

无人机航拍图像目标检测技术研究综述

钟帅, 王丽萍*

中国人民公安大学侦查学院, 北京 100038

摘要 近年来,无人机凭借其卓越的灵活性、机动性以及目标的精准感知能力,在众多领域发挥着重要作用,将基于深度学习的目标检测方法 with 无人机系统相结合,已成为当前研究和应用的焦点之一。然而,在实际模型部署和检测过程中,无人机系统往往易受到计算资源限制、复杂背景干扰以及图像分辨率不足等因素的影响,这些因素会导致模型性能下降,难以实现高效准确的检测任务。鉴于此,本文首先系统阐述了无人机航拍图像目标检测领域所面临的难点与挑战;随后,针对上述问题,重点综述了基于深度学习的检测方法和优化策略;接着,对现有的航拍图像数据集及其特性进行详细概述,为后续研究提供数据支持;最后,总结展望了该领域在未来的研究方向和发展趋势,指出亟待深入探究的关键问题,以为后续研究提供有益的参考和指导。

关键词 目标检测; 无人机; 航拍图像; 深度学习

中图分类号 TP391.41;V279

文献标志码 A

DOI: 10.3788/LOP242120

Review of Research on Object Detection in UAV Aerial Images

Zhong Shuai, Wang Liping*

College of Investigation, People's Public Security University of China, Beijing 100038, China

Abstract In recent years, unmanned aerial vehicles (UAVs) have become pivotal across various fields due to their remarkable flexibility, maneuverability, and precise target perception capabilities. Integrating deep learning-based object detection methods with UAV systems has become a major focus of contemporary research and applications. However, during practical deployment and detection, UAV systems often face limitations such as restricted computational resources, interference from complex backgrounds, and low image resolution, which can degrade model performance and hinder the efficiency and accuracy of detection tasks. This study systematically describes the difficulties and challenges associated with object detection in UAV aerial image. It then reviews existing deep-learning methods and optimization strategies aimed at overcoming these issues. Additionally, the study gives a detailed overview of the existing aerial image datasets, to provide data support for further research. Finally, the study summarizes future research directions and development trends, emphasizing critical issues that demand further exploration. This study aims to serve as a valuable reference and guide for advancing future studies.

Key words object detection; unmanned aerial vehicle; aerial image; deep learning

1 引言

在智能化、自动化不断深入的时代背景下,无人机技术已经成为融合新一代电子信息技术与航空工业技术的关键技术。将无人机技术与图像检测技术相结合,不仅可以实现对较大范围区域的即时监测和精确分析,帮助决策者及时捕捉监控区域的最新动态,还可以提供翔实的数据支持。在智能交通管控领域,采用无人机搭载数字图像处理技术,可以实现航拍路面图像的自动拼

接与问题识别,为道路状况监测、路面病害识别和智能交通管理提供技术支持和保障^[1]。在生态环境保护与建设方面,基于无人机视角下采集的森林图像数据,结合高效的机器学习算法,可以实现对森林类型的智能化识别与分类,有效提升森林资源管理的效率^[2]。在灾害应对与治理工作中,将无人机技术与低空遥感技术相融合,可以实现对崩塌地质灾害体的连续监测,为灾害的预防和整治提供有力的科学依据^[3]。在军事指挥作战场景下,将数据处理与分析技术融入无人机作战效能

收稿日期: 2024-10-15; 修回日期: 2024-11-11; 录用日期: 2024-11-26; 网络首发日期: 2024-12-06

基金项目: 中国人民公安大学刑事科学技术双一流创新研究专项(2023SYL06)

通信作者: *20216953@ppsuc.edu.cn

评估与优化系统中,可以进一步提升无人机对地攻击、自主作战的能力,为战场态势感知、目标定位、打击效果评估等关键环节提供坚实的数据支持^[4]。

为准确高效地实现目标识别与定位,研究人员致力于对现有目标检测技术进行深入研究和创新。目标检测技术的发展历程如图 1 所示。传统的目标检测技术,如尺度不变特征变换(SIFT)^[5]、VJ 检测器(VJ Det)^[6]、方向梯度直方图(HOG)^[7]、可变形部件模

型(DPM)^[8]等,多依赖于大量的专业知识和繁琐的人工特征设计过程,对特征提取和表征的能力有限,检测速度和检测精度均不太理想,难以适应无人机航拍视角下的多样化检测需求。随着深度学习技术的不断演进,以卷积神经网络(CNN)^[9]为基础衍生的一系列目标检测算法为图像处理领域注入了新的动力,这些算法能够自动从数据中学习复杂特征,有效提升目标检测的效率和精度。

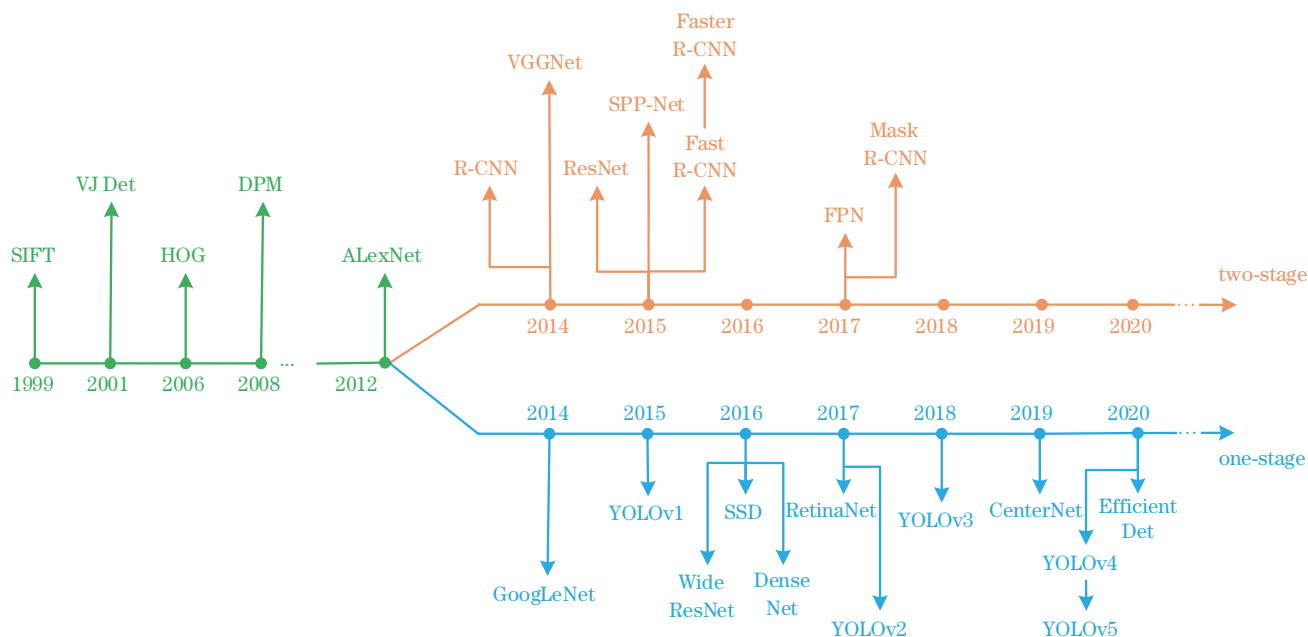


图 1 目标检测技术的发展历程

Fig. 1 Development process of target detection technology

在深度学习领域,根据处理任务的流程和复杂度,可以将检测算法分为两阶段目标检测算法和单阶段目标检测算法两大类。两阶段目标检测算法的处理流程如图 2 所示。首先,在原始图像上扫描并筛选出若干可能含有目标的候选区域框,接着对这些区域的深度特征进行提取、分类和边界框回归处理,实现对目标对

象的精确分类与定位。该阶段的典型代表算法有 R-CNN^[10]、SPP-Net^[11]、Fast R-CNN^[12]、Faster R-CNN^[13]、FPN^[14]、Mask R-CNN^[15]等。一阶段目标检测算法的处理流程如图 3 所示,该算法摒弃了传统候选区域框的生成步骤,转而采取直接在图像中预测目标类别和坐标信息的策略,这一策略使得整个检测过程更加迅

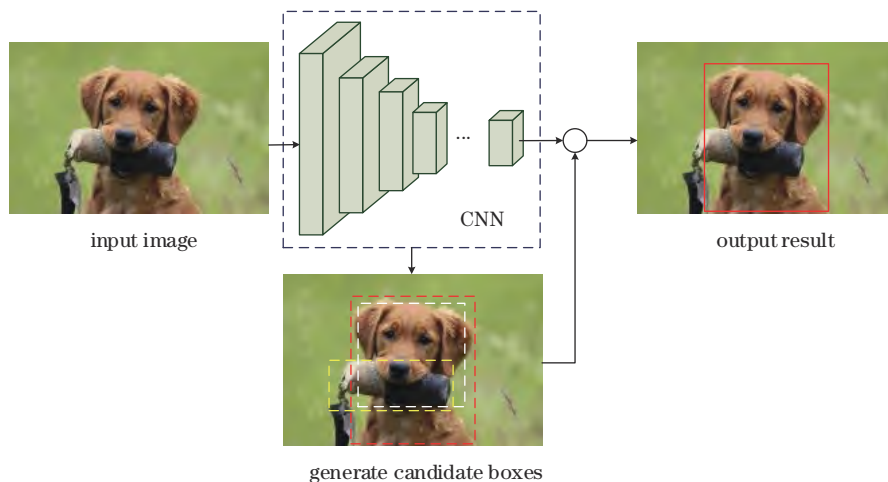


图 2 基于深度学习的两阶段目标检测算法

Fig. 2 Two-stage object detection algorithm based on deep learning

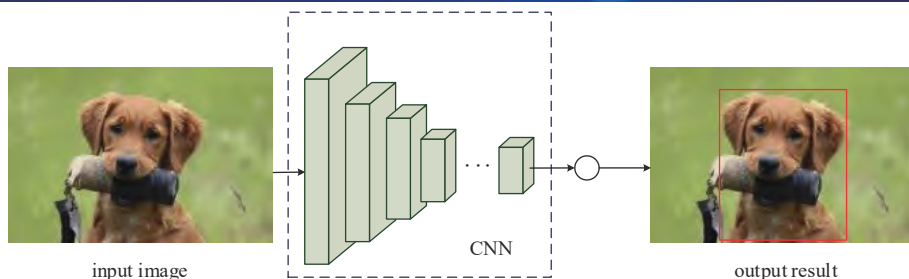


图 3 基于深度学习的单阶段目标检测算法

Fig. 3 One-stage object detection algorithm based on deep learning

速和高效。该阶段的典型代表算法有 SSD^[16]、RetinaNet^[17]、CenterNet^[18]、EfficientDet^[19]以及 YOLO 等系列算法^[20-25]等。特别地, YOLO 系列算法自推出以来, 经过不断更新迭代, 目前已发展至 YOLOv11 版本^[26]。YOLOv11 在继承历代版本核心优势的基础上, 创新性地引入了改良的锚点生成机制和多尺度预测技术, 对网络结构进行了深度优化, 显著提升了检测性能和效率, 受到众多研究学者的广泛关注。

整体而言, 相较于传统的目标检测方法, 深度学习驱动下的目标检测技术摒弃了繁琐的手工特征提取步骤, 能够自动从海量数据中提取关键信息。这一技术革新能使模型在应对实时性和检测性能要求更高的复杂场景时, 展现出更大的发展潜力和应用前景。但是, 当将其与无人机技术相融合, 执行小目标检测这类细粒度识别任务时, 面临着诸多挑战和限制。首先, 无人机的设计需要充分兼顾轻便和灵活部署的性能要求, 这就导致其携带的传感器和计算资源相对有限, 进而容易引发算力不足的问题。其次, 航拍图像中的目标对象往往尺寸较小、分辨率低, 在密集场景中容易发生相互遮挡的情况, 这给目标的准确识别带来了极大挑战。特别是在城市街道等人流密集区域, 如何从复杂的背景中提取出关键特征信息, 实现对目标对象的精准分类和定位, 已经成为当前的研究热点之一。此外, 在恶劣天气等极端条件下, 无人机拍摄的图像在清晰度、对比度等方面都可能不够理想, 这些因素极大地增加了图像检测过程中误判、漏判的风险, 亟须采取相应的技术手段进行优化和提升。

鉴于此, 本文通过系统阐述无人机航拍图像在目标检测领域面临的难点与挑战, 深入探讨解决这些问题常用的深度学习改进方法和优化策略。同时, 对当前可用的航拍图像数据集进行系统梳理, 详细阐述它们的构成、规模及应用范围, 旨在为研究人员提供全面的数据集概览。最后, 针对当前无人机航拍图像目标检测研究存在的缺陷与不足, 对其未来的发展趋势和重点研究方向进行分析和展望。

2 无人机航拍图像目标检测的难点与挑战

无人机航拍图像目标检测任务的执行主要依赖于

其搭载的高分辨率摄像头以及其他先进的传感器设备, 这些设备兼具轻量、便捷与高效能等多重优点, 极大地提高了无人机在空中作业时的机动性与灵活性, 为地面目标的监测提供了全新的视角和更广阔的视野。无人机所能搭载的传感器设备在算力资源上存在限制, 并且其特殊的成像机制对航拍环境提出了更高要求。此外, 飞行视角的不断变化很容易导致目标对象在视觉特性上出现遮挡、形变等现象, 这为实际场景下的目标检测任务带来了诸多挑战。因此, 针对无人机航拍图像目标检测过程中面临的各种难题与挑战进行全面剖析与总结, 能够为未来的研究方向提供明确的优化方向, 也有助于针对性地创新和发展更加先进的检测算法和技术手段, 从而全面提升无人机航拍图像目标检测技术的整体效能。

2.1 无人机平台资源限制

在无人机飞行过程中, 为确保其轻便、灵活和低能耗, 通常会尽可能缩减其体积和重量。然而, 在面对复杂环境下的高分辨率图像检测任务时, 又要求无人机必须具备强大的存储与处理能力, 这对其算力资源提出了严苛要求。因此, 在资源受限的无人机平台中, 如何通过算法优化来简化模型结构、降低模型的复杂度, 最终实现其检测精度与运算效率的进一步提升, 成为无人机技术发展中亟待解决的关键挑战。

2.2 复杂环境对图像质量的影响

无人机航拍环境的多变性对图像质量提出了巨大挑战, 尽管现代无人机技术的发展已经有所进步, 能在一定程度上适应这些变化, 但当其面对强光、暗光或恶劣天气等极端航拍条件时, 图像的噪声问题和模糊现象就变得尤为突出。这些因素会直接导致图像的清晰度受损以及目标对象的细节信息丢失, 极大地干扰了无人机在图像处理过程中的特征提取和目标识别能力, 进而削弱了航拍图像的检测精度和可靠性。

1) 强光照环境

在强光照射下, 航拍图像极易出现曝光现象, 特别是在高反光区域, 细节损失尤为严重, 这往往会导致过曝区域的关键信息部分或完全丢失, 从而导致目标对象的特征信息提取困难。如图 4 所示, 无人机在强灯光照射下捕获的街景图像出现了严重的曝光现象, 并且越靠近强光源区域, 失真程度越明显, 行人目标对象



图 4 强光环境下的街景图像

Fig. 4 Street image in intense light environment

的轮廓信息变得越模糊,原有的细节和真实感也丢失。

2) 暗光环境

在昏暗的环境中,航拍图像的整体亮度和对比度明显降低,导致目标轮廓的辨识度明显下降,与周围环境的区分度减小,易与图像背景混为一体。此外,由于航拍光线不足,捕获的图像还可能出现颜色偏差,造成色彩失真,这进一步增大了分析和处理航拍图像的难度。如图 5 所示,无人机在夜间对公园广场进行拍摄时,传感器的灵敏度明显降低,光源的照射角度也导致右下角的行人产生了明显的阴影,行人轮廓等细节特征变得模糊不清,增加了目标对象的辨识难度。

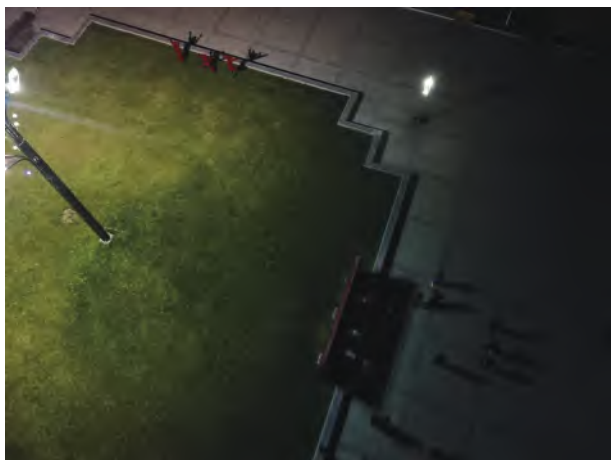


图 5 暗光环境下的公园图像

Fig. 5 Park image in low light environment

3) 极端天气条件

无人机在暴风雨、大雪、浓雾等恶劣天气下作业时,其传感器镜头可能会遭受雨水、沙尘等其他物理性遮挡,进而降低图像的清晰度和分辨率。此外,大气湍流可能会导致无人机拍摄图像时出现抖动现象,进而导致捕获的图像质量明显下降。另外,降雪与浓雾则可能遮蔽目标对象的具体细节,导致图像的辨识难度增大,检测性能降低。图 6 所示是无人机在大雪天气



图 6 雪后停车场图像

Fig. 6 Parking lot image after snowfall

下航拍的停车场图像。大雪将车辆目标部分或完全覆盖,额外的噪声和干扰降低了航拍图像质量,导致无人机在执行航拍图像目标检测任务时,无法有效区分目标与背景,从而降低了检测的准确率和鲁棒性。

2.3 目标变化引发的检测难点

在执行无人机空中监测与图像采集任务时,目标对象所处环境的复杂性及其运动状态的多变性对无人机航拍图像目标检测任务提出了更为严苛的要求与挑战。由于无人机飞行高度及拍摄视角的不同,同一航拍图像中不同目标对象在尺寸、形态以及空间分布密度上表现出了明显的异质性。此外,目标对象运动状态的不断变化进一步增大了捕获目标特征信息的难度。因此,通过对这些难点进行深入分析,能够帮助研究人员针对不同特点的检测任务,选择恰当的航拍图像数据集,以适应各种复杂环境下的目标检测需求。

1) 目标运动状态变化

目标对象处于静止状态时,无人机所采集的图像通常具有较高的稳定性。然而,当目标物体开始进行加速、减速或改变运动轨迹时,无人机视角下获取的航拍图像往往易出现模糊和抖动等失真现象,导致目标轮廓变得不清晰,并且颜色和纹理细节也会丢失,影响了图像的可用性。图 7 所示是无人机在高架桥上空捕获的图像。可见,受车辆运动状态持续变化的影响,图像中的目标出现了明显的拖影现象,车辆的大小、形状



图 7 无人机捕获的高架桥上行驶的汽车图像

Fig. 7 Image of cars driving on the overpass taken by UAV

等关键细节变得模糊不清,削弱了目标识别的精确性。

2) 目标对象尺寸变化

在无人机航拍过程中,由于飞行高度和拍摄视角的不同,图像中目标对象的尺度变化不一,常出现近大远小的情况。近景拍摄的目标对象占据的像素多,细节丰富,而在高空远景拍摄的目标对象在图像中占据的像素点数量较少,能够提取的特征信息相对匮乏,这在很大程度上降低了目标识别的准确性。图 8 所示是无人机在城市高空拍摄的街景图像。由于飞行视角的差异,靠近无人机的车辆呈现出较大的尺寸,且轮廓信息清晰,细节特征更加明显。相比之下,位于街道远端的物体则尺寸较小,轮廓和特征信息相对模糊,并且随着距离增加,目标的细节逐渐减弱。这种目标尺寸的差异性使得后续的图像分析和处理变得更加复杂。



图 8 无人机拍摄的城市街景图像

Fig. 8 Image of the city street landscape taken by UAV

3) 目标对象分布不均

无人机航拍图像覆盖的区域范围较广,常常出现多样化的场景密度。具体而言,在一张图片中,目标在部分区域可能呈现高度密集性,而在其他区域则可能分布稀疏,在极端情况下甚至仅包含单个目标,与密集区域形成强烈对比,造成不均匀的区域分布,给目标对象的准确识别和计数带来了挑战。图 9 所示是无人机



图 9 无人机航拍的校园广场图像

Fig. 9 Image of the campus square taken by UAV

在学校上空中拍摄的图像。图像左侧区域的学生较为密集,而右侧区域则相对稀疏,这种不同区域学生密度对比鲜明的场景增大了实际检测过程中出现漏检的可能性。

4) 目标对象遮挡现象

在复杂多变的拍摄环境下,遮挡现象的发生通常源于多种因素。一方面,自然环境中的物体(如树木、建筑物等)成为遮挡源时,会减少目标对象在图像中的可视面积,导致关键特征信息丢失。另一方面,在密集场景下,三维空间中的目标对象重叠投影到二维图像平面时,目标之间也会出现相互遮挡的视觉效果。遮挡现象的出现不仅增大了图像处理的复杂性,还易导致检测算法在对目标分类时产生混淆,降低检测精度。图 10 所示是无人机在公园航拍的图像。图中,在乒乓球运动的部分人群目标被树木遮挡,导致实际检测过程中的目标对象特征信息缺失,此时极易出现漏检情况。



图 10 无人机航拍的公园图像

Fig. 10 Image of the park taken by UAV

3 无人机航拍图像目标检测方法

尽管深度学习算法在目标检测领域取得了令人瞩目的成就,但将这些算法应用于无人机航拍图像分析任务时,检测性能往往会因为受到无人机航拍视角独特性的影响和作业场景复杂性的制约而不尽如人意。为此,当前的研究动向主要聚焦于对现有模型的优化与升级,旨在有针对性地增强算法在无人机航拍图像目标检测任务中的具体表现,以便更好地应对相关挑战。因此,本章节针对无人机航拍图像目标检测过程中存在的诸多难点与挑战,从数据增强、注意力机制、特征融合机制和模型轻量化 4 个维度,系统剖析和评述现有优化方法和改进策略,以期后续研究提供有益借鉴。

3.1 数据增强

在目标检测领域,数据增强技术因能够有效提高

模型在多样化场景中的检测能力而受到了众多学者的广泛关注。DeVries 等^[27]提出了 Cutout 数据增强技术,该方法在训练图像中引入随机遮挡策略,通过随机选择一个或多个区域填充 0 像素值来模拟目标对象在现实场景中被部分遮挡的情况,以提升模型识别遮挡图像的鲁棒性,降低过拟合风险。但是,该技术仅针对单一图像进行操作,无法利用样本间的协同效应来进一步增强模型的泛化能力。为解决这一难题,Zhang 等^[28]在此基础上进一步拓展,提出了 Mixup 技术。Mixup 技术不再局限于单一图像的处理,而是将两个随机选取的样本及其对应的标签按照一定比例进行混合,提高了模型的性能。随后,Yun 等^[29]结合 Cutout 和 Mixup 两类数据增强策略的优点,提出了 CutMix 技术,该技术既保留了 Mixup 技术中的标签平滑特性,又增加了空间信息的多样性,有效地促进了模型对局部特征的深入学习。Yolo 团队的 Bochkovskiy 等^[23]开创性地提出了 Mosaic 数据增强策略。区别于以往主要聚焦于单一图像或仅两两图像混合的增强技

术,Mosaic 策略将这一理念扩展至 4 张不同的图像。它通过将 4 张涵盖不同语义信息的图片拼接成一张训练图片,使模型在单次前向传播过程中充分关注多个不同应用场景下的目标特征信息,进一步增强了在复杂场景下的适应能力。

在无人机航拍图像目标检测领域,由于受天气变化、光照条件、拍摄角度等多种因素的限制,从单一场景中采集的图像往往存在样本数量不足和质量不高的问题。这些问题限制了检测算法在各种环境条件下对多样性目标特征的全面学习,进而降低了模型的泛化能力,使其在面对复杂多变的检测场景时表现不佳。采用数据增强策略,通过对数据集中的原始图像数据进行缩放、裁剪、旋转、加雾、随机拼接等一系列操作,可以在不增加数据采集成本的前提下拓展样本的分布范围,扩充数据集规模,从而帮助模型更好地适应各种复杂航拍场景,提升航拍图像目标检测性能。图 11 展示了常用数据增强策略对图像进行处理后的结果。

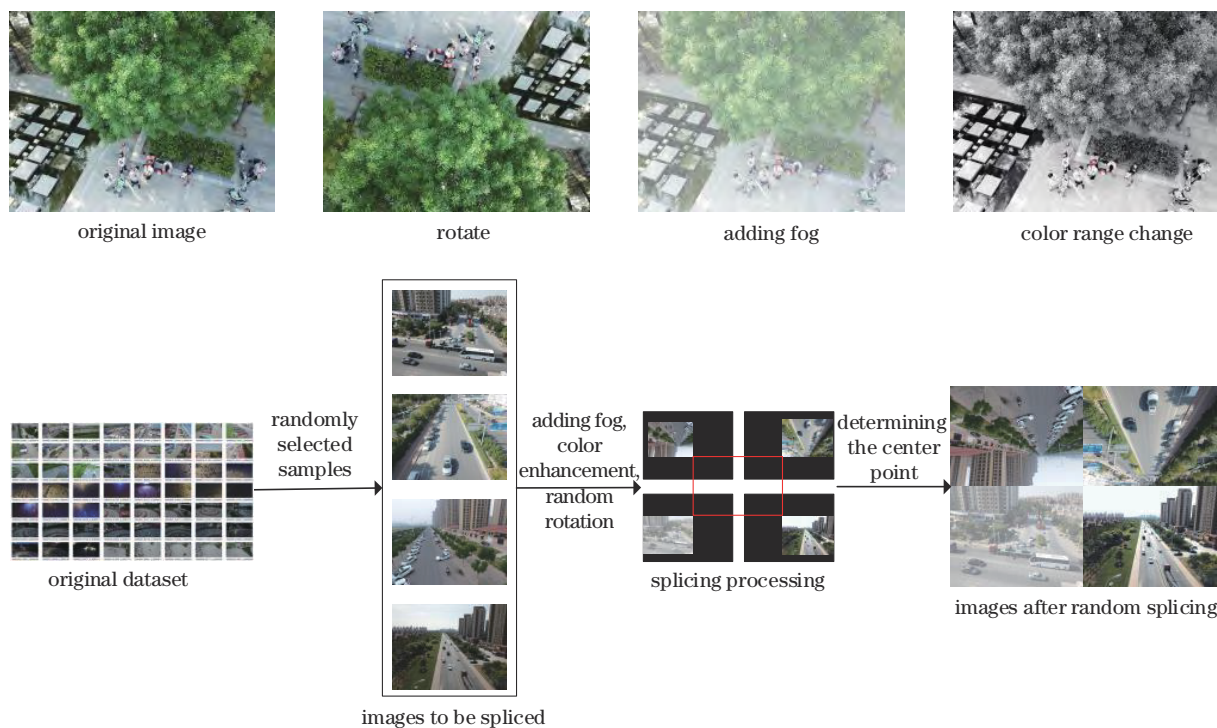


图 11 常用数据增强策略对图像处理的结果

Fig. 11 Results of image processing using common data augmentation strategies

在无人机视角下的目标识别与分类任务中,采用多样化的数据增强手段丰富和扩展航拍图像数据集,提高模型的鲁棒性和泛化能力,已经成为增强模型表现力的关键策略。Zhu 等^[30]将随机缩放、裁剪、平移和旋转等多种传统数据增强技术与 Mixup、Mosaic 等前沿技术相结合,实现了对数据集的进一步扩充,有效增强了模型在应对目标尺寸剧烈变化时的适应能力,缓解了图像几何失真问题。Zhang 等^[31]为了解决无人机航拍图像数据集中普遍存在的亮度失衡、数据分布不

均以及目标尺寸较小的问题,提出了 ECopy-Paste 方法。该方法通过巧妙地结合图像亮度转换、复制-粘贴、大规模抖动等方法,在扩展训练数据集的基础上进一步提高了模型在识别小尺寸目标时的准确率和鲁棒性。Gwon 等^[32]利用生成对抗网络进行图像间的转换,专注于修复和重建无人机航拍图像中可能出现的损伤细节,生成高质量的 RGB 图像数据,为航拍图像的分析 and 应用提供更为精确和可靠的保障。此外,为了解决图像识别模型在处理具有相似背景的图像时普

遍存在泛化能力不足的问题,周璇等^[33]在 Mosaic 数据增强方法的基础上融入混合数据增强策略。这一策略的处理流程为:首先设定图像的输出尺寸,随后随机划定分割线,对选定的图像进行裁剪、缩放处理和拼接操作,最终采用像素级混合技术创建新的训练样本。这一策略极大地丰富了数据集中的目标特征信息,增强了模型对复杂场景图像的识别能力。

针对特殊气象条件下的航拍图像目标检测任务,杨慧剑等^[34]采用中心点合成雾图的方法对原始图像进行雾化处理。该方法的原理为:在图像中选择一个特定点作为雾的起始扩散点,同时确保雾效随着接近该中心点而逐渐增强,以此模拟真实场景中的雾景。此后,经过处理的雾图被用于训练模型,以提升无人机在雾天执行航拍图像目标检测任务时的鲁棒性。殷旭平等^[35]针对无人机在雨雾恶劣天气下捕获的图像细节丢失严重的问题,提出了风格迁移的无人机图像目标检测方法,该方法引导模型更专注于学习目标物的轮廓特征,有效提升了无人机在雨雾天气条件下作业时的目标识别能力。为了有效应对无人机在极端气候条件下巡检时由紊流导致的图像抖动问题,陈奎等^[36]采取了一系列数据增强手段对图像进行镜像、裁剪、缩放、旋转以及高斯模糊等操作,并结合 Mosaic-9 技术对图像数据进行深入处理,有效缓解了图像中存在的目标尺度不平衡的问题,显著提升了无人机在复杂环境中执行检测任务时的适应能力。为解决无人机视角下捕获的图像运动模糊的问题,唐勋昊等^[37]提出了运动模糊图像复原的目标检测策略。该策略运用条件生成对抗网络实现运动模糊图像的清晰化处理,并将高性能的 MobileViT 作为生成器的核心特征提取架构,有效增强了模型对图像背景信息的辨识能力以及对局部细节特征的还原能力。

3.2 注意力机制

在实际应用场景中,无人机航拍覆盖的范围较广,待检测目标往往只占据图像中的一小部分,并且常常伴有遮挡问题,这给模型在复杂环境下对目标对象进行准确分类和定位带来了极大挑战。为了应对这一挑战,研究人员引入了注意力机制,用于模拟人类视觉系统中的注意力聚焦过程,使模型能在复杂背景中自动聚焦于关键区域,减少对非关键区域的资源分配。这种策略不仅有助于模型更精细地捕获目标对象的特征细节,如形状、颜色、尺寸,还能使其更全面地掌握目标之间的上下文信息,提高无人机航拍图像目标检测的效率和精确度。

随着无人机图像处理技术的不断进步,众多融合了注意力机制的先进方法应运而生。这些方法不仅提高了图像处理的速度和准确度,还进一步增强了无人机视觉系统对场景的理解能力。Cai 等^[38]为了克服无人机航拍图像中背景噪声干扰以及目标特征提取不足等挑战,基于特征金字塔提出了引导注意力网络。通

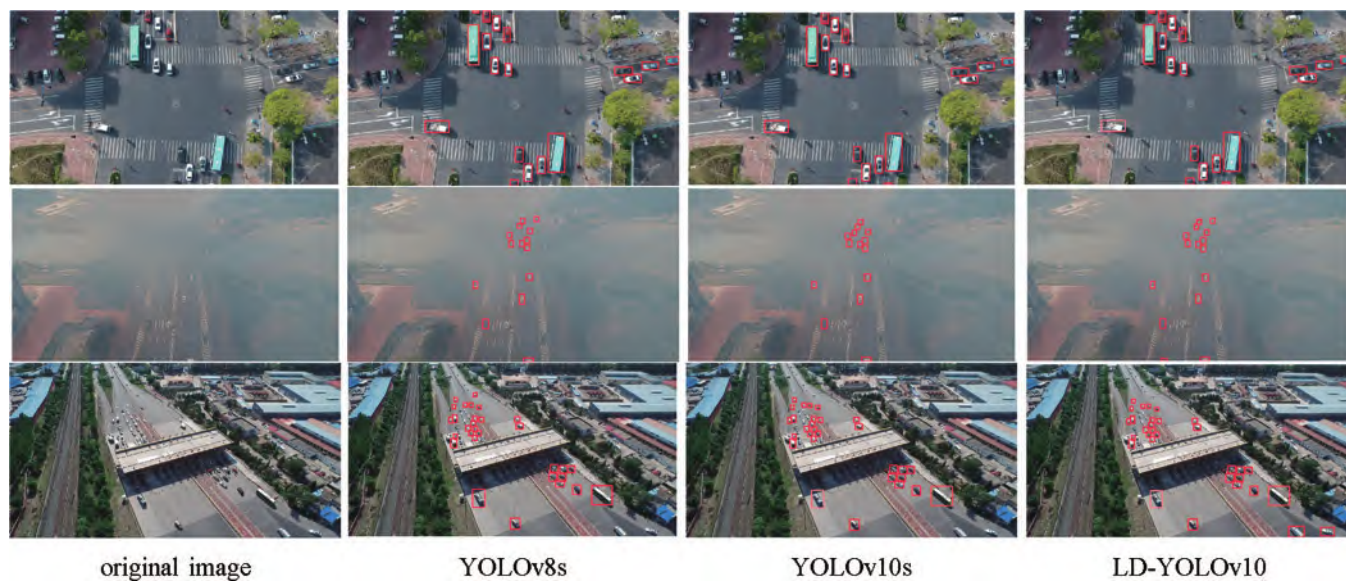
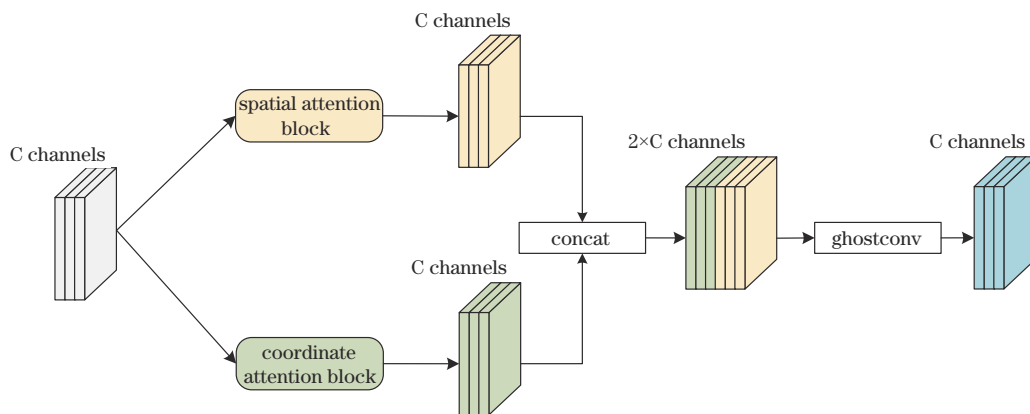
过前景注意力模块,该网络可以更全面地考虑目标对象的整体与局部特征,有效增强了模型在目标检测时对背景信息的提取能力。此外,为应对无人机目标检测场景中由光线、硬件设备等限制引起的图像模糊和失真问题,常凯旋等^[39]提出了一种融合 BiFormer 自注意力机制的 BRA 模块。该模块遵循稀疏性原则,在分析图像时更专注于处理与检测任务紧密相关的关键区域,从而自适应地增强输入信息的语义关联性,有效提升了模型的表征能力。为了使网络更高效地提取航拍图像序列中不同位置的细节特征,Qiu 及其团队^[40]在 YOLOv10 网络架构^[41]中融入 AIFI 模块提出了 LD-YOLOv10 算法。AIFI 模块利用自注意力机制深入挖掘高级特征信息,将特征图转换为特征向量后送入多头注意力机制和前馈网络进行处理,从而使得网络中的每个注意力单元能够独立地聚焦于不同的特征子空间,极大地提升了模型对输入特征多维度信息的捕捉能力。不同算法在 VisdroneDET-2021 和 UAVDT 数据集上的实验对比结果验证了 LD-YOLOv10 算法的显著优势,如图 12、图 13 所示。

为了灵活应对车辆重识别场景中由拍摄角度、目标状态变化所引起的目标遮挡、低辨识度以及尺寸多变等系列问题,Yin 等^[42]提出了一种双重混合注意力机制。该机制通过整合空间和通道两个维度上的注意力,有效捕获了像素级对应关系和通道间的依赖性,确保模型在不同尺度的车辆检测中均能保持良好的性能。针对无人机视角下的目标检测任务面临的目标分布高度密集以及复杂背景干扰等挑战,Zhang 等^[43]设计了全局-局部特征引导(GLFG)模块。该模块通过融合无人机航拍图像中的全局视野和局部细节信息,使模型更专注于对关键目标区域的处理,有效增强了模型在复杂场景中准确识别和定位不同尺度、密度目标对象的能力。此外,为了增强模型在处理航拍图像时对细节特征的提取能力,Liu 等^[44]构建了全新的空间坐标自注意力模块(SCSA),如图 14 所示。SCSA 通过修改、优化空间注意力和坐标注意力,并将它们结合形成新的注意力模块,实现了空间、位置和通道信息的充分融合,有效增强了模型对小目标对象位置信息的感知能力以及对细节特征信息的捕获能力。

上述先进方法在无人机图像处理领域的集成与应用极大地拓宽了注意力机制的适用范畴,为航拍图像目标检测提供了多样化的技术手段。这些技术相互促进,共同推动了图像处理技术的深入发展及创新,同时也为无人机应对复杂环境下的诸多挑战开辟了多元化的技术路线,助力该领域的持续进步与发展。

3.3 特征融合机制

受无人机飞行高度变化的影响,航拍视角下的目标对象尺寸会表现出极大的多变性。当无人机在较低空域作业时,目标在图像中通常占据较大的像素区域面积,并且细节特征信息表现丰富。然而,随着飞行高

图 12 不同算法在 VisdroneDET-2021 数据集上的可视化对比^[40]Fig. 12 Visualization comparison of different algorithms on VisdroneDET-2021 dataset^[40]图 13 不同算法在 UAVDT 数据集上的可视化对比^[40]Fig. 13 Visualization comparison of different algorithms on UAVDT dataset^[40]图 14 SCSA 模块示意图^[44]Fig. 14 Schematic diagram of spatial-coordinate self-attention module^[44]

度的增加,同一目标对象在航拍图像中所占据的像素区域逐渐变小,能呈现出的细节特征信息也随之减少。这种由飞行高度变化导致的目标尺寸多变性,使得仅采用单一尺度的特征提取方法难以全面捕获图像关键区域的所有细节。多尺度特征融合机制能确保模型同

时考虑目标的局部细节特征和全局上下文信息,有效提高检测的准确性和鲁棒性,进而使得模型能够在复杂多变的航拍场景中实现更精准的目标检测。图 15 是目前主流的特征融合网络结构示意图,为深入解析该技术的内部机制提供了直观的学术参考。

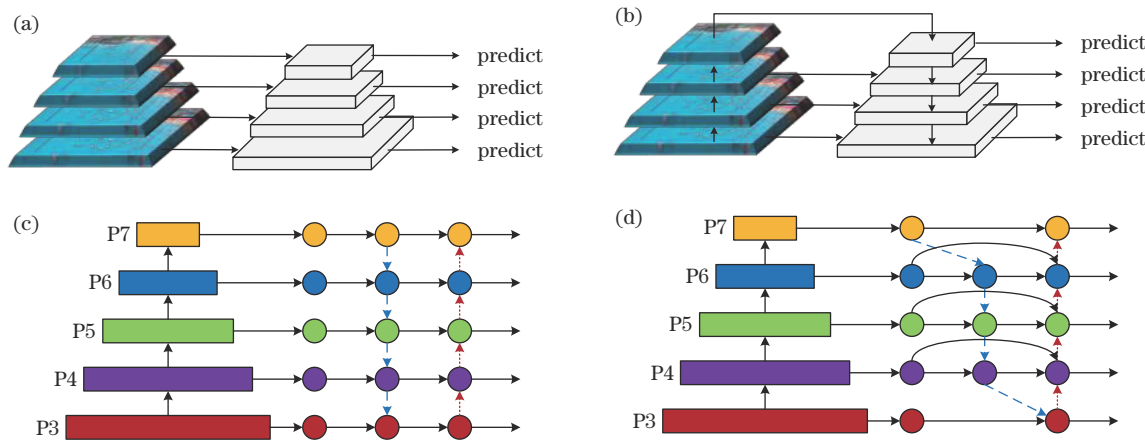


图 15 主流的特征融合网络结构示意图。(a)图像金字塔结构示意图;(b)FPN结构示意图;(c)PANet结构示意图;(d)BiFPN结构示意图

Fig. 15 Schematic diagrams of mainstream feature fusion network structures. (a) Image pyramid structure diagram; (b) feature pyramid network (FPN) structure diagram; (c) PANet structure diagram; (d) BiFPN structure diagram

在无人机航拍图像目标检测任务中,精细优化现有的特征融合策略或创新性地发展新型融合网络架构已成为现阶段的研究热点。这种方法能够帮助模型提升其在处理航拍图像时对目标尺寸变化的适应能力,使其较好地融合不同尺度的目标特征信息。Hong等^[45]为了克服无人机航拍图像中微小目标检测困难的问题,设计了一种新型的尺度选择金字塔网络(SSPNet)。该网络通过结合上下文注意模块、尺度增强模块和尺度选择模块,有效控制相邻层的数据流,并抑制FPN中梯度计算的不一致性,实现了不同尺度特征信息的高效融合,为微小目标在复杂背景下的精确识别提供了强有力的技术支撑。尽管SSPNet在多尺度特征融合上取得了一定进展,但当其在计算资源有限的无人机边缘设备上运行时,计算效率仍有较大的提升空间。鉴于此,基于图15(b)所示的传统FPN框架,Min等^[46]进一步进行优化和扩展,创新性地提出了特征金字塔网络E-FPN。该网络通过引入额外的F2级特征金字塔、残差网络分支以及CBAM(convolutional block attention module)模块^[47],不仅有效地解决了无人机边缘设备计算资源受限的难题,还增强了模型对航拍图像中目标尺寸频繁变化的适应能力。

如图15(c)所示,PANet(path aggregation network)^[48]作为FPN的进阶版本,通过引入额外的路径来增强特征传递,能够进一步解决FPN在处理复杂场景时遇到的挑战。Zhao等^[49]基于PANet网络结构提出了以内容感知为核心的多尺度网络结构C3-PANet。该网络

利用上采样CARAFE(content-aware reassembly feature enhancement)^[50]技术对特征图进行预测和重组,拓宽了模型的感受野,增强了其对目标特征的提取效果。Cheng等^[51]通过引入坐标注意力对PANet进行改进,构建了CA-PANet网络。该网络能够更好地促进低维和高维特征的融合,有助于模型在处理无人机视角下的多尺度目标时更精准地识别其形状、尺寸、纹理等关键信息,降低误检、漏检率。

尽管基于PANet的优化策略在提升检测精度方面取得了一定效果,但其单一方向的特征融合方式在执行多层级特征融合时,可能无法充分挖掘和利用各层级的特征信息,从而限制了其对复杂场景的检测性能。为了克服这一局限,研究人员提出了图15(d)所示的双向特征金字塔网络(BiFPN)^[19],并将其广泛应用于实际任务中。Zhang等^[52]将BiFPN网络结构应用于无人机航拍图像目标检测任务中,通过对多个特征图进行加权处理确保了特征信息在网络的不同层级间实现双向传递和高效融合,进而增强了模型在无人机航拍图像检测中对目标尺寸变化的适应能力。张瑞芳等^[53]设计了双向加权多尺度动态网络结构BiEO-Neck,其在加权特征融合网络中采用双线性插值对下采样后的特征图进行上采样,有效增强了模型对浅层语义信息的表达能力,以及对小目标对象形状轮廓等细节信息的感知能力。此外,针对航拍图像检测中目标对象像素区域小、分布密集且尺度变化明显等问题,Jiang等^[54]提出了一种新的双向密集特征金字塔网络(BDFPN)。该网络通过扩大特征金字塔网络规模并

引入跳跃连接,实现了浅层到深层多尺度特征信息的深度融合,进一步提升了模型在提取小目标细粒度特

征信息时的工作效率,有效增强了检测系统的整体性能。BDFPN 的网络结构如图 16 所示。

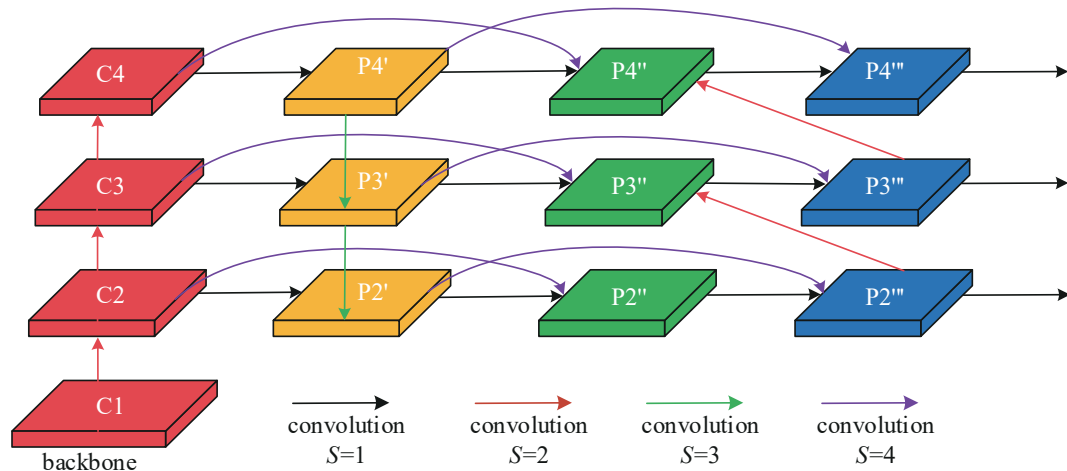


图 16 BDFPN 结构示意图^[54]

Fig. 16 Schematic diagram of BDFPN structure^[54]

上述先进的特征融合技术均是以 FPN、PANet、BiFPN 等多种流行算法为基础,针对无人机航拍图像目标检测中所面临的特定挑战,提出的创新性的优化方案。尽管如此,不同的网络架构在特定环境下执行检测任务时,仍显现出了一定的局限性。为了有效应对复杂多变的航拍任务需求,在实际应用中,研究人员需要根据不同的航拍场景灵活地选择并优化特征融合技术,才能更好地提升无人机航拍图像目标检测的性能和鲁棒性。

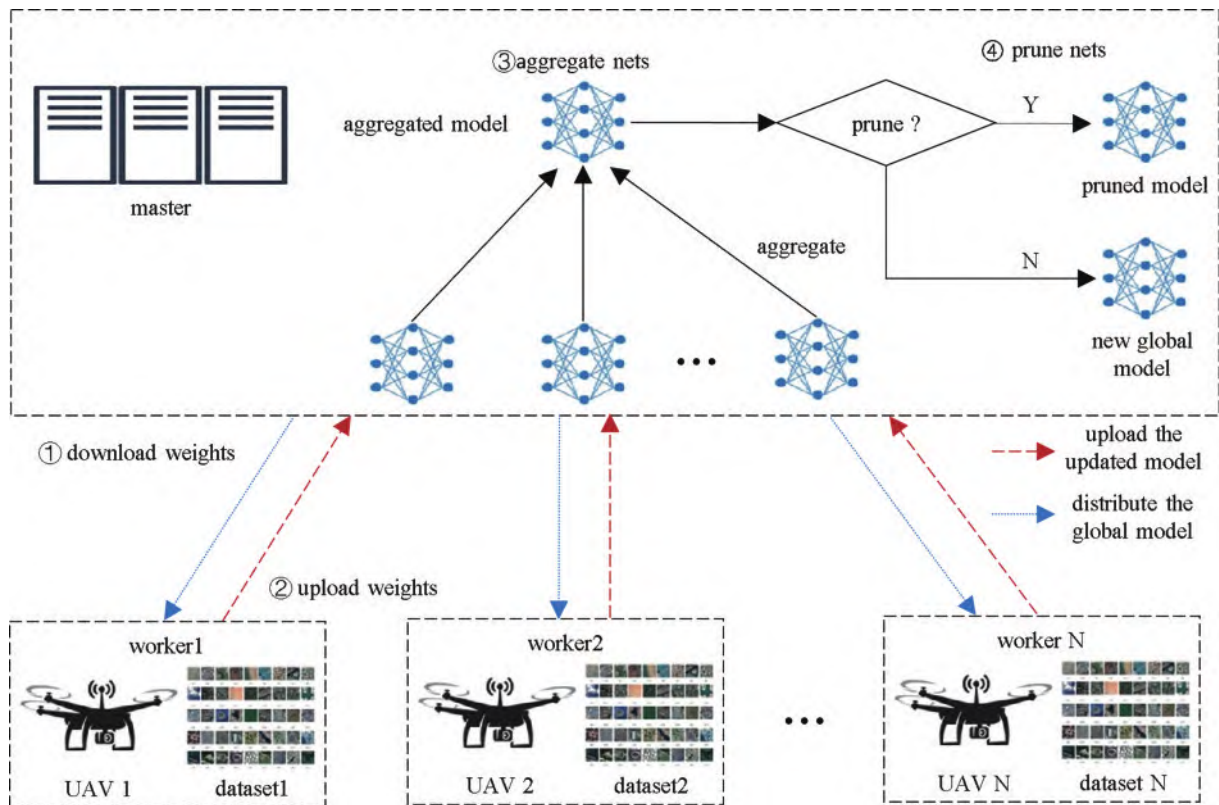
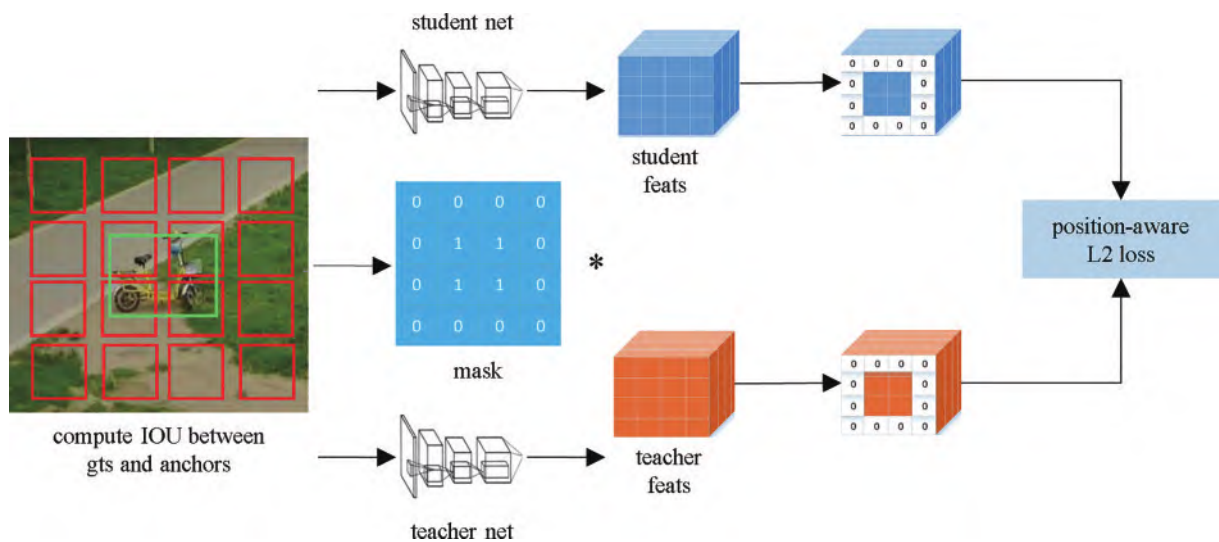
3.4 模型轻量化

无人机平台普遍面临着算力资源不足的问题,这一问题导致难以直接使用计算繁杂的深度学习模型对图像数据进行实时处理。为此,研究人员采用剪枝、知识蒸馏以及优化网络架构等多种轻量化策略对现有的深度学习模型进行优化,以达到降低资源消耗、提升处理速度的目的,从而确保无人机在执行图像检测任务时能够高效运作。

在无人机航拍图像目标检测领域,剪枝技术是一种被众多研究者青睐的高效的神经网络优化手段,其主要作用是剔除网络中的冗余神经元或对整体性能贡献微小的连接和神经元,从而降低模型的参数数量和计算复杂度,实现模型压缩和加速,方便其在资源受限的无人机平台上高效部署和运行。针对在资源有限的无人机群体中实施联邦学习的挑战,Song 等^[55]提出一种创新的剪枝联邦学习方法。该方法基于跨网络所有层的滤波器重要性度量策略,将一次性滤波器剪枝技术巧妙地融合到标准的联邦学习流程中,在全局网络层面上更精准地衡量滤波器的重要性,实现了一种无需设备间数据交换的高效通信技术。基于该技术,模型的资源消耗大幅减少了 88.4%。剪枝联邦学习原理如图 17 所示。针对无人机边缘设备上小麦穗计

数模型部署困难这一挑战,刘旭等^[56]基于 YOLOv5s 模型提出了一种创新性的轻量化方法。该方法将剪枝计算与知识蒸馏相结合,不仅维持了算法在麦穗计数上较高的准确率,还显著降低了模型对算力和存储资源的需求,为无人机的实时麦穗计数功能提供了可靠的技术保障。

知识蒸馏技术是另一种流行的模型精简方法,其巧妙地将大型复杂教师网络的深厚知识传递给一个更加小巧且高效的学生网络,使得学生网络不仅继承了教师网络的丰富经验,还在保持较高识别精度的基础上显著减小了模型体积,同时降低了对计算资源的需求。该技术的广泛应用为无人机视角下的航拍图像分析和处理提供了新的可能性。然而,若在特征图层面直接实施知识蒸馏策略,模型所实现的性能提升可能比较有限。为此,Liu 等^[57]提出一种自适应关键蒸馏位置(AKDP)方法,其具体实现原理如图 18 所示。该方法首先采用基于锚点的特征图蒸馏技术计算预设锚点与真实边界框之间的 IOU,通过 IOU 确定特征掩码,进而指示教师网络和学生网络特征图中的关键区域;随后,利用位置感知的 L2 距离测量教师网络与学生网络之间的差距,实现更高效的知识传递,从而显著提升模型的性能表现。此外,Xiao 等^[58]提出了一种名为 KD-RNet 的知识蒸馏方法,该方法通过协同训练策略将教师网络的类别和位置信息传递给学生网络,不仅有效提升了模型的检测精度和重叠度,还成功实现了对航拍图像旋转检测模型的压缩,减小了模型的体积和复杂度。Zhou 等^[59]为优化模型在无人机上的部署效率,采用邻域协同蒸馏技术,在特征细化前后的关键环节引入邻域协作机制,融合不同层级教师网络的特征信息,有效克服了传统一对一特征细化过程中易丢失

图 17 联邦学习与模型剪枝的原理图^[55]Fig. 17 Principle diagram of federated learning and model pruning^[55]图 18 AKDP算法原理图^[57]Fig. 18 Principle diagram of the adaptive key distillation position algorithm^[57]

重要附加信息的问题,确保了从教师网络到学生网络的知识全面迁移。轻量化后的学生网络模型与轻量化前相比,运算速度提升了92%,参数数量减少了83%。

除了知识蒸馏技术,大量学者也正积极投身于网络架构轻量化方法的研究与探索。这些方法专注于通过重构网络结构来降低模型的复杂性以及对资源的需求,从而确保在图像检测准确性不受影响的前提下,提升无人机在实时图像处理上的性能。在无人机森林防

火监测领域,Zhou等^[60]提出了一种轻量化的无人机监测方法。该方法通过将YOLOv5模型的骨干网络替换为MobileNetV3,并结合半监督知识蒸馏的训练策略,使模型体积从107.6 MB大幅降低至6.3 MB,同时提升了模型的检测精度。为了进一步优化无人机平台的实时检测性能,Cao等^[61]提出一种适用于无人机平台的轻量级实时检测算法MELF-YOLOv5-X。该算法将MobileNetV3作为YOLOv5的骨干网络,同时通过融合高效的通道注意力机制以及添加额外

的预测头模块进一步缓解模型的计算压力,确保了无人机在执行实时监测任务时的响应速度和检测精度。在无人机红外目标检测领域,Zhao 等^[62]基于轻量级网络 GhostHGNetV2 设计了无人机红外目标检测算法 ITD-YOLOv8。该算法引入了 VoVGSCSP 模块,该模块通过融合全局上下文信息和多尺度特征来增强颈部结构,提升模型的感知能力。同时,该算法采用了 AXConv 轻量级卷积技术,用不同大小的卷积核替换传统的固定尺寸的卷积核,实现了更高效、灵活的卷积处理。在增强模型对红外目标感知能力的基础上,该算法使参数量减少了 41.9%,运

算量减少了 25.9%。针对航拍图像中绝缘子缺陷检测的特定需求,黄福珍等^[63]提出了 GCP-YOLOv8s 无人机绝缘子缺陷检测算法。该算法用 GSConv 替换标准卷积操作,同时在 C2f 模块中引入 FasterNet 模块,有效解决了小缺陷特征提取困难的问题,并成功地将模型尺寸压缩至 1.47×10^7 ,将参数量降低至 7.2×10^6 ,极大地增强了该算法在实际应用中的部署灵活性。

表 1 展示了在当前无人机航拍图像目标检测的应用场景下,研究人员采用不同轻量化方法改进网络结构的对比分析。

表 1 不同轻量化方法改进网络结构的对比分析

Table 1 Comparative analysis of network structure improvements using different light-weight methods

Ref.	Light-weight method	Network structure	Improving strategy	Improvement effect	Dataset
[56]	Pruning and knowledge distillation	YOLOv5s	Training of pruned student models using L1 strategy for model parameter pruning and knowledge distillation by outputting feature knowledge types	Model size is reduced by 76% and the number of parameters is reduced by 79.61%	Wheat dataset collected by drone
[59]	Knowledge distillation	MJPNet-T	Using MJPNet-T as the teacher network, apply the neighborhood collaboration distillation method to perform multi-level feature distillation on the MJPNet-S student model for model light-weighting	The model's computing speed is increased by 92% and the number of parameters is decreased by 83%	DroneRGBT, RGBT-CC
[60]	Replacing backbone network and knowledge distillation	YOLOv5	Replacing the backbone network of YOLOv5 with MobileNetV3 structure and training the model by combining semi-supervised knowledge distillation strategy	Model volume is reduced from 107.7 MB to 6.3 MB and computation is reduced to 7.8 FGLOPs	FLAME
[61]	Replacing backbone network	YOLOv5	Replacing of YOLOv5s backbone network with MobileNetV3 architecture and introducing of channel attention mechanism and additional detection head	Compared with YOLOv5s, the number of parameters and operations are reduced by 51.4% and 38.6%, respectively Compared with YOLOv5l, the number of parameters and operations are reduced by 87.4% and 47.4%, respectively	VisDrone-DET-2021
[62]	Replacing backbone network	YOLOv8n	Replacing YOLOv8 backbone network with GhostHGNetV2 and replacing regular convolutional blocks with AXConv light-weight convolutional operations	The number of parameters and operations are reduced by 41.9% and 25.9%, respectively	HIT-UAV
[63]	Replacing backbone network	YOLOv8s	Introducing GSConv to replace the normal convolution in the network and replacing the Bottleneck module in C2f with a FasterNet block module	The model's parameters have been reduced to 7.2×10^6 , a decrease of 36.3%. Its size has been compressed to 1.47×10^7	CPLID

综上所述,轻量化技术的研究与探索不仅极大地推动了无人机实时检测技术的进一步发展,还为多样化的检测场景贡献了丰富的实践经验与高效的解决策

略。这些创新性的方法在增强无人机航拍图像目标检测上实现了效率与精度的双重提升,已经成为推动行业发展的关键驱动力。

4 无人机航拍图像目标检测数据集

数据集资源作为目标检测研究领域的关键基石,直接影响着模型的最终训练成效和性能评价。为确保算法在特定应用场景中展现出最佳性能,研究人员需精心挑选与场景需求高度契合的数据集进行实验,以优化模型的检测能力。

在航拍图像目标检测领域,无人机高空作业的独特优势使其能够从多个角度捕捉地面目标,实现对复杂场景的全方位监控,可为地理信息测绘、交通流量管理、城市空间规划以及灾害应急响应等提供高效且实时的解决策略。然而,要保障检测算法在复杂环境中的高效识别能力,选择高质量的数据集进行模型训练与调优尤为关键。这些数据集不仅需要涵盖丰富的目标类别和多样化的环境条件,还要充分考虑实际检测过程中可能会遇到的噪声、遮挡、光照变化等影响,以确保模型训练的鲁棒性和泛化能力。现有的无人机航拍图像数据集大多经过了严格筛选与预处理,具备精确的标注信息和较高的数据质量。本研究在充分调研相关学术文献及专家研究成果的基础上,对无人机航拍图像目标检测过程中使用的主流数据集进行具体分析和阐述,以期为进一步的研究和实践提供参考。

1) Stanford Drone 数据集

Stanford Drone 数据集^[64]是由斯坦福大学计算机视觉与几何实验室精心打造的,是一个专注于室外场景的无人机航拍数据集。该数据集采用无人机俯瞰的视角,收集了 8 个不同场景下的目标数据,涵盖 Bicyclist、Pedestrian、Skateboarder、Cart、Car 和 Bus 共 6 个不同类别的标注信息,每个目标均通过二维边界框进行标注,并附有类别标签和唯一识别码。为了便于区分,不同类别的标注框使用了不同的颜色。此外,该数据集具备目标轨迹预测和多重目标跟踪功能,在计算机视觉研究领域发挥着重要作用,为智能交通管理、智能安防监控等领域提供了宝贵的数据支持。

2) CARPK 数据集

CARPK 数据集^[65]是由 Hsieh 等研究者在无人机视角下构建的停车场数据集,旨在为车辆检测与计数研究提供一个大规模、高精度的数据资源。该数据集包含了 1573 张图片,覆盖了 4 个独特的停车场。他们采用精确的边界框对 89777 辆汽车进行了标注。此外,该数据集是通过 PHANTOM 3 PROFESSIONAL 无人机在约 40 m 的飞行高度采集得到的,图像分辨率达到了 1280 pixel×720 pixel,确保了从空中视角捕获的车辆图像具有较高的清晰度,对于推动智能交通系统领域车辆检测和计数技术的发展与应用具有重要意义。

3) UAVDT 数据集

UAVDT 数据集^[66]由中国科学院的 Due 等提出,是一个专注于无人机视觉领域的大规模数据集。该数据集涵盖了目标检测(DET)、单目标跟踪(SOT)以及

多目标跟踪(MOT)三个关键的计算机视觉应用方向,包含从 10 h 原始视频中精选的约 80000 个代表性视频帧,而且每个帧都带有手动标注的边界框和详细标签(如车辆类别和遮挡情况等)。此外,该数据集还包括轿车、卡车、大巴车三个目标类别,涉及广场、街道、收费站、高速公路、交叉路口和 T 形路口等城市地区的多种复杂场景,图像分辨率为 1080 pixel×540 pixel,适用于无人机智能交通管控等领域的应用研究。

4) VisDrone 2019 数据集

VisDrone 2019 数据集^[67]由天津大学机器学习和数据挖掘实验室的 AISKEYEYE 团队收集发布,涵盖 10209 幅静态图像,涉及 pedestrian、people、bicycle、car、van、truck、tricycle、awning-tricycle、bus、motor 等 10 个目标类别。这些图像不仅覆盖了城市街道、乡村道路、公园、桥梁等多种不同的场景,还涵盖晴天、多云、阴天和雨天等不同的天气状况和光照环境,为无人机视觉研究提供了丰富的数据资源。

5) AU-AIR 数据集

AU-AIR 数据集^[68]是一个专为低空交通监控设计的多模态无人机感知数据集。该数据集由 8 个视频序列构成,共计提取了 32823 个高清帧,图像最大分辨率为 1920 pixel×1080 pixel。这些序列覆盖了人类、轿车、公共汽车、面包车、卡车、自行车、摩托车及拖车等 8 种目标类别,融合了无人飞行器在航拍过程中记录的多传感器信息(如时间戳、GPS 坐标、海拔信息、IMU 数据及飞行速度),为静态图像和动态视频流中的目标检测与跟踪任务提供了丰富的数据资源,促进了计算机视觉和机器人算法在自主空中监视方面的应用。

6) DroneCrowd 数据集

DroneCrowd 数据集^[69]是由 Wen Longyin 团队于 2021 年提出的拥挤人群数据集,主要用于评估在无人机航拍视角下执行目标检测、人群计数、人群密度估计等任务的算法的性能。该数据集由 112 个视频片段组成,共计 33600 幅分辨率为 1920 pixel×1080 pixel 的高清图像,每幅图像都经过精细的头部标注和轨迹标注处理。其中,超过 480 万个头部标注点被用于计算密度图和人群计数,20800 人的轨迹标注被用于行人重识别和运动轨迹分析。该数据集提供的丰富元数据信息,如拍摄时间、地点以及天气条件等多样化因素,为研究者在不同环境因素影响下探讨人群行为模式提供了全方位的考量与数据支撑。

7) FLAME 森林火焰数据集

FLAME 森林火焰数据集^[70]是由北亚利桑那大学联合其他机构精心打造的一套森林火灾图像数据集。该数据集利用多种不同型号的无人机,在亚利桑那州松树林的预定燃烧活动中采集了大量的火灾图像资料,涵盖了 254 pixel×254 pixel、1280 pixel×720 pixel、3480 pixel×2160 pixel 等多种分辨率的视频记录以及

热成像图片,并对火焰和非火焰帧进行了详细标注。该数据集旨在支持森林火灾识别与监控预警等相关领域的研究,为火情图像的二分类识别和精确分割火灾边界提供重要的数据支持。

8) DroneVehicle 数据集

DroneVehicle 数据集^[71]是由天津大学的 Sun 团队精心收集并发布的大型数据集,专注于处理无人机视角下的 RGB-红外多模态车辆检测与计数任务。该数据集由 56878 幅分辨率为 840 pixel×712 pixel 的图像组成,其中 RGB 图像和红外图像各占一半,涵盖了轿车、卡车、大巴车、面包车、货车等 5 种目标类别。这些图像记录了从白昼到黑夜多种光照条件下的拍摄环境,并充分考虑了实际场景中可能出现的目标遮挡、尺度变化等问题,对于推动无人机视觉领域的智能城市交通管理研究具有重要意义。

9) SeaDronesSee 数据集

SeaDronesSee 数据集^[72]由 Leon Amadeus Varga 团队精心打造,是一个专注于海上搜索与救援(SAR)应用的大规模无人机图像库。该数据集涵盖了目标检测、单目标跟踪以及多光谱目标检测三大核心领域,主

要提供了 5~260 m 不同飞行高度、0°~90°各异视角的航拍图像数据,全面记录了不同天气和光照条件下船只、岛屿等多样化形态的海上元素。这些海上元素主要标注 ignored、swimmer、boat、jetski、life_saving_appliances、buoy 等 6 个类别的目标信息,尤其是对人类的位置和状态进行了精确标注,对于开发支持 SAR 任务的复杂自主无人机系统具有重要意义。

10) HIT-UAV 数据集

HIT-UAV 数据集^[73]是由哈尔滨工业大学(HIT)研究团队提出的首个公用高空无人机红外热成像数据集。该数据集涵盖学校、停车场、道路和操场等多样化场景,覆盖 Person、Car、Bicycle、OtherVehicle、DontCare 等 5 类对象标签,共计 2898 幅红外热成像图片。每幅图像均提供了无人机作业过程中的飞行高度、相机视角、拍摄日期和日光强度等关键信息,而且图像中的对象实例均采用定向和标准两种类型的边界框手动标注,能有效应对航拍图像中的对象实例严重重叠时的识别难题,对于推动红外图像处理和智能分析等领域的研究与发展具有重要意义。

无人机航拍图像常用数据集总结如表 2 所示。

表 2 无人机航拍图像常用数据集

Table 2 Common datasets for aerial images captured by UAV

Dataset	Application scenario	Category count	Content introduction	Release year	Data source	Link
Stanford Drone	Trajectory prediction and object detection and tracking tasks under multiple target categories	6	Video data of 20000 object trajectory interaction information	2016	Eight crowded scenarios such as college campuses, neighborhoods, etc.	https://cvgl.stanford.edu/projects/uav_data/
CARPK	Vehicle detection and counting tasks in parking lots	1	89777 cars in 1573 parking lot images	2017	Four different parking lots	https://lafi.github.io/LPN/
UAVDT	Intelligent traffic control, vehicle tracking and detection tasks	3	Approximately 80000 representative video frames selected from a 10-hour video	2018	Plazas, streets, highways, intersections, and other urban areas	https://sites.google.com/view/grli-uavdt/
VisDrone 2019	Object detection tasks in complex scenes	10	288 video clips, 261908 video frames, and 10209 static images	2019	Urban streets, rural roads, parks, bridges, and other scenes under various weather conditions	https://github.com/five-days/VisDrone-Dataset
AU-AIR	Object detection tasks under low-altitude traffic surveillance	8	32823 high-definition frames and 132034 object instances extracted from 8 video sequences	2020	Real outdoor scenes images collected by multi-modal sensors	https://bozcani.github.io/auairdataset
DroneCrowd	Object detection and tracking tasks, crowd counting and density estimation	5	A total of 112 videos and 33600 high-definition frames	2021	Urban streets, public squares, and other diverse scenes	https://github.com/VisDrone/DroneCrowd

表 2 (续)

Dataset	Application scenario	Category count	Content introduction	Release year	Data source	Link
FLAME	Binary classification of fire images and precise segmentation for fire boundary detection tasks	2	Videos of various resolutions recorded by different types of drones, as well as thermal imaging pictures	2021	Fire image data collected during the prescribed burning activity in the pine forest of Arizona	https://paperswithcode.com/dataset/flare
DroneVehicle	RGB-infrared multimodal vehicle detection and counting tasks	5	A total of 56878 images, with RGB and infrared images each accounting for half	2020	Images captured under varying lighting conditions, ranging from daylight to night-time	https://github.com/VisDrone/DroneVehicle
SeaDronesSee	Maritime search and rescue scene for object detection and tracking tasks	6	5630 training sets, 859 validation sets, 1796 test sets, and related maritime video clips	2022	Various scenes related to the entire rescue process under different weather and lighting conditions	https://seadronessee.cs.uni-tuebingen.de/
HIT-UAV	Object detection in thermal infrared imagery captured by UAV	5	2898 infrared thermal images extracted from 43470 video frames	2022	Diverse scenarios such as schools, parking lots, roads and playgrounds	https://gitcode.com/gh_mirrors/hi/hit-uav-Infrared-thermal-dataset

5 结 论

本文深入探讨了无人机航拍图像目标检测过程中面临的诸多挑战,对无人机平台资源限制、航拍环境对图像质量的负面影响,以及目标变化引发的检测难点进行了系统分析。同时,针对上述问题,从数据增强、注意力机制、特征融合机制和模型轻量化 4 个维度重点阐述了现有的能够优化无人机航拍图像目标检测的研究方法与优化思路。之后,系统介绍了无人机视角下的航拍图像目标检测数据集,并对不同数据集的具体特性进行横向比较,以探讨它们在不同检测任务中适用场景的差异性。虽然现阶段研究人员在无人机航拍图像目标检测领域取得了长足进步,但未来需要针对以下几点进行进一步深入探讨。

1) 复杂背景干扰问题

在无人机航拍图像目标检测领域,复杂背景下的遮挡物(如建筑物和树木等),往往会对目标识别与分类造成严重干扰。为应对此问题,研究人员采用了两种主要策略:一是模型结构的优化,通过改进模型架构来增强其在复杂环境中的检测性能。然而,这种优化方法常常会导致模型复杂度增加,不仅提高了运算成本,还可能引发过拟合问题,降低模型的泛化能力。二是数据集质量的提升,即通过增加大量相似场景的训练样本数量来改善模型对复杂背景的适应能力。然而,这种方法对大量特定场景标注数据的依赖性较强,不仅加重了人工标注的负担,而且在数据资源有限时,仍难以彻底消除复杂背景下的检测误差。

鉴于上述策略的局限性,未来的研究方向可着眼于

背景信息提取技术的探索。首先,通过滤波和去噪技术对图像进行预处理,降低随机噪声,提升图像质量。然后,运用背景建模技术,如基于像素或基于样本的方法,建立一个动态更新的背景模型。在这一过程中,可依据时空一致性原则,结合像素的历史信息和空间邻域信息,有效区分静态背景与动态前景。在完成背景建模后,采用特征提取和选择技术筛选出能够显著区分前景与背景的特征,如颜色、纹理、形状和运动模式等。接着,利用深度学习算法对这些特征进行训练,构建适用于复杂背景的目标检测模型。最终,在多种场景和遮挡程度下进行大量的实验评估,根据评估结果不断调整和优化背景建模参数,提升模型对复杂背景信息的解析能力。

2) 多模态融合

当前,无人机视角下的图像检测任务大多是基于单一视觉模态的检测方法开展的,这种方法获取的信息维度单一,而且易受天气、光照等复杂环境变化的影响,导致检测性能不够稳定。未来可考虑采用多模态融合技术,设计一个综合多传感器数据的多模态融合框架。首先,对雷达、可见光摄像头、红外设备等传感器收集的数据进行预处理,包括数据清洗、同步和对齐等,以确保数据的质量和一致性。接着,利用深度学习模型提取不同传感器数据的特征,以全面捕捉目标对象的形状、纹理、背景及环境温度等多维信息。特征提取后,引入注意力机制为这些特征分配权重,通过对模型进行训练来自动学习每种特征对于最终检测任务的

重要性(如夜间或低光照条件下,红外特征需要获得更高的权重)。随后,将加权后的特征进行融合,形成更具代表性的特征表示。最后,将融合后的特征输入检测网络(基于锚框的检测器或无锚框检测器)进行目标检测,并通过端到端的训练优化整个融合框架的性能,从而有效克服单一视觉模态在应对光照变化、恶劣天气等复杂情况时的局限性,提高无人机在复杂环境中的目标检测准确率和鲁棒性。

3) 分布式处理

无人机在高空作业时,其传统的工作模式主要依赖于单体无人机系统(UAS)。该系统配备有专门的传感器和处理器,以执行特定的目标检测任务。然而,这种单体架构普遍面临算力资源不足、单点故障风险、数据处理时效性差、环境适应能力弱以及执行任务灵活性低等问题。面对这些挑战,迫切需要将分布式计算架构引入无人机航拍图像的目标检测与识别领域。首先,采用多节点协同处理策略将复杂的检测任务分散至多个无人机节点,实现计算资源的负载均衡,减轻数据链路的传输压力和存储系统的负担。在此基础上,运用联邦学习等分布式数据处理算法,在确保数据隐私的前提下,实现节点间信息的高效整合。接着,设计故障容忍机制,确保在某个节点出现故障时,其他节点能够迅速接管其任务,保障系统的持续稳定运行。最后,通过实时性能监控和评估,根据系统响应速度和环境适应能力等关键指标,动态调整算法参数和任务分配策略,确保无人机在复杂多变的环境下保持高效和灵活的作业能力,实现资源的优化配置和任务的高效分配。

参 考 文 献

- [1] 韩建峰, 张妍. 基于无人机航拍路面的拼接算法研究[J]. 激光与光电子学进展, 2020, 57(20): 201003.
Han J F, Zhang Y. Research on stitching algorithm based on UAV based on aerial photography[J]. Laser & Optoelectronics Progress, 2020, 57(20): 201003.
- [2] 赵桂玲, 李鹏年, 郭泉荣, 等. 基于改进蚁群算法的无人机遥感影像森林类型分类[J]. 激光与光电子学进展, 2022, 59(16): 1628002.
Zhao G L, Li P N, Guo Q R, et al. Classification of forest types using UAV remote sensing images based on improved ant colony algorithm[J]. Laser & Optoelectronics Progress, 2022, 59(16): 1628002.
- [3] 段涛, 邓宇峰, 郭猛猛, 等. 无人机低空遥感技术在崩塌地质灾害动态监测中的应用[J]. 能源技术与管理, 2024, 49(3): 31-34.
Duan T, Deng Y F, Guo M M, et al. Application of low-altitude remote sensing technology of UAV in dynamic monitoring of collapse geological disasters[J]. Energy Technology and Management, 2024, 49(3): 31-34.
- [4] 邵明军, 刘树光, 李姗姗. 基于优化随机森林的对地攻击无人机自主作战效能评估[J]. 航空兵器, 2023, 30(6): 81-88.
- Shao M J, Liu S G, Li S S. Assessment of autonomous combat effectiveness of ground-attack UAV based on optimized random forest[J]. Aero Weaponry, 2023, 30(6): 81-88.
- [5] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [6] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[C]//Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR, December 8-14, 2001, Kauai, HI, USA. New York: IEEE Press, 2001: 511-518.
- [7] 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), June 20-25, 2005, San Diego, CA, USA. New York: IEEE Press, 2005: 886-893.
- [8] 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.
- [9] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [10] 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, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 580-587.
- [11] 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.
- [12] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1440-1448.
- [13] 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.
- [14] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.
- [15] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN [C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2980-2988.
- [16] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer International Publishing, 2016, 9905: 21-37.

- [17] 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.
- [18] Zhou X Y, Wang D Q, Krähenbühl P. Objects as points [EB/OL]. (2019-08-16)[2024-11-10]. <https://arxiv.org/abs/1904.07850>.
- [19] 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), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 10778-10787.
- [20] 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), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [21] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6517-6525.
- [22] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2024-11-10]. <https://arxiv.org/abs/1804.02767v1>.
- [23] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2024-11-10]. <https://arxiv.org/abs/2004.10934v1>.
- [24] Li C, Li L L, Jiang H L, et al. YOLOv6: a single-stage object detection framework for industrial applications[EB/OL]. (2022-09-07)[2024-08-09]. <https://arxiv.org/abs/2209.02976>.
- [25] 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]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 17-24, 2023, Vancouver, BC, Canada. New York: IEEE Press, 2023: 7464-7475.
- [26] Khanam R, Hussain M. YOLOv11: an overview of the key architectural enhancements[EB/OL]. (2024-10-23)[2024-11-10]. <https://ui.adsabs.harvard.edu/abs/2024arXiv241017725K>.
- [27] DeVries T, Taylor G W. Improved regularization of convolutional neural networks with Cutout[EB/OL]. (2017-08-15)[2024-11-10]. <https://arxiv.org/abs/1708.04552>.
- [28] Zhang H Y, Cisse M, Dauphin Y N, et al. Mixup: beyond empirical risk minimization[EB/OL]. (2017-10-25)[2024-11-10]. <https://arxiv.org/abs/1710.09412v2>.
- [29] Yun S, Han D, Chun S, et al. CutMix: regularization strategy to train strong classifiers with localizable features [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Republic of Korea. New York: IEEE Press, 2019: 6022-6031.
- [30] Zhu X K, Lyu S C, Wang X, et al. TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]//2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), October 11-17, 2021, Montreal, BC, Canada. New York: IEEE Press, 2021: 2778-2788.
- [31] Zhang L L, Liu X L, Tong Q. ECopy-Paste: an effective data augmentation method for object detection of UAV [C]//2022 International Conference on Big Data, Information and Computer Network (BDICN), January 20-22, 2022, Sanya, China. New York: IEEE Press, 2022: 469-474.
- [32] Gwon G H, Lee J H, Kim I H, et al. Image-to-image translation-based structural damage data augmentation for infrastructure inspection using unmanned aerial vehicle[J]. Drones, 2023, 7(11): 666.
- [33] 周璇, 葛琦, 邵文泽. 高分辨率特征增强的无人机航拍小目标检测[J]. 数据采集与处理, 2024, 39(4): 908-921.
- [34] 杨慧剑, 孟亮. 基于改进的YOLOv5的航拍图像中小目标检测算法[J]. 计算机工程与科学, 2023, 45(6): 1063-1070.
- [35] Yang H J, Meng L. A small target detection algorithm based on improved YOLOv5 in aerial image[J]. Computer Engineering & Science, 2023, 45(6): 1063-1070.
- [36] 殷旭平, 钟平, 薛伟, 等. 通过风格迁移的浓雾天气条件下无人机图像目标检测方法[J]. 航空兵器, 2021, 28(3): 22-30.
- [37] Yin X P, Zhong P, Xue W, et al. The method of UAV image object detection under foggy weather by style transfer[J]. Aero Weaponry, 2021, 28(3): 22-30.
- [38] 陈奎, 刘晓, 贾立娇, 等. 基于轻量化网络与增强多尺度特征融合的绝缘子缺陷检测[J]. 高电压技术, 2024, 50(3): 1289-1300.
- [39] Chen K, Liu X, Jia L J, et al. Insulator defect detection based on lightweight network and enhanced multi-scale feature fusion[J]. High Voltage Engineering, 2024, 50(3): 1289-1300.
- [40] 唐勋昊, 樊绍胜. 基于运动模糊图像复原的输电线路小目标检测算法[J]. 激光与光电子学进展, 2024, 61(22): 2212007.
- [41] Tang X H, Fan S S. Small-target detection method for transmission lines based on motion blurred image restoration[J]. Laser & Optoelectronics Progress, 2024, 61(22): 2212007.
- [42] Cai Y Q, Du D W, Zhang L B, et al. Guided attention network for object detection and counting on drones[C]//Proceedings of the 28th ACM International Conference on Multimedia, October 12-16, 2020, Seattle, WA, USA. New York: ACM Press, 2020: 709-717.
- [43] 常凯旋, 黄建华, 孙希延, 等. 基于双模态图像融合的无人机光学小目标检测算法[J]. 激光与光电子学进展, 2025, 62(4): 0428001.
- [44] Chang K X, Huang J H, Sun X Y, et al. Optical small

- target detection method by drone based on dual-mode image fusion[J]. *Laser & Optoelectronics Progress*, 2025, 62(4): 0428001.
- [40] Qiu X Y, Chen Y J, Cai W H, et al. LD-YOLOv10: a lightweight target detection algorithm for drone scenarios based on YOLOv10[J]. *Electronics*, 2024, 13(16): 3269.
- [41] Wang A, Chen H, Liu L H, et al. YOLOv10: real-time end-to-end object detection[EB/OL]. (2024-05-23)[2024-11-10]. <https://ui.adsabs.harvard.edu/abs/2024arXiv240514458W>.
- [42] Yin W J, Peng Y P, Ye Z C, et al. A novel dual mixing attention network for UAV-based vehicle re-identification [J]. *Applied Sciences*, 2023, 13(21): 11651-11667.
- [43] Zhang Y Z, Wu C Y, Zhang T, et al. Self-attention guidance and multiscale feature fusion-based UAV image object detection[J]. *IEEE Geoscience and Remote Sensing Letters*, 2023, 20: 6004305.
- [44] Liu C, Yang D G, Tang L, et al. A lightweight object detector based on spatial-coordinate self-attention for UAV aerial images[J]. *Remote Sensing*, 2022, 15(1): 83-103.
- [45] Hong M B, Li S W, Yang Y C, et al. SSPNet: scale selection pyramid network for tiny person detection from UAV images[J]. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 8018505.
- [46] Min X L, Zhou W, Hu R, et al. LWUAVDet: a lightweight UAV object detection network on edge devices[J]. *IEEE Internet of Things Journal*, 2024, 11(13): 24013-24023.
- [47] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. *Computer vision-ECCV 2018. Lecture notes in computer science*. Cham: Springer International Publishing, 2018, 11211: 3-19.
- [48] Liu S, Qi L, Qin H F, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 8759-8768.
- [49] Zhao X F, Xia Y T, Zhang W W, et al. YOLO-ViT-based method for unmanned aerial vehicle infrared vehicle target detection[J]. *Remote Sensing*, 2023, 15(15): 3778-3793.
- [50] Wang J Q, Chen K, Xu R, et al. CARAFE: content-aware reassembly of features[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Republic of Korea. New York: IEEE Press, 2019: 3007-3016.
- [51] Cheng Q Q, Li X H, Zhu B, et al. Drone detection method based on MobileViT and CA-PANet[J]. *Electronics*, 2023, 12(1): 223-238.
- [52] Zhang H, Shao F M, He X H, et al. Research on object detection and recognition method for UAV aerial images based on improved YOLOv5[J]. *Drones*, 2023, 7(6): 402-431.
- [53] 张瑞芳, 杜伊婷, 程小辉. 无人机视角下小目标检测算法 BiEO-YOLOv8s[J]. *激光与光电子学进展*, 2025, 62(4): 0437002.
- Zhang R F, Du Y T, Cheng X H. BiEO-YOLOv8s algorithm for small target detection from UAV view[J]. *Laser & Optoelectronics Progress*, 2025, 62(4): 0437002.
- [54] Jiang L J, Yuan B X, Du J W, et al. MFFSODNet: multiscale feature fusion small object detection network for UAV aerial images[J]. *IEEE Transactions on Instrumentation and Measurement*, 2024, 73: 5015214.
- [55] Song Q P, Cao J B, Li Y, et al. On efficient federated learning for aerial remote sensing image classification: a filter pruning approach[M]//Luo B, Cheng L, Wu Z G, et al. *Neural information processing. Lecture notes in computer science*. Singapore: Springer Nature Singapore, 2023, 14450: 184-199.
- [56] 刘旭, 宋作杰, 耿霞. 面向无人机边缘计算的小麦麦穗计数轻量化模型研究[J]. *山东农业大学学报(自然科学版)*, 2024, 55(3): 453-465.
- Liu X, Song Z J, Geng X. Lightweight model of wheat ear counting for UAV edge calculation[J]. *Journal of Shandong Agricultural University (Natural Science Edition)*, 2024, 55(3): 453-465.
- [57] Liu B Y, Chen H X, Huang Z, et al. ZoomInNet: a novel small object detector in drone images with cross-scale knowledge distillation[J]. *Remote Sensing*, 2021, 13(6): 1198.
- [58] Xiao Y C, Zhang J P. Knowledge distillation for oriented object detection on aerial images[EB/OL]. (2022-06-20) [2024-10-12]. <https://ui.adsabs.harvard.edu/abs/2022arXiv220609796X>.
- [59] Zhou W J, Yang X, Dong X N, et al. MJPNNet-S*: multistyle joint-perception network with knowledge distillation for drone RGB-thermal crowd density estimation in smart cities[J]. *IEEE Internet of Things Journal*, 2024, 11(11): 20327-20339.
- [60] Zhou M D, Wu L, Liu S, et al. UAV forest fire detection based on lightweight YOLOv5 model[J]. *Multimedia Tools and Applications*, 2024, 83(22): 61777-61788.
- [61] Cao L J, Song P D, Wang Y C, et al. An improved lightweight real-time detection algorithm based on the edge computing platform for UAV images[J]. *Electronics*, 2023, 12(10): 2274.
- [62] Zhao X F, Zhang W W, Zhang H, et al. ITD-YOLOv8: an infrared target detection model based on YOLOv8 for unmanned aerial vehicles[J]. *Drones*, 2024, 8(4): 161.
- [63] 黄福珍, 王天赐. 基于轻量化 GCP-YOLOv8s 的绝缘子缺陷检测[J]. *激光与光电子学进展*, 2025, 62(2): 0212004.
- Huang F Z, Wang T C. Insulator defect detection based on lightweight GCP-YOLOv8s[J]. *Laser & Optoelectronics Progress*, 2025, 62(2): 0212004.
- [64] Robicquet A, Sadeghian A, Alahi A, et al. Learning social etiquette: human trajectory understanding in crowded scenes[M]//Leibe B, Matas J, Sebe N, et al. *Computer vision-ECCV 2016. Lecture notes in computer science*. Cham: Springer International Publishing, 2016,

9912: 549-565.

- [65] 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), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 4165-4173.
- [66] Du D W, Qi Y K, Yu H Y, et al. The unmanned aerial vehicle benchmark: object detection and tracking[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision - ECCV 2018. Lecture notes in computer science. Cham: Springer International Publishing, 2018, 11214: 375-391.
- [67] 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), October 27-28, 2019, Seoul, Republic of Korea. New York: IEEE Press, 2019: 213-226.
- [68] Bozcan I, Kayacan E. AU-AIR: a multi-modal unmanned aerial vehicle dataset for low altitude traffic surveillance[C]//2020 IEEE International Conference on Robotics and Automation (ICRA), May 31-August 31, 2020, Paris, France. New York: IEEE Press, 2020: 8504-8510.
- [69] Wen L Y, Du D W, Zhu P F, et al. Detection, tracking, and counting meets drones in crowds: a benchmark[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 7808-7817.
- [70] Shamsoshoara A, Afghah F, Razi A, et al. Aerial imagery pile burn detection using deep learning: the FLAME dataset[J]. Computer Networks, 2021, 193: 108001.
- [71] Sun Y M, Cao B, Zhu P F, et al. Drone based RGBT vehicle detection and counting: a challenge[EB/OL]. (2020-03-05)[2024-08-09]. <https://arxiv.org/abs/2003.02437>.
- [72] Varga L A, Kiefer B, Messmer M, et al. SeaDronesSee: a maritime benchmark for detecting humans in open water[C]//2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), January 3-8, 2022, Waikoloa, HI, USA. New York: IEEE Press, 2022: 3686-3696.
- [73] Suo J S, Wang T Y, Zhang X Z, et al. HIT-UAV: a high-altitude infrared thermal dataset for unmanned aerial vehicle-based object detection[J]. Scientific Data, 2023, 10(1): 227.