ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 40,No. 2,Mar. 2025,pp. 274－302
DOI:10. 16337/j. 1004-9037. 2025. 02. 002

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

面向无人机的低空视觉数据集研究综述

孙一铭[1]，赵柯嘉[1]，王　硕[1]，陈振国[1]，阮　媛[1]，叶子凡[1]，陈星睿[1]，李　欣[1]，褚瑞麟[1]，宋生敏[1]，胡亦添[1]，郭周鹏[1]，王　森[1]，胡清华[2]，朱鹏飞[2]

(1. 东南大学自动化学院,南京 210096;2. 天津大学智能与计算学部,天津 300354)

摘　要：在无人机技术与人工智能的跨域协同驱动下,依托国家低空经济政策与空域开放试点改革,低空视觉感知在智慧城市及巡检搜救等方面发挥了重要作用。高质量的低空视觉数据是低空智能感知领域的关键基础资源,公开数据集的发布与应用对低空感知技术的深入推进起到了重要作用。尽管已有大量面向低空视觉感知的数据集被提出,但对其系统化的整理与分析尚不充分。针对这一问题,本文全面调研了近11年间公开发布的低空无人机视觉相关数据集,基于不同的数据特征和应用场景对其进行分类探究,并选取具有代表性的数据集进行详细分析。本文涵盖了单机感知、多机协同感知、多任务感知、多源感知、复杂环境特性以及无人机具身智能等多个领域,为便于研究者理解与使用,本文以图表形式对所有数据集的基本信息进行了归纳总结,并从以下两个主要维度对其发展趋势进行了系统分析:(1)元数据分析,包括数据集规模分布、场景分布及支持任务类型等特点;(2)基本信息分析,涉及图像视频总量、目标类别分布和标注实例数量等关键指标。通过分析,充分展示了低空视觉感知数据集质量的显著进步,同时指出尽管已初步形成低空数据体系化架构,但是低空数据标注成本与效率失衡、多源数据复用性不足、极端环境覆盖薄弱以及具身智能数据割裂等问题依旧存在。最后,本文对低空数据集未来发展方向进行了展望。

关键词：低空应用;多机协同;多源感知;多任务感知;无人机具身智能

中图分类号：TP391.4　　**文献标志码**:A

Research Review on Low-Altitude Visual Datasets for Unmanned Aerial Vehicles

SUN Yiming[1], ZHAO Kejia[1], WANG Shuo[1], CHEN Zhenguo[1], RUAN Yuan[1], YE Zifan[1], CHEN Xingrui[1], LI Xin[1], CHU Ruilin[1], SONG Shengmin[1], HU Yitian[1], GUO Zhoupeng[1], WANG Sen[1], HU Qinghua[2], ZHU Pengfei[2]

(1. School of Automation, Southeast University, Nanjing 210096, China; 2. College of Intelligence and Computing, Tianjin University, Tianjin 300354, China)

Abstract: Driven by the cross-domain synergy of unmanned aerial vehicle (UAV) technology and artificial intelligence, and supported by national low-altitude economic policies and pilot reforms for airspace opening, the low-altitude visual perception has played a significant role in smart cities, inspection, rescue, and other applications. High-quality low-altitude visual data serve as the crucial foundational resource in the field of low-altitude intelligent perception, and the release and application of public datasets have been

基金项目:新一代人工智能国家科技重大专项(2022ZD0116500);国家自然科学基金(62222608,62436002)。
收稿日期:2025-02-08;**修订日期**:2025-03-18

pivotal in advancing low-altitude perception technologies. Despite the proposal of numerous datasets for low-altitude visual perception, systematic organization and analysis of these datasets remain inadequate. To address this issue, this paper conducts a comprehensive survey of publicly released low-altitude UAV vision-related datasets over the past 11 years, categorizes and explores them based on different data characteristics and application scenarios, and selects representative datasets for detailed analysis. This review covers multiple domains, including single-UAV perception, multi-UAV cooperative perception, multi-task perception, multi-source perception, complex environmental characteristics, and UAV embodied intelligence. To facilitate researchers' understanding and use, the paper summarizes the basic information of all datasets in graphical form and systematically analyzes their development trends from two main dimensions: (1) metadata analysis, including dataset size distribution, scenario distribution, and supported task types; and (2) basic information analysis, involving total image and video counts, target category distribution, and annotation instance numbers. The analysis fully demonstrates the significant progress in the quality of low-altitude visual perception datasets. Meanwhile, it points out that, despite the initial formation of a systematic framework for low-altitude data, issues such as the imbalance between cost and efficiency in low-altitude data annotation, insufficient reusability of multi-source data, inadequate coverage of extreme environments, and fragmented embodied intelligence data still exist. Finally, this paper proposes outlooks for the future development of low-altitude datasets.

Key words: low-altitude applications; multi-drone collaboration; multi-source perception; multi-task perception; UAV embodied intelligence

引　言

低空空域通常指地面以上1 000 m以内的飞行区域,根据地区特点和需求,该界限可扩展至3 000 m[1]。低空空域受地形起伏、建筑遮蔽及局地气象扰动等多元动态约束,同时承载无人机等新型航空器的密集作业需求,形成空域资源争夺与异构障碍物共存的复杂态势。近年来,国家高度重视低空经济发展[2]:在顶层设计层面,2023~2025年间通过中央经济工作会议、《政府工作报告》等将低空经济纳入战略性新兴产业培育序列;在产业促进维度,实施财税激励引导社会资本向飞行器研发、智能空管等核心环节集聚;在安全治理领域,建立涵盖适航认证、空域准入和飞行监管的全周期框架,显著降低运行风险[3]。

低空视觉感知是低空经济产业发展的关键技术支撑,也是无人机等各类低空智能飞行器获取外界信息的重要手段之一。当下,低空视觉感知技术凭借其独特优势,在目标检测、跟踪、计数、分割、具身导航等关键任务领域实现了深度且广泛的应用,而具身智能的进一步突破将推动无人机从"任务执行工具"向"自主决策体"演进,展现出了极高的实用价值与发展潜力。无人机低空视觉数据集是无人机视觉感知及具身智能等任务的基石,对低空感知大模型的发展起着不可或缺的作用。随着低空视觉感知技术的广泛应用,科研人员和相关机构针对不同任务的独特需求构建了大量具有高度针对性的特定数据集。低空视觉感知技术的突破也得益于高质量标注数据集的支撑,如目标检测、分割等任务通过多场景样本与精准标注,显著提升了算法性能。具身智能领域的环境信息数据集为无人机导航与路径规划提供了决策依据。数据集的持续优化推动着低空视觉感知技术向更高精度和更广应用领域演进[4]。

本文重点围绕无人机低空视觉感知数据集进行探究。首先,从设备类型、任务需求、模态类型、环境特性和应用需求5个核心维度,对低空视觉感知数据集进行了细致的分类与阐释。然后,针对单机、

多机协同、多任务学习、多源融合、复杂环境特性以及无人机具身智能等领域内的典型数据集，从定义、特点、典型实例以及应用场景这4个方面进行了深度剖析。图1展示了低空视觉感知典型数据集的研究发展脉络。最后，本文对低空视觉感知数据集的研究现状进行了总结，并结合当前的技术发展趋势和实际应用需求，对未来发展进行了展望。本文致力于构建一个系统、全面的低空数据集认知框架，这不仅有助于科研人员准确把握低空视觉感知数据集的前沿研究方向，还能为相关应用的开发提供坚实的数据支撑和实践指导，进而推动低空领域的技术创新以及产业的升级发展。

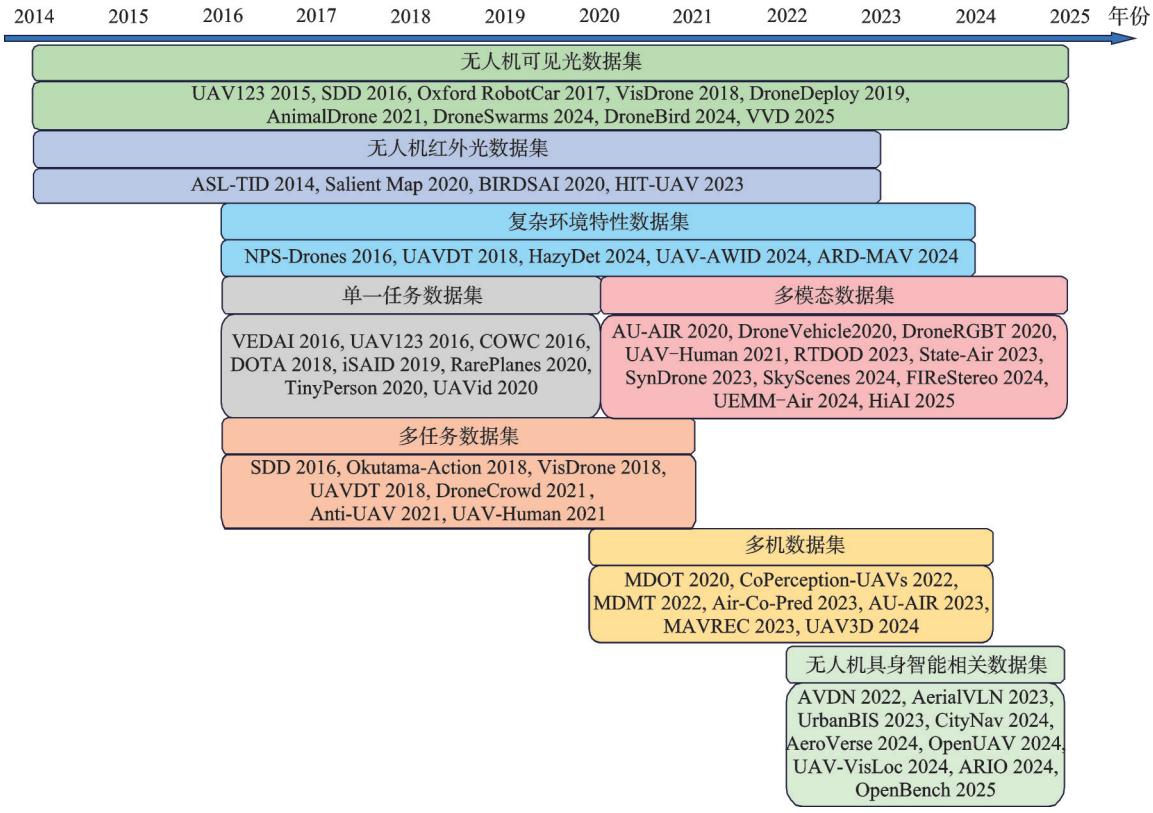


图1 低空视觉数据集研究脉络

Fig.1 Research trajectory of low-altitude visual datasets

1 低空视觉数据集概述

随着低空经济的迅速发展和无人机技术的不断升级，低空视觉数据集作为支撑无人机智能感知与导航的核心基础资源，越来越至关重要。然而，随着任务需求和场景复杂度的提升，单一维度的数据已经无法满足日益增长的多任务需求。因此，构建一个标准化、全面的数据资源框架显得尤为迫切。为此，本文提出了一种基于五大方向的低空视觉感知数据集分类体系，分别从设备类型、任务需求、模态类型、环境特性和应用需求等方面进行全面构建。这一分类体系旨在精准适配不同领域和场景的多样化需求，为未来无人机感知与导航技术的不断发展提供有力的数据支持。

（1）设备差异：单机与多机协同场景对数据同步性、视角覆盖度的要求截然不同。单目标协同与多机群智能的研究分化，驱动了单机独立感知与多机协同感知数据集的构建差异。

（2）任务需求：目标检测需密集标注的边界框，而多任务学习（如目标检测＋目标跟踪）需时序与空间信息的联合标注，目标检测、跟踪等单一任务与检测-分割-计数等多任务联合优化的算法边界，决定

了数据集标注粒度的设计准则。

(3) 环境特性:复杂环境场景会显著增加图像中的干扰信息,提高增加识别和追踪的难度,直接影响数据集的动态范围设计与模型泛化能力评价指标。

(4) 模态类型:可见光、红外、深度等多源数据的获取与融合需求,可突破单一传感器的物理局限(如夜间或遮挡场景),对应复杂场景下的感知容错性提升。

(5) 应用需求:视觉感知与具身导航的功能侧重差异,导致数据集需平衡环境感知信息与无人机本体运动参数的耦合关系。

本文分类框架及典型数据集如表1所示。该分类方式不仅体现了数据集在领域相关性(如工业巡检侧重多机协同,灾害救援依赖恶劣天气数据)和场景适应性(如城市环境需多任务支持,野外导航需

表1 低空视觉数据集分类框架

Table 1 Classification framework for low-altitude visual datasets

分类方向	数据集分类	典型数据集名称
设备类型	单机数据集	UAV123(Mueller等[5], 2016),Oxford RobotCar(Barnes等[6],2017),VisDrone(Du等[7], 2018),AnimalDrone(Zhu等[8],2021),DroneSwarms(Cao等[9],2024),DroneBird(Cao等[10], 2024),Varied Drone Dataset(Xiao等[11],2025)
	多机数据集	MDOT(Zhu等[12],2021),CoPerception-UAVs(Hu等[13],2022),Air-Co-Pred(Wang等[14], 2024),MDMT(Liu等[15],2023),AU-AIR(Bozcan等[16],2020),MAVREC(Dutta等[17],2024), UAV3D(Ye等[18],2025)
任务需求	单一任务数据集	VEDAI(Razakarivony等[19],2016),UAV123(Mueller等[5],2016),COWC(Mundhenk等[20],2016),DOTA(Xia等[21],2018),iSAID(Zamir等[22],2019),RarePlanes(Shermeyer等[23],2020),TinyPerson(Yu等[24],2020),UAVid(Lyu等[25],2020)
	多任务数据集	Stanford Drone Dataset(Robicquet等[26],2016),Okutama-Action(Barekatain等[27],2017), VisDrone(Zhu等[7],2018),UAVDT(Du等[28],2018),DroneCrowd(Wen等[29],2021), Anti-UAV(Jiang等[30],2021),UAV-Human(Li等[31],2021)
模态类型	单源数据集	UAV123(Mueller等[5], 2016),CARPK数据集(Hsieh等[32], 2017),UAVDT数据集(Du等[28], 2018),ASL-TID(Portmann等[33], 2014),Salient Map(Li等[34], 2020),BIRDSAI数据集(Bondi等[35], 2020),HIT-UAV数据集(Suo等[36], 2023)
	多源数据集	AU-AIR数据集(Bozcan等[16], 2020),DroneVehicle数据集(Sun等[37], 2022),DroneRGBT数据集(Peng等[38], 2020),UAV-Human数据集(Li等[31], 2021),RTDOD数据集(Feng等[39], 2023),SkyScenes数据集(Khose等[40], 2024),SynDrone数据集(Lenhard等[41], 2024), FIReStereo数据集(Dhrafani等[42], 2024),UEMM-Air数据集(Yao等[43], 2024),HiAI数据集(Xiao等[44],2025)
环境特性	复杂场景航拍数据集	UAVDT(Du等[28],2018),HazyDet(Feng等[45],2024),UAV-AWID(Munir等[46],2024)
	复杂场景无人机数据集	NPS-Drones(Li等[47],2016),ARD-MAV(Guo等[48],2024)
应用需求	无人机导航具身数据集	AVDN(Fan等[49],2022),AerialVLN(Liu等[50],2023),UrbanBIS(Yang等[51],2023),ARIO(Wang等[52],2024),UAV-VisLoc(Xu等[53],2024),OpenUAV(Wang等[54],2024),AeroVerse(Yao等[55],2024),CityNav(Lee等[56],2024),VLA-3D(Zhang等[57],2024), UAV-VLPA-nano-30(Sautenkov等[58],2025),OpenBench(Wang等[59],2025)

多源融合)上的差异,还反映了技术发展中数据驱动的核心作用。下文将详细介绍低空视觉感知数据集的分类方式。

1.1 按设备类型分类

单机数据集:单机数据集通过单一智能体独立采集数据,专注于特定场景和任务,提供单一视角、单源或有限多源的感知信息,支持目标检测、跟踪、定位及路径规划等算法的开发与验证。图2展示了单机数据集经典的组成实例。根据数据采集场景和任务目标的不同,单机数据集可以大致分为两类:一类是通用场景数据集,例如VisDrone[7]、斯坦福无人机数据集(Stanford drone dataset, SDD)[26]等,这些数据集覆盖了多种常见场景和任务,适用于通用算法的开发与验证;另一类是特定任务数据集,例如UAV123[5]、AnimalDrone[8]、DroneSwarms[9]等,这些数据集针对特定任务设计,具有较高的专业性和针对性。

图2 单机数据集样例

Fig.2 Examples of single-drone datasets

多机协同感知数据集(多机器人或多传感器跨视角组成):多机协同感知数据集通过多台机器或传感器协同采集数据,涵盖丰富的场景和多样的任务,提供全方面的感知信息,支持目标跟踪、检测、路径规划等算法的开发与验证,是多机协同感知研究的重要数据资源。现有的典型数据集可以大致分为仿真类型,如CoPerception-UAVs[13]、Air-Co-Pred[14];真实类型,如MDOT[12]、MDMT[15]。图3展示了典型的多机数据集样例。

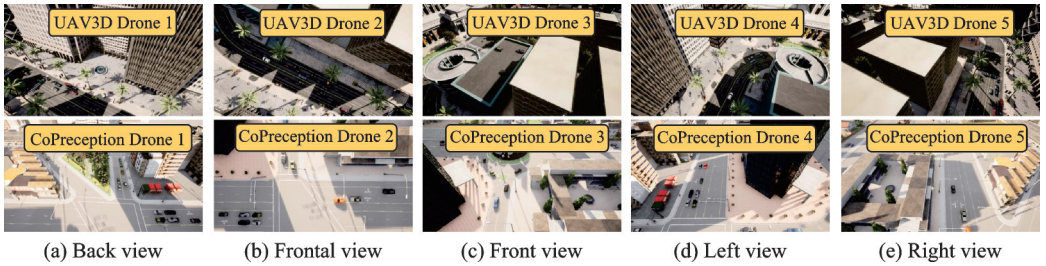


图3 多机数据集样例

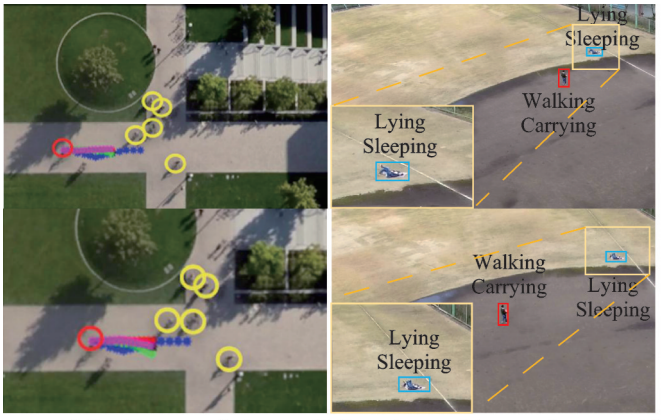
Fig.3 Examples of multi-drone datasets

低空单机数据集在无人机、航空监测等领域具有较高适用性,尤其是在单一设备的任务执行和特定环境下的数据采集与分析中表现出较高的效率。然而,其局限性也较为明显。视角单一性限制了感知能力与范围,难以对环境进行全方位多角度的感知。相比之下,多机协同感知数据集通过多节点动态协作,能够有效弥补低空单机数据集的固有缺陷。此外,多机系统的弹性架构支持动态资源调配,数据集的可迁移性也因多设备、多场景的协同标注得到增强。

1.2 按任务需求分类

单一任务数据集:低空视觉领域的单一任务数据集专注于单个视觉感知任务,如目标检测、语义分割等,且包含明确的标注与目标信息,如目标的边界框、分割掩码等。部分典型单一任务数据集如下:VEDAI数据集[19]采集自美国犹他州自动地理参考中心(Automated Geographic Reference Center, AGRC)的航空图像,适用于航空图像中车辆检测的任务;COWC数据集[20]由劳伦斯-利弗莫尔国家实验室(Lawrence Livermore National Laboratory, LLNL)构建并维护,支持包含车辆标注信息的遥感图像数据的处理与分析。

多任务数据集:多任务数据集同时支持多个视觉感知任务,如目标检测与跟踪、行为分析与目标检测、计数与分割等。多任务数据集的设计更多考虑任务间的关联性,标注设计围绕到多个任务,需要包含多个信息维度,如目标边界框、时序标注、身份标识号码(Identity document, ID)等。如SDD数据集[26]、VisDrone数据集[7]等。图4汇总了常见的多任务数据集样例。



(a) Tracking + prediction (Standford Drone) (b) Detection, tracking + action recognition (OkutamaAction)

图4 多任务数据集样例

Fig.4 Examples of multi-task datasets

1.3 按模态类型分类

单源(单一模态)数据集:无人机感知技术发展推动数据集构建呈现光谱特性差异化特征,可见光数据集已在密集场景解析及群体行为分析等方向形成体系化基准。相较之下,低空红外感知领域针对复杂光照缺失、目标热辐射特征弱化等挑战,逐步构建起面向低信噪比场景的基准体系。Li等[34]通过2 975帧昼夜双模态数据揭示大气衰减导致的低对比度目标辨识难题;Suo团队[36]提出的HIT-UAV数据集,为大倾角巡航下的异源热目标检测提供验证框架。这些数据集通过标注模态创新与场景耦合强化,共同推进了热物理特征驱动的小目标检测与动态干扰抑制技术发展。

多源数据集:在无人机多源数据领域,随着研究的深入与应用需求的增长,一系列具有代表性的数据集不断涌现。2020年,Peng等[38]提出DroneRGBT数据集,通过高对齐精度的3 600对图像推动可见光-热红外(RGB-thermal, RGB-T)人群计数研究。2022年,Sun等[37]提出DroneVehicle数据集,包含28 439对可见光-红外(RGB-Infrared)图像。SynDrone数据集[41]提供72 000帧RGB、深度与激光雷达(Light laser detection and ranging, LiDAR)数据,可以支持多源分割任务。2025年,Xiao等[44]提出了一款包含150组时空对齐的可见光-红外双模态视频序列的无人机高空多源目标跟踪数据集。

1.4 按环境特性分类

图像数据采集时所处的环境条件与人为采集的因素统称为图像数据集的环境特性,主要包括光

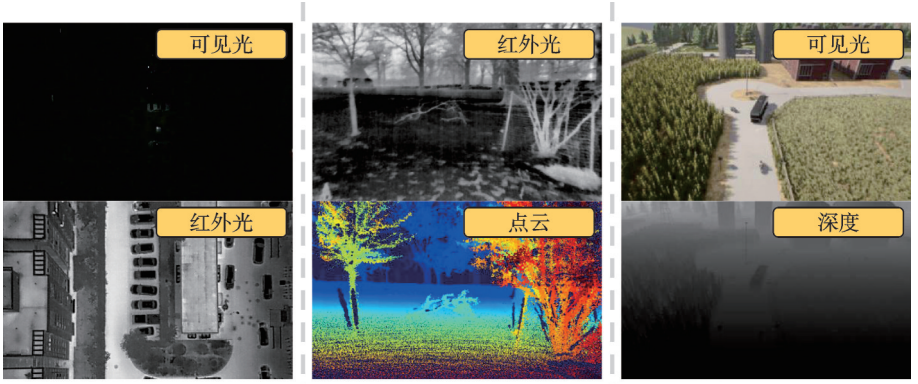


图5 按模态分类的数据集样例

Fig.5 Examples of datasets classified by modalities

照、背景、天气、设备和距离等。这些特性直接影响图像的质量，进而影响图像场景典型性与数据集多样性，最终影响模型训练偏向。其中，恶劣天气和拍摄设置是影响模型性能和任务难度的主要因素。图6汇总了不同环境特性下的数据集样例。

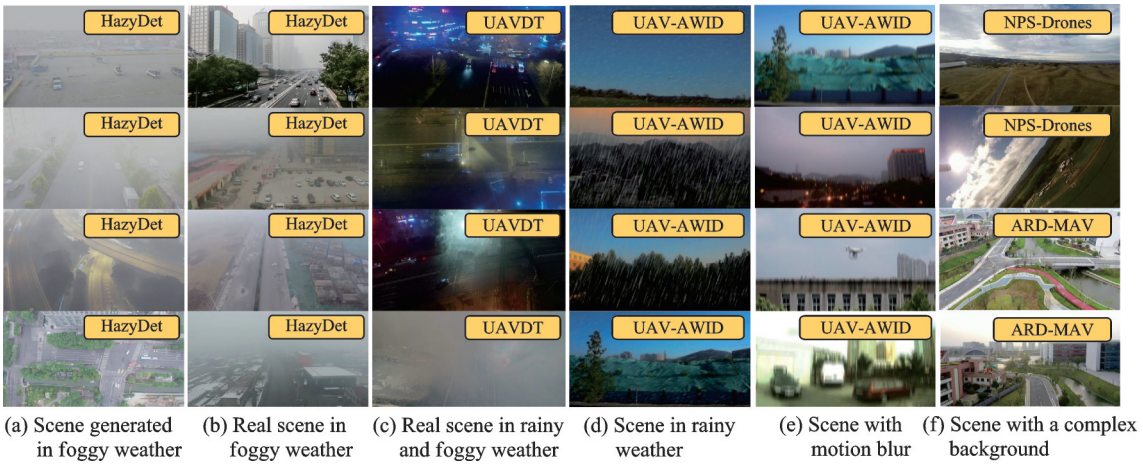


图6 复杂环境特性数据集样例

Fig.6 Samples of complex environment characteristics dataset

恶劣天气对低空视觉感知任务的影响是多方面的，会显著提高所采集数据集的复杂度以及与常规数据集的差异性，进而削弱模型的特征捕捉能力，给感知算法性能造成严重负面影响[60]。在低能见度天气，光的传播过程受到了严重的影响，导致目标物体的边缘和细节难以分辨，HazyDet数据集[45]和UAVDT数据集[28]针对这一问题提供了丰富的数据量。

拍摄设置对低空视觉感知任务的影响也同样至关重要，尤其是在反无人机任务中。对空中目标进行检测时，由于检测物体本身偏小，所以细微的图像环境变化会带来较为严重的性能影响，需要保证模型从困难的环境中学习到稳定的信息。抖动是常见问题之一，会导致目标物体的边缘和细节信息丢失。拍摄设备未能正确对焦会导致失焦模糊，进而导致目标物体模糊不清。针对这些问题，NPS-Drones数据集[47]和ARD-MAV数据集[48]提供了丰富的数据量。

1.5 按应用需求分类

无人机数据集按应用需求可分为无人机视觉感知数据集与无人机具身智能数据集两大类,二者在技术特征、任务需求及数据集构建维度呈现显著差异。传统视觉感知数据集如前文所述,通常集中在对图像的局部特征提取和分析上,而具身智能不仅需要处理视觉数据的局部特征,还需对环境进行全局理解和语义认知,以构建环境的三维模型和语义地图,进而根据视觉感知结果来自主规划路径、避开障碍物,并实时调整飞行姿态和速度。如图7所示,2024年Lee等[56]发布的CityNav数据集通过集成3D城市扫描数据与自然语言指令,实现地理语义与运动轨迹的耦合建模;2024年Wang等[54]提出的OpenUAV则通过六自由度轨迹同步记录,捕捉无人机动力学特性与环境的交互过程。

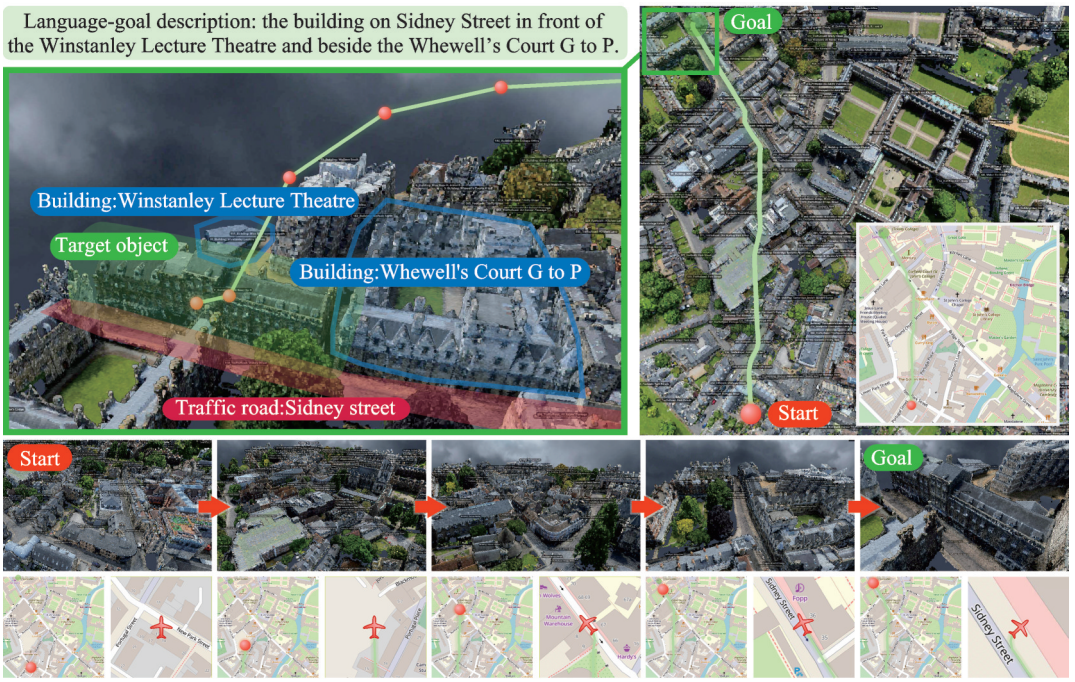


图7 CityNav数据集用于语言指引目标导航[56]

Fig.7 CityNav dataset designed for language-guided goal-oriented navigation[56]

在应用任务复杂度方面,相较于传统无人机视觉感知,无人机具身导航面临着更复杂的应用场景和任务需求,如在搜索救援、电力巡检等领域,无人机需要在复杂的环境中自主完成导航、目标识别、任务规划和执行等一系列复杂任务。2023年Rahnemoonfar等[61]提出的RescueNet数据集通过构建高分辨率灾后图像与像素级语义标注,支持无人机实现了对自然灾害场景的精细化语义理解与损伤评估;2024年Yao等[64]提出的AeroVerse数据集通过构建层次化语义地图(SkyAgent-Scene3k)与多目标规划任务(SkyAgent-Plan3k),支持无人机在物流配送中的路径优化与资源调度;AerialVLN数据集[50]采用动态环境仿真技术,模拟雨雪、风速等扰动因素,提升无人机在低空电力巡检任务中的抗干扰能力。

2 典型低空视觉感知数据集

本文针对不同的分类方式,选取各个类别典型的低空视觉数据集进行介绍,并将各数据集的特点汇总如表2所示。

表2 典型低空视觉数据集(k=1 000)

Table 2 Typical low-altitude visual datasets(k=1 000)

数据集	发布年份	数据数量	类别数量	标注规模	特点(场景类型、支撑任务等)
VisDrone	2018	2×10^7	10	超2×10^7	覆盖中国14个城市昼夜、晴天、阴天、雨天等场景,支持目标检测、跟踪、计数任务。
AnimalDrone	2021	53 600	1	超4×10^6	包含森林、草原、农场等场景及不同天气条件下的动物影像,适用动物计数与监测等任务。
DroneSwarms	2024	9 100	1	242 200	包含城市、山区、天空环境及不同时间、天气条件、姿态变化,适于微小目标检测。
DroneVehicle	2020	56 878	5	953 087	覆盖城市道路、居民区、停车场等场景,包含RGB-T模态,支持目标检测、语义分割任务。
UAV123	2015	112 578	28	112 578	附带小目标、长视频和拍摄角度等属性标注,支持目标跟踪、检测与定位等任务。
DOTA	2018	11 268	18	188 282	数据包含来自不同传感器的日间航拍图像,支持目标检测、实例分割任务。
Oxford RobotCar	2017	2×10^6	—	—	包含雷达、激光雷达、相机和GPS等数据,覆盖晴雨雪日夜四季环境,支持自动驾驶、定位与地图构建、环境感知等。
DroneDeploy	2019	5 000	30	—	覆盖农业、建筑、能源等多个行业的航拍场景,通过对原始图像的校正和拼接,生成高精度的地理参考图像和三维模型;适用于语义分割、地形测绘和环境监测等任务。
VVD	2025	6 831	8	361 489	覆盖晴阴雾、傍晚夜间70个场景,支持车辆检测、分类、交通统计和违规监控等任务。
MDOT	2021	259 800	9	超259 000	覆盖多种城市开放场景,包含昼夜,遮挡等多特征,支持多目标跟踪任务。
UAV3D	2025	500 000	17	3.3×10^6	覆盖城市复杂场景,采用仿真数据标注,支持目标检测与跟踪任务。
MDMT	2023	39 600	3	超2.2×10^6	覆盖真实开放城市场景,包含遮挡情况、昼夜环境条件,支持目标检测、多目标跟踪任务。
AU-AIR	2020	32 800	8	超1×10^6	覆盖城市场景,包含8类目标与位置姿态信息,支持目标检测、目标跟踪和规划控制任务。
MAVREC	2024	537 000	10	超1×10^6	覆盖欧洲街道场景,包含无人机、相机立体3D视角,规模庞大,支持目标检测任务。
CoPerception-UAVs	2022	131 900	—	1.94×10^6	城市街道仿真数据集,涉及前后左右下5个视角,支持3D目标检测、实例分割任务。
Air-Co-Pred	2024	32 000	3	超430 000	精简城市街道仿真数据集,高精度注释,支持3D目标检测任务。
Stanford Drone Dataset	2016	超929 000帧	6	19 564+185 000+40 000	覆盖100个真实大学校园户外场景,支持多目标跟踪及轨迹预测任务。
UAVDT	2018	80 000	3	841 500+超100 000	覆盖城市道路、交通枢纽及停车场等复杂车辆交通场景,适用于目标检测和目标跟踪任务。

续表

数据集	发布年份	数据数量	类别数量	标注规模	特点(场景类型、支撑任务等)
DroneCrowd	2021	33 600	1	超4 800 000	包含多种城市、室外场景及拥挤场所的高分辨率视频,适于人群密度估计、行为分析等任务。
CARPK	2017	1 448	1	89 777	覆盖4个无人机视角停车场场景,支持目标计数与定位任务。
Salient Map	2020	2 975	2	8 624	覆盖城市交通监控与紧急管理的昼夜时间场景,支持目标检测与显著性分割任务。
HIT-UAV	2023	2 898	5	24 899	覆盖学校、停车场、道路和操场的昼夜时间场景,支持目标检测与显著性分割任务。
BIRDSAI	2020	62 000+100 000帧	9	—	覆盖非洲保护区(稀树草原、水源、河流)夜间场景,支持目标检测、跟踪及跨域适应任务。
ASL-TID	2014	4 381	26	—	包含城乡无人机俯视/倾斜视角昼夜全天候热成像数据,适用于人员检测与跟踪。
DroneRGBT	2020	3 600	1	175 698	覆盖校园、街道等多种场景,涵盖不同光照条件,支持RGB-T人群计数任务。
AU-AIR	2020	32 823	8	132 034	覆盖道路交叉口多种天气光照条件场景,支持无人机低空交通场景多源目标检测任务。
UEMM-Air	2024	120 000对	13	—	包含城乡无人机俯视/倾斜视角的全天候多源数据,支持检测、分割、跨模态检索等任务。
SkyScenes	2024	33 600	28	—	覆盖城市和乡村场景,包含视觉、深度信息,支持空中场景理解、语义分割等任务。
SynDrone	2023	72 000帧	28	72 000	覆盖城乡多视角场景,RGB、点云等多源数据,支持检测、分割、3D重建及多源学习任务。
FIReStereo	2024	204 594对	3	35 000	覆盖城市、森林、昼夜雨雾等场景多源数据,支持深度估计、即时定位与地图构建(Simultaneous localization and mapping, SLAM)及多源学习任务。
UAV-Human	2021	67 428个视频	155	—	覆盖城乡室内外多种天气多源数据,支持动作识别、姿态估计、行人重识别、属性识别任务。
RTDOD	2023	16 192对	10	16 200	覆盖城乡晴雾夜间条件场景,包含RGB及热成像,支持域增量目标检测任务。
HiAI	2025	150对视频	9	超12 000	覆盖城市、郊区、湖边昼夜雾等场景,支持多源目标跟踪、微小目标、复杂场景任务。
VEDAI	2016	1 210	9	3 640	数据来自美国犹他州AGRC的航空图像,适用于航空影像车辆检测任务。
COWC	2016	53	1	32 716+58 247	数据来自加拿大、美国等地的航空图像,适用于航空影像车辆检测任务。
RarePlanes	2021	>50 000	1	630 000	覆盖了城区、郊外、海边等区域,由卫星拍摄或合成,适用于飞机及其属性检测任务。
TinyPerson	2019	2 369	1	72 651	包含海边行人密集的场景,适用于微小对象的检测任务。

续表

数据集	发布年份	数据数量	类别数量	标注规模	特点(场景类型、支撑任务等)
iSAID	2020	2 806	15	655 451	覆盖城市遥感船舶、储罐、棒球场、网球场等目标,支持城市场景实例分割任务。
UAVid	2020	300	8	—	主要覆盖复杂城市街景环境,适用于城市场景的语义分割任务。
Okutama-Action	2017	43 min视频	12	—	来自校园内场景,适用于目标跟踪与人体动作识别等任务。
Anti-UAV	2021	318视频对	1	—	覆盖建筑、云雾、树林、昼夜等场景与RGB-T双模态,适用于反无人机检测与跟踪任务。
UAVid	2020	300	8	—	主要覆盖复杂城市街景环境,适用于城市场景的语义分割任务。
HazyDet	2024	11 600	3	383 000	覆盖人工生成与真实场景下的雾天环境,可用于复杂场景下的车辆检测。
UAV-AWID	2024	24 876	2	22 581	包含人工生成模糊与雨天数据,可用于无人机检测。
NPS-Drones	2016	70 250	1	超50 000	包含云层、矮房与天空场景下,可用于无人机多机小目标检测任务。
ARD-MAV	2024	106 665	1	107 317	包含复杂背景、目标遮挡、运动模糊、小目标等挑战场景,可用于无人机单机检测任务。
OpenBench	2025	—	—	—	包含小、中、大仿真环境与动态地图更新,集成动态语义元素,可用视觉语言模型(Vision language models, VLMs)添加语义信息。
UAV-VLPA-nano-30	2025	30张卫星影像	—	—	覆盖城区、郊区、自然景观等地理场景,包含飞行路径、动作指令序列与地理元数据。
VLA-3D	2025	11 500	477	—	包含多源数据标注,构建了结构化场景图,支持语言-场景对齐、错误检测等任务。
CityNav	2024	—	—	32 637轨迹+5 850对象	包含真实城市三维扫描数据,数据均配有对应自然语言描述。
AerialAgent-Ego10k	2024	10 000	12	—	包含无人机第一人称视角1 920像素×1 080像素分辨率城市图像和对应细粒度文本描述。
CyberAgent-Ego500k	2024	500 000	—	—	由UE(Unreal engine)引擎生成,包含RGB图像、深度图、位姿矩阵,可模拟多种环境光照条件、颗粒特效。
SkyAgent-Scene3k	2024	3 000	—	—	三维场景理解数据集,支持不同时间、天气与多样场景,支持添加车辆、行人等动态元素。
SkyAgent-Reason3k	2024	3 000对	6	—	覆盖几何推理、拓扑推断、功能预测、物理仿真、时序分析和多源融合6类典型认知任务。
SkyAgent-Nav3k	2024	1 258+657对	—	—	包含RGB、深度图及姿态信息和自然语言描述,支持模拟行人、车辆流动等动态干扰。
SkyAgent-Plan3k	2024	3 000个任务	—	—	任务规划数据集,构建城市商业区、住宅区、工业园等异构城市场景的语义地图,包含临时封闭路段或天气影响等动态影响。

续表

数据集	发布年份	数据数量	类别数量	标注规模	特点(场景类型、支撑任务等)
SkyAgent-Act3k	2024	1 200 h	—	—	连续决策数据集采集,包含1 200 h无人机飞行控制序列,构建马尔可夫决策过程模型。
OpenUAV	2024	12 149条轨迹	89	—	包含城市、乡村、自然景观等22种不同场景,涵盖运动参数、惯性测量单位(Inertial measurement unit, IMU)、RGB-D视觉、点云等信息,支持光照、天气、植被波动等环境效应。
UAV-VisLoc	2024	6 742+11	—	—	覆盖11个气候带特征区域,多种地形特征,记录纬度、经度、相对高程、航向角等元数据。
ARIO	2024	约3×10^6帧	—	—	覆盖258个机器人321 064项任务,统一格式兼容数据,采用分层时间戳架构实现数据同步。

注:表中"数据数量"处除有单位和数据类型说明的,其余均为图像数据。

2.1 单机数据集

2.1.1 定义与特点

单机数据集是指由单一低空飞行设备(如无人机、航拍相机等)独立采集的视觉感知数据集,其核心目标是通过单设备的多源或多视角数据捕捉低空场景的动态信息。单机数据集帧数主要集中在10 000~100 000之间,类型数量基本在10以内,说明目前的单机数据集大多仍规模小、专用性强,缺少通用的大规模数据集。总体来说,单机数据集具有以下特点:

(1) 高分辨率与动态性:单机低空设备一般都配有能够捕捉高分辨率图像或视频的高精度摄像机或传感器,因此数据集合能够适应动态物体探测和跟踪在低空复杂环境中的需求[62]。

(2) 灵活性与场景多样性:单机低空数据集可覆盖城市、农田、森林等各种地形地貌,同时涵盖灯光变化、天气干扰、遮挡挑战等不同现实世界场景。

(3) 标注一致性:数据标注通常是从单个设备的角度出发,其内容涵盖目标检测框、语义分割掩码和轨迹追踪等类型,并对细粒度与任务要求的高度适配进行标记[63]。

2.1.2 典型单机数据集

(1) VisDrone是由天津大学于2018年创建的无人机视觉数据集[7]。该数据集为无人机视觉研究提供了一个大规模、高质量的基准数据集,有效推动了计算机视觉研究与无人机技术的紧密结合。VisDrone数据集包含2 000多万张图像及视频序列,拥有2 000万个标注目标,覆盖了行人、车辆和非机动车等多类目标。该数据集的特点是多场景、多天气、不同密度,支持目标检测和跟踪及计数等任务。如表3所示,VisDrone数据集面向目标检测任务,选取$AP_{0.5}$、AP_{vt}、AP_t和AP_s四种平均精确度指标,将基于归一化高斯Wasserstein(Normalized Gaussian Wasserstein, NWD)的检测模型(表3中带*模型)与其Baseline模型进行对比,以验证目标检测的精准度。NWD模型达到了$AP_{0.5}$ 35.0+、AP_{vt} 2.9+、AP_t 10.0+和AP_s 21.0+的精度指标。

表3 VisDrone数据集上Baseline模型与NWD模型(带*)的比较结果

Table 3 Comparison of the Baseline models with the NWD models (with *) on VisDrone dataset

对比方法	Faster R-CNN	Faster R-CNN*	Cascade R-CNN	Cascade R-CNN*
$AP_{0.5}$	38.0	**38.5**	38.5	**40.3**
AP_{vt}	0.1	**3.8**	0.5	**2.9**
AP_t	6.2	**10.2**	6.8	**11.1**
AP_s	20.0	**21.4**	21.4	**22.2**

(2) AnimalDrone是由天津大学朱鹏飞团队[8]于2021年创建的无人机视觉数据集,旨在为动物计数与监测任务提供高质量的基准数据。该数据集由两个子集组成,无人机实地采集得到的AnimalDrone-PartA和从互联网上收集的AnimalDrone-PartB,总计53 644帧,具有超过400万个目标。如表4所示,对比CSRNet、CFF、GFAN等计数方法,使用平均绝对误差(Mean absolute error, MAE)和均方误差(Mean squared error, MSE)作为指标,对计数算法的精准度进行评价,在AnimalDrone数据集PartA子集上,GFAN方法指标达到MAE 6.9和MSE 8.8。AnimalDrone数据集提供了丰富的动物种类、多样化的自然环境背景以及复杂的天气条件,同时涵盖了动物姿态变化、群体行为、遮挡和远距离拍摄等挑战性场景,为最新的监测和计数算法优化提供了大量的数据支持。

表4 最新计数算法在AnimalDrone-PartA数据集上的定量指标

Table 4 Quantitative metrics of the latest counting algorithm on AnimalDrone-PartA dataset

对比方法	Overall		Low-density		High-density		Low-height		Med-height		High-height		Bird-view	
	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
CSRNet	14.7	17.8	15.3	17.8	14.2	16.7	25.2	25.5	11.8	14.1	14.1	16.8	15.8	17.6
Switch-CNN	18.6	22.7	16.7	20.1	20.8	30.1	16.7	21.4	17.5	19.7	20.7	26.8	17.8	21.2
C-MTL	48.4	56.4	41.5	49.6	56.5	63.4	50.4	53.9	45.8	52.3	50.7	61.8	51.5	56.4
DA-Net	42.2	54.4	36.9	46.4	48.5	62.4	74.6	75.2	46.4	62.2	24.4	27.9	38.3	51.5
CFF	10.1	13.2	**7.3**	**9.5**	13.2	16.3	**8.3**	**9.9**	9.2	12.9	12.6	15.0	10.6	13.8
GFAN	**6.9**	**8.8**	8.1	10.4	5.7	**7.1**	11.7	13.4	4.5	5.7	7.5	**9.3**	7.4	**9.5**
GFAN-w/o-warp	7.5	9.7	8.2	11.0	6.6	8.4	13.1	15.2	4.3	5.8	8.6	10.2	**7.3**	10.3
GFAN-w/o-cnt	8.3	10.0	9.4	11.3	8.2	9.6	12.4	16.0	5.6	6.7	8.4	10.3	8.1	10.0

(3) DroneSwarms是由天津大学朱鹏飞团队[9]于2024年创建的目前平均尺寸最小的反无人机物体检测数据集,由9 109张图像和242 218个带注释的无人机实例组成,平均每张图像包含26.59个无人机实例,手动标记精度高。DroneSwarms包含各种户外环境和基于不同时间和天气条件的不同照明条件。与现有的反无人机数据集不同,DroneSwarms包含241 249个大小为32像素或以下的微小目标,约占99.60%,平均大小仅约为7.9像素。

(4) UAV123是由德国达姆施塔特技术大学的研究团队于2015年创建的无人机目标跟踪数据集[5]。该数据集旨在从无人机的角度为目标跟踪研究提供高质量的基准数据集,从而促进无人机视频跟踪技术的发展。UAV123数据集包含123个视频序列,总计超过110 000帧图像数据,使其成为仅次于ALOV300++的第二大目标跟踪数据集。

(5) Oxford RobotCar是由英国牛津大学于2017年创建的大规模自动驾驶数据集[6]。该数据集旨在提供多样化的基准数据集,用于研究自动驾驶和低空飞行器的长期情境感知。牛津ROBOTCAR数据集包含了城市驾驶数据,时间跨度长,空间尺度大,涵盖多种天气和光照条件。该数据集支持长期定位、环境感知、路径规划等任务。

(6) 多种类无人机数据集(Varied drone dataset,VDD)是由安徽大学人工智能实验室于2025年创建的无人机视频车辆检测大规模基准数据集[11]。该数据集包含70个在500 m高度拍摄的无人机视频,涵盖了各种场景和天气条件,共标注了361 489个车辆实例,每个车辆实例都使用定向边界框进行标注,并细分为8个特定的车辆类别。VVD数据集旨在促进无人机视频中车辆检测、分类和相关任务的研究和应用。

2.1.3 *应用场景分析*

单机数据集在低空视觉感知领域的典型应用场景主要有以下3个方面:

(1) 目标检测与跟踪:VisDrone与UAVDT等高质量针对性数据集通过密集标注为车辆与行人检测算法的训练提供了重要支撑,这些标注在交通监控领域具有显著的应用价值。此外,这类数据集在反无人机安防领域也展现出强大的应用潜力。

(2) 语义分割与场景理解:高分辨率航拍单机数据集在农田监测、城市规划等场景中发挥着关键作用。在农业检测领域,研究人员可实现对农田植被覆盖的像素级精确分割,从而为作物健康状况评估提供可靠依据;在城市规划领域,单机数据集为建筑物、道路、绿地等场景元素提供的精细化语义分割,给三维建模和空间分析提供了重要的基础数据支撑。

(3) 视觉自主导航:通过精确的轨迹标注,为动态场景数据集训练无人机路径规划模型提供了重要的支撑。在动态和复杂环境下,无人机需要对周围的障碍物进行实时感知,并对安全路径进行规划,有效提升无人机的环境适应性,这也为早期的具体智能研究提供了重要的数据基础。

2.2 多机协同数据集

2.2.1 定义与特点

多机协同指多个感知设备或系统通过信息共享与协同合作,共同完成对环境或目标的感知任务。多机协同通过多源异构数据的时空配准与特征级融合,显著提升复杂场景中的态势感知精度与决策可靠性。该技术通过克服单机的遮挡限制,利用多机协同的跨视角互补能力,实现精准的感知,在城市立体安防、广域灾害监测等高可靠性需求场景中具有重要的应用价值。本研究对典型多机协同数据集进行了统计。目前,大多数多机协同数据集包含2~3架无人机,大于10架的多机协同数据集尚为欠缺,显示出当前研究主要聚焦在中小规模的协同任务中。多机协同数据集具有以下特点:

(1) 跨视角数据融合:多机协同能够整合来自不同位置的无人机摄像头或无人机与地面摄像头协作所产生的多视角数据。例如在城市安防场景中,通过多架无人机从不同视角跟踪嫌疑目标车辆,可以提供比单一无人机视角更为全面的全局信息,进而显著提升嫌疑目标检测与跟踪的准确性。

(2) 多源数据互补:多机协同不仅涵盖视觉数据,还包含激光雷达、毫米波雷达等多源数据。这些异构数据的融合,有效增强了系统的感知鲁棒性和准确性。在多机器人协同感知场景中,视觉传感器与IMU、激光雷达与IMU以及视觉传感器和激光雷达与IMU的组合被广泛运用,有力地提升了环境感知能力。

2.2.2 典型多机协同数据集

(1) MDOT是天津大学朱鹏飞团队[12]在2021年由多机协同采集的千万级高精度视频数据集,包含双机协同拍摄的92组113 918帧画面和三机协同采集的63组145 875帧影像,覆盖城市开放场景中行人与车辆的立体化多视角跟踪需求。数据集创新标注10种场景属性(昼夜/运动模糊/遮挡程度等),构建多维度评估体系,作为首个多无人机单目标跟踪基准平台,为算法研发提供标准化测试环境与技术验证场景。如表5所示,对比ATOM、Tomp101、Stark和TranMDOT方法在单无人机跟踪(无人机1、无人机2)与多无人机跟踪(整体表现)上的效果,使用准确度作为指标进行评判,并对比了不同方法的运行帧率。MDOT数据集在面向目标跟踪的任务中,Stark和TransMDOT方法准确度达到了60%+。MDOT数据集丰富的场景、多维立体化视角为多目标跟踪提供了优化数据。

表5 在MDOT-Two上的目标跟踪对比结果

Table 5 Comparison results of target tracking on MDOT- Two dataset

方法	准确度/%			运行帧率/(帧·s^{-1})
	无人机1	无人机2	整体表现	
ATOM	63.8	55.2	59.5	64.1
Tomp101	65.4	53.8	59.6	**25.2**
Stark	67.0	57.4	61.7	28.0
TransMDOT	**71.1**	**64.7**	**68.2**	26.6

(2) CoPerception-UAVs是2022年由上海交通大学的Hu等[13]与南加州大学联合提出的首个

大规模无人机群协同感知数据集，通过AirSim[64]与CARLA[65]仿真框架生成，组织方式与nuScenes[66]一致。该无人机群在纪律模式（静态阵列）与动态模式（自由飞行）下的混合采集，数据集包含131 900张多源（RGB/LiDAR）航拍图像及1 940 000个3D物体边界框，覆盖城市开放场景中无人机群的动态协作感知需求。数据集创新性标注了像素级语义分割标签、2D/3D物体检测框及相机内外参矩阵，构建了支持3D目标检测、多目标跟踪、三维重建与语义分割的标准化评测体系。

（3）MDMT数据集是2022年由天津大学朱鹏飞团队[15]发布的面向多无人机多目标跟踪的大规模数据集，旨在解决遮挡情况下的目标跟踪问题。该数据集包含88组视频序列，共39 678帧图像，完成标注2 204 620个目标框，其中近四分之一包含目标遮挡情况。数据采集场景涵盖城市道路、郊区道路、停车场等多类复杂环境，包含昼夜交替及多种天气条件下的动态感知需求。如表6所示，对比使用不同主干网络（Faster-RCNN、Tood、Carafe、Cascade RPN、AutoAssign、MIA-Net）和跟踪方法（QDTrack、ByteTrack）在不同单无人机以及多无人机上的跟踪效果，使用MOTA（Multiple object tracking accuracy）和idF1（Identity F_1-score）进行评判，MDMT数据集在面向多目标跟踪任务上，在MIA-Net等方法中idF1指标达到了60%+。数据集提供像素级目标检测框、跨视角身份关联标注及无人机内外参信息，构建了支持多目标跟踪、跨设备身份匹配与遮挡解析的标准化评测体系，为多无人机协同感知算法研究提供了高多样性、强挑战性的真实场景验证平台。

表6 在MDMT数据集上的对比结果

Table 6 Comparison results on MDMT dataset %

方法	无人机1		无人机2		整体	
	MOTA	idF1	MOTA	idF1	MOTA	idF1
Faster-RCNN+QDTrack	52.12	66.23	43.02	57.68	47.57	61.96
Faster-RCNN+ByteTrack	53.88	67.71	47.98	64.14	50.92	65.93
Tood+Bytetrack	50.95	66.42	48.02	63.92	49.49	65.18
Carafe+QDTrack	53.20	66.28	43.06	57.46	48.13	61.87
Cascade RPN+QDTrack	51.05	65.00	45.75	58.66	48.40	61.82
AutoAssign+ByteTrack	49.52	67.38	44.52	63.75	47.01	65.56
Carafe+Bytetrack	54.13	68.22	**48.42**	64.93	51.38	66.58
MIA-Net (AutoAssign+ByteTrack)	**51.90**	**69.67**	**47.46**	**66.81**	**49.68**	**68.24**
MIA-Net (Carafe+Bytetrack)	**54.92**	**68.82**	**48.23**	**65.12**	**51.58**	**66.97**

（4）AU-AIR是由丹麦奥胡斯大学Bozcan等[16]提出的首个多传感器无人机交通监控数据集，通过低空飞行平台（Parrot Bebop 2）在真实城市场景中采集。该数据集包含8段总计2 h的航拍视频，提取32 823帧带标注的RGB画面及对应传感器数据，覆盖5～30 m高度范围和45°～90°垂直俯视视角的多样化交通场景。数据集标注8类交通对象共计132 034个实例，并同步记录每帧的时间戳、GPS坐标、海拔高度、IMU姿态角及三维速度向量。数据集首次将多传感器数据（视觉/时空/运动/惯性）与对象检测标注相结合，构建面向实时无人机交通监控的基准测试平台，为低空交通监测与多源数据融合提供了标准化数据支撑。

（5）MAVREC是由美国佛罗里达大学、北卡罗来纳大学夏洛特分校和丹麦哥本哈根大学联合团队[17]在2024年提出的首个大规模多视图无人机交通监控数据集。该数据集通过低空无人机与手持地面相机的协同采集，覆盖欧洲农村与城市混合景观的11个地理区域，包含2.5 h超高清（2 700像素×1 520像素）视频序列，总计537 030帧画面及1 102 604个标注实例。首次实现多视图时空同步的立体化数据采集，涵盖10类交通对象及复杂场景属性。

(6) Air-Co-Pred[14]是2024年中国科学院空天信息创新研究院联合上海人工智能实验室等机构,在CARLA[65]仿真平台上构建的首个多无人机协同轨迹预测数据集。该数据集通过4架无人机在50 m高度对100 m×100 m城市道路场景进行协同采集,生成32 000帧同步RGB图像(分辨率1600像素×900像素),创新性地标注了车辆、行人、自行车3类对象的2D/3D标注框及运动轨迹,覆盖长距离(>50 m)、小目标(<0.5 m)、多目标遮挡等复杂场景,为无人机群协同决策和智能交通系统开发提供了高保真数据支撑。

(7) UAV3D是由美国乔治亚州立大学Ye团队[18]在2025年基于CARLA-AirSim联合仿真构建的千万级3D感知基准数据集,包含4类多无人机协同采集的500万张高精度RGB图像和330万个3D标注框。数据集覆盖城市复杂场景中行人与车辆的立体化多视角跟踪需求,通过五机协同编队(交叉形布局,60 m高度)实现多场景泛化,创新性标注10种场景属性,并构建多任务统一评估框架,作为首个支持单/协作3D目标检测与跟踪的无人机基准平台,为算法研发提供高保真测试环境与跨模态验证场景。

2.2.3 应用场景分析

(1) 城市交通监控与管理:AU-AIR、MAVREC和Air-Co-Pred等数据集覆盖真实与仿真场景下的城市场景,提供了丰富的街道交通数据。无人机群体通过协同飞行、多传感器的集成(如RGB摄像头、LiDAR等),可以高效全面地监控城市中的交通情况,进行实时交通状况分析与预测。

(2) 多机协同搜索与救援:MDMT和CoPerception-UAVs等数据集能够有效赋能无人机在复杂环境下的搜索与救援能力。通过多机协同,多个无人机可以覆盖更广泛的区域,协同进行灾难现场的目标跟踪与识别。数据集中的多机视角和遮挡问题标注,使得算法能够在视觉受限或遮挡环境下仍保持高效的目标检测与追踪能力。

(3) 物流配送与仓储管理:利用MDOT和UAV3D数据集中的多机协同数据,无人机群体在物流配送和仓储管理中能够发挥重要作用。在仓储管理方面,多个无人机协同作业,可以在大型仓库中高效地进行物品盘点、分类与搬运,同时避免传统人工操作中的误差和低效问题。在物流配送方面,多机协同能够提供更灵活的运输路径选择和实时配送信息,显著提升配送效率和准确性。

2.3 多任务学习数据集

2.3.1 定义与特点

多任务学习数据集是指针对多种任务需求进行联合标注的数据集合,其一般包含较多数据与多种标注类型,支持模型同时学习多个相关任务。本文对典型多任务数据集进行了调研,当前多任务数据集识别种类大部分为1~10种,达到10种以上类别的数据集较少,总体中小型数据集占比较多;多任务数据集大多为检测任务和追踪任务,分割任务相对较少,包含检测与追踪任务的数据集为主体趋势。多任务学习数据集的核心特点是任务间的关联性和数据的多样性与标注的共享性,具体如下:

(1) 任务关联性:数据集的任务间通常存在关联,并围绕任务内在关联部分对数据集进行标注,为多个任务提供依赖。

(2) 数据多样性:数据集往往覆盖多环境、多尺度、多传感器和动态交互等复杂低空场景,以提高多任务的泛化性。

(3) 标注共享性:数据集包含多种数据与标注适用于多种任务,通过共享数据与标注表示减少冗余计算。

2.3.2 典型多任务学习数据集

(1) SDD[26]是斯坦福大学团队于2016年发布的无人机数据集,主要面向行人轨迹预测与行为交互分析任务。SDD数据集以其高分辨率的视频和丰富的标注信息著称,包含了不同天气、时间和光照条件下多种场景的动态对象。

(2) VisDrone[7]是天津大学朱鹏飞团队发布的大规模无人机低空视觉基准数据集,如前文介绍,该数据集支持包括目标检测、实例分割、多目标跟踪及人群计数等多种任务。VisDrone弥补了无人机视角数据规模不足、任务单一且标注粒度粗糙等问题,成为低空视觉领域的重要基准。

(3) DroneBird[10]是天津大学朱鹏飞团队在2024年提出的首个大规模无人机视角视频鸟类数据集。该数据集包含了3 686 409个鸟类标注点以支持鸟群计数任务,提供了9 389条轨迹标注用于支持目标追踪分析任务,轨迹时长从1帧到500帧不等。DroneBird涵盖了多种自然环境及不同天气条件下的鸟类活动影像。数据集中的标注特别关注了飞行姿态、群体行为及栖息地的高精度标注,为鸟类学研究和环境保护提供了丰富的数据支持,推动了无人机视角下鸟类研究的发展。表7对比了CSRNet、MAN、PET等不同方法在DroneBird数据集的计数任务效果,使用MAE和均方根误差(Root mean square error, RMSE)指标对效果进行评判。结果显示,E-MAC方法在MAE与RMSE上达到了最佳的计数误差指标性能,在鸟类计数任务上相比现有方法大幅提高了指标性能,表明DroneBird数据集能成为评估不同算法在航拍环境中目标计数性能的典型数据基准。

表7 DroneBird数据集上MAE和RMSE指标的定量比较

Table 7 Quantitative comparison of MAE and RMSE metrics on DroneBird dataset

方法	类型	DroneBird	
		MAE↓	RMSE↓
CSRNet	图片	66.11	79.33
MAN	图片	39.11	50.08
PET	图片	45.10	52.35
Gramformer	图片	49.11	65.50
EPF	视频	97.22	133.01
STGN	视频	92.38	124.67
E-MAC	视频	**38.72**	**42.92**

(4) DroneCrowd是由天津大学朱鹏飞团队[29]在2021年提出的首个大规模无人机视角下的拥挤人群视频数据集。该数据集由112个视频片段组成,包含70个不同场景中捕获的33 600个高分辨率帧(即1 920像素×1 080像素)。数据集标注信息丰富,提供了20 800个人物轨迹标注,包含480万个头部注释和多个视频级属性序列。DroneCrowd涵盖了多种城市环境、室外场景以及拥挤的场所,还特别标注了人群密度、个体行为以及互动情况,旨在为人群密度估计、行为分析和紧急情况下的人群控制提供重要的数据支持。如表8所示,对比使用DA-Net、CSRNet等方法在DroneCrowd数据集的人群密度估计任务效果,分别在大规模、小规模、多云、晴天等多种条件下使用MAE和MSE指标进行评判,结果显示,DM-Count、CSRNet等方法估计误差指标达到了MAE14.1+和MSE14.9+,表明DroneCrowd数据集提供了丰富多样的测试条件,可作为人群密度估计任务上的典型数据基准。

表8 DroneCrowd数据集上密度图估计误差

Table 8 Estimation errors of density maps on DroneCrowd dataset

方法	运行帧率/(帧·s^{-1})	整体		大规模		小规模		多云		晴天		夜晚		拥挤人群		稀疏人群	
		MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
StackPooling	0.73	68.8	77.2	68.7	77.1	68.8	77.3	66.5	75.9	74.0	83.4	65.2	67.4	95.7	101.1	53.1	59.1
DA-Net	2.52	36.5	47.3	41.5	54.7	28.9	33.1	45.4	58.6	26.5	31.3	29.5	34.0	56.5	68.3	24.9	28.7
CSRNet	3.92	19.8	25.6	17.8	25.4	22.9	25.8	12.8	16.6	19.1	22.5	42.3	45.8	20.2	24.0	19.6	26.5
CAN	7.12	22.1	33.4	18.9	26.7	26.9	41.5	**11.2**	**14.9**	14.8	17.5	69.4	73.6	**14.4**	**17.9**	26.6	39.7
DM-Count	10.04	18.4	27.0	19.2	29.6	17.2	22.4	11.4	16.3	**12.6**	**15.2**	51.1	55.7	17.6	21.8	18.9	29.6
STNNNet	3.41	**15.8**	**18.7**	**16.0**	**18.4**	**15.6**	**19.2**	14.1	17.2	19.9	22.5	**12.9**	**14.4**	18.5	21.6	**14.3**	**16.9**

(5) Okutama-Action[27]由慕尼黑工业大学等机构于2018年联合发布的数据集,涵盖目标跟踪与人体动作检测任务。该数据集包含43个全标注的视频序列,涵盖12种户外动作类别。每个视频序列的时长约为1 min。数据集的独特之处在于其动态行为转换、显著的尺度变化、视角变化等复杂情况,使其成为行为检测模型性能提升的重要数据资源。

(6) UAV-Human[31]是新加坡科技设计大学与山东大学于2021年联合发布的大规模人类行为理解数据集。该数据集包含67 428个多传感器视频序列,其中不同的视频片段分别用于姿态估计、行人重识别、属性识别等功能,共涉及119个动作识别主体。该数据集旨在推动无人机在复杂环境中的人类行为分析技术的发展,支持动作识别、姿态估计和人物再识别等多种任务。

(7) Anti-UAV[30]是大连理工大学团队在2021年发布的大规模多源无人机跟踪基准数据集,包含目标检测和目标跟踪两个子数据集。该数据集包含318个RGB-T的视频对,视频的场景包含白天和夜晚,红外和可见光以及各种各样的飞行背景。该数据集为无人机目标检测与跟踪任务提供了重要的数据基础。

2.3.3 *数据标注复杂性与应用场景分析*

(1) 数据标注复杂性:多任务学习数据集的标注复杂度显著高于单任务数据集。一方面多任务学习数据集需同时满足不同任务的基础标注类型;另一方面多任务标注需协调不同任务的标注规范。此外,跨场景泛化要求高,如无人机视角下光照、尺度变化大,需标注多样性数据以覆盖复杂环境。

(2) 典型应用场景:①城市安防场景,针对密集人群和交通场景(如VisDrone数据集),需要实现高精度的目标检测与行为异常识别,为城市公共安全提供支持;②交通监控场景,UAVDT数据集可用于车辆轨迹预测和交通流量分析,辅助交通管理和事故预防;③反无人机系统场景,利用Anti-UAV数据集进行跨模态目标检测与实时跟踪,保障关键场所安全,构建反制无人机的监控体系。

2.4 多源融合数据集

2.4.1 *定义与特点*

低空场景多源融合数据集是指在低空飞行环境下,通过多种传感器(如雷达、激光雷达、光学摄像头和红外传感器等)采集的各种数据的组合。数据集融合了不同模态的数据,旨在服务于低空感知、动态监测与智能决策任务的算法训练与验证。图8展示了典型多源学习数据集按照年份、类别数量、数据类型及多源融合类型的分布统计结果。图8(a)说明了当前多源数据集呈现上升趋势,图8(b)说明了目前数据集针对性较强,涵盖的类别分布集中在少数常见类别中,图8(c)表明目前数据还是以图像数据为主,图8(d)展示了低空多源数据集当前不同模态的分布情况。多源融合数据集具有以下特点:

(1) 多传感器融合:低空场景多源融合数据集是通过多种传感器采集的各种数据的综合,主要用于支持低空飞行中的目标检测、航迹追踪和环境感知等任务。

(2) 高精度时空同步:多种类型的传感器数据具有时空同步性,以保证数据的精确对齐,避免因时间错位影响数据融合效果。

(3) 复杂环境与动态因素:数据集一般都涵盖多种场景和动态物体,以提升系统的实时响应能力。同时多源数据往往会受到不同传感器噪声与误差的影响,所以多源数据一般需要对数据进行处理和校正来确保数据集的准确性和有效性。

2.4.2 *典型多源融合数据集*

(1) DroneVehicle是由天津大学朱鹏飞与孙一铭等[37]构建的面向无人机平台的大规模RGB-红外多源车辆检测数据集。该数据集包含28 439对高质量配准的RGB和红外图像、953 087个定向边界框,涵盖城市道路、居民区等多种场景。表9对比了多种目标检测方法在DroneVehicle数据集中的检测性能,其UA-CMDet框架通过融合RGB和红外多源数据信息,平均精度均值(mean Average precision, mAP)较传统检测器(如仅使用单模态的RetinaNet/Faster R-CNN)提升4.2%~6.7%,在数据集的各个

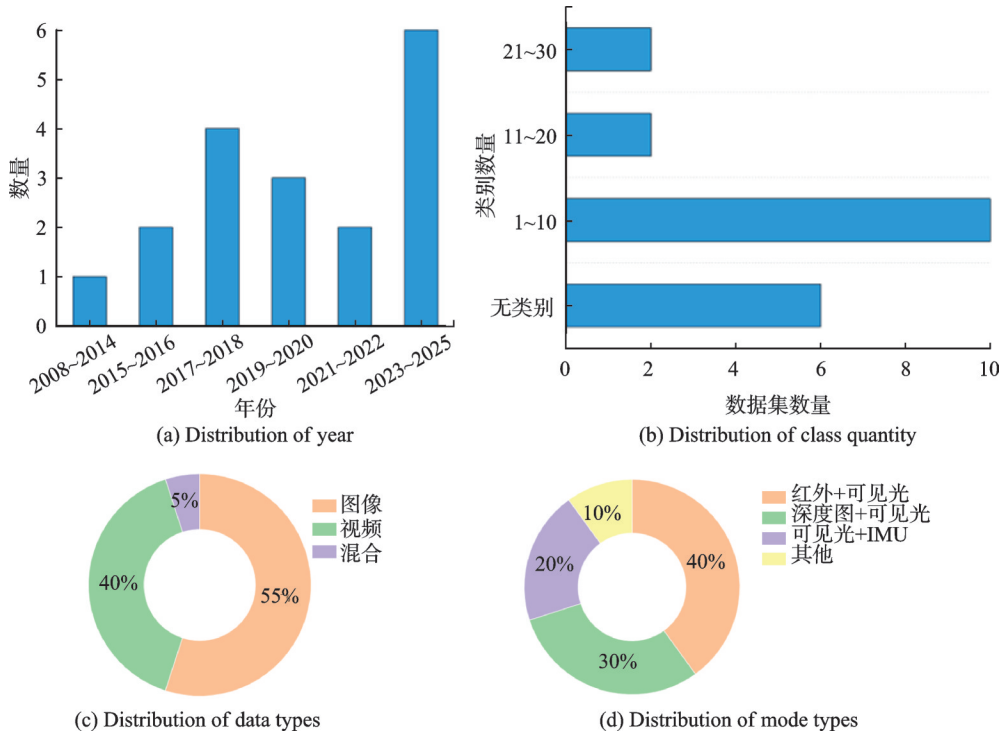


图8 典型多源数据集分布

Fig.8 Distribution of typical multi-modal dataset

表9 DroneVehicle数据集上不同检测方法的检测精度比较

Table 9 Comparison of detection precision by different detection methods on DroneVehicle dataset %

对比方法	模态	汽车	货车	卡车	巴士	小型货车	mAP
RetinaNet(OBB)	RGB	67.50	13.72	28.24	62.05	19.26	38.16
Faster R-CNN(OBB)	RGB	67.88	26.31	38.59	66.98	23.20	44.59
Faster R-CNN(Dpool)	RGB	68.23	26.4	38.73	69.08	26.38	45.76
RolTransformer	RGB	68.13	29.08	44.17	70.55	27.64	47.91
RetinaNet(OBB)	Infrared	79.86	28.05	32.84	67.32	16.44	44.90
Faster R-CNN(OBB)	Infrared	88.63	35.16	42.51	77.92	28.52	54.55
Faster R-CNN(Dpool)	Infrared	**88.94**	36.79	47.91	78.28	32.79	56.91
RolTransformer	Infrared	88.85	41.49	51.53	79.48	34.39	59.15
UA-CMDet(Ours)	RGB+Infrared	87.51	**46.80**	**60.70**	**87.08**	**37.95**	**64.01**

子类中的识别精度也有明显提升(如货车、卡车等)。DroneVehicle作为首个无人机RGB-T车辆检测数据集为多模态融合算法研究提供了标准化基准,推动了无人机感知从“单模态受限”向“全时段可靠”的技术跨越。

(2) DroneRGBT[38]是天津大学朱鹏飞团队发布的基于无人机平台的RGB-T人群计数数据集,旨在解决复杂场景下传统单模态数据难以应对的挑战。该数据集包含3 600对高质量配准的RGB和红外图像、175 698个标注实例,涵盖校园、街道、公园等多种场景。数据集在设计和采集时结合实际应用需求,涵盖了3种光照条件(暗光、黄昏、亮光)、3种高度范围以及不同人群密度。DroneRGBT数据集作为

首个无人机RGB-T人群计数数据集,单张图的平均目标标注量达到48.8个,为探索无人机应急响应、公共安全等需全天候精准人群感知的任务提供关键技术支撑。

(3) RTDOD是由Feng等[39]于2023年提出的面向无人机多源域增量目标检测的大规模RGB-热成像融合数据集。该数据集旨在解决复杂真实场景(如雾天、雨天、夜间等极端天气与光照条件)下目标检测的鲁棒性问题,通过同步采集的RGB与热成像视频对,覆盖晴、雾、傍晚和夜间4种动态变化的域场景。数据集共包含约16 200对图像,涵盖10类常见目标,总计179 672个标注实例。

(4) SynDrone[41]是2023年帕多瓦大学信息工程系研究团队发布的一个大规模多源无人机数据集,通过合成技术模拟了无人机在不同飞行高度和视角下采集的复杂城市场景,涵盖了高分辨率RGB图像、深度图和LiDAR数据,支持深度感知、语义分割及3D重建等多种任务。该数据集包含28个类别的语义标注,提供了72 000个标注样本,同时提供实例级别标注和3D边界框,便于分析动态与静态对象。

(5) SkyScenes是由Khose等[40]于2024年提出的面向无人机视角的合成航空场景理解数据集。该数据集基于CARLA模拟器构建,聚焦于航空场景感知任务,旨在通过高可控性的数据生成机制解决真实航空数据标注成本高、条件多样性不足的问题。数据集共包含33 600张图像,覆盖4种城市与4种乡村场景,5种天气与光照条件,3种飞行高度和4种相机俯仰角,同时提供28类像素级语义分割、实例分割及深度标注。

(6) UAV-Human[31]是由新加坡科技设计大学与山东大学于2021年联合发布的一个面向无人机环境下人类行为理解的大规模多传感器数据集。该数据集通过搭载多种传感器,在多个城乡地区连续3个月采集数据,涵盖了白天与夜间的多样化场景,数据模态包括RGB、深度图、红外、鱼眼视频、夜视视频及骨架序列,数据集内容信息在前文已详细介绍。

(7) UEMM-Air是由Yao等[43]于2025年提出的面向无人机多源感知的合成数据集。该数据集通过虚幻引擎模拟多样化无人机飞行场景,利用自动化采集系统同步获取多源数据,覆盖城市、高速公路、桥梁等复杂环境及不同高度、视角与光照条件。数据集共包含12万对图像,涵盖13类常见目标,包含6种严格对齐的模态数据。

(8) FIReStereo是由Dhrafani等[42]于2024年发布的无人机系统在视觉退化环境下深度感知的红外立体数据集。该数据集聚焦于森林与城市交界区域的复杂场景,包含204 594组同步立体热成像图像,覆盖4类环境和16条轨迹,涵盖多种退化条件,并提供基于LiDAR-SLAM生成的密集深度图(35 706张标注)、IMU数据及传感器标定参数。

(9) HiAI是由Xiao等[44]于2025年提出的面向高空无人机多源目标跟踪的数据集。该数据集图像来自于搭载混合传感器的无人机平台,包含150组严格时空对齐的可见光-红外双模态视频序列。数据集包含了汽车、行人、电动车等9类典型目标对象,覆盖了不同天气下的城市道路、郊区环境,可用于微小目标检测。

2.4.3 数据融合挑战与应用场景分析

(1) 数据融合挑战:在多源数据集中使用不同传感器获取的数据(2D图像、3D点云、时序信号等)在数据结构、分辨率和语义层面存在显著差异。例如,视觉数据与LiDAR点云的时空对齐需依赖高精度标定,而红外与可见光图像的像素级融合需解决分辨率不匹配问题。同时由于边缘设备计算资源有限,难以支持复杂模型,轻量化融合算法设计(如知识蒸馏[67]、边缘计算优化[68])与弱监督学习算法设计(如自监督预训练)是目前重要挑战。

(2) 典型应用场景:在城市安防场景与人群计数方面,利用RGBT人群计数数据集(DroneRGBT)和多源行为分析数据集(UAV-human)数据集,可实现烟雾/夜间复杂场景下的人群密度监测与异常行为识别;在交通管理方面,由于低光场景仅使用可见光的漏检率大幅上升,单一的红外图像由于缺乏颜色信息和空中场景的复杂背景误检率增高[37],利用RGBT车辆检测数据集(DroneVehicle),训练出模态

互补的双光模型可用于全天候的车辆巡检;在环境灾害响应方面,利用热红外深度数据集(FIReStere),在可见光失效时使用热图像与LiDAR结合可实现火灾场景的立体环境建模与受困者定位,为无人机在浓烟中执行受困者热信号搜索提供了可靠的三维环境建模支持。

2.5 复杂环境特性下的数据集

2.5.1 定义与重要性

复杂环境下的低空视觉感知数据集指在各种典型恶劣环境中采集的或者基于现有理想环境数据集人工制作的图像数据集,该数据集旨在扩充现存图像数据集,提升目标检测、跟踪、识别等任务的处理难度,力求反映真实世界的各种环境,增强模型对真实场景图像的性能[69]。如图9所示,本文对典型复杂环境特性数据集的数据来源和规模分布进行统计。从图9(a)可看出,复杂环境下的图像采集面临诸多挑战,现有低空视觉感知数据集中包含了大量人工合成数据,以提升数据集的多样性和复杂性;图9(b)体现出复杂场景图像需要经过人工筛选,且同种环境的数据场景相对有限,导致数据集整体规模相对有限。复杂环境特性数据集的重要性体现在:

(1) 提升算法适应性:通过真实场景数据实拍或模拟实际场景,极大地丰富了训练数据的场景多样性与场景复杂度,增强了模型对不同场景的理解能力。

(2) 支持反无人机系统:通过提供复杂环境下的无人机图像数据,大大提升复杂场景下模型对目标无人机的识别与追踪能力,提升反无人机系统的反应速度。

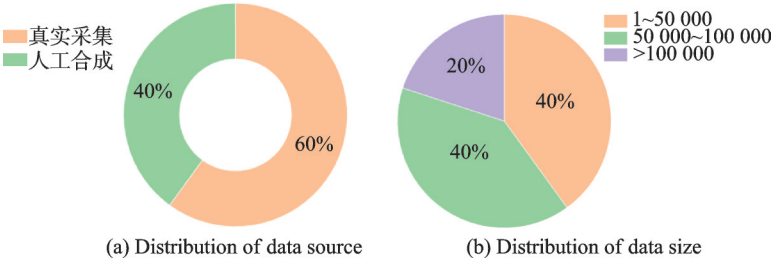


图9 典型复杂环境特性数据集分布

Fig.9 Distribution of typical complex environmental characteristics dataset

2.5.2 典型复杂环境特性数据集

(1) HazyDet[45]是一个无人机空对地航拍视角的复杂环境场景数据集,主要面向无人机航拍图像拍摄角度高、拍摄角度变化频繁导致检测目标尺寸偏小且变化较大的问题。HazyDet数据集包含了模拟雾天场景的合成图像中收集的383 000个真实世界检测目标。

(2) UAVDT[28]是由中国科学院深圳先进技术研究院发布的低空无人机空对地视觉数据集,包含大量夜间复杂天气场景。夜间场景中恶劣复杂天气会影响图像的曝光度,尤其是雨天环境图像可能出现严重的眩光问题。尽管UAVDT数据集关注到这一挑战,但是其包含的复杂场景图像的比例相对有限,并不足以完全支撑模型应对不同环境特性的需求。

(3) UAV-AWID是由Munir等[46]于沙特阿拉伯采集并构建的数据集,通过对相同的2 600张图像进行3种不同的加噪操作以模拟雨天环境、模糊环境以及人工噪声环境,并对应生成3个子数据集,每个子数据集包括高、中、低3种训练难度,分别对应加噪操作中的不同参数。该数据集中包含大量无人机图像与鸟群图像,支持评估Faster-RCNN[70]、RetinaNet[71]等感知模型的恶劣环境适应能力,为赋能感知模型在复杂环境中的鲁棒性提供了新的视角。

(4) NPS-Drones[47]是美国普渡大学与海军研究生学院于2016年发布的聚焦于复杂场景下无人机空对空检测任务的数据集,包含了70 250帧图像,标注了超过50 000个目标检测框。该数据集覆盖了

多种复杂场景与多种天空环境,如城市、郊区、开阔地带,无云层和厚云层等环境,数据集中的目标大多呈现小尺寸、低分辨率或模糊状态。

(5) ARD-MAV是Guo等[48]提出的无人机空对空感知数据集,包含60个视频和106 665帧图片。所有视频都由DJI M300和MAVIC2在中低空场景飞行拍摄,目标无人机为DJI Phantom4。数据集包含了复杂背景、非平面场景、目标遮挡、相机剧烈运动、运动模糊以及小目标无人机等多种有挑战性的环境。数据集的图片分辨率为1 920像素×1 080像素,目标平均尺寸仅为图像尺寸的0.02%。

2.5.3 应用场景分析

随着低空经济的蓬勃发展,人们对低空环境的探索力不断提升,复杂环境特性下的低空视觉感知数据集的应用需求与场景逐渐扩大。首先,在传统目标检测或跟踪任务中,这类数据集可以用于训练更为鲁棒且泛化的模型,有利于应对复杂与极端环境。其次,理解复杂环境中的任务特征有助于模型更好理解目标的本质,从而更好地优化模型已有的性能。最后,在无人机空对空监测领域,该类数据集可以有效训练和优化无人机检测、跟踪和识别算法,帮助系统在复杂天气和光照条件下准确识别和拦截非法无人机,增强安全保障。

2.6 无人机具身智能相关数据集

2.6.1 定义与重要性

具身智能是指智能体通过与物理环境的实时交互,实现感知、认知、决策和行动的智能范式,其核心在于将智能嵌入到物理实体中,使其能够从与环境的动态交互中学习和优化行为策略。在无人机领域,具身智能的引入可帮助无人机在不确定性场景以及局部环境可观测条件下理解四维时空的关联并做出准确行动,实现自主感知、行动端到端全链闭环,完成感知、决策、规划和控制等多层面结合性任务。图10展示了对无人机具身智能领域典型数据集进行的时空分布与模态特征分析。由图10(a)所示,数据集数量在2024年呈现指数级增长,印证了该领域研究热度的显著提升。如图10(b)所示,数据集涵盖多种异构数据类型,表明多源协同感知已成为无人机具身智能的核心技术路径,此外文本数据的高占比反映了任务指令语义解析的重要性。

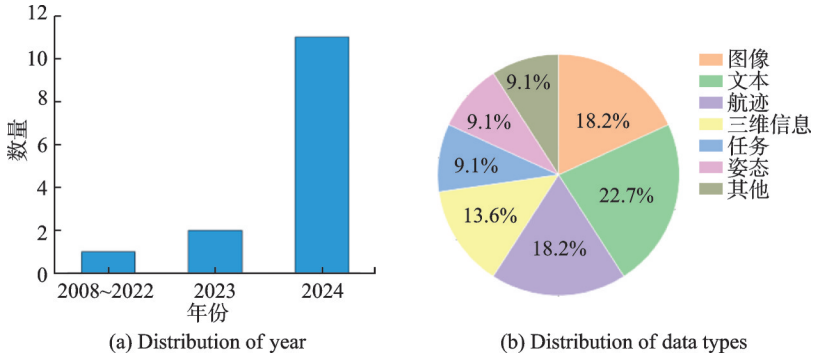


图10 典型无人机具身智能数据集分布

Fig.10 Distribution of typical embodied intelligence datasets for UAVs

2.6.2 典型具身智能相关数据集

(1) OpenBench是由Wang等[59]于2024年提出的面向智能物流任务的户外语义导航基准测试框架,旨在为自主配送系统提供长期操作能力与任务理解性能的标准化评估体系。该框架基于Gazebo仿真平台构建了3种不同规模的仿真环境,每个环境包含带有标注的建筑物模型与对应的开放街道地图(Open street map, OSM)数据,以模拟真实住宅区的复杂空间拓扑结构与动态导航需求。此外,该框架实现了跨模态能力融合,结合大语言模型(Large language models, LLMs)的语义推理能力与VLMs的

视觉定位能力，构建了无GPS依赖的全局定位与地图对齐机制，解决了传统方法在城区环境中的定位漂移问题。

（2）UAV-VLPA-nano-30是Sautenkov等[58]于2025年发布的面向全球尺度无人机路径规划任务的开放基准数据集，由30张高分辨率卫星影像构成。该数据集通过多源标注体系，建立了从自然语言指令到地理空间坐标的映射关系，支持像素坐标与无人机控制指令的精准转换。卫星影像覆盖城乡结合区、密集城区、郊区农田及自然景观等多样化地理场景，每张影像均包含由领域专家手工规划的全局飞行路径及其对应的动作指令序列，并附带完整的地理元数据，为端到端路径规划提供高可信度基准。

（3）VLA-3D是由Zhang等[57]于2024年发布的大规模三维语义场景理解与导航数据集，旨在推动多源具身智能在复杂室内环境中的应用。该数据集包含三维点云、对象级语义标注、可遍历空间标注及结构化场景图，整合了来自ScanNet[72]、Matterport3D[73]、Habitat-Matterport 3D（HM3D）[74]、3RScan[75]、ARKitScenes[76]五个数据集的11 500个真实室内场景的三维扫描数据，覆盖多房间复杂场景，涵盖286 000余个对象实例、477个细粒度语义类别、23.5×10^6个对象间空间关系及9.7×10^6个合成指称语句。

（4）CityNav是由Lee等[56]于2024年提出的无人机视觉语言导航基准数据集。该数据集基于SensatUrban数据集[77]的真实城市三维扫描数据构建，通过高精度三维重建技术实现了城市场景的高保真度还原。数据集包含32 637组自然语言指令与对应的人类示范轨迹，涵盖真实城市场景中5 850个具有地理参照的自然语言描述对象。该数据集为语言引导的空中导航任务提供了基准测试平台。

（5）AeroVerse是由中国科学院空天信息创新研究院Yao等[55]于2024年提出的新型无人机视觉语言导航基准框架。该基准框架包含多个数据集，如AerialAgent-Ego10k，它基于UrbanBIS遥感影像数据集[51]构建的首个真实城市场景无人机第一视角图像-文本预训练数据集，包含10 000张多分辨率航拍影像与细粒度场景描述文本，支持视觉语言模型的领域自适应预训练；CyberAgent-Ego500k依托Unreal Engine 4与AirSim仿真平台生成的虚拟城市对齐数据集，包含4类典型城市场景的500 000帧数据，每帧集成RGB-D图像、六自由度位姿参数及结构化场景描述文本，通过场景生成引擎构建从虚拟到现实的跨域表征迁移学习基准。

（6）OpenUAV是由北京航空航天大学Wang等[54]于2024年构建的无人机视觉语言导航数据集。该数据集利用Unreal Engine 4的渲染能力提供了多样化的场景和高保真的视觉效果，支持动态光照、植被波动等真实环境效应。数据集中包括22种不同场景，其数据分为目标对象集与轨迹数据集，涵盖89种不同的目标对象类别与12 149条真实飞行轨迹。

（7）UAV-VisLoc由北京邮电大学Xu等[53]于2024年提出的大规模无人机视觉定位基准数据集。该数据集包含6 742张无人机第一视角航拍图像和11幅高分辨率卫星地图，图像内容涵盖城市、村镇、河流等多样化的地形特征。该数据集的数据采用了固定翼无人机和多旋翼无人机，通过GNSS/IMU融合系统获取精确的地理坐标、相对高程和航向角等元数据，为视觉定位任务提供了绝对坐标真值。

（8）ARIO数据集是由Wang等[52]于2024年提出的大规模多源、多任务数据集，旨在系统解决现有具身智能数据资源在标准化程度、模态完整性及数据规模等方面的关键瓶颈。该数据集集成真实环境示教数据、多物理仿真平台数据及开源数据集转化数据，形成涵盖258个机器人系列、321 064项任务的超大规模数据集（总计3.03×10^6 episodes）。此外，数据集首次实现跨模态感知数据的系统性耦合，完整地融合了视觉、听觉、触觉及语言5类异构感知模态，并采用分层时间戳编码方案，为多模态感知决策模型提供完备数据支撑。

（9）AerialVLN是西北工业大学Liu等[50]在2023年构建的无人机视觉语言导航数据集，依托Unreal Engine 4与AirSim插件开发的高保真3D模拟器，生成25个异构化城市场景，每个场景含超870类语义实体，数据包含8 446条连续飞行路径、25 338条结构化指令、4 470个语义单元，实现子路径-子指令时空对齐。AerialVLN数据集主要应用于无人机自主导航、场景理解和语义分割、动态环境适应以及开

放场景下的视觉语言导航等研究。

(10) UrbanBIS是由深圳大学Yang等[51]在2023年提出的城市级三维理解基准数据集。该数据集基于航空摄影测量技术构建,通过113 346个视点的高精度航拍影像进行三维重建,形成包含2.5亿样本点的三维点云数据库、3 370个独立建筑单体。数据集提供亚米级分辨率航拍影像、三维稠密点云及三角网格模型,构建了分层语义标注体系。该数据集适用于城市数字孪生建模、大规模场景语义理解、建筑群形态分析等应用场景。

(11) AVDN是由加州大学圣克鲁兹分校团队[49]提出的首个面向无人机视觉与对话导航的多源数据集。该数据集基于xView数据集[78]的高分辨率卫星图像构建连续状态空间模拟器,生成无人机俯视视角的连续视觉观测序列,包含3 064条飞行轨迹及异步人机对话数据。数据集使用多源协同架构,采用地理区域划分策略,使用异步交互模式模拟了非实时监控场景,适用于自然语言指令动态修正、多源场景理解及注意力驱动的导航决策等任务。

2.6.3 特点与应用场景分析

(1) 无人机具身智能数据集特点:① 多源数据融合:无人机具身智能数据集普遍集成多源异构感知模态,通过多传感器同步采集与时空对齐,实现高精度环境建模与状态估计;② 动态环境适应性:通过参数化引擎或动态仿真技术模拟光照变化、雨雪扰动、植被波动等复杂条件,提升无人机在不确定性场景下的鲁棒性;③ 语义认知与空间推理:引入自然语言指令、结构化场景标注及多步推理任务,强化无人机对全局语义地图与时空关联的认知能力。

(2) 无人机具身智能数据集应用场景:① 无人机自主导航与避障:支持无人机在密集障碍环境或非视距条件下实现路径规划与动态避障,可应用于物流配送、紧急物资投送等任务;② 语义引导的任务执行:基于自然语言指令或对话交互,实现地标识别、目标搜索及语义分割;③ 动态环境适应:通过仿真极端天气或高动态目标,提升电力巡检、野生动物监测等任务中的抗干扰能力。

3 总结与未来展望

近年来,随着无人机技术的快速发展和低空智能应用场景的不断拓展,低空视觉感知数据集的研究与构建取得了显著进展,这将为低空视觉算法的发展提供重要支撑。本节将对于低空视觉感知数据集进行总结,并展望未来的发展方向。

3.1 研究现状总结

通过对当前工作的整理发现,现有的数据集已经涵盖了单无人机、多无人机、单任务与多任务等多种应用场景。同时,目前的研究人员认为多源融合会逐渐成为低空视觉感知未来的一个重要趋势,针对可见光、红外、深度等多源数据集的构建能够显著增强无人机在复杂环境下的感知鲁棒性。此外,数据集的场景适应性也在逐步提升,如SDD[26]等数据集已经开始纳入城市、工业、农业等不同应用场景。尽管目前低空视觉感知数据集已取得显著进展,但在支持无人机的协同化需求上仍然面临以下挑战:

(1)标注成本与效率失衡:当前数据集高度依赖人工标注,尤其在多任务与多源场景下,标注流程复杂且一致性难以保障,导致规模化扩展受限。以VisDrone数据集[7]为例,54万个实例标注需耗费3 000 h人工工时,而多源数据集DroneRGBT[38]的RGB-热红外图像的像素级标注误差仍达5~8像素,导致数据集规模化扩展受限。

(2)多源数据复用性不足:尽管多源数据集已初步涌现,但缺乏统一的标注标准与对齐协议。如:可见光与红外图像的时间同步偏差(如UAV-Human[31]存在20 ms级时序误差),深度信息与RGB数据的空间配准误差等问题(如SynDrone[41]的LiDAR-视觉外参标定残差大于0.1 m),严重制约了多源模型的训练效果。

(3)极端环境覆盖薄弱:现有数据集多以理想天气与静态场景为主,雨、雪、雾等恶劣气象条件下的

数据稀缺(HazyDet[45]的雾天数据中仅15%为真实采集),且动态环境建模(如突发障碍物、移动目标交互)尚未形成系统性框架,限制了算法的实际部署能力。

(4)具身智能数据割裂:无人机导航与群体协同需深度融合机体状态与环境感知,但当前数据集多孤立标注感知目标,缺乏飞行器动力学参数与场景交互的联合表征,例如,AerialVLN[50]虽提供无人机位姿数据,却未关联动态障碍物响应;MAVREC[17]的多视角数据未模拟5G NR(New radio)链路下的通信延迟,难以支撑端到端自主决策模型的训练。

3.2 未来发展方向展望

为了改进现有低空视觉感知数据集的不足,并结合低空应用场景的特殊性和技术发展趋势,未来研究还需要关注以下几个关键方向:

(1)数据多样性与标注效率的提升优化:低空场景的动态性和复杂性要求数据集能够覆盖更多不同的环境和任务。目前来看极端天气下的数据难以采集,研究人员可以通过物理仿真技术来生成极端天气下的合成数据来补充不同类型的数据;在标注效率方面,研究人员可以尝试使用预训练模型生成伪标签,从而减少人工标注工作,提高效率。

(2)多源数据融合与标准化建设:低空视觉通常需要依赖可见光、红外、深度等多源数据来进行协同分析,针对多源数据融合,未来应该研究设计跨模态对齐的数据集架构。同时,数据集的标准化问题也需要得到重视,随着不同场景、任务的数据集层出不穷,研究人员需要为多源数据存储、标注与评估制定统一标准,以便更好地复用数据和进行算法对比验证。

(3)恶劣天气与动态场景的深度覆盖:现有数据集对恶劣天气和动态环境的覆盖仍然不足。未来的研究应该重点关注如何采集极端气象条件下的数据,并与气象部门合作构建可量化的标注体系,并设计针对性的标注策略。此外,针对动态场景这方面,研究人员应该构建包含突发障碍物、移动目标交互的场景库,模拟真实飞行中的环境变化,以此来支撑无人机的真实应用需求。

(4)面向具身智能的数据基准设计:为了满足无人机自主导航与群体协同的需求,低空感知数据集需深度融合具身信息。例如,在为数据集标注环境目标的同时,同步记录无人机的位姿、速度、控制指令等状态数据,构建“环境-无人机本体”联合表征的数据标注框架;针对多机协同感知与决策规划任务方面,研究人员需要构建时空同步的多视角感知与决策协同优化的框架,以此来支持无人机集群具身算法的开发与研究。

参考文献:

[1] 浦口区人民政府. 低空空域知识科普[EB/OL]. (2024-09-10). http://www.pukou.gov.cn/hdjlpd/zsk/202409/t20240911_4762453.html.

[2] 人民日报. 首次写入政府工作报告——“低空经济”加速起飞[EB/OL]. (2024-04-02). https://www.gov.cn/yaowen/liebiao/202404/content_6943071.html.

[3] 陈勇, 杨健, 张余, 等. 面向低空经济的无人机通信频谱管理政策、标准与技术[J]. 数据采集与处理, 2025, 40(1): 2-26.
CHEN Yong, YANG Jian, ZHANG Yu, et al. Spectrum management regulations, standards, and technologies of unmanned aerial vehicle communication for low altitude economy[J]. Journal of Data Acquisition and Processing, 2025, 40(1): 2-26.

[4] 常宇轩, 杨文, 吴金建. 低代价高动态大视场低慢小飞行器检测与跟踪[J]. 数据采集与处理, 2025, 40(1): 86-101.
CHANG Yuxuan, YANG Wen, WU Jinjian. Low-cost, high-dynamic, large field-of-view detection and tracking of low-slow-small aerial vehicles[J]. Journal of Data Acquisition and Processing, 2025, 40(1): 86-101.

[5] MUELLER M, SMITH N, GHANEM B. A benchmark and simulator for UAV tracking[C]//Proceedings of Computer Vision—ECCV 2016. Cham: Springer International Publishing, 2016: 445-461.

[6] BARNES D, GADD M, MURCUTT P, et al. The Oxford radar RobotCar dataset: A radar extension to the Oxford RobotCar dataset[C]//Proceedings of 2020 IEEE International Conference on Robotics and Automation (ICRA). Paris, France: IEEE, 2020: 6433-6438.

[7] DU D, WEN L, ZHU P, et al. VisDrone-DET2020: The vision meets drone object detection in image challenge results[C]//Proceedings of Computer Vision—ECCV 2020 Workshops. Cham: Springer International Publishing, 2020: 692-712.

[8] ZHU P, PENG T, DU D, et al. Graph regularized flow attention network for video animal counting from drones[J]. IEEE Transactions on Image Processing, 2021, 30: 5339-5351.

[9] CAO B, YAO H, ZHU P, et al. Visible and clear: Finding tiny objects in difference map[C]//Proceedings of Computer Vision—ECCV 2024. Cham: Springer International Publishing, 2025: 1-18.

[10] CAO B, LU Q, FENG J, et al. Efficient masked AutoEncoder for video object counting and a large-scale benchmark[EB/OL]. (2024-11-20). https://arxiv.org/abs/2411.13056v2.

[11] XIAO Y, WANG J, ZHAO Z, et al. UAV video vehicle detection: Benchmark and baseline[J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 5609814.

[12] ZHU P, ZHENG J, DU D, et al. Multi-drone-based single object tracking with agent sharing network[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(10): 4058-4070.

[13] HU Y, FANG S, LEI Z, et al. Where2comm: Communication-efficient collaborative perception via spatial confidence maps [C]//Proceedings of the Advances in Neural Information Processing Systems. LA, USA: MIT, 2022: 4874-4886.

[14] WANG Z, CHENG P, CHEN M, et al. Drones help drones: A collaborative framework for multi-drone object trajectory prediction and beyond[EB/OL]. (2024-05-23). https://arxiv.org/abs/2405.14674.

[15] LIU Z, SHANG Y, LI T, et al. Robust multi-drone multi-target tracking to resolve target occlusion: A benchmark[J]. IEEE Transactions on Multimedia, 2023, 25: 1462-1476.

[16] BOZCAN I, KAYACAN E. AU-AIR: A multi-modal unmanned aerial vehicle dataset for low altitude traffic surveillance[C]//Proceedings of 2020 IEEE International Conference on Robotics and Automation (ICRA). Paris, France: IEEE, 2020: 8504-8510.

[17] DUTTA A, DAS S, NIELSEN J, et al. Multiview aerial visual recognition (MAVREC): Can multi-view improve aerial visual perception?[C]//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2024: 22678-22690.

[18] YE H, SUNDERRAMAN R, JI S. UAV3D: Large-scaleA 3D perception benchmark for unmanned aerial vehicles[EB/OL]. (2024-10-14). https://arxiv.org/abs/2410.11125.

[19] RAZAKARIVONY S, JURIE F. Vehicle detection in aerial imagery: A small target detection benchmark[J]. Journal of Visual Communication and Image Representation, 2016, 34: 187-203.

[20] MUNDHENK T N, KONJEVOD G, SAKLA W A, et al. A large contextual dataset for classification, detection and counting of cars with deep learning[C]//Proceedings of Computer Vision—ECCV 2016. Cham: Springer International Publishing, 2016: 785-800.

[21] XIA G S, BAI X, DING J, et al. DOTA: A large-scale dataset for object detection in aerial images[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 3974-3983.

[22] ZAMIR W S, ARORA A, GUPTA A, et al. iSAID: A large-scale dataset for instance segmentation in aerial images[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. CA, USA: IEEE, 2019: 28-37.

[23] SHERMEYER J, HOSSLER T, ETTEN A V, et al. RarePlanes: Synthetic data takes flight[C]//Proceedings of 2021 IEEE Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA: IEEE, 2021: 207-217.

[24] YU X, GONG Y, JIANG N, et al. Scale match for tiny person detection[C]//Proceedings of 2020 IEEE Winter Conference on Applications of Computer Vision (WACV). Snowmass, CO, USA: IEEE, 2020: 1246-1254.

[25] LYU Y, VOSSELMAN G, XIA G S, et al. UAVid: A semantic segmentation dataset for UAV imagery[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 165: 108-119.

[26] ROBICQUET A, SADEGHIAN A, ALAHI A, et al. Learning social etiquette: Human trajectory understanding in crowded scenes[C]//Proceedings of Computer Vision—ECCV 2016. Cham: Springer International Publishing, 2016: 549-565.

[27] BAREKATAIN M, MARTÍ M, SHIH H F, et al. Okutama-action: An aerial view video dataset for concurrent human action detection[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Honolulu, HI, USA: IEEE, 2017: 2153-2160.

[28] DU D, QI Y, YU H, et al. The unmanned aerial vehicle benchmark: Object detection and tracking[C]//Proceedings of the

European Conference on Computer Vision (ECCV). Munich, Germany: Springer, 2018: 370-386.

[29] WEN L, DU D, ZHU P, et al. Detection, tracking, and counting meets drones in crowds: A benchmark[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: IEEE, 2021: 7808-7817.

[30] JIANG N, WANG K, PENG X, et al. Anti-UAV: A large multi-modal benchmark for UAV tracking[J]. IEEE Transactions on Multimedia, 2021, 25: 486-500.

[31] LI T, LIU J, ZHANG W, et al. UAV-human: A large benchmark for human behavior understanding with unmanned aerial vehicles[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: IEEE, 2021: 16261-16270.

[32] HSIEH M R, LIN Y L, HSU W H. Drone-based object counting by spatially regularized regional proposal network[C]//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017: 4165-4173.

[33] PORTMANN J, LYNEN S, CHLI M, et al. People detection and tracking from aerial thermal views[C]//Proceedings of 2014 IEEE International Conference on Robotics and Automation (ICRA). Hong Kong, China: IEEE, 2014: 1794-1800.

[34] LI M, ZHAO X, LI J, et al. Object detection in UAV-borne thermal images using boundary-aware saliency maps[J]. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2020, XLIII-B2-2020: 1233-1238.

[35] BONDI E, JAIN R, AGGRAWAL P, et al. BIRDSAI: A dataset for detection and tracking in aerial thermal infrared videos [C]//Proceedings of 2020 IEEE Winter Conference on Applications of Computer Vision (WACV). Snowmass, CO, USA: IEEE, 2020: 1736-1745.

[36] SUO J, WANG T, ZHANG X, et al. HIT-UAV: A high-altitude infrared thermal dataset for unmanned aerial vehicle-based object detection[J]. Scientific Data, 2023, 10(1): 227.

[37] SUN Y, CAO B, ZHU P, 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.

[38] PENG T, LI Q, ZHU P. RGB-T crowd counting from drone: A benchmark and MMCCN network[C]//Proceedings of Computer Vision—ACCV 2020. Cham: Springer International Publishing, 2021: 497-513.

[39] FENG H, ZHANG L, ZHANG S, et al. RTDOD: A large-scale RGB-thermal domain-incremental object detection dataset for UAVs[J]. Image and Vision Computing, 2023, 140: 104856.

[40] KHOSE S, PAL A, AGARWAL A, et al. SKYSCENES: A synthetic dataset for aerial scene understanding[C]//Proceedings of Computer Vision—ECCV 2024. Cham: Springer Nature Switzerland, 2025: 19-35.

[41] LENHARD T R, WEINMANN A, FRANKE K, et al. SynDroneVision: A synthetic dataset for image-based drone detection [EB/OL]. (2024-11-08). https://arxiv.org/abs/2411.05633v1.

[42] DHRAFANI D, LIU Y, JONG A, et al. FIReStereo: Forest InfraRed stereo dataset for UAS depth perception in visually degraded environments[J]. IEEE Robotics and Automation Letters, 2025, 10(4): 3302-3309.

[43] YAO L, LIU F, XU S, et al. UEMM-air: Make unmanned aerial vehicles perform more multi-modal tasks[EB/OL]. (2024-06-10). https://arxiv.org/abs/2406.06230v3.

[44] XIAO Y, CAO D, LI C, et al. A benchmark dataset for high-altitude UAV multi-modal tracking[J]. Journal of Image and Graphics, 2025, 30(2): 361-374.

[45] FENG C, CHEN Z, KOU R, et al. HazyDet: Open-source benchmark for drone-view object detection with depth-cues in hazy scenes[EB/OL]. (2024-09-10). https://arxiv.org/abs/2409.19833v1.

[46] MUNIR A, SIDDIQUI A J, ANWAR S, et al. Impact of adverse weather and image distortions on vision-based UAV detection: A performance evaluation of deep learning models[J]. Drones, 2024, 8(11): 638.

[47] LI J, YE D H, CHUNG T, et al. Multi-target detection and tracking from a single camera in unmanned aerial vehicles (UAVs) [C]//Proceedings of 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Daejeon, Korea: IEEE, 2016: 4992-4997.

[48] GUO H, ZHENG Y, ZHANG Y, et al. Global-local MAV detection under challenging conditions based on appearance and motion[J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(9): 12005-12017.

[49] FAN Y, CHEN W, JIANG T, et al. Aerial vision-and-dialog navigation[EB/OL]. (2022-05-24). https://arxiv. org/abs/2205.12219.

[50] LIU S, ZHANG H, QI Y, et al. AerialVLN: Vision-and-language navigation for UAVs[C]//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE, 2023: 15338-15348.

[51] YANG G, XUE F, ZHANG Q, et al. UrbanBIS: A large-scale benchmark for fine-grained urban building instance segmentation[C]//Proceedings of Special Interest Group on Computer Graphics and Interactive Techniques Conference. Los Angeles, CA, USA: ACM, 2023: 1-11.

[52] WANG Z, ZHENG H, NIE Y, et al. All robots in one: A new standard and unified dataset for versatile, general-purpose embodied agents[EB/OL]. (2024-08-20). https://arxiv.org/abs/2408.10899.

[53] XU W, YAO Y, CAO J, et al. UAV-VisLoc: A large-scale dataset for UAV visual localization[EB/OL]. (2024-05-20). https://arxiv.org/abs/2405.11936.

[54] WANG X, YANG D, WANG Z, et al. Towards realistic UAV vision-language navigation: Platform, benchmark, and methodology[EB/OL]. (2024-10-09). https://arxiv.org/abs/2410.07087.

[55] YAO F, YUE Y, LIU Y, et al. AeroVerse: UAV-agent benchmark suite for simulating, pre-training, finetuning, and evaluating aerospace embodied world models[EB/OL]. (2024-08-28). https://arxiv.org/abs/2408.15511.

[56] LEE J, MIYANISHI T, KURITA S, et al. CityNav: Language-goal aerial navigation dataset with geographic information [EB/OL]. (2024-06-20). https://arxiv.org/abs/2406.14240.

[57] ZHANG H, ZANTOUT N, KACHANA P, et al. VLA-3D: A dataset for 3D semantic scene understanding and navigation [EB/OL]. (2024-11-05). https://arxiv.org/abs/2411.03540.

[58] SAUTENKOV O, YAQOOT Y, LYKOV A, et al. UAV-VLA: Vision-language-action system for large scale aerial mission generation[EB/OL]. (2025-01-09). https://arxiv.org/abs/2501.05014.

[59] WANG J, HUO D, XU Z, et al. OpenBench: A new benchmark and baseline for semantic navigation in smart logistics[EB/OL]. (2025-02-13). https://arxiv.org/abs/2502.09238.

[60] YANG Y, AVILES-RIVERO A I, FU H, et al. Video adverse-weather-component suppression network via weather messenger and adversarial back propagation[C]//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE, 2023: 13154-13164.

[61] RAHNEMOONFAR M, CHOWDHURY T, MURPHY R. RescueNet: A high resolution UAV semantic segmentation dataset for natural disaster damage assessment[J]. Scientific Data, 2023, 10(1): 913.

[62] MENOUAR H, GUVENC I, AKKAYA K, et al. UAV-enabled intelligent transportation systems for the smart city: Applications and challenges[J]. IEEE Communications Magazine, 2017, 55(3): 22-28.

[63] 陈琳, 刘允刚. 面向无人机的视觉目标跟踪算法: 综述与展望[J]. 信息与控制, 2022, 51(1): 23-40.
CHEN Lin, LIU Yungang. UAV visual target tracking algorithms: Review and future prospect[J]. Information and Control, 2022, 51(1): 23-40.

[64] SHAH S, DEY D, LOVETT C, et al. AirSim: High-fidelity visual and physical simulation for autonomous vehicles[C]//Proceedings of Field and Service Robotics. Cham: Springer International Publishing, 2018: 621-635.

[65] DOSOVITSKIY A, ROS G, CODEVILLA F, et al. CARLA: An open urban driving simulator[C]//Proceedings of Machine Learning Research. California, USA: PMLR, 2017: 1-16.

[66] CAESAR H, BANKITI V, LANG A H, et al. nuScenes: A multimodal dataset for autonomous driving[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 11618-11628.

[67] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network[EB/OL]. (2015-03-02). https://arxiv.org/abs/1503.02531v1.

[68] CHANG H, CHEN Y, ZHANG B, et al. Multi-UAV mobile edge computing and path planning platform based on reinforcement learning[EB/OL]. (2021-02-02). https://arxiv.org/abs/2102.02078v3.

[69] AKHTAR N, JALWANA M A A K, BENNAMOUN M, et al. Attack to fool and explain deep networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(10): 5980-5995.

[70] REN S, HE K, 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.

[71] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017: 2999-3007.

[72] DAI A, CHANG A X, SAVVA M, et al. ScanNet: Richly-annotated 3D reconstructions of indoor scenes[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 2432-2443.

[73] CHANG A, DAI A, FUNKHOUSER T, et al. Matterport3D: Learning from RGB-D data in indoor environments[C]//Proceedings of 2017 International Conference on 3D Vision (3DV). Qingdao, China: IEEE, 2017: 667-676.

[74] RAMAKRISHNAN S K, GOKASLAN A, WIJMANS E, et al. Habitat-matterport 3D dataset (HM3D): 1000 large-scale 3D environments for embodied AI[EB/OL]. (2021-09-16). https://arxiv.org/abs/2109.08238.

[75] WALD J, AVETISYAN A, NAVAB N, et al. RIO: 3D object instance re-localization in changing indoor environments[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea: IEEE, 2019: 7657-7666.

[76] BARUCH G, CHEN Z, DEHGHAN A, et al. ARKitScenes—A diverse real-world dataset for 3D indoor scene understanding using mobile RGB-D data[EB/OL]. (2021-11-17). https://arxiv.org/abs/2111.08897.

[77] HU Q, YANG B, KHALID S, et al. Towards semantic segmentation of urban-scale 3D point clouds: A dataset, benchmarks and challenges[C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: IEEE, 2021: 4975-4985.

[78] LAM D, KUZMA R, MCGEE K, et al. xView: Objects in context in overhead imagery[EB/OL]. (2018-02-22). https://arxiv.org/abs/1802.07856.

作者简介:

孙一铭(1996-),男,助理研究员,研究方向:无人机视觉、人工智能,E-mail:sunyiming@seu.edu.cn。

赵柯嘉(2003-),男,博士研究生,研究方向:无人机视觉、人工智能。

王硕(2002-),男,硕士研究生,研究方向:无人机视觉、人工智能。

陈振国(2003-),男,硕士研究生,研究方向:无人机视觉。

阮媛(2002-),女,硕士研究生,研究方向:无人机视觉、人工智能。

叶子凡(2002-),男,硕士研究生,研究方向:无人机视觉、人工智能。

陈星睿(2003-),男,硕士研究生,研究方向:无人机视觉、人工智能。

李欣(2002-),男,硕士研究生,研究方向:无人机视觉、人工智能。

褚瑞麟(2004-),男,硕士研究生,研究方向:机器视觉。

宋生敏(2003-),男,硕士研究生,研究方向:无人机视觉、人工智能。

胡亦添(2003-),男,硕士研究生,研究方向:无人机视觉、人工智能。

郭周鹏(2000-),男,博士研究生,研究方向:小目标检测、人工智能。

王森(2000-),男,博士研究生,研究方向:无人机具身智能。

胡清华(1976-),男,教授,研究方向:人工智能。

朱鹏飞(1986-),**通信作者**,男,教授,研究方向:无人机视觉、人工智能,E-mail:zhupengfei@tju.edu.cn。

(编辑:张黄群)