

DA-YOLO: 基于双重注意力的松枯病检测模型

邹素华¹, 郑秀玲¹, 杨鹏¹, 刘雪丽², 崔博琰¹

¹北华航天工业学院计算机学院, 河北 廊坊

²北华航天工业学院马克思主义学院, 河北 廊坊

收稿日期: 2024年2月23日; 录用日期: 2024年3月11日; 发布日期: 2024年3月19日

摘要

松枯病(PWD)是一种传播迅速、杀伤力极强的森林病害, 对我国森林生态安全构成严重威胁, 并造成巨大的林业经济损失。考虑到我国森林面积广阔, 人工巡查监测难度大且成本高, 因此利用无人机遥感技术监测病树成为控制松枯病传播的有效途径。尽管目前松枯病检测算法取得了相对较好的性能, 但由于松枯病的强传染性, 检测效果仍需进一步提高。基于此, 本文提出了一种基于YOLOv5的双重注意力混合模型——DA-YOLO, 用于更有效地检测病害树木区域。该算法使用基于自注意力的CoT模块加强骨干特征网络的提取能力, 并结合ECA注意力机制整体提升定位精度。实验结果显示, 在使用PWD遥感数据集时, 该模型的AP@0.5:0.95较之基线提高了5.2个百分点。并将本文提出的算法DA-YOLO与Faster R-CNN、RetinaNet、YOLOv5、YOLOv6、YOLOx、YOLOv7等算法的模型复杂度和精度进行对比, 并分析了它们在检测松枯萎线虫树木方面的效果。实验结果表明, 本文提出的DA-YOLO模型在检测方面具有明显的优势。

关键词

松枯病检测, YOLOv5, Transformer, 双重注意力机制

DA-YOLO: A Detection Model for Pine Wilt Disease Based on Dual Attention

Suhua Zou¹, Xiuling Zheng¹, Peng Yang¹, Xueli Liu², Boyan Cui¹

¹School of Computer, North China Institute of Aerospace Engineering, Langfang Hebei

²School of Marxism, North China Institute of Aerospace Engineering, Langfang Hebei

Received: Feb. 23rd, 2024; accepted: Mar. 11th, 2024; published: Mar. 19th, 2024

Abstract

This Pine wilt disease (PWD) is a rapidly spreading and highly lethal forest disease that poses a

文章引用: 邹素华, 郑秀玲, 杨鹏, 刘雪丽, 崔博琰. DA-YOLO: 基于双重注意力的松枯病检测模型[J]. 软件工程与应用, 2024, 13(2): 145-154. DOI: 10.12677/sea.2024.132015

serious threat to China's forest ecological security and causes significant economic losses in forestry. Considering the vast forest area in China, the difficulty and high cost of manual inspection and monitoring, the use of drone remote sensing technology to monitor diseased trees has become an effective way to control the spread of pine wilt disease. Although the current pine wilt disease detection algorithm has achieved relatively good performance, the detection effect still needs to be further improved due to the strong infectivity of pine wilt disease. Based on this, this article proposes a dual attention hybrid model called DA-YOLO based on YOLOv5, which is used to more effectively detect areas of diseased trees. This algorithm uses a self attention based CoT module to enhance the extraction ability of the backbone feature network, and combines ECA attention mechanism to improve the overall positioning accuracy. The experimental results show that when using the PWD remote sensing dataset, the model's AP@0.5 0.95 is 5.2 percentage points higher than the baseline. And the model complexity and accuracy of the algorithm DA-YOLO proposed in this article were compared with Faster R-CNN, RetinaNet, YOLOv5, YOLOv6, YOLOx, YOLOv7 and other algorithms, and their performance in detecting pine wilt nematode trees was analyzed. The experimental results show that the DA-YOLO model proposed in this paper has significant advantages in detection.

Keywords

Pine Wilt Detection, YOLOv5, Transformer, Dual Attention Mechanism

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

由松材线虫(PWN)引起的松枯病(PWD), 因其高致病性、迅速传播和广泛传播途径等特点对我国的森林资源造成了严重破坏[1]。由于森林面积广阔, 仅依靠人工巡查来监测病木, 成本高且监测力度有限, 无法满足大规模森林的监测需求。因此, 寻找高效、经济、准确的监测方法已成为森林病害防控领域的紧迫需求。随着地理信息系统和传感器技术的进步, 遥感技术已经能够高效地检测到大范围害虫侵害[2], 同时, 它在宏观层面上也已经能够进行自动化的实时监测, 显示出其在应用上的广阔前景。在众多遥感技术中, 无人机遥感技术以其快速响应和广泛覆盖的能力, 能够在较短时间内对广大区域进行监测, 同时收集到高质量的图像资料, 这对于及时识别受病虫害影响的树木尤为有效。

在过去几十年中, 无人机遥感技术在监测松树枯萎病等疫情方面的发展经历了持续的演进, 催生了多种图像检测算法。传统的机器学习算法如支持向量机(SVM)、随机森林(RF)和人工神经网络(ANN)等, 通过与光谱和空间特征的结合, 不断得到改进并在多光谱(MSI)和高光谱(HSI)数据集的松树损伤识别中取得成功。尽管如此, 这些经典机器学习方法在特征选择和组合方面需要复杂的操作, 并且往往无法充分利用图像的深层信息。随着深度学习在计算机视觉领域的飞速发展, 这一技术与农业信息感知的结合为农作物病虫害检测开辟了新的研究方向, 并迅速成为农业信息化领域的热点。深度学习能够自动学习图像的高级抽象特征, 从而简化了传统特征工程的复杂性。因此, 利用智能分析技术对遥感图像进行处理, 以识别森林中的受损树木, 不仅变得可行, 而且提高了检测的准确性和效率。许可等人[3]采用了将三种注意力机制与 YOLOv5 结合, 实现了对变色松树和枯死疫木检测效果的提升。李博等人[4]使用改进后的 RetinaNet 算法自动识别无人机遥感影像上的松材线虫病变色疫木。尽管在当前的研究进展中, 针对松枯病的检测算法已经展现出一定的性能优势, 然而, 鉴于松枯病具有较高的传染性和对森林生态系统

的潜在破坏性，这些检测算法在实际应用中的效果仍然有待进一步提升。在现实操作中，当无人机以较高速度或较大高度飞行时，所检测到的树木病虫害目标在图像中呈现出较小的尺寸，这影响了检测的精确度。同时，考虑到害虫侵害可能在图像中以不同尺寸呈现，实现对多尺度害虫的有效识别变得至关重要。

2. 方法

2.1. 所提出的模型 DA-YOLO

在深度学习目标检测框架 YOLOv5 中，骨干特征提取网络采用了全卷积的网络架构，通过堆叠多个卷积模块来提取特征，这在提取全局特征以及确保病害树木的准确检测方面存在一定的局限性。然而，这种设计限制了网络在获取全局上下文信息方面的能力。在本文的研究中，针对植物病虫害检测(PWD)任务，由于单粒度目标尺度较小，且部分检测目标与周围环境具有较高的相似性，这增加了检测的难度。CSPDarkNet53 主要依赖于卷积操作进行特征提取，这在提取全局特征以及确保病害树木的准确检测方面存在一定的局限性。因此，针对 PWD 检测任务的特性，本文提出了改进措施。

在本文的研究中，为了提升目标检测算法在特定任务中的性能，我们对 YOLOv5 框架中的骨干特征提取网络进行了创新性改进。具体而言，在骨干网络加入一种基于 Transformer 的 CoT 模块[5]，该结构中的自注意力机制能够在不同尺度的特征图上同时进行关注，从而有效地提取图像的全局信息。通过 CoT 模块，能够确保模型在处理每张图像时，不仅能够捕捉到局部细节，还能充分考虑整体上下文信息，以提高对小尺度且相似背景下的病虫害目标的检测准确性和鲁棒性。此外，本文使用了另一种注意力机制——高效通道注意力(ECA) [6]协同自注意力增强模型检测效果。新的模型结构(DA-YOLO)如图 1 所示。

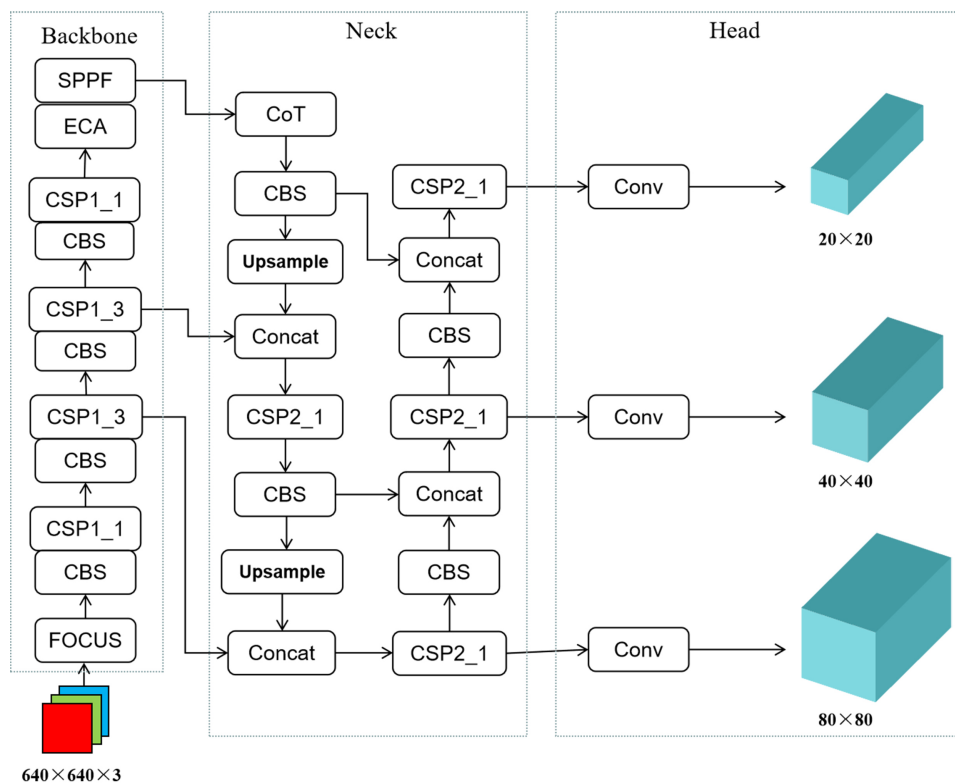


Figure 1. Overall structure of DA-YOLO

图 1. DA-YOLO 的总体结构

2.2. 基线网络

本研究对比了几种基线模型在检测性能上的表现，具体实验结果列于表 1 中。综合来看，YOLOv5 在检测效果方面表现较为出色。因此，本研究选择将 YOLOv5 作为基准网络。

Table 1. Experimental result of different detection models
表 1. 不同检测模型的实验结果

方法	AP@0.5(%)	AP@0.5:0.95(%)	Parameters (M)	GFLOPs
YOLOv5	97.6	90.8	7.1	15.8
Faster-RCNN	75.2	66.2	41.1	78.1
RetinaNET	95.9	92.5	36.1	81.6
YOLOV6	95.9	80.8	17.1	21.8
YOLOv7	82.5	55.9	6.5	13.9
YOLOX	96.0	84.3	8.9	13.3

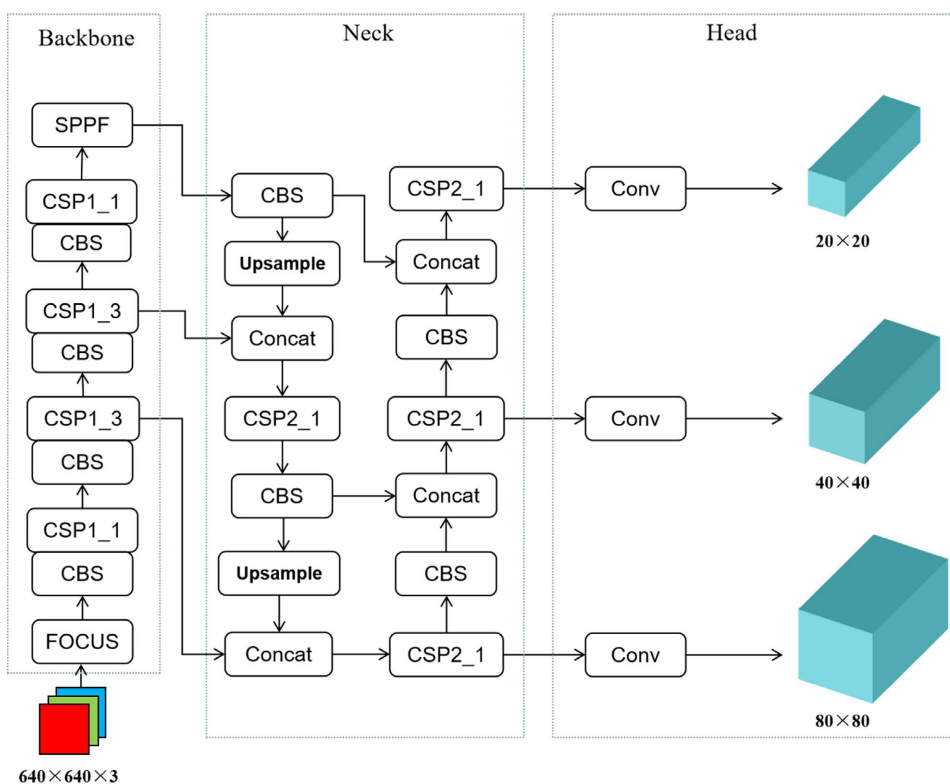


Figure 2. Network architecture of YOLOv5
图 2. YOLOv5 的网络架构

YOLOv5 的网络结构可以分为以下三个部分：骨干段、颈部段和头部输出端，具体结构图如图 2。YOLOv5 使用 CSPDarknet53 作为骨干网络，通过卷积操作将图像转化为特征图，对输入图像进行特征提取，提取出图像中的语义和结构信息。Yolov5 根据不同的尺度构建了一个特征金字塔 PANet。这个特征金字塔由多个不同分辨率的特征图层组成，每个图层都对应一个不同尺度的感受野。在每个尺度的特征图层上，Yolov5 使用 Anchor 框来预测目标的类别和位置。对于每个 Anchor 框，检测器会输出预测的边

界框的坐标、类别以及置信度等信息。由于一个目标可能在不同的尺度和位置上被多个 Anchor 框检测到,需要进行非最大抑制(NMS)来剔除冗余的框。NMS 操作会选择置信度最高的框,并消除与其重叠超过一定阈值的其他框。最终, YOLOv5 将经过 NMS 处理后的检测框输出作为最终的目标检测结果。每个检测框包含目标的类别、坐标和置信度等信息。

2.2. 骨干特征提取网络的改进

Transformer 因其出色的全局建模能力在各种自然语言处理(NLP)任务中得到了广泛应用。受 Transformer 结构的启发, ViT、DETR 等模型也借鉴了 Transformer 的结构, 以实现长距离依赖关系的建模。然而, 在原始的 Transformer 中, Self-Attention 结构(图 3(a))仅仅根据查询(query)和键(key)之间的交互来计算注意力矩阵, 因此忽略了相邻键(key)之间的关联。CoTNet 构建了一个新的 Transformer 模块 Contextual Transformer (CoT), 该模块超越了传统的自注意力机制, 将上下文信息挖掘和自注意力学习集成在一起, 通过充分利用了相邻键之间的上下文信息以促进自注意力学习, 并增强输出聚合特征图的表达能力。CoT 架构如图 3(b)所示。

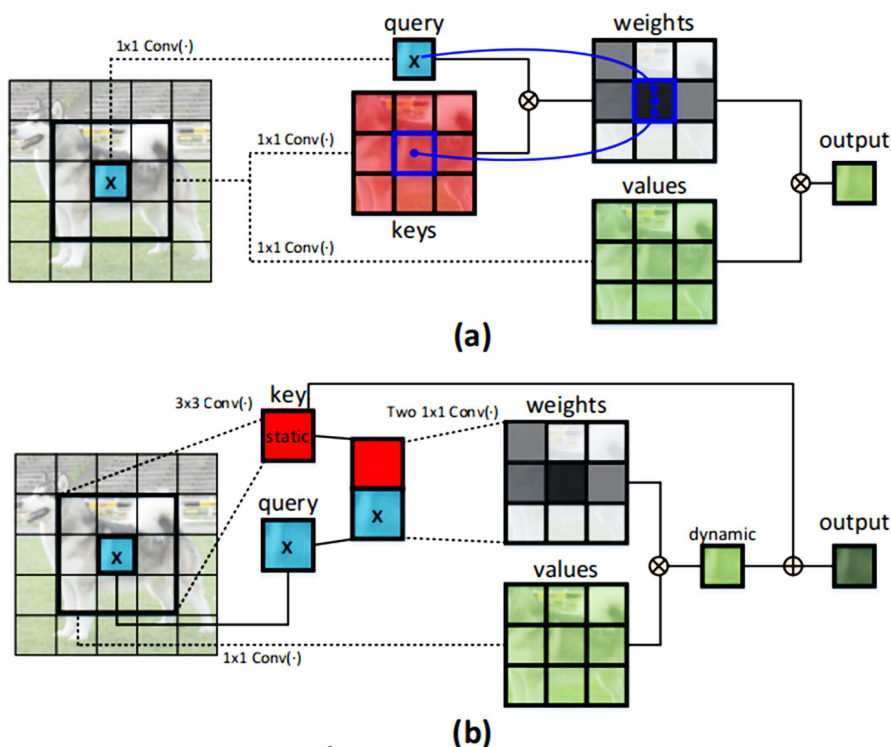


Figure 3. Traditional self-attention mechanism (a) and Context Transformer (CoT) module (b)
图 3. 传统的自注意力机制(a)和 Contextual Transformer (CoT)模块(b)

2.3. ECA 机制

在传统的 Transformer [7]架构中, 自注意力机制被用于捕捉输入序列中不同位置之间的依赖关系, 并且在每一层中都需要进行计算。然而, 这种方式忽略了通道间的关联性, 以及不同通道之间的重要性。ECA 注意力机制可以增强 Transformer 对通过 ECA 注意力机制, 模型可以更好地关注和加权每个通道之间的相关性, 从而在保持模型有效性的同时提高模型的预测性能。输入特征的抽取能力, 通过突出不同通道中的重要信息, 提高模型对输入序列的建模能力。ECA 模块如图 4 所示。

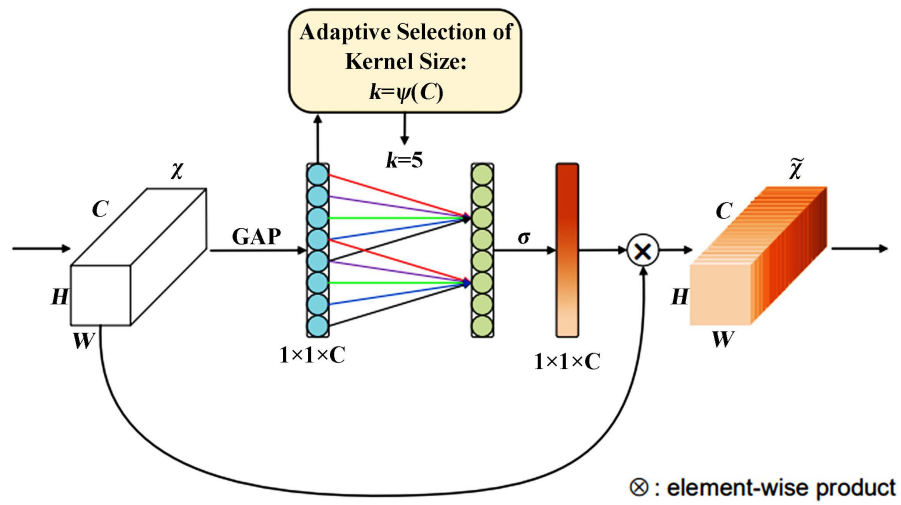


Figure 4. Diagram of efficient channel attention (ECA) module
 图 4. 高效通道注意力(ECA)模块的图示

3. 实验

3.1. 数据集获取

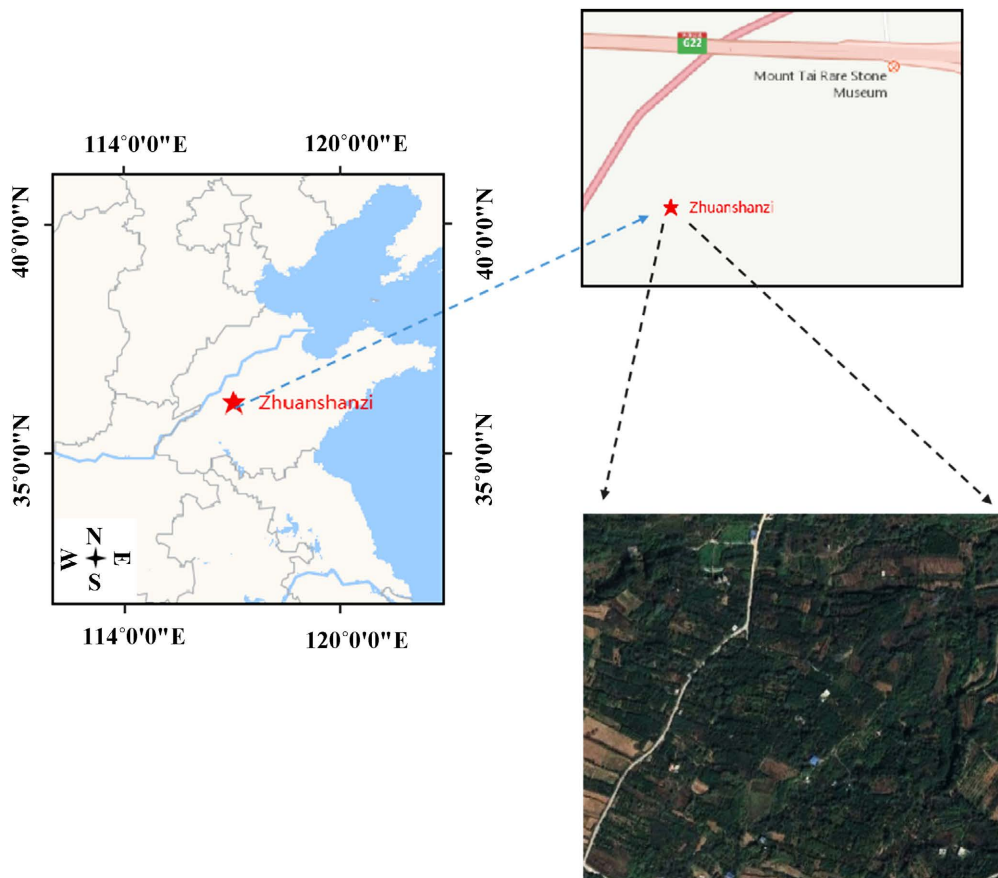


Figure 5. Data acquisition area
 图 5. 数据采集区

本文所用的图像数据收集于山东省泰安市转山子林场(北纬 31°14', 东经 117°36', 海拔 40 m)。图 5 显示了数据采集区域。

本文使用搭载 4800 万像素可见光摄像头的无人机, 获取分辨率为 6000 × 4000 像素的遥感图像数据。高分辨率图像随后裁剪为 640 × 640 像素的图像块。其中, 随机选取 10,000 个图像块作为训练集图像, 1200 个图像块作为验证集图像。采用 LabelImg 软件进行图像标注。图 6 是分辨率为 6000 × 4000 像素的无人机林场图像示例。

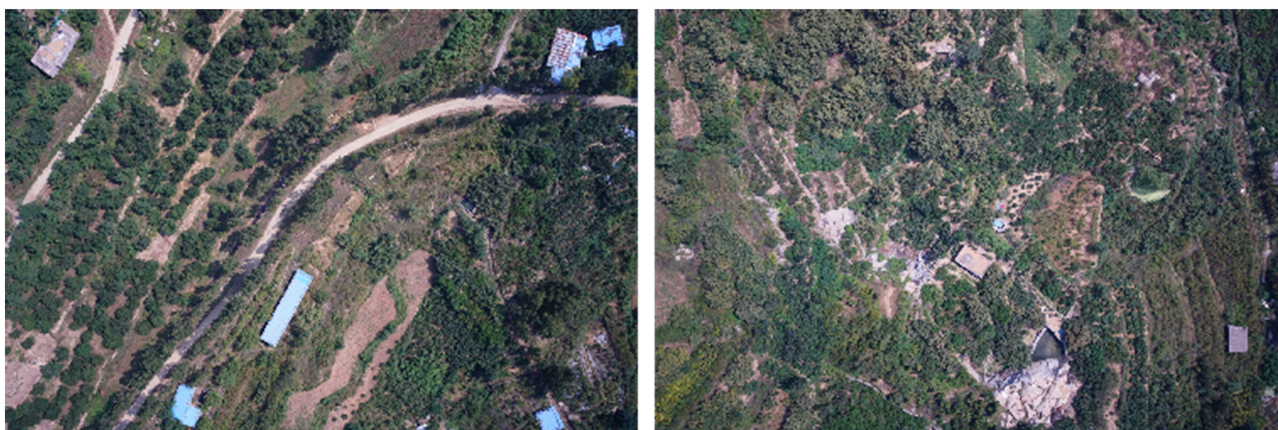


Figure 6. Examples of UAV tree images

图 6. 无人机树图像示例

3.2. 实验配置和评价指标

实验硬件和软件环境配置如表 2 所示。

Table 2. System resulting data of standard experiment

表 2. 标准试验系统结果数据

实验平台	配置
操作系统	Linux 3.10.0
CPU	Intel(R) Xeon(R) Gold 6138 CPU @ 2.00GHz
GPU	Tesla V100-PCIE-32GB
GPU 加速器	CUDA 10.2
深度学习框架	PyTorch 1.10.1
编译器	PyCharm & Anaconda
语言	Python 3.7

为了正确评估新模型, 我们使用以下几个性能评估指标包括平均精度(AP)、参数量和模型参数量每秒千兆浮点运算(GFLOPs)。AP 是单个目标类别的平均精度, 衡量模型的整体检测效果。AP 的计算公式如下:

$$AP = \int_0^1 Precision(Recall) d(Recall)$$

其中 Precision 表示为预测框中正确预测的比例, Recall 表示为所有实框中的预测比例。

3.3. 实验结果

3.3.1. 与其他算法的比较

为了评估我们提出的模型的有效性，我们对比了主流目标检测模型 Faster R-CNN [8]、RetinaNet [9]、YOLOv5、YOLOv6 [10]、YOLOv7 [11]和 YOLOx [12]在松枯病数据集的检测效果。本实验使用 AP@0.5、AP@0.5:0.95 这几个指标来衡量算法在 PWD 检测精度方面的性能。

Table 3. Comparative assessment of different detection models

表 3. 不同检测模型的性能比较

方法	AP@0.5(%)	AP@0.5:0.95(%)	Parameters (M)	GFLOPs
YOLOv5	97.6	90.8	7.1	15.8
Faster-RCNN	75.2	66.2	41.1	78.1
RetinaNET	95.9	92.5	36.1	81.6
YOLOV6	95.9	80.8	17.1	21.8
YOLOv7	82.5	55.9	6.5	13.9
YOLOX	96.0	84.3	8.9	13.3
DA-YOLO	97.2	96.0	8.19	16.8

从表 3 中的数据可以看出，与其他检测器相比本文提出的 DA-YOLO 模型在 AP@0.5:0.95 上达到最优为 96%。与基线网络 YOLOv5 相比，DA-YOLO 模型的 AP@0.5:0.95 提高了 5.8 个百分点，这表明本文的模型对于更加严格的 IoU (Intersection over Union) 阈值(0.5 到 0.95 范围内)的松枯病检测性能有所改善，能够更准确地检测目标的位置和边界框的精度提高。

3.3.1. 消融实验

为了更好地评估本文提出的每个模块的有效性，首先，我们使用 YOLOv5 网络作为基线，然后将 CoT 模块加入骨干网络，并与 ECA 进行结合，在 PWD 数据集测试各个模块的有效性。消融实验的结果如表 4 所示。在基线网络上加入 CoT 模块后，AP@0.5:0.95 增加了 4.6 个百分点。而且完善 ECA 模块后，AP@0.5:0.95 增加了 0.6 个百分点。

Table 4. Detection effects of different modules on the model

表 4. 不同模块对模型的性能检测效果

方法	AP@0.5(%)	AP@0.5:0.95(%)	Parameters (M)	GFLOPs
基线模型	97.6	90.8	7.1	15.8
+CoT	97.2	95.4	8.19	16.8
+ECA	97.2	96.0	8.19	16.8

3.3.2. 改进算法的性能评估

图 7 展示了模型训练的曲线。在经过 100 次迭代后，损失函数值呈现出较小的波动，表明模型已进入相对稳定的状态。同时，模型的精度、召回率和 mAP 等性能指标尽管仍在缓慢下降，但变化幅度已极为有限，这表明模型已达到相对平衡状态，可认为模型已基本收敛。

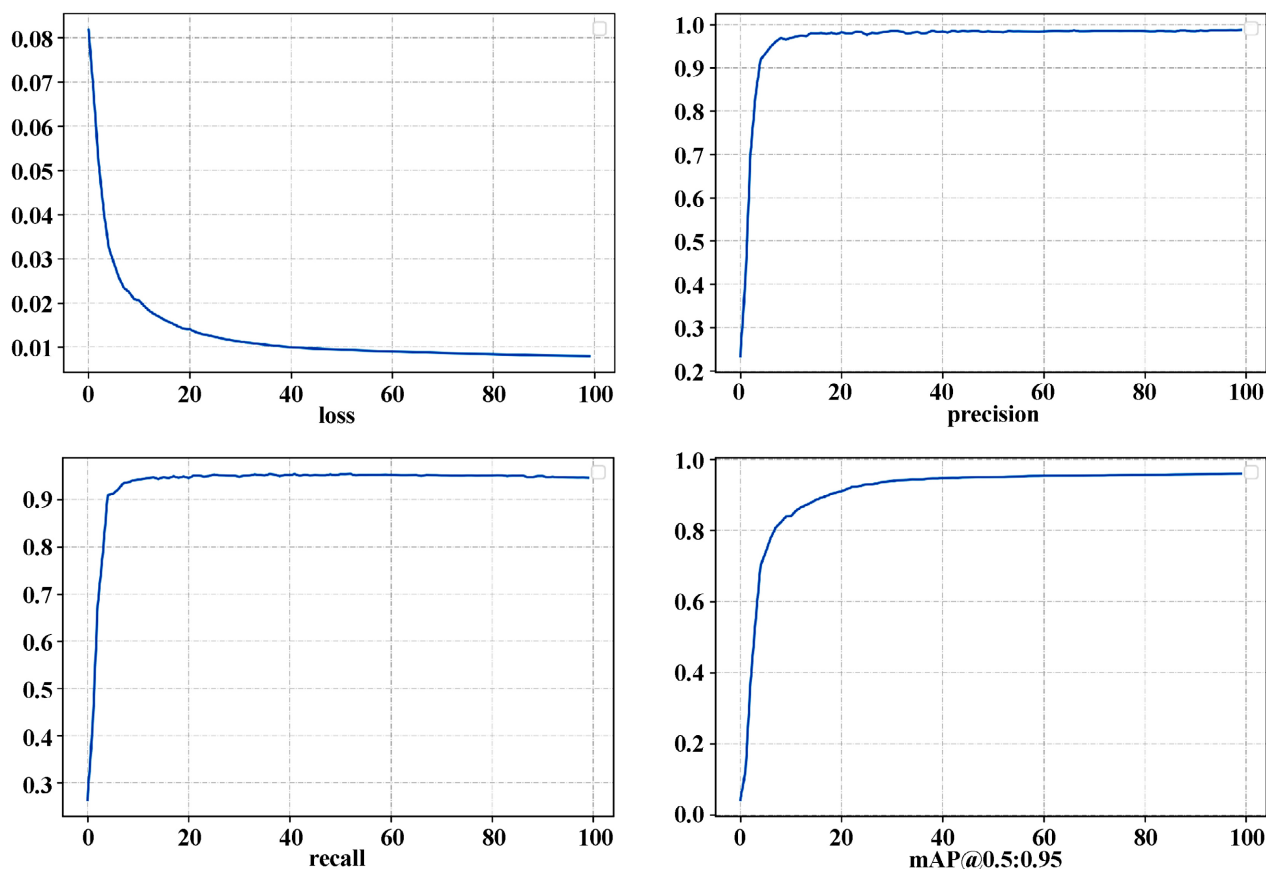


Figure 7. Curves for model training

图 7. 模型训练曲线

4. 结论

本文提出了一种基于 YOLOv5 的无人机遥感监测林场松材线虫病害的模型(DA-YOLO)。首先,针对 CNN 在全局特征提取中的不足,在骨干网络加入 CoT 模块,利用输入键之间的上下文信息来促进自注意力学习,从而提高检测网络的表征能力,并且引入高效通道注意力模块(ECA)来提升松枯病定位和识别精度。实验结果表明,本文的算法的检测精度超越众多目标检测器。后续工作将进一步对模型进行压缩来实现无人机平台的实时检测。

基金项目

校级研究生创新资助项目(项目编号: YKY-2022-41); 校级研究生创新资助项目(项目编号: YKY-2022-39); 校级研究生创新资助项目(项目编号: YKY-2023-30); 校级研究生创新资助项目(项目编号: YKY-2023-38); 河北省高等学校人文社会科学研究项目(项目编号: SZ2023091);

参考文献

- [1] Pan, C.S. (2011) Development of Studies on Pinewood Nematodes Diseases. *Journal of Xiamen University*, **50**, 476-483.
- [2] Duarte, A., Borralho, N., Cabral, P., *et al.* (2022) Recent Advances in Forest Insect Pests and Diseases Monitoring Using UAV-Based Data: A Systematic Review. *Forests*, **13**, Article 911. <https://doi.org/10.3390/f13060911>
- [3] 许可, 季卓, 夏凯, 等. 结合注意力机制与 YOLOv5 的松材线虫病受害木检测[J]. 林业工程学报, 2023, 8(3):

- 156-164. <https://doi.org/10.13360/j.issn.2096-1359.202208018>
- [4] 李博, 姚黎帆, 王广彪, 等. 基于改进 RetinaNet 算法和无人机影像的松材线虫病疫木检测[J]. 现代信息科技, 2023, 7(14): 71-74+79. <https://doi.org/10.19850/j.cnki.2096-4706.2023.14.016>
- [5] Li, Y., Yao, T., Pan, Y., *et al.* (2022) Contextual Transformer Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **45**, 1489-1500. <https://doi.org/10.1109/TPAMI.2022.3164083>
- [6] Wang, Q., Wu, B., Zhu, P., *et al.* (2020) ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 11534-11542. <https://doi.org/10.1109/CVPR42600.2020.01155>
- [7] Vaswani, A., Shazeer, N., Parmar, N., *et al.* (2017) Attention Is All You Need. In: Guyon, I., *et al.*, Eds., *Advances in Neural Information Processing Systems* 30, Long Beach, 04-09 December 2017, 1-11.
- [8] Ren, S., He, K., Girshick, R., *et al.* (2015) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In: Cortes, C., *et al.*, Eds., *Advances in Neural Information Processing Systems* 28, Montreal, 07-12 December 2015, 1-9.
- [9] Lin, T.Y., Goyal, P., Girshick, R., *et al.* (2017) Focal Loss for Dense Object Detection. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, 22-29 October 2017, 2980-2988. <https://doi.org/10.1109/ICCV.2017.324>
- [10] Li, C., Li, L., Jiang, H., *et al.* (2022) YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. arXiv preprint arXiv:2209.02976.
- [11] Wang, C.Y., Bochkovskiy, A. and Liao, H.Y.M. (2023) YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver, 17-24 June 2023, 7464-7475. <https://doi.org/10.1109/CVPR52729.2023.00721>
- [12] Ge, Z., Liu, S., Wang, F., *et al.* (2021) YOLOX: Exceeding YOLO Series in 2021. arXiv preprint arXiv:2107.08430.