

引用格式: 罗旭东, 吴一全, 陈金林. 无人机航拍影像目标检测与语义分割的深度学习研究方法研究进展[J]. 航空学报, 2024, 45(6): 028822. LUO X D, WU Y Q, CHEN J L. Research progress on deep learning methods for object detection and semantic segmentation in UAV aerial images[J]. Acta Aeronautica et Astronautica Sinica, 2024, 45(6): 028822 (in Chinese). doi: 10.7527/S1000-6893.2023.28822

# 无人机航拍影像目标检测与语义分割的深度学习研究方法研究进展

罗旭东, 吴一全\*, 陈金林

南京航空航天大学 电子信息工程学院, 南京 211106

综述

**摘要:** 无人机(UAV)因其小型轻便、操作简单等优点获得了广泛的应用。将深度学习方法与无人机系统相结合,有助于在清晰度高、视野范围广的无人机航拍影像上快速准确地检测出所需目标,相关课题已成为当前的研究热点之一。对近十年来无人机航拍影像目标检测与语义分割的深度学习研究方法研究进展进行了综述。首先概述了无人机及其航拍影像的特点和广泛的应用场景,简述了无人机航拍影像目标检测与语义分割方法的发展历程。然后对基于深度学习的无人机航拍影像目标检测与语义分割方法按照不同的网络模型进行分类,分别总结了它们的改进策略、应用场景、贡献和局限性。随后收集梳理了近些年无人机航拍影像的数据集,归纳了常用的卷积神经网络模型的评价指标。最后指出了本领域目前存在的相关问题,并对未来的研究趋势进行了展望。

**关键词:** 无人机航拍影像; 深度学习; 目标检测; 语义分割; 数据集; 性能评价指标

**中图分类号:** V279; TP391

**文献标识码:** A

**文章编号:** 1000-6893(2024)06-028822-30

无人机(Unmanned Aerial Vehicle, UAV)操作简单、灵活机动、性能优良,使用时受场地和环境限制小,适合在复杂的场景中执行任务<sup>[1-2]</sup>。由于无人机拍摄时飞行高度和视角的不同,与水平角度拍摄的自然影像相比,无人机航拍影像包含更加丰富的小目标,且图像内物体排列无序、方向随机、背景复杂。无人机系统便于携带多种类型的传感器进行日常巡检,具有数据采集强度高、运营成本低和便于运输等优点。无人机系统与先进的深度学习方法相结合,使其更加智能、便捷、高效,现已广泛应用于农作物监测<sup>[3-5]</sup>、交通监管<sup>[6]</sup>、城市规划<sup>[7]</sup>、市政管理<sup>[8-9]</sup>、输电线路巡检<sup>[10-11]</sup>、搜索与救援<sup>[12-13]</sup>和国防军事<sup>[14-15]</sup>等领域:①无人机在农业信息化

建设中表现出巨大的发展和应用前景,利用无人机传输的影像信息可以获取农田的相关数据,通过挂载的不同设备可以轻松实现喷洒农药、作物播种、定量施肥等;②无人机的应用丰富了交通管理部门的管理方式,提高了工作效率;利用图像识别技术,可以帮助无人机在飞行过程中捕获非法停车和占道,监测早晚高峰拥堵情况等,便于相关部门及时取证和快速恢复道路畅通;③无人机在电力线巡检、应急响应和输电线路铁塔建设等方面的应用逐步增多;相关工作经目标检测方法的赋能,可以实时检测出存在故障的绝缘子、防振锤和螺钉等目标,帮助工作人员快速了解故障情况;④无人机在现代战争中具有极其重要的作用;自阿富汗战争

收稿日期: 2023-04-06; 退修日期: 2023-05-04; 录用日期: 2023-06-01; 网络出版时间: 2023-06-12 09:47

网络出版地址: <https://hkxb.buaa.edu.cn/CN/Y2024/V45/I6/28822>

基金项目: 国家自然科学基金(61573183)

\* 通信作者: E-mail: nuaaimage@163.com

中无人机的首次使用,到近期俄乌冲突中的突出表现,察打一体无人机在复杂多变的战场上所起的作用不容小觑,其侦查、监测、目标捕获和实时打击的能力,极大地缩短了从发现到摧毁目标的时间,能够应用于多种战争场景。

图像目标检测的方法可以划分为基于传统特征与基于深度学习网络两大类。传统目标检测方法包括特征检测和分割检测2个方向<sup>[16-20]</sup>。其中特征的设计与选择极大地依赖先验条件,其准确性、客观性、鲁棒性和泛化性都受到了不同程度的制约。同时,传统的目标检测方法大多采用滑动窗口策略,导致计算时间长、效率低、处理复杂且准确度低。语义分割不同于目标检测,它对图像中的每一个像素点进行分类,确定其所属的类别,实现区域划分,转换为具有突出显示的感兴趣区域掩膜。传统的语义分割方法包括阈值分割<sup>[21]</sup>、区域分割<sup>[22]</sup>和边缘分割<sup>[23]</sup>等。这些方法常采用硬编码的规则,难以准确表示复杂类别,存在效率和准确性低的问题。随着计算机算力的提升和数据集规模的发展,传统方法已无法满足人们的需求。从2014年开始,以R-CNN<sup>[24]</sup>为代表的深度学习目标检测方法凭借其优秀的检测效果受到了众多研究人员的关注,在2015年,Long等<sup>[25]</sup>提出了第一个深度学习语义分割模型,即全卷积网络(Fully Convolutional Networks, FCN)。基于深度学习的目标检测和语义分割方法具有良好的特征表达能力,与传统方法相比能够更好地处理模糊、复杂和非线性的数据,提取出更多具有辨识性的抽象特征。此外,在拥有良好检测精度的同时大大缩短了时间开销。受无人机飞行高度的影响,无人机航拍影像相比于地面自然影像包含了更多的小目标,它们往往方向随机、分布不均并且目标之间经常排列密集和相互遮挡<sup>[26-27]</sup>。无人机飞行途中,不同的位置信息使得获取的图像背景各异,包含了不同的光照变化、天气条件和地形地势。由于拍摄时机载相机的抖动可能会降低图像的清晰度,捕获的图像中存在大量噪声干扰,使得目标模糊不清。拍摄时视角的改变,使得同一物体也会存在不同的表现形式、尺寸大小和纹理信息。以上特

点使得当前目标检测方法在无人机航拍影像上表现欠佳,原有的基于深度学习的目标检测方法仍存在很大的改进空间。为了提高原有检测器在无人机航拍影像上的检测性能,已提出了大量的改进方法,基于无人机航拍影像的目标检测与语义分割现已成为研究热点之一。

随着无人机和深度学习技术的快速发展,其结合愈发紧密。针对无人机航拍影像中目标的特点和实时性的需求,已提出了众多有针对性的改进方法。近些年,基于无人机航拍影像目标检测与语义分割的相关综述有:Mittal等<sup>[28]</sup>分析了现有的两阶段检测器和单阶段检测器在低空无人机航拍影像数据集上的表现,列出了其目标检测任务中面临的挑战,但文中所提方法仅限于2020年之前且各类中列举的方法数量有待扩充。Diez等<sup>[29]</sup>总结了基于无人机航拍影像中树木检测、树种分类和森林异常等方面的深度学习方法,但囿于单一的林业方向且缺少对方法局限性的讨论。江波等<sup>[1]</sup>按照无人机航拍影像的特点将总结方法所针对的问题分为复杂背景问题、小目标问题、大视场问题和旋转问题4类,但未包含针对上述组合问题的综合改进方法,且迁移学习部分的论述仍需完善;更应注意的是,在此之后尚有大量相关论文发表,该综述内容有待补充和更新。Zhu等<sup>[30]</sup>回顾了ECCV2018和ICCV2019会议中,基于Vis-Drone无人机数据集所提交的目标检测方法,分别从图像目标检测、视频目标检测、单目标跟踪和多目标跟踪4个方面进行了总结。程擎等<sup>[31]</sup>从无人机航拍影像的特点出发,介绍了相关的语义分割方法,包括小目标检测分割、模型实时性和多尺度特征融合等方面,概括了无人机航拍影像语义分割的相关应用,包括线检测、农业和建筑物提取。

本文概述了无人机航拍影像目标检测与语义分割方法的发展历程,综述了不同场景下基于各种网络模型的无人机航拍影像目标检测与语义分割方法,归纳了无人机航拍影像数据集和目标检测与语义分割模型的评价指标,最后从模型和数据集2个维度阐述了本领域当前存在的有关问题,并对下一步的工作进行了展望。

## 1 无人机航拍影像目标检测与语义分割方法发展概况

在2012年之前,局部或全局特征常用来描述图像中的内容。纹理因具有良好的区分性,被广泛应用于无人机航拍影像分类中,有效实现了彩色、多光谱和高光谱图像的特征提取。在此基础上,通过对特征数据的后处理操作进一步提高了分类精度,可以更有效地对图像内容进行语义建模。尽管基于特征的方法性能良好,但忽略了高阶局部特征和它们之间复杂的空间相关性。2014年,R-CNN的出现克服了上述缺点,多层卷积模块的叠加,更有利于从训练数据中学习深层的抽象特征,取得更好的检测效果。因此在无人机航拍影像中基于深度神经网络的检测方法日益受到普遍的认可<sup>[32-35]</sup>。

传统的目标检测方法由区域选择、特征提取和分类器组成。首先,在给定的图像上寻找候选区域,然后在这些区域中提取特征,最后,使用训练好的分类器进行分类。由于目标可能出现在图像中的任意位置,并且目标的大小和长宽比例等事先无法确定,所以需要设置不同尺度的滑动窗口对整幅图像进行遍历。这种穷举的策略虽然包含了目标可能出现的位置,但是存在时间复杂度高、冗余窗口多和区域匹配差等问题,严重影响后续特征提取的速度和效果。实际上,由于受到时间复杂度问题的限制,滑动窗口的尺寸一般是几个固定的预设值,对于长宽比浮动较大的目标,即便是遍历整幅图像也很难获取相匹配的特征区域。在特征提取阶段,常采用局部二值模式(Local Binary Pattern, LBP)<sup>[36]</sup>、尺度不变特征变换(Scale-Invariant Fea-

ture Transform, SIFT)<sup>[16]</sup>和方向梯度直方图(Histogram of Oriented Gradient, HOG)<sup>[17]</sup>等特征。由于目标形态多样、光照变化不确定、目标背景复杂,使得设计具有鲁棒性的特征非常困难。然而,提取特征的质量将直接影响分类的准确性。综上,传统检测方法的效果不稳定,容易受到多种因素的影响,难以投入实际使用。自深度学习技术出现之后,基于卷积神经网络的目标检测方法取得了巨大的突破。其中最受关注的2个方向有:①以R-CNN为代表的基于区域建议的深度学习目标检测方法,例如R-CNN、Fast R-CNN<sup>[37]</sup>、Faster R-CNN<sup>[38]</sup>和Cascade R-CNN<sup>[39]</sup>等。它们是两阶段检测器,需要先使用启发式方法或卷积神经网络(Convolutional Neural Networks, CNN)产生区域建议,然后在该区域内进行分类与回归操作。②以YOLO(You Only Look Once)<sup>[40]</sup>为代表的基于回归思想的深度学习目标检测方法,例如SSD(Single Shot MultiBox Detector)<sup>[41]</sup>、YOLO、RetinaNet<sup>[42]</sup>、EfficientDet<sup>[43]</sup>、CenterNet<sup>[44]</sup>和FCOS(Fully Convolutional One Stage)<sup>[45]</sup>等。它们是单阶段检测器,仅使用一个CNN网络就直接预测不同目标的类别与位置信息。在语义分割方面,FCN在抽象的特征信息中恢复出每个像素的所属类别,即从图像级的分类进一步延伸至像素级。尽管与传统的语义分割方法相比性能有所提升,但仍存在分类结果不够准确,对图像中细节信息不敏感,像素之间相关性考虑不足等问题。在后续的研究中,已提出了多种更加优秀的网络结构,其中常见的语义分割模型包括U-Net<sup>[46]</sup>、SegNet<sup>[47]</sup>和Mask R-CNN<sup>[48]</sup>等。目标检测与语义分割方法的发展历程如图1所示,2008年及以前为传

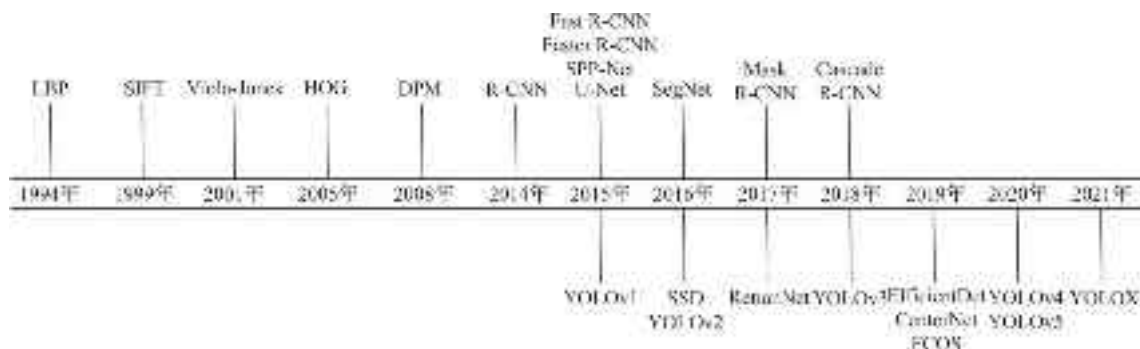


图1 目标检测与语义分割方法的发展历程

Fig. 1 Development history of object detection and semantic segmentation methods



统目标检测方法;2014年及以后为深度学习方法,时间轴的上方为两阶段目标检测与语义分割方法,下方为单阶段目标检测方法。本文将依据不同的网络模型进行归类,阐述它们在无人机航拍影像中的对比、改进和应用。

## 2 无人机航拍影像目标检测的深度学习 方法

### 2.1 无人机航拍影像目标检测的两阶段方法

最早的两阶段检测方法是R-CNN,它由候选区域建议器和CNN组成,奠定了以后的目标检测方法的基础。R-CNN采用选择性搜索算法对输入图像进行分割,获取了大量的目标候选区域。由于需要对所有的候选区域进行特征提取,因此计算复杂且非常耗时。网络中的全连接层需要保证输入的大小统一,图像经过裁剪和缩放等处理会导致畸变,影响检测效果。He等<sup>[49]</sup>提出了SPP-Net网络,在R-CNN的末端添加了空间金字塔池化(Spatial Pyramid Pooling, SPP)模块,避免了对输入大小的限制,使得网络可以输入任意尺寸的图像。该网络通过对整幅图像的处理,获取了固定长度的特征向量,检测速度明显提升。Fast R-CNN借鉴了SPP-Net中金字塔池化的思想,提出了感兴趣区域(Region of Interest, RoI)池化层,并且使用Softmax层替换了SVM用于分类。但是,该算法仍采用选择性搜索策略提取候选区域,过程复杂耗时。Faster R-CNN提出了创新的区域建议网络(Region Proposal Network, RPN),利用与Fast R-CNN共享的卷积层直接获取候选区域,大幅降低了计算成本,提高了检测速度。Cascade R-CNN针对检测过程中交并比(Intersection over Union, IoU)阈值的选取问题进行了改进。它由3个级联的检测头组成,每个检测头在训练和测试时设置了不同大小的IoU阈值,边界框回归在经过3次修正之后,得到了良好的预测效果。

无人机航拍影像目标检测方法的相关研究中,常使用的两阶段检测器有Faster R-CNN和Cascade R-CNN。

#### 2.1.1 无人机航拍影像目标检测的改进Faster R-CNN算法

奚祥书等<sup>[50]</sup>研究了基于多光谱数据的不同降维方法处理对无人机航拍影像中树木冠层检测效果的影响。在实验过程中,通过特征波段选择、特征提取和波段组合的方法生成5种不同的数据集,用于FPN-Faster-R-CNN、YOLOv3和Faster R-CNN检测器的训练。结果表明,若降维后影像中目标物体颜色与背景差异明显且轮廓清晰,会更有利于深度学习网络提取树木冠层的特征,获得更好的预测结果。

为了准确快速地确定害虫在叶片上的啃食区域,以便做出及时的防治。Du等<sup>[51]</sup>在Faster R-CNN的基础上提出了Pest R-CNN检测模型。Pest R-CNN由特征提取模块、区域建议网络和预测头组成。特征提取模块中,增加了输出特征图的数量,增强了多尺度特征融合,扩张了感受野,提高了泛化性。在区域建议网络中,添加了通道注意力模块(Channel Attention Module, CAM)和空间注意力模块(Spatial Attention Module, SAM),用于向害虫啃食位置提供更高的权重,突出该位置的语义信息。

为了提高无人机图像中绝缘子缺陷识别的检测精度,Tang等<sup>[52]</sup>提出了一种改进的基于Faster R-CNN的检测方法。进行了如下3处改进:①使用ResNet-50网络作为模型主干;②使用FPN网络融合多层特征;③使用RoIAlign方法对RoI池化层进行替换。这些改进有利于缓解梯度消失或梯度爆炸的问题,使浅层网络中的位置信息和深层网络中的语义信息更好地融合,解决了RoI池化层两次量化所引入的误差。

由于无人机图像中车辆等目标尺寸普遍较小,在经过多层的卷积神经网络之后容易导致细节信息严重丢失。针对上述问题,Wang等<sup>[53]</sup>改进了Faster R-CNN网络结构,使其能够更好地提取小目标的详细信息。首先,增加了主干输出特征图的数量,补充了小目标的纹理特征,然后,根据训练图像中待检测目标的直方图分布,增加了锚框数量,调整了锚框参数。

在分析无人机航拍影像时,由于无人机飞行

时相机会抖动,影响影像的质量。常见的解决方案是在不同尺度上分析输入图像,以获得尽可能多的信息来正确地检测和跟踪视频序列中的目标。Avola等<sup>[54]</sup>提出了MS-Faster R-CNN网络,这是一种多数据流(Multi-Stream,MS)结构。首先,利用MS卷积神经网络从给定的某帧画面中获取目标的多尺度特征,然后,通过Faster R-CNN提取特征映射,预测目标的边界框,最后,使用Deep Sort跟踪算法实现整个视频中目标的跟踪。

Ding等<sup>[55]</sup>提出了一种在无人机航拍影像中针对倒塌建筑物的目标检测方法。通常情况下倒塌建筑物分布在边界模糊且没有特定几何形状的大范围内,导致标记数据的主观性影响很大。尽管非极大值抑制(Non-Maximum Suppression,NMS)方法在后处理过程中消除了大多数检测到的重叠预测框,但是不规则的目标形状会造成预测框之间相互嵌套。此外,以倒塌的建筑物为目标的检测结果难以评价,可能出现一个

真实框对应多个预测框,或者多个真实框对应一个预测框的问题。因此在Faster R-CNN的框架上引入了可形变卷积(Deformable Convolution,DConv),用于学习不规则几何特征的相关信息,提高对任意形状倒塌建筑物的适应性。提出了目标相交比(Intersected Proportion of Objects,IPO)替代传统的IoU指标,用于衡量预测框之间的重合程度,并作为NMS的判断依据。

大型起重设备表面的金属结构复杂,并且不同位置之间可能存在较大的光照差异。为了实现这类设备表面裂纹的准确检测,Zhou等<sup>[56]</sup>提出了一种适用于无人机航拍影像的检测方法。该方法采用Faster R-CNN对裂纹区域进行检测,通过最大熵阈值分割、Canny边缘检测算子和投影特征提取等方法识别裂纹的长度、宽度、面积和纵横比等参数。

表1总结了上述不同场景下改进的Faster R-CNN无人机航拍影像目标检测方法。

表1 不同场景下改进的Faster R-CNN无人机航拍影像目标检测方法

Table 1 Improved Faster R-CNN object detection method for UAV aerial images in different scenarios

文献	改进策略	应用场景	贡献	局限性
[51]	增加主干网络输出、增强特征融合、添加注意力模块	无人机航拍影像中害虫在叶片上啃食区域的检测	增强了模型特征提取与特征融合的能力	模型的特征融合过程更加复杂,所添加的注意力模块增加了模型参数
[52]	更换主干网络、增强特征融合、修改池化层	无人机电力线巡检中绝缘子的缺陷检测	有利于缓解梯度消失和梯度爆炸的问题,有利于多尺度特征融合	添加的模块增加了模型参数
[53]	增加主干网络输出、数据集预处理	无人机航拍影像中车辆的检测	有利于小目标的特征提取	特征融合部分结构复杂,模型预处理过程繁琐
[55]	添加DConv卷积操作、优化NMS方法	无人机航拍影像中倒塌建筑物的检测	有利于学习不规则几何特征的相关信息,提高对任意形状倒塌建筑物的适应性	可形变卷积增加了模型的计算量

## 2.1.2 无人机航拍影像目标检测的改进Cascade R-CNN算法

无人机航拍影像中的目标小而密集,场景复杂,这些一直都是目标检测的难点。Huang等<sup>[57]</sup>针对上述问题,提出了一种基于Cascade R-CNN的目标检测方法。该方法针对不同的目标类别对检测头进行了细分,使其可以更好地提取边缘帧并对边缘帧进行精确地调整,提供更准确的感兴趣区域,提高检测结果的可靠性。

为了准确定位和识别高压输电线路中防振

锤等部件的缺陷,Bao等<sup>[58]</sup>提出了一种基于Cascade R-CNN的防振锤缺陷检测方法。首先,通过对数据集进行预处理,提高网络的泛化性,然后,使用ResNeXt-101网络作为主干,加入FPN模块进行多尺度特征融合,最后,使用Focal Loss损失函数改善RPN模块的分类损失,解决正负样本不平衡的问题。

由于无人机飞行高度和姿态的变化,无人机图像中目标尺度存在较大的差异,这给目标检测带来了很大的挑战。针对上述问题Lin等<sup>[59]</sup>提出了ECascade-RCNN目标检测网络。该网络由

Trident-FPN主干、RPN和级联双头检测器组成。依据无人机航拍影像中目标尺寸的分布特点,重新聚类RPN中的锚框,获得更精细化的参数。

李鑫等<sup>[60]</sup>提出了一种改进的 Cascade R-CNN多目标检测方法,用于解决无人机航拍影像中小目标、物体遮挡和正负样本不平衡的问题。以 ResNet-101 网络为主干,设计了一种新的特征金字塔网络。在后处理过程中使用 Soft-NMS 方法抑制冗余预测框,缓解了目标遮挡的问题。

在计算目标回归的损失时,利用 Focal Loss 思想,提升了模型的泛化性。针对无人机航拍影像中目标尺度变化大、分布差异明显、背景复杂和存在大量遮挡等特点,张瑞倩等<sup>[61]</sup>在 Cascade R-CNN 和 Faster R-CNN 算法的基础上,添加了多尺度空洞卷积,用于扩大模型的感受野,提升在复杂背景下的检测精度。

表 2 总结了上述不同场景下改进的 Cascade R-CNN 无人机航拍影像目标检测方法。

表 2 不同场景下改进的 Cascade R-CNN 无人机航拍影像目标检测方法

Table 2 Improved Cascade R-CNN object detection method for UAV aerial images in different scenarios

文献	改进策略	应用场景	贡献	局限性
[57]	改进检测头	无人机航拍影像中密集小目标的检测	有利于更好地提取边缘帧并对边缘帧进行精确地调整,提供更准确的感兴趣区域	模型的检测头结构复杂
[58]	数据集预处理、更换主干网络、修改损失函数	无人机电力线巡检中防震锤等部件缺陷的检测	提高了模型特征提取与特征融合的能力	模型预处理过程复杂耗时
[59]	优化主干网络、调整锚框参数	无人机航拍影像中多尺度目标的检测	提高了模型对不同尺度目标的检测能力	模型的特征融合部分结构复杂
[60]	更换主干网络、增强特征融合、修改 NMS 方法、修改损失函数	无人机航拍影像中多目标的检测	缓解了无人机图像中小目标、物体遮挡和正负样本不平衡的问题	增加了模型主干的复杂度
[61]	优化主干网络、增大模型感受野	无人机航拍影像中多目标的检测	扩大了模型的感受野,提升了复杂背景下的检测精度	空洞卷积的结果缺乏连续性

本节主要介绍了基于 Faster R-CNN 和 Cascade R-CNN 的无人机航拍影像目标检测方法。这些方法的改进策略大多集中在模型的主干和颈部,其中也包括一些针对激活函数和损失函数的改进。改进的目的是为了加快模型的运算速度,同时保持良好的检测精度。经过优化后的检测器,在无人机航拍影像上有着更加优秀的表现。

## 2.2 无人机航拍影像目标检测的单阶段方法

随着目标检测技术在工业中的应用越来越多,效率与实时性的需求不断提高。SSD 在一定程度上克服了之前 YOLO 算法难以检测小目标和定位不准确的问题。它使用卷积层在不同尺度的特征图中获得运算结果,受 Faster R-CNN 中锚框的启发,预设了不同长宽比的先验框,降低了训练难度。需要注意的是,SSD 的一些参数无法在训练过程中学习得到,必须人工设置。此

外,SSD 虽然采用了特征金字塔的层次结构,但仍难以处理大尺度的变化,在检测小物体时尤其明显。YOLOv2<sup>[62]</sup>与 YOLO 相比,使用了新的分类网络,借鉴了 Network in Network<sup>[63]</sup>的思路,  $3 \times 3$  与  $1 \times 1$  的卷积操作交替进行,以压缩特征图的通道数,使用批归一化策略加快了模型的收敛速度,保留了跳跃连接路径用于存储浅层的网络信息。与 YOLO 相比,YOLOv2 的检测精度虽然提升了许多,但是难以投入实际使用。He 等提出了 Focal Loss 损失函数,通过控制正负样本和难易分类样本之间的权重来解决单阶段目标检测器中的样本不平衡问题。设计了 RetinaNet 检测器验证有效性,它的结构简单,但生成预测框数量非常多,通常需要额外的阶段来完成图像分类任务。YOLOv3<sup>[64]</sup>的主干为 DarkNet53,使用的残差结构有效地增加了网络深度并缓解了梯度消失的问题,在颈部融合了不同尺度的特征图,它们分别位于主干中间层、中下层和底层,这



有利于更好地捕获目标的浅层位置信息和深层语义信息,提高检测精度。YOLOv3使用Leaky-ReLU激活函数,解决了Dead ReLU问题,但是该激活函数在深层网络中性能容易下降。EfficientDet采用EfficientNet为主干,加权双向特征金字塔网络(Bidirectional Feature Pyramid Network, BiFPN)为颈部,按照模型尺寸从小到大划分为D0~D7,共计8个版本。模型通过增加输入图片的分辨率和深层的网络结构获取了更抽象的特征信息,但是在训练和预测时,需要花费更多的时间来完成图像的推理工作。基于锚框的检测器的性能非常容易受到锚框大小、纵横比和数量等因素的影响。在处理小目标时,锚框往往还需要依据对象的特征而重新设定,这会阻碍模型的泛化能力。并且,锚框也会涉及一些复杂的计算,例如计算真实框与预测框之间的IoU值等。针对上述问题,提出了CenterNet和FCOS检测模型。CenterNet采用关键点估计来寻找物体的中心点,并通过回归的方法获取物体的其他属性。FCOS采用特征金字塔网络处理多尺度目标的边界框,使用中心度分支弥补预测像素点与对应边界框中心点的误差。此外,FCOS可以被拓展应用到两阶段检测方法中,例如作为Faster R-CNN的区域建议网络取得了良好的性能提升。YOLOv4<sup>[65]</sup>、YOLOv5和YOLOX<sup>[66]</sup>是YOLOv3的后续版本,改进主要分布在模型的主干、颈部和头部,整体在运算速度与检测精度方面均有所提升,并且对同一模型按照总参数量的大小进行了细分,使其可以更好地应用于不同的需求场景。

在无人机航拍影像目标检测方法的研究中,常使用的单阶段检测器有SSD、YOLOv2、RetinaNet、YOLOv3、EfficientDet、CenterNet、FCOS、YOLOv4、YOLOv5和YOLOX等。

### 2.2.1 无人机航拍影像目标检测的改进SSD算法

Rampriya等<sup>[67]</sup>利用无人机航拍影像制作了铁路障碍物检测数据集,包含了6种铁路上常见的障碍物类别。利用该数据集对CenterNet、SSD、Faster R-CNN和YOLOv3检测器进行了评估。

此外,对比了EfficientNet、MobileNet V2和ResNet-50 3种不同主干SSD模型的检测效果。实验结果表明,基于MobileNet V2的SSD模型在检测精度和召回率等方面表现最佳。

为了提高无人机航拍影像中小目标检测的准确性,Liu等<sup>[68]</sup>提出了CBSSD目标检测方法。CBSSD在VGG-16的基础上,增加了ResNet-50网络作为辅助主干,提高了特征提取能力,有利于保留更加丰富的语义信息。CBSSD模型拥有更高的识别率和更低的误检率,在低光照的条件下仍能保持良好的检测效果。裴伟等<sup>[69]</sup>为提高模型对小目标的检测精度,减少漏检与重复检测,提出了R-SSD和CI-SSD目标检测方法。R-SSD使用ResNet-50替换了原先的VGG-16网络,提高了特征提取能力,降低了训练难度。CI-SSD利用空洞卷积和反卷积操作扩大了模型的感受野,融合了不同层间的特征信息,增强了模型的泛化性。无人机捕获的航拍影像分辨率高,其中车辆等小尺寸目标特征点分布较为稀疏。针对该问题,李旭等<sup>[70]</sup>提出了一种基于目标空间分布特征的无人机航拍影像车辆检测网络DF-Net。DF-Net由E-SSD检测器和航拍车辆密度估计网络组成。E-SSD以SSD为基础,调整了锚框参数,使用深度可分离卷积替换了原先的 $3\times 3$ 卷积,并且在特征图下采样操作前添加了抗混叠低通滤波器。航拍车辆密度估计网络基于条件生成对抗网络的思想,用于获取车辆准确的分布特征,便于后续的图像切分处理。

无人机视频车辆检测容易受到视频质量、天气环境等客观因素的影响,导致检测结果较差。针对该问题,Wang等<sup>[71]</sup>提出了一种基于深度学习的无人机航拍影像车辆检测方法。首先,对原始采集到的图像进行HSV(Hue、Saturation、Value)空间亮度平移,以增强对不同光照条件和样本多样性的适应能力,然后,以SSD检测器为基础,引入焦点损失进行优化,用于提高模型的检测精度。

表3总结了上述不同场景下改进的SSD无人机航拍影像目标检测方法。

表 3 不同场景下改进的 SSD 无人机航拍影像目标检测方法

Table 3 Improved SSD object detection method for UAV aerial images in different scenarios

文献	改进策略	应用场景	贡献	局限性
[68]	优化主干网络	无人机航拍影像中小目标的检测	提高了特征提取能力,保留了更加丰富的语义信息	并行的模型主干结构复杂
[69]	更换主干网络、增大模型感受野、增强特征融合	无人机航拍影像中小目标的检测	提高了主干的特征提取能力,降低了网络的训练难度,增强了模型的泛化性	空洞卷积的结果缺乏连续性
[70]	调整锚框参数、优化主干网络	无人机航拍影像中车辆的检测	减少了模型参数,提升了检测速度,有利于获取车辆准确的分布特征	使用的深度可分离卷积会降低模型的检测精度
[71]	数据集预处理、修改损失函数	无人机航拍影像中车辆的检测	增强了对不同光照条件和样本多样性的适应能力	图像预处理过程复杂、耗时

### 2.2.2 无人机航拍影像目标检测的改进 YOLOv2 算法

Jawaharlalnehru 等<sup>[72]</sup>针对无人机航拍影像目标检测中多尺度目标定位精度低、检测速度慢和目标漏检等问题,提出了一种改进的 YOLOv2 算法。为了使锚框参数适应具体的检测任务,重新对自制的航拍检测数据集进行了聚类。在网络的训练过程中,每间隔 10 个轮次改变模型输入

的大小,提高了对不同尺度图像的鲁棒性。

Javed 等<sup>[73]</sup>利用通道剪枝和深度可分离卷积等压缩技术改进了 Tiny-YOLOv2 检测器,提出了 QuantYOLO 目标检测方法,适用于资源和功率受限的无人机部署平台。QuantYOLO 很好地平衡了吞吐量和检测精度之间的关系,实现了低功耗和实时的目标检测。

表 4 总结了上述不同场景下改进的 YOLOv2 无人机航拍影像目标检测方法。

表 4 不同场景下改进的 YOLOv2 无人机航拍影像目标检测方法

Table 4 Improved YOLOv2 object detection method for UAV aerial images in different scenarios

文献	改进策略	应用场景	贡献	局限性
[72]	调整锚框参数、调整图像输入批量大小	无人机航拍影像中多尺度目标的检测	提高了对不同尺度图像的鲁棒性	模型训练过程复杂,需要花费更多的时间
[73]	优化主干网络	资源和功率受限的无人机部署平台	降低了模型参数,提高了运行速度	使用的深度可分离卷积会降低模型的检测精度

### 2.2.3 无人机航拍影像目标检测的改进 RetinaNet 算法

Liu 等<sup>[74]</sup>提出了一种基于 RetinaNet 的目标检测模型,用于无人机电力线巡检中相关目标的检测。该模型采用 K-Means++ 算法在目标数据集上重新聚类,获取了更加匹配的锚框。使用 DenseNet-121 网络作为模型的主干,有助于提高检测精度,使得模型更加轻量化。

Purcell 等<sup>[75]</sup>研究了 RetinaNet 算法在无人机视频片段中检测鲨鱼种类的能力。分别使用了 ResNet-50 和 MobileNet V1 网络作为模型主

干。其中,MobileNet V1 网络在处理小目标时性能会下降,因此在部署前需要仔细对其参数进行调整。为确保检测结果可靠,建议最佳的无人机飞行高度应保持在 25~50 m。Takaya 等<sup>[76]</sup>研究了使用 RetinaNet 网络检测海滩上海洋废弃物的可行性。首先,使用无人机在目标区域按照指定高度采集航拍影像,然后,对采集到的图片进行标注,划分为人为海洋碎片和自然物体,最后,评估 RetinaNet 在其上的表现。实验结果证实了方法的可行性,但是难以检测塑料制品碎片这样的小目标,并且模型灵敏度低,存在过多的假阴性。



李洪瑶等<sup>[77]</sup>提出了一种多无人协同的目标检测方法。该方法在单架无人机上利用改进的RetinaNet网络检测特定目标,然后对多架无人机的预测结果进行拼接以完成大规模巡查任务。改进包括如下3点:①增加了锚框数量并调整了

宽高比;②优化了正负样本的阈值区间和损失函数;③利用迁移学习的策略,扩充了用于训练的图片数量。

表5总结了上述不同场景下改进的RetinaNet无人机航拍影像目标检测方法。

表5 不同场景下改进的RetinaNet无人机航拍影像目标检测方法

Table 5 Improved RetinaNet object detection method for UAV aerial images in different scenarios

文献	改进策略	应用场景	贡献	局限性
[74]	调整锚框参数、更换主干网络	无人机电力线巡检中相关目标的检测	获取了更加匹配的锚框	模型主干网络结构冗余
[75]	更换主干网络	无人机航拍影像中海洋生物的检测	降低了模型参数,提高了运行速度	在处理小目标时性能会下降
[77]	调整锚框参数、修改损失函数、数据集预处理	多无人协同的目标检测方法	使用了更加匹配的锚框,提高了小目标的召回率	预测结果的拼接过程复杂、耗时

#### 2.2.4 无人机航拍影像目标检测的改进YOLOv3算法

Cheng等<sup>[78]</sup>基于无人机航拍影像,对Faster R-CNN和YOLOv3目标检测算法进行了对比,在相同的数据集和硬件环境下,根据比较平均准确率和检测时间来帮助选择符合实际需求的算法。如果用户需要更高的检测精度,Faster R-CNN更加适合;如果对检测速度的要求更加紧迫,那么在保证一定准确率的前提下,YOLOv3是更好的选择。

Junos等<sup>[79]</sup>以无人机航拍影像中棕榈果为目标提出了YOLO-P检测模型。为了提高模型的鲁棒性,使用亮度调整、旋转形变和模糊处理等方法对训练图片进行了数据增强,用于模拟实际的自然环境情况。YOLO-P在YOLOv3-Tiny的基础上,在主干中使用了密集卷积网络、Swish激活函数、增加了小目标检测层并且依据目标数据集利用K-Means算法重新确定锚框的大小。

针对无人机电力线巡检过程中绝缘子缺陷检测准确率低的问题,Yang等<sup>[80]</sup>提出了一种改进的YOLOv3网络。为了提高对小目标的识别精度,将YOLOv3中原先单向信息流的特征金字塔结构改进为双向融合网络。为了使预测框更加逼近真实框,引入了高效交并比(Efficient Intersection over Union,EIOU)损失。针对检测精度、运行时间和存储空间之间的平衡问题,Liu

等<sup>[11]</sup>提出了MTI-YOLO目标检测方法。该方法针对无人机电力线巡检中绝缘子等目标,在YOLOv3-Tiny的基础上添加了多尺度特征检测头、多尺度特征融合结构和SPP模块。以上改进,有助于获取关于绝缘子等目标更加有针对性的语义信息,提高了对不同尺寸绝缘子的检测精度,改善了特征的表达效果。

Zhu等<sup>[81]</sup>提出了利用无人机进行路面破损的检测方法,搭建了图像采集平台,并对无人机的飞行设置进行了研究。针对6种道路病害问题,采集了相关的航拍影像,制作了路面航拍图像数据集,对比了Faster R-CNN、YOLOv3和YOLOv4等目标检测算法对路面破损的分类和定位效果。实验结果表明,YOLOv3对阴影、树木和路面标记等不同环境下的裂缝检测具有较强的鲁棒性。

Sahin等<sup>[82]</sup>研究了YOLO网络架构的改变对无人机航拍影像中小目标检测效果的影响,并在YOLOv3的基础上进行了改进,提出了YOLO-Drone目标检测方法。该方法将原先3个不同尺度的输出层进行了扩充,增加至5个。这有助于获取更多的位置信息,提高对小目标的定位效果,加强模型多尺度特征融合的能力。

刘芳等<sup>[83]</sup>提出了一种基于YOLOv3的无人机航拍影像目标检测方法,用于解决无人机航拍影像中小目标检测困难的问题。首先,利用多尺度卷积(Mutil-Scale Convolution,MSConv)替换

标准卷积,构建了自适应特征提取网络MSDarkNet-53,然后,在主干的残差结构中添加了卷积块注意力模块(Convolutional Block Attention Module,CBAM)来获取空间和通道维度中的特征权重,最后,扩展了主干的输出,并将所有的特征图调整至相同大小,用于对小目标的精细检测。针对无人机航拍影像中小目标漏检率和误

检率高的问题,蒲良等<sup>[84]</sup>提出了一种基于YOLOv3-Tiny的目标检测方法。该方法在主干中的不同位置添加了 $1\times 1$ 和 $3\times 3$ 的卷积层,增加了主干输出特征图的数量,并对K-Means聚类算法进行了优化。

表6总结了上述不同场景下改进的YOLOv3无人机航拍影像目标检测方法。

表6 不同场景下改进的YOLOv3无人机航拍影像目标检测方法

Table 6 Improved YOLOv3 object detection method for UAV aerial images in different scenarios

文献	改进策略	应用场景	贡献	局限性
[79]	数据集预处理、优化主干网络	无人机航拍影像中棕榈果的检测	提高了模型特征提取能力	模型预处理过程复杂、耗时
[80]	增强特征融合、修改损失函数	无人机电力线巡检中绝缘子的检测	提高了对小目标的识别精度	使用的损失函数计算复杂
[11]	增加主干网络输出、增强特征融合、增大模型感受野	无人机电力线巡检中绝缘子的检测	增强了网络的特征融合能力	模型的颈部结构复杂
[82]	增加主干网络输出、增强特征融合	无人机航拍影像中小目标的检测	优化了对小目标的定位效果,加强了模型多尺度特征融合的能力	模型的颈部结构复杂
[83]	优化主干网络、添加注意力模块、增加主干网络输出	无人机航拍影像中小目标的检测	提高了多尺度特征的表达能力,增强了模型特征融合的能力	注意力模块的添加增加了模型参数
[84]	优化主干网络、增加主干网络输出、调整锚框参数	无人机航拍影像中小目标的检测	提高了模型特征提取能力	模型的颈部结构复杂

#### 2.2.5 无人机航拍影像目标检测的改进 EfficientDet算法

无人机航拍影像的小目标检测在实际应用中是一项具有挑战性的任务,检测结果容易受复杂背景、高密度区域和随机纹理等因素的干扰。Tseng等<sup>[85]</sup>对比了HOG-SVM、EfficientDet和Faster R-CNN在无人机航拍影像中水稻幼苗的检测效果。实验结果表明,使用支持向量机分类的模型,它的计算复杂度会随样本数量的大小呈指数增长,在实际应用中存在一定的缺陷。基于卷积神经网络的模型具有更好的泛化性和良好的实时性,在不同成像条件的测试集上,EfficientDet拥有最佳的检测精度。

Aldahoul等<sup>[86]</sup>建立了一个基于EfficientDet-D7用于人体检测和人体活动识别的系统。首先,对比了EfficientDet、YOLO和Faster R-CNN在COCO数据集上训练用于人体检测的效果,然后,对比了不同版本的EfficientDet,如D0、D4和D7,在UCF-ARG无人机航拍影像数据集中对挥

手、行走和奔跑等行为的识别能力,最后,通过添加水平翻转、模糊处理、高斯噪声和明暗变化等数据增强方法,验证了EfficientDet-D7的鲁棒性。由于无人机高空成像的原因,在航拍影像中可能仅有 $0.1\%\sim 0.2\%$ 的部分为待检测的目标。Dousai等<sup>[87]</sup>针对感兴趣区域覆盖率低的问题,提出了一种基于EfficientDet结构的目标检测模型。该模型的颈部由BiFPN和全连接特征金字塔网络(Fully Connected Feature Pyramid Network,FC-FPN)级联组成,取得了在HERIDAL数据集上以人类为目标的最高准确率结果。

表7总结了上述不同场景下改进的EfficientDet无人机航拍影像目标检测方法。

#### 2.2.6 无人机航拍影像目标检测的改进 CenterNet算法

Xia等<sup>[88]</sup>提出了一种基于MobileNet-CenterNet的绝缘子检测模型,用于解决无人机电力线巡检过程中绝缘子等目标检测精度低、实时性差等问题。该模型使用MobileNet V1轻量化网络作为

表 7 不同场景下改进的 EfficientDet 无人机航拍影像目标检测方法

Table 7 Improved EfficientDet object detection method for UAV aerial images in different scenarios

文献	改进策略	应用场景	贡献	局限性
[86]	数据集预处理	无人机航拍影像中人体检测和人体活动识别	对比了不同模型之间的检测性能,对比了不同版本 EfficientDet 的检测性能	模型预处理过程复杂、耗时
[87]	优化主干网络、增强特征融合	无人机航拍影像中小目标的检测	提高了小目标的检测性能	级联的结构增加了模型的复杂度

主干,在不同位置添加了CBAM注意力模块,利用DIOU-NMS(Distance-Intersection over Union-NMS,DIOU-NMS)抑制冗余预测框,并增加了3个基于转置卷积的上采样层。

与自然影像相比,无人机航拍影像中目标类型分布不均,并且物体的方向、形状和比例差异更大。针对上述问题,Albaba等<sup>[89]</sup>提出了SyNet目标检测网络。SyNet利用图像增强技术处理样本类别不均衡的问题,通过单阶段与多阶段检测方法相互协同的策略,结合CenterNet和Cascade R-CNN检测器,提升了对小目标的预测效果。刘鑫等<sup>[90]</sup>针对在复杂背景下小目标特征信息学习困难的问题,提出了MSA-CenterNet检测方法。MSA-CenterNet在CenterNet的基础上添加了自适应(Modified Self-Adaptive,MSA)

基础模块和升维全局上下文注意力模块(Global Context-Block,GC-Block),并在模型的头部使用了深度可分离卷积和Mish激活函数。这些改进有助于抑制冗余的特征表达,强化关键点的语义信息。由于高分辨率的无人机航拍影像在输入网络时,图像压缩造成了小目标的遗失,不利于目标特征的提取。针对上述问题,王胜科等<sup>[91]</sup>提出了一种基于CenterNet的检测模型。该方法通过剪裁对图像进行预处理,用于获取合适的输入尺寸。在主干中引入了双重注意力机制,以提高对小目标的特征表达能力。使用GIOU(Generalized Intersection over Union)作为NMS的判定依据,加强了模型后处理的能力。

表8总结了上述不同场景下改进的CenterNet无人机航拍影像目标检测方法。

表 8 不同场景下改进的 CenterNet 无人机航拍影像目标检测方法

Table 8 Improved CenterNet object detection method for UAV aerial images in different scenarios

文献	改进策略	应用场景	贡献	局限性
[88]	更换主干网络、添加注意力模块、修改损失函数、增加主干网络输出	无人机电力线巡检中绝缘子的检测	降低了模型参数和计算复杂度,使得网络更加轻量化	转置卷积的添加会延长模型的训练时间
[89]	数据集预处理、优化主干网络	无人机航拍影像中杂乱目标的检测	提升了对小目标的预测效果,并且对不同尺度的目标具有更强的鲁棒性	并行结构使得模型过于复杂
[90]	优化主干网络、添加注意力模块、改进检测头、修改激活函数	无人机航拍影像中小目标的检测	提高了复杂背景下小目标的特征提取能力	添加的模块增加了模型的复杂度
[91]	数据集预处理、添加注意力模块、优化NMS方法	无人机航拍影像中小目标的检测	采用剪裁的方法对图像进行预处理,用于获取合适的输入尺寸	模型预处理过程复杂、耗时

### 2.2.7 无人机航拍影像目标检测的改进FCOS算法

利用无人机进行交通监控具有视野广阔、机动性强、不影响交通等优点,但是飞行高度、拍摄角度和复杂背景等给目标的检测带来了不小的挑战。Akshatha等<sup>[92]</sup>对比了FCOS和YOLOv3在无人机航拍影像中对车辆的检测效果。基于

平均精度的评价指标分析表明,FCOS的表现比YOLOv3更好。Wang等<sup>[93]</sup>利用无人机采集电力线图像,建立了绝缘子缺陷检测数据集。该数据集由14 074幅高分辨率图像和19 113个注释组成。基于该数据集,对比了FCOS、Faster R-CNN和YOLOv4在缺陷检测任务中的性能。此外,给出了一些有助于提高检测效果的建议:①可以利用数据增强的方法弥补训练样本不足的问题;



② 通过对原始图像的剪裁来提高小目标的检测效果;③ 根据数据集中目标的特点对锚框参数进行微调。

由于无人机特殊的高空视角,使得图像中包含了大量的小目标。Gao等<sup>[94]</sup>提出了一种针对无人机图像的单级检测器。它采用了FOCS中无锚框的思想,使得正负样本的判定更加合理;使用匹配分数图策略,有助于充分利用特征图中的相似信息;通过Soft-NMS方法缓解密集排列造成的漏检问题,有利于倾斜目标的检测。

针对无人机输电线路巡检中目标尺度差异大、背景复杂等特点,Zhao等<sup>[95]</sup>提出了一种基于FCOS的目标检测方法。该方法由特征提取网络、特征金字塔网络和每个特征层对应的检测头组成。在特征金字塔网络中,改变了自上而下结

构中的部分卷积连接,增加了自下而上的分支结构,有利于充分利用语义信息和位置信息,提高小目标的检测性能。在检测头通过高斯分布计算符合目标分布的中心权重,对预测结果进行修正,实现了正负样本标签的自适应分配,减小了复杂背景对检测精度的影响。

针对无人机航拍影像噪声干扰强、目标密集排列的特点,张智等<sup>[96]</sup>提出了一种结合多帧检测的车辆检测方法。该方法对FCOS网络中后处理的操作进行了改进,使用Soft-NMS抑制相互重叠的预测框。设计了多帧目标回归模块,用于结合当前帧画面和邻近帧中同一物体的特征表示,突出了目标信息,削弱了背景噪声干扰。

表9总结了上述不同场景下改进的FCOS无人机航拍影像目标检测方法。

表 9 不同场景下改进的FCOS无人机航拍影像目标检测方法

Table 9 Improved FCOS object detection method for UAV aerial images in different scenarios

文献	改进策略	应用场景	贡献	局限性
[94]	采用无锚框思想、增强特征融合、优化NMS方法	无人机航拍影像中小目标的检测	提高了模型针对小目标的检测能力	Soft-NMS中高斯惩罚函数的计算复杂
[95]	增加主干网络输出、改进检测头	无人机输电线路巡检中多尺度目标的检测	有助于充分利用语义信息和位置信息,提高了小目标的检测性能	添加的分支结构增加了模型的复杂度
[96]	优化NMS方法、增强特征融合	无人机航拍影像中车辆的检测	有助于抑制相互重叠的预测框,提高了模型的后处理能力	Soft-NMS中高斯惩罚函数计算复杂

#### 2.2.8 无人机航拍影像目标检测的改进 YOLOv4算法

针对无人机航拍影像背景复杂、目标小并且存在相互遮挡等情况,Tan等<sup>[97]</sup>对YOLOv4进行了改进,提出了YOLOv4\_DRONE检测模型。首先,在主干中添加了感受野模块(Receptive Field Block,RFB),扩张了模型的感受野,然后,在特征金字塔结构中加入超轻量级子空间注意力机制(Ultra-Lightweight Subspace Attention Mechanism,ULSAM),最后,使用Soft-NMS方法缓解因遮挡而造成的目标漏检问题。

Cheng等<sup>[98]</sup>针对无人机航拍过程中因相机抖动造成的影像模糊问题、光照不均导致的曝光问题以及传输过程中产生的噪声问题,提出了一种改进的YOLO模型。该方法采用了仿射变换、高斯模糊处理和灰度变换等多种数据增强方

法,加强了YOLOv4模型的数据预处理能力,有效地缓解了因数据量较少而造成的训练困难问题。

Guo等<sup>[10]</sup>提出了一种基于改进的YOLOv4模型用于输电线路中防振锤故障的检测。由于无人机航拍影像中防振锤这类的小目标,在经过YOLOv4主干网络的多次下采样之后,容易造成小目标特征被遗漏的问题。因此扩展了主干输出的特征层,并添加了RFB模块用于加强网络的特征提取能力。在模型的颈部,融合了多种尺度大小的特征信息,检测精度有了明显提升。Bao等<sup>[99]</sup>建立了基于无人机航拍影像的输电线路异常减震器数据集,其中含了4种不同类型的减震器,存在生锈、破损和正常等不同情况。针对数据集中小目标和复杂背景的特点,设计了PAM-YOLO检测网络。PAM-YOLO以YOLOv4为基础,在模型的颈部添加了平行混合注

意力(Parallel Mixed Attention, PMA)模块,并使用K-Means聚类算法重新求取了更加适合的锚框。

在无人机拍摄的果树冠层图像中,由于存在相互粘连和遮挡等情况,影响了检测器的准确性。为了提高现实场景中果树冠层的识别精度,Zhu等<sup>[100]</sup>提出了一种改进的YOLOv4检测方法。使用MobileNet V3网络作为主干,在模型的不同位置添加了CBAM注意力模块,在颈部的末尾添加了自适应空间特征融合(Adaptively Spatial Feature Fusion, ASFF)模块,并且在训练期间使用余弦退火衰减策略调整学习率的大小。针对YOLOv4算法在无人机航拍影像中对松材线虫病变树木识别效率低的问题,黄丽明等<sup>[101]</sup>对YO-

LOv4的主干和颈部进行了改进。在主干使用了MobileNet V2网络用于减少参数数量,在颈部使用深度可分离卷积替换部分标准卷积并且简化了PANet的结构,提高了识别速度。

为了实现无人机对桥梁裂缝的实时检测,Yu等<sup>[102]</sup>在YOLOv4的基础上提出了YOLOv4-FPM检测模型。YOLOv4-FPM利用Focal Loss的思想对损失函数进行了优化,采用剪枝算法简化了网络结构,使用多尺度数据集训练网络扩展了可预测范围。实验结果表明,改进后的方法在检测精度和模型尺寸方面均取得了提升,可以有效检测不同大小图片中的裂缝目标。

表10总结了上述不同场景下改进的YOLOv4无人机航拍影像目标检测方法。

表 10 不同场景下改进的YOLOv4无人机航拍影像目标检测方法

Table 10 Improved YOLOv4 object detection method for UAV aerial images in different scenarios

文献	改进策略	应用场景	贡献	局限性
[97]	增大模型感受野、添加注意力模块、优化NMS方法	无人机航拍影像中小目标的检测	扩张了模型的感受野,有利于提取不同尺度的特征	空洞卷积的结果缺乏连续性
[98]	数据集预处理	用于处理无人机航拍影像模糊、曝光和噪声的问题	加强了模型的数据预处理能力,有效缓解了因数据量较少而造成的训练困难问题	模型预处理过程复杂、耗时
[10]	增加主干网络输出、扩大模型感受野、增强特征融合	无人机电力线巡检中防振锤故障的检测	有效增加了模型的感受野,融合了多种尺度大小的特征信息	空洞卷积的结果缺乏连续性
[99]	添加注意力模块、调整锚框参数	无人机电力线巡检中目标缺陷的检测	增强了模型的特征提取能力,获取了更加准确的锚框参数	注意力模块的添加增加了模型参数
[100]	更换主干网络、添加注意力模块、增强特征融合、调整学习率参数	无人机航拍影像中果树冠层的检测	有助于网络更好地聚焦于目标特征,增强了网络检测多尺度目标的能力	模型的颈部结构复杂
[101]	更换主干网络、简化颈部网络	无人机航拍影像中松材线虫病变树木的检测	降低了模型参数,提高了模型的检测速度	使用的深度可分离卷积会影响模型的检测精度
[102]	修改损失函数、简化网络结构、数据集预处理	无人机航拍影像中桥梁裂缝的检测	缓解了复杂背景带来的干扰,增强了多尺度目标检测的鲁棒性	剪枝算法会降低模型的检测精度

### 2.2.9 无人机航拍影像目标检测的改进YOLOv5算法

Wu等<sup>[103]</sup>针对无人机航拍影像中小目标比例高和背景干扰多等问题,对YOLOv5算法进行改进与优化,提出了PWR-YOLOv5,用于检测防振锤和绝缘子等部件的腐蚀情况。PWR-YOLOv5引入了加权自适应路径聚合网络(Weight Adaptive Path Aggregation Network, WA-PANet)、金字塔分裂注意力(Pyramid Split Attention, PSA)

模块和感受性特征增强网络(Receptive Feature Enhancement Network, RFENet)。此外,采用EIOU损失函数计算边界框回归的损失,用于提高网络的定位精度和收敛速度。在以往的模型中,常采用特征金字塔结构来改善无人机航拍影像中小目标的漏检问题,但是关于小目标的信息可能已经在特征图中因下采样处理而丢失。针对上述问题,Zhu等<sup>[104]</sup>基于YOLOv5提出了UavTinyDet检测网络。UavTinyDet的颈部为扩展特征金字塔(Expanded Feature Pyramid, EFP)模

块,增加了针对小目标的检测层,保留了充分的位置信息,对小目标的检测更加友好。刘树东等<sup>[105]</sup>提出了一种基于倒置残差注意力的无人机航拍影像小目标检测方法。该方法使用倒置残差 C3(Inverted Residuals, IRC3)模块和倒置残差注意力 C3(Inverted Residuals Attention, IRAC3)模块替换了原先的残差结构,并且添加了多尺度特征融合(Multi-scale Feature Fusion, MFF)模块。以上改进有利于获取丰富的小目标位置信息和深层语义信息,提高了模型的定位效果。无人机航拍影像中小目标实例多、背景复杂等特点导致检测模型特征提取困难,冒国韬等<sup>[106]</sup>提出了一种基于多尺度分割注意力的目标检测方法 MSA-YOLO(Multi-scale Split Attention-YOLO)。其在 YOLOv5 的基础上引入了多尺度分割注意力单元(Multi-scale Split Attention Unit, MSAU)和自适应加权特征融合方法(Adaptive Weighted-feature Fusion, AWF),使得模型可以更具指向性地提取目标区域的关键信息,动态调节各个特征层之间的权重大小。

目前,许多检测方法虽然拥有较高的检测精度,但是其参数多、复杂度高,难以达到实时处理要求。针对上述问题,Li等<sup>[107]</sup>设计了一种满足无人机实时检测需求的轻量级方法。以 YOLOv5s 为基础,添加了挤压和激发(Squeeze-and-Excitation, SE)模块来提高网络的表达能力,删除了部分主干中冗余的  $1 \times 1$  卷积来减少模型的参数数量,在 SPP 模块中增加了一个  $3 \times 3$  的最大池化层来提高模型的接收范围。

针对风力机叶片表面缺陷检测中存在的检测能力不足、模型推理时间长、对小目标和长条形状缺陷识别准确率低的问题,Zhang等<sup>[108]</sup>提出了一种基于 YOLOv5 的无人机航拍影像目标检测方法 SOD-YOLO。首先,对风力机叶片图像进行前景分割和霍夫变换,建立缺陷检测数据集,然后,主干增加了小目标检测层,利用 K-Means 算法对锚框重新聚类,并添加 CBAM 注意力模块,最后,采用通道剪枝算法降低模型尺寸,提高模型的运算速度。在无人机对风力发电机进行巡检过程中,精准定位其桨叶前端是一项具有挑战性的任务,与此同时,受机载模块有限的

算力制约,常见方法的检测效率低下。白健鹏等<sup>[109]</sup>提出了一种基于轻量级的 YOLOv5 风力发电机桨叶检测与定位方法。该方法使用 ShuffleNet V2 轻量化网络作为主干,减少了参数数量。设计了桨叶叶尖空间定位方法,利用检测所得的像素坐标、无人机位置与姿态信息和空间平面的几何关系,对桨叶进行精准定位。

针对路面修复区域检测难度大、效率低的问题,Liu等<sup>[7]</sup>提出了 M-YOLO 目标检测算法。M-YOLO 使用 MobileNet V3 网络替换了 YOLOv5s 的主干,降低了模型尺寸,提高了目标的检测速度。为了防止图像失真和小目标检测精度不足的问题,在模型头部添加了 SPP-Net 网络结构。

针对枯死树木目标小、细节信息不明显的特征,Wang等<sup>[110]</sup>在 YOLOv5 框架的基础上提出了一种新的轻量级检测网络 LDS-YOLO。LDS-YOLO 的特征提取模块重复利用了之前层的特征信息,在小样本数据集上表现出非常好的抗过拟合性。使用了深度可分离卷积减少模型参数,使其便于在智能终端上进行部署。使用了 SoftPool 池化操作对 SPP 模块进行优化,以确保小目标在检测过程中不会被遗漏。Liu等<sup>[111]</sup>分析了基于无人机航拍影像中的玉米雄穗检测面临的挑战,针对玉米雄穗目标小、形态变化大、相互重叠等情况,提出了 YOLOv5-Tassel 目标检测方法。该方法采用 YOLOv5 主干提取雄穗特征,然后使用 BiFPN 特征金字塔结构进行多尺度特征融合。在模型的颈部添加了 SimAM 注意力模块,并且增加了检测头的数量,增强了模型的检测能力。无人机航拍影像中麦穗分布稠密、重叠严重、尺寸小、背景信息复杂,容易导致误检和漏检等情况,从而降低了麦穗的检测精度,无法对其进行准确计数。针对上述问题,鲍文霞等<sup>[112]</sup>提出了一种基于 TPH-YOLO 检测器的麦穗计数方法。TPH-YOLO 以 YOLOv5 为基础,添加了协调注意力(Coordinate Attention, CA)模块,使用了 Transformer 预测头。在训练过程中,采用 Retinex 算法进行图像增强处理,通过迁移学习策略先后在不同数据集上进行训练,提高了模型的泛化能力和检测精度。

表 11 总结了上述不同场景下改进的 YOLOv5 无人机航拍影像目标检测方法。



表 11 不同场景下改进的 YOLOv5 无人机航拍影像目标检测方法

Table 11 Improved YOLOv5 object detection method for UAV aerial images in different scenarios

文献	改进策略	应用场景	贡献	局限性
[103]	增强特征融合、添加注意力模块、修改损失函数	无人机航拍影像中小目标的检测	提高了对不同尺度目标的检测精度,提高了网络的定位精度和收敛速度	使用的损失函数计算复杂
[104]	增加主干网络输出	无人机航拍影像中小目标的检测	提高了模型针对小目标的检测性能	模型颈部结构复杂
[105]	优化主干网络、添加注意力模块、增强特征融合	无人机航拍影像中小目标的检测	有利于获取更加丰富的小目标位置信息和深层的语义信息	模型的预处理过程复杂、耗时,模型主干结构复杂
[106]	添加注意力模块、增强特征融合	无人机航拍影像中小目标的检测	实现了浅层位置信息与深层抽象信息的高效融合	模型颈部结构复杂
[107]	添加注意力模块、优化主干网络、优化颈部网络	一种轻量级无人机航拍影像目标检测方法	提高了模型的特征表达能力,扩大了模型的感受野	添加的注意力模块和池化层增加了模型的复杂度
[108]	数据集预处理、增加主干网络输出、调整锚框参数、添加注意力模块、简化网络结构	无人机航拍影像中风力机叶片缺陷的检测	提高了模型针对小目标的检测性能	使用的 K-Means 算法容易使锚框参数限于局部最优解
[109]	更换主干网络	无人机航拍影像中风力机桨叶的检测与定位	减少参数数量,提升检测速度	模型的检测精度仍需提升
[7]	更换主干网络、优化颈部网络	无人机航拍影像中路面修复区域的检测	减少参数数量,提升检测速度	模型的检测精度仍需提升
[110]	优化主干网络、增强特征融合	无人机航拍影像中枯死树木的检测	减少模型参数,使其便于在智能终端上进行部署	使用的深度可分离卷积会影响模型的检测精度
[111]	增强特征融合、添加注意力模块、增加检测头数量	无人机航拍影像中玉米雄穗的检测	增强了模型多尺度特征融合的能力	模型检测头结构复杂
[112]	添加注意力模块、优化检测头、数据集预处理	无人机航拍影像中麦穗的检测与计数	有利于抑制复杂背景的干扰,提高模型的泛化能力和检测精度	模型的主干结构复杂,预处理过程复杂耗时

### 2.2.10 无人机航拍影像目标检测的改进 YOLOX 算法

Ru 等<sup>[113]</sup>提出了一种适用于无人机航拍影像的轻量化电力线绝缘子检测模型 ECA-YOLOX-Tiny。该检测模型在 YOLOX-Tiny 的主干中添加了高效通道注意力(Efficient Channel Attention, ECA)模块调整了输入图像的分辨率大小,在训练阶段采用了余弦退火算法调整学习率参数。上述改进有利于获取关于小目标更多的特征信息,加快了模型的收敛速度。

无人机航拍影像中的麦穗具有目标小、易遮挡、分布密集和背景复杂等特点,为了快速且准确地对其进行识别,Yao 等<sup>[114]</sup>对 YOLOX-m 进行了优化,设计了一种改进的 Mosaic 数据增强策略。所提方法使用 BiFPN 网络作为模型的颈部,添加了 SE 模块,使用可以学习的权重来调整不同输入特征的重要性,实现更有效的多尺度特征融合。

Hou 等<sup>[115]</sup>提出了一种基于 YOLOX 的无人机

航拍影像山体滑坡检测方法 YOLOX-Pro,提高了针对不同地貌环境下的检测精度。YOLOX-Pro 使用 Focal Loss 损失函数,用于解决大小样本分布不均的问题。对比了 SE 模块、CBAM 模块和 CA 模块在模型中的改善效果。实验结果表明,CA 模块更有助于提高模型的检测精度,加强了特定区域的识别能力。

利用无人机获取的航拍影像检测烟雾目标可以有效地监测早期森林火灾,Zhan 等<sup>[116]</sup>以 YOLOX-l 为基础,提出了 PDAM-STPNet 检测网络。该网络使用并行空间域注意力机制(Parallel Spatial Domain Attention Mechanism, PDAM),利用局部和全局注意力子模块,捕获了图像中的位置和语义信息,使得预测结果更加可靠。在预处理过程中,设计了组件拼接(Component Stitching)数组增强技术,增加了小目标样本的数量,提高了模型的泛化性。

表 12 总结了上述不同场景下改进的 YOLOX 无人机航拍影像目标检测方法。

表 12 不同场景下改进的 YOLOX 无人机航拍影像目标检测方法

Table 12 Improved YOLOX object detection method for UAV aerial images in different scenarios

文献	改进策略	应用场景	贡献	局限性
[113]	添加注意力模块、调整输入图像分辨率大小、调整学习率参数	无人机电力线巡检中绝缘子的检测	增强了模型特征提取能力,提高了绝缘子检测的准确率	扩大输入图像的尺寸会增加模型的训练和预测时间
[114]	数据集预处理、添加注意力模块、增强特征融合	无人机航拍影像中麦穗的检测	有利于保留更多小目标的特征	模型预处理过程复杂、耗时
[115]	修改损失函数、添加注意力模块	无人机航拍影像中山体滑坡的检测	解决大小样本分布不均的问题,加强了特定区域的识别能力	添加的注意力模块增加了模型的复杂度
[116]	优化主干网络、添加注意力模块、数据集预处理	无人机航拍影像中烟雾的检测	增加了小目标样本的数量,提高了预测的可靠性	模型预处理过程复杂、耗时

本节介绍了基于 SSD、YOLOv2、RetinaNet、YOLOv3、EfficientDet、CenterNet、FCOS、YOLOv4、YOLOv5 和 YOLOX 的无人机航拍影像目标检测方法。因为单阶段检测器拥有良好的检测速度,所以相比于两阶段方法更适合应用于无人机航拍影像。上述检测方法经过改进升级后,更加适用于无人机航拍影像中目标的检测,胜任于多种多样的无人机应用场景。

### 3 无人机航拍影像语义分割的深度学习算法

针对 FCN 结果不够准确、图像细节信息不敏感的问题,Badrinarayanan 等<sup>[47]</sup>提出了 SegNet 语义分割模型。它的编码和解码部分分别由 13 个卷积层和 5 个上采样层组成,并且使用池化索引(Pooling Indices)对特征图进行上采样处理,这有利于保留更多的物体轮廓信息,帮助图像重建。但是,对低分辨率的特征图进行去池化处理时,容易忽略邻近像素间的相关性。Ronneberger 等<sup>[46]</sup>提出了 U-Net 语义分割模型,可以划分为特征提取网络与特征融合网络,分别对应为编码器与解码器。编码器用于获取图像中物体对应的语义信息,然后解码器将特征恢复至输入尺寸,得到最终的预测结果。U-Net 采用完全对称的 U 型结构,有助于特征间的相互融合,但是在实际分割中,大物体的边缘信息和小物体本身很容易在深层网络中因下采样操作而丢失。He 等<sup>[48]</sup>提出了 Mask R-CNN 算法,其在 Faster R-CNN 的基础上增加了全卷积神经网络以生成对应的掩膜分支,使用了 RoIAlign 层,很好地保留了特征点准确的空间位置,有助于提升掩膜的精度。

在处理无人机航拍影像中,语义分割得到了广泛的应用,其中常见的包括 SegNet、U-Net 和 Mask R-CNN 算法。

#### 3.1 无人机航拍影像语义分割的改进 SegNet 算法

Zhong 等<sup>[117]</sup>提出了一种基于多尺度特征融合的网络结构——W-SegNet,用于像素级路面破损检测。建立了多场景、多类型的无人机沥青路面破损数据集。W-SegNet 网络由 2 个对称的编码器组成,并且在编码器和解码器之间添加了由一系列上采样、级联和卷积操作组成的特征融合模块。这有助于更好地利用浅层位置信息和深层语义信息,提高特征图的分辨率,有利于检测出裂缝、坑洞和补丁等不同大小的多尺度路面问题。

Qi 等<sup>[118]</sup>提出了一种基于 SegNet 网络的果树冠层图像分割方法,用于从无人机航拍影像中快速提取冠层信息。所提方法使用了 Elu 激活函数和 Adam 优化器,通过批归一化处理、随机化数据集序列输入和 Dropout 层提高了模型的抗过拟合性。与 SegNet 相比,训练过程更稳定、收敛速度更快、鲁棒性更强、在不同光照条件下识别目标轮廓更精细。Li 等<sup>[119]</sup>针对几种波段的组合,研究了基于无人机三通道彩色图像的多光谱波段信息对向日葵倒伏状态特征提取的影响。改进了 SegNet 网络的输入,使其适用于多波段图像的处理。实验结果表明,添加近红外波段有助于提高分类精度,而加入红外波段则会造成精度的负增长。Li 等<sup>[120]</sup>提出了 CD-SegNet 网络模型,用于无人机航拍影像中棉花的像素分割。CD-SegNet 减少了 SegNet 网络中编码块和解码块的

卷积数量,降低了模型复杂度,提高了分割效率。引入了空洞卷积,在保持特征空间分辨率的同时扩大了模型的感受野,减少了因上采样而造成的

信息丢失。

表 13 总结了上述不同场景下改进的 SegNet 无人机航拍影像语义分割方法。

表 13 不同场景下改进的 SegNet 无人机航拍影像语义分割方法

Table 13 Improved SegNet semantic segmentation method for UAV aerial images in different scenarios

文献	改进策略	应用场景	贡献	局限性
[117]	增强特征融合	无人机航拍影像中路 面破损的检测	有助于更好地利用浅层位置信 息和深层语义信息,提高了特征 图的分辨率	编码器和解码器结构复杂
[118]	修改激活函数、修改模型优化器、 添加正则化操作	无人机航拍影像中果 树冠层的检测	训练过程更稳定,收敛速度更 快,鲁棒性更强,在不同光照条 件下识别目标轮廓更精细	Adam 自适应优化器在深层网 络中存在性能退化的问题
[119]	改进网络的输入	无人机航拍影像中向 日葵倒伏状态的检测	改进后的模型适用于多波段图 像的处理	添加的近红外波段输入会增加 模型的运算量
[120]	降低模型复杂度、扩大模型感受野	无人机航拍影像中棉 花的检测	降低了模型复杂度,提高了分割 效率	空洞卷积的结果缺乏连续性

3.2 无人机航拍影像语义分割的改进 U-Net 算法

Jeon 等<sup>[121]</sup>评估了利用无人机航拍影像和深度学习技术来检测海洋生态系统中海草分布情况的可行性。在 U-Net 模型中添加了 Z-Score 和 Min-Max 归一化方法,以验证不同归一化策略对模型的影响。实验结果表明,Z-Score 和 Min-Max 分别在黑白图像和光学图像上表现出优异的性能。对比了 U-Net、SegNet、PSPNet 和 Deeplab V3+ 网络在无人机航拍影像上的检测效果。其中,结构相对简单、参数较少的 U-Net 模型表现最佳。

Huang 等<sup>[122]</sup>针对高原山区烟叶种植分散、分布不均等特点,提出了一种基于深层语义分割模型的无人机航拍影像烟草种植区域提取方法。首先,利用无人机捕获图片,制作了烟草语义分割数据集,然后,使用 MobileNet 系列轻量级网络替换 DeeplabV3+、PSPNet、SegNet 和 U-Net 模型的编码器部分,用于加快模型训练。实验结果表明,U-Net 在小样本数据集下的烟草语义分割性能优于其他模型,并且对硬件要求不高,便于在实际应用中广泛推广。Narvaria 等<sup>[123]</sup>应用 U-Net 卷积神经网络结构为小麦、棉花、玉米、草和土壤等不同作物分类寻找最佳的输入光谱波段组合。分析了增强植被指数(Enhanced Vegetation Index,EVI)和标准化植被指数(Normalized

Difference Vegetation Index,NDVI)在作物分类中的作用。研究了各纹理参数,例如角二阶矩(Angular Second Moment,ASM)、逆差距(Inverse Difference Moment,IDM)、熵(Entropy,ENT)和对比度(Contrast,CON),在语义分割中起到的效果。通过对光谱植被指数和纹理特征的各种组合进行测试与对比,最大限度地提高了模型的精度。Zhai 等<sup>[124]</sup>提出了一种基于无人机航拍影像和语义分割算法的棉花田地残留种植地膜检测方法。该方法优化了 U-Net 的网络结构,减少了其中的卷积运算,加快了模型的运行时间。在下采样层中,使用 Inception 模块替换了普通的 3×3 卷积,在增加网络深度和宽度的同时,还解决了参数过多的问题,减少了模型的计算量。准确提取小麦倒伏面积可以为灾后产量损失评估和抗倒伏小麦育种提供重要的技术支持。目前,小麦倒伏评估方法面临着实时性和准确性之间的矛盾,同时也缺乏有效的倒伏特征提取方法。针对上述问题 Yu 等<sup>[125]</sup>基于 U-Net 网络提出了一种改进的分割模型。为了获得更大的感受野,保留更多的语义信息,在保持网络结构的基础上,使用对合算子代替主干中的卷积运算。在编码阶段,加入了密集连接模块,有利于减少参数,节省内存开销,缓解了梯度消失的问题。

表 14 总结了上述不同场景下改进的 U-Net 无人机航拍影像语义分割方法。



表 14 不同场景下改进的 U-Net 无人机航拍影像语义分割方法

Table 14 Improved U-Net semantic segmentation method for UAV aerial images in different scenarios

文献	改进策略	应用场景	贡献	局限性
[121]	更换主干网络、数据集预处理	无人机航拍影像中海草分布的检测	验证了不同的归一化策略对模型的影响	增加了模型预处理的复杂度
[123]	对比不同各纹理参数在语义分割中的效果	无人机航拍影像中不同作物的检测	寻找最佳的输入光谱波段组合	增加的纹理参数会提升模型的计算复杂度
[124]	优化主干网络、增强特征融合	无人机航拍影像中棉花田地残留种植地膜的检测	降低了模型主干参数,提高了模型的检测速度	使用的 Inception 模块会降低模型的检测精度
[125]	优化主干网络、增大模型感受野	无人机航拍影像中小麦倒伏的检测	扩大了模型的感受野,保留了更多的语义信息,提高了分割精度	使用的密集连接模块结构冗余多

### 3.3 无人机航拍影像语义分割的改进 Mask R-CNN 算法

Pi 等<sup>[126]</sup>为了从无人机航拍影像中生成像素级分割用于灾害发生后的损害评估,创建了一个名为 Volan2019 的数据集。该数据集包含 875 张带注释的图片,共计 9 个相关类别。提出了一种采用迁移学习训练、验证和测试 Mask R-CNN 和 PSPNet 网络的方法,分析了预测置信度和精度之间的相关性,设计并实现了一种基于数据均衡的有针对性的数据增强方法。

Zheng 等<sup>[127]</sup>以超高分辨率可见光、近红外和数字地表模型(Digital Surface Model, DSM)图像为研究对象,提出了基于深度卷积神经网络的草莓冠层自动提取和生物量预测方法。利用 Mask R-CNN 语义分割算法,分别使用 VGG-16 和 ResNet-50 网络作为主干,用于特征提取。实验结果表明,2 种主干均取得了良好的效果,验证了所提方法的可行性。Li 等<sup>[128]</sup>提出了一种改进的 Mask R-CNN 实例分割算法——ACE R-CNN,通过结合无人机航拍影像和激光雷达(Light Detection and Ranging, LiDAR)数据进行个体树种识别。ACE R-CNN 的主干网络为 AC-Net,它通过注意力互补模块(Attention Complementary Module, ACM)有选择性地强调目标特征。引入了边缘损失函数和边缘滤波器来计算边缘损失,提高个体树种的识别准确率,加快了模型的收敛速度。

由于轨道环境复杂、道砟遮挡以及无人机拍摄视点的变化,轨枕分割是一项充满挑战性的工作。Singh 等<sup>[129]</sup>利用 Mask R-CNN 方法对无人

机拍摄的低空图像中混凝土轨道轨枕进行实例分割,重点解决了小目标、复杂背景和噪声干扰等问题。

人工进行建筑外墙裂缝检测风险大、效率低,针对该问题,Chen 等<sup>[130]</sup>提出了一种基于无人机航拍影像的建筑外墙裂缝检测方法。首先,利用无人机采集了不同光照条件下住宅楼外墙图像,建立了裂缝数据集,然后,比较了 YOLOv3、YOLOv4、Faster R-CNN 和 Mask R-CNN 等方法的检测效果,其中 Mask R-CNN 的表现最佳,最后,基于裂纹在图像中的特点,对 Mask R-CNN 进行了有针对性的改进。由于裂纹在图像中所占比例较小,当图像大小被压缩时,目标容易因其细长且不明显的特征而丢失。因此改进后的方法使用 DenseNet 结构替换了原先的残差网络。DenseNet 中的批归一化操作不仅防止了梯度消失,还有利于增强网络的泛化性,加快网络的训练速度。为解决传统人工桥梁检测危险性高、影响交通、费用昂贵等问题,余加勇等<sup>[131]</sup>制作了关于桥梁裂缝、锈蚀和脱落病害的数据集,提出了基于无人机和 Mask R-CNN 检测网络的桥梁结构裂缝识别方法。该方法利用图像二值化、连通域去噪、边缘检测、裂缝骨架化和裂缝宽度计算等策略加强了模型的后处理能力,提高了对裂缝的检测精度和召回率,实现了裂缝形态及宽度信息的自动获取。

表 15 总结了上述不同场景下改进的 Mask R-CNN 无人机航拍影像语义分割方法。

本节主要介绍了基于 SegNet、U-Net 和 Mask R-CNN 的无人机航拍影像语义分割方法。通过优化主干网络、增强特征融合、数据集预处

表 15 不同场景下改进的 Mask R-CNN 无人机航拍影像语义分割方法

Table 15 Improved Mask R-CNN semantic segmentation method for UAV aerial images in different scenarios

文献	改进策略	应用场景	贡献	局限性
[128]	优化主干网络、添加注意力模块、修改损失函数	无人机航拍影像中小目标的检测	有选择性地强调从彩色图像和冠层高度模型分支提取的加权特征	添加的注意力模块增加了模型的复杂度
[130]	更换主干网络	无人机航拍影像中建筑外墙裂缝的检测	防止了梯度消失,有利于增强网络的泛化性,加快网络的训练速度	DenseNet模块结构冗余
[131]	数据集预处理	无人机航拍影像中桥梁裂缝、锈蚀和脱落病害的检测	提高了对裂缝的检测精度和召回率	模型预处理过程复杂、耗时

理和修改损失函数等方法,加强了目标的检测效果,使其更加适用于多种场景下目标的分割。

#### 4 无人机航拍影像数据集

基于深度学习的机器视觉方向经典的数据集包括 MNIST 手写数字数据集、Fashion MNIST 数据集、CIFAR-10 和 CIFAR-100 数据集、ILSVRC 竞赛的 ImageNet 数据集、用于检测和分割的 PASCAL VOC 和 COCO 数据集等。在训练模型时,数据集常分为训练集、验证集和测试集。其中,训练集是模型拟合的数据样本,用于调试卷积神经网络;验证集是模型训练过程中单独划分出的样本集,用于调整模型的超参数和初步评估模型的训练效果;测试集则用来评估模型最终的泛化能力。近年来,基于无人机航拍影像的目标检测与语义分割已成为研究热点,相关的数据集也在快速地更新之中。本节将对近期基于无人机航拍影像建立的数据集进行收集梳理。

Puertas 等<sup>[132]</sup>制作了一个环形交叉路口的无人机图像数据集。该数据由 61 896 幅大小为 1 920 像素×1 080 像素的彩色图像组成,拍摄时飞行高度为 100~120 m。标注的格式为 Pascal VOC 格式,划分有 6 个目标类别。

Gasienica-Józkowy 等<sup>[133]</sup>建立了无人机漂浮物(Aerial-drone Floating Objects, AFO)数据集,用于海上救援等工作。AFO 数据集包含了 3 647 幅图片和 39 991 个标注的对象,共计 6 个类别。其中图片的分辨率大小在 1 280 像素×720 像素~3 840 像素×2 160 像素之间,拍摄时无人机飞行高度为 30~80 m。

Behera 等<sup>[134]</sup>制作了基于高分辨率无人机航拍影像的数据集——NITRDrone,用于道路分割任务。数据集由 16 个大小为 8 GB 的视频序列和

1 000 幅图像组成,划分为 6 个目标类别,拍摄时无人机的飞行高度为 5~80 m 不等,图片分辨率为 1 280 像素×720 像素~3 000 像素×4 000 像素不等。

Du 等<sup>[135]</sup>建立了一个基于无人机航拍影像的检测与跟踪(UAV Detection and Tracking, UAVDT)数据集,用于目标检测、单目标跟踪和多目标跟踪等任务。UAVDT 数据集由 100 个视频序列组成,包含了多种常见的场景和不同的目标类别。

Bozic-Stulic 等<sup>[13]</sup>建立了 HERIDAL 数据集,用于搜索与救援工作。该数据集包含了尽可能多的真实场景,涵盖了实际搜救过程中可能遇到的多种情况。数据集的拍摄高度为 30~40 m,由 68 750 张大小为 4 000 像素×3 000 像素的图片组成。

Nigam 等<sup>[136]</sup>建立了 AeroScapes 语义分割数据集,从 141 个无人机拍摄的户外场景序列中抽取了 3 269 幅图片,并划分为 12 个不同类别。拍摄时,无人机的飞行高度为 5~50 m 之间,图片的大小为 1 280 像素×720 像素。

Robicquet 等<sup>[137-138]</sup>建立了 Campus 数据集,用于目标检测、多目标跟踪和大规模轨迹预测。该数据集通过无人机在户外环境中拍摄,包含 100 多个不同的俯视场景,20 000 个参与各种类型交互的目标。数据集中目标分为 6 类,拍摄时飞行高度约为 80 m,图片分辨率为 1 400 像素×1 904 像素。

Hsieh 等<sup>[139]</sup>建立了一个大规模停车场(Car Parking Lot, CARPK)数据集,用于目标检测和计数。CARPK 数据集包含来自 4 个不同停车场拍摄的无人机图像,所有目标均标记出左上角和右下角的坐标,共计 89 777 辆汽车。拍摄时,无人机的

飞行高度约为 40 m, 图片的大小为 1 000 像素 × 600 像素。该数据集是第一个基于无人机视图的停车场数据集。

Barekatain 等<sup>[140]</sup>建立了一个人体动作检测数据集 Okutama-Action。该数据集由时长 43 min 的带有完整注释的视频序列做成, 包含 12 个典型的户外动作类别。拍摄时, 无人机的飞行高度为 10~45 m, 相机倾斜角为 45° 或 90°, 图片分辨率为 3 840 像素 × 2 160 像素。

Sambolik 等<sup>[12]</sup>建立了 SARD 数据集, 用于无人机图像中复杂环境下的搜索和救援任务。该数据集由 1 981 幅图片组成, 包含了典型的运动类型, 并且模拟了不同年龄段和不同身体素质所带来的行为差异。拍摄时, 无人机的飞行高度为 5~50 m, 拍摄角度为 45°~90°。

Mueller 等<sup>[141]</sup>建立了 UAV123 数据集, 用于无人机图像中目标的检测与跟踪。该数据集包含 123 个视频序列, 分为 3 个子集, 涵盖了多种户外场景和目标类别, 拥有常见的视觉跟踪任务挑战, 例如长期的完全或部分遮挡、明暗变化和杂乱背景等。拍摄时, 无人机飞行高度为 5~25 m, 图片大小为 1 280 像素 × 720 像素。

Du 等<sup>[142-143]</sup>组织了基于无人机图像的目标检测挑战比赛, 发布了 VisDrone-DET 数据集, 共计 8 599 张图片和 10 个不同的目标类别。该数据集具有丰富的注释、遮挡情况和真实场景, 并且一些极具相似性的类别使得基于该数据集的目标检测任务更具挑战性。

表 16 对上述的无人机航拍影像数据集进行了总结与归纳。

表 16 常见的无人机航拍影像数据集  
Table 16 Common UAV aerial image datasets

文献	图片数量/张	无人机飞行高度/m	图像分辨率/像素	目标类别
[132]	61 896	100~120	1 920×1 080	汽车、自行车、卡车和公交车
[133]	3 647	30~80	1 280×720~3 840×2 160	人、滑板、游艇、浮标、帆船和皮划艇
[134]	1 000	5~80	1 280×720~3 000×4 000	公路、植被、行人、汽车、路灯和背景
[135]	8 000	10~30、30~70、70 以上	1 080×540	汽车、卡车和公共汽车
[13]	68 750	30~40	4 000×3 000	行人
[136]	3 269	5~50	1 280×720	行人、自行车、汽车、无人机、轮船、动物、障碍物、建筑物、植被、公路、天空和背景
[137-138]	20 000	80	1 400×1 904	行人、骑自行车的人、汽车、玩滑板的人、高尔夫球车和公交车
[139]	1 573	40	1 000×600	汽车
[140]		10~45	3 840×2 160	走路、坐下、站立、奔跑、躺下、搬运、推搡、阅读、喝水、打电话、握手和拥抱
[12]	1 981	5~50	1 000×1 000	奔跑、行走、站立、坐下和躺下
[141]		5~25	1 280×720	汽车、卡车、船舶、行人和飞机
[142-143]	8 599		2 000×2 000	行人、人、汽车、货车、公交车、卡车、摩托车、自行车、带遮阳棚的三轮车和不带遮阳棚的三轮车

## 5 目标检测与语义分割模型的评价指标

混淆矩阵可以用来对分类器的结果进行可视化处理, 对于常见的二分类问题来说, 它的混淆矩阵规模为 2×2。如表 17 所示, 其中 TP (True Positive)、FP (False Positive)、FN (False Negative) 和 TN (True Negative) 分别表示真阳性、假阳性、假阴性和真阴性。目标检测模型的评价指标主要包括: 精确率 (Precision,

表 17 混淆矩阵  
Table 17 Confusion matrix

真实值	预测值=Positive	预测值=Negative
True	TP (True Positive)	FN (False Negative)
False	FP (False Positive)	TN (True Negative)

P)、召回率 (Recall, R)、F1 分数 (F1 Score)、平均精度 (Average Precision, AP) 和平均精度均值 (Mean Average Precision, mAP) 等, 它们的



意义与计算公式如表 18 所示。语义分割模型的评价指标主要包括:像素精度(Pixel Accuracy, PA)、平均像素精度(Mean Pixel Accuracy, MPA)和平均交并比(Mean Intersection over Union, MIoU)等,它们的意义与计算公式如表 19 所示。此外,使用总参数量(Total Parameters)和浮点运算次数(Floating Point Operations, FLOPs)来衡量模型的大小和计算复杂度,当网络层数加深时,模型愈发复杂,相关参数的数值和对硬件的需求也会随之上升。使用每秒帧率(Frame Per Second, FPS)来衡量模型的检测速度,当处理一幅图像的时间缩短时,模型的 FPS 值会增大,表示速度得到了提升。

表 18 目标检测模型的评价指标

Table 18 Evaluation metrics for object detection models

评价指标	意义	计算式
精确率 $P$	预测真正样本与所有预测正样本之间的比值	(1)
召回率 $R$	预测真正样本与实际正样本之间的比值	(2)
F1 分数 (F1 Score)	精确率和召回率的调和平均值	(3)
平均精度 AP	某个类别精确率的平均值	(4)
平均精度均值 mAP	所有类别 AP 的平均值	(5)

表 19 语义分割模型的评价指标

Table 19 Evaluation metrics for semantic segmentation models

评价指标	意义	计算式
像素精度 PA	所有像素类别预测正确的数量与像素总数的比例	(6)
平均像素精度 MPA	所有类别像素精度的平均值	(7)
平均交并比 MIoU	所有类别交集与并集的比例的 平均值	(8)

$$P = \frac{TP}{TP + FP} \quad (1)$$

$$R = \frac{TP}{TP + FN} \quad (2)$$

$$F1 = \frac{2RP}{P + R} \quad (3)$$

$$AP = \int P(R) dR \quad (4)$$

$$mAP = \frac{1}{C} \sum_i AP_i \quad (5)$$

$$PA = \frac{\sum_i p_{ii}}{\sum_i \sum_j p_{ij}} \quad (6)$$

$$MPA = \frac{1}{K+1} \sum_i \frac{p_{ii}}{\sum_j p_{ij}} \quad (7)$$

$$MIoU = \frac{1}{K+1} \sum_i \frac{p_{ii}}{\sum_j p_{ij} + \sum_j p_{ji} - p_{ii}} \quad (8)$$

式中:  $C$  为目标检测数据集中的类别数;  $AP_i$  为第  $i$  类的平均精度;  $K+1$  为语义分割数据集中的类别数(包含  $K$  个目标类加 1 个背景类);  $p_{ii}$  为本属于第  $i$  类预测为第  $i$  类的像素点数量;  $p_{ij}$  为本属于第  $i$  类预测为第  $j$  类的像素点数量;  $p_{ji}$  为本属于第  $j$  类预测为  $i$  类的像素点数量。

## 6 主要问题与展望

无人机与深度学习技术相结合,在多个领域中表现出了一加一大于二的实际效果,使得无人机系统更加智能、高效、便捷。但在以往的研究与实际应用中,仍存在诸多问题与挑战。

1) 模型的轻量化问题。在大多数针对现有模型的改进过程中,通过增加主干的输出、添加其余模块和扩充特征融合结构等方法提高模型的检测精度,会导致模型更加复杂,参数激增,增加了模型的训练难度,延长了模型的检测时间。

2) 小目标检测效果不理想。由于无人机飞行高度的影响,捕获的影像中目标往往尺寸极小,所占的像素比例极低。图像在模型中,经过多次的下采样处理,会导致模型遗漏小目标的特征,进而降低了其检测精度。

3) 复杂背景干扰。无人机的起降基本不依赖所处的环境,在任意平坦的场地均可实现,因此无人机常常在各种环境下飞行,拍摄的影像包含了不同的地理位置。与自然影像相比,无人机航拍影像的背景更加复杂,带来的干扰问题会更加棘手。

4) 影像视场大的问题。无人机航拍影像往往视场较大,并且分辨率较高,这对检测方法的实时性提出了不小的挑战。大视场会导致图像在导入模型时花费更多的时间,高分辨率会导致

处理图像时更加复杂困难。

5) 目标分布不均。无人机拍摄的影像中包含的目标往往会存在2种极端的分布情况。例如,牧场中的牛羊等分布稀疏,极大的场地上可能仅存在单个目标;港口、停车场等环境下,船舶和汽车大多以较为密集的方式排列,相互之间距离非常近。上述极端分布情况会降低算法的召回率,错误地抑制掉真实样本。

6) 目标旋转的问题。无人机航拍影像中包含的物体大多不是水平分布排列,而是以随机的角度分散在图像各处。带有倾斜角的物体在检测时更容易受到背景的干扰,预测框将会包含更多的无关信息,这对目标特征的提取带来了不小的困难。

7) 数据集中样本类别不平衡。在多目标无人机航拍影像数据集中容易出现该问题,即有些样本类型的数量远大于其他样本数,影响模型的训练效果,导致模型可能无法学习到数据集中某些类别的抽象特征,降低了模型的平均精度均值。

针对上述问题,对未来无人机航拍影像目标检测与语义分割方法的研究趋势进行如下展望:

1) 模型的优化升级中,需要更多地关注计算复杂度和检测速度等因素。一个高性能的模型需要在检测精度和运算速度之间做到良好的平衡。如何在保持甚至降低模型复杂度的同时提升检测精度是一个充满挑战的研究。

2) 模型的特征提取部分,大多使用的是多层堆叠的残差网络。残差网络有助于缓解梯度问题,可以通过增加深度的方法来改善模型的检测性能。如何对传统的残差结构进行优化,以加强它针对小目标的特征提取能力有待深入研究。

3) 在图像处理中经常使用注意力模块来排除复杂背景的干扰,提升模型的检测性能。通过以往的研究,注意力机制已有多种多样的实现形式,主要分为通道注意力与空间注意力模块等。如何设计更加有效且轻量化的注意力模块非常有意义。

4) 对于处理大视场影像中的多尺度目标问题,采用特征金字塔结构是一种常见的解决方案。已经提出了如FPN、PANet、NAS-FPN<sup>[144]</sup>

和BiFPN等金字塔结构。如何加强不同尺度特征的融合,加强特征重用和特征提取仍需进一步研究。

5) 在目标检测算法中,使用边界框回归(Bounding Box Regression, BBR)来定位目标,这是决定目标定位性能的关键一步,良好的损失函数对于边界框回归至关重要。现已经提出了多种基于交并比的损失函数,例如IOU损失、GIOU损失、DIOU损失、CIOU损失和EIOU损失等。一个设计良好的损失函数有利于更好地衡量预测值和真实值之间的差异,指导下一步训练向正确的方向进行。

6) 在模型优化器的选择上还存在进一步改进的空间。之前的研究中,优化器选择表现的过于单一,仅为带动量的SGD优化器或Adam优化器等。后续可以尝试自适应与非自适应方法相结合的策略来优化网络模型参数,使其更好地逼近或达到最优值。

7) 在数据集的制作方面,应使用多种不同来源的综合数据验证模型,避免单一类别和单一背景下收集而来的图像。此外,包含图片数量的多少也是衡量数据集的一个重要指标。在后续数据集的制作中,需要注意多类别、多背景和图片数量等问题。

## 7 结 语

基于无人机航拍影像的目标检测与语义分割方法研究越来越受到众多科研人员的关注,相关方法的改进策略也在快速迭代与优化之中。本文综述了近些年关于无人机航拍影像中目标检测与语义分割方向的研究,按照不同的深度学习方法将其分类,这样更有利于进行方法之间横向与纵向的相互比较,有助于快速了解同一种方法在无人机航拍影像目标检测与语义分割方面的研究进展。最后,对现有研究存在的问题进行了讨论,分析了问题产生的原因以及对模型本身的影响,并且对未来该研究的发展方向进行了展望,以期提供有价值的参考。

## 参 考 文 献

- [1] 江波, 屈若锟, 李彦冬, 等. 基于深度学习的无人机航拍目标检测研究综述[J]. 航空学报, 2021, 42(4): 524519.

- JIANG B, QU R K, LI Y D, et al. Object detection in UAV imagery based on deep learning: Review[J]. *Acta Aeronautica et Astronautica Sinica*, 2021, 42(4): 524519 (in Chinese).
- [2] 刘传洋, 吴一全. 基于深度学习的输电线路视觉检测方法研究进展[J]. *中国电机工程学报*, 2023, 43(19): 7423-7446.
- LIU C Y, WU Y Q. Research progress of vision detection methods based on deep learning for transmission lines [J]. *Proceedings of the CSEE*, 2023, 43(19): 7423-7446 (in Chinese).
- [3] OSCO L P, DOS SANTOS DE ARRUDA M, GONÇALVES D N, et al. A CNN approach to simultaneously count plants and detect plantation-rows from UAV imagery [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2021, 174: 1-17.
- [4] VEERANAMPALAYAM SIVAKUMAR A N, LI J T, SCOTT S, et al. Comparison of object detection and patch-based classification deep learning models on mid-to late-season weed detection in UAV imagery [J]. *Remote Sensing*, 2020, 12(13): 2136.
- [5] WU J T, YANG G J, YANG H, et al. Extracting apple tree crown information from remote imagery using deep learning [J]. *Computers and Electronics in Agriculture*, 2020, 174: 105504.
- [6] 朱子健, 刘琪, 陈红芬, 等. 基于并行融合网络的航拍红外车辆小目标检测方法[J]. *光子学报*, 2022, 51(2): 190-202.
- ZHU Z J, LIU Q, CHEN H F, et al. Infrared small vehicle detection based on parallel fusion network[J]. *Acta Photonica Sinica*, 2022, 51(2): 190-202 (in Chinese).
- [7] LIU Y C, SHI G, LI Y X, et al. M-YOLO based detection and recognition of highway surface oil filling with unmanned aerial vehicle[C]//2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP). Piscataway: IEEE Press, 2022: 1884-1887.
- [8] DING W, ZHANG L. Building detection in remote sensing image based on improved YOLOv5[C]//2021 17th International Conference on Computational Intelligence and Security (CIS). Piscataway: IEEE Press, 2021: 133-136.
- [9] ZHANG R, WEN C B. SOD-YOLO: A small target defect detection algorithm for wind turbine blades based on improved YOLOv5 [J]. *Advanced Theory and Simulations*, 2022, 5(7): 2100631
- [10] GUO J, XIE J H, YUAN J Z, et al. Fault identification of transmission line shockproof hammer based on improved YOLOv4[C]//2021 International Conference on Intelligent Computing, Automation and Applications (ICAA). Piscataway: IEEE Press, 2021: 826-833.
- [11] LIU C Y, WU Y Q, LIU J J, et al. MTI-YOLO: A light-weight and real-time deep neural network for insulator detection in complex aerial images[J]. *Energies*, 2021, 14(5): 1426.
- [12] SAMBOLEK S, IVASIC-KOS M. Automatic person detection in search and rescue operations using deep CNN detectors [J]. *IEEE Access*, 2021, 9: 37905-37922.
- [13] BOŽIĆ - ŠTULIĆ D, MARUŠIĆ Ž, GOTOVAC S. Deep learning approach in aerial imagery for supporting land search and rescue missions[J]. *International Journal of Computer Vision*, 2019, 127(9): 1256-1278.
- [14] 索文凯, 胡文刚, 伍锡山, 等. 基于光学视觉辅助无人机自主降落研究综述[J]. *激光杂志*, 2019, 40(4): 9-13.
- SUO W K, HU W G, WU X S, et al. Research on autonomous landing of UAV based on optical vision[J]. *Laser Journal*, 2019, 40(4): 9-13 (in Chinese).
- [15] 张洲宇, 曹云峰, 范彦铭. 低空小型无人机空域冲突视觉感知技术研究进展[J]. *航空学报*, 2022, 43(8): 025645.
- ZHANG Z Y, CAO Y F, FAN Y M. Research progress of vision based aerospace conflict sensing technologies for small unmanned aerial vehicle in low altitude[J]. *Acta Aeronautica et Astronautica Sinica*, 2022, 43(8): 025645 (in Chinese).
- [16] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [17] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR' 05). Piscataway: IEEE Press, 2005: 886-893.
- [18] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Object detection with discriminatively trained part-based models[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(9): 1627-1645.
- [19] VIOLA P, JONES M J. Robust real-time face detection[J]. *International Journal of Computer Vision*, 2004, 57(2): 137-154.
- [20] PAPAGEORGIOU C, POGGIO T. A trainable system for object detection [J]. *International Journal of Computer Vision*, 2000, 38(1): 15-33.
- [21] MARDIA K V, HAINSWORTH T J. A spatial thresholding method for image segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1988, 10(6): 919-927.
- [22] ADAMS R, BISCHOF L. Seeded region growing[J].



- IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994, 16(6): 641-647.
- [23] GIANNAKEAS N, KARVELIS P S, EXARCHOS T P, et al. Segmentation of microarray images using pixel classification—comparison with clustering-based methods[J]. Computers in Biology and Medicine, 2013, 43(6): 705-716.
- [24] 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. Piscataway: IEEE Press, 2014: 580-587.
- [25] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Piscataway: IEEE Press, 2015: 3431-3440.
- [26] LUO X D, WU Y Q, ZHAO L Y. YOLOD: A target detection method for UAV aerial imagery[J]. Remote Sensing, 2022, 14(14): 3240.
- [27] LUO X D, WU Y Q, WANG F Y. Target detection method of UAV aerial imagery based on improved YOLOv5[J]. Remote Sensing, 2022, 14(19): 5063.
- [28] MITTAL P, SINGH R, SHARMA A. Deep learning-based object detection in low-altitude UAV datasets: A survey[J]. Image and Vision Computing, 2020, 104: 104046.
- [29] DIEZ Y, KENTSCH S, FUKUDA M, et al. Deep learning in forestry using UAV-acquired RGB data: A practical review[J]. Remote Sensing, 2021, 13(14): 2837.
- [30] ZHU P F, WEN L Y, DU D W, et al. Detection and tracking meet drones challenge[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(11): 7380-7399.
- [31] 程擎, 范满, 李彦冬, 等. 无人机航拍图像语义分割研究综述[J]. 计算机工程与应用, 2021, 57(19): 57-69.
- CHENG Q, FAN M, LI Y D, et al. Review on semantic segmentation of UAV aerial images[J]. Computer Engineering and Applications, 2021, 57(19): 57-69 (in Chinese).
- [32] 李子豪, 王正平, 贺云涛. 基于自适应协同注意力机制的航拍密集小目标检测算法[J]. 航空学报, 2023, 44(13): 244-254.
- LI Z H, WANG Z P, HE Y T. Aerial-photography dense small target detection algorithm based on adaptive cooperative attention mechanism[J]. Acta Aeronautica et Astronautica Sinica, 2023, 44(13): 244-254 (in Chinese).
- [33] 王殿伟, 胡里晨, 房杰, 等. 基于改进 Double-Head RCNN 的无人机航拍图像小目标检测算法[J/OL]. 北京航空航天大学学报. (2023-01-10) [2023-03-13]. <https://doi.org/10.13700/j.bh.1001-5965.2022.0591>.
- WANG D W, HU L C, FANG J, et al. Small object detection algorithm based on improved Double-Head RCNN for UAV aerial images [J/OL]. Journal of Beijing University of Aeronautics and Astronautics. (2023-01-10) [2023-03-13]. <https://doi.org/10.13700/j.bh.1001-5965.2022.0591> (in Chinese).
- [34] 谢学立, 李传祥, 杨小冈, 等. 基于动态感受野的航拍图像目标检测算法[J]. 光学学报, 2020, 40(4): 107-119.
- XIE X L, LI C X, YANG X G, et al. Dynamic receptive field-based object detection in aerial imaging[J]. Acta Optica Sinica, 2020, 40(4): 107-119 (in Chinese).
- [35] 张智, 易华挥, 郑锦. 聚焦小目标的航拍图像目标检测算法[J]. 电子学报, 2023, 51(4): 944-955.
- ZHANG Z, YI H H, ZHENG J. Focusing on small objects detector in aerial images[J]. Acta Electronica Sinica, 2023, 51(4): 944-955 (in Chinese).
- [36] OJALA T, PIETIKÄINEN M, HARWOOD D. A comparative study of texture measures with classification based on featured distributions[J]. Pattern Recognition, 1996, 29(1): 51-59.
- [37] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2016: 1440-1448.
- [38] 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.
- [39] CAI Z W, VASCONCELOS N. Cascade R-CNN: Delving into high quality object detection[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 6154-6162.
- [40] 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). Piscataway: IEEE Press, 2016: 779-788.
- [41] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot MultiBox detector[C]//European Conference on Computer Vision. Cham: Springer, 2016: 21-37.
- [42] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327.
- [43] TAN M X, PANG R M, LE Q V. EfficientDet: Scalable and efficient object detection[C]//2020 IEEE/CVF

- Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 10778-10787.
- [44] ZHOU X Y, KOLTUN V, KRÄHENBÜHL P. Tracking objects as points[C]//VEDALDI A, BISCHOF H, BROX T, et al. European Conference on Computer Vision. Cham: Springer, 2020: 474-490.
- [45] TIAN Z, SHEN C H, CHEN H, et al. FCOS: Fully convolutional one-stage object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2020: 9626-9635.
- [46] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation [C] //International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2015: 234-241.
- [47] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39 (12): 2481-2495.
- [48] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//16th IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2017: 2961-2969.
- [49] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [50] 奚祥书, 夏凯, 杨垠晖, 等. 结合多光谱影像降维与深度学习的城市单木树冠检测[J]. 遥感学报, 2022, 26(4): 711-721.
- XI X S, XIA K, YANG Y H, et al. Urban individual tree crown detection research using multispectral image dimensionality reduction with deep learning[J]. National Remote Sensing Bulletin, 2022, 26 (4): 711-721 (in Chinese).
- [51] DU L, SUN Y Q, CHEN S, et al. A novel object detection model based on Faster R-CNN for spodoptera frugiperda according to feeding trace of corn leaves[J]. Agriculture, 2022, 12(2): 248.
- [52] TANG J P, WANG J, WANG H L, et al. Insulator defect detection based on improved Faster R-CNN[C]//2022 4th Asia Energy and Electrical Engineering Symposium (AEEES). Piscataway: IEEE Press, 2022: 541-546.
- [53] WANG M, LUO X, WANG X, et al. Research on vehicle detection based on Faster R-CNN for UAV images [C] //IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium. Piscataway: IEEE Press, 2021: 1177-1180.
- [54] AVOLA D, CINQUE L, DIKO A, et al. MS-Faster R-CNN: Multi-stream backbone for improved Faster R-CNN object detection and aerial tracking from UAV images[J]. Remote Sensing, 2021, 13(9): 1670.
- [55] DING J J, ZHANG J H, ZHAN Z Q, et al. A precision efficient method for collapsed building detection in post-earthquake UAV images based on the improved NMS algorithm and Faster R-CNN[J]. Remote Sensing, 2022, 14(3): 663.
- [56] ZHOU Q F, DING S Q, QING G W, et al. UAV vision detection method for crane surface cracks based on Faster R-CNN and image segmentation [J]. Journal of Civil Structural Health Monitoring, 2022, 12(4): 845-855.
- [57] HUANG H, LI L L, MA H B. An improved Cascade R-CNN-based target detection algorithm for UAV aerial images[C]//2022 7th International Conference on Image, Vision and Computing (ICIVC). Piscataway: IEEE Press, 2022: 232-237.
- [58] BAO W X, REN Y X, LIANG D, et al. Defect detection algorithm of anti-vibration hammer based on improved Cascade R-CNN[C]//2020 International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI). Piscataway: IEEE Press, 2021: 294-297.
- [59] LIN Q Z, DING Y, XU H, et al. ECascade-RCNN: Enhanced Cascade RCNN for multi-scale object detection in UAV images[C]//2021 7th International Conference on Automation, Robotics and Applications (ICARA). Piscataway: IEEE Press, 2021: 268-272.
- [60] 李鑫, 刘帅男, 杨桢, 等. 基于改进 Cascade R-CNN 的输电线路多目标检测[J]. 电子测量与仪器学报, 2021, 35 (10): 24-32.
- LI X, LIU S N, YANG Z, et al. Multi-target detection of transmission lines based on improved Cascade R-CNN [J]. Journal of Electronic Measurement and Instrumentation, 2021, 35(10): 24-32 (in Chinese).
- [61] 张瑞倩, 邵振峰, Aleksei Portnov, 等. 多尺度空洞卷积的无人机影像目标检测方法[J]. 武汉大学学报(信息科学版), 2020, 45(6): 895-903.
- ZHANG R Q, SHAO Z F, PORTNOV A, et al. Multi-scale dilated convolutional neural network for object detection in UAV images [J]. Geomatics and Information Science of Wuhan University, 2020, 45(6): 895-903 (in Chinese).
- [62] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2017: 6517-6525.
- [63] LIN M, CHEN Q, YAN S. Network in network [DB/

- OL]. arXiv preprint: 1312.4400, 2013.
- [64] REDMON J, FARHADI A. YOLOv3: An incremental improvement[DB/OL]. arXiv preprint: 1804.02767, 2018.
- [65] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [DB/OL]. arXiv preprint: 2004.10934, 2020.
- [66] GE Z, LIU S, WANG F, et al. YOLOx: Exceeding YOLO series in 2021 [DB/OL]. arXiv preprint: 2107.08430, 2021.
- [67] RAMPRIYA R S, SUGANYA R, NATHAN S, et al. A comparative assessment of deep neural network models for detecting obstacles in the real time aerial railway track images[J]. Applied Artificial Intelligence, 2022, 36(1): 2018184..
- [68] LIU W J, QIANG J, LI X X, et al. UAV image small object detection based on composite backbone network[J]. Mobile Information Systems, 2022, 2022: 7319529.
- [69] 裴伟, 许晏铭, 朱永英, 等. 改进的SSD航拍目标检测方法[J]. 软件学报, 2019, 30(3): 738-758.
- PEI W, XU Y M, ZHU Y Y, et al. The target detection method of aerial photography images with improved SSD[J]. Journal of Software, 2019, 30(3): 738-758 (in Chinese).
- [70] 李旭, 宋世奇, 殷晓晴. 基于目标空间分布特征的无人机航拍车辆实时检测技术研究[J]. 中国公路学报, 2022, 35(12): 193-204.
- LI X, SONG S Q, YIN X Q. Real-time vehicle detection technology for UAV imagery based on target spatial distribution features [J]. China Journal of Highway and Transport, 2022, 35(12): 193-204 (in Chinese).
- [71] WANG X Q. Vehicle image detection method using deep learning in UAV video [J]. Computational Intelligence and Neuroscience, 2022, 2022: 8202535.
- [72] JAWAHARLALNEHRU A, SAMBANDHAM T, SEKAR V, et al. Target object detection from unmanned aerial vehicle (UAV) images based on improved YOLO algorithm[J]. Electronics, 2022, 11(15): 2343.
- [73] JAVED M G, RAZA M, GHAFAR M M, et al. QuantYOLO: A high-throughput and power-efficient object detection network for resource and power constrained UAVs[C]//2021 Digital Image Computing: Techniques and Applications (DICTA). Piscataway: IEEE Press, 2021: 1-8.
- [74] LIU J, JIA R, LI W, et al. High precision detection algorithm based on improved RetinaNet for defect recognition of transmission lines [J]. Energy Reports, 2020, 6: 2430-2440.
- [75] PURCELL C R, WALSH A J, COLEFAX A P, et al. Assessing the ability of deep learning techniques to perform real-time identification of shark species in live streaming video from drones[J]. Frontiers in Marine Science, 2022, 9: 981897.
- [76] TAKAYA K, SHIBATA A, MIZUNO Y, et al. Unmanned aerial vehicles and deep learning for assessment of anthropogenic marine debris on beaches on an island in a semi-enclosed sea in Japan [J]. Environmental Research Communications, 2022, 4(1): 015003.
- [77] 李洪瑶, 李小强, 韩心中, 等. 基于决策融合的多无人机协同目标检测识别算法[J]. 系统工程与电子技术, 2022, 44(3): 746-754.
- LI H Y, LI X Q, HAN X Z, et al. Cooperative object detection and recognition algorithm for multiple UAVs based on decision fusion [J]. Systems Engineering and Electronics, 2022, 44(3): 746-754 (in Chinese).
- [78] CHENG Z, CHEN J Y, ZHANG X, et al. Comparative study of two target detection algorithms in UAV aerial photography detection[C]//Proc SPIE 12057, Twelfth International Conference on Information Optics and Photonics, 2021, 12057: 889-894.
- [79] JUNOS M H, KHAIRUDDIN A S M, THANNIRMA-LAI S, et al. Automatic detection of oil palm fruits from UAV images using an improved YOLO model[J]. The Visual Computer, 2022, 38(7): 2341-2355.
- [80] YANG Z S, XU Z, WANG Y H. Bidirection-fusion-YOLOv3: An improved method for insulator defect detection using UAV image[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 3521408.
- [81] ZHU J Q, ZHONG J T, MA T, et al. Pavement distress detection using convolutional neural networks with images captured via UAV [J]. Automation in Construction, 2022, 133: 103991.
- [82] SAHIN O, OZER S. YOLODrone: Improved YOLO architecture for object detection in drone images[C]//2021 44th International Conference on Telecommunications and Signal Processing (TSP). Piscataway: IEEE Press, 2021: 361-365.
- [83] 刘芳, 韩笑. 基于多尺度深度学习的自适应航拍目标检测[J]. 航空学报, 2022, 43(5): 325270.
- LIU F, HAN X. Adaptive aerial object detection based on multi-scale deep learning[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(5): 325270 (in Chinese).
- [84] 蒲良, 张学军. 基于深度学习的无人机视觉目标检测与跟踪[J]. 北京航空航天大学学报, 2022, 48(5): 872-880.
- PU L, ZHANG X J. Deep learning based UAV vision object detection and tracking [J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(5): 872-880 (in Chinese).
- [85] TSENG H H, YANG M D, SAMINATHAN R, et al. Rice seedling detection in UAV images using transfer



- learning and machine learning[J]. Remote Sensing, 2022, 14(12): 2837.
- [86] ALDAHOUL N, KARIM H A, SABRI A Q M, et al. A comparison between various human detectors and CNN-based feature extractors for human activity recognition via aerial captured video sequences[J]. IEEE Access, 2022, 10: 63532-63553.
- [87] DOUSAI N M K, LONČARIĆ S. Detecting humans in search and rescue operations based on ensemble learning[J]. IEEE Access, 2022, 10: 26481-26492.
- [88] XIA H Y, YANG B H, LI Y L, et al. An improved CenterNet model for insulator defect detection using aerial imagery[J]. Sensors, 2022, 22(8): 2850.
- [89] ALBABA B M, OZER S. S<sub>y</sub>Net: An ensemble network for object detection in UAV images[C]//2020 25th International Conference on Pattern Recognition (ICPR). Piscataway: IEEE Press, 2021: 10227-10234.
- [90] 刘鑫, 黄进, 杨涛, 等. 改进 CenterNet 的无人机小目标捕获检测方法[J]. 计算机工程与应用, 2022, 58(14): 96-104.
- LIU X, HUANG J, YANG T, et al. Improved small object detection for UAV acquisition based on CenterNet[J]. Computer Engineering and Applications, 2022, 58(14): 96-104 (in Chinese).
- [91] 王胜利, 任鹏飞, 吕昕, 等. 基于中心点和双重注意力机制的无人机高分辨率图像小目标检测算法[J]. 应用科学学报, 2021, 39(4): 650-659.
- WANG S K, REN P F, LÜ X, et al. Small target detection algorithm of UAV high resolution image based on center point and dual attention mechanism[J]. Journal of Applied Sciences, 2021, 39(4): 650-659 (in Chinese).
- [92] AKSHATHA K R, BISWAS S, KARUNAKAR A K, et al. Anchored versus anchorless detector for car detection in aerial imagery[C]//2021 2nd Global Conference for Advancement in Technology (GCAT). Piscataway: IEEE Press, 2021: 1-6.
- [93] WANG H X, HUANG Z, CHEN Y Q, et al. Defect detection from power line images using advanced deep detectors[C]//2021 13th International Conference on Wireless Communications and Signal Processing (WCSP). Piscataway: IEEE Press, 2021: 1-5.
- [94] GAO Y X, HOU R M, GAO Q A, et al. A fast and accurate few-shot detector for objects with fewer pixels in drone image[J]. Electronics, 2021, 10(7): 783.
- [95] ZHAO L J, LIU C A, ZHANG Z, et al. Transmission line object detection method based on label adaptive allocation[J]. Mathematics, 2022, 10(12): 2150.
- [96] 张智, 郑锦. 结合帧间目标回归网络的无人机视频车辆检测[J]. 西安电子科技大学学报, 2021, 48(4): 151-158.
- ZHANG Z, ZHENG J. Interframe target regression network for vehicle detection in UAV video[J]. Journal of Xidian University, 2021, 48(4): 151-158 (in Chinese).
- [97] TAN L, LV X Y, LIAN X F, et al. YOLOv4\_Drone: UAV image target detection based on an improved YOLOv4 algorithm[J]. Computers & Electrical Engineering, 2021, 93: 107261.
- [98] CHENG Y B. Detection of power line insulator based on enhanced YOLO model[C]//2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC). Piscataway: IEEE Press, 2022: 626-632.
- [99] BAO W X, REN Y X, WANG N A, et al. Detection of abnormal vibration dampers on transmission lines in UAV remote sensing images with PMA-YOLO[J]. Remote Sensing, 2021, 13(20): 4134.
- [100] ZHU Y C, ZHOU J, YANG Y H, et al. Rapid target detection of fruit trees using UAV imaging and improved light YOLOv4 algorithm[J]. Remote Sensing, 2022, 14(17): 4324.
- [101] 黄丽明, 王懿祥, 徐琪, 等. 采用 YOLO 算法和无人机影像的松材线虫病异常变色木识别[J]. 农业工程学报, 2021, 37(14): 197-203.
- HUANG L M, WANG Y X, XU Q, et al. Recognition of abnormally discolored trees caused by pine wilt disease using YOLO algorithm and UAV images[J]. Transactions of the Chinese Society of Agricultural Engineering, 2021, 37(14): 197-203 (in Chinese).
- [102] YU Z W, SHEN Y G, SHEN C K. A real-time detection approach for bridge cracks based on YOLOv4-FPM[J]. Automation in Construction, 2021, 122: 103514.
- [103] WU J, SUN Y X, WANG X L. Corrosion detection method of transmission line components in mining area based on multiscale enhanced fusion[J]. Mobile Information Systems, 2022, 2022: 7408265.
- [104] ZHU J R, WANG X D, LIU Y, et al. UavTinyDet: Tiny object detection in UAV scenes[C]//2022 7th International Conference on Image, Vision and Computing (ICIVC). Piscataway: IEEE Press, 2022: 195-200.
- [105] 刘树东, 刘业辉, 孙叶美, 等. 基于倒置残差注意力的无人机航拍图像小目标检测[J]. 北京航空航天大学学报, 2023, 49(3): 514-524.
- LIU S D, LIU Y H, SUN Y M, et al. Small object detection in UAV aerial images based on inverted residual attention[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(3): 514-524 (in Chinese).
- [106] 冒国韬, 邓天民, 于楠晶. 基于多尺度分割注意力的无人机航拍图像目标检测算法[J]. 航空学报, 2023, 44(5): 268-278.
- MAO G T, DENG T M, YU N J. Object detection in

- UAV images based on multi-scale split attention [J]. *Acta Aeronautica et Astronautica Sinica*, 2023, 44(5): 268-278 (in Chinese).
- [107] LI Z R, NAMIKI A, SUZUKI S, et al. Application of low-altitude UAV remote sensing image object detection based on improved YOLOv5 [J]. *Applied Sciences*, 2022, 12(16): 8314.
- [108] ZHANG R, WEN C B. SOD-YOLO: A small target defect detection algorithm for wind turbine blades based on improved YOLOv5 [J]. *Advanced Theory and Simulations*, 2022, 5(7): 2100631.
- [109] 白健鹏, 王巍, 陈雨溪, 等. 基于轻量型 YOLOv5 的风机桨叶检测与空间定位[J]. *智能系统学报*, 2022, 17(6): 1173-1181.
- BAI J P, WANG W, CHEN Y X, et al. Detection and spatial location of wind turbine blades based on lightweight YOLOv5 [J]. *CAAI Transactions on Intelligent Systems*, 2022, 17(6): 1173-1181 (in Chinese).
- [110] WANG X W, ZHAO Q Z, JIANG P, et al. LDS-YOLO: A lightweight small object detection method for dead trees from shelter forest [J]. *Computers and Electronics in Agriculture*, 2022, 198: 107035.
- [111] LIU W, QUIJANO K, CRAWFORD M M. YOLOv5-tassel: Detecting tassels in RGB UAV imagery with improved YOLOv5 based on transfer learning [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2022, 15: 8085-8094.
- [112] 鲍文霞, 谢文杰, 胡根生, 等. 基于 TPH-YOLO 的无人机图像麦穗计数方法[J]. *农业工程学报*, 2023, 39(1): 155-161.
- BAO W X, XIE W J, HU G S, et al. Wheat ear counting method in UAV images based on TPH-YOLO [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2023, 39(1): 155-161 (in Chinese).
- [113] RU C Y, ZHANG S H, QU C N, et al. The high-precision detection method for insulators' self-explosion defect based on the unmanned aerial vehicle with improved lightweight ECA-YOLOX-tiny model [J]. *Applied Sciences*, 2022, 12(18): 9314.
- [114] YAO Z S, LIU T, YANG T L, et al. Rapid detection of wheat ears in orthophotos from unmanned aerial vehicles in fields based on YOLOX [J]. *Frontiers in Plant Science*, 2022, 13: 851245.
- [115] HOU H Y, CHEN M X, TIE Y B, et al. A universal landslide detection method in optical remote sensing images based on improved YOLOX [J]. *Remote Sensing*, 2022, 14(19): 4939.
- [116] ZHAN J L, HU Y W, CAI W W, et al. PDAM-STPNNet: A small target detection approach for wild-land fire smoke through remote sensing images [J]. *Symmetry*, 2021, 13(12): 2260.
- [117] ZHONG J T, ZHU J Q, JU H Y, et al. Multi-scale feature fusion network for pixel-level pavement distress detection [J]. *Automation in Construction*, 2022, 141: 104436.
- [118] QI L J, ZHOU J R, WAN J J, et al. Canopy recognition of cherry fruit tree based on SegNet network model [C]// *Proc SPIE 11915, International Conference on Optics and Image Processing (ICOIP 2021)*, 2021, 11915: 92-104.
- [119] LI G A, HAN W T, HUANG S J, et al. Extraction of sunflower lodging information based on UAV multi-spectral remote sensing and deep learning [J]. *Remote Sensing*, 2021, 13(14): 2721.
- [120] LI F, BAI J Y, ZHANG M Y, et al. Yield estimation of high-density cotton fields using low-altitude UAV imaging and deep learning [J]. *Plant Methods*, 2022, 18(1): 1-11.
- [121] JEON E I, KIM S, PARK S, et al. Semantic segmentation of seagrass habitat from drone imagery based on deep learning: A comparative study [J]. *Ecological Informatics*, 2021, 66: 101430.
- [122] HUANG L, WU X Q, PENG Q Z, et al. Depth semantic segmentation of tobacco planting areas from unmanned aerial vehicle remote sensing images in plateau mountains [J]. *Journal of Spectroscopy*, 2021, 2021: 6687799.
- [123] NARVARIA A, KUMAR U, JHANWEE K S, et al. Classification and identification of crops using deep learning with UAV data [C]// *2021 IEEE International India Geoscience and Remote Sensing Symposium (InGARSS)*. Piscataway: IEEE Press, 2022: 153-156.
- [124] ZHAI Z Q, CHEN X G, ZHANG R Y, et al. Evaluation of residual plastic film pollution in pre-sowing cotton field using UAV imaging and semantic segmentation [J]. *Frontiers in Plant Science*, 2022, 13: 991191.
- [125] YU J, CHENG T, CAI N, et al. Wheat lodging extraction using Improved\_Unet network [J]. *Frontiers in Plant Science*, 2022, 13: 1009835.
- [126] PI Y L, NATH N D, BEHZADAN A. Detection and semantic segmentation of disaster damage in UAV foot-age [J]. *Journal of Computing in Civil Engineering*, 2021, 35(2): 04020063.
- [127] ZHENG C W, ABD-ELRAHMAN A, WHITAKER V M, et al. Deep learning for strawberry canopy delineation and biomass prediction from high-resolution images [J]. *Plant Phenomics*, 2022, 2022: 9850486.
- [128] LI Y B, CHAI G Q, WANG Y T, et al. ACE R-CNN: An

- attention complementary and edge detection-based instance segmentation algorithm for individual tree species identification using UAV RGB images and LiDAR data[J]. *Remote Sensing*, 2022, 14(13): 3035.
- [129] SINGH A K, DWIVEDI A K, SUMANTH M, et al. An efficient approach for instance segmentation of railway track sleepers in low altitude UAV images using Mask R-CNN[C]//IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium. Piscataway: IEEE Press, 2022: 4895-4898.
- [130] CHEN W, CHEN C Y, LIU M, et al. Wall cracks detection in aerial images using improved Mask R-CNN[J]. *Computers, Materials & Continua*, 2022, 73(1): 767-782.
- [131] 余加勇, 李峰, 薛现凯, 等. 基于无人机及 Mask R-CNN 的桥梁结构裂缝智能识别[J]. *中国公路学报*, 2021, 34(12): 80-90.
- YU J Y, LI F, XUE X K, et al. Intelligent identification of bridge structural cracks based on unmanned aerial vehicle and Mask R-CNN[J]. *China Journal of Highway and Transport*, 2021, 34(12): 80-90 (in Chinese).
- [132] PUERTAS E, DE-LAS-HERAS G, FERNÁNDEZ-ANDRÉS J, et al. Dataset: Round about aerial images for vehicle detection[J]. *Data*, 2022, 7(4): 47.
- [133] GAŚIENICA-JÓZKOWY J, KNAPIK M, CYGANIEK B. An ensemble deep learning method with optimized weights for drone-based water rescue and surveillance [J]. *Integrated Computer-Aided Engineering*, 2021, 28(3): 221-235.
- [134] BEHERA T K, BAKSHI S, SA P K, et al. The NITRDrone dataset to address the challenges for road extraction from aerial images[J]. *Journal of Signal Processing Systems*, 2023, 95(2): 197-209.
- [135] DU D W, QI Y K, YU H Y, et al. The unmanned aerial vehicle benchmark: Object detection and tracking[C]//European Conference on Computer Vision. Cham: Springer, 2018: 375-391.
- [136] NIGAM I, HUANG C, RAMANAN D. Ensemble knowledge transfer for semantic segmentation[C]//2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE Press, 2018: 1499-1508.
- [137] ROBICQUET A, SADEGHIAN A, ALAHI A, et al. Learning social etiquette: Human trajectory understanding in crowded scenes [C]//European Conference on Computer Vision. Cham: Springer, 2016: 549-565.
- [138] ROBICQUET A, ALAHI A, SADEGHIAN A, et al. Forecasting social navigation in crowded complex scenes[DB/OL]. arXiv preprint: 1601.00998, 2016.
- [139] HSIEH M R, LIN Y L, HSU W H. Drone-based object counting by spatially regularized regional proposal network[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2017: 4165-4173.
- [140] BAREKATAIN M, MARTÍ M, SHIH H F, et al. Okutama-action: An aerial view video dataset for concurrent human action detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE Press, 2017: 2153-2160.
- [141] MUELLER M, SMITH N, GHANEM B. A benchmark and simulator for UAV tracking[C]//14th European Conference on Computer Vision (ECCV). Cham: Springer, 2016: 445-461.
- [142] DU D W, ZHU P F, WEN L Y, et al. VisDrone-DET2019: The vision meets drone object detection in image challenge results [C]//2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). Piscataway: IEEE Press, 2020: 213-226.
- [143] ZHU P F, DU D W, WEN L Y, et al. VisDrone-VID2019: The vision meets drone object detection in video challenge results [C]//2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). Piscataway: IEEE Press, 2020: 227-235.
- [144] GHIASI G, LIN T Y, LE Q V. NAS-FPN: Learning scalable feature pyramid architecture for object detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 7029-7038.

(责任编辑: 李丹)



## Research progress on deep learning methods for object detection and semantic segmentation in UAV aerial images

LUO Xudong, WU Yiquan\*, CHEN Jinlin

*College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China*

**Abstract:** Unmanned Aerial Vehicles (UAV) have been widely used due to their advantages of small size, light weight, and simple operation. The combination of the deep learning method and UAV system can help to quickly and accurately detect the desired target on UAV aerial images with high definition and a wide field of view. Related topics have become one of the current research hotspots. This paper reviews the research progress of deep learning methods for object detection and semantic segmentation in UAV aerial images in the past ten years. First, the characteristics and wide range of application scenarios of UAVs and their aerial images are summarized, and the development process of target detection and semantic segmentation methods for UAV aerial images is briefly described. Then, the object detection and semantic segmentation methods of UAV aerial images based on deep learning are classified according to different network models, and their improvement strategies, application scenarios, contributions, and limitations are summarized. Subsequently, the data sets of aerial images taken by UAVs in recent years are collected and sorted out, and the evaluation indicators of commonly used convolutional neural network models are summarized. Finally, the existing problems in this field are pointed out, and prospects for future research trends are presented.

**Keywords:** Unmanned Aerial Vehicle (UAV); deep learning; object detection; semantic segmentation; data set; performance evaluation indicator

---

**Received:** 2023-04-06; **Revised:** 2023-05-04; **Accepted:** 2023-06-01; **Published online:** 2023-06-12 09:47

**URL:** <https://hkxb.buaa.edu.cn/CN/Y2024/V45/I6/28822>

**Foundation item:** National Natural Science Foundation of China (61573183)

\* **Corresponding author.** E-mail: nuaaimage@163.com