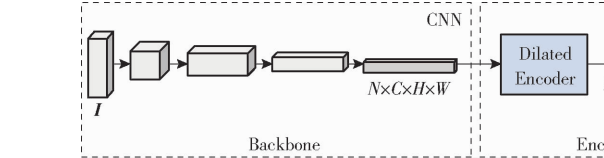


图 5 YOLOF 检测器结构图
Fig. 5 Structure diagram of YOLOF detector

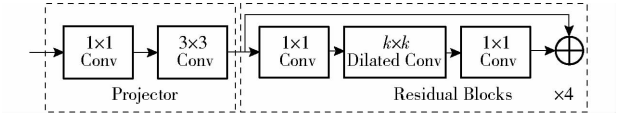


图 4 膨胀编码器结构
Fig. 4 Dilated encoder structure

膨胀编码器包含 Projector 和 Residual Blocks 模块。膨胀编码器首先使用 1×1 卷积层将特征图进行压缩,再使用 3×3 卷积层细化上下文语义信息。然后,连续堆叠 4 个带有不同膨胀率(2、4、6、8)卷积的残差块,生成具有不同感受野的输出特征图。

在 RetinaNet 中,如果一个先验框和真实框之间的最大交并比(Intersection over union, IoU)大于一个给定的阈值,则把这个先验框定义为正样本,这个策略称为 Max-IoU 匹配。在单一级别的特征图上产生正样本,框的数量相比于使用 FPN 结构的检测器会减少很多,使用 Max-IoU 策略会导致正 anchor 的不平衡,不利于小目标的检测^[15]。为此,YOLOF 提出了均衡匹配策略,即对每一个真实框,采用最接近的 A 个候选框作为正样本,不论真实框的尺寸,正样本数目均相同,这可以确保所有真实框平等地参与训练。对于大 IoU 负样本和小 IoU 正样本通过设置阈值进行过滤。

通过上述两种解决方案,只使用单级别特征图的简单目标检测模型 YOLOF 也取得了良好的检测效果,其整体结构如图 5 所示。其中, N 表示每批输入的图像数量, C 表示通道总数, H 表示特征图的高, W 表示特征图的宽, T 表示害虫类别数, A 表示候选框数量。

本文使用在 ImageNet 数据集上预训练的 ResNet50^[20]作为特征提取骨干网络,提取图像特征;然后使用 Dialated Encoder 细化上下文的语义信息,获取不同感受野的特征;在 Decoder 阶段,包含两个并行分支,用于目标分类和边框回归任务,在回归分支中增加了一个隐式的对象预测,最终分类结果是通过将分类分支输出量与相应的隐式对象相乘得到。

2.2 自适应坐标注意力机制

注意力机制在很多计算机视觉任务中都表现出了良好的性能,其中 SE(Squeeze and excitation)通道注意力机制^[21]以其简单、高效的特点被广泛使用。

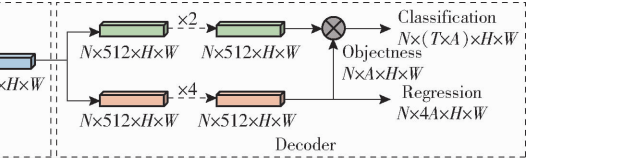


图 5 YOLOF 检测器结构图
Fig. 5 Structure diagram of YOLOF detector

然而,SE通道注意力机制忽略了位置信息的重要性。HOU等^[18]在2021年提出了坐标注意力机制,通过嵌入位置信息到通道注意力中以获取更丰富的特征信息。ZHOU等^[22]使用类响应图论证了全局平均池化(Global average pooling, GAP)和全局最大池化(Global max pooling, GMP)的定位能力。坐标注意力机制将输入特征图在高度和宽度两个方向上进行GAP操作,从而获得高度和宽度两个方向的特征图。然而,GAP的简单性使其难以很好地捕捉各种输入的复杂信息^[23]。CBAM^[24]注意力机制中通过简单相加的方式将GAP和GMP组合起来提高性能。在此基础上,本文进行了更深入的研究,在原有CA注意力机制的GAP路径中增加GMP路径,并且使用可学习参数自适应更新不同路径的权重,提出自适应坐标注意力机制(Adaptive coordinate attention, ACA),提升模型在FieldPest5数据集上的检测效果。

自适应坐标注意力机制将输入特征图 $X \in \mathbf{R}^{C \times H \times W}$ 在高度和宽度两个方向上进行GAP和GMP操作,并且使用可学习参数 p_1 和 p_2 (初始化为1)确定GAP和GMP的路径权重,获得高度和宽度两个方向的特征图,具体公式为

$$z_c^h(h) = p_1 \frac{\sum_{0 \leq i < W} x_c(h, i)}{W} + p_2 \varphi_{0 \leq i < W}(x_c(h, i)) \quad (2)$$

$$z_c^w(w) = p_1 \frac{\sum_{0 \leq j < H} x_c(j, w)}{H} + p_2 \varphi_{0 \leq j < H}(x_c(j, w)) \quad (3)$$

式中 z_c^h ——高度为 h 的第 c 通道输出
 z_c^w ——宽度为 w 的第 c 通道输出
 x_c ——第 c 通道输入
 i ——宽度方向上的迭代变量
 j ——高度方向上的迭代变量
 φ ——求最大值函数

将式(2)、(3)中的输出特征图进行拼接并进行卷积操作,此时特征图维度变为 C/e , e 为超参数,再经过归一化和非线性激活操作后得到中间特征图,为

$$m = \delta(F_1([z^h, z^w])) \quad (4)$$

式中 m ——空间信息在水平方向和垂直方向进行编码得到的中间特征映射

δ ——非线性激活函数

F_1 —— 1×1 卷积变换

将 m 沿着高度和宽度维度分成两个独立的张量,并且使用卷积变换将其转换为与输入 X 具有相同通道数的张量,再经过激活函数 σ 后得到注意力权重为

$$\begin{cases} u^h = \sigma(F_h(m^h)) \\ u^w = \sigma(F_w(m^w)) \end{cases} \quad (5)$$

式中 u^h ——高度上的注意力权重
 u^w ——宽度上的注意力权重
 F_h ——高度上的卷积变换
 F_w ——宽度上的卷积变换
 m^h ——沿着高度方向分解的张量
 m^w ——沿着宽度方向分解的张量
 σ ——激活函数 Sigmoid

将 u^h 和 u^w 与输入做乘法加权操作,得到最终的输出为

$$y_c(i, j) = x_c(i, j) u_c^h(i) u_c^w(j) \quad (6)$$

式中 y_c ——第 c 通道的输出特征图
 u_c^h —— c 通道在高度上的注意力权重
 u_c^w —— c 通道在宽度上的注意力权重

自适应坐标注意力机制的结构如图6所示,将自适应坐标注意力机制引入到Projector模块 3×3 卷积层后,命名为Projector_ACA模块。

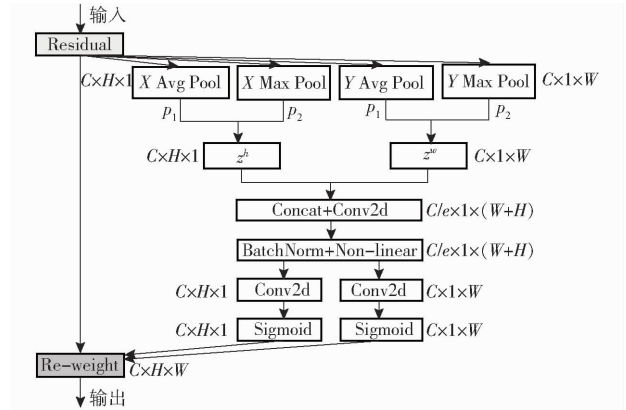


图6 自适应坐标注意力机制结构

Fig.6 ACA structure

2.3 Dilated_Dwise_ACA 编码器

膨胀卷积在普通卷积的基础上注入空洞,以此来增加感受野。使用超参数膨胀率可以指定卷积核中的间隔数量。YOLOF中使用带有不同膨胀率的膨胀卷积构建残差块,实现不同感受野的输出。膨胀卷积等效为普通卷积核大小的计算公式为

$$k_e = k + (k - 1)(r - 1) \quad (7)$$

式中 k_e ——膨胀卷积的等效卷积核大小
 r ——膨胀率

YOLOF在 3×3 普通卷积的基础上注入空洞,膨胀率设置为2、4、6、8,等效为普通卷积核大小为 5×5 、 9×9 、 13×13 和 17×17 。膨胀卷积通过调整膨胀率可以获得更大尺度的信息,这在大物体检测上或许有良好的效果,但是对于一些小尺寸物体的检测,效果一般^[25]。GoogLeNet^[26]中提出了

Inception 模块获取不同尺度上的特征信息,从而提升网络性能。在借鉴 Inception 模块思想以及不过度增加 YOLOF 复杂度的基础上,将 Residual 模块中的膨胀卷积与更小感受野的 3×3 深度可分离卷积和 1×1 普通卷积进行融合,提升 YOLOF 在小尺度害虫上的检测效果,融合后的模块称为 Residual_Concat,如图 7 所示。综合 2.2 节中介绍的 Projector_ACA 模块和本节提出的 Residual_Concat 模块,提出 Dilated_Dwise_ACA 编码器,如图 7 所示。

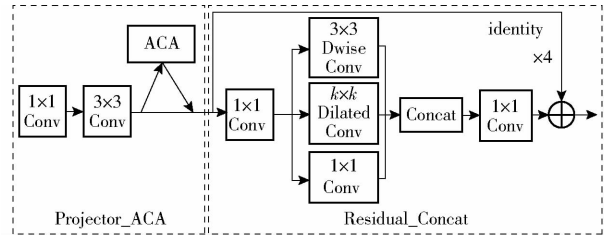


图 7 Dilated_Dwise_ACA 编码器

Fig. 7 Dilated_Dwise_ACA encoder

2.4 损失函数

IoU 表示预测框和真实框的交集和并集的比值,在目标检测中能够较好地反映预测框和真实框之间的检测效果。由于 IoU 无法精确地反映预测框和真实框的重合度,并且无法衡量无重叠框之间的距离。为此,YOLOF 中采用 GIoU 作为位置损失函数,GIoU 在 IoU 的基础上加上一个惩罚项,缓解 IoU 损失在检测框不重叠时出现的梯度无法计算的问题^[27]。当两个框之间的距离越大时,惩罚项也越大。

ZHENG 等^[17]为了解决 GIoU 收敛速度较慢等问题,提出 DIoU (Distance intersection over union) 损失函数,通过在 IoU 的基础上添加惩罚项,最小化两个检测框中心点之间的距离。DIoU 损失值可以直接最小化两个目标框的距离,并且比 GIoU 损失值收敛更快。为了使预测框的宽和高更快地与真实框的宽和高接近,将重叠面积、中心点距离和长宽比考虑进来,提出了 CIoU 损失函数。计算公式为

$$C_{IoU}(Q,B)=I_{ou}(Q,B)-\rho^2(Q_{ctr},B_{ctr})/v^2-gd$$

(8)

$$L_{CIoU}=1-C_{IoU}$$

(9)

其中

$$d=\frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}}-\arctan\frac{w^{pt}}{h^{pt}}\right)^2$$

(10)

$$g=\frac{v}{1-I_{ou}+d}$$

(11)

式中 w^{gt} ——真实框的宽 Q ——预测框
 h^{gt} ——真实框的高 B ——真实框
 w^{pt} ——预测框的宽
 h^{pt} ——预测框的高

Q_{ctr} ——预测框的中心点坐标
 B_{ctr} ——真实框的中心点坐标
 ρ ——欧氏距离函数
 g ——权重参数
 d ——长宽比相似性度量指标
 I_{ou} ——交并比
 C_{IoU} ——Complete IoU
 L_{CIoU} ——CIoU 损失值
 v —— Q 和 B 最小包围框对角线长度

为了更好地获得预测框的回归位置,本文使用 CIoU 损失函数作为改进模型的位置损失函数,类别损失函数依然使用 Focal loss。

2.5 YOLOF_PD 检测模型

在原 YOLOF 检测模型中增加 CutOut 数据增强方式以缓解图像遮挡问题,其次使用 2.2 节中介绍的 ACA 注意力机制和 2.3 节中的 Residual_Concat 多尺度特征融合结构构建 Dilated_Dwise_ACA Encoder,提升模型对小尺寸害虫的检测性能,然后使用 2.4 节中的 CIoU loss 代替 GIoU loss 进一步提升性能,得到最终的害虫检测模型 YOLOF_PD,其整体结构如图 8 所示。

3 实验与讨论

3.1 实验设置

实验环境为:CPU 型号为 Intel(R) Core(TM) i9-10900K, GPU 为 24 GB 显存的 NVIDIA RTX3090, RAM 为 32 GB 的 DDR4, CUDA 版本为 11.4,深度学习框架为 Pytorch 1.7。每幅图像的尺寸设置为 1 333 像素 $\times$ 800 像素。为了防止网络过拟合,使用 1.3 节中介绍的数据增强方式增加训练样本的多样性。实验过程中,使用随机梯度下降(Stochastic gradient descent,SGD)优化器训练网络,动量设置为 0.9。模型批量大小设置为 2,并且训练 36 个迭代周期,初始学习率设置为 0.002 5,在第 24 和第 33 个迭代周期时缩小 10 倍。前 1 500 次迭代时使用 linear warm-up 学习率训练策略^[28]。

本文通过对比一阶段检测模型(RetinaNet^[14]、FCOS^[12]、ATSS^[29])以及两阶段检测模型(Faster R-CNN^[10]、Dynamic R-CNN^[30]、Double Head R-CNN^[31]、Cascade R-CNN^[11])来评估 YOLOF_PD 的性能。所有比较方法的参数都与原始设置一致,不进行调整。

3.2 评价指标

为了公平地评估各模型的检测性能,本文所有实验采用的评价指标包括每个类别的平均精度(Average precision, AP)以及 PASCAL VOC^[19]中常

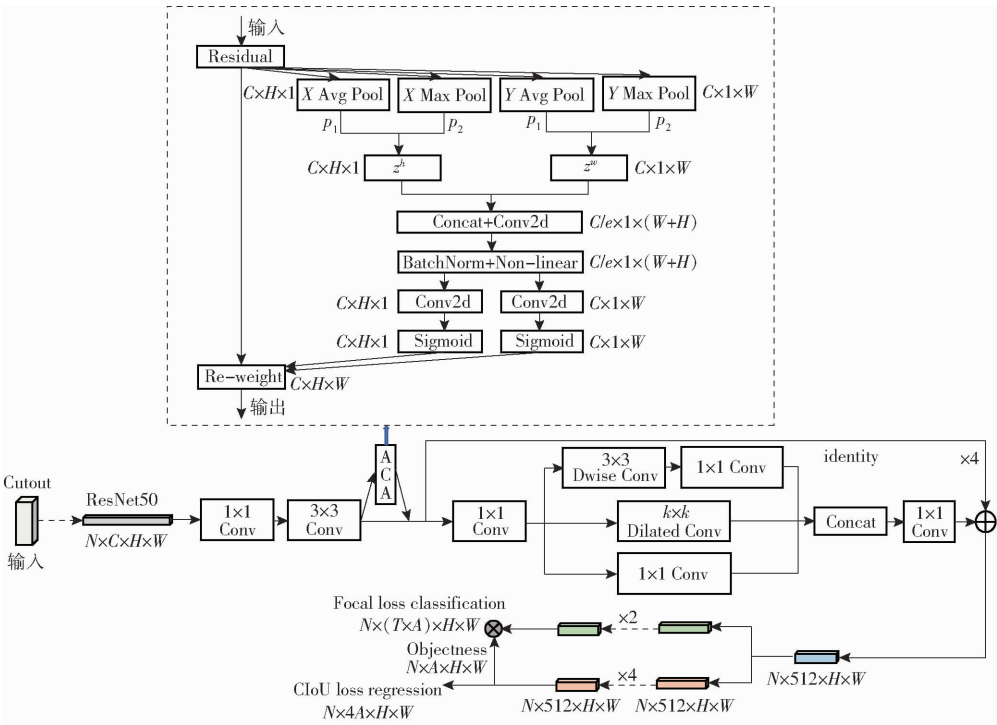


图 8 YOLOF_PD 模型结构图

Fig. 8 Structure diagram of YOLOF_PD model

用的 mAP(IoU 阈值设置为 0.5)。除此之外,与其他模型对比时,还使用模型参数量、浮点运算量、显存占用量来衡量模型复杂度,使用每秒传输图像帧数衡量模型检测效率。

3.3 对比实验

3.3.1 不同骨干网络对比实验

使用不同的主干网络来评估原始 YOLOF 模型在 FieldPest5 数据集上的性能。表 2 显示了使用 ResNet50^[27]、ResNet101^[27]、ResNest50^[32]、ResNest101^[32] 和 MobileNet V2^[33] 作为骨干网络的实验结果。

表 2 使用不同骨干网络的 YOLOF 性能

Tab.2 Performance of YOLOF by using different backbone networks

骨干网络	mAP/%	模型参数量	浮点运算量
ResNet50	91.6	4.216×10^7	1.032×10^{11}
ResNet101	90.6	6.112×10^7	1.880×10^{11}
ResNest50	91.4	4.412×10^7	1.359×10^{11}
ResNest101	89.1	6.481×10^7	2.158×10^{11}
MobileNet V2	89.3	2.078×10^7	2.630×10^{10}

从表 2 中可以看出,使用更深骨干网络的 YOLOF 模型并没有表现出更好的性能。例如,使用 ResNet101 相比使用 ResNet50 作为骨干网络的 YOLOF,在 FieldPest5 测试集上的 mAP 低 1 个百分点,使用 ResNest101 相比使用 ResNest50 作为骨干网络的 YOLOF 的 mAP 低 2.3 个百分点。原因可能

是部分害虫的尺寸较小,增加网络的深度对小物体的识别没有帮助,但却增加了计算负担。使用轻量化的 MobileNet V2 作为骨干网络的 YOLOF 在参数量和浮点运算量方面具有优势,但是相比使用 ResNet50 作为骨干网络的 YOLOF,mAP 低 2.3 个百分点。综上所述,本研究使用 ResNet50 作为 YOLOF 模型的骨干网络。

3.3.2 注意力机制对比实验

为了验证本文提出的 ACA 注意力机制的有效性,分别与原 CA 注意力机制、使用 GMP 代替 GAP 的 CA 注意力机制、使用 GMP 与 GAP 的和代替 GAP 的 CA 注意力机制、ECA 注意力机制^[34] 以及 SE 注意力机制^[20] 进行对比,所有注意力机制引入到 YOLOF 模型中的位置均为 Projector 模块 3×3 卷积后,在 FieldPest5 测试集上的 mAP 以及模型参数量对比结果如表 3 所示。

表 3 注意力机制对比实验结果

Tab.3 Comparison experiments results of different attention mechanisms

注意力机制	mAP/%	模型参数量
CA(GAP)	91.8	4.218×10^7
CA(GMP)	92.0	4.218×10^7
CA(GMP + GAP)	91.5	4.218×10^7
ACA	92.3	4.218×10^7
ECA	91.8	4.216×10^7
SE	92.0	4.219×10^7

从表 3 中可以看出,在 YOLOF 中引入 GMP 代

替 GAP 的 CA 注意力机制,在 FieldPest5 测试集上的 mAP 更高,但是将 GAP 和 GMP 简单的相加并没有取得更好的结果。本研究提出的 ACA 注意力机制能够自适应调整 GAP 和 GMP 的权重,将 ACA 注意力机制引入到 YOLOF 中,在测试集上 mAP 达到 92.3%,相比于改进前的 CA 注意力机制,提升 0.5 个百分点,并且模型的参数量保持不变。在与 ECA

和 SE 注意力机制的对比中也有良好的表现。热力图可以通过颜色变化程度,直观反映出热点分布等信息。为了更加直观地体现出本研究提出的 ACA 注意力机制的有效性,以热力图的形式对 YOLOF 中引入注意力机制后的特征进行可视化,如图 9 所示,颜色越趋近于深红色,说明模型在该区域的响应越大。

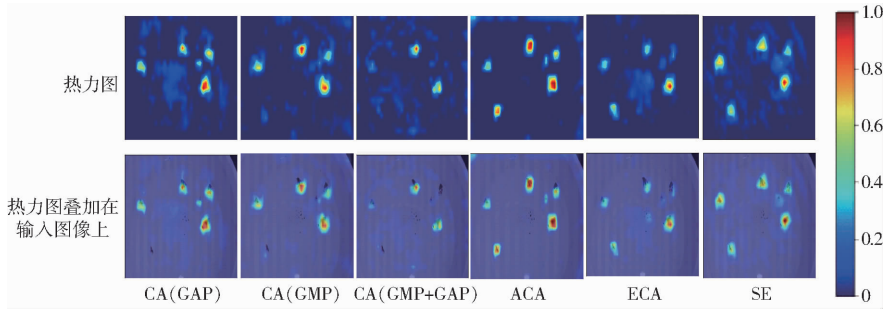


图 9 YOLOF 中引入注意力机制后的特征热力图

Fig. 9 Feature heat map after introducing attention mechanism in YOLOF

从图 9 中可以看出,在 YOLOF 中引入原 CA 注意力机制,对于小尺寸的害虫的响应能力较差,在大尺寸害虫上出现了响应范围不够准确的情况;引入 GMP 代替 GAP 的 CA 注意力机制能够稍微关注到小尺寸害虫,并且能够加强模型对大尺寸害虫的响应强度;而将 GAP 和 GMP 简单相加代替 GAP 的效果较差;引入本研究改进的 ACA 注意力机制,能够更好地关注到小尺寸害虫,对于尺寸更大的害虫关注到的主体部分也更多。引入 ECA 注意力机制出现了部分害虫响应强度较低的情况;对比表现良好的 SE 注意力机制,引入 ACA 注意力机制对于小尺寸害虫的响应强度更大,在大尺寸害虫上的响应能力也相当。

3.3.3 回归损失函数对比实验

原 YOLOF 中使用 GIoU loss 作为回归损失函数,使用 2.4 节中提到的 DIoU 和 CIoU 损失函数替换 GIoU loss 进行对比实验,在测试集上的 mAP 如表 4 所示。

表 4 回归损失函数对比实验结果

Tab. 4 Regression loss function comparison test results

		%
回归损失函数	mAP	
GIoU	91.6	
DIoU	91.9	
CIoU	92.4	

从表 4 中可以看出,在 YOLOF 模型中使用 DIoU loss 和 CIoU loss 比使用 GIoU loss 在 FieldPest5 测试集上的 mAP 更高。CIoU 在 DIoU 的基础上将重叠面积、中心点距离和长宽比考虑进来,取得了最优的

检测性能,相比于 GIoU,mAP 提升 0.8 个百分点。

3.3.4 消融实验

本文提出的 YOLOF_PD 检测器在 YOLOF 的基础上改进了 4 个要素,包括增加 Cutout 数据增强方法、使用 CIoU loss 替换 GIoU loss、使用 Projector_ACA 模块和 Residual_Concat 模块。为了分析每个组成部分在提升模型检测性能方面的有效性,进行消融实验。在 FieldPest5 测试集上的 mAP 如表 5 所示,训练过程中各个模型在验证集上的 mAP 变化情况如图 10 所示。

表 5 模型在 FieldPest5 测试集上的消融实验结果

Tab. 5 Ablation experiments results performed on

FieldPest5 test set		%
模型	mAP	
YOLOF	91.6	
YOLOF + Cutout	92.6	
YOLOF + CIoU loss	92.4	
YOLOF + Projector_ACA	92.3	
YOLOF + Residual_Concat	92.7	
YOLOF + Cutout + CIoU loss + Projector_ACA	93.4	
YOLOF_PD	93.7	

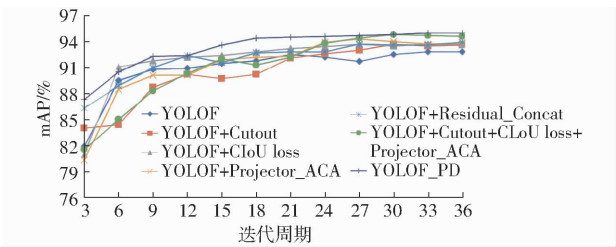


图 10 训练时各模型在验证集上的 mAP 变化曲线

Fig. 10 Variation curves of mAP for each model on validation set during training process

使用 Cutout 数据增强方式能够在一定程度上缓解害虫遮挡问题,从表 5 中可以看出,使用 Cutout 数据增强,YOLOF 模型的 mAP 从原来的 91.6% 上升到 92.6%。使用 CIoU loss 也带来了一定性能的提升,mAP 上升 0.8 个百分点。在 Project 模块中加入 ACA 注意力机制,模型在 FieldPest5 数据集上的 mAP 上升到 92.3%。使用 Residual_Concat 模块,能够更好地提取多尺度特征,mAP 增加 1.1 个百分点。综合 Cutout、CIoU loss 和 Project_ACA 模块,模型的检测性能进一步提升,mAP 从原来的 91.6% 上升到 93.4%。最终,综合改进 4 个要素的 YOLOF_PD 模型达到了最好的检测性能,在 FieldPest5 测试集上的 mAP 上升到 93.7%,较改进前提升 2.1 个百分点。

从图 10 中可以看出,迭代 36 次后,模型在 FieldPest5 验证集上的 mAP 逐渐收敛。使用 Cutout 数据增强、CIoU loss、Project_ACA 模块以及 Residual_Concat 模块比 YOLOF 在验证集上的 mAP 更高,并且 YOLOF_PD 模型取得了最优的表现,验证了本文对 YOLOF 一系列改进的有效性。为了分析 YOLOF 和 YOLOF_PD 在 5 类害虫上的具体表现,统计每一类害虫的 AP 值,如图 11 所示。

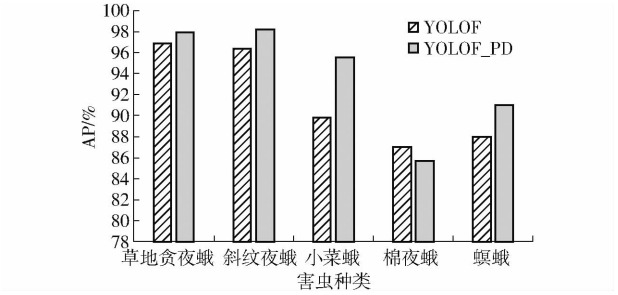


图 11 模型在具体害虫上的 AP
Fig. 11 AP of model on specific pests

从图 11 中可以看出,总体上看 YOLOF_PD 在

表 6 不同模型综合性能对比

Tab.6 Comparison of comprehensive performance of different detectors					
模型	mAP/%	帧速率/(f·s ⁻¹)	模型参数量	浮点运算量	显存占用量/MB
Faster R-CNN	91.4	23.6	4.114 × 10 ⁷	2.163 × 10 ¹¹	7 701
Dynamic R-CNN	91.6	23.6	4.114 × 10 ⁷	2.163 × 10 ¹¹	7 699
Double Head R-CNN	92.2	12.4	4.673 × 10 ⁷	4.905 × 10 ¹¹	9 591
Cascade R-CNN	91.7	20.6	6.894 × 10 ⁷	2.441 × 10 ¹¹	8 185
RetinaNet	92.5	23.5	3.619 × 10 ⁷	2.165 × 10 ¹¹	6 913
FCOS	92.9	26.5	3.185 × 10 ⁷	2.067 × 10 ¹¹	7 263
ATSS	93.4	23.8	3.190 × 10 ⁷	2.118 × 10 ¹¹	7 803
YOLOF_PD	93.7	42.4	4.284 × 10 ⁷	1.039 × 10 ¹¹	5 385

表 6 中模型使用的骨干网络均为 ResNet50,在两阶段模型中,Faster R-CNN、Dynamic R-CNN 的性能相当,Double Head R-CNN 在 FieldPest5 测试

具体害虫上的 AP 值优于改进前的 YOLOF。尽管改进后的 YOLOF_PD 在棉夜蛾上的 AP 降低 1.3 个百分点,但是在草地贪夜蛾、斜纹夜蛾、小菜蛾和螟蛾上的 AP 都有明显的提升。在害虫尺寸最小的小菜蛾上提升最为明显,AP 上升 5.7 个百分点,这也验证了对于 YOLOF 的改进在一定程度上提升了对小尺寸目标检测的性能。

为了更好地观察 YOLOF_PD 相比于 YOLOF 在害虫检测方面的差异性,将部分害虫检测结果可视化,如图 12 所示,YOLOF 漏检部分使用红圈框出。

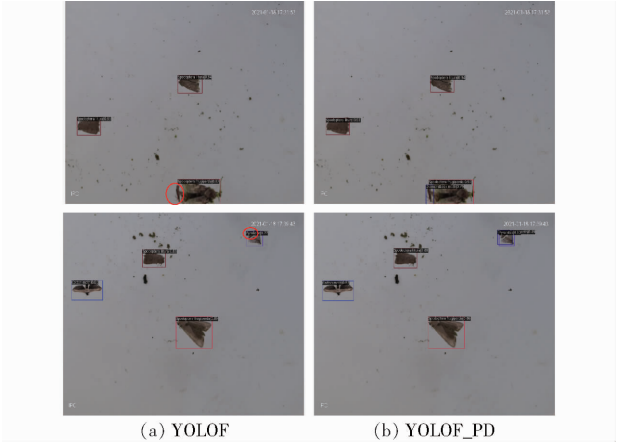


图 12 模型在 FieldPest5 测试集上的检测样例
Fig. 12 Model detection on FieldPest5 test set samples

从图 11 中可以看出,YOLOF 对于小尺寸的小菜蛾检测效果较差,尤其是当小菜蛾和其他害虫之间存在遮挡时容易漏检,改进后的 YOLOF_PD 较好地提升了遮挡情况下小尺寸害虫的检测效果。

3.3.5 与其他模型对比实验

为了更好地评估本文提出的 YOLOF_PD 的性能,与 3.1 节中提到的两阶段和一阶段的模型进行定量比较,实验结果如表 6 所示。

集上的 mAP 为 92.2%,但是帧速率只有 12.4 f/s,检测效率较低。RetinaNet、FCOS、ATSS 和 YOLOF_PD 使用 Focal loss 解决样本不平衡和难样本问题,

在 FieldPest5 测试集上的 mAP 高于两阶段检测模型。FCOS 在参数量方面占优,ATSS 检测效果较好,而本研究提出的 YOLOF_PD 检测模型 mAP 为 93.7%、帧速率为 42.4 f/s、浮点运算量为 1.039×10^{11} 、显存占用量为 5 385 MB,综合表现最佳。

4 结论

(1)针对田间害虫分布广、数据不易收集等问

题,设计诱虫灯装置,构建了包含 3 409 幅害虫图像的数据集 FieldPest5。

(2)在 YOLOF 中引入了改进的 ACA 注意力机制,在 FieldPest5 测试集上 mAP 达到 92.3%。

(3)提出了 Dilated_Dwise_ACA 编码器,提高了 YOLOF 在小尺寸害虫上的检测性能。实验结果表明,本文方法在 FieldPest5 数据集上 mAP 达到 93.7%、帧速率达到 42.4 f/s,优于 Faster R-CNN、RetinaNet 等模型。

参 考 文 献

[1] CHENNG X, ZHANG Y, CHEN Y, et al. Pest identification via deep residual learning in complex background[J]. Computers and Electronics in Agriculture, 2017, 141: 351-356.

[2] PATTNAIK G, SHRIVASTAVA V K, PARVATHI K. Transfer learning-based framework for classification of pest in tomato plants[J]. Applied Artificial Intelligence, 2020, 34(13): 981-993.

[3] 翟肇裕,曹益飞,徐焕良,等.农作物病虫害识别关键技术研究综述[J].农业机械学报,2021,52(7):1-18.

[4] ZHAI Zhaoyu, CAO Yifei, XU Huanliang, et al. Review of key techniques for crop disease and pest detection[J]. Transactions of the Chinese Society for Agricultural Machinery,2021,52(7):1-18. (in Chinese)

[5] 王林惠,兰玉彬,刘志壮,等.便携式柑橘虫害实时检测系统的研制与试验[J].农业工程学报,2021,37(9):282-288.

[6] WANG Linhui, LAN Yubin, LIU Zhizhuang, et al. Development and experiment of the portable real-time detection system for citrus pests[J]. Transactions of the CSAE,2021,37(9):282-288. (in Chinese)

[7] TENG Y, ZHANG J, DONG S, et al. MSR-RCNN: a multi-class crop pest detection network based on a multi-scale super-resolution feature enhancement module[J]. Frontiers in Plant Science, 2022, 13: 810546.

[8] LI W, WANG D, LI M, et al. Field detection of tiny pests from sticky trap images using deep learning in agricultural greenhouse[J]. Computers and Electronics in Agriculture, 2021, 183: 106048.

[9] WANG R, JIAO L, XIE C, et al. S-RPN: sampling-balanced region proposal network for small crop pest detection[J]. Computers and Electronics in Agriculture, 2021, 187: 106290.

[10] JIAO L, XIE C, CHEN P, et al. Adaptive feature fusion pyramid network for multi-classes agricultural pest detection[J]. Computers and Electronics in Agriculture, 2022, 195: 106827.

[11] LIU J, WANG X. Plant diseases and pests detection based on deep learning: a review[J]. Plant Methods, 2021, 17(1): 1-18.

[12] REN S, HE K, 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, 2016, 39(6): 1137-1149.

[13] CAI Z, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6154-6162.

[14] TIAN Z, SHEN C, CHEN H, et al. FCOS: fully convolutional one-stage object detection[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision and Pattern Recognition, 2019: 9627-9636.

[15] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.

[16] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]// Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 2017: 2980-2988.

[17] CHEN Q, WANG Y, YANG T, et al. You only look one-level feature[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13039-13048.

[18] DEVRIES T, TAYLOR G W. Improved regularization of convolutional neural networks with cutout[J]. arXiv preprint arXiv: 1708.04552, 2017.

[19] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.

[20] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.

[21] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal visual object classes (voc) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.

[22] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

[23] HU J, SHEN L, SUN G. Squeeze-and-excitation net-works[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.

[24] ZHOU B, KHOSLA A, LAPEDRIZA A, et al. Learning deep features for discriminative localization[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2921-2929.

166 – 174. (in Chinese)

[19] 王细萍,黄婷,谭文学,等. 基于卷积网络的苹果病变图像识别方法[J]. 计算机工程,2015,41(12):293 – 298.
WANG Xiping,HUANG Ting,TAN Wenxue,et al. Apple lesion image recognition method based on convolutional network[J]. Computer Engineering,2015,41(12):293 – 298. (in Chinese)

[20] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C] // Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3 – 19.

[21] 刘斌,徐皓玮,李承泽,等. 基于快照集成卷积神经网络的苹果叶部病害程度识别[J]. 农业机械学报,2022,53(6): 286 – 294.
LIU Bin,XU Haowei,LI Chengze,et al. Apple leaf disease identification method based on snapshot ensemble CNN [J]. Transactions of the Chinese Society for Agricultural Machinery,2022,53(6):286 – 294. (in Chinese)

[22] 王美华,吴振鑫,周祖光. 基于注意力改进 CBAM 的农作物病虫害细粒度识别研究[J]. 农业机械学报,2021,52(4): 239 – 247.
WANG Meihua,WU Zhenxin,ZHOU Zuguang. Fine-grained identification research of crop pests and diseases based on improved CBAM via attention[J]. Transactions of the Chinese Society for Agricultural Machinery,2021,52(4):239 – 247. (in Chinese)

[23] CHEN J, ZHANG D, ZEB A, et al. Identification of rice plant diseases using lightweight attention networks[J]. Expert Systems with Applications, 2021, 169: 114514.

(上接第 294 页)

[23] QIN Z, ZHANG P, WU F, et al. FcaNet: frequency channel attention networks [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 783 – 792.

[24] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C] // Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3 – 19.

[25] LI Y, CHEN Y, WANG N, et al. Scale-aware trident networks for object detection [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 6054 – 6063.

[26] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1 – 9.

[27] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 658 – 666.

[28] GOYAL P, DOLLÁR P, GIRSHICK R, et al. Accurate, large minibatch sgd: training imagenet in 1 hour [J]. arXiv preprint arXiv:1706.02677, 2017.

[29] ZHANG S, CHI C, YAO Y, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 9759 – 9768.

[30] ZHANG H, CHANG H, MA B, et al. Dynamic R – CNN: towards high quality object detection via dynamic training [C] // Proceedings of the European Conference on Computer Vision, 2020: 260 – 275.

[31] WU Y, CHEN Y, YUAN L, et al. Rethinking classification and localization for object detection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10186 – 10195.

[32] ZHANG H, WU C, ZHANG Z, et al. ResNest: split-attention networks [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 2736 – 2746.

[33] SANDLER M, HOWARD A, ZHU M, et al. MobileNet v2: inverted residuals and linear bottlenecks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510 – 4520.

[34] WANG Q, WU B, ZHU P, et al. ECA – Net: efficient channel attention for deep convolutional neural networks [C] // Proceeding of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2020:11531 – 11539.