中图法分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1006-8961(XXXX)XX-0001-38

论文引用格式:Bi Shifan, Ye Liang, Wang Zhixiang, Zhang Ziyang, Hong Hanyu, Sang Nong. XXXX. Research progress on object detection and tracking methods for nighttime UAV aerial imagery. Journal of Image and Graphics, XX(XX):0001-0038(毕诗帆, 叶亮, 王志祥, 张梓阳, 洪汉玉, 桑农. XXXX. 夜间无人机航拍图像目标检测与跟踪方法研究进展. 中国图象图形学报, XX(XX):0001-0038)[DOI:10.11834/jig.250459]

夜间无人机航拍图像目标检测与跟踪方法研究进展

毕诗帆[1,2],叶亮[1,2],王志祥[1,2],张梓阳[1,2],洪汉玉[1,2],桑农[3]

1. 武汉工程大学电气信息学院,武汉　430205; 2. 武汉工程大学光学信息与模式识别湖北省重点实验室,武汉　430205; 3. 华中科技大学多谱信息智能处理技术全国重点实验室,武汉　430074

摘　要:视觉目标检测与跟踪技术已在白天场景中取得显著突破,为无人机(unmanned aerial vehicle,UAV)在智能领域的广泛应用提供了强大支撑。然而,这些方法在夜间场景下往往表现不佳,检测与跟踪精度显著下降。夜间作为无人机应用中不可或缺的场景,其复杂性与挑战性凸显了开展针对夜间无人机航拍图像目标检测与跟踪研究的必要性和现实意义。针对夜间无人机目标航拍图像检测与跟踪技术的现状及发展趋势,本文分析了感知能力有限、可视化特征不足、硬件平台资源受限以及复杂成像条件等因素所带来的挑战。从夜间无人机航拍图像目标检测研究出发,综述了夜间图像增强、域适应学习、多模态感知融合和轻量化模型等方法的研究进展。在夜间无人机航拍图像目标跟踪方面,重点综述了基于深度学习的五类范式,包括先增强后跟踪、域自适应、视觉提示学习、课程学习和多模态融合,系统总结了相关方法的优缺点及所应对的挑战。随后,介绍了夜间及全天候无人机航拍图像目标检测与跟踪常用的评价指标与典型数据集,并在构建的夜间无人机车辆目标检测集DroneVehicle-Night上进行性能评估与对比分析;同时,从VisDrone2019的测试集中筛选昼夜样本,对现有检测方法的夜间适应性进行了对比测试;此外,还汇总了包含四类跟踪范式在内的20种算法在夜间无人机航拍图像目标跟踪数据集UAVDark135与NAT2021上的性能评估结果。最后,对夜间无人机航拍图像目标检测与跟踪未来的发展方向进行了展望,为该领域的后续研究提供参考。本文实验所用到的算法、构建的数据集已经汇总至https://github.com/bsfsf/DroneVehicle-Night和https://doi.org/10.57760/sciencedb.32435以便后续研究者使用。

关键词:深度学习;无人机航拍图像;夜间;目标跟踪;目标检测

Research progress on object detection and tracking methods for nighttime UAV aerial imagery

Bi Shifan[1,2], Ye Liang[1,2], Wang Zhixiang[1,2], Zhang Ziyang[1,2], Hong Hanyu[1,2], Sang Nong[3]

1. School of Electrical and Information Engineering, Wuhan Institute of Technology, Wuhan 430205, China; 2. Hubei Proviacial Key Laboratory of Optical Information and Pattern Recognition, Wuhan Institute of Technology, Wuhan 430205, China; 3. State Key Laboratory of Multispectral Information Intelligent Processing Technology, Huazhong University of Science and Technology, Wuhan 430074, China

Abstract: Nighttime perception remains a critical bottleneck for the autonomous operation of unmanned aerial vehicles (UAVs). Under low-light conditions, weak and uneven illumination amplifies noise and glare, background interference

收稿日期:2025-09-22;**修回日期**:2025-12-26

* **通信作者**:叶亮,通信作者,男,讲师,主要研究方向为图像处理与图像分析、计算机视觉。E-mail:yeliang@wit.edu.cn;叶亮　yeliang@wit.edu.cn

基金项目:国家自然科学基金项目(62171329);武汉工程大学科学研究基金项目(K2023057)

Supported by: National Natural Science Foundation of China(62171329) and Wuhan University of Engineering Scientific Research Fund Project(K2023057)

becomes increasingly complex, and targets often appear small, low-contrast, or partially occluded. Meanwhile, UAV platforms impose strict constraints on onboard computation and power consumption, requiring detection and tracking methods to strike a balance between robustness and efficiency. This paper presents a comprehensive review and empirical study of the application of deep learning in UAV nighttime target detection and tracking, synthesizing methodological trends, experimental findings, and future research directions. We first characterize the unique challenges of UAV nighttime vision in contrast to daytime scenarios, highlighting four interrelated issues: 1) limited perceptual signals and variable illumination, leading to degraded pixel-level cues; 2) feature degradation and background interference caused by artificial light sources and shadows; 3) restricted onboard resources, making it difficult to deploy large-scale models; and 4) complex imaging conditions, where target scale and appearance fluctuate frequently. These constraints underscore the need for specialized algorithmic paradigms rather than direct transfer of daytime methods. From the perspective of UAV nighttime target detection, this paper systematically reviews four major research directions. First, nighttime image enhancement methods employ dedicated restoration networks to improve brightness and contrast before detection. While these approaches can effectively enhance visual quality, the two-stage pipeline may introduce artifacts and does not always optimally align with detector objectives. Second, domain adaptation methods leverage adversarial learning, style transfer, or feature alignment to bridge the day-night distribution gap. Such methods reduce dependence on annotated nighttime data but remain sensitive to inter-domain diversity and often require carefully designed pseudo-labeling or regularization. Third, multimodal perception fusion—most commonly RGB-infrared (IR) fusion—utilizes complementary thermal cues to mitigate illumination deficiencies. State-of-the-art fusion networks (e. g. M^2D-LIF) have demonstrated significant mAP improvements on UAV nighttime benchmarks, with some models achieving over 80% mAP@0. 5 on the DroneVehicle-Night subset. Fourth, lightweight models (e. g. CSPDarkNet-based backbones) focus on balancing accuracy and efficiency under limited computational budgets. Experiments indicate that backbone choice is critical for enabling real-time nighttime detection without significant accuracy loss. For nighttime target tracking, this paper categorizes four representative paradigms: 1) Enhancement-before-tracking: low-light enhancers are used as a preprocessing stage (sometimes co-optimized) to improve input quality and thereby downstream tracker performance. 2) Domain adaptation: extending static feature alignment to temporal modeling, sometimes combined with synthetic nighttime sequences or foundation models, to enhance robustness. 3) Prompt learning: reframing day-night transfer as a fast adaptation task, leveraging small-scale prompts for efficient transfer with minimal parameter updates. 4) Curriculum learning: ordering training samples from easy to difficult (e. g. progressively increasing darkness or motion blur) to stabilize training and improve performance under extreme degradation. 5) Multimodal fusion. We further analyze the interplay among these paradigms and the tension between restoration-oriented and task-oriented objectives. Subsequently, the commonly used evaluation metrics and benchmark datasets for nighttime and all-weather UAV detection and tracking are introduced, and a nighttime UAV vehicle detection dataset, DroneVehicle-Night, is constructed. To support this survey, two complementary empirical studies are conducted. The first study, based on DroneVehicle-Night (derived from DroneVehicle), contains tens of thousands of paired RGB-IR annotated images covering urban roads, residential areas, and parking lots, with five classes of vehicles annotated. Experiments reveal that under identical training conditions, IR consistently outperforms RGB (e. g. CSPDarkNet-based detectors achieve higher mAP@0. 5 in IR), while multimodal fusion methods such as M2D-LIF achieve over 83. 9% mAP@0. 5, underscoring the value of modality complementarity in nighttime detection. The second study evaluates the transferability of generic UAV detectors to nighttime. From VisDrone2019, we constructed a paired subset (521 nighttime/513 daytime images). While generic detectors achieve 30%-45% mAP@0. 5 on daytime data, their performance drops by 8%-12% at night, with small and low-contrast objects exhibiting the most severe degradation. These findings confirm that specialized nighttime and modality-aware methods are effective, while simple day-night transfer remains insufficient. To further reflect the state of the literature on nighttime tracking, this paper summarizes the performance of 20 representative trackers on UAVDark135 and NAT2021 (results cited from original publications). Among them, Mamba-based methods (e. g. MambaNUT-Small, MambaTrack) and several domain adaptation methods (e. g. DARTer, DCPT) consistently rank among the top, achieving robust AUC - accuracy trade-offs while maintaining reasonable frame rates—highlighting the importance of efficient sequence modeling and domain-aware adaptation in nighttime tracking. Based on the survey and empirical evidence, sev-

eral key conclusions are drawn. First, 1)sensor modality selection is critical: IR and RGB-IR fusion significantly enhance robustness to low-light and complex backgrounds, whereas RGB-only pipelines remain fragile. Second, 2)backbone design strongly influences the accuracy-efficiency balance: compared to heavier backbones such as ResNet-50, CSPDarkNet variants offer more attractive real-time trade-offs on embedded platforms. Third, 3)explicit temporal modeling and sequence-oriented architectures are indispensable for robust nighttime tracking, beyond frame-by-frame strategies. Fourth, 4)day-night transfer remains a major challenge—as evidenced by the VisDrone experiments, which highlight the urgent need for specialized nighttime designs and datasets. Finally, we identify four promising research directions: 1)cross-modal vision-language fusion, leveraging large-scale language/vision models to enhance multimodal alignment and interactive tracking; 2)construction of richer multimodal nighttime datasets, extending beyond urban scenes to cover wild, mountainous, and aquatic environments; 3) unsupervised and weakly supervised learning, to reduce annotation costs and improve cross-domain robustness; and 4) Lightweight Architecture Design. Beyond accuracy, UAV tasks demand real-time and efficient models. Future work may explore long-sequence modeling with Mamba, combined with NAS, dynamic inference, and efficient operators to balance accuracy, latency, and energy. In conclusion, this survey and empirical study provide a clear roadmap toward robust UAV nighttime perception and practical deployment. All algorithms and datasets used in this work are publicly available at: https://github. com/bsfsf/DroneVehicle-Nightandhttps://doi. org/10. 57760/sciencedb. 32435, to facilitate future research.

Key words: deep learning; unmanned aerial vehicles aerial image; nighttime; target tracking; object detection

0 引 言

无人机视觉目标检测与跟踪已逐渐成为智能无人系统研究的核心问题，其目标在于使无人机在复杂动态环境中自主完成识别、定位与持续跟踪，从而实现对环境的高效感知与任务决策支持。在实际应用中，该技术已在搜索和救援任务(Paulin 等，2024)、战场态势侦察(蒋超 等，2021)、无人自主飞行(Wang 等，2025)以及双目定位(Zheng 等，2023)等场景中展现出重要价值。然而，夜间作为一种典型且极具挑战性的任务场景，其低照度、高噪声和目标可见性弱等特征直接导致感知精度与环境理解能力下降，进而影响无人机任务的稳健执行。因此，要充分挖掘无人机视觉的潜力，亟需探索面向夜间环境的高效检测与跟踪技术，从而提升多功能性与环境适应能力(Liu 等，2021；Zhao 等，2018；Bolme 等，2010)。

已有学者对目标检测与跟踪方向进行了较为系统的综述。相关检测类综述研究主要聚焦于弱监督检测、遥感小目标检测、自动驾驶三维检测、无人机目标检测及通用视觉检测等典型任务，其研究焦点集中在检测算法整体框架、尺度自适应建模以及特征表示与增强策略等方面。这类综述系统梳理了检测任务的发展脉络，但在特定复杂成像条件(如低光照、夜间等)下的研究仍相对薄弱。

相较之下，近年来的目标跟踪综述工作主要关注不同跟踪任务范式及模态扩展，包括单目标跟踪、多模态跟踪以及卫星视频跟踪等方向，但整体仍以通用视觉任务为主，对无人机夜间等特定应用场景的系统性研究相对不足。

总体而言，现有综述已有较为系统的总结，但针对夜间无人机航拍图像的目标检测与跟踪研究尚无系统性梳理。为此，本文在综合前人工作的基础上，进一步整理了2019至2025年间国内外具有代表性的综述论文，总结其任务范围、适用场景以及分类角度(见表1)，以突出夜间无人机视觉这一兼具学术价值与工程挑战的前沿研究方向。

本文结构框架如图1所示，共分为六个部分：1)引言，简要介绍无人机视觉目标检测与跟踪的定义、应用价值及夜间场景的局限性，并对已发表综述进行简要梳理；2)夜间无人机航拍图像目标检测与跟踪的挑战，从夜间环境特性和无人机视角等角度分析主要问题；3)夜间无人机航拍图像目标检测与跟踪方法，围绕夜间无人机航拍场景下目标检测与跟踪的共性问题，系统梳理相关方法的研究进展，主要可归纳为五类代表性技术路线：轻量化模型、多模态感知融合、域自适应学习、图像增强以及提示学习与课程学习等，并对各类方法的核心思想、适用场景及优化方向进行总结与分析；4)数据集及评价体系，介

表1　与已发表目标检测与跟踪综述对比

Table 1　Comparison with the total number of published object detection and tracking

任务类别	文献	年份	任务范围	涵盖夜间场景	包含无人机视角	分类角度
检测	曹家乐等人(2022)	2022	通用视觉目标检测	×	×	按“传感器数量(单目/双目)”划分大类,单目下按照“锚点机制/预测模式”细分,双目下按照“特征空间”细分,同时补充子模块设计(特征、预测头、损失)。
	陈震元等人(2023)	2023	弱监督目标检测	×	×	按“核心网络架构改进方向”分为3类,分别优化候选框生成器、新增分割分支、优化自训练检测网络。
	袁翔等人(2023)	2023	遥感影像小目标检测	√	√	按遥感影像类型分为光学、SAR、红外三类小目标检测算法。
	石争浩等人(2023)	2023	航空遥感场景下的目标检测	×	√	以“框架+场景改进”双维度,聚焦大尺寸、多方向、尺度与小目标问题。
	潘晓英等人(2023)	2023	通用小目标检测	×	√	以“技术策略+问题导向”分为8类,涵盖数据增强、特征融合、注意力等主流技术,同时补充场景化与主流检测器改进。
	冷佳旭等人(2023)	2023	无人机视角下的目标检测	√	√	按“解决的不均衡问题”分为5大类方法,涵盖数据增强、多尺度融合等主流策略。
	李昌财等人(2024)	2024	自动驾驶车载三维目标检测	√	×	按“传感器类型与数量”分为5类,每类再按“数据预处理/模型架构”细分。
跟踪	李玺等人(2019)	2019	单目标跟踪综述	×	×	从“网络结构、网络功能、网络训练”三维度分类,补充4类融合算法。
	Marvasti-Zadeh等人(2023)	2021	通用视觉目标跟踪	×	√	从9个维度划分,既覆盖技术本质(如CNN/SNN),也贴合任务需求(如航拍/长期跟踪)。
	李成龙等人(2023)	2023	多模态视觉跟踪	√	√	以“信息融合方式”为核心,重点将RGB-T分为结合式/判别式融合两大类细分,其他按“融合阶段/数据处理方式”补充分类。
	Javed等人(2023)	2023	DCF与Siamese网络两大范式的视觉目标跟踪	×	√	以“范式划分+技术原理”为核心,DCF按“特征通道+优化策略”细分,Siamese按“训练流程+功能改进”细分,同时补充跨范式的骨干网络、目标状态估计等子模块。
	Jiao等人(2023)	2023	通用视觉目标跟踪	×	√	梳理单/多目标跟踪及短/长跟踪中的技术进展,分为网络架构、利用方式、跟踪挑战改进、高级主题四类。
	高桃峰等人(2025)	2025	卫星视频单目标跟踪	×	×	按“跟踪策略(生成式/判别式)划分,判别式进一步按“相关滤波、深度学习(CNN/Transformer)”细分。
检测与跟踪	本文	2025	夜间无人机目标检测与跟踪	√	√	以“夜间场景+无人机视角”为核心,将检测分4类(增强/域适应/多模态/轻量化),跟踪分5类(先增强后跟踪/域自适应/提示学习/课程学习/RGB-T跟踪)。

绍典型的夜间与全天候无人机检测与跟踪数据集,以及构建的夜间无人机车辆检测数据集DroneVehicle-Night,并简要说明检测与跟踪的评价指标;5)夜间无人机航拍图像检测与跟踪性能评价,

基于DroneVehicle-Night开展针对性的评价分析，并在VisDrone2019-DET(Zhu等，2022)中筛选昼夜样本，对比测试通用算法在夜间场景下的适应性，同时汇总夜间无人机跟踪四类范式在内的20种算法在UAVDark135(Li等，2023)和NAT2021(Ye等，2022)数据集上的性能表现；6)总结和展望，总结目前夜间无人机航拍图像目标检测与跟踪的趋势和展望未来可能的发展方向。

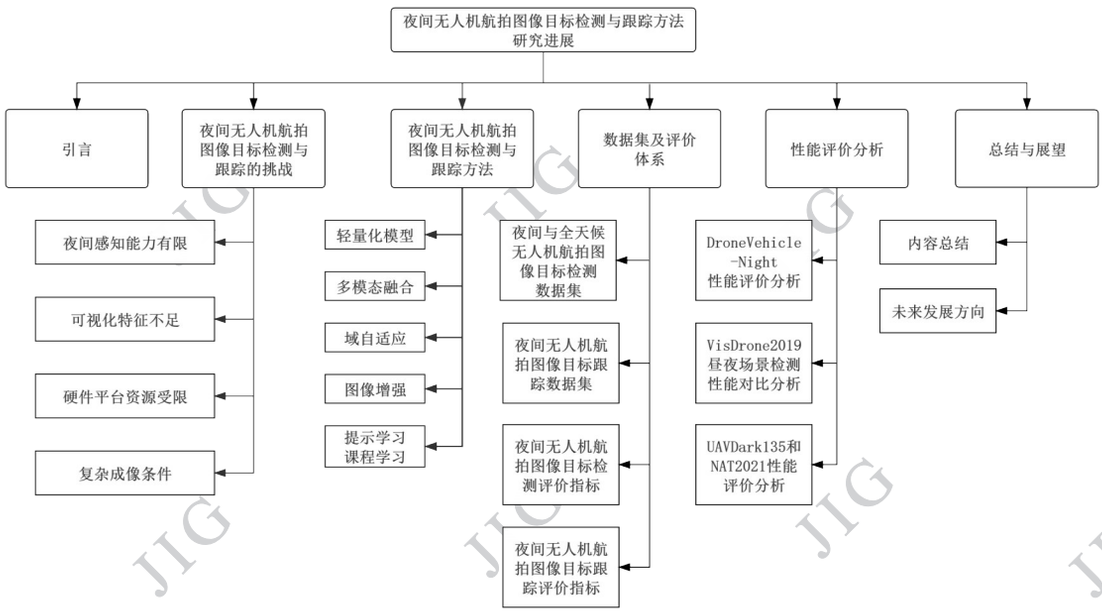


图1　本文结构框图

Fig. 1　Overall framework of this review

1　夜间无人机航拍图像目标检测与跟踪的挑战

无人机作为空中机动平台，能够通过多视角、多尺度的观测实现高效感知。这些特点不仅拓展了检测与跟踪的应用边界，同时也带来了更大的技术挑战。例如，频繁视角切换往往会导致感知结果的不稳定；再加之载荷受限、能耗约束以及外界环境干扰，无人机视觉系统在复杂条件下的稳健性面临严峻考验。尤其在夜间低照度环境中，这些问题被进一步放大，成为制约无人机可靠感知与任务执行的关键瓶颈。具体挑战描述如下(Gong等，2011)，如图2所示。

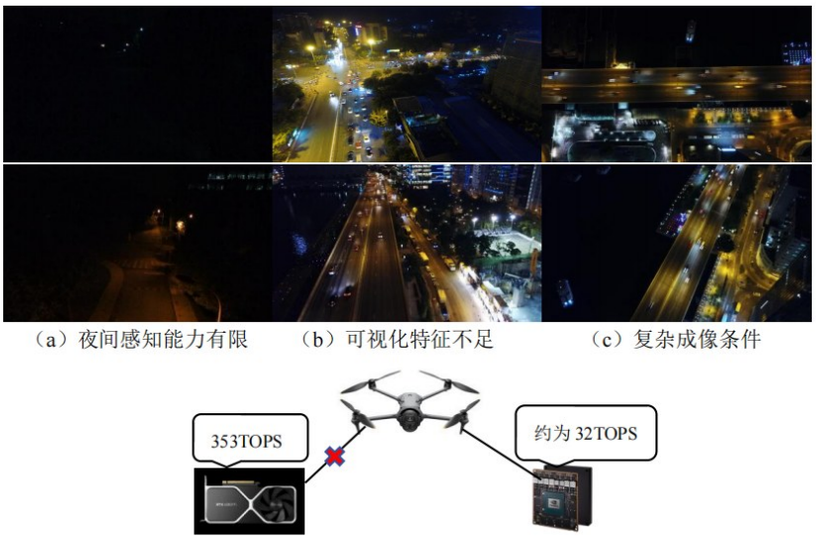
1)夜间感知能力有限。自然光照度不足，目标与背景的对比度显著降低，容易导致图像细节丢失与目标特征模糊，从而使无人机搭载的可见光相机难以获取高质量图像。在感知层面，即使引入具备夜视功能的相机，可在一定程度上缓解光照不足的问题，但由于不同物体的热辐射特性差异较大，在复杂环境中，地表、植被及建筑等背景往往会产生相似的热辐射分布，缺乏显著区分度，从而进一步干扰目标的有效检测与稳定跟踪。

2)可视化特征不足。夜间环境中的各种光源(如路灯、车灯、建筑物灯光等)往往形成复杂的背景光，易对目标产生干扰，导致目标检测出现误判。同时，夜间的阴影区域分布广泛，造成目标的外观与轮廓特征的显著变形，从而增加特征提取和匹配的难度。此外，昼夜场景间的特征分布差异(跨域差异)，导致当前算法对夜间场景的泛化能力受限，且短期内难以收集并标注足够多的夜间低光照数据集用于模型训练，这极大地阻碍了相关无人机应用在夜间的推广。

3)复杂成像条件。在飞行过程中获取图像时，地面目标(如行人、车辆等)在图像中占比极小，显著增加了小目标检测的难度。同时，由于无人机在执行任务过程中姿态多变，其拍摄的图像常具有不同的俯仰角和视角，导致同一目标在不同帧中的外观

特征和尺度存在显著差异，从而进一步加大了检测与跟踪的复杂性。此外，图像中目标的空间分布往往较为复杂，常呈现密集、稀疏或成簇分布。更为关键的是，无人机与目标之间的相对运动在捕捉高速运动目标时极易引发运动模糊，造成图像质量下降。

4）平台计算资源受限。受限于载重、体积、功耗和散热条件，无人机通常仅能搭载轻量化的机载处理器。以NVIDIA Jetson AGX Xavier为例，其最大算力约为32TOPS（tera operations per second），仅约为消费级GPU（central processing unit）（如GeForce RTX 4060 Ti，算力可达353TOPS）的1/11。尽管部分深度学习检测与跟踪算法在高性能GPU（graphics processing unit）上可实现50－250帧/秒的推理速度，但在资源受限的无人机平台上，多数算法仍难以稳定满足25－30帧/秒的实时性要求。检测任务在部分场景中可以通过图像回传至地面服务器进行处理，从而对实时性要求相对宽松，但若需机载执行则必须依赖轻量化设计以适应算力约束；相比之下，跟踪任务因其本质上要求对连续帧目标进行实时关联，实时性无法妥协，因此算法设计普遍强调轻量化。由此可见，在有限算力条件下实现高精度与低延迟的检测与跟踪，是无人机夜间感知研究的关键挑战之一。



（a）夜间感知能力有限　（b）可视化特征不足　（c）复杂成像条件

（d）平台计算资源受限

（(a) limited nighttime perception；(b) insufficient visual features；(c) complex imaging conditions；(d) limited computational resources on the platform）

图2　夜间无人机航拍图像目标检测与跟踪挑战

Fig. 2　Challenges in object detection and tracking from nighttime UAV aerial imagery

2　夜间无人机航拍图像目标检测与跟踪方法

2.1　夜间无人机航拍图像目标检测方法

2012年，AlexNet（Krizhevsky等，2017）在ImageNet竞赛中取得突破性成绩，使卷积神经网络（convolutional neural network，CNN）受到广泛关注，并推动目标检测领域进入快速发展阶段。主流的深度学习目标检测算法主要分为两类：两阶段检测（two-stage detection）算法与单阶段检测（single-stage detection）算法。两阶段检测算法首先生成候选区域（region proposal），然后利用CNN提取特征并完成分类与边界框回归（bounding box regression）。典型代表包括R-CNN（region-based convolutional neural networks）（Girshick等，2014）、Fast R-CNN（fast region-based convolutional network）（Girshick等，2015）、Faster R-CNN（faster regions with convolutional neural networks）（Ren等，2017）、Cascade R-CNN（Cai等，2018）以及Mask R-CNN（mask region-based convolutional neural network）（He等，2017）等。单阶段检测算法则采用端到端回归框架，直接在整幅图

像上预测目标类别与位置，在速度与精度之间取得较好平衡，代表方法有SSD（single shot multibox detector）（Liu等，2016）与YOLO（you only look once）系列（Redmon等，2016；Redmon等，2017；Redmon等，2018）等。

近年来，目标检测在边界框预测机制上取得一定进展，可大致分为基于锚点（anchor-based）和无锚点（anchor-free）两类。基于锚点的方法通过预设锚框进行回归调整，对多尺度目标检测有效，但需人工设计参数。无锚点方法如FCOS（fully convolutional one-stage）（Tian等，2019）直接预测目标中心和边界，简化流程并提高效率。同时，基于Transformer的方法也逐渐兴起。DETR（detection Transformer）（Carion等，2020）实现端到端检测，避免了锚点和NMS（non-maximum suppression）等人工设计，但计算量大、收敛慢。为此，后续提出了改进方法DINO（self-distillation with no labels）（Zhang等，2022）引入去噪训练策略和动态锚框机制，提升收敛速度和检测精度，而最新的DINOv3（Siméoni等，2025）在大规模预训练和特征密度上进一步优化，使Transformer检测在性能和效率上更加均衡。

尽管上述方法在多种应用场景中取得了显著成果，但在夜间无人机环境下依然受到第1章所述挑战的制约，对目标检测性能造成了显著限制。为应对这些问题，近年来研究者在1）夜间图像增强；2）域适应学习；3）多模态感知融合；4）轻量化模型等方向取得了重要进展，本节将围绕上述四个方向梳理代表性研究成果。

2.1.1　夜间图像增强

该方向的研究基于一个直接假设：若能把夜间低光图像恢复到“接近白天”的视觉质量，常规检测器即可无缝迁移。因此大量工作集中在像素层面的低光增强与去噪上。初期的深度学习低光增强方法LLNet（low-light network）（Lore等，2017）采用堆叠稀疏深度自动编码器（deep neural networks-stacked sparse denoising autoencoder，SSDA）对低光图像进行提亮和去噪。KinD（Zhang等，2019）将图像分解为照明分量和反射率分量，分别用于光照调节和去除噪声与光照不均，有效增强了夜间图像。Xu等人（2022）则提出信噪比（signal-to-noise ratio，SNR）感知增强网络，从空间变换的角度实现像素级增强。尽管这些方法在低光增强任务中表现出色，但由于未与目标检测任务深度耦合，且多依赖特定光照条件或静态场景，其在夜间无人机航拍图像目标检测中往往难以发挥理想作用，从而亟需面向检测的专用增强策略。

Yao等人（2025）在检测训练中引入增强过程的端到端优化，显著提升夜间交通灯识别精度。但是性能对输入尺寸的依赖性较强。Vinoth等人（2024）结合深度信息进行像素亮度与对比度优化，并通过TensorRT加速YOLOv8（Jocher等，2023）推理，在ExDark（exclusively dark）（Loh等，2019）数据集上实现精度与实时性的平衡。Feng等人（2024）设计了一种面向机器视觉的高效语义引导模块（efficient semantic-guided Machine Vision-oriented module，EMV），在潜在特征空间而非原始图像空间进行增强，移除了上采样步骤，满足实时检测需求。但依赖特定轻量Retinex（Land等，1977）分解编码器，因此模块对结构变化适应性较差。Shen等人（2025）采用共享编码器提取特征，并设计双解码器（分别用于低光增强和目标检测）进行联合训练，确保低光增强过程直接服务于检测性能的提升。Zhang等人（2024）基于Zero-DCE（zero-reference deep curve estimation）（Guo等，2020）框架的零参考学习（zero-shot learning，ZSL）思想，应用亮度增强曲线、融合原始图像与增强图像的特征，有效突出了目标的有效信息并抑制背景噪声。Zhao等人（2024）对YOLOv7（Wang等，2023）进行了改进，在数据层面引入了光照适应性增强技术，并在特征层面融入了边缘增强与注意力增强机制，从而实现了良好的夜间检测性能。Lashkov等人（2023）创新性地应用暗通道先验（dark channel prior，DCP）来提高夜间图像对比度，有效解决了低光条件下点光源光晕效应导致的过度增强问题。

在无人机视角下，Wang等人（2021）通过集成最优双曲正切曲线（hyperbolic tangent function，HTC）和改进的BM3D（block-matching and 3D filtering）（Dabov等，2007）算法，实现了良好的夜间无人机行人检测，但该方法在应对不同光照时需要自适应获取图像处理参数。Yuan等人（2023）通过分离无人机图像的亮度和色彩通道和对比度有限自适应直方图均衡（contrast limited adaptive histogram equalization，CLAHE）（Reza等，2004）来增强对比度，实现对比度增强和噪声抑制的平衡。Li等人（2024）针对夜

间无人机视觉感知存在统计特征偏差、伪影放大、训练数据缺失等三大挑战，提出零参考启发与恢复网络（zero-referenced enlightening and restoration network，ZERNet），通过估计夜间启发图进行像素级变换与自正则化恢复，不依赖成对训练数据即实现对结构细节和语义信息的保留与增强。Yue等人（2024）受IA-YOLO（image adaptive YOLO）（Liu等，2022）启发提出了夜间无人机检测驱动的暴露校正网络DEDet（detection-driven exposure-correction network for nighttime dronedet），根据无人机拍摄图像中非均匀照明调整低光图像的像素值，生成适合无人机检测的图像，并构建NightDrone数据集，填补了夜间无人机检测数据集的空白。Chen等人（2026）结合Retinex（Land等，1977）分解原理与差分注意力机制（differential attention mechanism，DAM）（Ye等，2025），通过两个独立的Softmax注意力图计算差分注意力分数，抑制了噪声传播并聚焦图像暗区域。

尽管图像增强能显著改善视觉质量，在检测性能方面显示出一定的改进，但它们存在两个问题：1）增强过程可能引入伪影/分布偏移，影响下游检测器；2）计算复杂度较高，难以适配资源受限的无人机平台；3）当可见光信息本身在语义级别缺失时，单纯的像素级恢复难以根本解决问题。

2.1.2 域适应学习

面对增强方法的局限，研究将关注点从像素恢复转移到“特征/分布对齐”上，即通过域自适应（domain adaptation，DA）减少昼夜域差，从而让检测器在源域（白天）学习到的语义能更好地迁移到夜间。

Huang等人（2022）首先通过无监督学习特征投影来增强提取的特征，然后通过概率模型动态选择更优特征。针对现有数据集往往倾向于过度拟合特定的域特征，Schutera等人（2021）通过无监督图像转换将夜间图像转换为白天风格，并直接使用适用日间场景的检测器。部分方法遇到了教师模型产生不准确标签等问题，Kennerley等人（2023）提出了两阶段一致性训练网络2PCNet（two-phase consistency network），在第一阶段将高置信度预测框与学生RPN（region proposal networks）提出的区域合并，第二阶段教师使用合并后的建议框生成新的标签。解决了昼夜无监督域自适应（unsupervised domain adaptation，UDA）目标检测中伪标签误差传播、学生网络偏向白天域和小目标检测精度低等痛点。Du等人（2024）提出了一种零样本昼夜域适应方法，通过学习光照不变的反射率特征以及进行两次序贯图像分解，使得检测器即使在没有真实夜间数据的情况下仍能保持较高性能。Zhang（2025）等人设计了一种可学习的文本提示机制，旨在捕捉目标域的语言先验以及检测任务所需的实例级知识。同时，生成伪目标域视觉特征，有效缓解了零样本域适应中的域偏差（domain shift）和检测偏差问题。这些研究为解决夜间目标检测中的跨域泛化（domain generalization）难题提供了新的思路。

在无人机视角下，域自适应研究多基于自然图像，针对夜间等不利条件下无人机图像的研究仍处于起步阶段。Wang等人（2024）提出了一种频域解耦（frequency domain disentanglement）方法，以提升跨域泛化能力。通过两个可学习滤波器分别提取域不变特征和域特定特征，并设计了图像级和实例级的对比损失，引导滤波器更有效地进行频域解耦，从而提高检测性能。Chen等人（2025）针对伪标签（pseudo-labeling）有限或复杂导致的域偏差问题，提出方差反馈平滑阈值策略，通过合并方差惩罚项动态调整选择阈值，显著提升伪标签质量并挖掘潜在有效标签。但该方法泛化能力有限，且在极低空无人机视角下性能显著下降。

域适应方法有效缩小了昼夜分布差，然而，当光照极度退化时，仅依赖特征对齐仍不足以保证检测鲁棒性。因此，研究者开始引入红外、激光雷达、事件流等额外感知模态，以实现跨模态互补与感知增强，推动了多模态感知融合的兴起。

2.1.3 多模态感知融合

研究者们通过融合激光雷达（light detection and ranging，LiDAR）、红外（infrared）、事件相机（event-based camera）与可见光相机（visible light camera）等多源信息，检测的鲁棒性得到了显著提升。

Dai等人（2023）利用NIR（near infrared）成像与LiDAR距离标注，实现了检测与距离估计的端到端联合优化，在夜间场景中获得了高精度、低漏检率的行人-距离检测结果，成为多模态融合在智能驾驶夜间感知中的代表性工作。Liang等人（2024）提出“RGB-LiDAR双稀疏融合”策略，以RGB提供语义补充、LiDAR提供空间补充，解决了跨模态维度不匹配问题，实现了高精度的夜间三维检测与定位。Li等

人(2025)通过融合单目视觉与LiDAR三维先验,引入低光增强模块,在暗光场景下自动触发图像增强。

事件相机通过编码异步事件,以对数刻度记录动态场景的稀疏时空表现,捕捉局部亮度的变化信息,具有低功耗、高动态范围以及微秒级时间分辨率等特性,使其在高速运动和复杂光照条件下具有天然优势。Li等人(2023)利用双视觉流的时序信息提升检测性能,通过异步注意力融合模块消除单模态退化、突破帧采样率的局限,有效应对运动模糊与低光挑战。Liu等人(2024)通过亮度分布估计信息损失,自适应提取事件特征进行补偿,但仍未充分利用事件相机的空间稀疏性与高时间分辨率,导致延迟较高、帧率偏低。Fan等人(2025)以尖峰神经网络(spiking neural network,SNN)为核心,提出的LIMF(leaky integrate-and-multi-fire)神经元解决了因"单尖峰输出"限制无法有效编码强刺激或微弱信号(如夜间低光下的光照变化)的问题。Jiang等人(2025)提出的可学习自适应选择与融合模块(learnable adaptive selection and fusion module,ASFM)在通道维度中进行特征排序和融合,实现高效的多模态融合。

结合可见光与红外成像能够有效利用模态互补性,增强算法在低光与复杂环境下的鲁棒性,构建高效的RGB-IR跨模态特征融合机制,是提升检测性能的核心问题。早期研究聚焦于特征对齐与共享。Liu等人(2022)提出跨模态特征学习模块,通过拆分与聚合策略显式区分共享与模态特定特征,增强跨语义级别的特征表达。Mazhar等人(2021)针对光照剧变和传感器非对称退化问题,设计动态估计传感器可靠性的机制,但在小目标场景下表现有限且计算复杂度较高。

随着研究深入,融合策略逐渐向更高层次的语义建模与动态交互转变。Xie等人(2022)通过局部与非局部特征的联合学习强化了跨模态细粒度关联,但仍忽略部分特征间依赖。Cao等人(2023)引入通道切换与空间注意力机制,对多模态特征进行自适应筛选与空间增强,有效缓解了夜间信息缺失问题。Sun等人(2024)提出以模态特异性为核心的重建思路,使模型在融合前先学会区分、保留并强化模态的独立信息。但上述方法均忽视模态间的冗余背景特征。Li等人(2024)引入狄利克雷分布(dirichlet distribution)建模各模态的类别不确定性和证据理论(dempster-shafer theory)动态整合彩色与热成像信息,使模型在噪声或模态缺失情况下仍保持更高的可靠性与鲁棒性。CALNet(He等,2023)通过交叉注意力计算跨模态相似度,并与k个最相似的模态进行交互,以降低跨模态异质性。Shen等人(2024)提出了一种查询引导的交叉注意力机制增强特征可辨性,在共享参数的前提下持续强化跨模态与模态内表征。Guo等人(2024)引入多光谱可变形交叉注意力,在多语义层自适应挖掘细粒度互补信息,解决模态错位问题。Berjawi等人(2025)进一步提出滤波多模态跨交叉注意力融合(filtered multi-modal cross attention fusion,FMCAF),结合频域滤波以抑制冗余频谱特征,兼顾局部与全局注意力,增强了模型泛化性与可移植性。但上述方法多依赖交叉注意力机制,堆叠的注意力带来了计算负担。

后续研究方向多聚焦于利用上下文信息和局部互补特征。Jin等人(2025)利用视觉识别中图像区域的内在关联性,在融合过程中引入全局语义约束以动态增强局部特征质量。Yang等人(2025)设计模态专属分支结构,使网络能够学习特定于模态的特征表示,同时捕获跨模态互补信息,在局部、全局与通道维度实现协同融合,该方法在各种光照条件下均展现出优越的检测性能。Hu等人(2025)利用结构与边缘信息强化跨模态互补,解决多光谱感知在动态光照下融合失衡的问题。Shen等人(2025)通过引入状态空间模型(state space model,SSM)的双路径特征融合机制,提升了小目标与低光场景的检测性能和推理速度。接着,Shen等人(2025)首次将差分反馈放大器引入多光谱融合,放大互补特征,抑制共享的背景噪声,实现了噪声与特征的动态平衡。Yang等人(2025)提出RGB-IR双编码器与三分支扩散模型(diffusion model),通过分阶段训练解决了夜间"RGB弱、IR强"的模态互补难题,设计的自适应特征增强器以适应不同场景特性。

在无人机视角下,Li等人(2023)提出跨模态知识蒸馏(knowledge distillation)框架,利用红外检测器的知识辅助可见光检测器优化,测试阶段仅部署可见光检测器,解决了多模态检测器无人机部署难题与传统知识蒸馏跨模态场景的局限性。Wang等人(2022)针对模态间冗余信息及人工照明分类(仅为白天和黑夜)的不足,提出基于直方图统计的精细照明分类。上述方法均未有效解决特征间的空间位置错位问题,缺乏对全局上下文信息的利用。因此,

Xie等人(2023)提出跨模态局部标定与全局上下文建模网络(cross-modal local calibration and global context modeling network,CLGNet),在多种照明条件下推断目标的存在与位置,并可灵活集成到不同检测器中。Yuan等人(2024)设计了跨模态交叉注意机制和自适应特征采样策略,在DroneVehicle(Sun等,2022)的夜间场景中实现了较优性能,但泛化能力仍有局限。Liu等人(2025)利用全局频域特征弥补局部空间信息缺失,有效缓解了夜间无人机视觉退化与尺度变化带来的检测难题。吴光宇等人(2025)受人类的选择注意力机制启发,提出了仿神经形态协同选择性注意(neuromorphic collaborative selective attention,NOSA)框架,通过早期关注(early focus)和后期关注(late focus)分别在数据域和特征域筛选关键信息,实现了跨模态感知的高效聚焦。

为进一步提升在夜间无人机视角下复杂环境的检测精度,Zhao等人(2025)借助Retinex(Land等,1977)理论,利用反射率的照度不变性,实现了不同照明强度和分布下的多模态特征的稳定对齐,有效解决了位置错位和特征退化问题。Wang等人(2024)构建了照明感知模块与特征加权机制,将不确定性量化为权重,引导网络关注于主导模态,以激励网络朝着有利于最佳目标检测的方向学习。Shang等人(2025)提出了跨模态跨域低光检测器CCLDet(cross-modality and cross-domain low-light detector),可根据目标光照自适应调整模态权重,并通过可见性损失将位置偏差映射为物体区域各点的光照强度差异,可提供更多可用于精确定位的信息。为兼顾检测精度与模型效率,Sun等人(2024)受低秩适应(low-rank adaptation,LoRA)微调方法(Hu等,2021)的启发,提出的低秩增强技术和动态光感知掩码模态能无偏移地提取多模态特征,以频域视角过滤幅度噪声、融合跨模态特征,但这种掩码方式忽略了目标的边缘细节信息,且过度依赖于IR模态。对此,刘奎等人(2025)设计超像素照明感知掩膜(super-pixel dynamic illumination- aware mask,SDIM)模块和低照度图像增强模块(low illumination image enhance,LLIE)分别缓解边缘特征丢失与可见光语义退化问题,通过多尺度交叉注意力融合机制解决跨模态特征异构性融合冲突。Kim等人(2024)引入环境上下文理解机制,利用多模态大语言模型(large language model,LLM)提取天气、能见度和光照等环境语义特征,并基于输入图像自适应选择特征并融入检测框架,使模型具备环境感知能力。

综上,多模态感知融合能提升夜间检测的稳健性,但通常以增加计算与采集复杂度为代价,限制了其在资源受限的无人机平台上的部署。

2.1.4 轻量化模型

在多模态融合提升检测性能的同时,研究者也意识到算法的复杂度与能耗问题成为制约无人机机载部署的关键瓶颈。因此,轻量化模型的设计成为夜间无人机检测研究的又一重要发展方向。

Liu等人(2024)将多头注意力机制(multi-head self-attention,MHA)中的Softmax替换为轻量化SiLU(sigmoid linear unit)激活函数,避免了复杂的矩阵计算。Wang等人(2024)提出红外与可见光融合网络,并通过轻量级的加法操作整合多源信息,减少计算开销。Xue等人(2024)从消除特征融合过程中的层间干扰出发,采用DSConv(depthwise separable convolution)和TensorRT量化技术,实现了机载边缘计算平台24FPS(frame per second)的处理速度。Zhong等人(2025)基于YOLOv11s(Jocher等,2024)提出PS-YOLO,通过引入Faster_C3k2简化骨干网络,删除对小物体效果不佳的C2PSA(cross stage partial with pyramid squeeze attention)模块,并在检测头中加入共享卷积和NWD(normalized wasserstein distance)损失函数,实现了在夜间密集小目标场景下的高精度和高推理速度。Fan等人(2025)基于YOLOv8(Jocher等,2023)提出了LUD-YOLO,利用特征的稀疏表示和自注意力机制减少了模型参数,并通过剪枝进一步降低网络复杂度。冯琪涵等人(2025)通过对前景区域进行超分辨率处理和引入分解式多维自注意力,有效减少了背景计算量和冗余特征交互,但对于激活图质量有较高依赖。Wang等(2025)提出的MINIAOD采用Ghost卷积构建轻量骨干网络,并集成GSConv(ghost shuffle convolution)模块构建特征增强网络。

综上所述,夜间无人机航拍图像目标检测方法的研究经历了从图像增强到特征域对齐,再到多模态感知融合与轻量化模型设计的持续演进。早期工作主要依赖像素级亮度提升与去噪技术以缓解低光退化,但在复杂光照与动态场景下鲁棒性有限。随后,域适应学习通过特征与分布对齐有效缩小了昼夜差距,却在极端低照条件下仍受限于可见光信息

的不足。为此，研究者引入红外、激光雷达、事件流等模态，推动了跨模态融合与协同感知的发展，大幅提升了夜间检测的稳定性与泛化能力。与此同时，轻量化网络结构与高效推理策略的兴起，为机载部署提供了现实可行的解决方案。总体来看，夜间无人机目标检测的研究正逐步迈向从“增强感知”到“融合互补”再到“高效部署”的综合发展阶段，为夜间无人机目标跟踪奠定了坚实的技术基础。

2.2 夜间无人机航拍图像目标跟踪方法

目标检测与跟踪在无人机夜间视觉感知任务中并非孤立存在。检测通常负责在单帧图像中发现并定位目标，为跟踪提供初始状态和类别信息；跟踪则在连续帧中延续目标身份，保持轨迹连续性，并在一定程度上缓解夜间环境下单帧检测易出现的漏检与漂移问题。因此，检测结果往往决定了跟踪的起点与可靠性，而跟踪的时序约束又能在检测之间补充目标状态，使整体感知过程更加稳定。复杂夜间场景对检测的影响也同样影响跟踪任务的效果。

目标跟踪技术经历了从早期生成式/判别式方法、基于相关滤波（correlation filters，CF）的方法。生成式方法通过建立目标外观或运动模型来匹配候选位置，代表性方法有光流（kanade-lucas-tomasi，KLT）（Shi 等，1994）、均值漂移（meanshift）（Comaniciu 等，2000）、卡尔曼滤波（kalman filter）（Kalman 等，1960）和粒子滤波（particle filter）（Isard 等，1998）。判别式方法则利用分类器区分目标与背景，如支持向量机（SVM）（Avidan 等，2024；江少杰 等，2017）、随机森林（random forest）（Saffari 等，2009）和提升方法（Boosting）（Grabner 等，2006）。相关滤波（DCF）方法依靠快速傅里叶变换（fast fourier transform，FFT）实现高效相关计算，典型代表包括 MOSSE（minimum output sum of squared error）（Bolme 等，2010）、KCF（kernel correlation filter）（Henriques 等，2015）、DSST（discriminative scale space tracking）（Danelljan 等，2014）、SRDCF（spatially regularized correlation filters）（Danelljan 等，2015）和 BACF（background-aware correlation filters）（Galoogahi 等，2017）。总体而言，这些早期方法在光照稳定、目标外观变化有限的常规场景中能够兼顾实时性与一定精度，但在夜间无人机场景下，其鲁棒性与泛化能力均显著受限，易出现漂移或跟踪失败。

进入深度学习时代，以孪生网络（siamese network）（Chopra 等，2005）为代表的相似性学习架构、注意力机制与 Transformer（Vaswani 等，2017）架构，以及新兴的 Mamba（Gu 等，2024）架构等技术，进一步提升了跟踪在复杂场景下的表征与建模能力。然而，将这些先进算法直接迁移到夜间无人机场景仍面临多重挑战，对跟踪算法的鲁棒性、实时性与适应性提出了更高的要求。

本节将系统综述深度学习的夜间无人机航拍图像目标跟踪研究的主要范式，包括：1）先增强后跟踪；2）域自适应；3）视觉提示学习；4）课程学习；5）多模态融合。同时，在每一小节对方法的优势与局限作出总结与未来研究建议。

2.2.1 引入图像增强器范式

通过低光图像增强器（LLIE）提升输入图像质量，在不对跟踪器进行重新训练的前提下提升夜间场景的跟踪性能。

Ye 等人（2021）提出 DarkLighter，通过轻量级映射估计网络 ME-Net（map estimation network）联合估计光照图和噪声图，实现迭代增强，兼容 CNN 跟踪器，但与视觉跟踪任务的协作能力较弱。针对这些不足，Ye 等人（2022）进一步提出任务驱动的低光增强器 SCT（spatial-channel Transformer），在空间通道中引入注意力模块捕捉全局信息，并将前馈神经网络（feed-forward neural network，FFN）替换为 ResNet（residual network），以保留局部上下文信息；同时通过非线性曲线投影与噪声项联合建模，实现正常像素照度调整与噪声像素去噪。Li 等人（2021）将低光图像增强器（LLIE）嵌入相关滤波（CF）的跟踪框架，利用照明变化的目标感知掩码来过滤增强引入的噪声，结合全局适应因子实现亮度调整与颜色信息保留。王法胜等人（2023）引入自适应图像增强模块，不影响各颜色通道比例前提下进行增强，并嵌入高斯形状掩膜，从而在保证 CPU（central processing unit）端实时性的同时缓解光照变化、边界效应与样本污染带来的性能退化。

此外，部分研究关注提升人类视觉感知以辅助地面操作员的目标筛选。Fu 等人（2022）提出自适应增强器 HighlightNet，凸显在线目标选择与潜在目标，显著提升峰值信噪比（peak signal-to-noise ratio，PSNR）和结构相似性（structure similarity index measure，SSIM）。Huang 等人（2025）针对潜在图像特征挖掘不足，通过引入金字塔注意力模块（pyramid

attention module，PAM)(Zhao 等，2017)提高网络获取全局信息和捕捉通道间依赖关系的能力，结合曲线投影实现夜间图像的高质量恢复。

针对夜间场景照明物引起的光照分布不均，Yao 等人(2024)将光分布与图像内容分离，既抑制过亮区域又增强暗区细节，在UAVDark135(Li 等，2023)测试中，性能显著优于使用相同跟踪器SiamRPN++(Li 等，2019)下的多种SOTA低光图像增强器(如RUAS(Liu 等，2021)，LIME(Guo 等，2017)，EnlightenGAN(Jiang 等，2021)，LLVE(Zhang 等，2021)，DCE++(Li 等，2022)，KinD++(Zhang 等，2021)，DarkLightNet(Ye 等，2021)，HighlightNet(Fu 等，2022)和SCT(Ye 等，2022))。揭示了现有方案在光分布不均场景下易出现过曝与暗区模糊等问题，这些视觉失真严重影响跟踪性能。邵延华等人(2025)通过自适应增强模块与光照质量因子动态调节时间正则化，提升了光照剧变与不均条件下的跟踪稳定性。

Zhang 等人(2025)将Mamba(Gu 等，2024)引入夜间无人机跟踪，提出的低光图像增强器(LLIE)既实现全局增强又保留局部细节，受(He 等，2025)启发，其跨模态Mamba网络通过视觉-语言(vision-language)融合引入语义信息，缓解了夜间数据稀缺问题。与当前先进的视觉-语言的方法CiteTracker(Li 等，2023)相比，该方法GPU占用减少了50%，推理速度达到42帧/秒。

现有去噪模块设计相对简单，可能仅采用单一算法，导致低光增强过程中图像噪声也可能被放大。Wang 等人(2024)设计了一种即插即用的实时去噪器CGDenoiser(conditional generative denoiser)，通过学习真实噪声的复杂分布，自适应生成与输入图像相关的噪声并去除。Xu 等人(2024)提出基于解耦表示的跟踪方法NiDR，通过解耦特征减少由图像退化引起的有害特征对跟踪性能的影响。

总体来看，低光增强器范式已在弱光场景下显著改善输入可视性，但仍存在两大核心挑战：1)增强过程在提升视觉质量的同时不可避免地破坏了图像固有分布，影响特征在跟踪模型中的有效表示；2)增强器和跟踪器分别聚焦于图像恢复与目标定位，任务目标本质差异导致两者难以实现协同最优。

2.2.2 域自适应范式

域自适应(DA)旨在特征层面缩小域间差异。尽管域自适应在其他视觉任务中发展迅速，其在目标跟踪领域的研究仍相对薄弱。

Ye 等人(2022)首次将无监督域自适应(UDA)引入夜间无人机跟踪，提出了UDAT(unsupervised domain adaptation)方法。在源域(白天)和目标域(夜间)数据上进行对抗训练，实现域不变特征的学习，然而，该方法未充分利用图像的尺度与平移不变性，在场景快速切换或目标密集时易出现目标丢失；依赖显著性检测生成样本可能遗漏小目标或复杂背景目标；缺乏伪监督机制，性能提升受到限制。针对这些不足，Wei 等人(2024)基于UDAT提出了TransffCAR，集成Swin Transformer V2(Liu 等，2022)与坐标注意力(coordinate attention，CA)机制(Hou 等，2021)，提升了下游任务中图像的分辨率和感受野，在快速切换的场景中自适应锁定目标，并利用多尺度特征融合和通道加权提高了小目标检测能力。Chen 等人(2023)借助平均教师(mean teacher)(Tarvainen 等，2017)框架提供的伪监督(pseudo supervision)信号，提出一种新的无监督域适应框架，嵌入低光增强器以改进伪标签质量，进而促进后续基于一致性约束的学习。

近年来，扩散概率模型(diffusion probabilistic model，DPM)在重建低分辨率图像的目标信息方面展现出卓越性能(Dhariwal 等，2021；Song 等，2021)。传统的一步式域自适应范式在应对夜间无人机跟踪中常见的低分辨率挑战时，难以直接感知和提取低分辨率目标特征。Zuo 等人(2024)首次将扩散模型(diffusion model)引入该领域，提出域自适应框架DaDiff(domain-aware diffusion model)，设计的跟踪层能有效整合并对齐低分辨率目标特征的域感知信息，将扩散过程与跟踪任务相关联，实现夜间低分辨率目标特征与白天特征空间的对齐。

目标域的训练数据通常远少于源域，且缺乏场景多样性，这种不平衡的训练分布会导致域判别器的过拟合。Zhang 等人(2023)基于特征表示可解耦为域不变内容与域特定风格的假设(Huang 等，2018)。采用编码-解码结构实现跨域内容与风格解耦及特征重构，将源域的语义信息有效迁移至目标域，从而平衡两域训练数据分布，实现跨光照条件的自适应能力增强。Chouhan 等人(2024)结合静态风格传输合成成对图像与重建辅助域适应技术，无需外部增强模型和目标域伪标签，即可实现输入与特

征层级的双重适应，为无监督域适应在夜间视觉任务中的应用提供了有价值的参考。

分割一切模型（segment anything model，SAM）（Kirillov 等，2023）依托大规模数据训练，具备强大的泛化能力和零样本迁移性能，但其计算开销使其难以直接部署在资源受限的无人机平台上。为此，Fu 等人（2024）提出基于 SAM 的域自适应框架 SAM-DA（SAM-powered domain adaptation），首次应用于夜间无人机跟踪。该方法利用 SAM 生成高质量的夜间样本，缓解夜间标注成本高、样本质量不足等问题；仅在训练时使用 SAM，推理阶段采用轻量化模型部署于无人机平台，推理速度达到 32FPS。

上述的域自适应方法普遍忽视连续帧的时间上下文，而时序信息对无人机跟踪尤为关键。Fu 等人（2024）提出时序域自适应训练框架 TDA（temporal domain adaptation）引入时序上下文对齐与提示驱动的目标挖掘机制，保证连续帧特征一致性，并借助文本提示实现精准目标定位。针对无人机多视角动态特征未被充分利用以及效率与精度难以兼顾的问题，Li 等人（2025）提出一种动态特征自适应方法 DARTer（dynamic adaptive representation tracker），融合静态与动态模板，周期性更新动态模板以整合不同时间的目标视角信息，从而增强多视角特征表达。模型可根据输入特征自适应激活 Transformer 的不同层，大幅减少冗余计算，在提升 6.3% 精度的同时，实现 74 FPS 的实时性能。Wu 等人（2025）提出了首个同时适配雾天和夜间场景的域自适应无人机跟踪器 LVPTrack，通过师生网络（teacher-student model）蒸馏跨域层次语义、伪标签投票缓解噪声干扰，并以动态聚合提示提升目标外观适应性。在此基础上，Yao 等人（2025）进一步提出统一多域自适应跟踪框架 UMDATrack（unified multi-domain adaptive track），结合扩散模型（diffusion model）生成的少量高质量无标签样本、轻量化域适配与置信度对齐机制，在单一模型中实现对夜间、雾天、雨天等多种恶劣天气的高效鲁棒跟踪，解决了现有方法单域适配与样本生成低效等核心问题。

总体来看，夜间无人机跟踪领域的域自适应研究已由早期单一特征对齐，发展为涵盖模型、数据、标签与时序的多维协同优化框架，并正进一步迈向统一多域适配方向。然而，仍存在若干关键挑战亟待突破：首先，轻量化部署仍是难题，SAM 及 Transformer 类模型的高计算成本与无人机资源限制之间存在矛盾；其次，现有域自适应方法多依赖可见光图像，受夜间环境影响显著且大多仅针对单一复杂环境设计，泛化能力不足；需要生成大量目标域样本，忽视域间目标的内在关联。

2.2.3 视觉提示学习和课程学习范式

近年来，提示学习（prompt learning，PL）技术引起广泛关注，从自然语言处理（nature language processing，NLP）成功扩展至视觉领域（Jia 等，2022；Zhu 等，2023）。通常只需向预训练模型添加少量可训练参数，即可引导其学习适应新任务的有效提示。课程学习则通过难度递进的训练信号提升极低光与运动模糊等条件下的鲁棒性。

Zhu 等人（2024）将夜间无人机跟踪重新定义为快速学习问题。通过迭代强调/削弱黑暗线索，结合门控特征聚合（gated feature aggregation，GFA）机制高效融合暗光信息，有效弥补了白天跟踪器在夜间的特征提取缺陷。在提示调优（prompt tuning）阶段，模型仅通过调整 prompt 参数并冻结基础模型，即可实现日间到夜间的高效迁移。现有提示学习方法大多只依赖空间定位的监督信息，这导致生成的提示往往较为模糊。Zhu 等人（2025）受视觉仿生学启发，提出了定向核引导提示学习方法，通过深入挖掘目标的拓扑结构与细粒度特征，有效解决了夜间航拍跟踪中提示模糊和目标定位不精准的问题。然而，此类提示学习方法（尤其当日间跟踪器基于 Transformer 时）仍引入额外可训练参数。基于 CNN 的跟踪器虽相对轻量，但受限于局部感受野，难以充分建模长时依赖关系。

Wu 等人（2025）首次将 Mamba（Gu 等，2024）用于夜间无人机跟踪。为缓解夜间高质量数据稀缺问题，他们借鉴课程学习（curriculum learning，CL）（Bengio 等，2009），提出自适应课程学习策略：训练早期以日间数据为主，促使模型学习通用特征；随后动态提升夜间样本权重，使训练过程从“易”逐步过渡到“难”，从而获得更适应夜间场景的鲁棒表示。依托 Mamba 的高效架构，该方法的参数量仅为 Transformer 方法的 1/20 至 1/10，推理速度可达 75FPS，而 Transformer 模型通常不足 40FPS。

总体而言，视觉提示学习与课程学习范式为夜间无人机跟踪提供了高效且可迁移的解决思路，在缓解数据稀缺与提升部署效率方面展现出潜力。然

而，提示学习的跨场景泛化与Mamba（Gu 等，2024）架构在极端低光条件下的稳定性仍缺乏系统验证，且自适应课程学习易受参数设定影响，增加了部署调试难度。未来可重点研究面向低光动态场景的提示生成策略与鲁棒的课程调度机制，以提升长期跟踪的稳定性与适应性。

2.2.4　多模态融合范式

随着传感器迭代升级、算力提升，在机载算力及成本允许的条件下逐渐出现了一些多模态融合的无人机跟踪方法。常见的多模态组合包括RGB-D（可见光+深度）、RGB-E（可见光+事件相机）、RGB-T（可见光+热红外）。

其中，RGB-D无人机跟踪利用深度图提供的几何结构信息来辅助遮挡处理与尺度估计，代表性工作如Yang等人（2023）等人提出的EMT跟踪器，其通过全局池化和全连接层计算模态权重，在GPU上可实现百帧级实时性能，并能在Jetson NX等嵌入式平台上高效运行。Liang等人（2025）开发了基于浅层特征融合的RGB-D跟踪器，通过1×1卷积生成自适应权重矩阵，加权融合两种模态。实验表明，两者均能在低光照和无人机快速运动条件下表现鲁棒，显示深度信息在无人机目标跟踪中的潜力。

事件相机以微秒级时间分辨率记录亮度变化，具备高动态范围和低延迟的优势，特别适用于高速运动与极端光照场景。代表性研究如Wang等人（2024）提出了第一个高分辨率事件视觉跟踪数据集EventVOT，并设计了基于Transformer的多模态教师-学生蒸馏框架，实现了仅依赖事件信号的高速低延迟跟踪。Hamann等人（2025）首次将事件相机引入任意点跟踪（tracking any point，TAP）任务，提出基于特征对齐损失的改进TAP框架，实现了对运动引起的事件特征变化的鲁棒表征。然而，现有RGB-E跟踪方法依赖复杂的跨模态融合结构，导致计算开销大且难以有效利用目标的历史外观变化。Zhang等人（2024）将多目标跟踪理念整合到RGB-E的单目标跟踪，以有效利用外观和运动信息。Yang等人（2025）构建了首个完全基于尖峰范式的帧-事件融合框架，通过随机拼接与时空正则化有效缓解了位置偏移与跨模态不对称性，且功耗显著降低。Sun等人（2025）基于Mamba（Gu 等，2024）的长程建模与线性复杂度优势，提出的历史解码器能够捕获长期外观变化，在长/短期RGB-E基准上取得优异性能。尽管深度相机、事件相机为无人机目标跟踪提供了丰富的几何、时间或三维信息，但它们在体积、重量、能耗、传感性能及数据集可用性等方面存在共性限制。

相比之下，RGB-T模态以其设备轻量、数据完备及算法成熟度高的优势，成为目前夜间无人机目标跟踪的主流选择。大多数RGB-T的单目标跟踪算法以MDNet（multi-domain network）（Nam 等，2016）或RT-MDNet（real-time multi-domain convolutional neural network）（Jung 等，2018）为基础框架，通过引入多模态模块或权重机制来提升融合效果。早期研究多采用加权融合策略，如Li等人（2017）利用重建残差估计模态置信度，Lan等人（2018）基于分类得分实现自适应模态加权，以实现RGB与IR信息的动态融合。然而，当模态置信度估计不准确时，这类方法容易在光照突变或热信号饱和的场景下出现目标丢失。早期RGB-T融合方法虽实现了模态权重的自适应调整，但受限于浅层结构与手工特征表达，其鲁棒性与泛化性有限。

随着深度学习的发展，研究重点逐渐转向特征级融合与表示学习。Li等人（2017）提出加权稀疏表示正则化的图学习方法，通过模态权重自适应整合RGB和IR特征；Li等人（2018）进一步设计了双流卷积神经网络（two-stream convolutional networks），自适应选择最具判别力的特征图进行融合。这类方法显著提升了融合的表达能力，但仍难以应对复杂的动态干扰场景。Li等人（2019）提出多适配器架构MANet（multi-adapter convolutional network），联合学习模态共享、模态特定与实例感知特征；Zhang等人（2021）通过属性驱动的残差分支实现多挑战条件下的自适应特征聚合。随后，Xiao等人（2022）提出“挑战属性解耦”思想，将光照变化、遮挡、热交叉等典型挑战映射至独立融合分支中，有效降低了对大规模数据的依赖。Hui等人（2023）通过收集、分发目标和环境相关的上下文来桥接RGB和IR搜索区域之间的跨模态交互。深层多模态表征的引入使RGB-T融合具备了更强的可解释性与泛化能力，但模型复杂度随之上升，实时性能受到一定制约。

近年来，研究趋势进一步转向动态融合与轻量化设计。Wang等人（2023）根据输入的特征分布自适应生成卷积核，实现模态间动态通信，在快速运动与遮挡场景中表现出更强的时空鲁棒性。Tang等

人(2023)在决策阶段引入动态加权与线性模板更新机制，建立决策级融合范式，从像素与特征之外实现模态响应的自适应整合，使模型能在不同环境条件下灵活平衡精度与实时性。Hou等人(2024)引入轻量级自适应模块，通过高效微调完成特征转换，并在此基础上实现结构对称的多模态特征融合。Shao等人(2025)针对域迁移导致的性能下降，强调快速参数更新以减少计算成本并稳定在线适配。高栋等人(2025)通过交互编码器提取互补信息、重构解码器抑制模态噪声，并结合位置线索增强空间建模，在VTUAV(visible-thermal UAV)(Zhang 等，2022)与HiAL(high attitude UAV multi-modal tracking dataset)(肖云 等，2025)等无人机跟踪数据集中取得显著性能提升，说明重构式融合在动态条件下能够更好兼顾模态互补性与鲁棒性。Ding等人(2025)以模板为核心构建不确定性，借对比方式更新模板，有效解决多数方法存在仅生成动态模态权重，却难以量化模态及融合质量等问题。

大多数研究对时间信息的利用非常有限，同时也忽视了空间与时间信息之间的相关性。Wang等人(2024)整合时空信息与多模态信息，有效解决了传统RGB-T跟踪中时空割裂、跨模态交互不足、跟踪漂移等问题。Li等人(2025)利用动态邻接矩阵自适应融合邻近帧的多模态信息，并通过时序扩散机制将干扰视作噪声加以抑制，从而实现跨时间的动态优化。Feng等人(2025)整合双模态为单紧凑特征，通过单分支架构实现了高效时空-模态联合建模。Hu等人(2025)则突破“稀疏时间更新”局限，提出的时间状态生成器基于交叉Mamba(Gu 等，2024)架构，实现了帧到帧的时间信息连贯传递。上述两种方法均统一适配RGB-D/T/E多任务，泛化能力强，为多模态跟踪提供通用解决方案。

针对多模态数据稀缺与迁移困难问题，提示学习在RGB-T跟踪中迅速兴起。Yang等人(2022)在像素级层面引入提示学习，缩小多模态与RGB分布差距，实现无微调跨模态迁移。Zhu等人(2023)不改变基础模型的特征提取能力，仅通过少量可学习的特征提示模块，在特征层面实现了多模态信息的互补与分布对齐，这也是首次将特征层面的提示学习引入多模态跟踪。上述方法均预设主/辅模态，仅单向传递辅助模态的提示信息，导致鲁棒性降低。Cao等人(2024)进一步提出双向跨模态适配器，能够感知开放场景中主导模态的动态变化，实现了动态、互逆的跨模态特征提示传递，以极低参数量达到较好性能，尤其适用于数据稀缺、模态关系多变的视觉任务，但该方法仅支持RGB-T模态。Hong等人(2024)打破了多模态提示学习仅适配特定模态的局限，将语言、掩码、深度/热成像/事件流等均视为提示，实现了对6类跟踪任务的统一适配。Chan等人(2025)设计了三分支结构，同时创新性引入Sigma机制，借助Mamba(Gu 等，2024)实现三分支之间的深度特征共享与协同优化，有效抑制冗余信息并充分释放多模态互补性，但整体结构较为复杂，在实时性方面仍存在不足。Zhu等人(2025)提出的RGB-D-T跟踪器RDTTrack利用提示学习在预训练RGB跟踪器上无缝集成深度与热红外信息，通过正交投影约束实现三模态的互补协同，突破了双模态方法的性能上限，同时提出了高质量的包含跨三种模态同步帧的数据集RGBDT500。尽管这些方法实现了较好的跟踪性能，但普遍忽视了模态间的特征差异。Lu等人(2024)通过风格-内容解耦的知识蒸馏机制缓解特征差异，实现96.4FPS(frame per second)的高效推理。Hu等人(2025)提出了LRPD(low-rank prompting and distillation)框架，采用对称架构的低秩提示学习(low-rank prompting，LoRA-P)实现RGB-T的双向跨模态交互，解决了模态间适应性不足，并结合提示驱动的师生蒸馏，优化轻量学生模型的跟踪表现。Lu等人(2025)进一步利用Mamba(Gu 等，2024)的选择性扫描机制高效筛选差异特征的互补特征，实现了差异越显著，融合增益越明显的跟踪效果，尤其是在低光等极端场景。Tang等人(2025)将并行混合训练改为串行渐进整合任务，自然契合持续学习(continual learning，CL)的核心目标，同时揭示了模态差异与性能降解的关联。另外，构建了首个整合RGB-T、RGB-D和RGB-E三类任务的基准UniBench300数据集，通过对齐训练与测试范式，将三次独立推理简化为单次推理，提供了高效统一的评估平台。

此外，多光谱作为光谱层面的模态扩展且随着多光谱成像设备在无人机平台中的普及，多光谱视觉逐渐成为全天候感知的重要手段。Li等人(2023)有效利用了波段关联性，通过光谱自表达模型学习各波段的重要性，并用于构造伪彩影像、提取深度特征以及加权融合跟踪结果。Li等人(2024)基于不同

材料在不同波长的光下反射/吸收特性不同，充分挖掘材料信息，显著抑制了跟踪漂移。上述方法都基于RGB的跟踪器进行改造以处理多光谱图像，Qin等人(2025)提出UNTrack框架，通过统一编码光谱、时空信息，实现小目标和遮挡场景下的鲁棒跟踪，并构建了首个多光谱无人机单目标跟踪的数据集MUST(multispectral UAV single object tracking dataset)。Li等人(2025)提出旋转感知光谱-空间建模方案，将旋转角度纳入多目标跟踪，显著提升了动态旋转目标的稳定性且满足无人机实时跟踪需求，并发布首个无人机多光谱多目标跟踪数据集MMOT(multispectral multi-object tracking dataset)。

总体而言，现有方法多基于地面或固定视角设计，少数研究针对无人机场景系统优化，但部分研究在具有俯视特征的数据中已展现出良好性能，表明其在无人机视角下具备较强的可迁移性与应用潜力。值得注意的是，在夜间低照度或复杂光照条件下，多模态可有效弥补可见光退化造成的感知盲区，为无人机在弱光环境下实现稳健目标定位提供了新思路。然而，多模态夜间无人机跟踪仍面临：1)数据的采集与标注成本高，现有数据集规模有限，缺乏覆盖复杂夜间场景和模态缺失情况的标准基准；2)模态差异与融合效率的矛盾依然突出，当前轻量化模型虽提升了实时性，却易丢失跨模态互补信息；3)极端场景的鲁棒性与模型泛化性不足，在光照突变、遮挡或快速机动条件下仍存在性能退化问题。未来，随着边缘设备适配、轻量化传感器与跨模态智能融合技术(如Mamba架构与Adapter范式)的发展，多模态融合有望推动夜间无人机视觉从“感知增强”到“认知理解”的跃迁，为空中智能定位和自主决策提供更具鲁棒性的技术支撑。

3 数据集及评价体系

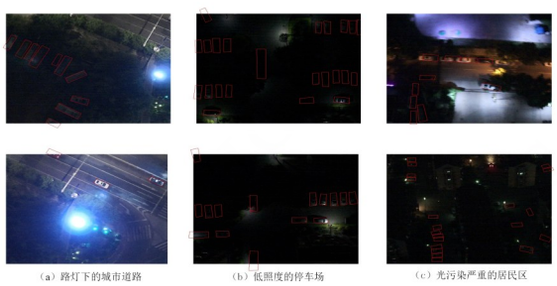
3.1 夜间与全天候无人机航拍图像目标检测数据集

夜间无人机目标检测的发展高度依赖高质量数据集与完善的评价体系。然而现有无人机视角数据集多集中于白天场景，夜间样本明显不足。为此，本节基于DroneVehicle(Sun等，2022)构建了夜间车辆检测数据集DroneVehicle-Night，并整理典型的全天候无人机航拍检测数据集，以便整体比较夜间数据与通用数据集的互补性。

3.1.1 DroneVehicle-Night介绍

在构建DroneVehicle-Night数据集时，本文针对第一章提出的夜间无人机检测挑战，在样本筛选与标注设计上进行了有针对性的优化。

首先，针对“夜间感知能力有限”和“可视化特征不足”的问题，从DroneVehicle数据集的可见光中筛选出低照度、人工光源照明、暗背景及非均匀光照等典型夜间场景样本，以覆盖真实夜间成像特性。图3展示了示例样本，如路灯照明下的城市道路(见图3a)、低照度的停车场(见图3b)以及光污染严重的居民区(见图3c)等.



((a)urban road under streetlights;(b)low-illumination parking lot;(c)residential area with severe light pollution)

图3 DroneVehicle-Night数据集样本示例

Fig. 3 Example samples from the DroneVehicle-Night dataset

其次，为应对“复杂成像条件”，数据集涵盖多种无人机飞行姿态与视角，目标在尺度、朝向与密度分布上具有高度多样性。为此，本文在标注层面引入了旋转边界框(oriented bounding box，OBB)格式，相较于传统的水平边界框(horizontal bounding box，HBB)更能精确描述倾斜视角下的车辆轮廓，并保留目标的空间结构信息。

总体而言，DroneVehicle-Night数据集包含10357对训练图像、868对验证图像和6013对测试图像，每对样本均含RGB与IR两种模态。数据集共标注276819个目标实例，覆盖五类目标：汽车、卡车、公共汽车、面包车与货车。场景涉及城市道路、居民区与停车场，充分体现了光照干扰的复杂性与空间分布的多样性。综上，DroneVehicle-Night数据集在样本筛选、模态设计与标注规范上均针对夜间无人机航拍图像检测的关键挑战进行了系统优化，可作为后续多模态融合感知算法的重要验证平台。

3.1.2　全天候无人机航拍图像目标检测数据集

除DroneVehicle-Night外，本节还综述了全天候或特定领域的无人机航拍图像目标检测数据集，并附相应的下载链接见表2，供后续研究使用。

1）Zhu等人（2022）构建了大规模无人机数据集VisDrone2019，由天津大学机器学习与数据挖掘的AISKYEYE团队采集。包含288个视频片段，由各种无人机拍摄的261908帧和10209张静态图像组成。其中62.4%的标注目标为面积小于32×32像素的小目标。其目标检测子集VisDrone-DET含6471张训练图像、548张验证图像、1610张测试图像及1580张挑战测试图像，类别涵盖人类与车辆等10类。

2）Du等人（2018）构建了无约束场景的无人机数据集UAVDT。专注于复杂场景，含77819张图像，标注了多达14种属性（如天气、飞行高度、车辆类别、遮挡程度等），为多维条件下的无人机航拍图像目标检测与跟踪提供了丰富素材。

3）Kraft等人（2021）构建了无人机环境垃圾检测数据集UAVVaste。包含772张图像，3718个标注框。仅标记为1类，即环境垃圾，填补了特定领域无人机检测数据的空白。

4）J Gasienica-Jozkowy等人（2021）构建了面向水域无人机搜救与监测的数据集AFO，包含3647张图像，119973个标注框，涵盖9个类别。并根据任务需求细分为三类标签方案：人类优先、大小物体平衡分类、全物体合并检测。

5）Sun等人（2022）构建了首个基于无人机的RGB-T跨模态车辆检测数据集DroneVehicle。共56878张图像（RGB与红外各占一半），包含五类车辆。原图比例为840×712像素，去除四周各100像素白边后为640×512，适用于低能见度与夜间检测任务。

6）Suo等人（2023）构建了首个高空无人机红外热成像数据集HIT-UAV，由哈尔滨工业大学研究团队收集。涵盖不同场景（学校、停车场、道路、操场），覆盖5类对象标签，共计2898张红外热图像。记录了飞行高度（60米-130米）、相机视角（30度-90度）、日光强度（白天或黑夜）等信息，并提供定向与标准两类边界框。该数据集在1）飞行高度与目标检测精度对无人机的关系；2）相机视角对无人机目标检测的影响；3）无人机在夜间搜索和救援任务的可行性等研究上均展现出巨大潜力。

7）Zhang等人（2025）构建了基于无人机的红外车辆检测数据集OSIV，由39583张红外图像和617370个标注组成。该数据集包括与DroneVehicle相同的五个车辆类别，分为训练（25246张图像）、验证（1663张图像）和测试（12674张图像）集。横跨建筑、停车场、道路、树木覆盖、雾、沙漠、草原、雪景、丘陵等九个不同场景，包括白天和黑夜。

8）Dhrafani等人（2025）构建了应用于无人机空中感知的立体热成像深度感知数据集FIReStereo，由204594张立体热成像图像组成。涵盖多种环境条件，其中56%来自树木茂密的野外环境。84%的图像在白天采集，其余则是在夜间采集。

上述数据集中（5）~（8）引入或直接使用了红外成像数据，能在弱光条件下通过多模态融合带来一定补充作用，或通过红外实现夜间目标检测。但是，红外数据语义信息不如可见光丰富，还可能引入非目标热源的干扰和非热源目标的漏检；多模态融合需要增加无人机载荷从而带来额外的硬件成本，算法的整体计算量和复杂度也会提升从而增加机载算力要求。因此，夜间或低照度环境下的目标检测依然面临诸多挑战，亟需进一步研究与突破。

3.2　夜间无人机航拍图像目标跟踪数据集

大多数跟踪研究集中于光照充足条件下，常用数据集包括OTB100（Wu等，2013）、GOT-10K（Huang等，2021）、LaSOT（Fan等，2019），因而对夜间场景的关注相对较少。为系统反映夜间及无人机视角下的研究基础，本节整理并简要介绍了现有夜间跟踪基准，涵盖无人机采集的数据集及部分地面视角的低光基准。

1）UAVDark：分为UAVDark70（Li等，2021）和UAVDark135（Li等，2023）两个版本，由无人机在夜间拍摄的跟踪测试序列，分别包含70和135个手工标注的视频序列，覆盖了道路、街道、高速公路、街区等多种场景，包括大量对象，如行人、汽车、卡车、自行车和建筑。

2）DarkTrack2021（Ye等，2022）：是一个具有挑战性的夜间无人机跟踪基准测试，它包含110个具有挑战性的序列，总共超过100 K帧，所有序列都是在夜间场景以30帧/秒捕获的。

3）NAT2021（Ye等，2022）：是专为无监督域自适应夜间无人机航拍图像目标跟踪而设计的开创性

表2　无人机航拍图像目标检测典型数据集

Table 2　Typical UAV aerial image object detection datasets

名称	模态	年份	数量	特点	分辨率	下载地址	引用次数
UAVDT	RGB	2018	77819张	分为低中高三个飞行高度且包含日间、夜间、雾三种天气	1080×540	https://sites.google.com/view/grli-uavdt	1003
VisDrone	RGB	2021	10209张	使用不同型号的无人机、不同场景、各种天气和照明条件拍摄。但类不平衡。	960×540~2000×1500	https://github.com/VisDrone/VisDrone-Dataset	872
UAVVaste	RGB	2021	772张	近地拍摄，所包含图像，目标框较小，适用于小目标检测。	3840×2160	https://github.com/PUTvision/UAVVaste	131
AFO	RGB	2021	3647张	适用于水上搜救任务，环境多样且小目标为主。	1280×720-3840×2160	https://www.kaggle.com/datasets/jang-sienicajzkowy/afo-aerial-dataset-of-floating-objects	96
DroneVehicle	RGB+IR	2022	56878张	RGB和红外图像各占一半，专注于车辆检测和计数。	840×712	https://github.com/VisDrone/DroneVehicle	353
HIT-UAV	IR	2023	2898张	首个无人机高空红外热数据集，可进行多项研究活动。	640×512	https://gitcode.com/open-source-toolkit/f9a26/tree/main	121
OSIV	IR	2025	39583张	第一个基于场景划分的红外无人机视角车辆检测数据集	704×704	https://github.com/rslab1111/OSIV	1
Firestereo	IR	2025	204594张	包含立体热成像图像、激光雷达数据、惯性测量单元数据以及在城市和森林环境中、不同条件下采集的真实深度图。	640×512	https://github.com/firestereo/firestereo	5

基准测试。包含180个跟踪序列的测试集和1400个跟踪序列的训练集。覆盖了道路、城市景观和校园等场景，包括各种对象，如汽车、卡车、人、团体和摩托等。为了提供长时跟踪性能评估，进一步构建了长时跟踪子集NAT2021-L-test，由23个序列，54K帧组成。

4）WebUAV-3M（Zhang等，2023）：是目前最大的多模态无人机跟踪的数据集。包含4500个视频中的超过330万帧，并提供223个高度多样化的目标类别。此外，涵盖了多任务（夜间跟踪、对抗性示例、多模态跟踪、数据不平衡）。

5）NUT-L（Fu等，2024）：是一个综合的长时夜间无人机跟踪基准数据集。由NAT2021-test和UAVDark135中的长时跟踪视频组合而成，包含42个序列和95274张图像。

6）LMOT（Wang等，2024）：创新性地开发了一种双摄像头系统，可以同时捕捉光线充足和夜间低光的视频帧。在空间和时间维度上高度对齐，具有两个关键优势：1）能够在光线充足的视频帧上进行标注，从而获得高质量的标注；2）光线充足的视频可以在训练阶段提供额外的监督信息，有利于提升黑暗场景中的性能。包含32个视频序列、超过35000帧和超过815000个边界框。

7）LLOT（Zhong等，2024）：包含269个视频序列，超过132000帧图像，它使用普通视角收集各种室内和室外低光场景，特别包含许多肉眼难以识别的物体，更接近实际具有挑战性的应用场景并详细标注了12个挑战属性。

8）NT-VOT211（Liu等，2024）：由211个不同的视频组成，超过211000个注释良好的帧，具有8个挑战属性。

表3给出了这些数据集的关键统计信息与常见夜间跟踪挑战，包括光照变化（illumination variation，IV）、尺度变化（scale variation，SV）、遮挡（occlusion，OCC）、形变（deformation，DEF）、运动模糊（motion blur，MB）、快速运动（fast motion，FM）、低分辨率

(low resolution, LR)、视角变化(viewpoint change, VC)、长宽比变化(aspect ratio change, ARC)、背景杂波(background clutter, BC)、相机运动(camera motion, CM)、完全遮挡(full occlusion, FO)、出视野(out-of-view, OV)、相似物(similar object, SOB)、低光照强度(low ambient intensity, LAI)、旋转(rotation, ROT)、部分遮挡(partial occlusion, PO)、运动变化(motion change, MOC)、小目标(small-scale goals, SSG)。

尽管日间跟踪算法在上述挑战条件下已有一定适应性，但在夜间尤其是无人机视角的极端低光和长时序条件下，现有数据集仍存在场景分布偏倚、模态匮乏(例如红外/热成像)与长序列样本不足等问题，制约了算法泛化性和稳健性的评估。

表3　夜间跟踪数据集简单对比

Table 3　A brief comparison of night tracking datasets

数据集	视频序列数	总帧数	数据划分	挑战属性数量	挑战属性名称	被引用量
UAVDark70/135	70/135	66k/125k	Test	5	LR,FM,IV,VC,OCC	57/71
DarkTrack2021	110	100k	Test	6	VC,FM,OCC,LR,LAI,OV	69
NAT2021	1603	470k	Train/Test	12	BC,ARC,CM,FM,OCC,FO,OV,SV,SOB,VC,IV,LAI	118
WebUAV-3M	4500	3.3M	Train/Val/Test	15	LR,PO,FO,OV,FM,CM,VC,ROT,DEF,BC,SV,ARC,IV,MB,CM	37
NUT-L	42	95k	Test	9	OCC,FM,VC,IV,LR,SOB,SV,LAI,FO	24
LMOT	32	35k	Train/Val/Test	5	OCC,OV,SOB,LAI,PO	7
LLOT	269	132k	Test	12	IV,SV,FM,OV,LAI,MB,ARC,LR,SOB,PO,ROT,CM	1
NT-VOT211	211	211k	Test	8	CM,DEF,FM,MB,SSG,SOB,OCC,OV	2

3.3　夜间无人机航拍图像目标检测评价指标

与MS COCO(Microsoft common objects in context)(Liu等，2014)数据集的评价体系类似，通常采用平均精度(average precision, AP)、平均精度均值(mean average precision, mAP)、精度率(precision, P)、召回率(recall, R)、模型参数(params)、每秒10亿次浮点运算(GFLOPs)、模型大小和帧率(FPS)等指标来评估模型性能。

GFLOPs用于衡量算法计算复杂度，数值越低表示推理所需的浮点运算量越小、运算速度越快；Params表示模型可训练参数数量，参数越少说明模型更轻量；模型大小指训练后生成的权重文件占用的存储空间。mAP的计算过程如下：

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

$$AP = \int_0^1 P(R)\,dR \tag{3}$$

$$mAP = \frac{1}{m}\sum_{i=1}^{m} AP_i \tag{4}$$

式中，TP(true positive)指正确预测的正样本，FP(false positive)指错误预测的正样本，FN(false negative)指预测错误的负样本。*AP*代表平均精度，即精度-召回率(precision-recall curve, PR)曲线下的面积。*AP*值越高，反映检测结果越好。m*AP*只有在指定了交并比(intersection over union, IoU)阈值时，该指标才有意义。其中m是类别数量，AP_i是第i个类别的*AP*值。在大多数研究中，主要使用m*AP*50和m*AP*50－95作为评价指标。m*AP*50表示IoU阈值为0.5时计算出的平均精度均值，该指标反映了模型在较为宽松的IoU阈值下的检测性能，通常用于评估模型的召回能力。m*AP*50－95表示IoU阈值在0.5~0.95、以0.05为步长的10个阈值下计算的平均精度均值。

3.4　夜间无人机航拍图像目标跟踪评价指标

夜间无人机单目标跟踪常用评价指标主要包括

精度(P)、归一化精度(normalized precision,NP)、成功率(success rate,SR)、成功率曲线下面积(area under curve,AUC)以及帧率(FPS)。

精度:基于中心位置误差(center location error,CLE),即预测中心与真实中心的欧氏距离。通常以CLE小于20像素的帧占比作为精度得分。

归一化精度:对中心误差按目标尺度进行归一化,并在阈值范围[0, 0.5]内绘制曲线并计算面积。该指标更适合目标尺度变化较大的场景。

成功率:基于预测框与真实框的IoU值,统计IoU高于阈值(如0.5)的帧占比。

成功率曲线下面积:对成功率曲线(success plot)积分得到的面积,可综合反映跟踪算法的准确性和稳定性,是多数数据集默认采用的核心指标。

帧率:衡量算法速度的关键指标,表示每秒处理的帧数。一般认为FPS≥25~30即满足实时性要求。

4 性能评价分析

4.1 DroneVehicle-Night性能评价分析

为评价现有方法在DroneVehicle-Night上的整体性能,本节选取现有部分方法在该数据集上进行性能评价分析。

4.1.1 实验设置

本节所有实验在单张配备48GB显存的NVIDIA RTX A6000显卡上进行。为确保对比充分,我们优先遵循各方法官方论文公开的训练设置。由于红外模态的目标标注更全面,因此使用红外模态的标签作为真值。实验基于MMRotate(Zhou等,2022)和MMDetection(Chen等,2019)进行,且检测器采用旋转边界框(OBB)检测头对车辆进行定位和分类。

4.1.2 评价结果分析

我们选取了部分具有竞争力的方法,包括4种单模态方法:FasterR-CNN(Ren等,2017)、RetinaNet(Lin等,2017)、YOLOv8s(Jocher等,2023)和S2A-Net(single-shot alignment network)(Han等,2022);以及7种包含夜间场景的多模态融合方法CFT(cross-modality fusion Transformer)(Fang等,2021)、CALNet(corss-modal conflict-aware learning network)(He等,2023)、C^{2}Former(calibrated and complementary Transformer)(Yuan等,2024)、ICAFusion(iterative cross-attention guided feature fusion)(Shen等,2024)、DAMSDet(dynamic adaptive multispectral detection Transformer)(Guo等,2024)、M^{2}D-LIF(mono-modality distillation method and the local illumination-aware fusion)(Zhao等,2025)、MS2Fusion(multispectral state-space feature fusion)(Shen等,2025)。表4汇总了不同方法与模态下按类的mAP50结果。

首先,在单模态对比中,红外(IR)模态普遍优于可见光(RGB)模态。以YOLOv8s为例,RGB模态下的mAP50为65.7%,而IR模态下为79.7%,约高出14%。这一差距清晰表明:在夜间航拍场景中,红外成像能够更稳定地刻画目标和背景差异,从而缓解低光、反光及噪声对特征提取的负面影响,使检测器更容易分辨目标轮廓和热特征,从而提升总体mAP。

其次,多模态融合方法在整体性能上明显优于单模态基线。CFT、ICAFusion、DAMSDet、M^{2}D-LIF和MS2Fusion的mAP50均在80%以上。以M^{2}D-LIF为代表的多模态融合策略,在若干难检测类别上(例如Truck、Freight-car、Van)取得了显著提升(Truck80.9%、Freight-car76.0%、Van68.9%),最终得到83.9%mAP50的总体结果,明显高于任一单模态结果。这说明在夜间条件下,RGB与IR的互补性可以通过合理的融合(如局部光照感知融合与蒸馏机制)得到充分利用,从而对小目标、遮挡或弱纹理目标提供更强的支持。

从整体上看,使用不同的骨干网络时,检测效果存在明显差异,具体表现为:以ResNet50(residual network 50 layers)(He等,2016)为骨干网络,无论是单模态还是多模态的FasterR-CNN、RetinaNet、S^{2}A-Net、C^{2}Former、DAMSDet,在各类别的检测精度以及整体的mAP50指标上,普遍低于基于CSPDarkNet(cross stage partial DarkNet)(Wang等,2020)系列骨干的模型。值得注意的是,CSPDarkNet的分层跨阶段结构在保持轻量化的同时,能够增强特征复用与跨尺度信息流动,这或许有助于在低光环境下提取更具判别性的多尺度特征。因此,可以推测:在夜间无人机检测任务中,选取兼顾特征表达能力与模型复杂度的骨干网络,有助于实现性能与部署效率的平衡。

表4　DroneVehicle-Night数据集上性能简单评估

Table 4　Simple performance evaluation on the DroneVehicle-Night dataset

方法	发表信息	主干网络	模态	Car	Truck	Freight-car	Bus	Van	mAP50
FasterR-CNN	TPAMI 2017	ResNet50	RGB	74.9	27.5	28.0	77.4	40.3	49.6
RetinaNet	ICCV 2017	ResNet50		72.5	12.2	23.4	49.4	31.0	37.7
S^2A-Net	TGRS 2021	ResNet50		72.2	21.1	23.8	78.1	36.8	46.4
YOLOv8s	Ultralytics 2023	CSPDarkNet53		88.6	45.7	52.5	89.6	52.3	65.7
FasterR-CNN	TPAMI 2017	ResNet50	IR	89.8	47.4	52.2	88.2	48.0	65.1
RetinaNet	ICCV 2017	ResNet50		89.1	18.2	35.3	70.3	32.9	49.2
S^2A-Net	TGRS 2021	ResNet50		89.4	41.8	56.6	88.9	45.8	64.5
YOLOv8	Ultralytics 2023	CSPDarkNet53		98.1	69.3	75.8	**96.5**	58.7	79.7
CFT	arXiv 2022	CSPDarknet53	RGB+IR	97.4	71.2	75.5	96.3	61.6	80.4
CALNet	ACM MM 2023	CSPDarkNet53		89.8	72.6	68.9	88.8	59.0	75.8
C^2Former	TGRS 2024	ResNet50		90.0	67.2	62.9	89.1	57.8	73.4
ICAFusion	PR 2024	CSPDarkNet53		98.1	77.2	**81.2**	96.1	64.0	83.3
DAMSDet	ECCV 2024	ResNet50		95.8	72.5	79.4	94.2	64.0	81.2
M^2D-LIF	ICCV 2025	CSPDarkNet53		97.6	**80.9**	76.0	96.1	**68.9**	**83.9**
MS2Fusion	Inf.Fusion 2025	CSPDarkNet53		**98.2**	75.4	79.2	96.2	65.2	82.8

注：加粗字体表示各列最优结果

4.2　VisDrone2019昼夜场景检测性能对比分析

现有无人机航拍图像目标检测研究大多在日间图像上展开，取得了较为理想的精度。然而，对于夜间场景，这些算法的实际表现仍然存在较大不确定性。为了验证这一问题，本节基于VisDrone2019-DET（Zhu等，2022）测试集设计了昼夜对比实验。与DroneVehicle-Night不同，本节的VisDrone2019数据集涵盖了更多的目标类型和更复杂的场景环境，但训练集中夜间样本仅占约18.3%，能够更全面地反映通用算法在不同光照条件下的适应性。

4.2.1　数据集昼夜情况划分

由于VisDrone2019-DET（Zhu等，2022）未提供夜间属性标注，本文通过人工筛选与亮度阈值过滤，从其测试集中选取了521张夜间图像及对应标注构建夜间基准，涵盖城市道路、停车场、路口等多种场景，并包含车辆、行人、非机动车等主要类别，兼顾路灯、无光源、车灯干扰等不同光照条件。

为便于对比，又从同一测试集中挑选了513张日间图像，使昼夜样本在场景类型和目标类别上保持一致，形成可对照的评价基准。

4.2.2　实验设置

本节参照VisDrone2019（Zhu等，2022）的划分方式，将数据集分为训练集、验证集和测试集，分别包含6471、548和1610张图像。硬件配置如表5所示。训练过程中采用随机尺寸缩放、翻转等常规数据增强，以缓解锚点类型不足带来的不平衡问题，并提高模型的鲁棒性与泛化能力。

4.2.3　昼夜测试对比分析

本节评估多种主流目标检测算法在昼夜两类典型光照条件下的性能。对比模型主要涵盖两大类：

1）单阶段算法，如YOLO系列的YOLOv5（Jocher等，2020）、YOLOv9（Wang等，2023）、YOLOv10（Wang等，2024）、YOLOv11（Jocher等，2024）、YOLOv13（Lei等，2025），且大多选择小型化版本，及其改进版本Gold-YOLO（Wang等，2023）和Hic-YOLOv5（Tang等，2024），

2）两阶段检测算法，包括QueryDet（Yang等，2022）和ClustDet（clustered detection）（Yang等，2019）。这些方法在实际无人机应用中具备较高的可移植性与通用性，因此能较全面反映当前检测算法在光照变化下的性能差异。

表5　硬件配置表

Table 5　Hardware Configuration Table

环境	参数
CPU	13th cen Intel(R)core(TM)i7-13700F
GPU	NVIDIA GeForce RTX 4080
显存	16GB
运行内存	32GB
操作系统	Ubuntu 20.04.6 LTS
语言	Python3.8.20
框架	Pytorch2.4.1
CUDA版本	CUDA12.1

表6展示了各算法在日间与夜间场景下的检测结果。从整体趋势来看，在光照充足的日间环境中，各模型表现较为理想，mAP50普遍分布在30%－45%的区间。其中，ClusDet（45.2%）在复杂多类别场景下表现最优，得益于针对航拍图像目标稀疏性设计的聚类特征融合机制，使模型能在复杂背景与多尺度目标下提取更具判别性的特征。YOLO系列中，YOLOv9和HIC-YOLOv5的mAP50均超过41%，值得注意的是，相较于YOLOv5-s，HIC-YOLOv5在mAP50上提高了6.5%，这主要归功于增加的小目标检测头与高分辨率特征图，使模型更适配无人机航拍图像并强化了空间域细节建模能力。

当场景由日间切换到夜间低光条件时，所有算法的检测性能均出现显著下降，更直观如图4所示。具体来看，9个对比模型的mAP50平均绝对下降9.1个百分点（相对下降约为23.1%），其中Recall下降更为明显，平均相对下降约24.1%，而Precision相对下降约11.0%，表明检测器在夜间更倾向于保守预测，导致漏检率上升。性能退化的根本原因在于夜间图像中纹理细节与局部梯度分布被弱化，浅层特征响应不稳定，使得依赖局部显著特征的模型难以维持检测精度。值得注意的是，两阶段检测算法在夜间的mAP50仍普遍高于单阶段模型，这说明其层次化结构与基于注意力的特征建模机制在弱光环境下对保持语义响应具有一定优势。而部分YOLO系列模型的性能衰减更为显著，说明其结构优化主要提升了可见光下的检测效率，但在低信噪比条件下缺乏鲁棒性设计。

表6　昼夜场景对比实验结果

Table 6　Results of the day and night scene comparison experiment

方法	日间			夜间		
	P%	R%	mAP50%	P%	R%	mAP50%
YOLOv5-s	46.9	35.5	34.7	43.9	26.5	26.0
YOLOv9-s	**52.5**	40.6	41.0	44.9	31.7	32.1
YOLOv10-m	51.3	40.9	40.0	48.0	31.6	32.2
YOLOv11-s	49.3	40.4	39.1	39.7	32.1	29.6
YOLOv13-s	45.7	37.5	35.9	42.1	26.3	27.3
Gold-YOLO	50.9	36.0	34.7	41.5	28.1	26.0
HIC-YOLOv5	50.0	41.2	41.2	43.6	32.7	33.4
QueryDet	52.3	46.1	44.8	**48.8**	33.4	**34.1**
ClusDet	51.0	**46.5**	**45.2**	47.9	**34.2**	33.8

注：加粗字体表示各列最优结果。

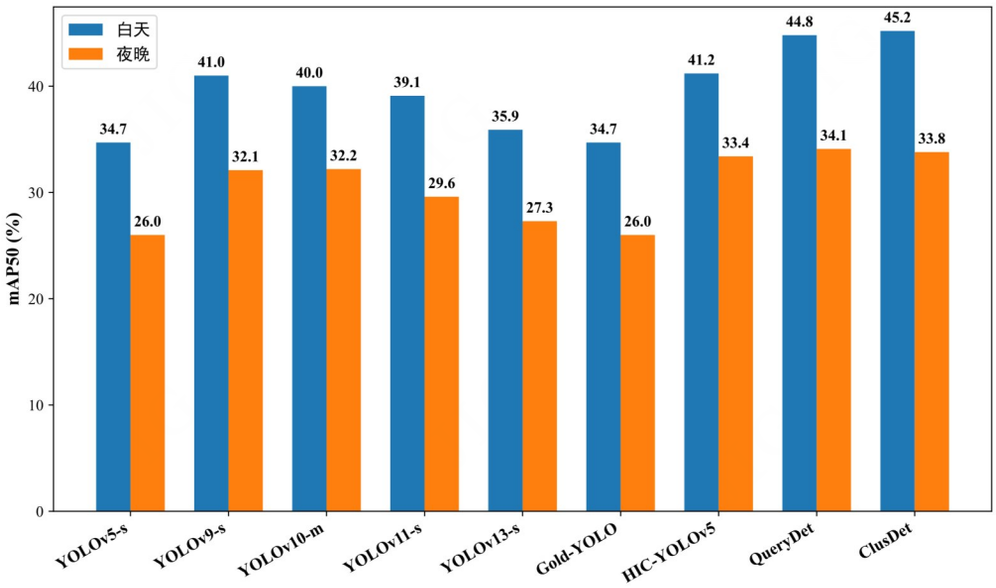


图4　检测算法昼夜对比

Fig. 4　Day-night comparison of general algorithms

与在DroneVehicle-Night上的实验不同，前者主要关注夜间典型算法及多模态融合方法的性能表现，旨在验证红外模态及模态互补在夜间无人机检测中的潜在优势。而本节基于VisDrone2019-DET的昼夜对比实验则从另一角度切入，揭示了在日间表现良好的目标检测算法，在夜间场景中性能显著衰减，尤其是在小目标和低对比度目标上存在明显漏检与误检。两组实验结果相互印证：一方面，表明夜间检测任务中引入多模态或光照自适应机制能够显著改善模型表现；另一方面，也凸显了直接迁移日间检测算法在夜间任务中存在的局限性。综合来看，夜间无人机航拍图像目标检测不应被视作日间检测算法的简单扩展，而应作为独立研究方向，算法设计需围绕低光增强、光照建模、跨模态融合及特征增强等方面进行针对性优化，以实现复杂光照环境下的稳健检测性能。

4.3　UAVDark135和NAT2021性能评价

为全面评估夜间无人机目标跟踪算法的性能，本节选取了20种具有代表性的主流方法，涵盖了先增强后跟踪方法、域自适应（DA）方法、视觉提示学习（PL）与课程学习（CL）方法以及孪生网络结构系列方法等。

表7汇总了这些方法在UAVDark135（Li等，2023）和NAT2021（Ye等，2022）数据集上的评估结果，图5展示了在UAVDark135数据集上精度-速度对比结果。

从整体趋势来看，基于Transformer与Mamba架构的模型在夜间跟踪任务中表现突出，说明长程依赖建模和动态特征调度对于复杂光照条件具有显著优势。动态特征自适应方法DARTer（Li等，2025）在AUC与Precision上取得了最优结果（NAT2021：53.2%/70.2%，UAVDark135：58.2%/71.6%），其自适应Transformer激活机制可针对夜间光照退化与运动模糊自动调整感受野范围，实现性能与计算效率的平衡。

相比之下，基于Mamba架构的MambaTrack（Zhang等，2025）与MambaNUT-Small（Wu等，2025）分别在AUC指标上取得了53.2%与52.4%的成绩，均位列前三。得益于Mamba的线性复杂度与时序依赖建模能力，两者在维持高精度的同时显著降低了推理复杂度。尤其是MambaNUT结合了课程式特征调度机制，使其在低光与遮挡场景下的鲁棒性显著提升。

提示学习类方法DCPT（darkness clue-prompted tracking）（Zhu等，2024）同样表现出较好的性能（UAVDark135：57.7%，NAT2021：52.6%），展示了在参数冻结条件下通过轻量提示模块实现知识迁移的潜力，且只调整小规模参数，就可以激活大模型原有的泛化能力，避免了“灾难性遗忘”。为未来基于大模型的无人机夜间视觉提供了可扩展方案。

表7　夜间无人机航拍图像目标跟踪方法与最先进跟踪方法的简要性能评估

Table 7　Brief performance evaluation of nighttime UAV aerial image object tracking methods and state-of-the-art tracking approaches

方法	发表信息	NAT2021			UAVDark135			Avg. FPS
		AUC	Prec	P_{norm}	AUC	Prec	P_{norm}	
Ocean(Zhang 等,2020)	ECCV 2020	38.6	58.1	49.9	48.1	60.8	60.3	43
SiamAPN++(Cao 等,2021)	IROS 2021	41.2	60.2	51.4	33.0	43.2	29.5	114
HiFT(Cao 等,2021)	ICCV 2021	37.0	54.5	46.7	34.8	45.4	32.9	123
SiamRPN++(Li 等,2019)	CVPR 2019	41.3	61.5	52.1	37.2	47.4	35.0	**152**
MAT(Zhao 等,2023)	CVPR 2023	47.7	64.8	58.8	47.7	57.2	57.6	56
SiamCAR(Guo 等,2020)	CVPR 2020	45.6	65.8	59.5	40.6	52.5	52.3	37
DarkLighter_SiamRPN++(Ye 等,2021)	IROS 2021	–	–	–	38.5	49.5	–	108
SCT+SiamRPN++(Ye 等,2022)	IRAL 2022	–	–	–	42.1	54.7	–	31
LDEnhancer_SiamRPN++(Yao 等,2024)	IROS 2024	–	–	–	43.4	56.0	41.4	30
HighlightNet_SiamAPN+(Fu 等,2022)	IROS2022	–	–	–	42.4	53.9	–	32
UDAT_SiamCAR(Ye 等,2022)	CVPR2022	48.1	67.9	60.9	48.6	60.7	61.3	36
PDST_SiamCAR(Zhang 等,2023)	ACM MM2023	50.2	69.1	63.1	50.3	63.3	64.0	–
TCTrack++(Cao 等,2023)	TPAMI2023	41.7	61.1	52.8	37.8	47.4	47.4	122
DCPT(Zhu 等,2024)	ICRA2024	52.6	69.0	63.5	57.7	70.3	70.1	35
SAM-DA(Fu 等,2024)	ICARM2024	47.1	67.3	59.2	47.6	60.4	59.4	37
TDA-Track(Fu 等,2024)	IROS2024	42.3	61.7	53.5	36.9	49.5	49.9	114
NiDR(Lei 等,2024)	TGRS2024	47.0	65.2	57.1	51.1	64.2	62.9	71
MambaTrack(Zhang 等,2025)	ICASSP 2025	**53.2**	67.5	56.8	57.2	67.6	61.4	42
MambaNUT-Small(Wu 等,2025)	IROS2025	52.4	70.1	**64.6**	57.1	70.0	69.3	72
DARTer(Li 等,2025)	ICMR2025	**53.2**	**70.2**	63.7	**58.2**	**71.6**	**72.1**	74

注:加粗字体表示各列最优结果,"–"表示无相应的实验数据。

从运行效率角度看,SiamRPN++(Li 等,2019)与SiamAPN++(Cao 等,2021)等基于孪生网络的模型依旧在实时性上占据优势,推理速度分别达到152FPS与114FPS,但其在低光场景中的精度明显低于其他四种类别。说明传统孪生网络虽具有速度优势,但在夜间复杂背景下的特征分辨能力有限。

5　总结和展望

本文围绕夜间无人机航拍图像目标检测与跟踪这一前沿领域展开系统性综述,深入剖析了该领域的研究价值与核心技术挑战,包括夜间低光照导致的感知能力有限、目标特征退化、无人机硬件算力受限以及小目标与视角变化带来的复杂问题等。

在夜间无人机航拍图像目标检测方面,梳理了夜间图像增强、域适应学习、多模态感知融合和轻量化模型等方法的进展,且在夜间无人机航拍图像数据集DroneVehicle-Night上进行性能评估和VisDrone2019数据集上开展的昼夜对比实验,验证了红外模态及多模态融合在夜间检测中的优势,同时揭示了通用检测算法在夜间场景的性能衰减问题,强调了针对夜间特性专门优化的必要性。在目标跟踪方面,归纳了基于深度学习的五类主流范式,即先增强后跟踪、域自适应、视觉提示学习、课程学习和多模态目标跟踪,详细阐述了各类方法的核心思想、代表工作及优缺点。结合UAVDark135、NAT2021等夜间无人机航拍图像目标跟踪数据集,对20种代表性算法的性能进行了评估汇总,展现了不同方法在

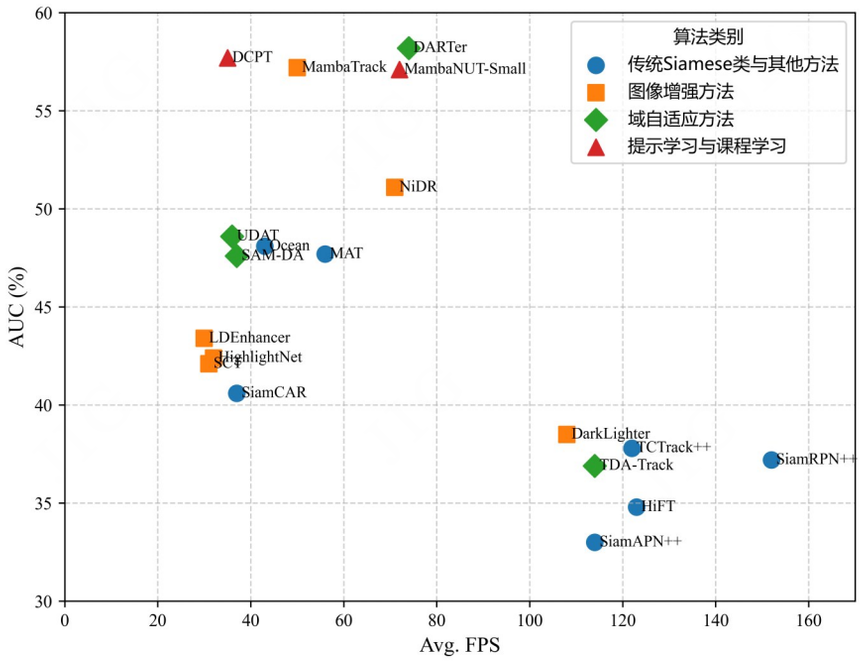


图5　UAVDark135数据集上速度与精度比较

Fig. 5　Comparison of speed and accuracy on the UAVDark135 dataset

夜间跟踪任务中的表现。

综合来看，夜间无人机航拍图像目标跟踪已形成多范式解决方案，但目标检测仍面临专用研究稀缺、数据集不足的瓶颈，且跨模态融合、轻量化设计等关键方向尚未实现充分突破。本文通过数据驱动与算法分析，为夜间无人机视觉任务提供了从理论到实践的系统性参考，也为后续技术创新明确了关键突破点。

针对如何进一步提升夜间无人机航拍图像目标检测与跟踪的性能，本文对未来的研究方向进行展望。

1）跨模态视觉-语言驱动的检测与跟踪协同优化。近年来大语言模型（large language model，LLM）与视觉任务的融合为夜间无人机航拍图像目标感知提供了全新路径。在目标检测中，跨模态技术可通过文本语义补全夜间退化的视觉特征，例如：当可见光图像中车辆因低光照呈现模糊轮廓时，语言模型可基于夜间行驶车辆通常伴随灯光轨迹的先验知识，辅助分类器过滤误检并提升小目标检测精度；在红外-可见光跨模态检测中，语言模型能关联热辐射高的目标可能为发动机等领域知识，优化跨模态特征融合策略。对于目标跟踪，能利用文本查询直接关联视觉目标，支持非实时标注场景下的目标回溯与跟踪初始化，尤其适用于夜间无预训练数据的突发任务；在长时跟踪中语言可作为记忆锚点，记录目标历史属性，缓解因视觉特征漂移导致的跟踪丢失；地面控制人员可通过自然语言实时调整跟踪策略，结合视觉反馈形成闭环交互，提升复杂场景下的决策效率。

2）构建夜间复杂场景无人机检测跟踪数据集。现有数据集在场景覆盖和目标类型上存在一定局限性，大多数目标集中于城市道路场景中的人和车辆。这类场景虽贴近常见应用，但也暴露出一些问题：如场景单一化，数据集中的背景高度依赖于城市基础设施，缺乏对野外、山区、水域等复杂场景的覆盖，导致算法在非结构化夜间场景（战场侦察、森林救援、海岸线巡逻）中泛化不足；跨域适应性不足，所列举出的数据集多数光照模式为人工光源主导，与其他的夜间环境差异显著，导致算法在不同光照分布下鲁棒性可能缺乏验证。因此，未来可以考虑扩展多场景数据集（如增加野外、室内等场景的夜间视频）、引入多模态数据（同步采集可见光、红外、激光雷达等数据）。

3）无监督/弱监督视觉目标检测与跟踪。训练高性能模型高度依赖大规模标注视频数据。然而手动逐帧标注成本高昂，且当模型应用于差异显著的

新场景时,需要重新标注训练数据,导致算法迭代更新成本高。在此背景下,研究无监督与弱监督训练方式成为关键突破口,其核心价值在于显著降低数据标注与模型适配成本。本文聚焦夜间无人机场景,部分方法对无监督域自适应跟踪技术展开研究。因此,未来可以进一步拓展技术边界:一方面,从域适应向域泛化延伸,使模型同时具备多个未知目标域的能力,而非局限于单一源域到目标域的迁移;另一方面,借鉴大语言模型的自监督学习思路,探索利用海量无标注视频数据(如互联网公开影像、无人机自采数据)构建自监督pretext任务,以摆脱对强监督信号的依赖。

4)轻量化架构设计。当前检测与跟踪方法的发展仍以提升数据集性能为主要目标,而在无人机等资源受限的实际应用中,轻量化与实时性同样至关重要。未来研究可从多个层面展开探索:在模型架构层面,需要突破对Transformer等复杂结构的依赖。已有研究表明,Mamba通过线性复杂度的状态空间模型(state space model,SSM)能够在长序列建模上展现出优于传统注意力机制的效率,这为夜间无人机检测与跟踪提供了新的轻量化思路。同时,神经网络架构搜索(neural architecture search,NAS)、动态推理等方法也有助于在性能与效率之间取得更优平衡;在算子层面,未来可以进一步关注高效卷积、轻量化注意力及硬件友好的归一化与激活函数优化,通过高效算子设计减少底层计算开销。总体而言,未来轻量化研究不仅限于单一模型压缩,更应面向架构、算子等全方位协同设计,以满足夜间无人机检测与跟踪对实时性、精度与能耗的多重要求。

参考文献(References)

Avidan S. 2004. Support vector tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(8): 1064-1072 [DOI: 10.1109/TPAMI.2004.53]

Bengio Y, Louradour J, Collobert R and Weston J. 2009. Curriculum learning//Proceedings of the 26th Annual International Conference on Machine Learning. Montreal, Canada: ACM: 41-48 [DOI: 10.1145/1553374.1553380]

Berjawi J, Dupas Y and Cerin C. 2025. Towards a generalizable fusion architecture for multimodal object detection//Proceedings of 2025 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Hawaii, USA: IEEE: 2192-2200

Bolme D S, Beveridge J R, Draper B A and Liu Y M. 2010. Visual object tracking using adaptive correlation filters//Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE: 2544-2550 [DOI: 10.1109/CVPR.2010.5539960]

Cai Z W and Vasconcelos N. 2018. Cascade R-CNN: delving into high quality object detection//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 6154-6162 [DOI: 10.1109/CVPR.2018.00644]

Cao B, Guo J L, Zhu P F and Hu Q H. 2024. Bi-directional adapter for multimodal tracking//Proceedings of the 39th Annual AAAI Conference on Artificial Intelligence. Pennsylvania, USA: AAAI Press: 927-935 [DOI: 10.1609/aaai.v38i2.27852]

Cao J L, Li Y L, Sun H Q, Xie J Huang K Q and Pang Y W. 2022. A survey on deep learning based visual object detection. Journal of Image and Graphics, 27(6): 1697-1722 (曹家乐, 李亚利, 孙汉卿, 谢今, 黄凯奇, 庞彦伟. 2022. 基于深度学习的视觉目标检测技术综述. 中国图象图形学报, 27(6): 1697-1722) [DOI: 10.11834/jig.220069]

Cao Y, Bin J C, Hamari J, Blasch E and Liu Z. 2023. Multimodal object detection by channel switching and spatial attention//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Vancouver, Canada: IEEE: 403-411 [DOI: 10.1109/CVPRW59228.2023.00046]

Cao Z A, Fu C H, Ye J J, Li B W and Li Y M. 2021. HiFT: hierarchical feature transformer for aerial tracking//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE: 15437-15466 [DOI: 10.1109/ICCV48922.2021.01517]

Cao Z A, Fu C H, Ye J J, Li B W and Li Y M. 2021. SiamAPN++: siamese attentional aggregation network for real-time UAV tracking//Proceedings of 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Prague, Czech Republic: IEEE: 3086-3092 [DOI: 10.1109/IROS51168.2021.9636309]

Cao Z A, Huang Z Y, Pan L, Zhang S W, Liu Z W and Fu C H. 2023. Towards real-world visual tracking with temporal contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(12): 15834-15849 [DOI: 10.1109/TPAMI.2023.3307174]

Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A and Zagoruyko S. 2020. End-to-end object detection with transformers//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 213-229 [DOI: 10.1007/978-3-030-58452-8_13]

Chan S X, Li Z D, Li W H, Lu S J, Shen C H and Zhang X Q. 2025. SMSTracker: tri-path score mask sigma fusion for multi-modal tracking//Proceedings of 2025 IEEE/CVF International Conference on Computer Vision (ICCV). Hawaii, USA: IEEE: 4776-4775

Chen H Y, Liu J P, Wang Y, Zhu J, Feng D J and Xie Y K. 2025. Teaching in adverse scenes: a statistically feedback-driven thresh-

old and mask adjustment teacher-student framework for object detection in UAV images under adverse scenes. ISPRS Journal of Photogrammetry and Remote Sensing, 227: 332-348 [DOI: 10.1016/j.isprsjprs.2025.06.009]

Chen J Y, Sun Q Y, Zhao C Q, Ren W Q and Tang Y. 2023. Rethinking unsupervised domain adaptation for nighttime tracking//Proceedings of the 30th International Conference on Neural Information Processing. Changsha, China: Springer: 391-404 [DOI: 10.1007/978-981-99-8181-6_30]

Chen K, Wang J Q, Pang J M, Cao Y H, Xiong Y, Li X X, Sun S Y, Feng W S, Liu Z W, Xu J R, Zhang Z, Cheng D Z, Zhu C C, Cheng T H, Zhao Q J, Li B Y, Lu X, Zhu R, Wu Y, Dai J F, Wang J D, Shi J P, Ouyang W L, Loy C C and Lin D H. 2019. MMDetection: open mmlab detection toolbox and benchmark [EB/OL]. [2025-08-13]. https://arxiv.org/pdf/1906.07155.pdf

Chen L, Deng H B, Liu G H, Law B, Li D F, Wu E Q and Zhu L M. 2026. Retinex-guided illumination recovery and progressive feature adaptation for real-world nighttime UAV-based vehicle detection. Expert Systems with Applications, 297: #129476 [DOI: 10.1016/j.eswa.2025.129476]

Chen Z Y, Wang Z D and Gong C. 2023. Image-level labeled weakly supervised object detection: a survey. Journal of Image and Graphics, 28(09): 2644-2660 (陈震元, 王振东, 宫辰. 2023. 图像级标记弱监督目标检测综述. 中国图象图形学报, 28(09): 2644-2660) [DOI: 10.11834/jig.220854]

Chopra S, Hadsell R and LeCun Y. 2005. Learning a similarity metric discriminatively, with application to face verification//Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). San Diego, USA: IEEE: 539-546 [DOI: 10.1109/CVPR.2005.202]

Chouhan A, Chandak M, Sur A, Chutia D and Aggarwal S P. 2024. RADA: reconstruction assisted domain adaptation for nighttime aerial tracking//Proceedings of the 27th International Conference on Pattern Recognition. Kolkata, India: Springer: 315-330 [DOI: 10.1007/978-3-031-78192-6_21]

Comaniciu D, Ramesh V and Meer P. 2000. Real-time tracking of non-rigid objects using mean shift//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662). Hilton Head, USA: IEEE: 142-149 [DOI: 10.1109/CVPR.2000.854761]

Dabov K, Foi A, Katkovnik V and Egiazarian K. 2007. Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Transactions on Image Processing, 16(8): 2080-2095 [DOI: 10.1109/TIP.2007.901238]

Dai X B, Hu J P, Luo C L, Zerfa H, Zhang H and Duan Y X. 2023. NIRPed: a novel benchmark for nighttime pedestrian and its distance joint detection. IEEE Transactions on Intelligent Transportation Systems, 24(7): 6932-6942 [DOI: 10.1109/TITS.2023.3257079]

Danelljan M, Häger G, Khan F and Felsberg M. 2014. Accurate scale estimation for robust visual tracking//Proceedings of the British Machine Vision Conference. Nottingham, UK: BMVA Press [DOI: 10.5244/c.28.65]

Danelljan M, Hager G, Khan F S and Felsberg M. 2015. Learning spatially regularized correlation filters for visual tracking//Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE: 4310-4318 [DOI: 10.1109/ICCV.2015.490]

Dhariwal P and Nichol A. 2021. Diffusion models beat GANs on image synthesis//Proceedings of the 35th International Conference on Neural Information Processing Systems. Virtually: Curran Associates Inc: 8780-8794 [DOI: 10.5555/3540261.3540933]

Dhrafani D, Liu Y F, Jong A, Shin U, He Y and Harp T. 2025. FireStereo: forest infrared stereo dataset for UAS depth perception in visually degraded environments. IEEE Robotics and Automation Letters, 10(4): 3302-3309 [DOI: 10.1109/LRA.2025.3536278]

Ding Z D, Li C L, Miao S Q and Tang J. 2025. Template-based uncertainty multimodal fusion network for RGBT tracking//Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-25). Montreal, Canada: IJCAI: 909-917 [DOI: 10.24963/ijcai.2025/102]

Du D W, Qi Y K, Yu H Y, Yang Y F, Duan K W, Li G R, Zhang W G, Huang Q M and Tian Q. 2018. The unmanned aerial vehicle benchmark: object detection and tracking//Proceedings of 15th European Conference on Computer Vision. Munich, Germany: Springer: 375-391 [DOI: 10.1007/978-3-030-01249-6_23]

Du Z P, Shi M J, Deng J K. 2024. Boosting object detection with zero-shot day-night domain adaptation//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 12666-12676 [DOI: 10.1109/CVPR52733.2024.01204]

Fan H, Lin L T, Yang F, Chu P, Deng G, Yu S J, Bai H X, Xu Y, Liao C Y and Ling H B. 2019. LaSOT: a high-quality benchmark for large-scale single object tracking//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 5369-5378 [DOI: 10.1109/CVPR.2019.00552]

Fan L W, Yang J J, Wang L, Zhang J P, Lian X K and Shen H. 2025. Efficient spiking neural network for RGB-event fusion-based object detection. Electronics, 14(6): 1105 [DOI: 10.3390/electronics14061105]

Fan Q S, Li Y T, Deveci M, Zhong K Y and Kadry S. 2025. LUD-YOLO: a novel lightweight object detection network for unmanned aerial vehicle. Information Science, 686: #121366 [DOI: 10.1016/j.ins.2024.121366]

Fang Q Y, Han D P and Wang Z K. 2022. Cross-modality fusion transformer for multispectral object detection [EB/OL].

https://arxiv.org/pdf/2111.00273.pdf

Feng Q H, Wang Z X, Sun C C and Shao Z W. 2025. Small object detection in drone images via foreground refinement and multidimensional inductive bias self-attention. Journal of Image and Graphics, 30(11): 3547-3563 (冯琪涵, 王志晓, 孙成成, 邵志文. 2025. 融合前景细化和多维归纳偏置自注意力的无人机图像小目标检测. 中国图象图形学报, 30(11): 3547-3563) [DOI: 10.11834/jig.250017]

Feng X K, Zhang D L, Hu S Y, Li X C, Wu M Q, Zhang J, Chen X T and Huang K Q. 2025. CSTrack: enhancing RGB-X tracking via compact spatiotemporal features//Proceedings of 2025 IEEE/CVF International Conference on Computer Vision (ICCV). Hawaii, USA: IEEE: 4766-4775

Feng X, Zeng J X, Wang S P and He Z W. 2024. Toward highly efficient semantic-guided machine vision for low-light object detection//Proceedings of the British Machine Vision Conference. Glasgow, UK: BMVA Press

Fu C H, Dong H L, Ye J J, Zheng G Z, Li S H and Zhao J L. 2022. HighlightNet: highlighting low-light potential features for real-time UAV tracking//Proceedings of 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Kyoto, Japan: IEEE: 12146-12153 [DOI: 10.1109/IROS47612.2022.9981070]

Fu C H, Wang Y H, Yao L L, Zheng G Z, Zuo H B and Pan J. 2024. Prompt-driven temporal domain adaptation for nighttime UAV tracking//Proceedings of 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Abu Dhabi, United Arab Emirates: IEEE: 9706-9713 [DOI: 10.1109/IROS58592.2024.10801335]

Fu C H, Yao L L, Zuo H B, Zheng G Z and Pan J. 2024. SAM-DA: UAV tracks anything at night with sam-powered domain adaptation//Proceedings of IEEE International Conference on Advanced Robotics and Mechatronics (ICARM). Tokyo, Japan: IEEE: 31-38 [DOI: 10.1109/ICARM62033.2024.10715901]

Galoogahi H K, Fagg A and Lucey S. 2017. Learning background-aware correlation filters for visual tracking//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE: 1144-1152 [DOI: 10.1109/ICCV.2017.129]

Gao D, Lai P J, Wang S L and Cheng G. 2025. RGB-T UAV object tracking based on feature-cooperative reconstruction. Acta Aeronautica et Astronautica Sinica, 46(24): #632017 (高栋, 赖普坚, 王世磊, 程塬. 2025. 基于特征协同重构的RGB-T无人机目标跟踪. 航空学报, 46(24): #632017) [DOI: 10.7527/S1000-6893.2025.32017]

Gao T F, He Y X, Ma X L, Lu Z L, Peng S Y and Liu Y. 2025. A review of single target tracking in satellite videos[J/OL]. Journal of Image and Graphics, 1-19 (高桃峰, 何银鑫, 马学良, 卢自来, 彭世勇, 刘洋. 卫星视频单目标跟踪综述[J/OL]. 中国图象图形学报, 1-19 [DOI:10.11834/jig.240750]

Gasienica-Jozkowy J, Knapik M and Cyganek B. 2021. An ensemble deep learning method with optimized weights for drone-based water rescue and surveillance. Integrated Computer-Aided Engineering, 28(3): 221-235 [DOI: 10.3233/ICA-210649]

Girshick R, Donahue J, Darrell T and Malik J. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE: 580-587 [DOI: 10.1109/CVPR.2014.81]

Girshick R. 2015. Fast r-cnn//Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE: 1440-1448 [DOI: 10.1109/ICCV.2015.169]

Gong H F, Sim J, Likhachev M and Shi J B. 2011. Multi-hypothesis motion planning for visual object tracking//Proceedings of 2011 International Conference on Computer Vision. Barcelona, Spain: IEEE: 619-626 [DOI: 10.1109/ICCV.2011.6126296]

Grabner H, Grabner M and Bischof H. 2006. Real-time tracking via online boosting//Proceedings of the British Machine Vision Conference 2006. Edinburgh, UK: BMVA Press: 6.1-6.10 [DOI: 10.5244/c.20.6]

Gu A and Dao T. 2024. Mamba: linear-time sequence modeling with selective state spaces[EB/OL]. [2025-08-13]. https://arxiv.org/pdf/2312.00752.pdf

Guo C L, Li C Y, Guo J C, Loy C C, Hou J H, Kwong S and Cong R M. 2020. Zero-reference deep curve estimation for low-light image enhancement//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 1777-1786 [DOI: 10.1109/CVPR42600.2020.00185]

Guo D Y, Wang J, Cui Y, Wang Z H and Chen S Y. 2020. SiamCAR: siamese fully convolutional classification and regression for visual tracking//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 6268-6276 [DOI: 10.1109/CVPR42600.2020.00630]

Guo J J, Gao C Q, Liu F C, Meng D Y and Gao X B. 2024. DAMSDet: dynamic adaptive multispectral detection transformer with competitive query selection and adaptive feature fusion//Proceedings of 18th European Conference Computer. Milan, Italy: Springer: 464-481 [DOI: 10.1007/978-3-031-73383-3_27]

Guo X J, Li Y and Ling H B. 2017. LIME: low-light image enhancement via illumination map estimation. IEEE Transactions on Image Processing, 26(2): 982-993 [DOI: 10.1109/TIP.2016.2639450]

Hamann F, Gehrig D, Febryanto F, Daniilidis K and Gallego G. 2025. ETAP: event-based tracking of any point//Proceedings of 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 27186-27196 [DOI: 10.1109/CVPR52734.2025.02532]

Han J M, Ding J, Li J and Xia G S. 2022. Align deep features for oriented object detection. IEEE Transactions on Geoscience and Remote Sensing, 60: #5602511 [DOI: 10.1109/TGRS.2021.3062048]

He K M, Gkioxari G, Dollar P and Girshick R. 2017. Mask R-CNN//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE: 2980-2988 [DOI: 10.1109/ICCV.2017.322]

He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

He X H, Cao K, Zhang J, Yan K Y, Wang Y Y, Li R, Xie C J, Hong D F and Zhou M. 2025. Pan-Mamba: effective pan-sharpening with state space model. Information Fusion, 115: #102779 [DOI: 10.1016/j.inffus.2024.102779]

He X, Tang C, Zou X and Zhang W. 2023. Multispectral object detection via cross-modal conflict-aware learning//Proceedings of the 31st ACM International Conference on Multimedia. Ottawa, Canada: ACM: 1465-1474 [DOI: 10.1145/3581783.3612651]

Henriques J F, Caseiro R, Martins P and Batista J. 2015. High-speed tracking with kernelized correlation filters. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(3): 583-596 [DOI: 10.1109/TPAMI.2014.2345390]

Hong L Y, Yan S L, Zhang R R, Li W Y, Zhou X Y, Guo P X, Jiang K X, Chen Y T, Li J L, Chen Z Y and Zhang W Q. 2024. OneTracker: unifying visual object tracking with foundation models and efficient tuning//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 19079-19091 [DOI: 10.1109/CVPR52733.2024.01805]

Hou Q B, Zhou D Q and Feng J S. 2021. Coordinate attention for efficient mobile network design//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 13708-13717 [DOI: 10.1109/CVPR46437.2021.01350]

Hou X J, Xing J Z, Qian Y J, Guo Y W, Xin S, Chen J H, Tang K, Wang M M, Jiang Z K, Liu L and Liu Y. 2024. SDSTrack: self-distillation symmetric adapter learning for multi-modal visual object tracking//Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 26541-26551 [DOI: 10.1109/CVPR52733.2024.02507]

Hu E J, Shen Y L, Wallis P, Allen-Zhu Z Y, Li Y Z, Wang S A and Chen W Z. 2021. LoRa: low-rank adaptation of large language models[EB/OL]. [2025-08-13]. https://arxiv.org/pdf/2106.09685v2.pdf

Hu K, He Y D, Li Y, Zhao J Y, Chen S and Kang Y. 2025. EI^{2}Det: edge-guided illumination-aware interactive learning for visible-infrared object detection. IEEE Transactions on Circuits and Systems for Video Technology, 35(7): 7101-7115 [DOI: 10.1109/TCSVT.2025.3539625]

Hu Q K, Li Y C and Yu W B. 2025. Exploiting multimodal prompt learning and distillation for RGB-T tracking//Proceedings of the 2025 International Conference on Multimedia Retrieval. Chicago, IL, USA: ACM: 469-477 [DOI: 10.1145/3731715.3733332]

Hu X T, Tai Y, Zhao X, Zhao C, Zhang Z Y, Li J, Zhong B N and Yang J. 2025. Exploiting multimodal spatial-temporal patterns for video object tracking//Proceedings of the 39th Annual AAAI Conference on Artificial Intelligence. Pennsylvania, USA: AAAI Press: 3581-3589 [DOI: 10.1609/aaai.v39i4.32372]

Huang L H, Zhao X and Huang K Q. 2021. GOT-10k: a large high-diversity benchmark for generic object tracking in the wild. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(5): 1562-1577 [DOI: 10.1109/TPAMI.2019.2957464]

Huang S C, Hoang Q V and Jaw D W. 2022. Self-adaptive feature transformation networks for object detection in low luminance images. ACM Transactions on Intelligent Systems and Technology (TIST), 13(1): 1-11 [DOI: 10.1145/3480973]

Huang X M, Wu Z H, Li Y, Shang C J and Shen Q. 2025. Pyramid attention enhancement network for nighttime UAV tracking//Proceedings of ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Hyderabad, India: IEEE: 1-5 [DOI: 10.1109/ICASSP49660.2025.10889408]

Huang X, Liu M Y, Belongie S and Kautz J. 2018. Multimodal unsupervised image-to-image translation//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 179-196 [DOI: 10.1007/978-3-030-01219-9_11]

Hui T R, Xun Z Z, Peng F G, Huang J S, Wei X M, Wei X L, Dai J, Han J Z and Liu S. 2023. Bridging search region interaction with template for RGB-T tracking//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 13630-13639 [DOI: 10.1109/CVPR52729.2023.01310]

Isard M and Blake A. 1998. Condensation—conditional density propagation for visual tracking. International Journal of Computer Vision, 29(1): 5-28 [DOI: 10.1023/A:1008078328650]

Javed S, Danelljan M, Khan F S, Khan M H, Felsberg M and Matas J. 2023. Visual object tracking with discriminative filters and siamese networks: a survey and outlook. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(5): 6552-6574 [DOI: 10.1109/TPAMI.2022.3212594]

Jia M L, Tang L M, Chen B C, Cardie C, Belongie S, Hariharan B and Lim S N. 2022. Visual prompt tuning//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 709-727 [DOI: 10.1007/978-3-031-19827-4_41]

Jiang C, Cui Y W and Wang H. 2021. A review of UAV battlefield situational awareness technology based on images. Measurement & Control Technology, 40(12): 14-19 (蒋超, 崔玉伟, 王辉. 2021. 基于图像的无人机战场态势感知技术综述. 测控技术, 40(12): 14-19) [DOI: 10.19708/j.ckjs.2021.12.001]

Jiang S J, Ning J F and Li Y S. 2017. Object tracking algorithm via weighted margin structured support vector machine. Journal of Image and Graphics, 22(9): 1261-1269 (江少杰, 宁纪锋, 李云

松. 2017. 加权间隔结构化支持向量机目标跟踪算法. 中国图象图形学报, 22(9): 1261-1269[DOI: 10.11834/jig.160651]

Jiang Y F, Gong X Y, Liu D, Cheng Y, Fang C, Shen X H, Yang J C, Zhou P and Wang Z Y. 2021. EnlightenGAN: deep light enhancement without paired supervision. IEEE Transactions on Image Processing, 30: 2340-2349[DOI: 10.1109/TIP.2021.3051462]

Jiang Y, Wang Y H, Zhao M H, Zhang Y J and Qi H. 2025. Nighttime traffic object detection via adaptively integrating event and frame domains. Fundamental Research, 5(4): 1633-1644[DOI: 10.1016/j.fmre.2023.08.004]

Jiao L C, Wang D, Bai Y D, Chen P H and Liu F. 2023. Deep learning in visual tracking: a review. IEEE Transactions on Neural Networks and Learning Systems, 34(9): 5497-5516[DOI: 10.1109/TNNLS.2021.3136907]

Jin G Y, Zhao T M, Yan J C and Tian T. 2025. Contextually-guided state space fusion for misaligned multi-spectral object detection// Proceedings of the 33rd ACM International Conference on Multimedia. Doblin, Ireland: ACM: 2526-2535[DOI: 10.1145/3746027.3754550]

Jocher G, Chaurasia A, Qiu J and Ultralytics. 2023. YOLOv8: a cutting-edge YOLO model for object detection[EB/OL]. [2025-08-13]. https://github.com/ultralytics/ultralytics

Jocher G, Chaurasia A, Qiu J and Ultralytics. 2024. YOLOv11: real-time object detection with YOLO architecture[EB/OL]. [2025-08-13]. https://github.com/ultralytics/ultralytics

Jocher G, Chaurasis A and Stoken A. 2020. YOLOv5: a PyTorch implementation of the YOLO object detector[EB/OL]. [2025-08-13]. https://github.com/ultralytics/yolov5

Jung I, Son J, Baek M and Han B. 2018. Real-time MDNet//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 89-104[DOI: 10.1007/978-3-030-01225-0_6]

Kalman R E. 1960. A new approach to linear filtering and prediction problems. Journal of Basic Engineering, 82(1): 35-45[DOI: 10.1115/1.3662552]

Kennerley M, Wang J G, Veeravalli B and Tan R T. 2023. 2PCNet: two-phase consistency training for day-to-night unsupervised domain adaptive object detection// Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 11484-11493[DOI: 10.1109/CVPR52729.2023.01105]

Kim H, Lee D, Park S and Ro Y M. 2024. Weather-aware drone-view object detection via environmental context understanding//Proceedings of 2024 IEEE International Conference on Image Processing (ICIP). Abu Dhabi, United Arab Emirates: IEEE: 549-555[DOI: 10.1109/ICIP51287.2024.10647388]

Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg A C, Lo W Y, Dollar P and Girshick R. 2023. Segment anything//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE: 3992-4003[DOI: 10.1109/ICCV51070.2023.00371]

Kraft M, Piechocki M, Ptak B and Walas K. 2021. Autonomous, onboard vision-based trash and litter detection in low altitude aerial images collected by an unmanned aerial vehicle. Remote Sensing, 13(5): #965[DOI: 10.3390/rs13050965]

Krizhevsky A, Sutskever I and Hinton G E. 2017. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6): 84-90[DOI: 10.1145/3065386]

Lan X Y, Ye M, Zhang S P and Yuen P C. 2018. Robust collaborative discriminative learning for RGB-infrared tracking//Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence. New orleans, Louisiana, USA: AAAI Press: 7008-7015[DOI: 10.5555/3504035.3504893]

Land E H. 1977. The retinex theory of color vision. Scientific American, 237(6): 108-129[DOI: 10.1038/scientificamerican1277-108]

Lashkov I, Yuan R Z and Zhang G H. 2023. Edge-computing-facilitated nighttime vehicle detection investigations with CLAHE-enhanced images. IEEE Transactions on Intelligent Transportation Systems, 24(11): 13370-13383[DOI: 10.1109/TITS.2023.3255202]

Lei M Q, Li S Q, Wu Y H, Zhou Y, Zheng X H, Ding G G, Du S Y, Wu Z Z and Gao Y. 2025. YOLOv13: real-time object detection with hypergraph-enhanced adaptive visual perception[EB/OL]. [2025-08-13]. https://arxiv.org/pdf/2506.17733.pdf

Lei X, Zhang Y, Xu C, Cheng W S and Yang W. 2024. NiDR: nighttime aerial tracking via decoupled representations. IEEE Transactions on Geoscience and Remote Sensing, 62: #5651014[DOI: 10.1109/TGRS.2024.3508136]

Leng J X, Mo M J C, Zhou Y H, Ye Y M, Gao C Q and Gao X B. 2023. Recent advances in drone-view object detection. Journal of Image and Graphics, 28(9): 2563-2586(冷佳旭, 莫梦竟成, 周应华, 叶永明, 高陈强, 高新波. 2023. 无人机视角下的目标检测研究进展. 中国图象图形学报, 28(9): 2563-2586)[DOI: 10.11834/jig.220836]

Li A, Ni S X, Chen Y N, Chen J X, Wei X, Zhou L and Guizani M. 2023. Cross-modal object detection via UAV. IEEE Transactions on Vehicular Technology, 72(8): 10894-10905[DOI: 10.1109/TVT.2023.3262129]

Li B W, Fu C H, Ding F Q, Ye J J and Lin F L. 2021. ADTrack: target-aware dual filter learning for real-time anti-dark UAV tracking//Proceedings of 2021 IEEE International Conference on Robotics and Automation (ICRA). Xi'an, China: IEEE: 496-502[DOI: 10.1109/ICRA48506.2021.9561564]

Li B W, Fu C H, Ding F Q, Ye J J and Lin F L. 2023. All-day object tracking for unmanned aerial vehicle. IEEE Transactions on Mobile Computing, 22(8): 4515-4529[DOI: 10.1109/TMC. 2022.

3162892］

Li B，Wu W，Wang Q，Zhang F Y，Xing J L and Yan J J. 2019. SiamRPN++：evolution of siamese visual tracking with very deep networks//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition（CVPR）. Long Beach，USA：IEEE：4277-4286［DOI：10.1109/CVPR.2019.00441］

Li C C，Chen G，Hou Z X，Huang K and Zhang W. 2024. Survey of 3D object detection algorithms for autonomous driving. Journal of Image and Graphics，29(11)：3238-3264（李昌财，陈刚，侯作勋，黄凯，张伟. 2024. 自动驾驶中的三维目标检测算法研究综述. 中国图象图形学报，29(11)：3238-3264）［DOI：10.11834/jig.230779］

Li C L，Lu A D，Liu L and Tang J.2023. Multi-modal visual tracking：a survey. Journal of Image and Graphics，28(01)：0037-0056（李成龙，鹿安东，刘磊，汤进. 2023. 多模态视觉跟踪方法综述. 中国图象图形学报，28(01)：0037-0056）［DOI：10.11834/jig.220578］

Li C L，Lu A D，Zheng A H，Tu Z Z and Tang J. 2019. Multi-adapter RGBT tracking//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision Workshop（ICCVW）. Seoul，Korea（South）：IEEE：2262-2270［DOI：10.1109/ICCVW.2019.00279］

Li C L，Sun X，Wang X，Zhang L and Tang J. 2017. Grayscale-thermal object tracking via multitask laplacian sparse representation. IEEE Transactions on Systems，Man，and Cybernetics：Systems，47(4)：673-681［DOI：10.1109/TSMC.2016.2627052］

Li C L，Wu X H，Zhao N，Cao X C and Tang J. 2018. Fusing two-stream convolutional neural networks for RGB-T object tracking. Neurocomputing，281：78-85［DOI：10.1016/j. neucom. 2017.11.068］

Li C L，Zhao N，Lu Y J，Zhu C L and Tang J. 2017. Weighted sparse representation regularized graph learning for RGB-T object tracking//Proceedings of the 25th ACM international conference on Multimedia. California，USA：ACM：1856-1864［DOI：10.1145/3123266.312328］

Li C Y，Guo C L and Loy C C. 2022. Learning to enhance low-light image via zero-reference deep curve estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence，44(8)：4225-4238［DOI：10.1109/TPAMI.2021.3063604］

Li D Z，Tian Y H and Li J N. 2023. SODFormer：streaming object detection with transformer using events and frames. IEEE Transactions on Pattern Analysis and Machine Intelligence，45(11)：14020-14037［DOI：10.1109/TPAMI.2023.3298925］

Li J H，Zheng H B，Cui Z M，Huang Z D，Liang Y S and Li P. 2025. D-LDRNet：monocular vision framework merging prior LiDAR knowledge for all-weather safe monitoring of vehicle in transmission lines. IEEE Transactions on Intelligent Vehicles，10(5)：3318-3330［DOI：10.1109/TIV.2024.3451245］

Li Q，Zhang C Q，Hu Q H，Zhu P F，Fu H Z and Chen L. 2024 Stabilizing multispectral pedestrian detection with evidential hybrid fusion. IEEE Transactions on Circuits and Systems for Video Technology，34(4)：3017-3029［DOI：10.1109/TCSVT.2023.3306870］

Li S L，Yao R，Zhou Y，Zhu H C，Sun K Y，Liu B，Shao Z W and Zhao J Q. 2025. Modality-guided dynamic graph fusion and temporal diffusion for self-supervised RGB-T tracking//Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence（IJCAI-25）. Montreal，Canada：IJCAI：1422-1430［DOI：10.24963/ijcai.2025/159］

Li T H，Xu T F，Wang Y，Qin H L，Lin X and Li J N. 2025. MMOT：the first challenging benchmark for drone-based multispectral multi-object tracking［EB/OL］. https://arxiv.org/pdf/2510.12565.pdf

Li X Z，Li X C and Hu S Y. 2025. DARTer：dynamic adaptive representation tracker for nighttime UAV tracking//Proceedings of the 2025 International Conference on Multimedia Retrieval. Chicago，USA：ACM：1998-2002［DOI：10.1145/3731715.3733473］

Li X，Huang Y Q，He Z Y，Wang Y W，Lu H C and Yang M H. 2023. CiteTracker：correlating image and text for visual tracking//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision（ICCV）. Vancouver，Canada：IEEE：9940-9949［DOI：10.1109/ICCV51070.2023.00915］

Li X，Zha Y F，Zhang T Z，Cui Z，Zuo W M，Hou Z Q，Lu H C and Wang H Z. 2019. Survey of visual object tracking algorithms based on deep learning. Journal of Image and Graphics，24(12)：2057-2080（李玺，查宇飞，张天柱，崔振，左旺孟，侯志强，卢湖川，王菡子. 2019. 深度学习的目标跟踪算法综述. 中国图象图形学报，24(12)：2057-2080）［DOI：10.11834/jig.190372］

Li Y Z，Niu Y Z，Xu R and He Y Q. 2024. Zero-referenced enlightening and restoration for UAV nighttime vision. IEEE Geoscience and Remote Sensing Letters，21：#8002105［DOI：10.1109/LGRS.2024.3353731］

Li Z F，Xiong F C，Zhou J，Lu J F and Qian Y T. 2023. Learning a deep ensemble network with band importance for hyperspectral object tracking. IEEE Transactions on Image Processing，32：2901-2914［DOI：10.1109/TIP.2023.3263109］

Li Z F，Xiong F C，Zhou J，Lu J F，Zhao Z and Qian Y T. 2024. Material-guided multiview fusion network for hyperspectral object tracking. IEEE Transactions on Geoscience and Remote Sensing，62：#5509415［DOI：10.1109/TGRS.2024.3366536］

Liang J T，Zhou J Q，Li W，Wang Y，Hu T J and Wu Q. 2025. Reconstructed and simulated dataset for aerial RGBD tracking. IEEE Robotics and Automation Letters，10(2)：2008-2015［DOI：10.1109/LRA.2025.3526565］

Liang S Y，Chen P B，Wu S C and Cao H T. 2024. Complementary fusion of camera and LiDAR for cooperative object detection and localization in low contrast environments at night outdoors. IEEE Transactions on Consumer Electronics，70(3)：6392-6403［DOI：10.1109/TCE.2024.3436852］

Lin T Y，Goyal P，Girshick R，He K M and Dollar P. 2017. Focal loss

for dense object detection//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE: 2999-3007 [DOI: 10.1109/ICCV.2017.324]

Lin T Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P and Zitnick C L. 2014. Microsoft COCO: common objects in context//Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: Springer: 740-755 [DOI: 10.1007/978-3-319-10602-1_48]

Liu B, Jin J L, Zhang Y H and Sun C. 2025. WRRT-DETR: weather-robust RT-DETR for drone-view object detection in adverse weather. Drones, 9(5): #369 [DOI: 10.3390/drones9050369]

Liu K, Sun H, Wu H, Ji K F and Kuang G Y. 2025. Dynamic brightness reconstruction for UAV visible-infrared fusion object detection. Acta Aeronautica et Astronautica Sinica, 46(24): #631968 (刘奎, 孙浩, 伍瀚, 计科峰, 匡纲要. 2025. 动态亮度重建的无人机可见学-红外融合目标检测. 航空学报, 46(24): #631968) [DOI: 10.7527/S1000-6893.2025.31968]

Liu R S, Ma L, Zhang J A, Fan X and Luo Z X. 2021. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 10556-10565 [DOI: 10.1109/CVPR46437.2021.01042]

Liu S W, He H, Zhang Z C and Zhou Y T. 2024. LI-YOLO: an object detection algorithm for UAV aerial images in low-illumination scenes. Drones, 8(11): #653 [DOI: 10.3390/drones8110653]

Liu S, Liu D Y, Srivastava G, Połap D and Woźniak M. 2021. Overview and methods of correlation filter algorithms in object tracking. Complex & Intelligent Systems, 7: 1895-1917 [DOI: 10.1007/s40747-020-00161-4]

Liu T S, Lam K M, Zhao R and Qiu G P. 2022. Deep cross-modal representation learning and distillation for illumination-invariant pedestrian detection. IEEE Transactions on Circuits and Systems for Video Technology, 32(1): 315-329 [DOI: 10.1109/TCSVT.2021.3060162]

Liu W Y, Ren G F, Yu R S, Guo S, Zhu J k and Zhang L. 2022. Image-adaptive YOLO for object detection in adverse weather conditions//Proceedings of 36th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI: 1792-1800 [DOI: 10.1609/AAAI.v36i2.20072]

Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Y and Berg A C. 2016. SSD: single shot multibox detector//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, Netherlands: Springer: 21-37 [DOI: 10.1007/978-3-319-46448-0_2].

Liu Y, Mahmood A and Khan M H. 2024. NT-VOT211: a large-scale benchmark for night-time visual object tracking//Proceedings of the 17th Asian Conference on Computer Vision. Hanoi, Vietnam: Springer: 314-332 [DOI: 10.1007/978-981-96-0901-7_19]

Liu Z W, Yang N, Wang Y, Li Y K, Zhao X M and Wang F Y. 2024. Enhancing traffic object detection in variable illumination with RGB-Event fusion. IEEE Transactions on Intelligent Transportation Systems, 25(12): 20335-20350 [DOI: 10.1109/TITS.2024.3456108]

Liu Z, Hu H, Lin Y T, Yao Z L, Xie Z D, Wei Y X, Ning J, Cao Y, Zhang Z, Dong L, Wei F R and Guo B N. 2022. Swin transformer v2: scaling up capacity and resolution//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE: 11999-12009 [DOI: 10.1109/CVPR52688.2022.01170]

Loh Y P and Chan C S. 2019. Getting to know low-light images with the exclusively dark dataset. Computer Vision and Image Understanding, 178: 30-42 [DOI: 10.1016/j.cviu.2018.10.010]

Lore K G, Akintayo A and Sarkar S. 2017. LLNet: a deep autoencoder approach to natural low-light image enhancement. Pattern Recognition, 61: 650-662 [DOI: 10.1016/j.patcog.2016.06.008]

Lu A D, Wang W Y, Li C L, Tang J, Luo B. 2025. RGBT tracking via all-layer multimodal interactions with progressive fusion mamba//Proceedings of the 39th Annual AAAI Conference on Artificial Intelligence. Pennsylvania, USA: AAAI Press: 5793-5801 [DOI: 10.1609/AAAI.v39i6.32618]

Lu A D, Zhao J C, Li C L, Xiao Y and Luo B. 2024. Breaking modality gap in RGB-T tracking: coupled knowledge distillation//Proceedings of the 32nd ACM International Conference on Multimedia. Melbourne, Australia: ACM: 9291-9300 [DOI: 10.1145/3664647.3680878]

Marvasti-Zadeh S M, Cheng L, Ghanei-Yakhdan H and Kasaei S. 2023. Deep learning for visual tracking: a comprehensive survey. IEEE Transactions on Intelligent Transportation Systems, 23(5): 3943-3968 [DOI: 10.1109/TITS.2020.3046478]

Mazhar O, Babuška R and Kober J. 2021. GEM: glare or gloom, i can still see you-end-to-end multi-modal object detection. IEEE Robotics and Automation Letters, 6(4): 6321-6328 [DOI: 10.1109/LRA.2021.3093871]

Nam H and Han B. 2016. Learning multi-domain convolutional neural networks for visual tracking//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE: 4293-4302 [DOI: 10.1109/CVPR.2016.465]

Pan X Y, Jia N X, Mu Y Z and Gao X R. 2023. Survey of small object detection. Journal of Image and Graphics, 28(09): 2587-2615 (潘晓英, 贾凝心, 穆元震, 高炫蓉. 2023. 小目标检测研究综述. 中国图象图形学报, 28(09): 2587-2615) [DOI: 10.11834/jig.220455]

Paulin G, Sambolek S and Ivasic-Kos M. 2024. Application of raycast method for person geolocalization and distance determination using UAV images in real-world land search and rescue scenarios. Expert Systems with Applications, 237: #121495 [DOI: 10.1016/j.eswa.2023.121495]

Qin H L, Xu T F, Li T H, Chen Z X, Feng T and Li J N. 2025. MUST:

the first dataset and unified framework for multispectral UAV single object tracking//Proceedings of 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 16882-16891 [DOI: 10.1109/CVPR52734.2025.01573]

Redmon J and Farhadi A. 2017. YOLO9000: better, faster, stronger//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 6517-6525 [DOI: 10.1109/CVPR.2017.690]

Redmon J and Farhadi A. 2018. YOLOv3: an incremental improvement [EB/OL]. [2025-08-13]. https://arxiv.org/pdf/1804.02767.pdf

Redmon J, Divvala S, Girshick R and Farhadi A. 2016. You only look once: unified, real-time object detection//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE: 779-788 [DOI: 10.1109/CVPR.2016.91]

Ren S Q, He K M, Girshick R and Sun J. 2017. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6): 1137-1149 [DOI: 10.1109/TPAMI.2016.2577031]

Reza A M. 2004. Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology, 38: 35-44 [DOI: 10.1023/B: VLSI. 0000028532.53893.82]

Saffari A, Leistner C, Santner J, Godec M and Bischof H. 2009. On-line random forests//Proceedings of 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops. Kyoto, Japan: IEEE: 1393-1400 [DOI: 10.1109/ICCVW.2009.5457447]

Schutera M, Mostafa H, Abhau J, Mikut R and Reischl M. 2021. Night-to-day: online image-to-image translation for object detection within autonomous driving by night. IEEE Transactions on Intelligent Vehicles, 6(3): 480-489 [DOI: 10.1109/TIV.2020.3039456]

Shang X P, Li N N, Li D J, Lv J W, Zhao W and Zhang R F. 2025. CCLDET: a cross-modality and cross-domain low-light detector. IEEE Transactions on Intelligent Transportation Systems, 26(3): 3284-3294 [DOI: 10.1109/TITS.2024.3522086]

Shao Y H, Chen H L, Fu G, Wu Y D and Ren Z W. 2025. Fuse image enhancement with a regularized correlation filter for target tracking of UAVs. Journal of Image and Graphics, 30(10): 3302-3318 (邵延华, 陈慧玲, 付贵, 吴亚东, 任珍文. 2025. 融合图像增强的正则化相关滤波无人机目标跟踪. 中国图象图形学报, 30(10): 3302-3318) [DOI: 10.11834/jig.240576]

Shao Z K, Hu Y F, Fan B and Liu H M. 2025. PURA: parameter update-recovery test-time adaption for RGB-T tracking//Proceedings of 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 22089-22098 [DOI: 10.1109/CVPR52734.2025.02057]

Shen J F, Chen Y F, Liu Y, Zuo X, Fan H and Yang W K. 2024. ICAFusion: iterative cross-attention guided feature fusion for multispectral object detection. Pattern Recognition, 145: #109913 [DOI: 10.1016/j.patcog.2023.109913]

Shen J F, Zhan H B, Dong S H, Zuo X, Yang W K and Ling H B. 2025. Multispectral state-space feature fusion: bridging shared and cross-parametric interactions for object detection. Information Fusion, 127: #103895 [DOI: 10.1016/j.inffus.2025.103895]

Shen J F, Zhan H B, Zuo X, Fan H, Yuan X H, Li J and Yang W K. 2025. IRDFusion: iterative relation-map difference guided feature fusion for multispectral object detection [EB/OL]. [2025-11-24]. https://arxiv.org/pdf/2509.09085.pdf

Shen X Y, Li H B, Li Y Q and Zhang W M. 2025. LDWLE: self-supervised driven low-light object detection framework. Complex & Intelligent Systems, 11: #82 [DOI: 10.1007/s40747-024-01681-z]

Shi J B and Tomasi C. 1994. Good features to track//Proceedings of 1994 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 593-600 [DOI: 10.1109/CVPR.1994.323794]

Shi Z H, Wu C W, Li C J, You Z Z, Wang Q and Ma C C. 2023. Object detection techniques based on deep learning for aerial remote sensing images: a survey. Journal of Image and Graphics, 28(09): 2616-2643 (石争浩, 仵晨伟, 李成建, 尤珍臻, 王泉, 马城城. 2023. 航空遥感图像深度学习目标检测技术研究进展. 中国图象图形学报, 28(09): 2616-2643) [DOI:10.11834/jig.221085]

Siméoni O, Vo H V, Seitzer M, Baldassarre F, Oquab M, Jose C, Khalidov V, Szafraniec M, Yi S, Ramamonjisoa M, Massa F, Haziza D, Wehrstedt L, Wang J Y, Darcet T, Moutakanni T, Sentana L, Roberts C, Vedaldi A, Tolan J, Brandt J, Couprie C, Mairal J, Jégou H, Labatut P and Bojanowski P. 2025. Dinov3 [EB/OL]. [2025-11-24]. https://arxiv.org/pdf/2508.10104.pdf

Song J M, Meng C L and Ermon S. 2021. Denoising diffusion implicit models//Proceedings of International Conference on Learning Representations. [DOI: 10.48550/arXiv.2010.02502]

Sun C Y, Zhang J Q, Wang Y, Ge H L, Xia Q C, Yin B C and Yang X. 2025. Exploring historical information for RGBE visual tracking with mamba//Proceedings of 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 6500-6509 [DOI: 10.1109/CVPR52734.2025.00609]

Sun X Y, Zhu Y H and Huang H. 2025. Specificity-guided cross-modal feature reconstruction for RGB-infrared object detection. IEEE Transactions on Intelligent Transportation Systems, 26(1): 950-961 [DOI: 10.1109/TITS.2024.3495028]

Sun X, Yu Y H and Cheng Q. 2024. Low-rank multimodal remote sensing object detection with frequency filtering experts. IEEE Transactions on Geoscience and Remote Sensing, 62: #5637114 [DOI: 10.1109/TGRS.2024.3446814]

Sun Y M, Cao B, Zhu P F and Hu Q H. 2022. Drone-based RGB-infrared cross-modality vehicle detection via uncertainty-aware learning. IEEE Transactions on Circuits and Systems for Video Technology, 32(10): 6700-6713[DOI: 10.1109/TCSVT.2022.3168279]

Suo J S, Wang T Y, Zhang X Z, Chen H Y, Zhou W and Shi W S. 2023. HIT-UAV: a high-altitude infrared thermal dataset for unmanned aerial vehicle-based object detection. Scientific Data, 10(1): #227[DOI: 10.1038/s41597-023-02066-6]

Tang S Y, Zhang S and Fang Y N. 2024. HIC-YOLOv5: improved YOLOv5 for small object detection//Proceedings of 2024 International Conference on Robotics and Automation (ICRA). Yokohama, Japan: IEEE: 6614-6619[DOI: 10.1109/ICRA57147.2024.10610273]

Tang Z Y, Xu T Y, Li H, Wu X J, Zhu X F and Kittler J. 2023. Exploring fusion strategies for accurate RGBT visual object tracking. Information Fusion, 99: #101881[DOI: 10.1016/j.inffus.2023.101881]

Tang Z Y, Xu T Y, Zhu X F, Cheng C Y, Zhou T, Wu X J and Kittler J. 2025. Serial over parallel: learning continual unification for multi-modal visual object tracking and benchmarking//Proceedings of the 33rd ACM International Conference on Multimedia. Dublin, Ireland: ACM: 1229-1238[DOI: 10.1145/3746027.3754879]

Tarvainen A and Valpola H. 2017. Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc: 1195-1204[DOI: 10.5555/3294771.3294885]

Tian Z, Shen C H, Chen H and He T. 2019. FCOS: fully convolutional one-stage object detection//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE: 9627-9636[DOI: 10.1109/ICCV.2019.00972]

Vaswani A, Shazeer N, Parmar N, UszKoreit J, Jones L, Gomez A N, Kaiser L and Polosukhin I. 2017. Attention is all you need//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc: 6000-6010[DOI: 10.5555/3295222.3295349]

Vinoth K P S. 2024. Lightweight object detection in low light: pixel-wise depth refinement and TensorRT optimization. Results in Engineering, 23: #102510[DOI: 10.1016/j.rineng.2024.102510]

Wang A, Chen H, Liu L H, Chen K, Lin Z J, Han J G and Ding G G. 2024. YOLOv10: real-time end-to-end object detection//Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc: 107984-108011[DOI: 10.48550/arXiv.2405.14458]

Wang C C, He W, Nie Y, Guo J Y, Liu C J, Wang Y H and Han K. 2023. Gold-YOLO: efficient object detector via gather-and-distribute mechanism//Proceedings of the 37th International Conference on Neural Information Processing Systems. Los Angeles, USA: Curran Associates Inc: 51094-51112[DOI: 10.5555/3666122.3668346]

Wang C Y, Bochkovskiy A and Liao H Y M. 2023. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 7464-7475[DOI: 10.1109/CVPR52729.2023.00721]

Wang C Y, Liao H Y M, Wu Y H, Chen P Y, Hsieh J W and Yeh I H. 2020. CSPNet: a new backbone that can enhance learning capability of CNN//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, USA: IEEE: 1571-1580[DOI: 10.1109/CVPRW50498.2020.00203]

Wang C Y, Yeh I H and Liao H Y M. 2024. YOLOv9: learning what you want to learn using programmable gradient information//Proceedings of the 18th European Conference on Computer Vision. Milan, Italy: Springer: 1-21[DOI: 10.1007/978-3-031-72751-1_1]

Wang F S, Li F, Yin S S, Wang X, Sun F M and Zhu B. 2023. All-day and real-time multi-regularized correlation filter for UAV object tracking. ACTA AUTOMATICA SINICA, 49(11): 2409-2425(王法胜, 李富, 尹双双, 王星, 孙福明, 朱兵. 2023. 全天实时跟踪无人机目标的多正则化相关滤波算法. 自动化学报, 49(11): 2409-2425)[DOI: 10.16383/j.aas.c220424]

Wang H Y, Liu X T, Li Y F, Sun M, Yuan D and Liu J. 2024. Temporal adaptive RGBT tracking with modality prompt//Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI Press: 5436-5444[DOI: 10.1609/aaai.v38i6.28352]

Wang H Y, Wang C P, Fu Q, Si B Q, Zhang D D, Kou R K, Yu Y and Feng C F. 2024. YOLOFIV: object detection algorithm for around-the-clock aerial remote sensing images by fusing infrared and visible features. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17: 15269-15287[DOI: 10.1109/JSTARS.2024.3447649]

Wang H Y, Wang C P, Fu Q, Si B Q, Zhang D D, Kou R K, Yu Y and Feng C F. 2025. MINIAOD: lightweight aerial image object detection. IEEE Sensors Journal, 25(5): 9167-9184[DOI: 10.1109/JSEN.2025.3530076]

Wang H Y, Wang C P, Fu Q, Zhang D D, Kou R K and Yu Y. 2024. Cross-modal oriented object detection of UAV aerial images based on image feature. IEEE Transactions on Geoscience and Remote Sensing, 62: #5403021[DOI: 10.1109/TGRS.2024.3367934]

Wang J W, Xu C, Yang W and Yu L. 2022. A normalized gaussian wasserstein distance for tiny object detection[EB/OL]. [2025-08-13]. https://arxiv.org/pdf/2110.13389.pdf

Wang K Y, Fu X Y, Ge C J, Cao C Z and Zha Z J. 2024. Towards generalized UAV object detection: a novel perspective from frequency domain disentanglement. International Journal of Computer Vision,

132：5410-5438［DOI：10.1007/s11263-024-02108-5］

Wang M Y，Wang Q H，Wang Z，Gao Y M，Wang J P，Cui C，Li Y，Ding Z M，Wang K W，Xu C and Gao F. 2025. Unlocking aerobatic potential of quadcopters：autonomous freestyle flight generation and execution. Science Robotics，10(101)：#9905［DOI：10.1126/scirobotics.adp9905］

Wang Q W，Chi Y K，Shen T，Song J，Zhang Z F and Zhu Y. 2022. Improving RGB-infrared object detection by reducing cross-modality redundancy. Remote Sensing，14(9)：#2020［DOI：10.3390/rs14092020］

Wang W J，Peng Y P，Cao G Z，Guo X Q and Kwok N. 2021. Low-illumination image enhancement for night-time UAV pedestrian detection. IEEE Transactions on Industrial Informatics，17(8)：5208-5217［DOI：10.1109/TII.2020.3026036］

Wang X Z，Ma K，Liu Q K，Zou Y H and Fu Y. 2024. Multi-object tracking in the dark//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition（CVPR）. Seattle，USA：IEEE：382-392［DOI：10.1109/CVPR52733.2024.00044］

Wang X，Shu X J，Zhang S L，Jiang B，Wang Y W，Tian Y H and Wu F. 2023. MFGNet：dynamic modality-aware filter generation for RGB-T tracking. IEEE Transactions on Multimedia，25：4335-4348［DOI：10.1109/TMM.2022.3174341］

Wang X，Wang S A，Tang C M，Zhu L，Jiang B and Tian Y H. 2024. Event stream-based visual object tracking：a high-resolution benchmark dataset and a novel baseline//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition（CVPR）. Seattle，USA：IEEE：19248-19257［DOI：10.1109/CVPR52733.2024.01821］

Wang Y C，Fu C H，Lu K H，Yao L L and Zuo H B. 2024. Conditional generative denoiser for nighttime UAV tracking//Proceedings of 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems（IROS）. Abu Dhabi，United Arab Emirates：IEEE：12971-12978［DOI：10.1109/IROS58592.2024.10802714］

Wei H R，Fu Y Y，Wang D Y，Guo R，Zhao X Y and Fang J. 2024. Unsupervised nighttime object tracking based on transformer and domain adaptation fusion network. IEEE Access，12：130896-130913［DOI：10.1109/ACCESS.2024.3378117］

Wu G Y，Wu Q H，Jiang T，Chen H N，Zhao Y X，Qu Y B，Hu J S and Guo Y. 2025. A selective attention approach for cross-modal target perception. Sci Sin Inform，55(10)：2471-2490（吴光宇，吴启晖，江涛，陈好男，赵宇心，屈毓锛，胡金水，郭耀. 2025. 面向跨模态目标感知的选择性注意力方法. 中国科学：信息科学），55(10)：2471-2490［DOI：10.1360/SSI-2025-0107］

Wu H J，Yao S Y，Huang F，Wang S，Zhang L C，Zheng Z R and Ren W Q. 2025. LVPTrack：high performance domain adaptive UAV tracking with label aligned visual prompt tuning//Proceedings of the 39th Annual AAAI Conference on Artificial Intelligence. Pennsylvania，USA：AAAI Press：8395-8403［DOI：10.1609/aaai.v39i8.32906］

Wu Y，Lim J and Yang M H. 2013. Online object tracking：a benchmark//Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland，USA：IEEE：2411-2418［DOI：10.1109/CVPR.2013.312］

Wu Y，Yang X Y，Wang X C，Ye H Z，Zeng D and Li S W. 2025. MambaNUT：nighttime UAV tracking via Mamba-based adaptive curriculum learning［EB/OL］.［2025-08-13］. https://arxiv.org/pdf/2412.00626v3.pdf

Xi Y，Jia W J，Miao Q G，Feng J M，Ren J C and Luo H. 2024. Detection-driven exposure-correction network for nighttime drone-view object detection. IEEE Transactions on Geoscience and Remote Sensing，62：#5605014［DOI：10.1109/TGRS.2024.3351134］

Xiao Y，Cao D，Li C L，Jiang B and Tang J. 2025. A benchmark dataset for high-altitude UAV multi-modal tracking. Journal of Image and Graphics，30(02)：0361-0374（肖云，曹丹，李成龙，江波，汤进. 2025. 基于高空无人机平台的多模态跟踪数据集. 中国图象图形学报，30(02)：0361-0374）［DOI：10.11834/jig.240040］

Xiao Y，Yang M M，Li C L，Liu L and Tang J. 2022. Attribute-based progressive fusion network for RGBT tracking//Proceedings of the 36th AAAI Conference on Artificial Intelligence. Virtually：AAAI Press：2831-2838［DOI：10.1609/AAAI.v36i3.20187］

Xie J，Anwer R M，Cholakkal H，Nie J，Cao J L，Laaksonen J and Khan F S. 2022. Learning a dynamic cross-modal network for multispectral pedestrian detection//Proceedings of the 30th ACM International Conference on Multimedia. Lisboa，Portugal：ACM：4043-4052［DOI：10.1145/3503161.3547895］

Xie J，Nie J，Ding B N，Yu M Y and Cao J L. 2023. Cross-modal local calibration and global context modeling network for RGB-infrared remote-sensing object detection. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing，16：8933-8942［DOI：10.1109/JSTARS.2023.3315544］

Xu X G，Wang R X，Fu C W and Jia J Y. 2022. SNR-aware low-light image enhancement//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition（CVPR）. New Orleans，USA：IEEE：17693-17703［DOI：10.1109/CVPR52688.2022.01719］

Xue C，Xia Y L，Wu M J，Chen Z Q，Cheng F Y and Yun L J. 2024. EL-YOLO：an efficient and lightweight low-altitude aerial objects detector for onboard applications. Expert Systems with Applications，256：#124848［DOI：10.1016/j.eswa.2024.124848］

Yang C H，Huang Z H and Wang N Y. 2022. QueryDet：cascaded sparse query for accelerating high-resolution small object detection//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition（CVPR）. New Orleans，USA：IEEE：13658-13667［DOI：10.1109/CVPR52688.2022.01330］

Yang F，Fan H，Chu P，Blasch E and Ling H B. 2019. Clustered object detection in aerial images//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision（ICCV）. Seoul，Korea：

IEEE: 8310-8319 [DOI: 10.1109/ICCV.2019.00840]

Yang F, Liang B B, Li W and Zhang J W. 2025. Multidimensional fusion network for multispectral object detection. IEEE Transactions on Circuits and Systems for Video Technology, 35(1): 547-560 [DOI: 10.1109/TCSVT.2024.3454631]

Yang J J, Fan L W, Zhang J P, Lian X K, Shen H and Hu D W. 2025. Fully spiking neural networks for unified frame-event object tracking[EB/OL]. [2025-11-24]. https://arxiv.org/pdf/2505.20834.pdf

Yang J Y, Gao S, Li Z, Zheng F and Leonardis A. 2023. Resource-efficient RGBD aerial tracking//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 13374-13383 [DOI: 10.1109/CVPR52729.2023.01285]

Yang J Y, Li Z, Zheng F, Leonardis A and Song J K. 2022. Prompting for multi-modal tracking//Proceedings of the 30th ACM International Conference on Multimedia. Lisboa, Portugal: ACM: 3492-3500 [DOI: 10.1145/3503161.3547851]

Yang L J, Yu M, Fang L J, Yang Y and Yue Y F. 2025. CDMFusion: RGB-T image fusion based on conditional diffusion models via few denoising steps in open environments//Proceedings of IEEE International Conference on Robotics and Automation (ICRA). Atlanta, GA, USA: IEEE: 6314-6320 [DOI: 10.1109/ICRA55743.2025.11128410]

Yao L L, Fu C H, Wang Y H, Zuo H B and Lu K H. 2024. Enhancing nighttime UAV tracking with light distribution suppression//Proceedings of 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Abu Dhabi, United Arab Emirates: IEEE: 5902-5909 [DOI: 10.1109/IROS58592.2024.10802568]

Yao S Y, Zhu R, Wang Z Q, Ren W Q, Yan Y Y and Cao X C. 2025. UMDATrack: unified multi-domain adaptive tracking under adverse weather conditions//Proceedings of 2025 IEEE/CVF International Conference on Computer Vision (ICCV). Hawaii, USA: IEEE: 6466-6475

Yao Z K, Liu Q, Zhao Z Z, Qin Y L, Zhu J L, Xia T Z, Li B and Wang L P. 2025. Night-time traffic light recognition based on enhancement-guided object detection. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 55(6): 4410-4422 [DOI: 10.1109/TSMC.2025.3552621]

Ye J J, Fu C H, Cao Z, An S, Zheng G Z and Li B W. 2022. Tracker meets night: a transformer enhancer for UAV tracking. IEEE Robotics and Automation Letters, 7(2): 3866-3873 [DOI: 10.1109/LRA.2022.3146911]

Ye J J, Fu C H, Zheng G Z, Cao Z and Li B W. 2021. Darklighter: light up the darkness for UAV tracking//Proceedings of 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Prague, Czech Republic: IEEE: 3079-3085 [DOI: 10.1109/IROS51168.2021.9636680]

Ye J J, Fu C H, Zheng G Z, Paudel D P and Chen G. 2022. Unsupervised domain adaptation for nighttime aerial tracking//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE: 8886-8895 [DOI: 10.1109/CVPR52688.2022.00869]

Ye T Z, Dong L, Xia Y Q, Sun Y T, Zhu Y, Huang G and Wei F R. 2025. Differential transformer[EB/OL]. [2025-11-25]. https://arxiv.org/pdf/2410.05258.pdf

Yuan M X and Wei X X. 2024. C²former: calibrated and complementary transformer for RGB-infrared object detection. IEEE Transactions on Geoscience and Remote Sensing, 62: #5403712 [DOI: 10.1109/TGRS.2024.3376819]

Yuan X, Cheng G, Li G, Dai W, Yin W X, Feng Y C, Yao X W, Huang Z L, Sun X and Han J W. 2023. Progress in small object detection for remote sensing images. Journal of Image and Graphics, 28(06): 1662-1684 (袁翔, 程塨, 李戈, 戴威, 尹文昕, 冯瑛超, 姚西文, 黄钟泠, 孙显, 韩军伟. 2023. 遥感影像小目标检测研究进展. 中国图象图形学报, 28(06): 1662-1684) [DOI: 10.11834/jig.221202]

Yuan Z K, Zeng J, Wei Z X, Jin L J, Zhao S J, Liu X H, Zhang Y Y and Zhou G J. 2023. CLAHE-based low-light image enhancement for robust object detection in overhead power transmission system. IEEE Transactions on Power Delivery, 38(3): 2240-2243 [DOI: 10.1109/TPWRD.2023.3269206]

Zhang C H, Huang G J, Liu L, Huang S, Yang Y N, Wan X, Ge S M and Tao D C. 2023. WebUAV-3M: a benchmark for unveiling the power of million-scale deep UAV tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(7): 9186-9205 [DOI: 10.1109/TPAMI.2022.3232854]

Zhang C H, Liu L, Wen H, Zhou X and Wang Y F. 2025. MambaTrack: exploiting dual-enhancement for night UAV tracking//Proceedings of ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Hyderabad, India: IEEE: 1-5 [DOI: 10.1109/ICASSP49660.2025.10890855]

Zhang F, Li Y, You S D and Fu Y. 2021. Learning temporal consistency for low light video enhancement from single images//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 4965-4974 [DOI: 10.1109/CVPR46437.2021.00493]

Zhang H, Li F, Liu S L, Zhang L, Su H, Zhu J, Ni L and Shen X Y. 2022. DINO: detr with improved denoising anchor boxes for end-to-end object detection[EB/OL]. [2025-11-24]. https://arxiv.org/pdf/2203.03605.pdf

Zhang J P, Li Z W, Wei R N and Wang Y H. 2023. Progressive domain-style translation for nighttime tracking//Proceedings of the 31st ACM International Conference on Multimedia. Ottawa, Canada: ACM: 7324-7334 [DOI: 10.1145/3581783.3612305]

Zhang N, Chai B R, Song J M, Tian T, Zhu P Y, Ma J Y and Tian J W. 2025. Omni-scene infrared vehicle detection: an efficient selective aggregation approach and a unified benchmark. ISPRS Journal of

Photogrammetry and Remote Sensing, 223: 244-260 [DOI: 10.1016/j.isprsjprs.2025.03.002]

Zhang P Y, Wang D, Lu H C and Yang X Y. 2021. Learning adaptive attribute-driven representation for real-time RGB-T tracking. International Journal of Computer Vision, 129: 2714-2729 [DOI: 10.1007/s11263-021-01495-3]

Zhang P Y, Zhao J, Wang D, Lu H C and Ruan X. 2022. Visible-thermal UAV tracking: a large-scale benchmark and new baseline// Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE: 8876-8885 [DOI: 10.1109/CVPR52688.2022.00868]

Zhang R H, Peng J T, Guo W T, Ma Y H, Chen J Z, Hu H Y, Li W H, Yin G D and Li Z W. 2024. A robust and real-time lane detection method in low-light scenarios to advanced driver assistance systems. Expert Systems with Applications, 256: #124923 [DOI: 10.1016/j.eswa.2024.124923]

Zhang T L, Debattista K, Zhang Q, Ding G G and Han J G. 2024. Revisiting motion information for RGB-Event tracking with MOT philosophy//Advances in Neural Information Processing Systems 37 (NeurIPS 2024. Vancouver, Canada: Cambridge University Press: 89346--89372 [DOI: 10.52202/079017-2836]

Zhang X, Wei F, Wang Y, Zhao W D, Li F Y and Chu X X. 2025. UPRE: zero-shot domain adaptation for object detection via unified prompt and representation enhancement//Proceedings of 2025 IEEE/CVF International Conference on Computer Vision (ICCV). Hawaii, USA: IEEE: 508-518

Zhang Y H, Guo X J, Ma J Y, Liu W and Zhang J W. 2021. Beyond brightening low-light images. International Journal of Computer Vision, 129: 1013-1037 [DOI: 10.1007/s11263-020-01407-x]

Zhang Y H, Zhang J W and Guo X J. 2019. Kindling the darkness: a practical low-light image enhancer//Proceedings of the 27th ACM International Conference on Multimedia. Nice, France: ACM: 1632-1640 [DOI: 10.1145/3343031.3350926]

Zhang Z P, Peng H W, Fu J L, Li B and Hu W M. 2020. Ocean: object-aware anchor-free tracking//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 771-787 [DOI: 10.1007/978-3-030-58589-1_46]

Zhao D W, Shao F M, Zhang S, Yang L, Zhang H, Liu S D and Liu Q. 2024. Advanced object detection in low-light conditions: enhancements to YOLOv7 framework. Remote Sensing, 16(23): #4493 [DOI: 10.3390/rs16234493]

Zhao H J, Wang D and Lu H C. 2023. Representation learning for visual object tracking by masked appearance transfer//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 18696-18705 [DOI: 10.1109/CVPR52729.2023.01793]

Zhao H S, Shi J P, Qi X J, Wang X G and Jia J Y. 2017. Pyramid scene parsing network//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 6230-6239 [DOI: 10.1109/CVPR.2017.660]

Zhao T Y, Liu B Y, Gao Y L, Sun Y M, Yuan M X and Wei X X. 2025. Rethinking multi-modal object detection from the perspective of mono-modality feature learning[EB/OL]. [2025-08-13]. https://arxiv.org/pdf/2503.11780.pdf

Zhao Z C, Zhang W, Xiao Y, Li C L and Tang J. 2025. Reflectance-guided progressive feature alignment network for all-day UAV object detection. IEEE Transactions on Geoscience and Remote Sensing, 63: #5404215 [DOI: 10.1109/TGRS.2025.3574963]

Zhao Z Q, Feng P, Guo J J, Yuan C H, Wang T J, Liu F, Zhao Z J, Cui Z M and Wu B. 2018. A hybrid tracking framework based on kernel correlation filtering and particle filtering. Neurocomputing, 297: 40-49 [DOI: 10.1016/j.neucom.2018.02.043]

Zheng Y, Zheng C L, Zhang X Y, Chen F, Chen Z and Zhao S Y. 2023. Detection, localization, and tracking of multiple MAVs with panoramic stereo camera networks. IEEE Transactions on Automation Science and Engineering, 20(2): 1226-1243 [DOI: 10.1109/TASE.2022.3176294]

Zhong H, Zhang Y, Shi Z G, Zhang Y and Zhao L.2025. PS-YOLO: a lighter and faster network for UAV object detection. Remote Sensing, 17(9): #1641 [DOI: 10.3390/rs17091641]

Zhong P Z, Guo X Y, Huang D F, Peng X J, Li Y, Zhao Q J and Li S W. 2024. Low-light object tracking: a benchmark[EB/OL]. [2025-08-13]. https://arxiv.org/pdf/2408.11463.pdf

Zhou Y, Yang X, Zhang G F, Wang J B, Liu Y Y, Hou L P, Jiang X, Liu X Z, Yan J C, Lyu C Q, Zhang W W and Chen K. 2022. MMRotate: a rotated object detection benchmark using Pytorch// Proceedings of the 30th ACM International Conference on Multimedia. Lisbon, Portugal: ACM: 7331-7334 [DOI: 10.1145/3503161.3548541]

Zhu J W, Lai S M, Chen X, Wang D and Lu H C. 2023. Visual prompt multi-modal tracking//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 9516-9526 [DOI: 10.1109/CVPR52729.2023.00918]

Zhu J W, Tang H Y, Cheng Z Q, He J Y, Luo B, Qiu S H, Li S M and Lu H C. 2024. DCPT: darkness clue-prompted tracking in nighttime UAVs//Proceedings of 2024 IEEE International Conference on Robotics and Automation (ICRA). Yokohama, Japan: IEEE: 7381-7388 [DOI: 10.1109/ICRA57147.2024.10610544]

Zhu P F, Wen L Y, Du D W, Bian X, Fan H, Hu Q H and Ling H B. 2022. Detection and tracking meet drones challenge. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(11): 7380-7399 [DOI: 10.1109/TPAMI.2021.3119563]

Zhu X F, Xu T Y, Pan Y F, Gu J J, Li X, Lu J W, Wu X J and Kittler J. 2025. Collaborating vision, depth, and thermal signals for multi-modal tracking: dataset and algorithm[EB/OL]. [2025-11-24]. https://arxiv.org/pdf/2509.24741.pdf

Zhu Z Q, Gao X B, Lu W, Li J, Wang Z Y and Ge M Q. 2025. DPTrack: drectional kernel-guided prompt learning for robust nighttime aerial tracking[EB/OL]. [2025-11-24]. https://arxiv.org/pdf/2510.15449.pdf

Zuo H B, Fu C H, Zheng G Z, Yao L L, Lu K H and Pan J. 2024. DaDiff: domain-aware diffusion model for nighttime UAV tracking// Proceedings of 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Abu Dhabi, United Arab Emirates: IEEE: 11094-11101 [DOI: 10.1109/IROS58592.2024.10802294]

作者简介

王志祥,男,硕士研究生,主要研究方向为目标检测。E-mail: 22403010147@wit.edu.com

张梓阳,男,硕士研究生,主要研究方向为低光图像增强。E-mail:1006400705@qq.com

洪汉玉,男,教授,主要研究方向为导航与制导、模式识别、智能控制系统。E-mail:hhyhong@163.com

桑农,男,教授,主要研究方向为模式识别和计算机视觉。E-mail:nsang@hust.edu.cn