

# 面向无人机航拍图像的目标检测研究综述

李琼<sup>1</sup>, 考月英<sup>1</sup>, 张莹<sup>1</sup>, 徐沛<sup>2</sup>

(1. 北京市科学技术研究院信息与人工智能技术研究所, 北京 100089;  
2. 中国科学院自动化研究所智能系统与工程研究中心, 北京 100190)

**摘要:** 随着无人机和计算机视觉技术的快速发展与深度融合, 面向无人机航拍图像的目标检测研究受到越来越多的关注, 已广泛应用于精准农业、动物监测、城市管理、应急救援等领域。与普通视角下拍摄的图像相比, 无人机航拍图像具有视野更广、目标尺寸显著缩小、视角和尺度灵活多变等特点, 无法完全适用普通视角下的目标检测方法。基于此, 首先详细回顾了普通视角下目标检测方法的研究进展, 包括传统方法、深度学习方法和基于大模型的方法, 随后综述了现有目标检测方法针对无人机航拍图像目标检测中的图像质量下降、尺度和视角变化、小目标检测难度大、复杂背景及遮挡、大视场中的不均衡, 以及实时性要求高等 6 大难点问题提出的创新策略和优化方法。此外, 归纳总结了无人机航拍图像目标检测数据集, 并在 2 个具有代表性的数据集上对现有方法进行性能分析。最后, 根据无人机航拍图像目标检测领域仍存在的问题, 展望了未来可能的研究方向, 为无人机航拍图像目标检测的发展和应用提供参考。

**关键词:** 无人机航拍图像; 深度学习; 计算机视觉; 目标检测; 多尺度目标

**中图分类号:** TP 391

**DOI:** 10.11996/JG.j.2095-302X.2024061145

**文献标识码:** A

**文章编号:** 2095-302X(2024)06-1145-20

## Review on object detection in UAV aerial images

LI Qiong<sup>1</sup>, KAO Yueying<sup>1</sup>, ZHANG Ying<sup>1</sup>, XU Pei<sup>2</sup>

(1. Institute of Information and Artificial Intelligence Technology, Beijing Academy of Science and Technology, Beijing 100089, China;  
2. Center for Research on Intelligent System and Engineering, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China)

**Abstract:** With the rapid development and deep integration of unmanned aerial vehicle (UAV) and computer vision technologies, research on object detection in UAV aerial images has gained increasing attention and has been widely applied in precision agriculture, animal monitoring, urban management, emergency rescue, and other fields. Compared to images captured from conventional perspectives, images acquired by UAVs feature a wider field of view, significantly reduced object size, and variations in viewpoint and scale, rendering conventional object detection methods inadequate. Accordingly, a detailed review of progress in object detection methods from a conventional perspective was first provided, including traditional methods, deep learning methods, and large-model-based methods. Subsequently, the innovative strategies and optimization methods proposed by existing object detection methods were summarized, specifically addressing six challenging issues specific to UAV aerial image object detection, i.e., image

---

收稿日期: 2024-08-02; 定稿日期: 2024-09-22

**Received:** 2 August, 2024; **Finalized:** 22 September, 2024

**基金项目:** 国家自然科学基金(62406030); 北京市科学技术研究院财政资助项目(24CE-BGS-01, 24CA004-03, 24CB012-01); 国家资助博士后研究人员计划(GZC20232995)

**Foundation items:** National Natural Science Foundation of China (62406030); Financial Program of BJAST (24CE-BGS-01, 24CA004-03, 24CB012-01); Postdoctoral Fellowship Program of CPSF (GZC20232995)

**第一作者:** 李琼(1997-), 女, 助理研究员, 硕士。主要研究方向为目标检测与识别。E-mail: liqiong@bjast.ac.cn

**First author:** LI Qiong (1997-), assistant researcher, master. Her main research interests cover object detection and recognition. E-mail: liqiong@bjast.ac.cn

**通信作者:** 考月英(1989-), 女, 高级工程师, 博士。主要研究方向为模式识别与机器视觉等。E-mail: kaoyueying@bjast.ac.cn

**Corresponding author:** KAO Yueying (1989-), senior engineer, Ph.D. Her main research interests cover pattern recognition and machine vision, etc.  
E-mail: kaoyueying@bjast.ac.cn

quality degradation, scale and viewpoint variation, small-object detection difficulty, complex background and occlusion, imbalance in large fields of view, and high real-time requirements. Additionally, UAV aerial image object detection datasets were consolidated and analyzed, with an evaluation of the performance of existing methods on two representative datasets. Finally, potential research directions for the future were outlined based on the unresolved issues in the field of UAV aerial image object detection, providing reference for the development and application of object detection in UAV aerial images.

**Keywords:** unmanned aerial vehicle aerial image; deep learning; computer vision; object detection; multi-scale objects

计算机视觉是一门从图像中提取信息的学科。通过计算机模拟人类视觉系统，对图像数据进行分析和处理，从而获得对数据信息的高级理解。目标检测作为计算机视觉领域中的核心任务之一，致力于在图像中识别并定位目标，不仅需要确定图像中是否存在给定类别的对象实例，还需要返回每个对象实例的精确位置<sup>[1]</sup>。目标检测是解决实例分割、目标跟踪、场景理解等其他众多高层次计算机视觉任务的基础，因此备受关注并沉淀了丰硕的研究成果，被应用于真实世界，如自动驾驶、机器人视觉、视频监控等<sup>[2]</sup>场景中。

无人机作为一种轻量便携、操作简单、成本低廉的设备，能够摆脱环境干扰和地理限制，在各种复杂或危险的场景中执行任务<sup>[3]</sup>。在无人机上搭载摄像机，可实时收集飞行过程中的视觉数据，提供比水平角度拍摄的地自然图像更丰富、可靠的信息源。对无人机航拍图像进行目标检测，有助于增强无人机的感知能力，提升其飞行和驱动能力，为人类决策提供依据，以便无人机更有效且高效地完成各项任务<sup>[4]</sup>。

近年来，随着微电子软硬件、深度学习等技术的发展，基于无人机航拍图像的目标检测得到了越来越多的关注和应用，如图 1 所示。



图 1 无人机航拍图像目标检测应用<sup>[5-8]</sup>

Fig. 1 Applications of object detection in UAV aerial images<sup>[5-8]</sup>

1) 精准农业。在农业生产规模化、专业化、优质高效的发展趋势下，配备 RGB 相机的无人机能

捕获到高分辨率的作物图像并进行检测处理，在植物识别、生长监控、产量估计、病虫害防治等方面发挥重要作用。无人机目标检测将信息技术和农业生产相结合，有助于管理人员做出定性和定量的准确判断，进行精准的田间管理。例如，CHEN 等<sup>[5]</sup>使用无人机对果园中的害虫进行拍摄和检测，实时确定害虫的位置，以规划最优的农药喷洒路线。

2) 动物监测。使用无人机替代人工巡逻和固定摄像机自动识别和定位动物，为动物管理提供了一种远距离、强机动性、非入侵式的监测方式，有助于在大空间尺度上对物种的数量、种群密度、分布、行为模式和栖息地偏好等进行评估，对动物保护和科学研究具有重要意义。例如，PROSEKOV 等<sup>[6]</sup>在西伯利亚冬季森林中，使用配备热红外成像相机的无人机进行大型动物监测，提供了有关欧洲麋鹿数量变化的可靠结果。

3) 城市管理。通过高空巡航和目标检测，无人机有助于监控城市基础设施运行状态，识别潜在安全风险，获取城市布局、交通状况等信息，为城市管理和规划提供支持。CHEN 等<sup>[7]</sup>针对无人机和自动驾驶领域提出了一种基于深度学习的无人机图像高效目标检测器 DW-YOLO，通过优化残差模块和增加卷积核数量，实现了对不同尺度物体的检测。

4) 应急救援。在自然灾害、人为事故或紧急情况下，无人机作为一个典型的辅助救援设备，提供高空视角和广阔视野范围，结合目标检测技术，可以快速搜索并定位灾害区域和遇险人员，将大幅减少灾害损失和人员伤亡。例如，LYGOURAS 等<sup>[8]</sup>在自动救援无人机的嵌入式系统上实现了深度学习算法，实现了对开放水域遇险人员的实时识别和精准检测。

面向无人机航拍图像的目标检测，是一个极具潜力的研究方向。目前，已经有很多研究人员综述了无人机目标检测方面的研究成果。2020 年，MITTAL 等<sup>[9]</sup>对基于深度学习的经典目标检测器及改进算法进行回顾，评估其在低空无人机数据集上

的性能, 但未对无人机航拍图像的检测难点进行分析。2021 年, WU 等<sup>[10]</sup>除了详细地介绍了无人机机载图像和视频中的目标检测深度学习算法外, 还对机载视频中的目标跟踪应用方面的深度学习算法进行了详细地综述, 并总结了一些算法的性能评估对比, 但杂糅了无人机航拍图像和普通视角图像目标检测算法。2021 年, RAMACHANDRAN 和 SANGAIAH<sup>[11]</sup>分类阐述了不同应用下的无人机目标检测方法, 并提出了一种基于鲁棒目标检测框架的精准农业安全车载处理系统, 但未考虑到无人机航拍图像与普通视角图像之间的差异。2022 年, BABARYKA 等<sup>[12]</sup>针对小目标(人和车), 对深度学习检测方法和经典数据集进行回顾。2023 年, TANG 等<sup>[13]</sup>指出无人机目标检测中存在小目标检测、复杂背景、目标旋转、尺度变化和类别不平衡问题, 并分类总结了深度学习目标检测的两阶段算法和一阶段算法应对上述问题的解决方案。文献[14-15]梳理了基于无人机航拍图像的方法在农业领域的应用。SRIVASTAVA 等<sup>[16]</sup>介绍了基于无人机航拍图像的深度学习方法在车辆检测方面的综述。DOLL 和 LOOS<sup>[17]</sup>比较了不同目标检测器在无人机拍摄的绵羊图像数据集上的性能, 为利用计算机视觉算法优化牲畜管理提供指导。ZHAO 等<sup>[18]</sup>综述了基于无人机的海上目标检测面临的挑战、相关方法及航拍数据集。但目前的综述文献仍缺乏基于无人机航拍图像的目标检测的最新研究进展。

基于此, 本文重点聚焦无人机航拍图像的特点和目标检测这一基础问题的难点, 对现有方法和数

据集进行全面的总结和分析。首先简要介绍经典的传统目标检测方法和深度学习目标检测方法在无人机航拍图像目标检测中的应用, 然后面向无人机航拍图像目标检测难点, 对各种改进方法进行分类介绍, 并分析其优势和短板; 接着整理了现有的无人机航拍图像数据集, 对各研究方法在典型数据集上的性能进行对比分析; 最后总结了无人机目标检测的研究现状, 并对未来的研究方向作出展望。

## 1 无人机航拍图像目标检测难点

无人机以适应地形、轻便敏捷、操作简单等优势而著称。如图 2 所示, 搭载摄像头的无人机能够以不同的速度在大场域中飞行, 拍摄到视野更广、范围更大、场景更丰富、分辨率更高的图像。无人机需要传递信息、快速检测、实时收集和传输数据。然而, 受限于设备稳定性、视角和尺度变化、环境干扰等, 无人机航拍图像呈现出与水平拍摄的普通视角图像不同的特点(图 3), 使得对无人机航拍图像进行目标检测成为一项具有挑战性的任务。



图 2 配备摄像头的无人机<sup>[19]</sup>  
Fig. 2 UAV equipped with a camera<sup>[19]</sup>

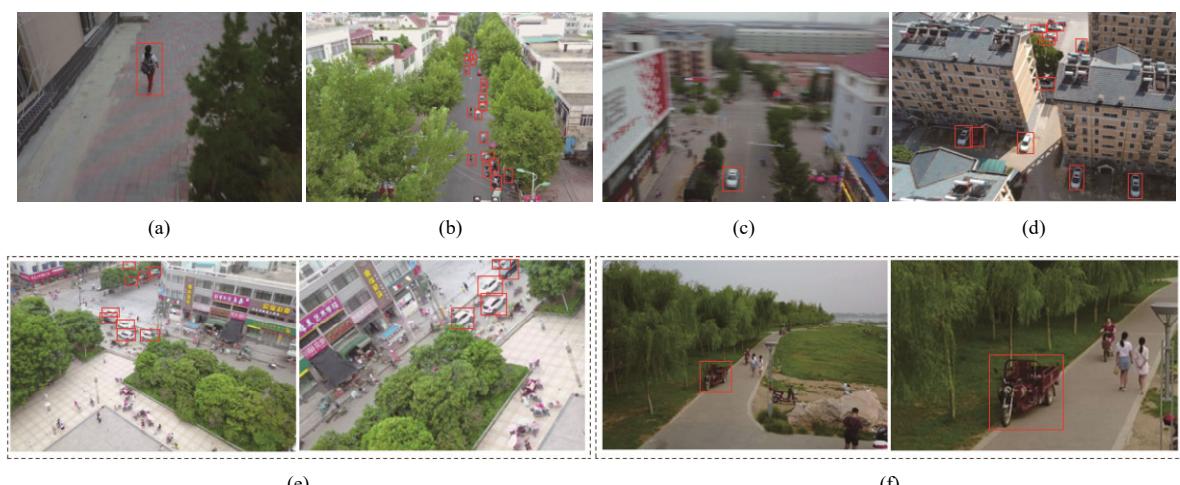


图 3 航拍图像目标检测存在的挑战<sup>[20]</sup> ((a)目标与背景占比不均衡; (b)小目标检测; (c)运动模糊; (d)目标模糊; (e)视角变化; (f)尺度变化)

Fig. 3 Challenges of object detection in aerial images<sup>[20]</sup> ((a) Unbalanced proportion between objects and background; (b) Small object detection; (c) Motion blur; (d) Blurred objects; (e) Variations in viewpoint; (f) Variations in scale)

1) 图像质量下降。无人机通常在室外环境中执行任务。一方面,无人机拍摄图像时易受到外部环境(光照、雨雪、云雾、粉尘等)的影响,导致航拍图像存在噪声;另一方面,飞行过程中可能受到气流、风力和振动等因素的干扰,在航拍图像中引入运动模糊和画面畸变。上述因素降低了图像质量和对比度,从而降低了目标检测精度<sup>[21]</sup>。

2) 尺度和视角变化。由于无人机在室外环境中飞行的灵活性,无人机与目标之间的距离可能会发生变化,导致同类目标的尺度差异较大,目标在图像中的覆盖面积也会因飞行高度的调整而不同;无人机的旋转会导致拍摄视角的变化,使得同一类目标在航拍图像中呈现的效果也存在较大差别<sup>[22]</sup>。尺度和视角的变化容易造成误检和漏检,要求目标检测算法具备对不同尺度、不同视角的目标进行有效识别和定位的能力。

3) 小目标检测。在复杂的航拍场景中,无人机摄像头被放置在较高的位置,拍摄距离较远、拍摄视角更宽,因此,航拍图像中的目标体积小、密度高,对算法准确识别和定位小目标提出了挑战<sup>[23]</sup>。

4) 背景复杂,目标遮挡。无人机飞行的大多数环境中存在树木、建筑物、电线等物体,对目标物体造成部分或完全遮挡,致使目标信息缺失;同时,这些物体的存在导致航拍图像背景较为复杂,包含大量噪声信息,给真实目标的检测带来干扰<sup>[24]</sup>。

5) 大视场中的不均衡。无人机航拍图像捕捉到的视野范围更广,覆盖的地理区域更大,导致模型训练受以下 3 类不均衡问题影响:①目标和背景的不均衡。无人机从空中拍摄的图像中,背景物体占据了较大区域,而目标物体覆盖面积较小,为目标检测增加了背景噪声干扰,使模型更关注背景信息<sup>[25]</sup>;②目标之间的不均衡。在真实世界的场景中,各类目标物体数量和识别难易程度往往存在较大差异,导致检测算法增加对数量较多或易识别的目标的偏好,泛化性降低<sup>[26]</sup>;③目标空间分布不均衡。无人机航拍图像的不同区域之间差别较大,目标往往集中在小部分区域,而其他区域包含少量目标或不包含目标,不同区域之间的检测难度不同,对图像整体采用相同的处理算法,会导致计算资源的浪费,限制目标检测精度<sup>[27]</sup>。

6) 实时性要求。动物监测、应急救援等应用场景需要快速准确地定位目标,因此对无人机航拍图像目标检测的实时性提出了要求。然而,受限于无人机的成本和负载能力,无法满足实时处理航拍高

分辨率图像所需的高算力,需要通过减少网络参数量、降低计算复杂程度等优化方式来实现无人机图像的实时检测处理<sup>[28]</sup>。

## 2 普通视角下的目标检测方法

面向无人机航拍图像的目标检测方法,并基于普通视角图像中的目标检测方法进行改进。本节回顾了普通视角下目标检测方法的研究基础和进展。

### 2.1 基于手工特征的目标检测方法

2014 年以前,传统的目标检测算法大多是基于手工特征构建的,检测流程大致分为 3 步,首先从原始图像中生成一系列可能包含目标的候选区域,然后分别提取其特征,最后使用训练好的分类器根据区域特征进行分类<sup>[29]</sup>。为了生成候选区域,通常使用滑动窗口扫描整个图像。由于目标可能具有不同尺寸和长宽比,因此,输入图像会被重新调整为不同大小,且使用多尺度的窗口在不同大小的图像上滑动。在特征提取阶段,通常使用尺度不变特征变换(scale-invariant feature transform, SIFT)<sup>[30]</sup>、加速鲁棒特征(speed up robust feature, SURF)<sup>[31]</sup>、方向梯度直方图(histogram of oriented gradients, HOG)<sup>[32]</sup>等低级视觉描述符来编码区域特征。其中,经典的 SIFT 特征是一种基于局部特征的描述符。首先构建尺度空间并检测关键点,然后计算关键点附近区域的梯度和方向信息用以描述目标,对旋转、尺度和光照变化具有较好的不变性。SURF 特征在 SIFT 特征的基础上进行改进,利用 Hessian 矩阵来检测关键点,并在尺度空间中进行关键点定位和方向分配,比 SIFT 更简单、高效。HOG 特征是一种用于描述图像局部纹理和形状的特征表示。可变形部件模型(deformable part-based model, DPM)<sup>[33]</sup>是一种基于部件的目标检测方法,将目标表示为由多个部件的组合体,利用部件之间的关系来描述目标。由于通常使用 HOG 来描述部件信息,因此 DPM 可以看作是 HOG 目标检测器的扩展。在区域分类阶段,通常使用 Adaboost<sup>[34]</sup>算法、支持向量机(support vector machine, SVM)<sup>[35]</sup>分类器等,为区域特征赋予类别标签。

尽管手工特征在早期目标检测中取得了一定进展,但一些问题限制了其精度的提升。首先,在生成候选区域的过程中,基于滑动窗口的穷举搜索计算量大,同时产生大量冗余,其中的负样本影响后续分类性能;其次,人工设计的特征仅描述低级

视觉信息, 难以在复杂场景中提取鲁棒的图像表征; 最后, 描述符的设计与分类器的训练是分离的, 难以获取目标检测的全局最优解。

## 2.2 基于深度学习的目标检测方法

随着深度学习的兴起, 利用深度神经网络自动学习特征表示在计算机视觉任务中取得了巨大成功。目前, 基于深度学习的目标检测算法已经占据主导地位, 且大致可以分为 2 类: 两阶段目标检测算法和单阶段目标检测算法。

### 2.2.1 两阶段目标检测算法

两阶段目标检测算法遵循“由粗到细”的处理模式<sup>[2]</sup>。首先使用选择性搜索或区域建议网络生成目标的候选区域, 再进行分类和定位的细化。

GIRSHICK 等<sup>[36]</sup>提出的 R-CNN (*regions with CNN features*)率先突破传统目标检测算法的限制。先通过选择性搜索筛选目标候选区域, 然后采用卷积神经网络(*convolutional neural network*, CNN)提取候选区域特征, 并将特征向量输入线性 SVM 以判断是否存在目标物体, 输入全连接网络得到具体位置信息。

HE 等<sup>[37]</sup>针对 RCNN 处理大量重叠候选区域时耗时多的问题, 提出空间金字塔池化网络(*spatial pyramid pooling networks*, SPP-Net)。在卷积层和全连接层之间引入空间金字塔(*spatial pyramid pooling*, SPP)层, 将整个图像输入 CNN 中, 能够生成任意区域固定长度的特征表示。

GIRSHICK<sup>[38]</sup>在 R-CNN 和 SPP-Net 的基础上提出 Fast R-CNN, 使用 VGG-16 代替 AlexNet 主干网络, 使用感兴趣区域(*region of interest*, RoI)池化层替代 SPP, 使用 Softmax 代替 SVM 的二分类, 整合特征提取、目标分类和位置回归, 检测速度和精度均有所提升, 但仍受候选区域提取算法速度限制。

对此, REN 等<sup>[39]</sup>提出了 Faster R-CNN, 使用区域建议网络(*region proposal network*, RPN)替代选择性搜索, 并将 RPN 设置在卷积层之后, 使整个网络共享卷积特征, 大大提升了训练效率。

HE 等<sup>[40]</sup>对 Faster R-CNN 进行扩展, 结合实例分割和目标检测构建了 Mask R-CNN。Mask R-CNN 在已有边界框识别分支的基础上, 增加了一个用于预测目标掩码的 Mask 分支, 在检测图像物体的同时, 为每个实例生产高质量的分割掩码, 提高了检测精度。

CAI 和 VASCONCELOS<sup>[41]</sup>针对 R-CNN 系列算

法在 IoU (*intersection over union*)阈值选取问题提出 Cascade R-CNN。Cascade R-CNN 由 3 个经过 IoU 阈值训练的检测头级联组成, 上一个检测头的输出将作为下一个检测头的输入。为每个检测头设置不同的 IoU 阈值, 可以实现逐渐精细的目标筛选和定位。

LIN 等<sup>[42]</sup>提出具有横向链接的特征金字塔网络(*feature pyramid networks*, FPN)。FPN 是一种自顶向下的架构, 顶层特征逐层与下层不同尺度的特征进行融合, 在不同深度分别进行预测。FPN 在小目标和多尺度目标检测方面具有显著优势。

两阶段目标检测算法先筛选可能包含目标的图像区域, 再对每个区域进行处理, 能够充分学习到目标特征, 更易取得高检测精度。但是, 由于网络结构复杂, 参数量和计算量较大, 两阶段算法检测速度相对缓慢, 难以适应对实时性要求较高的场景。

### 2.2.2 单阶段目标检测算法

单阶段目标检测算法将目标检测视为回归问题, 无需筛选候选区域, 直接在原始图像上预测物体的位置和类别。

REDMON 等<sup>[43]</sup>提出的 YOLO (*you only look once*)是最早的单阶段目标检测算法。YOLO 算法以整张图像作为输入, 将图像划分为  $S \times S$  个区域, 同时预测每个区域的边界框和类别概率。YOLO 将提取候选区域和目标识别合并, 使用一个卷积神经网络实现了端到端的目标检测。在 YOLO 基础上, REDMON 和 FARHADI<sup>[44-45]</sup>分别提出了改进模型 YOLOv2 和 YOLOv3, 使用更高效的特征提取网络, 对目标检测与分类执行联合训练, 使模型检测精度和速度均有所提升。此后, BOCHKOVSKIY 等<sup>[46]</sup>提出高效而强大的检测网络 YOLOv4。Ultralytics LLC 团队发布了 YOLOv5 的代码<sup>[47]</sup>, 采用自适应锚框计算和自适应图像缩放, 同时引入 Focus 结构, 对 YOLOv4 做出改进。LI 等<sup>[48]</sup>采用一种新的主干网络 EfficientRep, 融合先进的标签分配、损失函数和数据增强技术, 提出了 YOLOv6。WANG 等<sup>[49]</sup>通过引入模型重参数化、优化标签分配策略、提出高效层聚合网络和采用带辅助头的训练方法, 对 YOLO 系列网络进行优化, 提出 YOLOv7。Ultralytics 公司基于 YOLOv5 进行改进, 发布了 YOLOv8<sup>[50]</sup>, 提高了检测性能和灵活性。2024 年, YOLOv9<sup>[51]</sup>和 YOLOv10<sup>[52]</sup>相继问世, 目标检测精度和速度进一步提升。

除了 YOLO 系列网络外, SSD (single shot multibox detector) 系列网络也是经典的单阶段目标检测的模型。LIU 等<sup>[53]</sup>首次提出 SSD 的概念, 以 VGG-16 为主干网络, 在网络末端添加逐渐减小的卷积特征层, 允许使用多尺度特征进行预测; 使用一组小卷积过滤器来预测一组固定的默认边界框的类别分数和框偏移量。为提升 SSD 的小目标检测性能, 文献[54-55]分别提出改进的 DSSD (de-convolutional single shot detector) 算法和 FSSD (feature fusion single shot multibox detector) 算法。

针对单阶段检测器精度低于两阶段检测器的问题, LIN 等<sup>[56]</sup>发现极端前景-背景类别不平衡是主要原因, 设计了 Focal Loss, 并提出密集训练器 RetinaNet。LAW 和 DENG<sup>[57]</sup>舍弃基于锚框的检测范式, 提出了一种新的无锚框目标检测方法 CornerNet, 性能优于之前的单阶段检测器。ZHOU 等<sup>[58]</sup>遵循基于关键点的检测范式, 提出改进的算法 CenterNet。TAN 等<sup>[59]</sup>研究了目标检测网络架构的效率, 开发了一种目标检测器族 EfficientDet, 以更少的参数和计算量实现了更高的精度。

此外, CARION 等<sup>[60]</sup>为了克服非最大值抑制的局限性, 提出了一种基于 Transformer<sup>[61]</sup>的目标检测网络 DETR (detection Transformer)。DETR 将目标检测视为集合预测问题, 无需预定义锚点和非最大值抑制, 使用一组损失函数在预测目标和真实目标之间进行二分匹配, 从而实现端到端的训练。文献[62-63]分别在 DETR 的基础上, 提出可变形 DETR 和 RT-DETR (real time DETR), 其中, RT-DETR 的精度超过了 YOLOv8。

单阶段目标检测算法结构相对简单, 可以直接将图像像素映射到边界框坐标, 在一步推理中检测所有目标, 检测速度显著提高, 适用于具有实时性要求的移动设备。然而, 随着网络深度的增加, 模型可能会丢失掉细节信息, 导致对小目标、密集目标、多尺度目标等检测精度较低。

### 2.3 基于大模型的目标检测方法

大模型又被称为基础模型, 通常是指在大量数据上进行预训练、包含数亿参数的模型, 主要包括大语言模型和多模态大模型<sup>[64]</sup>。近年来, 大语言模型已经在众多文本任务中展现出强大性能。针对以目标检测为例的视觉任务, 一些研究人员对多模态大模型进行探索。

与传统限定类别的检测不同, 大模型主要针对目标检测中的开集检测, 即在文本辅助下检测任意

类别目标。LI 等<sup>[65]</sup>提出一种基础语言-图像预训练 (grounded language-image pre-training, GLIP) 模型, 通过将目标检测重构为短语定位(在图像中标定短语描述的物体), 统一预训练目标检测和短语定位任务, 学习对象级、语言感知的和语义丰富的视觉表示, 性能超过之前最先进的目标检测模型。LIU 等<sup>[66]</sup>将基于 Transformer 的检测器 DINO (DETR with improved denoising anchor boxes) 与定位预训练结合, 提出 Grounding DINO 模型, 在多个阶段执行语言-视觉模态融合, 使模型可以根据文本输入检测任意物体。KAUL 等<sup>[67]</sup>采用两阶段目标检测框架, 分别探索了基于文本、视觉和多模态的分类器在开集目标检测中的性能。研究发现, 基于多模态的分类器能够获得比全监督检测器更好的性能。ZANG 等<sup>[68]</sup>提出一种“先生成后检测”的多模态模型 ContextDET, 主要由视觉编码器、预训练大语言模型和视觉解码器构成, 对视觉-语言上下文进行端到端的可微建模, 在开集目标检测中取得较高精度。XU 等<sup>[69]</sup>提出支持多模态查询的目标检测大模型 MQ-Det (multi-modal queried object detection), 将视觉查询合并到基于语言查询的检测大模型中, 引入一个即插即用的类可扩展门控感知器模块, 利用图像示例增强文本描述, 提升开集目标检测性能。

多模态大模型结合自然语言处理和计算机视觉, 能够对文本、图像、视频等多种数据进行处理, 实现对多模态信息的综合分析, 在目标检测领域具有巨大应用潜力。

## 3 面向无人机航拍图像的目标检测方法

本节从 3 个方面介绍面向无人机航拍图像的目标检测方法。

### 3.1 面向无人机航拍图像的传统目标检测方法

早期, 航拍图像中的目标检测方法大多使用滑动窗口搜索图像, 利用手工特征或基于简单学习的特征表征目标信息。

ZHAO 和 NEVATIA<sup>[70]</sup>为了检测航拍图像中的客车, 将其视为 3D 目标识别问题, 考虑视点和阴影的变化, 提出了一种基于阴影、色彩强度和贝叶斯网络的检测方法。KLUCKNER 等<sup>[71]</sup>在 Haar-like 和 HOG 等特征上使用在线增强建立一个高效的航拍图像车辆检测器。MORANDUZZO 和 MELGANI<sup>[72-73]</sup>基于 SIFT 进行特征提取, 识别图像中的一组关键

点, 然后利用 SVM 分类器区分车辆关键点和其他特征点, 合并属于同一辆车的关键点, 实现车辆检测。STOKKEL<sup>[74]</sup>使用滑动窗口提取 HOG 特征, 并使用线性 SVM 进行分类, 对空旷区域中的人员进行检测。MORANDUZZO 等<sup>[75]</sup>提出一种通用的目标检测方法, 结合不同阶的图像梯度特征设计非线性滤波器, 并进行高斯过程建模, 对滑动窗口给出的区域是否存在目标进行估计。XU 等<sup>[76]</sup>针对 Viola-Jones+SVM 方法和 HOG+SVM 方法检测速度下降的问题, 提出一种混合自适应切换策略, 根据检测速度下降趋势的不同, 综合二者的优点, 实现更高的检测效率。WANG 等<sup>[77]</sup>首先利用 SVM 对航拍图像进行是否包含植被的分类, 然后设计 HOG 来描述棕榈树特征, 利用 HOG 训练和优化 SVM 分类器, 训练好的 SVM 使用自适应移动窗口检测棕榈树, 获取了良好的检测性能。

### 3.2 面向无人机航拍图像的深度学习目标检测方法

传统目标检测方法的性能依赖于手工特征的准确性。基于深度学习的方法通过自动提取具有区分性的特征, 能更好地适应不同应用场景。因此, 越来越多研究人员将其应用于无人机航拍图像的目标检测中。

MAIRE 等<sup>[78]</sup>将基于简单线性迭代聚类分割的区域建议方法和基于 CNN 的分类相结合, 自动检测航拍图像中的海洋物种。AMMOUR 等<sup>[79]</sup>先分割图像以得到候选区域, 然后使用预训练的 CNN 作为特征提取工具, 结合 SVM 分类器区分图像区域是否存在车辆, 并进行微调以准确定位车辆, 使检测准确率和速度上均有所提升。LI 等<sup>[80]</sup>将无人机捕获的视频流解码为分离帧, 发送给 YOLO 目标检测网络, 获得目标边界框和类别预测, 实现无人机的自主目标检测。HONG 等<sup>[81]</sup>分别采用 Faster R-CNN, R-FCN, SSD, Retinanet 和 YOLO 等 5 种不同的基于深度学习的目标检测方法, 对无人机航拍图像中的鸟类进行检测, 并评估了各方法的检测性能。MAKAROV 等<sup>[82]</sup>使用有标注的航拍图像训练 YOLOv2, 以实现无人机对汽车、大型车辆、建筑物、飞机、直升机等 6 类物体的检测。CHENG 等<sup>[83]</sup>基于无人机航拍图像数据集, 对 Faster R-CNN 和 YOLOv3 进行研究, 分别评估 2 种算法的目标检测平均准确率和耗时, 发现 Faster R-CNN 更适合对目标检测精度要求较高的场景, 而 YOLOv3 更适合对目标检测实时性要求较高的场景。ZHU 等<sup>[84]</sup>利

用无人机拍摄路面图像并进行破损标注, 分别使用 Faster R-CNN, YOLOv3 和 YOLOv4 对航拍图像中的路面破损进行检测。KU 等<sup>[85]</sup>采用 YOLOv7 作为车辆和海洋漂浮物的目标检测模型, 获得了较为准确和鲁棒的目标检测结果。

尽管基于普通视角设计的目标检测算法已经在无人机航拍图像目标检测中取得了较高精度, 但是由于航拍图像尺寸和视野范围较大, 且拍摄角度和拍摄环境不固定, 通用深度学习目标检测算法的性能受到限制, 难以直接应用于无人机检测的现实场景。

### 3.3 面向无人机航拍图像的改进目标检测方法

目前, 有许多工作基于无人机航拍图像的特点, 对目标检测中遇到的图像质量下降、不均衡、尺度和视角变化、小目标、复杂背景与遮挡、实时性检测等挑战展开研究。

#### 3.3.1 针对图像质量下降的改进

无人机航拍图像可能受到快速运动、光照变化、雾霾、粉尘等影响造成模糊或畸变。为了去除此类干扰, 需要采取相应的去噪和校正措施, 以提高图像质量和检测精度。

图像增强是一种常见的处理手段。针对光照影响, WANG 等<sup>[86]</sup>使用双曲正切曲线将低照度图像的亮度拉伸到理想水平, 并构建反锐化滤波器, 对图像去噪和锐化, 利用 CNN 提取图像特征进行行人检测, 获得了更高的精度和鲁棒性。WANG 等<sup>[87]</sup>基于生成对抗网络(generative adversarial network, GAN)的 LighterGAN 模型, 将具有低照度特征的图像转换为具有足够照度的特征来实现增强, 缓解图像光照不足问题。LIU 等<sup>[88]</sup>提出一种自适应车辆检测算法 DCNet。首先判断图像是否模糊, 构建模糊图像数据集; 然后, 设计基于 GAN 的改进算法 Drone-GAN, 增强模糊图像中的目标特征, 使用于检测的特征对光照变化、遮挡和模糊具有鲁棒性。

此外, WANG<sup>[89]</sup>对图像原始样本执行 HSV (hue-saturation-value)空间亮度平移操作, 丰富样本集, 增加分类器对不同样本和光照条件的适应性。MINH 等<sup>[90]</sup>提出 FFA-Net (feature fusion attention network)<sup>[91]</sup>去雾方法和 PAA (probabilistic anchor assignment)<sup>[92]</sup>目标检测方法。其中, FFA-Net 使用特征注意力模块和局部残差学习模块灵活处理和关注不同区域和特征, 提高图像去雾效果; PAA 在训练过程中指定一种新的锚框, 并修改损失函数, 确定有利于模型的正样本。ZHANG 等<sup>[93]</sup>将多种属

性融合到一个耐噪声的哈希框架中，对每个目标的内部结构进行灵活编码，并通过协同利用  $l_F$  和  $l_1$  范数，使计算出的哈希码对低质量的航拍图像和嘈杂的语义标签具有鲁棒性。ZHU 等<sup>[94]</sup>基于对图像模糊物理特性和网络模型适应性的分析，提出一种自适应多尺度融合去模糊生成对抗网络，利用图像模糊程度指导网络多尺度特征以自适应的权重进行融合，抑制融合过程中的对齐误差，提高图像去模糊效果。

### 3.3.2 针对尺度和视角变化的改进

由于无人机的飞行高度和拍摄视角并不固定，导致航拍图像中的目标的形状和尺寸发生变化。当前，有许多方法致力于应对航拍图像目标检测中的尺度和视角变化。

针对无人机航拍图像中的视角变化，CHENG 等<sup>[95]</sup>提出一种旋转不变模型，在现有 CNN 结构上引入旋转不变层来提高目标检测性能，并施加正则化约束来强制训练样本旋转前后的特征映射相接近。DING 等<sup>[96]</sup>提出 ROI Transformer，在 ROI 上应用空间变换，并在有向包围盒标注的监督下学习变换参数，提取旋转不变特征。PAN 等<sup>[97]</sup>基于 CenterNet 提出动态细化网络。其中，特征选择模块来根据目标方向调整感受野，动态细化头则能够以目标感知的方式进行动态预测。

针对无人机航拍图像中的目标尺度变化，CHEN 等<sup>[98]</sup>设计了一种混合检测器 RRNet。放弃先验锚点，并提出 AdaResampling 自适应重采样策略来对目标进行逻辑增强，在密集场景中的多尺度目标检测中取得更好的结果。WANG 等<sup>[99]</sup>提出 SAMFR (spatial attention for multi-scale feature refinement)，利用设计感受野扩展块扩大高层语义特征的感受野，引入空间细化模块修复多尺度物体的空间细节，细化了传统的基于特征金字塔的多尺度特征，很好地应对尺度变化问题。LIN 等<sup>[100]</sup>提出一种改进的目标检测网络 ECascade-RCNN，采用 Trident-FPN 主干提取多尺度特征，同时采用新的双头注意力机制来增强检测器的性能。LI 等<sup>[101]</sup>提出一种轻量级模型 DSYOLOv3，不仅融合 YOLOv3 不同层的特征，还结合通道注意力机制和空间金字塔池化模块融合不同尺度的通道特征，实现对不同尺度目标的检测。WANG 等<sup>[102]</sup>提出了 SPB-YOLO 的实时检测器，利用注意力机制更好地理解无人机图像中不同尺度目标的宽-高依赖关系，提高对不同尺度目标的检测灵敏度。SUN 等<sup>[103]</sup>以头部解耦的

单阶段目标检测器为基线，提出一种考虑不同分类和定位偏好的解耦特征金字塔网络 (decoupled feature pyramid network, DFPN)。ZHANG 等<sup>[104]</sup>提出采用全尺度特征聚合模块对多组特征图进行尺度自适应和信息聚合，并将聚合特征细分为多个子层特征，使其自主学习目标特征的通道和空间布局。此外，提出并行超分辨率语义增强模块来重建深度特征图并融合空间上下文信息，增强模型对模糊目标的分类能力。LIU 等<sup>[105]</sup>设计了一个尺度不变特征解缠模块，用于分离尺度相关特征和尺度不变特征，并采用对抗学习来增强解缠效果，提升检测的鲁棒性。

### 3.3.3 针对小目标检测的改进

由于无人机拍摄的图像分辨率有限，在与目标距离较远时，目标在航拍图像中的占比较小，易使深层网络丢失目标信息。因此，研究人员提出许多改进方案，使检测模型更关注小目标。

LIANG 等<sup>[106]</sup>提出一种基于特征融合和缩放的单目标检测器，在反卷积模块中添加一个额外的缩放分支形成特征金字塔，同时对原始特征融合分支进行调整，使 2 个特征金字塔共同作用于小目标检测任务。此外，将目标空间关系融入到目标重检测中，进一步提升小目标检测性能。

ZHANG 等<sup>[107]</sup>在主干网络 ResNet 中引入可变形卷积层，以学习更有区分度的特征，同时使用交错级联架构处理特征，用来消除遮挡的影响，细化目标位置和大小。

LIU 等<sup>[108]</sup>提出一种多分支并行特征金字塔网络 (multi-branch parallel feature pyramid networks, MPFPN)，利用并行分支恢复深层网络中缺失的小目标特征，同时引入空间注意力模块使网络更关注小尺寸目标。

JADHAV 等<sup>[109]</sup>以 ResNet 模型为基础，利用更密集的具有大尺度方差的锚点尺度来检测密集分布的小目标，同时获取通道依赖关系，以获得更好的特征表示。

LIU 等<sup>[110]</sup>提出一种高分辨率检测网络 (high-resolution detection network, HRDNet)，核心思想是分别使用深层网络和浅层网络来处理低分辨率和高分辨率图像，通过融合多深度、多尺度特征，提高小目标检测精度和效率。

针对航拍图像中的小目标漏检问题，TIAN 等<sup>[111]</sup>提出一种双神经网络目标检测方法 (dual neural network object detection, DNOD)。先使用单阶段检

测器以最优质信度阈值检测目标, 在不忽略疑似区域的情况下实现高准确率检测; 然后利用 VGG 主干网络提取特征, 对疑似目标区域的特征进行二次识别, 快速筛选漏检目标。

SHANG 等<sup>[112]</sup>提出一种基于 YOLOv5s 的小目标检测算法。首先, 在 YOLOv5s 中添加小目标检测层; 然后, 采用增强的加权双向特征金字塔网络替换 YOLOv5s 中的路径聚合网络, 提高对不同尺度目标的检测能力; 同时, 引入非参数注意力机制, 增强网络特征提取能力。上述改进的综合作用使算法在小目标上的检测性能较 YOLOv5s 有较大提升。

ZHAO 等<sup>[113]</sup>提出一个基于 YOLOv7 的小目标检测算法。通过引入非跨行卷积模块、多尺度注意模块和可变形注意模块, 提高网络的特征提取能力和泛化能力。

LI 等<sup>[114]</sup>提出了一种高效的 YOLOv7-UAV 算法, 增加浅层预测头 P2 来提高对小目标特征的提取性能, 去除深层预测头 P5 来降低深层特征的影响, 对双向特征金字塔网络进行加权交叉连接以融合多尺度特征图。与 YOLOv7 相比, YOLOv7-UAV 在检测性能和速度上均有明显提升。

### 3.3.4 针对复杂背景及目标遮挡的改进

无人机从空中拍摄图像, 环境中可能出现建筑物、树木、道路等多种干扰物体, 导致复杂的背景信息和目标区域遮挡, 需要采用有效的算法设计来应对这些挑战, 提高检测准确性和鲁棒性。

YANG 等<sup>[115]</sup>提出一种高性能的无人机车辆检测器, 使用单发细化神经网络(single-shot refinement neural network, RefineDet)作为基础网络, 采用自顶向下的结构来提供上下文信息, 同时设计多尺度邻接连接模块, 为车辆检测提供有效的上下文信息。

ZHANG 等<sup>[116]</sup>提出了一种多尺度和遮挡感知网络。使用多尺度特征自适应融合网络自适应地融合低层位置信息和高层语义信息, 同时使用基于区域注意力的三头部网络增强感兴趣区域, 减少遮挡的影响。

LIU 等<sup>[117]</sup>提出一种基于空洞卷积和注意力机制的多尺度特征融合算法 D-A-FS SSD (dilated-attention-feature fusion SSD)。在特征提取阶段引入空洞卷积以扩大感受野, 提高网络对目标分布和尺度差异信息的表达, 同时使用注意力网络增强目标特征; 在多尺度检测阶段, 融合低层特征和高层特

征, 进一步提升算法精度。

CAI 等<sup>[118]</sup>提出一种融合背景注意力和前景注意力的无锚点引导注意力网络, 从杂乱的背景中学习具有判别力的特征。

XI 等<sup>[119]</sup>提出了一种细粒度目标聚焦网络(fine-grained target focusing network, FiFoNet), 从不同层次的特征聚合出目标中的子部分特征, 以提供更细粒度的表示, 同时通过目标掩模阻挡背景干扰, 更关注前景目标。此外, 使用具有不同尺寸核的卷积层提取上下文信息以增强目标特征表示。

### 3.3.5 针对大视场中不均衡的改进

相比地面拍摄, 无人机航拍图像的覆盖范围较大, 因此, 图像中真实目标占据面积较小, 而背景占据较大区域, 导致目标与背景之间比例不均衡, 降低了检测模型对目标的关注, 影响检测性能。为了获取大视场中的小目标信息, 研究人员提出裁剪法, 将原始图像分成多个小区域, 以缓解背景干扰。例如, LI 等<sup>[120]</sup>提出一种密度图引导的目标检测网络(density-map guided object detection network, DMNet), 设计了基于密度图的有效裁剪方法, 利用目标之间的空间和上下文信息提高检测性能。LI 等<sup>[121]</sup>提出一种基于包的单阶段检测器(bag-based single-stage detector, BSSD), 定义了用于覆盖目标的简单且自适应的包。训练过程中, 采用 OHEM (online hard example mining)<sup>[122]</sup>进一步减少目标和背景的不平衡。YANG 等<sup>[123]</sup>提出一种基于 YOLOX 的目标检测算法 VAMYOLOX, 增加一个预测头来检测密集小目标, 提高对目标的检测能力, 并引入三重关注模块来重新设计网络的颈部, 增强网络提取关键特征的能力。LENG 等<sup>[124]</sup>受人类视觉系统对困难目标投入更多精力的现象启发, 提出了 Pareto 聚焦检测(pareto refocus detection, PRDet)网络。使用反向注意力探索模块, 通过抑制对常用检测器而言更显著的特征来发现目标, 同时引入区域特定上下文学习模块, 加强对图像中目标区域的理解。

在实际场景中, 不同目标类别的出现频率以及识别的难易程度可能存在较大差异, 导致目标之间不均衡, 使检测模型偏好数量较多和容易识别的目标。HONG 等<sup>[125]</sup>考虑到目标之间的类别不平衡和困难样本, 将预训练模型分类错误的目标区域作为困难样本, 提出基于困难区域的训练。从原有数据集和外部数据集中提取目标图像块, 构建目标池, 使用这些图像块扩充数据集, 以平衡各类别的比例, 利用普通样本和困难样本共同训练检测模型,

取得了具有竞争力的检测结果。YU 等<sup>[126]</sup>提出了一种双采样器和头部检测网络(dual sampler and head detection network, DSHNet)，其中包括类偏置采样器和双边箱头，以双路径方式分别处理高频和低频类别，显著提升了数量较少的目标类别的检测性能。YAMANI 等<sup>[127]</sup>提出多样不确定聚合方法，旨在选择具有更多样化目标类别的图像，同时根据每个类的表现调整不确定性计算，以解决类别不平衡问题，提升单阶段目标检测器性能。HOU 等<sup>[128]</sup>提出一种基于无锚点检测器的检测框架。其中，样本平衡策略模块用以平衡正负样本和难易样本，具有中心权重的超分辨率 GAN 可以有效增强局部特征图，改善小目标表征。

由于拍摄场景多样性和目标特异性，图像不同区域的目标数量存在较大差别，目标在空间分布上存在不均衡现象，导致不同区域的检测难度不同。为了缓解航拍图像检测中的目标分布不均匀问题，文献 [129-130] 分别提出了聚类检测 (clustered detection, ClusDet) 网络和聚类区域估计网络 (cluster region estimation network, CRENet)，将聚类引入检测框架中，搜索包含密集目标的区域，并使检测器关注这些区域，有效提升了目标检测精度和效率。DENG 等<sup>[131]</sup>提出一种端到端的全局-局部自适应网络 (global-local self-adaptive network, GLSAN)，先粗略地检测出被原始图像限制的区域和目标，然后以自适应的方式继续细化对微小尺度目标的检测。LIAO 等<sup>[132]</sup>提出一种无监督聚类引导目标检测网络 (unsupervised cluster guided network, UCGNet)，利用无监督聚类模块分割出目标聚集区域，并将其输入检测器，使用全局合并模块合并所有密集区域的候选框来生成最终的预测结果。LI 等<sup>[133]</sup>提出一种自适应聚类目标检测方法。其中，自适应聚类子网络通过提取图像中潜在的小目标聚集区域，对原始图像进行划分；分割填充子网络对分割图像进行尺寸校正，以适应检测网络的输入要求；局部和全局检测网络分别对分割图像和原始图像中的目标进行检测，融合检测结果有效提升模型的检测精度和泛化性能。

### 3.3.6 针对实时性要求的改进

为了快速响应、及时更新目标信息和提供即时决策支持，要求无人机目标检测具有实时性。然而无人机设备计算和存储资源有限，难以实时运行大规模高精度深度学习模型。因此，研究人员对模型进行压缩，同时提出许多改进算法，帮助无人机在

不损失精度的情况下提升检测速度。

KYRKOU 等<sup>[134]</sup>探讨了不同 CNN 检测器在无人机目标检测中的应用，提出一种基于 Tiny-YOLO 的轻量级网络 DroNet，减少了网络层数和滤波器数量，可以在无人机嵌入式平台上高效执行车辆检测任务。

ZHANG 等<sup>[135]</sup>通过对通道缩放因子施加 L1 正则化来加强卷积层的通道级稀疏性，并修剪信息量较少的特征通道，获得参数量和计算量更少的深度目标检测器 SlimYOLOv3。

YOLOv3, YOLOv4 和 YOLOv5 是目标检测中经常用的 3 个系列。为了提升检测速度，许多研究将 YOLO 的主干替换为尺寸更小的网络<sup>[136]</sup>。例如文献[137-138]分别将 YOLOv3 和 YOLOv4 的主干网络替换为 MobileNet，通过去除网络冗余以进一步压缩模型，并利用知识蒸馏算法提高压缩模型的检测精度，在保证精度的同时提升无人机火灾检测和幸存者检测的实时性。

QIN 等<sup>[139]</sup>提出一个基于 YOLOv3-Tiny 的架构 Ag-YOLO。Ag-YOLO 由一个衍生自 ShuffleNet v2 的主干网络、1 个 ResBlock 颈部和 1 个 YOLOv3 头部组成，并经过模型压缩，在资源受限的硬件上实现了更高的精度和更快的速度。

DENG 等<sup>[140]</sup>在 YOLOv5s 的基础上，利用轻量化检测算法 LAI-YOLOv5s，用最大检测头替换最小检测头，设计深度特征映射交叉路径融合网络，将最深层的特征交叉路径与主干网络融合在一起，丰富特征的语义信息；设计基于 VoVNet20 和基于 ShuffleNetV2 的 2 个改进模块，提高特征提取能力，减少模型的计算量和参数量。

CAO 等<sup>[141]</sup>不仅使用尺寸更小的 MobileNetV3 作为 YOLOv5 主干网络，还同时引入了高效通道注意机制，提高网络的特征提取能力，并设计了 2 种不同参数尺度的颈部结构，以满足不同嵌入式设备的要求。与 YOLOv5s 相比，该算法的检测精度和速度具有显著提升。

### 3.3.7 综合改进

在无人机目标检测领域，存在小目标检测、复杂背景及目标遮挡、样本不均衡等问题，限制了航拍图像的检测性能。一些算法综合考虑多方面的挑战，为提升检测准确性和鲁棒性提供了新的思路和解决方案。

WU 等<sup>[142]</sup>将航拍图像中面临的视角变化、高度变化等干扰均看作细粒度领域，提出一种干扰解缠

特征变换(*nuisance disentangled feature transform*, NDFT)对抗性训练框架, 学习特定于任务的、领域不变的特征。该网络可以直接迁移到其他细粒度领域中, 解决无人机图像中目标检测的特定挑战。

TANG 等<sup>[143]</sup>针对密集目标、类别不平衡和多尺度问题, 提出一种点估计网络(*points estimate network*, PENet)。其中, 掩码重采样模块用来扩充不平衡的数据集, 粗粒度无锚点检测器结合非最大值抑制算法生成小目标高质量聚类的中心点, 并送入细粒度无锚点检测器进行精确检测。同时, 将原始图像输入细粒度检测器, 检测航拍图像中的大目标。组合细粒度检测器的 2 种输出, 得到最终的检测结果, 性能优于之前的模型。

ALBABA 和 OZER<sup>[144]</sup>针对目标尺度和视角变化、普通视角图像与航拍图像之间的纹理和形状特征差异, 结合单阶段和多阶段目标检测的优点, 使用具有预训练特征提取器的 CenterNet 和 Cascade R-CNN 以及集成策略, 降低多阶段目标检测器的高漏检率, 提高单阶段检测器的性能。

HUANG 等<sup>[145]</sup>针对航拍图像中前景目标比例较低及相似类别区分度低的问题, 提出一种统一前景封装的多代理检测网络(*multi-proxy detection network with unified foreground packing*, UFPMP-Det), 设计了统一前景封装模块, 首先通过聚类算法对粗检测得到的前景子区域进行合并以抑制背景, 然后将合并后的子区域进行尺度自适应放大, 显著提高检测精度和速度。

LUO 等<sup>[146]</sup>提出了一个基于 YOLOv5l 的改进方法 YOLO-UAV, 使用 3 个非对称卷积模块分别对 YOLOv5 主干网络中不同位置的残差块进行替换。在 Focus 模块之后加入改进的高效通道注意力模块, 帮助网络聚焦重要特征。同时, 使用组空间金字塔池化降低模型参数量和过拟合风险。与 YOLOv5l 相比, YOLO-UAV 在小目标、密集排列、稀疏分布、复杂背景等挑战下仍具有较好的检测性能。

LI 和 LIU<sup>[147]</sup>提出一种基于 YOLOv7X 的改进算法 YOLOv7X+, 在 YOLOv7X 主干网络的浅层引入混合残差空洞卷积注意力模块以获取小目标细节, 同时添加调制卷积模块, 获取丰富的全局信息特征图, 并通过在颈部引入动态卷积, 应对无人机图像多角度拍摄引起的目标尺度形状变化。

## 4 数据集及性能评估

无人机航拍图像数据集为目标检测算法提供

了大量的有标注图像, 帮助模型准确地检测和识别各种目标。与普通视角图像数据集相比, 用于目标检测的无人机航拍图像数据集在数量和数据量上都有所欠缺。本文梳理了部分代表性无人机视角下的目标检测数据集(表 1), 并在 2 个常用的公开数据集上对现有算法进行性能评估(表 2 和表 3)。

### 4.1 数据集

1) CARPK 数据集<sup>[148]</sup>于 2017 年发布, 是我国台湾大学提出的用于停车场车辆计数的数据集。使无人机飞行高度在 40 m 左右, 在 4 个停车场的不同场景下拍摄近 9 万辆汽车, 并对每个汽车目标进行边界框注释, 即记录边界框左上角和右下角的点坐标。

2) UAVDT 数据集<sup>[149]</sup>于 2018 年发布, 是中国科学院大学针对复杂场景下的车辆检测和追踪提出的大规模数据集。使用无人机分别以 10~30 m, 30~70 m 和 >70 m 的 3 种高度飞行, 共拍摄 10 h 的原始视频, 包括各种常见的场景, 如广场、收费站、高速公路等。从中获取 100 个视频序列, 8 万个代表帧, 人工标注了边界框及天气条件、飞行高度、车辆类别等 14 种属性。数据集包含汽车、卡车和公共汽车等 3 种类别, 而卡车和公共汽车的数量不到整个数据集的 10%。

3) VisDrone 数据集<sup>[20]</sup>于 2018 年发布, 是天津大学提出的大规模无人机航拍图像数据集, 图像来自不同场景、天气和光照条件, 被应用于 VisDrone-DET 挑战赛<sup>[150~152]</sup>中。该数据集包含 8 599 张图像(6 471 张作为训练集, 548 张作为验证集, 1 580 张作为测试集), 重点检测 10 个预定的类别(行人、人、汽车、面包车、公共汽车、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), 具有边界框、目标类别、遮挡等人工标注。各类别目标的数量存在不均衡现象, 其中, 带遮阳棚的三轮车、不带遮阳棚的三轮车、公共汽车等类别数量较少, 行人和汽车的数量则远超其他类别。

4) DAC-SDC 数据集<sup>[153]</sup>于 2018 年发布, 是 University of Notre Dame 提出的大规模目标检测数据集, 包含来自 95 个类别的 15 万张图像。该数据集中的目标占据航拍图像的 1%~2%, 且图像数量在亮度/信息量上的分布接近高斯分布, 即大多数图像包含中等亮度和信息量, 少部分图像具有过大/过小的亮度/信息量。数据集中的目标类别以人、汽车和骑车为主, 超过所有类别数量的 2/3。

5) AU-AIR 数据集<sup>[154]</sup>于 2020 年发布, 是 Aarhus

University 使用无人机在真实世界室外环境中收集的多模态传感器数据，包含视觉、时间、位置、高度、惯性测量单元数据、速度等。无人机以不同的飞行高度和相机角度拍摄 8 个视频流(总时长超过 2 h)，选取 32 823 个视频帧，进行边界框和人、汽车、货车、卡车、摩托车、自行车、公共汽车、拖车等 8 个类别的标注。在交通监控的背景下，汽车的出现次数显著高于其他类别，且汽车，货车和卡车占据大部分的标注。

6) UVSD 数据集<sup>[116]</sup>于 2020 年发布，是山东大学基于无人机车辆检测和分割问题提出的公共数据集。该数据集包含 5 874 张图像，共计 98 600 辆车辆实例。由于图像是在不同场景、不同角度、不同高度、不同光照条件下拍摄的，因此，车辆实例具有视点变化、尺度变化、光照变化等局部遮挡、分布密集等特点。除了像素级实例注释外，UVSD 还包括有向边界框和水平边界框注释，可以用于语义分割和车辆检测任务。

7) MOHR 数据集<sup>[155]</sup>于 2021 年发布，是哈尔滨工业大学(深圳)针对无人机图像多尺度目标检测提出的大规模基准数据集。使用无人机分别在 200 m, 300 m 和 400 m 采集来自郊区、山区、雪地和沙漠地区的不同分辨率的图像共 10 631 张。对这些图像进行建筑、汽车、卡车、塌陷、洪灾等 5 个类别的标注，共获取 90 014 个带有标签和边界框的目标实例。其中，数量最多的汽车占据全部标注的 46%，数量最少的洪灾仅占全部标注的 3%。

8) DroneVehicle 数据集<sup>[156]</sup>于 2022 年发布，是天津大学针对低光照下的无人机车辆检测提出的大规模 RGB-红外车辆检测数据集，覆盖了城市道路、居民区、停车场等多种场景。该数据集包含 28 439 个 RGB-红外图像对，对图像中的汽车、公共汽车、面包车、货车、卡车等 5 个类别进行标注，标注的汽车数量接近货车数量的 4 倍。

9) Manipal-UAV person detection dataset<sup>[157]</sup>于 2022 年发布，是马尼帕尔高等教育学院针对小目标检测提出的数据集。使用飞行在不同高度、位置和天气条件下的 2 架无人机采集 33 个视频，并对其采样得到 13 462 张图像，包含大量尺度、姿态、光照及遮挡各异的目标，对图像中 153 112 个目标实例进行标注。

10) SeaDronesSee 数据集<sup>[158]</sup>发布于 2022 年，是图宾根大学针对缺少海上无人机数据问题提出

的大规模开放水域中的人员检测数据集。使用无人机在 5~260 m 范围内的不同高度、0°~90°范围内的不同视角拍摄图像，收集和注释了超 5.4 万帧图像，40 万个实例。除了边界框和类别注释外，该数据集还提供了高度、相机角度、速度、时间等元信息。在船、救生衣、游泳者、穿救生衣的游泳者等 6 个类别中，船和游泳者的数量接近其他 4 个类别数量总和的 2 倍。

11) RTDOD 数据集<sup>[159]</sup>于 2023 年发布，是中国科学院大学针对复杂场景中彩色图像信息不足问题提出的大规模目标检测数据集。使用安装在无人机上的校准的彩色热成像摄像机在不同天气条件下同步捕获 RGB 和热成像视频，提取 16 200 对 RGB-T 图像对，对人、狗、自行车、运动球、汽车、船、摩托车、卡车、婴儿车和公共汽车等 10 个类别的目标进行注释。各类别数量分布并不均衡，其中，汽车和人的比例最高，占总数的一半以上，运动球和婴儿车的比例最低，在训练集和测试集中均只有 200~300 个样本。

12) WAID 数据集<sup>[160]</sup>于 2023 年发布，是北京林业大学基于野生动物监测提出的公开数据集。WAID 包含来自不同环境条件的 14 375 张图像，涵盖羊、牛、海豹、骆驼、西藏野驴和斑马等 6 种野生动物种类和沙漠、草原、沙滩等栖息地类型。每张图像均具有类别和目标边界框标注。其中，牛的数量最多，超出骆驼、西藏野驴、斑马数量 5 倍以上。

13) NVD<sup>[161]</sup>于 2023 年发布，是吕勒奥理工大学针对北欧地区不同环境、不同积雪条件下的车辆检测提出的大规模数据集。包括 120~250 m 范围内不同高度下拍摄的 8 450 帧图像，具有 26 313 个汽车标注。

14) UEMM-Air 数据集<sup>[162]</sup>于 2024 年发布，是河海大学提出的合成多模态目标检测数据集。使用虚幻引擎模拟各种飞行场景及目标类型，包括城市、公园、高速公路等 13 个大类场景、上百种车型，使无人机在不同高度、场景和模态下收集数据。该数据集的 2 万对图像包含 5 种模态，分别为 RGB、红外、分割、表面法向量和 IMU 参数。此外，还提出一种启发式的图像自动标注算法，避免错误标记重叠目标。由于不同车型在现实场景中出现的频率不同，因此标注的不同目标类别的数量具有较大差异。

表1 无人机视角下的目标检测数据集  
Table 1 UAV-view object detection datasets

数据集	图像数量/ 张	标注数量/ 个	飞行高度/m	分辨率	目标类别	任务	发布时间/ 年	下载链接
CARPK <sup>[148]</sup>	1 448	89 777	40	1280×720	汽车	车辆计数	2017	<a href="https://lafi.github.io/LPN/">https://lafi.github.io/LPN/</a>
UAVDT <sup>[149]</sup>	80 000	840 000	10~30、 31~70、 >70	1080×540	汽车、卡车和 公共汽车	车辆检测 和追踪	2018	<a href="https://sites.google.com/site/daviddo0323/">https://sites.google.com/site/daviddo0323/</a>
VisDrone <sup>[20]</sup>	8 599	540 000	-	2000×1500	行人、人、 汽车、面包车、 公共汽车、 卡车等 10 个类别	目标检测	2018	<a href="https://github.com/VisDrone/VisDrone-Dataset">https://github.com/VisDrone/VisDrone-Dataset</a>
DAC-SDC <sup>[153]</sup>	150 000	-	-	640×360	人、汽车、船、 建筑等 12 个 大类及 95 个子类	目标检测	2019	<a href="https://github.com/xyzxinyizhang/2018-DAC-System-Design-Contest">https://github.com/xyzxinyizhang/2018-DAC-System-Design-Contest</a>
AU-Air <sup>[154]</sup>	32 823	132 000	5~30	1920×1080	人、汽车、货车、 卡车、摩托车等 8 个类别	交通监视	2020	<a href="https://bozcani.github.io/auairdataset">https://bozcani.github.io/auairdataset</a>
UVSD <sup>[116]</sup>	5 874	98 600	10~150	960×540~ 5280×2970	车辆	车辆检测 和分割	2020	<a href="https://github.com/liuchunsense/UVSD">https://github.com/liuchunsense/UVSD</a>
MOHR <sup>[155]</sup>	10 631	90 014	200、300 和 400	5472×3078; 7360×4192; 8688×5792	建筑、汽车、 卡车、塌陷、 洪灾等 5 个类别	多尺度 目标检测	2021	/
DroneVehicle <sup>[156]</sup>	56 878	819 000	-	840×712	汽车、公共汽车、 面包车、货车和 卡车等 5 个类别	车辆检测	2022	<a href="https://github.com/VisDrone/DroneVehicle">https://github.com/VisDrone/DroneVehicle</a>
Manipal-UAV person detection dataset <sup>[157]</sup>	13 462	153 112	10~50	1280×720	人	小目标 人员 检测	2022	<a href="https://github.com/Akshathakrbhat/Manipal-UAV-Person-Dataset">https://github.com/Akshathakrbhat/Manipal-UAV-Person-Dataset</a>
SeaDroneSee <sup>[158]</sup>	54 000	400 000	5~260	3840×2160~ 5456×3632	船、救生衣、 游泳者等 6 个类别	海上 人员 检测	2022	<a href="https://seadroneesee.cs.uni-tuebingen.de">https://seadroneesee.cs.uni-tuebingen.de</a>
RTDOD <sup>[159]</sup>	32 400	179 672	-	1280×720	人、狗、自行车、 运动球、摩托车、 船等 10 个类别	目标 检测	2023	<a href="https://github.com/fengt96/RTDOD">https://github.com/fengt96/RTDOD</a>
WAID <sup>[160]</sup>	14 375	-	-	1018×572	羊、牛、海豹等 6 个类别	野生 动物 监测	2023	<a href="https://github.com/xiaohuicui/WAID">https://github.com/xiaohuicui/WAID</a>
NVD <sup>[161]</sup>	8 450	26 313	120~250	1920×1080~ 3840×2160	车辆	车辆 检测	2023	<a href="https://nvd.ltu-ai.dev/">https://nvd.ltu-ai.dev/</a>
UEMM-Air <sup>[162]</sup>	20 000	-	5~50	1920×1080	城市、公园等 13 大类场景， 上百种车型	车辆 检测	2024	<a href="https://github.com/le12Leon/UEMM-Air">https://github.com/le12Leon/UEMM-Air</a>

表2 UAVDT 数据集上的性能评估  
Table 2 Performance evaluation on UAVDT dataset

对应问题	方法	基础网络	训练/测试	输入尺寸/像素	AP	AP <sub>50</sub>	AP <sub>75</sub>	发表时间/年
图像质量下降	TRÂN 等 <sup>[90]</sup>	FFA-Net+PAA	23 384/2 181	-	12.50	-	-	2022
尺度和视角变化	DSYOLOv3 <sup>[101]</sup>	YOLOv3	24 143/16 592	608×608	9.80	23.40	5.00	2021
	DFPN <sup>[103]</sup>	-	23 258/15 069	640×640	17.10	29.30	18.10	2023
小目标检测	DNOD <sup>[111]</sup>	YOLOv4	23 258/15 069	1080×540	14.20	31.90	11.00	2021
	DNOD <sup>[111]</sup>	EfficientDet-D7	23 258/15 069	1080×540	12.90	32.00	10.90	2021
复杂背景及遮挡	FiFoNet <sup>[119]</sup>	-	23 258/15 069	最长边 1536	21.30	36.80	22.50	2022
样本不均衡	ClusDet <sup>[129]</sup>	ResNet50	23 258/15 069	600×1000	13.70	26.50	12.50	2019
	DMNet <sup>[120]</sup>	ResNet 50	23 258/15 069	600×1000	14.70	24.60	16.30	2020
	BSSD <sup>[121]</sup>	ResNet 101	23 829/16 580	640×640	19.27	30.71	19.96	2021
	BSSD <sup>[121]</sup>	ResNet 101	23 829/16 580	640×640	18.14	29.37	19.83	2021
	DSHNet <sup>[126]</sup>	ResNet-50	23 258/15 069	600×1000	17.80	30.40	19.70	2021
	GLSAN <sup>[131]</sup>	ResNet-50	23 258/15 069	600×1000	19.00	30.50	21.70	2021
	UCGNet <sup>[132]</sup>	Yolov5	23 258/15 069	1080×540	19.10	36.70	18.00	2021
	PRDet <sup>[124]</sup>	ResNet-50	23 258/15 069	600×1000	19.80	34.10	21.30	2023
	LI 等 <sup>[133]</sup>	DetNet-59	23 258/15 069	1080×540	15.30	29.30	16.20	2023
综合改进	NDFT <sup>[142]</sup>	ResNet101	23 258/15 069	1080×540	52.03	-	-	2019
	PENet <sup>[143]</sup>	-	23 258/15 069	-	<b>67.30</b>	<b>76.30</b>	<b>74.60</b>	2020
	UFPMP-Det <sup>[145]</sup>	ResNet-50	23 258/15 069	600×1000	24.60	38.70	28.00	2022

注: 加粗数据为最优值。

**表 3 VisDrone 数据集上的性能评估**  
**Table 3 Performance evaluation on VisDrone dataset**

对应问题	方法	基础网络	训练/测试	AP	AP <sub>50</sub>	AP <sub>75</sub>	AR <sub>1</sub>	AR <sub>10</sub>	AR <sub>100</sub>	AR <sub>500</sub>	发表时间/年
尺度和视角变化	DCNet <sup>[88]</sup>	CenterNet	6 471/1 610	29.43	-	-	-	-	-	-	2021
	RRNet <sup>[98]</sup>	Hourglass	6 741/1580	29.13	55.82	27.23	<b>1.02</b>	<b>8.50</b>	35.19	46.05	2019
	SAMFR <sup>[99]</sup>	DetNet-59	6 471/548	33.72	58.62	33.88	0.53	3.40	22.60	46.03	2019
	SAMFR <sup>[99]</sup>	DetNet-59	6 471/1 580	20.18	40.03	18.42	0.46	3.49	21.60	30.82	2019
	ECascade-RCNN <sup>[100]</sup>	Trident-FPN	6 371/521	28.40	-	-	-	-	-	-	2021
	DSYOLOv3 <sup>[101]</sup>	YOLOv3	6 471/548	22.30	44.50	20.30	-	-	-	-	2021
	SPB-YOLO <sup>[102]</sup>	YOLOv5	6 471/1 580	40.10	-	-	-	-	-	-	2021
	DFPN <sup>[103]</sup>	-	6 471/548	30.30	51.90	30.50	-	-	-	-	2023
	Zhang 等 <sup>[107]</sup>	ResNet50+RPN	6 471/1 580	22.61	45.16	19.94	0.42	2.84	17.10	35.27	2019
	MPFPN <sup>[108]</sup>	ResNet-101	6 471/1 580	29.05	54.38	26.99	0.55	5.81	<b>35.57</b>	45.69	2020
小目标检测	Jadhav 等 <sup>[109]</sup>	ResNet-50	6 471/1 580	11.19	25.65	8.78	0.56	4.87	17.19	24.09	2020
	HRDNet <sup>[110]</sup>	ResNeXt50+101	3 564/1 725	35.51	62.00	35.13	-	-	-	-	2021
	DNOD <sup>[111]</sup>	YOLOv4	6 471/1 610	54.88	-	-	-	-	-	-	2021
	DNOD <sup>[111]</sup>	EfficientDet-D7	6 471/1 610	53.76	-	-	-	-	-	-	2021
	Shang 等 <sup>[112]</sup>	YOLOv5s	6.471/1 610	36.40	-	-	-	-	-	-	2023
复杂背景及遮挡	Zhao 等 <sup>[113]</sup>	YOLOv7	6 471/548	<b>56.80</b>	-	-	-	-	-	-	2023
	YOLOv7-UAV <sup>[114]</sup>	YOLOv7	-	45.30	-	-	-	-	-	-	2024
	D-A-FS SSD <sup>[117]</sup>	VGG16	6 471/548	36.70	-	-	-	-	-	-	2020
	FiFoNet <sup>[119]</sup>	-	6 471/548	36.91	63.80	36.11	-	-	-	-	2022
	ClusDet <sup>[129]</sup>	ResNeXt101	6 471/548	32.40	56.20	31.60	-	-	-	-	2019
	Hong 等 <sup>[125]</sup>	ResNet-101	6.471/1 610	29.13	54.70	27.38	0.32	1.48	9.46	44.53	2019
	Hong 等 <sup>[125]</sup>	ResNet-101	6 471/548	37.15	65.54	36.56	0.32	1.47	7.28	<b>53.78</b>	2019
样本不均衡	CRENet <sup>[130]</sup>	Hourglass-104	6 471/548	33.70	54.30	33.50	-	-	-	-	2020
	DMNet <sup>[120]</sup>	ResNeXt 101	6 471/548	29.40	49.30	30.60	-	-	-	-	2020
	DSHNet <sup>[126]</sup>	ResNet-50	6 471/548	30.30	51.80	30.90	-	-	-	-	2021
	GLSAN <sup>[131]</sup>	ResNet-50	6 471/548	32.50	55.80	33.00	-	-	-	-	2021
	UCGNet <sup>[132]</sup>	Yolov5	6 471/548	32.80	53.10	33.90	-	-	-	-	2021
	VAMYOLOX <sup>[123]</sup>	Darknet53	6 471/548	29.40	47.00	-	-	-	-	-	2023
	PRDet <sup>[124]</sup>	ResNeXt-101	6 471/548	40.20	62.00	43.50	-	-	-	-	2023
实时检测	Li 等 <sup>[133]</sup>	DetNet59	6 471/548	31.40	54.50	27.30	-	-	-	-	2023
	SlimYOLOv3 <sup>[135]</sup>	YOLOv3	6 471/548	23.90	-	-	-	-	-	-	2019
	LAI-YOLOv5s <sup>[140]</sup>	YOLOv5	6 471/548	-	40.40	-	-	-	-	-	2023
	Cao 等 <sup>[141]</sup>	YOLOv5	6 471/549	27.70	46.90	-	-	-	-	-	2023
	NDFT <sup>[142]</sup>	ResNet101	6 471/548	52.77	-	-	-	-	-	-	2019
综合改进	PENet <sup>[143]</sup>	-	6 471/548	41.10	58.00	<b>44.30</b>	-	-	-	-	2020
	SyNet <sup>[144]</sup>	CenterNet	6 471/1580	25.10	48.40	26.20	-	-	-	-	2021
	UFPMP-Det <sup>[145]</sup>	ResNeXt-101	6 471/548	40.10	<b>66.80</b>	41.30	-	-	-	-	2022
	YOLO-UAV <sup>[146]</sup>	YOLOv5l	6 471/548	30.50	-	-	-	-	-	-	2022
	YOLOv7X <sup>[147]</sup>	YOLOv7	6 471/548	-	60.30	-	-	-	-	-	2023

注：加粗数据为最优值。

## 4.2 性能评估

### 4.2.1 评价指标

采用平均精度(average precision, AP)和平均召回率(average recall, AR)指标来衡量无人机图像目标检测性能。AP 表示某个 IoU 阈值下的准确率-召回率曲线下的面积, AP<sub>50</sub> 和 AP<sub>75</sub> 表示在计算精度时设置的 IOU 阈值分别为 0.50 和 0.75, AP=

AP<sup>IoU=0.50:0.05:0.95</sup>, 表示在步长为 0.05 的情况下, IoU 阈值从 0.50 到 0.95 的获得的 10 个平均精度的平均值。对于 AR, AR<sub>n</sub> 表示最大检测次数为 n 时的平均召回率。

### 4.2.2 算法性能

表 2 和表 3 分别列出了先进的目标检测算法在 2 个常用的无人机航拍图像数据集 UAVDT 和

VisDrone 上的检测结果。由表 2 可知, PENet<sup>[143]</sup>在 AP, AP<sub>50</sub> 和 AP<sub>75</sub> 指标上均取得了最好的性能, 且超出其他模型 10%以上。由于 UAVDT 是在不同天气条件和不同视角下拍摄的, 因此存在严重的样本不平衡和多尺度问题。对此, PENet 首先扩充了不平衡数据集, 然后采用“由粗到细, 粗细结合”的无锚点检测方式, 应对不同尺寸的目标。其消融实验表明, 重采样模块、粗粒度检测模块和细粒度检测模块在 UAVDT 检测中均发挥重要的作用。

由表 3 可知, RRNet<sup>[98]</sup>在 AR<sub>1</sub> 和 AR<sub>10</sub> 指标上取得了最佳性能, 在 AR<sub>50</sub> 和 AR<sub>75</sub> 指标上取得了接近最佳的性能。与 PENet 类似, RRNet 同样使用重采样策略增强目标, 并将无锚点检测器与再回归模块结合, 从而更精准地识别密集场景中的目标, 有效防止漏检。但 RRNet 检测精度不够高, 存在大量误检。文献[113]使用先进的 YOLOv7 模型, 引入非跨行卷积模块、多尺度注意模块和可变形注意模块, 通过多方面的优化增强网络对小目标特征的提取能力, 因此取得了最高的平均精度。

此外, NDFT<sup>[142]</sup>将图像中的无关信息看作细粒度领域, 学习领域鲁棒的特征, 在 UAVDT 学习后迁移至 VisDrone, 在 2 个数据集上均取得了较好的性能。测试结果表明, NDFT 提取了任务有关但领域无关的特征, 摆脱了图像干扰, 并具有很好的迁移性。

## 5 结论

无人机搭载摄像头能够获取覆盖范围更大、信息更丰富的航拍图像, 因此, 面向无人机航拍图像的目标检测受到越来越多的关注, 为各领域的智能化决策提供基础支持。相较于普通视角下的图像, 无人机视角下的图像具有目标小且聚集、视角和尺度变化等特点, 通用的目标检测方法无法直接适用。基于此, 本文回顾了普通视角下目标检测的传统方法、深度学习方法和基于大模型的方法, 并系统探讨了针对无人机航拍图像目标检测 6 个难点问题的改进策略和优化方法, 为无人机目标检测的发展和应用提供重要参考。

尽管无人机视角下的目标检测已经取得了显著进展, 但当前研究仍然存在诸多问题, 如航拍数据收集和标注困难、低质量图像影响检测性能、难以在简易无人机平台上实现实时高精度检测等。因此, 本文对未来的研究趋势做出如下展望:

1) 数据集制作。现有无人机目标检测数据集受限于拍摄场景, 不同类别的目标数量存在差异, 导致模型在训练过程中存在一定偏好。一方面需要积极应对数据集中存在的长尾问题, 可以采用生成式算法或数据训练均衡化策略减轻不均衡数据的影响; 另一方面应综合使用不同数据集来训练和验证模型, 进一步提升模型的通用性。

2) 基于深度学习的图像修复方法。深度学习方法能有效处理大规模数据, 自动提取图像特征, 具有较强的适应性和灵活性。然而, 深度学习方法的性能依赖于数据质量和优化策略。因此, 可以通过图像增广方法适用于不同质量的数据, 或者将深度学习方法与图像预处理算法结合, 通过对去噪、增强对比度、锐化等技术, 改善图像的质量。

3) 多模态融合。现有方法大多仅使用单一模态, 特别是 RGB 彩色图像作为输入, 在复杂、极端的环境下存在局限性。未来, 无人机目标检测系统可以协同利用红外、多光谱、高动态范围成像等数据, 通过设计高效的多模态特征提取和融合网络及模态对齐算法, 克服不同环境对检测性能的影响。

4) 减少标签依赖。为了应用大规模无标注数据, 可以结合上下文信息、目标长宽比、目标之间的共现关系等先验知识设计辅助任务, 自适应地调整任务优先级, 挖掘数据的潜在信息, 减少对标签的依赖, 实现半监督、无监督或自监督, 从大规模无标注中学习泛化性更强的模型。

5) 快速高效检测。受限于无人机的计算和存储资源, 如何在检测性能和速度之间取得良好平衡是一项具有挑战性的工作<sup>[163]</sup>。可以通过设计轻量的特征提取网络、结构化和非结构化剪枝相结合、软硬件协同设计等方式, 保持模型精度, 减少模型参数数量和计算量。

6) 联合优化。设计端到端的深度学习模型, 同时考虑目标检测和图像修复任务, 使数据和模型相互促进, 提高整体性能; 开发包含传感器配置、数据处理、模型设计的全流程优化框架, 实现端到端的联合优化; 优化通信与计算策略, 探索多无人机协同的分布式目标检测。

7) 利用大模型优势。大模型技术的快速发展为面向无人机航拍图像的目标检测带来了新的机遇和可能性。①对现有视觉大模型进行微调, 迁移至无人机图像目标检测任务中; ②利用大模型指导数据增强, 利用生成式大模型生成多样化的训练数

据；③利用大模型的跨模态理解能力，实现基于文本描述的零样本或跨模态目标检测；④融合图像、文本等多种信息源，结合自然语言处理和计算机视觉的技术优势，训练泛化性和灵活性更强、更适用于开集目标检测的多模态大模型。

## 参考文献 (References)

- [1] LIU L, OUYANG W L, WANG X G, et al. Deep learning for generic object detection: a survey[J]. International Journal of Computer Vision, 2020, 128(2): 261-318.
- [2] ZOU Z X, CHEN K Y, SHI Z W, et al. Object detection in 20 years: a survey[J]. Proceedings of the IEEE, 2023, 111(3): 257-276.
- [3] MOHSAN S A H, KHAN M A, NOOR F, et al. Towards the unmanned aerial vehicles (UAVs): a comprehensive review[J]. Drones, 2022, 6(6): 147.
- [4] KANELAKIS C, NIKOLAKOPOULOS G. Survey on computer vision for UAVs: current developments and trends[J]. Journal of Intelligent & Robotic Systems, 2017, 87(1): 141-168.
- [5] CHEN C J, HUANG Y Y, LI Y S, et al. Identification of fruit tree pests with deep learning on embedded drone to achieve accurate pesticide spraying[J]. IEEE Access, 2021, 9: 21986-21997.
- [6] PROSEKOV A, VESNINA A, ATUCHIN V, et al. Robust algorithms for drone-assisted monitoring of big animals in harsh conditions of Siberian winter forests: recovery of European elk (*Alces alces*) in Salair mountains[J]. Animals, 2022, 12(12): 1483.
- [7] CHEN Y F, ZHENG W Q, ZHAO Y Y, et al. DW-YOLO: an efficient object detector for drones and self-driving vehicles[J]. Arabian Journal for Science and Engineering, 2023, 48(2): 1427-1436.
- [8] LYGOURAS E, SANTAVAS N, TAITZOGLOU A, et al. Unsupervised human detection with an embedded vision system on a fully autonomous UAV for search and rescue operations[J]. Sensors, 2019, 19(16): 3542.
- [9] 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.
- [10] WU X, LI W, HONG D F, et al. Deep learning for unmanned aerial vehicle-based object detection and tracking: a survey[J]. IEEE Geoscience and Remote Sensing Magazine, 2022, 10(1): 91-124.
- [11] RAMACHANDRAN A, SANGAIAH A K. A review on object detection in unmanned aerial vehicle surveillance[J]. International Journal of Cognitive Computing in Engineering, 2021, 2: 215-228.
- [12] BABARYKA A, KATERYNCHUK I, KLYMASH M, et al. Deep learning methods application for object detection tasks using unmanned aerial vehicles[C]//2022 IEEE 16th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering . New York: IEEE Press, 2022: 808-811.
- [13] TANG G Y, NI J J, ZHAO Y H, et al. A survey of object detection for UAVs based on deep learning[J]. Remote Sensing, 2023, 16(1): 149.
- [14] SU J Y, ZHU X Y, LI S H, et al. AI meets UAVs: a survey on AI empowered UAV perception systems for precision agriculture[J]. Neurocomputing, 2023, 518: 242-270.
- [15] BOUGUETTAYA A, ZARZOUR H, KECHIDA A, et al. Deep learning techniques to classify agricultural crops through UAV imagery: a review[J]. Neural Computing and Applications, 2022, 34(12): 9511-9536.
- [16] SRIVASTAVA S, NARAYAN S, MITTAL S. A survey of deep learning techniques for vehicle detection from UAV images[J]. Journal of Systems Architecture, 2021, 117: 102152.
- [17] DOLL O, LOOS A. Comparison of object detection algorithms for livestock monitoring of sheep in UAV images[C]// Workshop Camera Traps, AI, and Ecology. Cham: Springer, 2023: 1-7.
- [18] ZHAO C J, LIU R W, QU J X, et al. Deep learning-based object detection in maritime unmanned aerial vehicle imagery: review and experimental comparisons[J]. Engineering Applications of Artificial Intelligence, 2024, 128: 107513.
- [19] DASMEHDIXTR. Drone Dataset (UAV)[EB/OL]. [2024-07-23]. [https://www.kaggle.com/datasets/dasmehdixtr/drone-dataset-uav?select=drone\\_dataset\\_yolo](https://www.kaggle.com/datasets/dasmehdixtr/drone-dataset-uav?select=drone_dataset_yolo).
- [20] LIU H H, YU Y H, LIU S Z, et al. A military object detection model of UAV reconnaissance image and feature visualization[J]. Applied Sciences, 2022, 12(23): 12236.
- [21] TENG S Z, ZHANG S L, HUANG Q M, et al. Viewpoint and scale consistency reinforcement for UAV vehicle re-identification[J]. International Journal of Computer Vision, 2021, 129(3): 719-735.
- [22] TIAN G Y, LIU J R, ZHAO H, et al. Small object detection via dual inspection mechanism for UAV visual images[J]. Applied Intelligence, 2022, 52(4): 4244-4257.
- [23] WANG C Y, SHI Z R, MENG L L, et al. Anti-occlusion UAV tracking algorithm with a low-altitude complex background by integrating attention mechanism[J]. Drones, 2022, 6(6): 149.
- [24] SHAN P, YANG R G, XIAO H M, et al. UAVNet: a balanced and enhanced UAV object detection and pose recognition network[J]. Measurement, 2023, 222: 113654.
- [25] 李斌, 张彩霞, 杨阳, 等. 复杂场景下深度表示的无人机目标检测算法[J]. 计算机工程与应用, 2020, 56(15): 118-123.
- [26] LI B, ZHANG C X, YANG Y, et al. Drone target detection algorithm for depth representation in complex scene[J]. Journal of Computer Engineering and Applications, 2020, 56(15): 118-123 (in Chinese).
- [27] YE T, QIN W Y, ZHAO Z Y, et al. Real-time object detection network in UAV-vision based on CNN and transformer[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 2505713.
- [28] ZHU P F, WEN L Y, DU D W, et al. VisDrone-DET2018: the vision meets drone object detection in image challenge results[C]//The European Conference on Computer Vision Workshops. Cham: Springer, 2019: 437-468.
- [29] WU X W, SAHOO D, HOI S C H. Recent advances in deep learning for object detection[J]. Neurocomputing, 2020, 396: 39-64.
- [30] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International journal of Computer Vision, 2004, 60(2): 91-110.
- [31] BAY H, ESS A, TUYTELAARS T, et al. Speeded-up robust features (SURF)[J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.
- [32] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2005: 886-893.
- [33] FELZENSZWALB P, MCALLESTER D, RAMANAN D. A discriminatively trained, multiscale, deformable part

- model[C]//2008 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2008: 1-8.
- [34] FREUND Y, SCHAPIRE R E. A decision-theoretic generalization of on-line learning and an application to boosting[J]. Journal of Computer and System Sciences, 1997, 55(1): 119-139.
- [35] HEARST M A, DUMAIS S T, OSUNA E, et al. Support vector machines[J]. IEEE Intelligent Systems and their Applications, 1998, 13(4): 18-28.
- [36] 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. New York: IEEE Press, 2014: 580-587.
- [37] 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.
- [38] GIRSHICK R. Fast R-CNN[C]//The IEEE International Conference on Computer Vision. New York: IEEE Press, 2015: 1440-1448.
- [39] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//The 28th International Conference on Neural Information Processing Systems. New York: ACM, 2015: 91-99.
- [40] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision. New York: IEEE Press, 2017: 2980-2988.
- [41] 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. New York: IEEE Press, 2018: 6154-6162.
- [42] 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. New York: IEEE Press, 2017: 936-944.
- [43] 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. New York: IEEE Press, 2016: 779-788.
- [44] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2017: 6517-6525.
- [45] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2024-06-01]. <https://arxiv.org/abs/1804.02767>.
- [46] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2024-06-01]. <https://arxiv.org/abs/2004.10934>.
- [47] Ultralytics. YOLOv5 in PyTorch[EB/OL]. [2024-07-23]. <https://github.com/ultralytics/yolov5>.
- [48] LI C Y, LI L L, JIANG H L, et al. YOLOv6: a single-stage object detection framework for industrial applications[EB/OL]. (2022-09-07)[2024-06-01]. <https://arxiv.org/abs/2209.02976>.
- [49] 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. New York: IEEE Press, 2023: 7464-7475.
- [50] Ultralytics. YOLOv8[EB/OL]. [2024-07-23]. <https://docs.ultralytics.com/zh/models/yolov8/>.
- [51] WANG C Y, YEH I H, LIAO H Y M. YOLOv9: learning what you want to learn using programmable gradient information[EB/OL]. (2024-02-29)[2024-06-01]. <https://arxiv.org/abs/2402.13616>.
- [52] WANG A, CHEN H, LIU L H, et al. YOLOv10: real-time end-to-end object detection[EB/OL]. (2024-05-23)[2024-06-01]. <https://arxiv.org/abs/2405.14458>.
- [53] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//The 14th European Conference on Computer Vision-ECCV 2016. Cham: Springer, 2016: 21-37.
- [54] FU C Y, LIU W, RANGA A, et al. DSSD: deconvolutional single shot detector[EB/OL]. (2017-01-23)[2024-06-01]. <https://arxiv.org/abs/1701.06659>.
- [55] LI Z X, YANG L, ZHOU F Q. FSSD: feature fusion single shot multibox detector[EB/OL]. (2024-02-23)[2024-06-01]. <https://arxiv.org/abs/1712.00960>.
- [56] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision. New York: IEEE Press, 2017: 2999-3007.
- [57] LAW H, DENG J. CornerNet: detecting objects as paired keypoints[C]//The 15th European Conference on Computer Vision . Cham: Springer, 2018: 765-781.
- [58] ZHOU X Y, WANG D Q, KRÄHENBÜHL P. Objects as points[EB/OL]. (2019-04-25)[2024-06-01]. <https://arxiv.org/abs/1904.07850>.
- [59] 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 . New York: IEEE Press, 2020: 10778-10787.
- [60] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//The 16th European Conference on Computer Vision. Cham: Springer, 2020: 213-229.
- [61] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//The 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 6000-6010.
- [62] ZHU X Z, SU W J, LU L W, et al. Deformable DETR: deformable transformers for end-to-end object detection[EB/OL]. [2024-06-01]. <https://dblp.uni-trier.de/db/conf/iclr/iclr2021.html#ZhuSLLWD21>.
- [63] ZHAO Y A, LV W Y, XU S L, et al. DETRs beat YOLOs on real-time object detection[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2024: 16965-16974.
- [64] YAO J, YI X J, WANG X T, et al. From instructions to intrinsic human values--a survey of alignment goals for big models[EB/OL]. (2023-11-04)[2024-06-01]. <https://arxiv.org/abs/2308.12014>.
- [65] LI L H, ZHANG P C, ZHANG H T, et al. Grounded language-image pre-training[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2022: 10955-10965.
- [66] LIU S L, ZENG Z Y, REN T H, et al. Grounding DINO: marrying DINO with grounded pre-training for open-set object detection[EB/OL]. (2024-07-19)[2024-06-01]. <https://arxiv.org/abs/2303.05499>.
- [67] KAUL P, XIE W D, ZISSERMAN A. Multi-modal classifiers for open-vocabulary object detection[EB/OL]. [2024-06-01]. <https://dl.acm.org/doi/10.5555/3618408.3619063>.
- [68] ZANG Y H, LI W, HAN J, et al. Contextual object detection with multimodal large language models[EB/OL]. (2024-08-12)[2024-06-01]. <https://arxiv.org/abs/2305.18279>.
- [69] XU Y F, ZHANG M D, FU C Y, et al. Multi-modal queried object detection in the wild[C]//The 37th International Conference on Neural Information Processing Systems. New York: ACM, 2023: 198.
- [70] ZHAO T, NEVATIA R. Car detection in low resolution aerial

- images[J]. *Image and Vision Computing*, 2003, 21(8): 693-703.
- [71] KLUCKNER S, PACHER G, GRABNER H, et al. A 3D teacher for car detection in aerial images[C]//2007 IEEE 11th International Conference on Computer Vision. New York: IEEE Press, 2007: 1-8.
- [72] MORANDUZZO T, MELGANI F. A SIFT-SVM method for detecting cars in UAV images[C]//2012 IEEE International Geoscience and Remote Sensing Symposium. New York: IEEE Press, 2012: 6868-6871.
- [73] MORANDUZZO T, MELGANI F. Automatic car counting method for unmanned aerial vehicle images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, 52(3): 1635-1647.
- [74] STOKKEL X L X. Detecting humans from a top-down perspective using an unmanned aerial vehicle[D]. Groningen: University of Groningen, 2015.
- [75] MORANDUZZO T, MELGANI F, BAZI Y, et al. A fast object detector based on high-order gradients and Gaussian process regression for UAV images[J]. *International Journal of Remote Sensing*, 2015, 36(10): 2713-2733.
- [76] XU Y Z, YU G Z, WANG Y P, et al. A hybrid vehicle detection method based on Viola-Jones and HOG + SVM from UAV images[J]. *Sensors*, 2016, 16(8): 1325.
- [77] WANG Y R, ZHU X L, WU B. Automatic detection of individual oil palm trees from UAV images using HOG features and an SVM classifier[J]. *International Journal of Remote Sensing*, 2019, 40(19): 7356-7370.
- [78] MAIRE F, ALVAREZ L M, HODGSON A. Automating marine mammal detection in aerial images captured during wildlife surveys: a deep learning approach[C]//The 28th Australasian Joint Conference on AI 2015: Advances in Artificial Intelligence. Cham: Springer, 2015: 379-385.
- [79] AMMOUR N, ALHICRHI H, BAZI Y, et al. Deep learning approach for car detection in UAV imagery[J]. *Remote Sensing*, 2017, 9(4): 312.
- [80] LI C L, SUN X M, CAI J H. Intelligent mobile drone system based on real-time object detection[J]. *Journal on Artificial Intelligence*, 2019, 1(1): 1-8.
- [81] HONG S J, HAN Y, KIM S Y, et al. Application of deep-learning methods to bird detection using unmanned aerial vehicle imagery[J]. *Sensors*, 2019, 19(7): 1651.
- [82] MAKAROV S B, PAVLOV V A, BEZBORODOV A K, et al. Multiple object tracking using convolutional neural network on aerial imagery sequences[C]//International Youth Conference on Electronics, Telecommunications and Information Technologies. Cham: Springer, 2021: 413-420.
- [83] CHENG Z, CHEN J Y, ZHANG X, et al. Comparative study of two target detection algorithms in UAV aerial photography detection[C]//The 12th International Conference on Information Optics and Photonics. Bellingham: SPIE, 2021, 12057: 889-894.
- [84] 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.
- [85] KU C, CHEN X Z, CHEN Y L. Robust object detection model for UAV application[C]//2023 International Automatic Control Conference. New York: IEEE Press, 2023: 1-5.
- [86] WANG W J, PENG Y P, CAO G Z, et al. Low-illumination image enhancement for night-time UAV pedestrian detection[J]. *IEEE Transactions on Industrial Informatics*, 2021, 17(8): 5208-5217.
- [87] WANG J S, YANG Y, CHEN Y, et al. LighterGAN: an illumination enhancement method for urban UAV imagery[J]. *Remote Sensing*, 2021, 13(7): 1371.
- [88] LIU Y, WANG J W, QIU T T, et al. An adaptive deblurring vehicle detection method for high-speed moving drones: resistance to shake[J]. *Entropy*, 2021, 23(10): 1358.
- [89] WANG X Q. Vehicle image detection method using deep learning in UAV video[J]. *Computational Intelligence and Neuroscience*, 2022, 2022(1): 8202535.
- [90] MINH T T, VAN BAO T, NGUYEN V D, et al. An object detection method for aerial hazy images[J]. *Can Tho University Journal of Science*, 2022, 14(1): 91-98.
- [91] QIN X, WANG Z L, BAI Y C, et al. FFA-Net: feature fusion attention network for single image dehazing[C]//The 34th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2020: 11908-11915.
- [92] KIM K, LEE H S. Probabilistic anchor assignment with IoU prediction for object detection[C]//The 16th European Conference on Computer Vision. Cham: Springer, 2020: 355-371.
- [93] ZHANG L M, WANG G F, CHEN M, et al. An enhanced noise-tolerant hashing for drone object detection[J]. *Pattern Recognition*, 2023, 143: 109762.
- [94] ZHU B Y, LV Q B, TAN Z. Adaptive multi-scale fusion blind deblurred generative adversarial network method for sharpening image data[J]. *Drones*, 2023, 7(2): 96.
- [95] CHENG G, ZHOU P C, HAN J W. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 54(12): 7405-7415.
- [96] DING J, XUE N, LONG Y, et al. Learning RoI transformer for oriented object detection in aerial images[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2019: 2844-2853.
- [97] PAN X J, REN Y Q, SHENG K K, et al. Dynamic refinement network for oriented and densely packed object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2020: 11204-11213.
- [98] CHEN C R, ZHANG Y, LV Q X, et al. RRNet: a hybrid detector for object detection in drone-captured images[C]//2019 IEEE/CVF International Conference on Computer Vision Workshop. New York: IEEE Press, 2019: 100-108.
- [99] WANG H R, WANG Z X, JIA M X, et al. Spatial attention for multi-scale feature refinement for object detection[C]//2019 IEEE/CVF International Conference on Computer Vision Workshop. New York: IEEE Press, 2019: 64-72.
- [100] 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 . New York: IEEE Press, 2021: 268-272.
- [101] LI Z K, LIU X L, ZHAO Y, et al. A lightweight multi-scale aggregated model for detecting aerial images captured by UAVs[J]. *Journal of Visual Communication and Image Representation*, 2021, 77: 103058.
- [102] WANG X R, LI W H, GUO W, et al. SPB-YOLO: an efficient real-time detector for unmanned aerial vehicle images[C]// 2021 International Conference on Artificial Intelligence in Information and Communication. New York: IEEE Press, 2021: 99-104.
- [103] SUN H K, CHEN Y X, LU X B, et al. Decoupled feature pyramid learning for multi-scale object detection in low-altitude remote sensing images[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2023, 16: 6556-6567.
- [104] ZHANG Y Z, WU C Y, ZHANG T, et al. Full-scale feature

- aggregation and grouping feature reconstruction-based UAV Image target detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 5621411.
- [105] LIU F, YAO L, ZHANG C Y, et al. Scale-invariant feature disentanglement via adversarial learning for UAV-based object detection[EB/OL]. (2024-05-31) [2024-06-01]. <https://arxiv.org/abs/2405.15465>.
- [106] LIANG X, ZHANG J, ZHUO L, et al. Small object detection in unmanned aerial vehicle images using feature fusion and scaling-based single shot detector with spatial context analysis[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(6): 1758-1770.
- [107] ZHANG X D, IZQUIERDO E, CHANDRAMOULI K. Dense and small object detection in UAV vision based on cascade network[C]//2019 IEEE/CVF International Conference on Computer Vision Workshop. New York: IEEE Press, 2019: 118-126.
- [108] LIU Y J, YANG F B, HU P. Small-object detection in UAV-captured images via multi-branch parallel feature pyramid networks[J]. IEEE Access, 2020, 8: 145740-145750.
- [109] JADHAV A, MUKHERJEE P, KAUSHIK V, et al. Aerial multi-object tracking by detection using deep association networks[C]//2020 National Conference on Communications. New York: IEEE Press, 2020: 1-6.
- [110] LIU Z M, GAO G Y, SUN L, et al. HRDNet: high-resolution detection network for small objects[C]//2021 IEEE International Conference on Multimedia and Expo. New York: IEEE Press, 2021: 1-6.
- [111] TIAN G Y, LIU J R, YANG W Y. A dual neural network for object detection in UAV images[J]. Neurocomputing, 2021, 443: 292-301.
- [112] SHANG J C, WANG J S, LIU S B, et al. Small target detection algorithm for UAV aerial photography based on improved YOLOv5s[J]. Electronics, 2023, 12(11): 2434.
- [113] ZHAO D W, SHAO F M, LIU Q, et al. A small object detection method for drone-captured images based on improved YOLOv7[J]. Remote Sensing, 2024, 16(6): 1002.
- [114] LI X M, WEI Y K, LI J H, et al. Improved YOLOv7 algorithm for small object detection in unmanned aerial vehicle image scenarios[J]. Applied Sciences, 2024, 14(4): 1664.
- [115] YANG J X, XIE X M, YANG W Z. Effective contexts for UAV vehicle detection[J]. IEEE Access, 2019, 7: 85042-85054.
- [116] ZHANG W, LIU C S, CHANG F L, et al. Multi-scale and occlusion aware network for vehicle detection and segmentation on UAV aerial images[J]. Remote Sensing, 2020, 12(11): 1760.
- [117] LIU Y Z, DING Z M, CAO Y, et al. Multi-scale feature fusion UAV image object detection method based on dilated convolution and attention mechanism[C]//2020 8th International Conference on Information Technology: IoT and Smart City. New York: ACM, 2020: 125-132.
- [118] CAI Y Q, DU D W, ZHANG L B, et al. Guided attention network for object detection and counting on drones[C]//The 28th ACM International Conference on Multimedia. New York: ACM, 2020: 709-717.
- [119] XI Y, JIA W J, MIAO Q G, et al. FiFoNet: fine-grained target focusing network for object detection in UAV images[J]. Remote Sensing, 2022, 14(16): 3919.
- [120] LI C L, YANG T J N, ZHU S J, et al. Density map guided object detection in aerial images[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New York: IEEE Press, 2020: 737-746.
- [121] LI X H, LI X D, LI Z J, et al. Robust vehicle detection in high-resolution aerial images with imbalanced data[J]. IEEE Transactions on Artificial Intelligence, 2021, 2(3): 238-250.
- [122] SHRIVASTAVA A, GUPTA A, GIRSHICK R. Training region-based object detectors with online hard example mining[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2016: 761-769.
- [123] YANG Y H, GAO X Z, WANG Y, et al. VAMYOLOX: an accurate and efficient object detection algorithm based on visual attention mechanism for UAV optical sensors[J]. IEEE Sensors Journal, 2023, 23(11): 11139-11155.
- [124] LENG J X, MO M J C, ZHOU Y H, et al. Pareto refocusing for drone-view object detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(3): 1320-1334.
- [125] HONG S, KANG S, CHO D. Patch-level augmentation for object detection in aerial images[C]//2019 IEEE/CVF International Conference on Computer Vision Workshop. New York: IEEE Press, 2019: 127-134.
- [126] YU W P, YANG T J N, CHEN C. Towards resolving the challenge of long-tail distribution in UAV images for object detection[C]//2021 IEEE Winter Conference on Applications of Computer Vision. New York: IEEE Press, 2021: 3257-3266.
- [127] YAMANI A, ALYAMI A, LUQMAN H, et al. Active learning for single-stage object detection in UAV images[C]//2024 IEEE/CVF Winter Conference on Applications of Computer Vision. New York: IEEE Press, 2024: 1849-1858.
- [128] HOU X Y, ZHANG K L, XU J H, et al. Object detection in drone imagery via sample balance strategies and local feature enhancement[J]. Applied Sciences, 2021, 11(8): 3547.
- [129] YANG F, FAN H, CHU P, et al. Clustered object detection in aerial images[C]//2019 IEEE/CVF International Conference on Computer Vision. New York: IEEE Press, 2019: 8310-8319.
- [130] WANG Y, YANG Y L, ZHAO X. Object detection using clustering algorithm adaptive searching regions in aerial images[C]//European Conference on Computer Vision. Cham: Springer, 2020: 651-664.
- [131] DENG S T, LI S, XIE K, et al. A global-local self-adaptive network for drone-view object detection[J]. IEEE Transactions on Image Processing, 2021, 30: 1556-1569.
- [132] LIAO J J, PIAO Y C, SU J H, et al. Unsupervised cluster guided object detection in aerial images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 11204-11216.
- [133] LI G X, WANG X J, LI Y, et al. Adaptive clustering object detection method for UAV images under long-tailed distributions[J]. Information Technology and Control, 2023, 52(4): 1025-1044.
- [134] KYRKOU C, PLASTIRAS G, THEOCHARIDES T, et al. DroNet: efficient convolutional neural network detector for real-time UAV applications[C]//2018 Design, Automation & Test in Europe Conference & Exhibition. New York: IEEE Press, 2018: 967-972.
- [135] ZHANG P Y, ZHONG Y X, LI X Q. SlimYOLOv3: narrower, faster and better for real-time UAV applications[C]//2019 IEEE/CVF International Conference on Computer Vision Workshop. New York: IEEE Press, 2019: 37-45.
- [136] LI M L, ZHAO X K, LI J S, et al. ComNet: combinational neural network for object detection in UAV-borne thermal images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(8): 6662-6673.
- [137] DONG J, OTA K, DONG M X. Real-time survivor detection in UAV thermal imagery based on deep learning[C]//2020 16th International Conference on Mobility, Sensing and Networking. New York: IEEE Press, 2020: 352-359.
- [138] WANG S Y, ZHAO J, TA N, et al. A real-time deep learning forest fire monitoring algorithm based on an improved Pruned + KD model[J]. Journal of Real-Time Image Processing, 2021, 18(6): 2319-2329.

- [139] QIN Z W, WANG W S, DAMMER K H, et al. Ag-YOLO: a real-time low-cost detector for precise spraying with case study of palms[J]. *Frontiers in Plant Science*, 2021, 12: 753603.
- [140] DENG L X, BI L Y, LI H Q, et al. Lightweight aerial image object detection algorithm based on improved YOLOv5s[J]. *Scientific Reports*, 2023, 13(1): 7817.
- [141] 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.
- [142] WU Z Y, SURESH K, NARAYANAN P, et al. Delving into robust object detection from unmanned aerial vehicles: a deep nuisance disentanglement approach[C]//2019 IEEE/CVF International Conference on Computer Vision. New York: IEEE Press, 2019: 1201-1210.
- [143] TANG Z Y, LIU X, SHEN G Y, et al. PENet: object detection using points estimation in aerial images[EB/OL]. [2020-01-22][2024-06-01]. <https://arxiv.org/abs/2001.08247>.
- [144] ALBABA B M, OZER S. SyNet: an ensemble network for object detection in UAV images[C]//2020 25th International Conference on Pattern Recognition. New York: IEEE Press, 2021: 10227-10234.
- [145] HUANG Y C, CHEN J X, HUANG D. UFPMP-Det: toward accurate and efficient object detection on drone imagery[C]//The 36th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2022: 1026-1033.
- [146] 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.
- [147] LI S Q, LIU W S. Small target detection model in aerial images based on YOLOv7X+[J]. *Engineering Letters*, 2024, 32(2): 436-443.
- [148] 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. New York: IEEE Press, 2017: 4165-4173.
- [149] DU D W, QI Y K, YU H Y, et al. The unmanned aerial vehicle benchmark: object detection and tracking[C]//The 15th European Conference on Computer Vision. Cham: Springer, 2018: 375-391.
- [150] 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.
- [151] 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. New York: IEEE Press, 2019: 213-226.
- [152] CAO Y R, HE Z Y, WANG L J, et al. VisDrone-DET2021: the vision meets drone object detection challenge results[C]//2021 IEEE/CVF International Conference on Computer Vision. New York: IEEE Press, 2021: 2847-2854.
- [153] XU X W, ZHANG X Y, YU B, et al. DAC-SDC low power object detection challenge for UAV applications[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(2): 392-403.
- [154] 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. New York: IEEE Press, 2020: 8504-8510.
- [155] ZHANG H J, SUN M S, LI Q, et al. An empirical study of multi-scale object detection in high resolution UAV images[J]. *Neurocomputing*, 2021, 421: 173-182.
- [156] SUN Y M, CAO B, ZHU P F, et al. Drone-based RGB-infrared cross-modality vehicle detection via uncertainty-aware learning[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(10): 6700-6713.
- [157] AKSHATHA K R, KARUNAKAR A K, SHENOY B S, et al. Manipal-UAV person detection dataset: a step towards benchmarking dataset and algorithms for small object detection[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2023, 195: 77-89.
- [158] 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. New York: IEEE Press, 2022: 3686-3696.
- [159] FENG H T, ZHANG L, ZHANG S Q, et al. RTDOD: a large-scale RGB-thermal domain-incremental object detection dataset for UAVs[J]. *Image and Vision Computing*, 2023, 140: 104856.
- [160] MOU C, LIU T F, ZHU C C, et al. WAID: a large-scale dataset for wildlife detection with drones[J]. *Applied Sciences*, 2023, 13(18): 10397.
- [161] MOKAYED H, NAYEBI STANEH A, DE K, et al. Nordic Vehicle Dataset (NVD): performance of vehicle detectors using newly captured NVD from UAV in different snowy weather conditions[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2023: 5314-5322.
- [162] LIU F, YAO L, XU S X, et al. UEMM-Air: a synthetic multi-modal dataset for unmanned aerial vehicle object detection[EB/OL]. [2024-07-01]. <https://arxiv.org/abs/2406.06230>.
- [163] 李利霞, 王鑫, 王军, 等. 基于特征融合与注意力机制的无人机图像小目标检测算法[J]. *图学学报*, 2023, 44(4): 658-666.  
LI L X, WANG X, WANG J, et al. Small object detection algorithm in UAV image based on feature fusion and attention mechanism[J]. *Journal of Graphics*, 2023, 44(4): 658-666 (in Chinese).