

引用格式:

张鹏程, 矫桂娥, 毕卓. 基于YOLOv7的轻量化农田害虫检测算法[J]. 湖南农业大学学报(自然科学版), 2025, 51(2): 103–112.

ZHANG P C, JIAO G E, BI Z. Lightweight farmland pest detection algorithm based on YOLOv7[J]. Journal of Hunan Agricultural University(Natural Sciences), 2025, 51(2): 103–112.

投稿网址: <http://xb.hunau.edu.cn>



基于YOLOv7的轻量化农田害虫检测算法

张鹏程¹, 矫桂娥^{1,2}, 毕卓^{2*}

(1.上海海洋大学信息学院, 上海 201306; 2.上海建桥学院信息技术学院, 上海 201306)

摘要: 针对现有的害虫检测算法存在计算量和参数量大、检测精度较低等问题, 本文提出了一种基于YOLOv7的轻量化农田害虫检测算法。首先, 将轻量级GhostNetV2和PConv模块分别引入主干网络和颈部网络, 在降低网络的参数量和计算量的同时减少通道的特征冗余; 其次, 引入可变形大核注意力机制(D-LKA), 增强模型对不规则形状的目标信息的捕捉能力; 然后, 在颈部网络运用尺度内特征交互模块AIFI提升尺度内和尺度间的特征交互能力; 最后, 针对特征融合导致的特征信息丢失的问题, 引入CARAFE上采样算子, 以提高模型的感知野, 增加特征信息流通, 减少特征损失。结果表明: 改进后的算法对农田害虫的检测精度达到了72.1%; 相较于YOLOv7, 其参数量下降43.4%, 计算量下降37.0%。本文提出的检测算法在实现模型轻量化的同时, 提高了检测结果的准确率, 可为农业智能机器的研究提供参考。

关键词: 害虫检测; YOLOv7; 轻量化; 注意力机制; 特征融合

中图分类号: TP391

文献标志码: A

文章编号: 1007-1032(2025)02-0103-10

Lightweight farmland pest detection algorithm based on YOLOv7

ZHANG Pengcheng¹, JIAO Guie^{1,2}, BI Zhuo^{2*}

(1.College of Information Technology, Shanghai Ocean University, Shanghai 201306, China; 2.College of Information Technology, Shanghai Jian Qiao University, Shanghai 201306, China)

Abstract: Considering that the existing pest detection algorithms face challenges such as large computation and parameter requirements, as well as low detection accuracy, an improved lightweight agricultural pest detection algorithm was proposed based on YOLOv7. Firstly, lightweight GhostNetV2 and PConv modules were introduced into the backbone and neck networks, reducing the network's parameter and computational load while minimizing the channel redundancy. Secondly, the deformable large kernel attention(D-LKA) mechanism was incorporated to enhance the model's ability to capture irregularly shaped target information. Additionally, the attention-based intra-scale feature interaction(AIFI) module was employed in the neck network to improve intra-scale and inter-scale feature interaction. Finally, to address the issue of feature loss due to feature fusion, the CARAFE upsampling operator was introduced to increase the model's receptive field, promote the feature information flow and reduce feature loss. The results showed that compared with YOLOv7, the improved algorithm achieved a pest detection accuracy of 72.1%, with a 43.4% reduction in parameter number and a 37.0% decrease in computational load. This proposed detection algorithm not only achieved model lightweighting but also enhanced detection accuracy, providing valuable reference for research in agricultural intelligent machinery.

Keywords: pest detection; YOLOv7; lightweight; attention mechanism; feature fusion

收稿日期: 2024-04-11

修回日期: 2024-06-11

基金项目: 国家自然科学基金项目(42376194)

作者简介: 张鹏程(2000—), 男, 安徽合肥人, 硕士研究生, 主要从事计算机视觉研究, 1769884480@qq.com; *通信作者, 毕卓, 博士, 副教授, 主要从事智能系统与嵌入式视觉系统研究, bi_zhuo@126.com

害虫对作物的侵害严重影响了农业生产的正常进行。因此,提高害虫检测的准确性和效率对于维护农田生态平衡和确保农产品安全至关重要。传统的虫害监测方式往往依赖于人工巡查,主要采用灯光、信息素诱集等方法,通常依赖人工目视识别害虫种类和计数,这类方法需要专业人员进行操作,存在工作效率低、成本高、监测范围有限等问题^[1-2]。此外,由于人工监测无法实时获取害虫种群的动态信息,常常出现虫情信息获取延误及防控不及时的情况,从而造成严重的经济损失^[3]。为解决这些问题,搭载图像传感器的无人机和农业机器人被应用于害虫监测。无人机和农业机器人属于可移动、低功耗设备,一般靠电池供电;大规模农田的网络条件通常较差,因此,其识别算法以本地嵌入式算法为主。由于深度学习模型的参数量会影响实时检测的效率和效果,也决定了模型的功耗水平,故对模型进行轻量化改进是农田害虫检测算法的重点。

近年来,随着人工智能的迅猛发展,深度学习领域涌现出许多目标检测算法。这些算法可分为单阶段目标检测算法和两阶段目标检测算法2类。在单阶段目标检测方面,代表性的算法包括YOLO^[4]系列算法和SSD^[5]算法,二者直接采用回归的方法计算目标的位置并进行分类,实现了较为高效的目标检测。在两阶段目标检测方面,涌现出了一系列经典算法,包括R-CNN^[6]、Fast R-CNN、Faster R-CNN^[7]以及Mask R-CNN^[8]。这类算法先生成候选区域,再对这些区域进行分类和回归,能够得到更为精准的检测结果。虽然两阶段目标检测算法在检测精度上相对较高,但其计算耗时较长,不满足实时检测需求。单阶段目标检测算法省略了区域建议网络的步骤,直接在输入图像上生成相应的预测框来检测目标,这种实时性和高效性使得单阶段目标检测算法在农田环境下更具优势。

在以往基于深度学习的农田害虫目标检测研究中,大多数研究侧重于采用大型网络以提高检测精度,这往往导致模型参数量过多,从而导致检测速度较慢。张博等^[9]将空间金字塔池化与YOLOv3深度卷积神经网络结合,将网络输入尺度设定为608×608并减少下采样次数以避免小目标特征丢失,解决了小目标害虫样本检测率低的问题。苗海

委等^[10]基于SSD模型,采用默认框生成策略,在对浅层特征进行回归的过程中使用较小的默认框,而深层特征回归则使用较大的默认框,为小目标检测、类别间相似度较大的检测任务提供了思路。候瑞环等^[11]利用三支注意力机制,通过旋转操作和残差变换建立维度间的依存关系,以提高特征通道权重;在PANet结构上增加将跳跃连接与跨尺度连接相结合的特征融合方式,以获取更丰富的语义信息和位置信息。赵辉等^[12]在YOLOv3的FPN层分别引入注意力机制ECA和CBAM模块,以解决特征层堆叠处的特征提取问题;将CSPNet与DarkNet53中的残差模块相结合,避免出现梯度信息重复。吕宗旺等^[13]用加权双向特征金字塔Bi-FPN结构替换YOLOv5s中的特征金字塔结构,以提高多尺度特征融合效率,提升网络检测精度。赵辉等^[14]将CSP Bottleneck与基于移位窗口的Transformer自注意力机制相结合,提高了模型获取密集害虫目标位置信息的能力。

深度学习的目标检测算法在农田害虫检测中表现出良好的识别效果,但其模型内存需求较大,导致检测速度较慢。轻量化深度学习网络可以显著减少模型内存需求,提高检测速度。然而,在农田复杂环境下,轻量化网络往往面临检测精度降低的问题。因此,本文提出一种基于YOLOv7的轻量化农田害虫目标检测算法:

- 1) 引入轻量级模块GhostNetV2,代替主干网络的ELAN模块,在减少网络整体参数量的同时提高网络对目标的特征提取能力;
- 2) 将ELAN-W模块部分卷积替换为PConv,进一步实现网络轻量化目标;
- 3) 引入D-LKA(可变形大核注意力)机制,通过融合可变形卷积让模型动态地适应图像内容,从而提升模型的检测性能;
- 4) 将空间金字塔池化模块SPPCSPC更换为AIFI,提升尺度内和尺度间的特征交互能力,在提升特征提取能力的同时降低计算量;
- 5) 引入CARAFE上采样算子,提高模型的感知野,进一步理解图像中的全局结构和语境,从而提高算法对复杂场景的理解能力,提高检测精度。

1 YOLOv7

YOLOv7^[15]模型的结构主要包括：输入端(input)、主干网络(backbone)、颈部网络(neck)和头部(head)这4个关键部分。在输入端将输入的图片进行归一化，有助于提高网络的收敛速度和稳定性。将输入的图片进行Mosaic数据增强后再输入主干网络中，有助于增加数据的多样性和鲁棒性，提高模型的泛化能力。主干网络用于对图像特征进行提取，主要由CBS模块、ELAN模块、MP模块组成。CBS模块用于提取图像的局部特征，增加特征图的非线性，提高网络的拟合能力。ELAN模块基于特征金字塔网络(FPN)的理念，整合来自不同层次的特征图，融合多层次的信息，并通过组卷积(group convolution)优化网络结构，提高网络的效率和速度，更有效地处理多尺度特征。MP模块用于下采样，使用最大池化层和卷积层的组合，来降低特征图的尺寸，同时保持特征的丰富性。采用主干网络进行特征提取后将其输入颈部网络中进行特征融合。颈部网络采用SPPCSPC模块，通过引入卷积和空间金字塔池化操作，使模型更全面地捕捉输入图

像中的特征，以增强其对不同尺度目标的有效检测。头部主要由多个检测头组成，负责对颈部网络提取的特征信息进行位置和类别的预测。

2 基于YOLOv7的轻量化农田害虫检测算法

YOLOv7自2022年发布以来，被广泛地应用于各种目标检测场景中。但在农田这种开放环境下直接使用YOLOv7需要考虑以下问题：首先，YOLOv7模型虽然具有较高的检测速度和精度，但其参数量和计算量仍然较大，不适合在低功耗的边缘设备上运行；其次，光线、目标堆叠、天气环境等因素影响农田害虫图像的质量，导致害虫的特征不明显，难以识别；然后，农田害虫中有些害虫体积较小，这些害虫在图像中占据的像素较少，容易被忽略或误判；最后，农田害虫的种类繁多，且形态、颜色、大小等特征差异较大，需要收集大量的高质量农田害虫图像，并进行精确的标注，以供YOLOv7模型进行训练和测试。针对上述问题，本文对YOLOv7的主干和颈部网络进行重新设计，提出一种参数量和计算量更低、检测精度更高的网络结构。改进后的YOLOv7网络结构如图1所示。

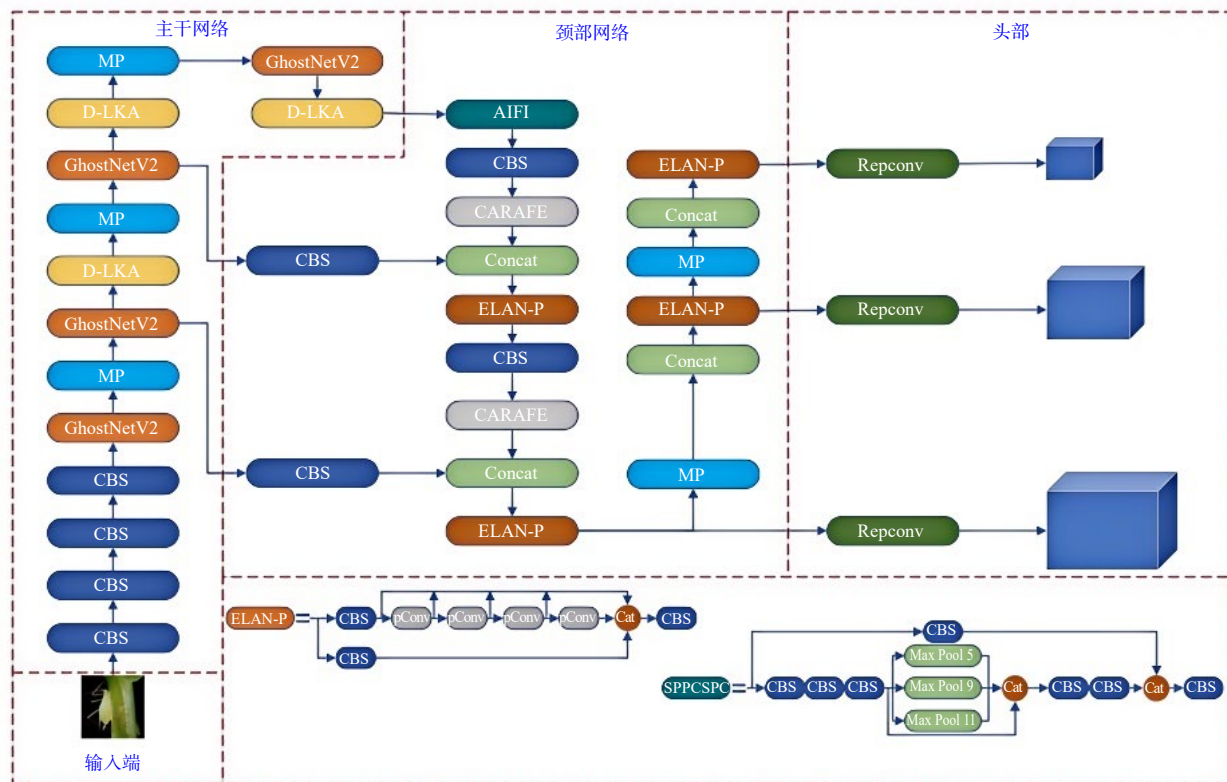


图1 改进后的YOLOv7网络

Fig.1 Improved YOLOv7 network

2.1 轻量级的ELAN-P模块和GhostNetV2模块

2.1.1 ELAN-P模块

为了构建高效的神经网络,许多研究者致力于降低浮点运算次数(FLOPs)。然而,降低FLOPs并不一定会导致相同幅度的延迟减少。这主要是因为每秒浮点运算的效率相对较低,而这种低效率主要是由深度卷积频繁进行内存访问引起的^[16]。因此,本文在YOLOv7的ELAN-W模块中引入部分卷积(partial convolution),以替换卷积核大小为 3×3 、步长为2的基础卷积,通过同时削减冗余计算和内存访问,更有效地提取空间特征。ELAN-P模块结构如图2所示。

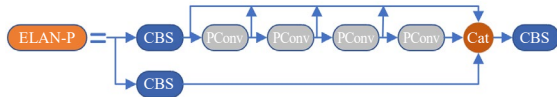


图2 ELAN-P模块结构

Fig.2 ELAN-P module structure

PConv只对输入通道的一部分应用基础卷积,而剩余的通道保持不变,利用了特征图之间的冗余性,同时减少了计算和内存访问。基础卷积在进行计算时的FLOPs为 hwk^2c^2 。其中, h 和 w 分别为特征图的高度和宽度, k 为卷积核的尺寸, c 为基础卷积通道数。而PConv的FLOPs为 $hwk^2c_p^2$,其中 c_p 为PConv的通道数。因为PConv仅处理1/4的通道,所以部分比 $r=c_p/c=1/4$,导致PConv的计算量与基础卷积的计算量之比为1/16。PConv具有较小的内存访问量,即 $2hwc_p+k^2c_p^2 \approx 2hwc_p$,而 $r=1/4$ 的基础卷积的内存访问量是 $2hwc+k^2c^2 \approx 2hwc$ 。这说明,相对于基础卷积,PConv的FLOPs仅为1/16,内存访问量也仅为1/4。因此,采用PConv替代基础卷积不仅使网络轻量化,而且能够显著提升推理速度。PConv结构如图3所示。

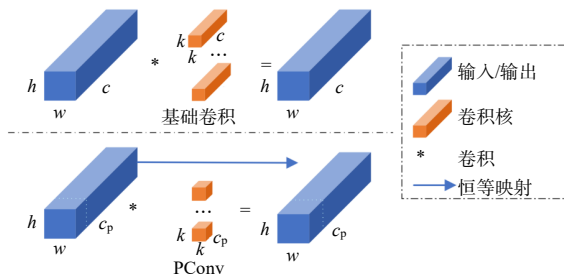


图3 PConv结构

Fig.3 PConv structure

2.1.2 GhostNetV2模块

GhostNetV2^[17]是一种新颖的视觉主干网络,本文以GhostNet为基础,引入一种新型注意力机制DFC(decoupled fully connected)。DFC注意力机制采用解耦的全连接结构,保持了轻量级卷积神经网络的高效实现,同时能够捕获长距离的空间信息,增强低成本操作的表征能力。GhostNetV2将DFC注意力和Ghost模块并联,形成全新的GhostNetV2模块,可以同时聚合局部和长距离的信息。用GhostNetV2模块替代YOLOv7主干网络的ELAN模块,能够在实现模型轻量化的同时增强其特征提取能力。

DFC注意力机制负责将输入的特征图进行压缩处理,随后在水平和垂直方向扩大感受野,并利用Sigmoid函数实现注意力权重的归一化,确保结构的有效性和稳定性。当步长为1时,为了增强扩展特征,DFC注意力分支和第1个Ghost模块同时作用于输入特征,然后将增强后的特征送入第2个Ghost模块得到输出特征,最后将其与输入特征相加得到最终特征。当步长为2时,需要在第2个Ghost模块之前加入一个深度可分离卷积(depthwise separable convolution)来调节特征的尺寸。GhostNetV2用于重新设计YOLOv7的主干网络,可以在降低参数量的同时提升主干网络的特征提取能力。GhostNetV2模块结构如图4所示。

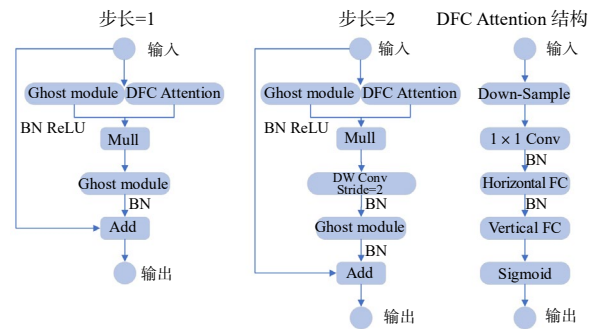


图4 GhostNetV2模块结构

Fig.4 GhostNetV2 module structure

2.2 D-LKA机制

为了更精确地提取害虫重叠区域和图像边缘区域的特征和空间位置信息,在主干网络中引入了D-LKA机制^[18]。D-LKA将大卷积核和可变形卷积的注意力机制融合,通过使用大卷积核来模拟自我关注的感受野,解决了传统自我关注机制的高计算成本的问题。此外,D-LKA通过可变形卷积来灵活调

整采样网格,使模型能够更好地适应不同的数据模式。可变形卷积(deformable convolution)用于增强模型对图像中的不规则形状的捕捉能力,通过添加额外的偏移量来调整标准卷积的采样位置,从而允许卷积核动态地适应图像的内容,有助于提升图像检测的精确性和边缘定义精度。

大卷积核通常用于捕捉图像中的丰富上下文信息,以模仿自我关注机制的感受野,但该模块占用的模型参数和计算量更少。通过深度可分离卷积和深度可分离的带扩张的卷积,能够有效地构建大卷积核。在D-LKA中,大卷积核与可变形卷积结合使用,进一步增强了模型对复杂图像模式的适应性;Conv2D为标准的2D卷积,Deform-DW Conv2D为带有偏移量的变形卷积,允许网络根据输入特征自适应地调整其感受野;偏移场由一个标准卷积层生成,用于指导变形卷积层调整其采样位置。D-LKA模块结构如图5所示。图5中, N 为输入通道数。

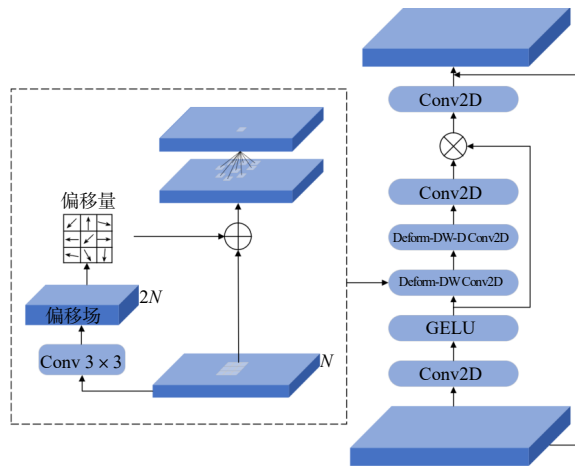


图5 D-LKA模块结构

Fig.5 D-LKA module structure

2.3 AIFI模块

AIFI^[19](attention based intra-scale feature interaction)模块是一种基于注意力的尺度内特征交互模块,它可以提高多尺度特征的表达能力和检测性能。AIFI模块的作用是:在尺度内,利用多头自注意力机制(MHSA)或可变形注意力(deformable attention)机制来增强特征的全局感知能力,捕捉特征之间的长距离依赖关系;在尺度间,利用卷积神经网络(CNN)来实现跨尺度的特征融合,保留特征的空间结构信息,同时降低计算复杂度;通过解耦尺度内交互和尺度间融合,实现高效的多尺度特征

处理,提升目标检测的精度和速度。AIFI框架如图6所示。

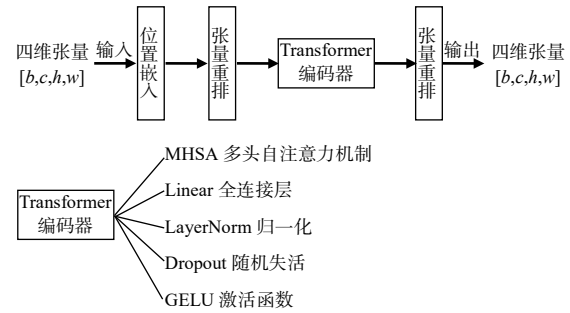


图6 AIFI框架

Fig.6 AIFI framework

AIFI的流程如下:

- 1) 输入特征。AIFI模块接受一个输入特征 $[b, c, h, w]$, 其中 b 是批量大小, c 是通道数, h 是高度, w 是宽度。
- 2) 位置嵌入。AIFI根据输入特征的宽度和高度生成2D位置嵌入。位置嵌入使用正弦、余弦函数来编码特征的位置信息, 可以增强特征的空间感知能力。
- 3) 输入张量重排, 将输入张量从 $[b, c, h, w]$ 重排为 $[b, h \times w, c]$, 以适应Transformer编码器的输入要求。
- 4) Transformer编码器。在尺度内增强特征的感知能力; 在尺度间进行特征融合, 保留空间结构信息。
- 5) 输出张量重排。Transformer编码器的输出从 $[b, h \times w, c]$ 重排为 $[b, c, h, w]$ 。通过卷积层, 对不同尺度的特征进行上采样或下采样, 使得它们具有相同的形状, 然后, 通过加法或者拼接的方式, 将它们融合为一个输出特征, 并使其形状与输入特征的形状相同。

2.4 CARAFE上采样算法

YOLOv7的上采样模块采用最近邻插值作为上采样方式, 其原理是将目标图像中每个像素点的灰度值设置为原始图像中与其位置最近的像素点的灰度值。尽管这种方法速度较快, 但缺点也很明显, 即在处理图像缩放时可能会导致图像边缘出现锯齿。为了提升农田害虫检测效果, 本文采用上采样算子CARAFE对YOLOv7的上采样模块进行改进, 其结构如图7所示。

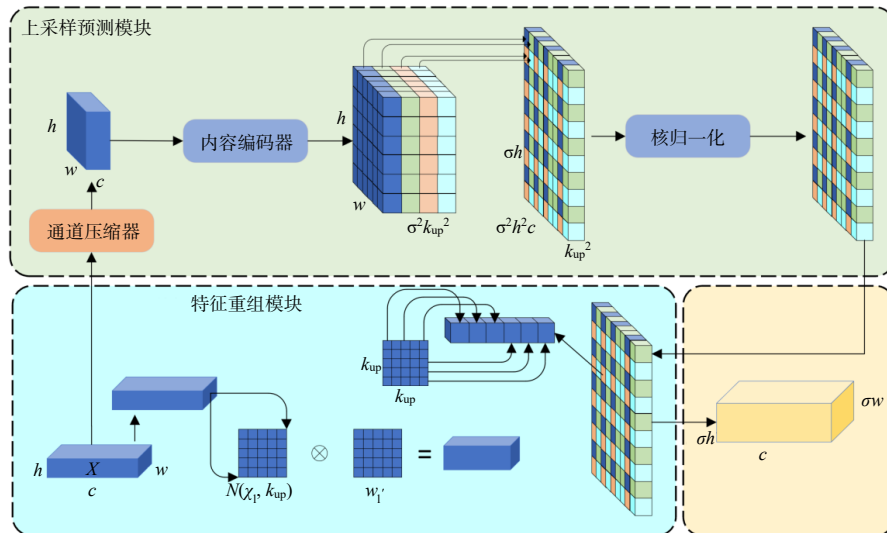


图7 CARAFE结构

Fig.7 CARAFE structure

图7中， X 为输入特征图， σ 为上采用率， k_{up} 为重组核尺寸， $N(\chi_l, k_{up})$ 为以位置 l 为中心的、尺寸为 k_{up} 的 X 的子区域， w_l 为预测感知核。

CARAFE^[20](context-aware reassembly of features) 上采样算子利用较大的感受野来聚合周围的上下文信息，可以实现更好的特征重建和上采样效果。此外，还可以根据输入特征的语义信息来动态生成自适应的上采样核，其优点在于只引入了很少的参数和计算量，不会影响速度和内存消耗。CARAFE 主要包括上采样预测模块和特征重组模块。输入特征图的通道数量在上采样预测模块被降低，然后通过内容编码卷积层得到重组核，接着将重组核展开并重新排列，再通过Softmax函数进行归一化。特征重组模块将输出特征图中的每个位置对应到输入特征图中，取出以该位置为中心的邻域，将其与该位置的预测上采样核做点积操作，得到上采样后的特征图。

与YOLOv7中使用的最近邻插值上采样方法相比，CARAFE上采样模块采用自适应生成的重组核对图像中的每个特征点执行上采样操作，使得特征点能够在更大的感受野范围内聚合上下文的语义信息。CARAFE有效地降低了上采样过程中图像特征信息的损失，同时增强了特征提取能力。这种自适应生成的重组核的应用有助于更好地保留和利用特征之间的关联，提高模型对复杂语义结构的理解和表示能力。

3 实验验证

3.1 实验环境与参数设置

本文所有实验均在同一实验环境中完成，具体如表1所示。训练网络模型时，优化器为随机梯度下降(SGD)，训练轮数(epoch)设置为300，批尺寸(batchsize)设置为16，图像输入尺寸为640像素×640像素，权重衰减系数设置为0.000 5，初始学习率设置为0.001，学习率动量设置为0.937。

表1 实验环境

Table 1 Experimental environment	
名称	配置信息
CPU	Intel Xeon Gold 6226R@2.90GHz
GPU	Nvidia GeForce RTX 3090 24 GB
内存	64 GB
操作系统	Windows 10
Python版本	3.10.13
加速环境	Cuda 11.7
深度学习框架	Pytorch 1.13.1

3.2 数据集

本文采用的数据集来自IP102数据集，它是一个用于害虫识别的大规模基准数据集(<https://github.com/xpwu95/IP102>)。该数据集包含超过75 000幅图像，分属102个类别，数据呈自然的长尾分布。此外，该数据集为约19 000幅图像标注了边界框，用于目标检测。本文从中挑选出蝼蛄、红蜘蛛、玉米螟、行军虫、甘蓝夜蛾、象鼻虫、普通刺吸虫、西

班牙蝇、水疱甲9类常见害虫图像4 675幅,将部分标签错误的图像删除,仅保留4 659幅图像,并按照8 : 1 : 1的数量比例随机划分训练集、验证集、测试集,完成数据集的构建。

3.3 评价指标

为验证本文所提出算法的性能,实验采用准确率(P)、召回率(R)、精度均值(P_{AP})、平均精度均值(P_{mAP})、10亿次浮点运算(GFLOPs)和参数量6个指标来准确客观地评价模型的性能。

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \tag{1}$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{2}$$

$$P_{AP} = \int_0^1 P(R) dR \tag{3}$$

$$P_{mAP} = \frac{1}{k} \sum_{i=1}^k P_{AP,i} \tag{4}$$

式中: N_{TP} 表示真阳性(true positive)即正样本被预测正确的数量, N_{FP} 表示假阳性(false positive)即负样本被预测为正样本的数量, N_{FN} 表示假阴性(false negative)即正样本错误地被检测成负样本的数量; $R_{AP,i}$ 为第*i*个类别的 P_{AP} ; P_{mAP} 表示所有类别的平均 P_{AP} , 其值越大表示识别目标的精度越高。 $P(R)$ 曲线用于直观展示模型在不同置信度阈值下的精确率和召回率的关系。 P_{AP} 即为 $P(R)$ 曲线下的积分面积, 面积越大, 表示识别精度越高。

3.4 消融实验

为了进一步验证基于YOLOv7的检测算法中各个改进点的效果,在同一数据集上对模型进行消融实验,结果如表2所示。

表2 消融实验结果

Table 2 Ablation experiment results										
模型	GhostNetV2	ELAN-P	AIFI	CARAFE	D-LKA	$P/\%$	$R/\%$	$P_{mAP}/\%$	GFLOPs	参数量/($\times 10^6$ 个)
YOLOv7						61.2	74.8	70.5	106.5	37.6
改进模型1	√					67.2	68.8	68.7	63.5	27.4
改进模型2	√	√				66.5	65.4	69.4	56.0	23.7
改进模型3	√	√	√			62.5	70.9	70.5	50.7	17.3
改进模型4	√	√	√	√		67.3	70.7	71.3	50.7	17.1
改进模型5	√	√	√	√	√	66.9	72.1	72.1	67.1	21.3

注: “√”表示算法中包含该改进模块。

改进模型1在主干网络引入GhostNetV2轻量化模块,相较于基准模型(YOLOv7),其 P_{mAP} 下降1.8个百分点,但参数量下降27.1%,计算量(以GFLOPs计)下降40.4%,说明GhostNetV2可以减少模型的冗余,但降低了部分检测精度。改进模型2在颈部网络引入PConv模块对原ELAN-W进行改进后, P_{mAP} 相较于改进模型1提升0.7个百分点,参数量下降了 3.7×10^6 个, GFLOPs下降7.5,说明PConv能在降低网络复杂度的同时提升网络精度。改进模型3用AIFI替换SPPCSPC模块后, P_{mAP} 提升至70.5%,模型参数量下降至 17.3×10^6 个, GFLOPs下降至50.7,表明AIFI模块能够通过自我注意力机制专注于处理高级图像特征,从而提高模型检测和识别方面的性能,同时减少不必要的计算消耗。改进模型4引入CARAFE上采样算子后, P_{mAP} 提升至71.3%,模型参数量下降至 17.1×10^6 个。改进模型5引入D-LKA后,相较于基准模型, P_{mAP} 提升1.6个百分点,模型

参数量下降43.4%,计算量下降37.0%,说明D-LKA注意力机制能够提升模型对图像中的不规则形状的捕捉能力。

3.5 注意力机制对比实验

为了进一步探究注意力机制对改进模型的影响,本文采用CBAM^[21]、SE^[22]、SimAM^[23]、CA^[24]、D-LKA这5种主流注意力机制进行消融实验。训练时控制相同参数,在本文数据集上进行实验,结果如表3所示。

表3 注意力机制对比实验结果

Table 3 Comparative experiment result of attention mechanism			
注意力机制	$P_{mAP}/\%$	GFLOPs	参数量/(10^6 个)
CBAM	68.0	50.6	17.1
SE	70.8	50.6	17.1
SimAM	68.7	50.6	17.0
CA	69.6	50.6	17.0
D-LKA	72.1	67.1	21.3

实验数据表明,本文采用的D-LKA机制相较于其他主流注意力机制具有明显优势,虽然参数量和计算量有小幅度的上升,但平均精度均值 P_{mAP} 相较于CBAM、SE、SimAM、CA分别高出了4.1个百分点、1.3个百分点、3.4个百分点、2.5个百分点。

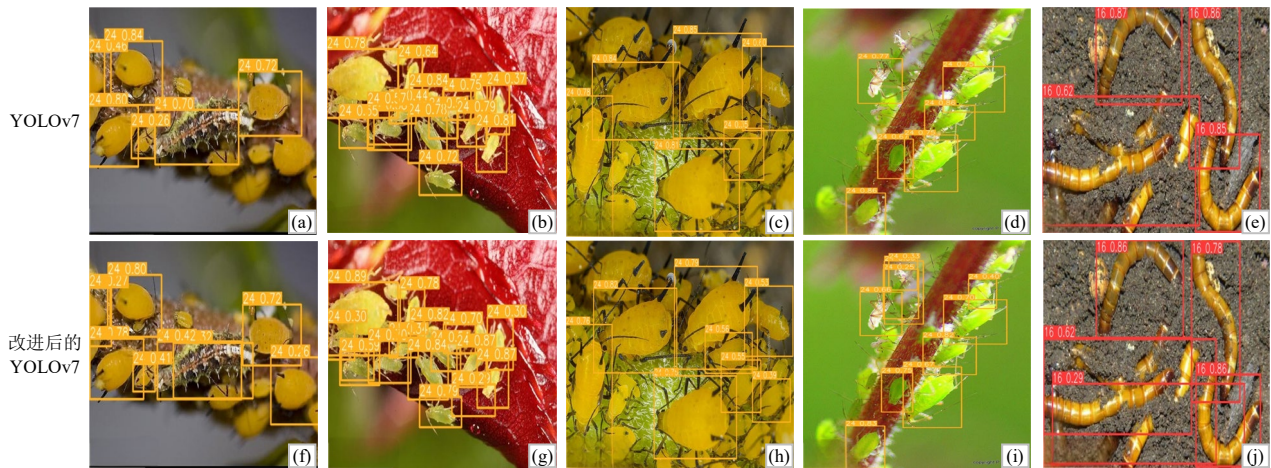
3.6 不同算法性能对比

改进后的YOLOv7算法在农田害虫场景下具有更高的检测精度和更小的模型参数量,为了验证本文算法的实际效果,在相同数据集下,将目前主流目标检测算法Faster R-CNN、SSD、YOLOv5s、YOLOv6、YOLOv7、YOLOv7-tiny、RT-DETR与本文算法进行对比实验,结果如表4所示。

表4 不同算法性能对比

算法	$P_{mAP}/\%$	GFLOPs	参数量/(10^6 个)
Faster R-CNN	59.1	74.3	174.4
SSD	63.7	78.5	24.3
YOLOv5s	70.8	15.8	7.1
YOLOv6	69.1	44.1	17.2
YOLOv7	70.5	106.5	37.6
YOLOv7-tiny	65.0	13.9	6.2
RT-DETR	71.7	110.0	32.0
改进后的YOLOv7	72.1	67.1	21.3

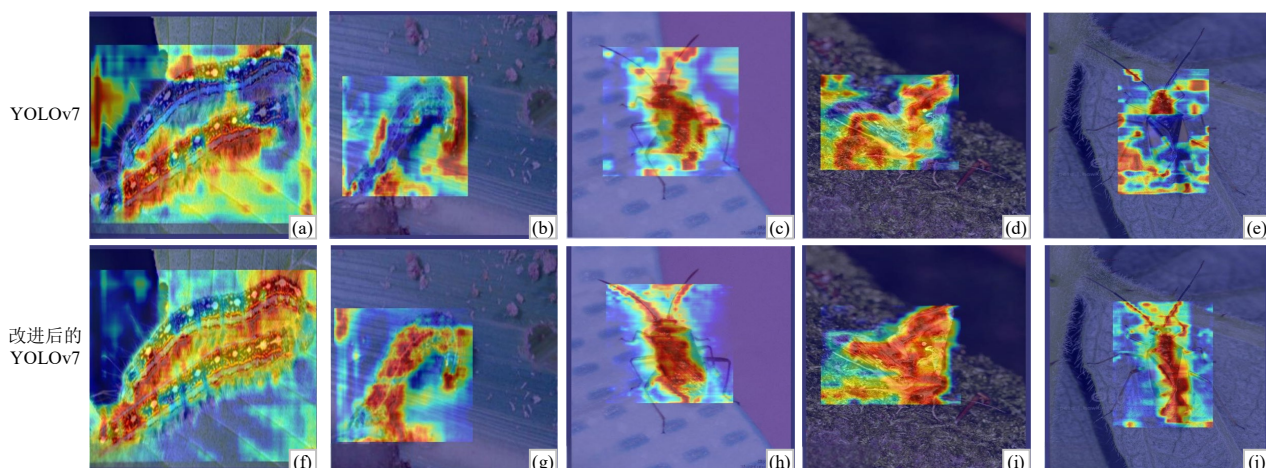
由表4可得,改进后的YOLOv7算法在8种算法中精度最高,同时计算量与参数量处于较低水平。对比YOLOv7-tiny算法,虽然其参数量和计算量较小,但其平均精度均值 P_{mAP} 仅为65.0%,在上述模型中处于较低水平;本文算法 P_{mAP} 比YOLOv5s的高1.3个百分点,比YOLOv6的高3个百分点,比YOLOv7的高1.6个百分点;本文算法的参数量和计算量远比Faster R-CNN的低, P_{mAP} 则高出13个百分点。RT-DETR虽然在精度上与本文算法接近,但在参数量和计算量上存在劣势。SSD虽然参数量较低,但计算量较大, P_{mAP} 精度仅为63.7%。本文提出的基于YOLOv7的轻量化农田害虫检测算法在保持较低的参数量和计算量、满足农田低功耗边缘设备的使用条件的同时, P_{mAP} 达到72.1%,有效地证明了本文改进的算法的先进性和可行性。改进后的算法与YOLOv7的检测效果对比以及热力图对比分别如图8、图9所示。由图8可见:改进后的YOLOv7算法在边缘区域以及对部分隐藏在土壤的害虫检测效果更好,在重叠、目标较小的害虫检测上表现更优秀。由图9可见:改进后的YOLOv7算法更加关注重叠的复杂特征,侧重于提取图片的边缘信息。



(a) 出现边缘目标漏检的情况较多; (b) 出现小目标漏检和重复检测的情况; (c) 出现重叠目标漏检的情况较多; (d) 出现因特征提取不足而漏检的情况较多; (e) 出现因土壤隐藏特征导致的漏检的情况; (f) 出现边缘目标漏检的情况较少; (g) 未出现小目标漏检和重复检测的情况; (h) 出现重叠目标漏检的情况较少; (i) 出现因特征提取不足而漏检的情况较少; (j) 未出现因土壤隐藏特征导致的漏检的情况

图8 YOLOv7改进前后的检测效果对比

Fig.8 Comparison of detection effects before and after improvement of YOLOv7



(a) 热力图分散, 模型可能无法聚焦目标; (b) 热力图破碎, 表明神经网络特征融合不足; (c) 热力图仅响应局部导致定位出现偏差; (d) 热力图偏移导致目标定位不准; (e) 热力图峰值弱导致目标模糊或类别难以区分; (f) 热力图与目标位置基本重合, 模型更加注重重叠特征; (g) 热力图完整, 神经网络特征完全融合; (h) 热力图基本覆盖整个目标, 模型能捕捉目标全局特征; (i) 热力图未发生偏移, 定位较为精确; (j) 热力图峰值较强, 目标轮廓清晰

图9 YOLOv7改进前后的热力图对比

Fig.9 Comparison of thermal map before and after improvement of YOLOv7

4 结论与讨论

受限于无人机和农业机器人设备资源有限、农田环境复杂等问题, 当前主流目标检测模型存在参数量大、检测精度低的弊端。为此, 本文提出基于YOLOv7的轻量化农田害虫检测算法。首先, 用GhostNetV2替换主干网络的ELAN模块, 并在颈部网络引入PConv重构ELAN-W模块, 以降低模型整体参数量和计算量。然后, 利用AIFI处理高级图像特征, 从而提高模型在对象检测和识别方面的性能, 同时减少计算消耗。其次, 结合CARAFE上采样算子, 减少特征损失, 提升检测精度。最后, 在主干网络用可变形大核注意力机制提升特征提取能力, 减少检测对象的位置信息丢失。结果表明, 本文提出的算法相较于YOLOv7参数量降低43.4%, 计算量(以GFLOPs计)降低37.0%, P_{mAP} 达到72.1%, 提升了1.6个百分点。本文的算法在低参数量和计算量情况下的检测效果有明显的改进。虽然本文算法检测效果较好, 但在不同季节、天气的农田环境下, 还需要进一步完善网络模型设计, 研究更加高效的算法来减少环境变化对检测结果造成的影响。

参考文献:

[1] 陈晨, 姚继刚, 李猛, 等. 农作物重要病虫害物联网监测系统建设[J]. 农业工程技术, 2022, 42(36): 19–20, 23.

[2] 何佳遥, 翟俊峰, 潘绪斌, 等. 基于物联网的农业害虫监测技术进展[J]. 植物检疫, 2023, 37(3): 1–5.

[3] 李衡霞, 龙陈锋, 曾蒙, 等. 一种基于深度卷积神经网络的油菜虫害检测方法[J]. 湖南农业大学学报(自然科学版), 2019, 45(5): 560–564.

[4] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[EB/OL]. 2015:1506.02640. <http://arxiv.org/abs/1506.02640>.

[5] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[C]//LEIBE B, MATAS J, SEBE N, et al. Computer Vision: ECCV 2016. Amsterdam, NL: ECCV, 2016: 21–37.

[6] XIE X, CHENG G, WANG J, et al. Oriented R-CNN for object detection[C]//Proceedings of the IEEE/CVF international conference on computer vision. Montreal, Canada: IEEE, 2021: 3520–3529.

[7] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28.

[8] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[C]//Proceedings of the IEEE international conference on computer vision. Venice, Italy: IEEE, 2017: 2961–2969.

[9] 张博, 张苗辉, 陈运忠. 基于空间金字塔池化和深度卷积神经网络的作物害虫识别[J]. 农业工程学报, 2019, 35(19): 209–215.

[10] 苗海委, 周慧玲. 基于深度学习的粘虫板储粮害虫图像检测算法的研究[J]. 中国粮油学报, 2019, 34(12): 93–99.

- [11] 候瑞环, 杨喜旺, 王智超, 等. 一种基于YOLOv4-TIA的林业害虫实时检测方法[J]. 计算机工程, 2022, 48(4): 255–261.
- [12] 赵辉, 李建成, 王红君, 等. 基于改进YOLOv3的水稻叶部病害检测[J]. 湖南农业大学学报(自然科学版), 2024, 50(1): 100–106.
- [13] 吕宗旺, 邱帅欣, 孙福艳, 等. 基于改进YOLOv5s的轻量化储粮害虫检测方法[J]. 中国粮油学报, 2023, 38(8): 221–228.
- [14] 赵辉, 黄镖, 王红君, 等. 基于改进YOLOv7的农田复杂环境下害虫识别算法研究[J]. 农业机械学报, 2023, 54(10): 246–254.
- [15] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE. 2023: 7464–7475.
- [16] CHEN J, KAO S, HE H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2023: 12021–12031.
- [17] TANG Y, HAN K, GUO J, et al. GhostNetv2: enhance cheap operation with long-range attention[J]. Advances in Neural Information Processing Systems, 2022, 35: 9969–9982.
- [18] AZAD R, NIGGEMEIER L, HUTTEMANN M, et al. Beyond self-attention: deformable large kernel attention for medical image segmentation[C]//Proceedings of the IEEE/CVF winter conference on applications of computer vision. Waikoloa, USA: IEEE, 2024: 1287–1297.
- [19] ZHAO Y A, LV W Y, XU S L, et al. DETRs beat YOLOs on real-time object detection[EB/OL]. 2023: 2304.08069. <https://doi.org/10.48550/arXiv.2304.08069>.
- [20] WANG J Q, CHEN K, XU R, et al. CARAFE: content-aware reassembly of features[C]//IEEE International Conference on Computer Vision. Seoul, Republic of Korea: IEEE, 2019: 3007–3016.
- [21] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Computer Vision-ECCV 2018. Munich, Germany: Springer, 2018: 3–19.
- [22] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City, USA: IEEE, 2018: 7132–7141.
- [23] YANG L, ZHANG R Y, LI L, et al. SimAM: a simple, parameter-free attention module for convolutional neural networks[C]//International conference on machine learning. Cambridge, UK: PMLR, 2021: 11863–11874.
- [24] 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. Piscataway, USA: IEEE, 2021: 13713–13722.

责任编辑: 伍锦花

英文编辑: 张承平