中图法分类号:TP391　　**文献标识码**: A　　**文章编号**: 1006-8961(2025)02-0361-14

论文引用格式: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]

基于高空无人机平台的多模态跟踪数据集

肖云[1,2],曹丹[3],李成龙[1,2*],江波[3],汤进[3]

1. 安徽大学人工智能学院,合肥　230601; 2. 安全人工智能安徽省重点实验室,合肥　230601;
3. 安徽大学计算机科学与技术学院,合肥　230601

摘　要: **目的**　无人机(unmanned aerial vehicle,UAV)因易操纵、灵活等特点,近年来在军事和民用等多个领域得到广泛应用。相对于低空无人机,高空无人机具有更广的视野,更强的隐蔽性,在情报侦察、灾害救援等方面具有更高的应用价值。然而,现有无人机多模态目标跟踪研究主要针对低空无人机,缺乏高空无人机多模态目标跟踪数据集,限制了该领域的研究和发展。**方法**　构建了一个用于评估高空无人机多模态目标跟踪方法的数据集HiAl(high altitude UAV multi-modal tracking dataset),该数据集主要由搭载混合传感器的无人机在500 m高空拍摄的可见光—红外多模态视频构成,两种模态数据经过精确配准和帧级标注,可以较好地评估不同多模态目标跟踪方法在高空无人机平台下的性能表现。**结果**　将主流的12种多模态跟踪方法在所提数据集与非高空无人机场景数据集上的表现进行了比较,方法TBSI(template-bridged search region interaction)在RGBT234数据集(RGB-thermal dataset)上PR(precision rate)值达到0.871,而在本文所提数据集上仅0.527,下降了39.5%,其SR(success rate)值由RGBT234数据集上的0.637,下降到本文所提数据集上的0.468,下降了26.5%。方法HMFT(hierarchical multi-modal fusion tracker)在所提数据集上的PR与RGBT234相比下降了23.6%,SR下降了14%。此外,利用HiAl数据集对6个方法进行重新训练实验,所有重训练方法的性能均得到提升。**结论**　本文提出一个基于高空无人机平台的多模态目标跟踪数据集,旨在促进多模态目标跟踪在高空无人机上的应用研究。HiAl数据集的在线发布地址为:https://github.com/mmic-lcl/Datasets-and-benchmark-code/tree/main。

关键词: 多模态目标跟踪;高空无人机;微小目标;高质量配准;数据集

A benchmark dataset for high-altitude UAV multi-modal tracking

Xiao Yun[1,2], Cao Dan[3], Li Chenglong[1,2*], Jiang Bo[3], Tang Jin[3]

1. School of Artificial Intelligence, Anhui University, Hefei　230601, China; 2. Anhui Provincial Key Laboratory of Security Artificial Intelligence, Hefei　230601, China; 3. School of Computer Science and Technology, Anhui University, Hefei　230601, China

Abstract: **Objective** Unmanned aerial vehicles (UAVs) have become crucial tools in both modern military and civilian contexts owing to their flexibility and ease of operation. High-altitude UAVs provide unique and distinct advantages over low-altitude UAVs, such as wider fields of view and stronger concealment, making them highly valuable in intelligence

收稿日期:2024-01-24;**修回日期**:2024-05-22;**预印本日期**:2024-05-29

*通信作者:李成龙　lcl1314@foxmail.com

基金项目:国家自然科学基金项目(U20B2068,62376004);安徽省高校协同创新项目(GXXT-2021-038);安徽省自然科学基金项目(2208085J18,2408085MF153);安徽省重点研发计划资助(2022i01020014)

Supported by: National Natural Science Foundation of China(U20B2068,62376004); University Synergy Innovation Program of Anhui Province(GXXT-2021-038); Natural Science Foundation of Anhui Province, China(2208085J18, 2408085MF153); Anhui Provincial Key Research and Development Program,China(2022i01020014)

reconnaissance, emergency rescue, and disaster relief tasks. However, tracking objects with high-altitude UAVs introduces considerable challenges, including UAV rotation, tiny objects, complex background changes, and low object resolution. The current research on multi-modal object tracking of UAVs focuses primarily on low-altitude UAVs, such as the dataset named VTUAV (visible-thermal UAV) for multi-modal object tracking of UAVs, which is shot in low-altitude airspace of 5–20 m and can fully show the unique perspective of UAVs. However, the scenes captured by high-altitude UAVs significantly differ from those captured by low-altitude UAVs. Thus, this dataset cannot provide strong support for the development of high-altitude UAV multi-modal object tracking, which is also the bottleneck of the lack of data support in the research field of multi-modal object tracking of high-altitude UAVs. Given the lack of an evaluation dataset to evaluate the multi-modal object tracking method of high-altitude UAVs, this limitation hinders research and development in this field. **Method** This study proposes an evaluation dataset named HiAl specifically for multi-modal object tracking methods of high-altitude UAVs captured at approximately 500 m. The UAV shooting this dataset is equipped with a hybrid sensor, which can capture video in both visible and infrared modes. The collected multimodal videos with high-quality videos were registered to provide a higher level of ground truth annotation and evaluate different multi-modal object tracking methods more fairly. First, the two video modalities were manually aligned to ensure that the same tracking object in each pair of videos occupied the same position within the frame. During the registration process, ensuring accurate registration of the area where the tracking object is located is the top priority, and under this premise, other areas in the image also become roughly aligned. Then, accurate ground truth annotations are provided to each frame of the video on the basis of the high alignment of the two modalities. The horizontal annotation boxes were used to label the position of the target in a way that best fits the contour of the tracked object. In the abovementioned modal alignment, two video modalities can share the same ground truth, which allows better evaluation of different multi-modal object tracking methods under the high-altitude UAV platform. Tracking attributes, scenes, and object categories was comprehensively considered during the data collection process to ensure the diversity and authenticity of the dataset. The dataset considers different lighting conditions and weather factors, including night and foggy days, for nine common object categories in high-altitude UAV scenes. The dataset has 12 tracking attributes; two are unique to UAVs, which have rich practical significance and high challenges. In contrast to existing multimodal tracking datasets, this dataset tracks mostly small targets, which is also a realistic challenge associated with high-altitude UAV shooting. **Result** The performances of 10 mainstream multi-modal tracking methods on this dataset are compared with those on a nonhigh-altitude UAV scene dataset. This study employs common quantitative evaluation metrics, namely, the precision rate (PR) and success rate (SR), to assess the performance of each method. Taking the two outstanding methods as examples, the PR value of the template-bridged search region interaction (TBSI) method on the RGB-thermal dataset (RGBT234) reached 0.871, whereas it was only 0.527 on the dataset proposed in this study, which decreased by 39.5%; its SR value decreased from 0.637 on RGBT234 to 0.468 on the dataset proposed in this study, which decreased by 26.5%. Compared with those of RGBT234, the PR and SR of the hierarchical multi-modal fusion tracker (HMFT) on the HiAl dataset decreased by 23.6% and 14%, respectively. In addition, the dataset HiAlto was used to retrain six methods. Comparative results indicated improved performance of all the retraining methods. For example, the PR value of the duality-gated mutual condition network (DMCNet) is increased from 0.485 before training to 0.524, and the SR value increased from 0.512 before training to 0.526. These experimental results reflect the high challenge and necessity of the dataset. **Conclusion** An evaluation dataset designed to assess the performance of multi-modal object tracking methods for high-altitude UAVs is introduced in this study. The multimodal data collected in the real scene and provided frame-level ground truth annotations were carefully registered to provide a dedicated dataset for high-quality multi-modal tracking of high-altitude UAVs. This proposed dataset HiAl can serve as a'standard evaluation tool for future research, offering researchers access to authentic and varied data to evaluate their algorithms performance. The experimental results of 10 mainstream tracking algorithms in HiAl with other datasets were compared, and the experimental results of retraining 6 tracking algorithms, including the limitations of existing algorithms in the multi-modal object tracking task of high-altitude UAVs, were analyzed. The potential research directions were extracted for researchers' reference. The HiAl dataset is available at https://github.com/mmic-lcl/Datasets-and-benchmark-code/tree/main.

Key words: multi-modal object tracking; high altitude unmanned aerial vehicle; small object; high quality alignment; dataset

0 引 言

无人机(unmanned aerial vehicle, UAV)具有体积小、成本低、灵活机动和观测视野广等特点,在情报侦察、敌方目标跟踪、电力巡检和灾害救援等军事和民用领域得到了广泛的应用。无人机目标跟踪任务是诸多应用领域中的研究热点之一,有着重要的研究意义(卓力 等,2021)。然而,基于可见光成像的无人机目标跟踪易受到云雾、光照条件等环境因素的影响,遭遇云雾遮挡、低光照和光照变化等挑战,造成视觉无人机目标跟踪方法的性能下降(Li 等,2023)。为了更有效地应对复杂环境下的跟踪挑战,研究人员引入红外、深度等多种模态数据,利用不同模态的互补优势,显著提高跟踪的鲁棒性。高空无人机多模态目标跟踪指的是利用高空无人机捕捉可见光和红外、深度等多模态视频,在给定初始目标框的情况下,对目标进行持续跟踪的过程。其中,依据无人机实际应用场景与《无人驾驶航空器飞行管理暂行条例》将 120～500 m 空域划分为高空,而 120 m 以下空域则称为低空。

数据集是推动科学研究发展的重要组成部分,Zhang 等人(2022b)提出了一个基于无人机的多模态目标跟踪数据集,其中有 500 个视频序列,应用场景包括长时跟踪和短时跟踪,以及目标分割预测。此外,该数据集场景多样,目标类型丰富,包含多种挑战类型,很大程度上缓解了无人机多模态目标跟踪领域数据集的匮乏问题。然而,该数据集主要拍摄于 5～20 m 低空,其拍摄场景与高空无人机在观测视野、环境复杂性等方面存在显著差异,难以充分体现高空无人机的独特优势,如观测视野广、隐蔽性高等。此外,由于该数据集拍摄于低空,连续帧之间目标位置和尺寸变化小,因此采用了每隔 10 帧对目标进行准确真值标注的稀疏标注方式,这种方法大幅降低了标注成本,但也在一定程度上牺牲了标注的准确性,并减少了对跟踪算法的挑战性。

针对上述不足,本文提出一个基于高空无人机平台的多模态目标跟踪数据集(high altitude UAV multi-modal tracking dataset, HiAl),旨在深化目标跟踪技术的研究,并推动多模态目标跟踪在无人机上的应用。其重点关注 9 类目标对象,包括 12 种挑战属性,其主要贡献如下:1)构建了一个基于高空无人机平台的多模态跟踪数据集,主要由无人机在 500 m 高空拍摄获得,包括 150 对可见光和红外视频对,总共超过 12 K 帧对。2)对所提数据集进行了精细的模态配准,并手工标注所有帧的目标边界框,以确保高质量的跟踪目标框标注。采用模态统一的标注方式有助于更准确地评估多模态跟踪方法的性能。3)在该数据集上对 12 种不同的多模态跟踪算法进行了综合评价和分析,分析现有算法在高空无人机多模态跟踪任务上的局限。

1 相关工作

1.1 无人机目标跟踪数据集

1.1.1 无人机视觉目标跟踪数据集

构建规范标准的数据集是无人机目标跟踪任务的重要组成部分。表 1 是现有主流无人机目标跟踪数据集的对比结果。UAV123 是由 Mueller 等人(2016)提出的无人机视觉目标跟踪数据集,包含 20 个长时视频序列和 123 个短时视频序列,总帧数约 113 K,是无人机目标跟踪领域的代表性数据集之一。Li 和 Yeung (2017) 提出 DTB70 (drone tracking benchmark),该数据集包括 70 个短时视频序列,总帧数约 15.8 K 帧,包括 11 种挑战属性。UAVDT(UAV detection and tracking benchmark)是 Du 等人(2018)提出的用于多种视觉任务的数据集,该数据集针对单目标跟踪任务提供 50 个视频序列。Zhu 等人(2021)提出 VisDrone(vision meets drones)数据集,包括 192 个视频序列,总帧数达到 221.9 K 帧,包括 10 种目标类别和 12 种挑战属性。UAVDark135(UAV dark tracking benchmark)是由 Ye 等人(2021)提出的用于夜间无人机目标跟踪的基准数据集,包括 135 个视频序列,总帧数约 125.4 K。2022 年,Zhang 等人(2022a)提出的 WebUAV-3M 数据集包括 4 500 个视频,总帧数达到 3.3 M,数据集中有 223 类目标对象,涵盖 17 种挑战。

1.1.2　无人机多模态目标跟踪数据集

为了克服可见光成像的局限，研究人员将多模态数据引入无人机目标跟踪任务。目前，多模态跟踪数据集主要包括RGB-T（RGB-thermal）和RGB-D（RGB-depth）两类。针对RGB-D目标跟踪，Yang等人（2023）首次提出基于无人机的RGB-D目标跟踪数据集D^2Cube，该数据集主要拍摄于2～5 m的超低空，包括手、球、杯子等34种类别，并标注了18种挑战属性，为无人机RGB-D目标跟踪研究提供充足的数据支持。本文主要针对RGB-T多模态无人机目标跟踪，下面重点介绍该领域现有的数据集。Jiang等人（2021）提出了一个多模态目标跟踪数据集Anti-UAV，收集了318对可见光与红外视频，涵盖6种无人机类别，主要来自大疆和派诺特，提供了7种挑战属性，对反无人机任务的研究具有重要价值。为了进一步推动无人机多模态目标跟踪领域的发展，Zhang等人（2022）提出了VTUAV数据集，包括500对可见光与红外视频序列，总帧数达到1.7 M，涵盖13种目标类别和13种挑战属性，并为视频提供帧级挑战标签，有助于促进不同挑战类别的多模态目标跟踪研究。VTUAV数据集可用于长时目标跟踪、短时目标跟踪和目标分割3种子任务，拓展了多模态目标跟踪的研究空间。然而，该数据集主要拍摄于5～20 m低空，目标尺寸较大，且与高空无人机视角存在显著差异，缺少高空无人机拍摄的多模态目标跟踪数据集在一定程度上限制了该领域的发展。

表1　无人机目标跟踪数据集对比表

Table 1　Comparison of UAV object tracking datasets

模态	数据集名称	样本尺寸/像素	高空	目标类别数	挑战属性数	标注方式	标注	配准	提出年份
RGB	UAV123	1 280×720	–	9	12	人工	帧级标注	–	2016
	DTB70	1 280×720	–	–	11	人工	帧级标注	–	2017
	UAVDT	1 080×540	是	3	9	半自动	帧级标注	–	2018
	VisDrone	3 860×2 160	是	10	12	人工	帧级标注	–	2019
	UAVDark135	–	是	5	12	人工	帧级标注	–	2021
	WebUAV-3M	–	–	223	17	半自动	帧级标注	–	2022
RGB-D	D^2Cube	1 280×720	否	34	18	人工	帧级标注	逐帧	2023
RGB-T	Anti-UAV	–	是	1	7	人工	帧级标注	–	2021
	VTUAV	1 920×1 080	否	13	13	人工	稀疏标注	首帧	2022
	HiAl	1 920×1 080	是	9	12	人工	帧级标注	逐帧	2025

注："–"表示具体信息未知或未提供。

1.2　无人机多模态目标跟踪方法

无人机多模态目标跟踪是计算机视觉领域中的重要研究课题之一。当前，主流的无人机多模态目标跟踪方法主要分为两类：传统方法和深度学习方法。传统方法包括稀疏表示方法和相关滤波方法等。稀疏表示方法有助于在跟踪过程中有效地建立可见光和热红外两种模态的联合特征。Lan等人（2020）利用稀疏表示对不同模态的目标进行特征表示，并根据模态相关性结合特征表示提升模态融合效率。Zhai等人（2019）通过对可见光和红外分别使用相关滤波器，然后利用低秩约束联合学习滤波器，实现跨模态融合。Yun等人（2019）提出了一种判别式融合相关学习模型，通过加权融合原则计算模态权重，有效结合多模态特征以提升跟踪性能。

随着深度学习技术的迅速发展，基于深度学习的多模态目标跟踪受到了广泛关注。Zhu等人（2020）提出了FANet（quality-aware feature aggregation network），通过在每个网络中聚合分层特征，有效提升了在高挑战性条件下的鲁棒性。Li等人（2020）针对可见光和红外两种模态数据的共同挑战及模态特定挑战提出了挑战感知网络CAT（challenge-aware RGBT tracking），应对可见光和红外模态数据的共同与特定挑战。为适应更多真实场景下的挑战，Zhang等人（2021）提出了自适应属性驱动表示的算法ADRNet（adaptive attribute-driven representation network）。Xiao等人（2022）提出了一种基于属性的渐进式融合网络APFNet（attribute-based progressive fusion network），自适应融合不同属

性的聚合特征与主干网络特征，实现鲁棒跟踪。Tu等人(2021)提出了一种多模态多边界度量学习框架，利用基于注意力机制的质量感知聚合模块建模简单样本与困难样本之间的关系，从而提升算法鲁棒性。Hui等人(2023)采用Transformer作为主干网络，通过融合模板作为模态交互的媒介，减少噪声信息的跨模态传播，显著提升了跟踪性能。Zhu等人(2023)提出视觉提示多模态跟踪算法ViPT(visual prompt multi-modal tracking framework)，通过学习模态相关提示，增强可见光模型的知识，使基础模型更好地适应多模态跟踪任务。Wang等人(2024)提出了一种新的时间自适应多模态跟踪框架，通过在线模板更新获取时态信息，并利用时空双流结构有效挖掘时空信息和多模态信息进行目标定位。

考虑到无人机场景下目标尺寸较小，Zhang等人(2022b)提出了一种多层级多模态融合跟踪算法HMFT(hierarchical multi-modal fusion tracker)。该算法结合了互补图像融合、通道注意力特征融合和空间注意力决策融合3种融合策略，实现了鲁棒的跟踪效果。Gao等人(2023)提出了一种多模态特征聚合网络，利用空间和通道注意力机制聚合异构模态特征信息，并优化该算法以适应无人机边缘计算平台。

2 数据集

为了促进多模态目标跟踪研究的发展，本文提出了一个高空无人机多模态目标跟踪数据集，该数据集聚焦于小汽车、行人等9类目标对象，包含了12种挑战属性，如无人机旋转、微小目标等。数据集由无人机搭载的混合传感器相机拍摄获得，包括150对可见光和红外视频序列，其中，127对拍摄于500 m左右高空，23对拍摄于120 m以下的低空。此外，数据集中还包含了多样化的交通场景、不同光照条件和多种天气情况下的数据，为无人机多模态目标跟踪的研究提供了丰富的实验基础。

2.1 硬件设置

拍摄本数据集的无人机型号为经纬Matrice300 RTK，如图1所示，其技术规格在表2中列出。经纬Matrice300 RTK搭载禅思Zenmuse H20T相机，可同时收集可见光与热红外两种模态的数据，可见光图像分辨率为1 920 × 1 080像素，热红外图像分辨率为640 × 512像素，这里的图像分辨率是指图像中像素点的数量，用“宽度 × 高度”的形式表示，单位是像素。



图1 经纬Matrice 300 RTK实图

Fig. 1 Image of Matrice 300 RTK

表2 Matrice 300 RTK的技术规格

Table 2 The technical specifications of Matrice 300 RTK

技术参数	参数值
尺寸(展开，不包含桨叶)	810 mm × 670 mm × 430 mm ($L \times W \times H$)
尺寸(折叠)	430 mm × 420 mm × 430 mm ($L \times W \times H$)
对角线电机轴距	895 mm
重量(不含电池)	3 600 g
最大载重	2 700 g
最大起飞重量	9 000 g
最大旋转角速度	俯仰轴：300 °/s，航向轴：100 °/s
最大上升速度	6 m/s
最大下降速度	5 m/s
最大飞行海拔高度	5 000 m(2 110桨叶)/7 000 m (2 195高原静音桨叶)
最大前飞时间(海平面)	45 min(负载重量为700 g)
最大悬停时间(海平面)	43 min(负载重量为700 g)

2.2 高质量配准与标注

鉴于无人机在拍摄可见光和热红外视频时视角与分辨率存在差异，因此，原始的多模态视频中目标对应的空间位置具有不一致性。为了确保目标的空间位置一致，保证数据集的准确性和有效性，本文对可见光和热红外视频进行了精细化的模态配准。对于单目标跟踪任务而言，目标出现的区域往往只是视频画面的局部区域，跟踪器通常只对目标区域进行处理。因此，配准的核心任务是确保可见光与热红外两种模态下目标区域的对齐精度。鉴于两种模

态的数据均以相同的帧速率采集，配准工作的首要步骤是进行时序对齐，通过精确调整确保可见光和热红外两种模态的视频序列起始于同一时间节点，这是实现目标区域精确配准的前提条件。随后，通过逐帧处理，对可见光和热红外两种模态的视频序列中目标所处的局部区域进行精细的空间对齐，确保对应帧之间的目标位置一致，配准完成后统一以1 920×1 080像素的图像分辨率导出。帧级的模态精细化配准为后续真值标注方案的实施提供了坚实的基础，确保了标注的准确性和一致性。

视觉目标跟踪任务需要数据集提供高质量的运动目标包围盒标注，这是准确评测跟踪器性能的前提，也是训练更加稳健的跟踪器的关键。为确保这一点，本文首先对两种模态的数据进行了精细的对齐工作，在此基础上采用模态共享的标注方案。具体来说，对于给定的特定跟踪目标的视频序列，从第一帧开始，对后续每一帧中的目标用覆盖其整体的最小边界框进行标注。当目标超出视野时，将目标真值的高度和宽度均设为0，以此确保标注的连贯性和准确性。在两种模态数据精细对齐的前提下，从视觉角度出发确保生成最佳包围框，以此来达到高质量的真值标注。

2.3 数据统计

数据集主要由无人机在500 m高空拍摄，重点关注9类目标对象，包含12种挑战属性，以短时目标跟踪为主，对数据集视频时长和序列帧数的统计见图2。数据集中包含了多样的场景、光照条件和丰富的对象类别，下面将具体分析数据集的特点。

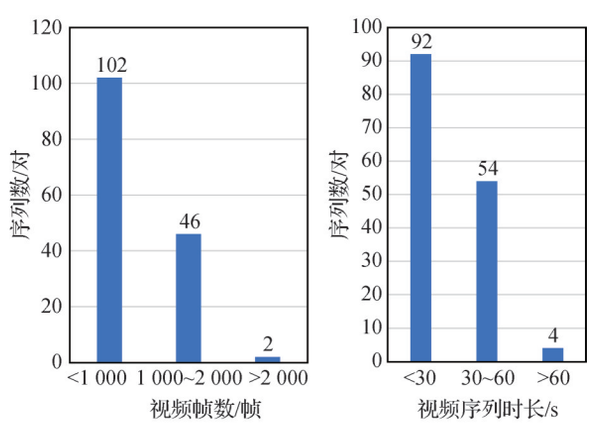


图2 HiAl中视频序列的帧数与时长统计

Fig. 2 Statistics on the number of frames and duration of video sequences in HiAl

2.3.1 挑战分析

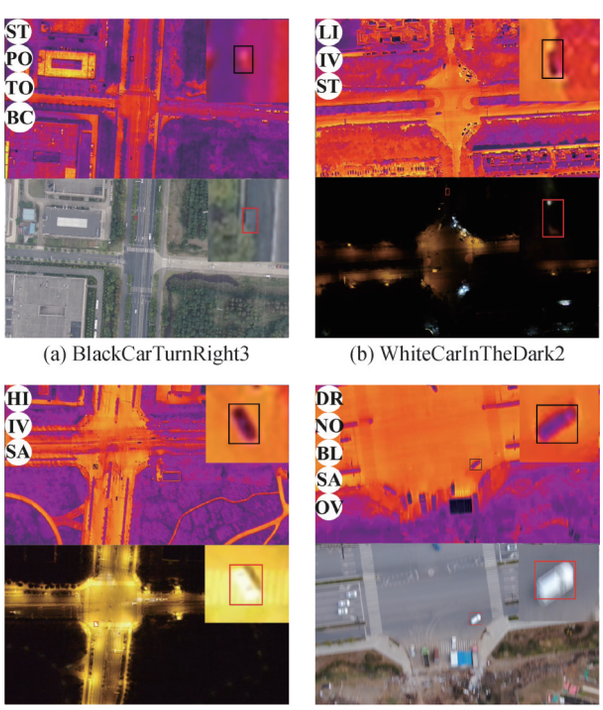
现有的多模态目标跟踪数据集已经涵盖了一系列挑战，如超出视野(out of view, OV)、背景杂乱(background clutter, BC)、相似物体干扰(similar appearance, SA)、模糊(blurred object, BL)、高照明度(high illumination, HI)、低照明度(low illumination, LI)、光照变化(illumination variation, IV)、完全遮挡(total occlusion, TO)、部分遮挡(partial occlusion, PO)和无遮挡(no occlusion, NO)等。然而，无人机平台与普通相机的应用场景存在显著差异，因此，在无人机平台下进行多模态目标跟踪会面临一些特殊的挑战。例如，无人机的旋转可能导致目标位置急剧变化，进而引发图像模糊问题。特别是在高空飞行时，目标尺寸会显著缩小，导致特征信息大幅减少，这为多模态目标跟踪带来了前所未有的难度。为了更全面地反映这些挑战，本文在传统挑战的基础上，特别引入了两种新的挑战类别：1)“无人机旋转”(drone rotation, DR)，以凸显无人机平台特有的问题；2)“微小目标”(small target, ST)，以强调高空无人机平台下目标尺寸显著减小的特殊挑战。本节通过表3详细定义了数据集中挑战属性的具体内容，而图3展示了各类挑战的典型示例。这些示例图中，不同序列涉及的挑战通过挑战名称英文缩写进行明确标注，目标区域的详细特征可见于图像的右上角。图4提供了各挑战属性数量的统计信息。

下面针对新增的两种挑战无人机旋转和微小目标进行详细解释。在无人机的实际跟踪过程中，旋转现象时有发生，这种旋转不仅展现了无人机的灵活性能，同时也导致了目标位置剧变、目标暂时性丢失以及图像模糊等复杂多变的跟踪难题。特别地，当无人机进行高速旋转时，可见光图像的清晰度会大幅下降，使得目标边界难以辨识，从而给目标跟踪任务带来了极大的挑战，本文将这一挑战定义为无人机旋转。在高空无人机平台下，目标尺寸通常较小，是与低空无人机及普通相机拍摄场景的主要区别之一。微小目标是高空无人机多模态目标跟踪的显著特性和重要挑战，本文将目标包围框的像素平均值低于32×32像素的视频序列视为微小目标挑战序列(Lin等，2014)。现有的无人机多模态目标跟踪数据集VTUAV中仅有15对微小目标挑战序列，而本文的数据集共150对视频，其中具有微小目标

表 3　HiAl 中挑战属性的详细定义

Table 3　Definition of challenge attributes in HiAl

挑战属性	定义
无人机旋转(DR)	由于无人机旋转导致目标的位置剧烈变化,图像模糊
微小目标(ST)	目标区域像素值低于 32 × 32 像素,目标特征微弱
无遮挡(NO)	目标未被遮挡
部分遮挡(PO)	目标局部被其他物体遮挡
完全遮挡(TO)	目标完全被其他物体遮挡
光照变化(IV)	目标所在区域光照明显变化
低照明度(LI)	目标所在区域光照强度低
高照明度(HI)	目标所在区域光照强度过高,导致难以辨别目标所在位置
模糊(BL)	图像信息模糊
相似物体干扰(SA)	存在与目标具有相似外观的其他对象
背景杂乱(BC)	目标与背景外观相似
超出视野(OV)	目标离开相机视野范围



(a) BlackCarTurnRight3　(b) WhiteCarInTheDark2

(c) TurnRightCar　(d) Motorbike2

图 3　挑战属性示例图

Fig. 3　Example images of challenge attributes

((a) BlackCarTurnRight3; (b) WhiteCarInTheDark2; (c) TurnRightCar; (d) Motorbike2)

挑战的序列达到 99 对。通过对 VTUAV 和 HiAl 的目标真值框所占绝对像素的统计对比,揭示了高空无人机数据集的复杂性。如图 5 所示,HiAl 的目标尺寸主要分布在 2 048 像素以下,其中微小目标的占比高达 49.4%。而 VTUAV 由于主要拍摄于 5～20 m 低空领域,其目标尺寸大多在 4 096 像素以上,微小目标仅占数据集总体的 1.8%。目标尺寸分布的显著差异彰显了高空无人机数据集的独特优势和高挑战性。

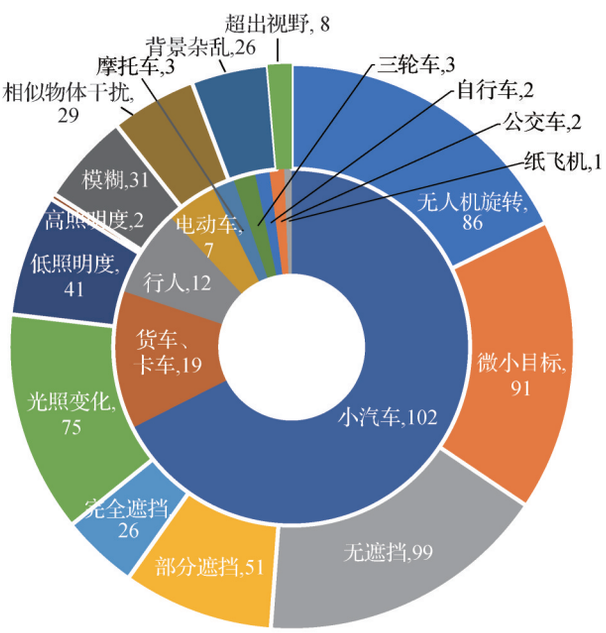


图 4　HiAl 中目标类别与挑战属性统计

Fig. 4　Statistics of object categories and attributes in HiAl

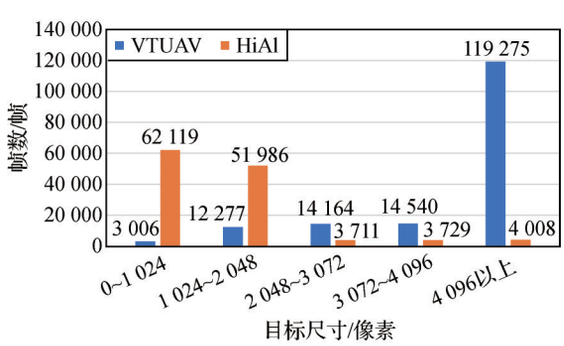


图 5　HiAl 与 VTUAV 的目标尺寸统计

Fig. 5　Statistics of object sizes in HiAl and VTUAV

2.3.2　场景与光照条件多样

为了使数据集更具现实意义,本数据集涵盖了丰富多样的场景和光照条件。在城市和郊区的各种具有不同特征的场景中收集数据,其中包括湖边、工地、高架桥、篮球场、草坪、塑胶跑道、小区、广场、街道等。日照条件包括白天、黑夜和傍晚,天气情况包括晴天和雾天,不同环境因素下的序列数统计如图 6

所示。数据集涵盖了41对在黑夜条件下以及13对在雾天环境下拍摄的视频。这些视频的引入不仅显著丰富了数据集的多样性，使得数据集中包含的挑战场景更为全面，同时也极大提升了数据集的整体挑战性，为深入研究高空无人机多模态目标跟踪算法提供了更为复杂且真实的实验环境。

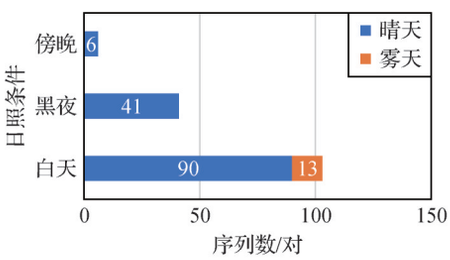


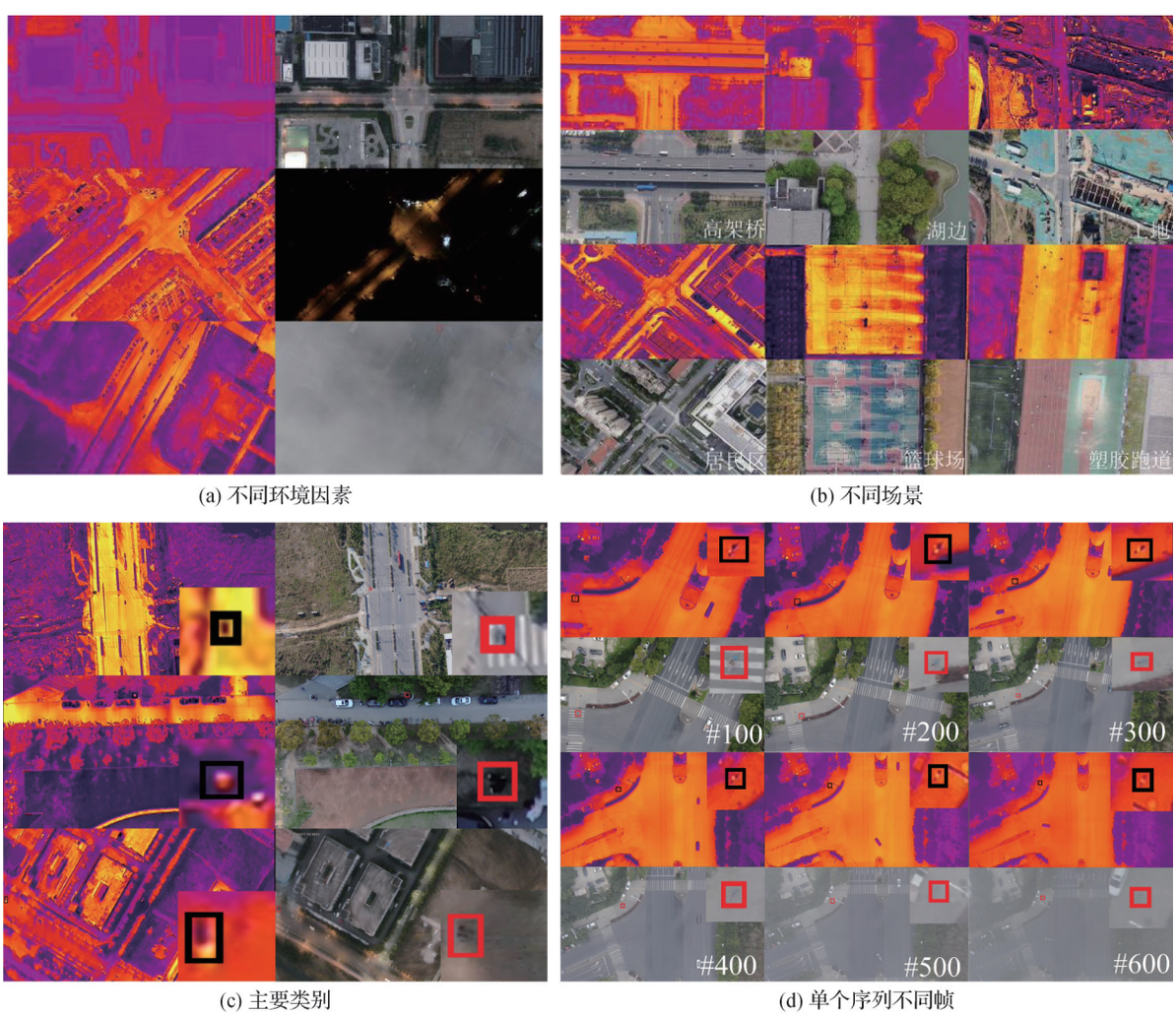
图6　HiAl中的不同环境因素统计图

Fig. 6　Statistical of different environmental factors in HiAl

图7(a)展示了不同环境因素下的示例帧对，由图可知，在黑夜及雾天环境下，可见光图像中目标信息稀缺，热红外图像则能提供更准确的目标信息。而在光照良好时，可见光图像中的目标特征更为突出，凸显了多模态数据的互补特性。图7(b)展示了高空无人机视角下的多样化观测场景。图7(c)给出了主要目标类别的典型示例，包括小汽车、行人和电动车，并涉及了无人机旋转、图像模糊以及微小目标跟踪等挑战。如图7(d)所示，随着无人机上升过程中高度的变化，其观测范围逐渐扩大，场景信息日益丰富，然而目标尺寸的持续缩小也为跟踪任务带来了更大的技术挑战。

2.3.3　对象类别丰富

本文构建的数据集涵盖了汽车、货车/卡车、公交车、行人、自行车、电动车等9大类目标对象，具体



(a) 不同环境因素　(b) 不同场景

(c) 主要类别　(d) 单个序列不同帧

图7　HiAl中的不同环境因素与不同场景的示例图

Fig. 7　Examples of different environmental factors and different scenarios in HiAl ((a) different environmental factors; (b) different scenes; (c) major categories; (d) different frames within a single sequence)

如图4所示。特别地，汽车、货车/卡车、公交车等大型目标主要集中在高空拍摄的视频序列中，为了进一步增强目标类别的丰富性和数据的多样性，引入了23对120 m以下低空拍摄的视频对，为保证数据集的小目标特性，这些低空拍摄的视频主要聚焦于行人、自行车、电动车等小型目标，确保了数据集中微小目标特性的体现。本文所提数据集在目标类别的分布上遵循长尾分布，与高空无人机在实际应用中采集到的目标分布情况高度一致。

3 评价与分析

在本节中，使用目标跟踪中常见的两种评估指标，以评估12种多模态跟踪算法在本数据集上的性能，针对其中6种算法进行了重训练实验。此外，给出了10种方法在两个主流多模态跟踪数据集GTOT（grayscale-thermal object tracking）（Li 等，2016）和RGBT234（RGB-thermal dataset）（Li 等，2019a）上的评估结果。最后，选取4个具有代表性的方法进行详细的对比分析实验，以验证和分析多模态数据之间的互补特性。图8给出了测试和评估的流程。

3.1 评估跟踪器

鉴于深度学习技术的快速发展，本文在所提数据集上评估了现有的12种主流多模态跟踪算法，包括HMFT、FANet、MaCNet（modal-aware attention network）（Zhang 等，2020）、CAT、MANet（multi-adapter network）（Li 等，2019b）、DMCNet（duality-gated mutual condition network）（Lu 等，2022）、MANet++（Lu 等，2021）、ADRNet、APFNet、ViPT、SuperDiMP（super discriminative model prediction）（Bhat 等，2019）和TBSI（template-bridged search region interaction）。

3.2 评价指标

由于数据集经过精细化模态配准，对可见光和热红外两种模态的视频可使用统一的目标真值标签。因此，本文采用跟踪算法的常用指标准确率（precision rate，PR）和成功率（success rate，SR）来评估算法性能。

准确率通过计算预测框与真实标注框中心点的欧氏距离L，并统计L在一定阈值内的帧所占的百分比来衡量。准确率图展示了阈值从小到大变化时符合条件的帧所占百分比的变化情况。通常阈值设定为20像素，但因本数据集多为微小目标，本文参考GTOT数据集，将阈值设置为5像素。距离L计算式为

$$L = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{1}$$

式中，(x_2, y_2)和(x_1, y_1)分别是预测边界框与真值框的中心位置坐标。

成功率是衡量目标跟踪算法性能的关键指标之一，定义为成功帧数与总帧数之间的比率。成功帧

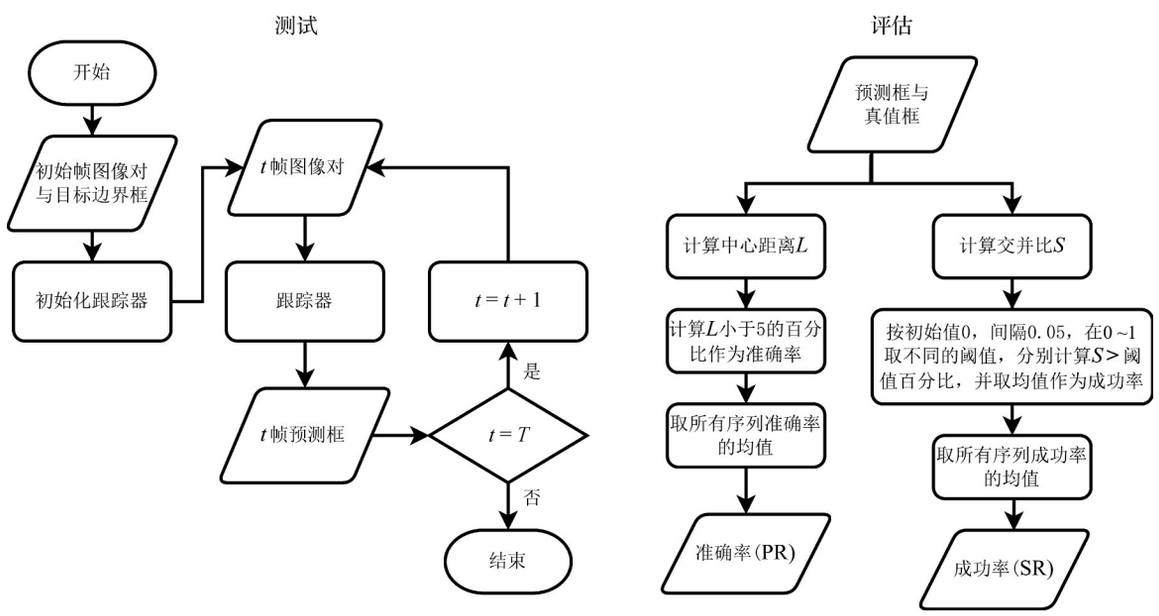


图8 测试与评测流程图

Fig. 8 Flowchart of test and evaluation

指预测框与真实标注框之间的交并比(intersection over union, IoU)大于设定阈值的帧。成功率图展示了阈值为0～1时,成功率的变化趋势,以评估算法在不同重叠率下的性能。交并比计算为

$$S = \frac{|\boldsymbol{r}_t \cap \boldsymbol{r}_a|}{|\boldsymbol{r}_t \cup \boldsymbol{r}_a|} \tag{2}$$

式中,$\boldsymbol{r}_t$表示预测边界框,$\boldsymbol{r}_a$表示真实标注框,∩和∪分别为交集和并集运算符。

3.3 总体评估结果

为评估多模态跟踪算法的总体性能,本文将整个HiAl数据集作为测试集评估12个主流多模态跟踪方法,图9为总体的准确率图和成功率图。结果显示,HMFT方法在HiAl上获得了最佳准确率分数0.602,该方法针对无人机平台下的场景特性,采用多层级模态融合策略,有效提升多模态跟踪算法性能。基于属性的渐进式融合网络APFNet获得最佳成功率分数0.514,APFNet通过其针对不同属性的融合分支来学习不同的融合参数,充分利用多模型挑战属性之间的信息,以实现鲁棒跟踪。高空无人机平台下场景与环境复杂,挑战繁多,目标纵横比变化频繁,目标尺度小,而APFNet能够针对特定挑战学习不同的多模态特征融合策略,以此实现更精确的目标框大小预测。

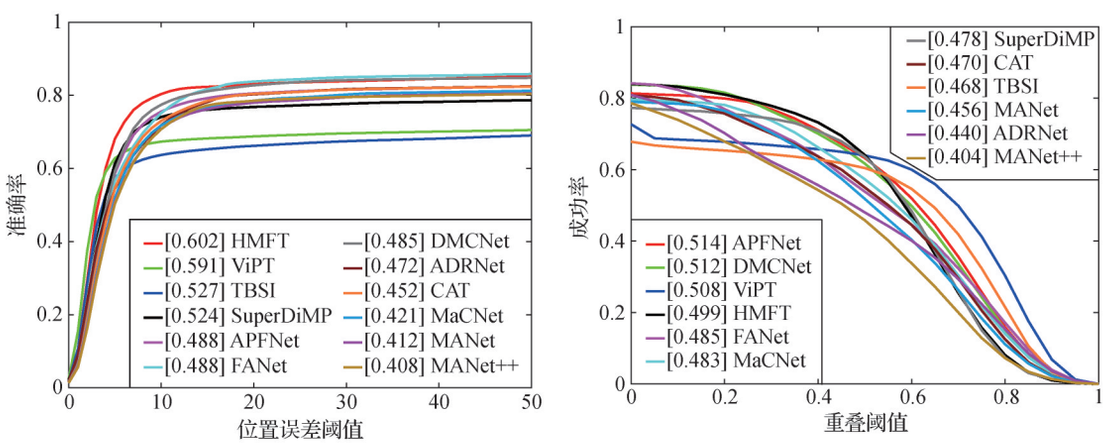


图9 跟踪器在HiAl上的总体准确率与成功率图

Fig. 9 Overall accuracy rate and success rate of trackers on HiAl

表4展示了10种方法在两个主流多模态跟踪数据集GTOT、RGBT234和HiAl上的实验结果,由表可知,10种方法在无人机数据集HiAl上的跟踪准确率与成功率普遍低于非无人机拍摄的数据集GTOT和RGBT234。由于地空视角差异大,且高空中捕获的目标通常尺寸小,特征微弱,跟踪的挑战更大。为进一步促进高空无人机多模态目标跟踪领域的发展,提出高空无人机数据集是必要的。

表4 跟踪器在GTOT、RGBT234和HiAl上的跟踪结果

Table 4 Tracking results on GTOT、RGBT234 and HiAl

方法	GTOT		RGBT234		HiAl	
	准确率	成功率	准确率	成功率	准确率	成功率
HMFT	**0.909**	**0.751**	0.788	0.569	**0.602**	0.499
DMCNet	0.908	0.733	0.839	0.593	0.485	0.512
APFNet	0.907	0.721	0.827	0.577	0.488	**0.514**
ADRNet	0.899	0.741	0.807	0.570	0.472	0.440
MANet++	0.897	0.725	0.795	0.560	0.408	0.404
MANet	0.895	0.709	0.777	0.535	0.412	0.456
FANet	0.891	0.728	0.787	0.553	0.488	0.485
MaCNet	0.878	0.697	0.790	0.550	0.421	0.483
CAT	0.860	0.693	0.804	0.555	0.452	0.470
TBSI	0.906	0.750	**0.871**	**0.637**	0.527	0.468

注:加粗字体表示各列最优结果。

3.4 重新训练

普通场景和高空无人机场景的图像特征差异显著,使用普通场景或低空无人机拍摄的数据集训练的跟踪器在高空场景下性能受限,高空无人机场景具有更广阔的视野和不同的光照条件,要求算法具备更强的泛化能力。因此,本文细致考虑数据集HiAl的关键属性,包括挑战类型、目标类别、日照情况等,进而将其分为100对视频序列的训练集和

50对视频序列的测试集。基于此训练了6个跟踪器，如表5所示，其中，训练后的DMCNet和MANet分别以0.524和0.529的分数获得了最高的准确率和成功率分数。相较于未采用HiAl数据集的训练情况，所有跟踪器的性能均显著提高，这验证了所提数据集在提升跟踪器性能方面的重要性。

表5　跟踪器在HiAl上重新训练的跟踪结果

Table 5　Results on HiAl of retraining experiments

	训练前		训练后	
	准确率	成功率	准确率	成功率
DMCNet	0.485	0.512	**0.524**	**0.526**
ADRNet	0.472	0.440	**0.483**	**0.457**
MANet++	0.408	0.404	**0.450**	**0.480**
MANet	0.412	0.456	**0.481**	**0.529**
FANet	0.488	0.485	**0.522**	**0.514**
CAT	0.452	0.470	**0.507**	**0.521**

注：加粗字体表示各行准确率与成功率的最优结果。

3.5　多模态互补性分析

为了分析多模态互补特性，本节对比了4个跟踪器在多模态和单模态数据上的跟踪性能，包括DMCNet、ViPT、HMFT和TBSI。使用多模态的模型参数对单模态数据进行跟踪，例如，跟踪可见光模态时，在保证其他条件不变的前提下，将多模态跟踪方法的输入图像统一为可见光图像，即热红外分支也以可见光图像作为输入，以此得到不同跟踪器在可见光模态下的跟踪结果，对热红外模态采取类似处理。如表6所示，4种跟踪器在多模态数据下均表现出最佳跟踪性能。算法HMFT有效结合了3种多模态数据融合方式，在跟踪的不同阶段充分挖掘模态互补特性，实现整体跟踪性能的显著提升，验证了该方法的有效性，也充分体现了多模态数据的互补优势。ViPT方法利用模态互补提示器充分提取下游任务不同语义级别的特征表示，学习两种模态的互补性，实现鲁棒跟踪。TBSI使用融合后的模板特征作为多模态特征交互的中间媒介，有效减少了噪声信息的跨模态传播。DMCNet通过对偶门控模块以充分利用所有模态的判别信息，实现鲁棒跟踪。通过对比4个方法的跟踪结果进一步验证了多模态数据之间的互补优势。

表6　多模态互补性分析实验结果

Table 6　Results of multi-modal complementarity analysis

方法	多模态		可见光		热红外	
	准确率	成功率	准确率	成功率	准确率	成功率
DMCNet	**0.485**	**0.512**	0.335	0.347	0.335	0.431
ViPT	**0.591**	**0.508**	0.547	0.470	0.234	0.230
HMFT	**0.602**	**0.499**	0.549	0.413	0.363	0.403
TBSI	**0.532**	**0.482**	0.463	0.414	0.165	0.159

注：加粗字体表示各行准确率与成功率的最优结果。

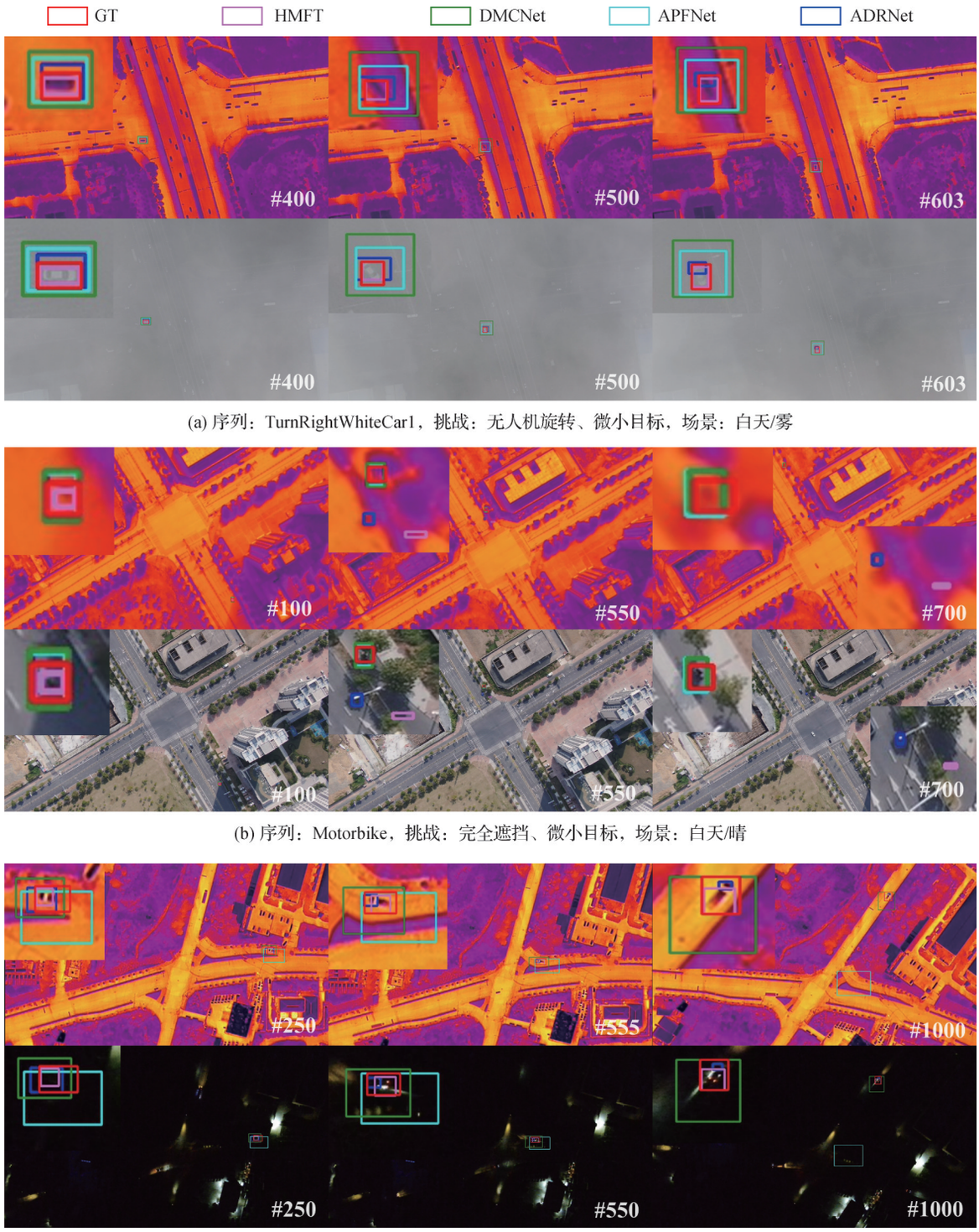
3.6　可视化结果分析

在本小节将对4种表现出色的深度跟踪器进行定性评价，图10展示了这些跟踪器在涵盖雾天、夜间和白天3种不同环境因素的代表性视频序列上的跟踪结果。

由图10(a)可见，在雾天场景下，HMFT通过有效利用热红外数据，能够较准确地预测出目标的位置与大小。面对第500帧的无人机旋转挑战时，HMFT显示出对目标纵横比变化的强大适应性。而APFNet与DMCNet面对上述挑战时难以准确预测出目标的尺寸大小。另外，如图10(b)所示，当遭遇遮挡挑战时，跟踪器容易丢失目标，并且在目标重新出现于画面中时，部分跟踪器难以恢复跟踪，如ADRNet、HMFT。在黑夜场景下，如图10(c)所示，可见光模态的信息减少，目标边界模糊，跟踪器难以对目标进行准确建模并精确估计目标尺寸大小。其中，APFNet针对5种常见挑战属性设计相应的挑战分支以解耦特征融合过程，从而减少模型对大规模训练数据的依赖。然而该方法面对极低光照的挑战时跟踪性能受限，漂移至干扰物上。在实际应用过程中，跟踪面临的挑战往往是复合存在的，如遮挡、纵横比变化、相似物体干扰等，跟踪器需要综合考虑多模态数据融合、改进特征提取机制、引入强化学习和挑战感知模型、动态更新目标模型和采用多任务学习等策略，有效改进无人机多模态跟踪算法在面对复杂挑战场景时的性能。

4　结　论

本文提出了一种高空无人机多模态目标跟踪数据集，旨在推动目标跟踪研究的进一步发展，并促进多模态目标跟踪在高空无人机上的应用。本文对数



(a) 序列：TurnRightWhiteCar1，挑战：无人机旋转、微小目标，场景：白天/雾

(b) 序列：Motorbike，挑战：完全遮挡、微小目标，场景：白天/晴

(c) 序列：MiniTruckTurnRightAndGoStraight，挑战：无人机旋转、光照变化，场景：夜晚/晴

图 10　在 3 个代表性视频序列上的可视化结果

Fig. 10　Visualization results of three representative sequences ((a) sequence: TurnRightWhiteCar1, attributes: DR, ST, scene: daytime/foggy; (b) sequence: Motorbike, attributes: TO, ST, scene: daytime/sunny; (c) sequence: MiniTruckTurnRightAndGoStraight, attributes: DR, IV, scene: night/sunny)

据集提供精细化模态配准和帧级真值标注，采用模态统一的标注方式以便于评估多模态算法性能。在所提数据集上对 12 个主流的多模态跟踪算法进行综合性能评估，分析了现有算法在高空无人机多模态目标跟踪任务上的局限性。

1）微小目标特征增强。高空无人机视野范围

广，捕获的目标尺寸较小，目标特征信息微弱，极大限制了目标跟踪性能。因此，通过自适应融合可见光与红外两种模态的互补信息，实现微小目标的特征增强，是高空无人机多模态目标跟踪的潜在研究方向。

2）高效、轻量化模型设计。基于深度学习的跟踪模型通常跟踪精度高，效率较低，计算复杂度高，存储量大，与无人机平台计算资源有限、存储空间小的特点存在冲突，限制了深度学习在无人机平台下的应用。因此，设计高精度、高实时性的轻量化模型是值得关注的研究方向。

3）运动感知跟踪算法研究。高空无人机具有运动灵活（旋转、俯冲和快速移动等）的特点，造成目标位置突变、目标外观剧烈变化，导致跟踪器发生漂移而丢失目标。如何通过视频帧时序信息补偿无人机运动带来的位置和外观变化，是应对上述挑战的有效方法，因此，研究运动感知的高空无人机多模态目标跟踪算法是值得研究的关键问题。

4）地—空协同目标跟踪。高空视角下目标小、干扰物体多，导致目标会被干扰物长时完全遮挡或部分遮挡，而使用传统的重检测机制难以有效解决此类挑战。考虑到地面相机能避免空中物体的遮挡，为高空无人机补充遮挡下的目标细节信息，从而可以解决高空无人机的遮挡问题。因此，协同地面相机实现鲁棒的高空无人机多模态目标跟踪具有重要的研究意义。

参考文献（References）

Bhat G, Danelljan M, Van Gool L and Timofte R. 2019. Learning discriminative model prediction for tracking//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 6181-6190 [DOI: 10.1109/ICCV.2019.00628]

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 the 15th European Conference on Computer Vision. Munich, Germany: Springer: 375-391 [DOI: 10.1007/978-3-030-01249-6_23]

Gao Z N, Li D D, Wen G J, Kuai Y L and Chen R. 2023. Drone based RGBT tracking with dual-feature aggregation network. Drones, 7(9): #585 [DOI: 10.3390/drones7090585]

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. Vancouver, Canada: IEEE: 13630-13639 [DOI: 10.1109/cvpr52729.2023.01310]

Jiang N, Wang K R, Peng X K, Yu X H, Wang Q, Xing J L, Li G R, Zhao J, Guo G D and Han Z J. 2021. Anti-UAV: a large multi-modal benchmark for UAV tracking [EB/OL]. [2024-01-24]. https://arxiv.org/pdf/2101.08466v2.pdf

Lan X Y, Ye M, Zhang S P, Zhou H Y and Yuen P C. 2020. Modality-correlation-aware sparse representation for RGB-infrared object tracking. Pattern Recognition Letters, 130: 12-20 [DOI: 10.1016/J.PATREC.2018.10.002]

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 C L, Cheng H, Hu S Y, Liu X B, Tang J and Lin L. 2016. Learning collaborative sparse representation for grayscale-thermal tracking. IEEE Transactions on Image Processing, 25(12): 5743-5756 [DOI: 10.1109/TIP.2016.2614135]

Li C L, Liang X Y, Lu Y J, Zhao N and Tang J. 2019a. RGB-T object tracking: benchmark and baseline. Pattern Recognition, 96: #106977 [DOI: 10.1016/J.PATCOG.2019.106977]

Li C L, Liu L, Lu A D, Ji Q and Tang J. 2020. Challenge-aware RGBT tracking//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 222-237 [DOI: 10.1007/978-3-030-58542-6_14]

Li C L, Lu A D, Liu L and Tang J. 2023. Multi-modal visual tracking: a survey. Journal of Image and Graphics, 28(1): 37-56 (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(1): 37-56) [DOI: 10.11834/jig.220578]

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

Li S Y and Yeung D Y. 2017. Visual object tracking for unmanned aerial vehicles: a benchmark and new motion models//Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, USA: AAAI Press: 4140-4146 [DOI: 10.1609/AAAI.V31I1.11205]

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]

Lu A D, Li C L, Yan Y Q, Tang J and Luo B. 2021. RGBT tracking via multi-adapter network with hierarchical divergence loss. IEEE Transactions on Image Processing, 30: 5613-5625 [DOI: 10.1109/TIP.2021.3087341]

Lu A D, Qian C, Li C L, Tang J and Wang L. 2022. Duality-gated

mutual condition network for RGBT tracking. IEEE Transactions on Neural Networks and Learning Systems, 1-14 [DOI: 10.1109/TNNLS.2022.3157594]

Mueller M, Smith N and Ghanem B. 2016. A benchmark and simulator for UAV tracking//Proceedings of 14th European Conference on Computer Vision. Amsterdam, the Netherlands: Springer: 445-461 [DOI: 10.1007/978-3-319-46448-0_27]

Tu Z Z, Lin C, Zhao W, Li C L and Tang J. 2021. M5L: multi-modal multi-margin metric learning for RGBT tracking. IEEE Transactions on Image Processing, 31: 85-98 [DOI: 10.1109/TIP.2021.3125504]

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. Washington, USA: AAAI: 5436-5444

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. Menlo Park, USA: AAAI Press: 2831-2838 [DOI: 10.1609/AAAI.V36i3.20187]

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. Vancouver, Canada: IEEE: 13374-13383 [DOI: 10.1109/CVPR52729.2023.01285]

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

Yun X, Sun Y J, Yang X X and Lu N N. 2019. Discriminative fusion correlation learning for visible and infrared tracking. Mathematical Problems in Engineering, 2019: #2437521 [DOI: 10.1155/2019/2437521]

Zhai S L, Shao P P, Liang X Y and Wang X. 2019. Fast RGB-T tracking via cross-modal correlation filters. Neurocomputing, 334: 172-181 [DOI: 10.1016/J.NEUCOM.2019.01.022]

Zhang C H, Huang G J, Liu L, Huang S, Yang Y N, Wan X, Ge S M and Tao D C. 2022a. 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 H, Zhang L, Zhuo L and Zhang J. 2020. Object tracking in RGB-T videos using modal-aware attention network and competitive learning. Sensors, 20(2): #393 [DOI: 10.3390/s20020393]

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(9): 2714-2729 [DOI: 10.1007/s11263-021-01495-3]

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

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. Vancouver, Canada: IEEE: 9516-9526 [DOI: 10.1109/CVPR52729.2023.00918]

Zhu P F, Wen L Y, Du D W, Bian X, Fan H, Hu Q H and Ling H B. 2021. 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 Y B, Li C L, Tang J and Luo B. 2020. Quality-aware feature aggregation network for robust RGBT tracking. IEEE Transactions on Intelligent Vehicles, 6(1): 121-130 [DOI: 10.1109/TIV.2020.2980735]

Zhuo L, Zhang S Y, Zhang H and Li J F. 2021. Survey on techniques of single object tracking in unmanned aerial vehicle imagery. Journal of Beijing University of Technology, 47(10): 1174-1187 (卓力, 张时雨, 张辉, 李嘉锋. 2021. 无人机影像单目标跟踪综述. 北京工业大学学报, 47(10): 1174-1187) [DOI: 10.11936/bjutxb2020030017]

作者简介

肖云,女,副教授,硕士生导师,主要研究方向为计算机视觉和多模态分析。E-mail:xiaoyun@ahu.edu.cn

李成龙,通信作者,男,教授,博士生导师,主要研究方向为计算机视觉、深度学习。E-mail: lcl1314@foxmail.com

曹丹,女,硕士研究生,主要研究方向为基于无人机的多模态目标跟踪。E-mail: cdan0712@163.com

江波,男,教授,博士生导师,主要研究方向为模式识别与计算机视觉。E-mail: jiangbo@ahu.edu.cn

汤进,男,教授,博士生导师,主要研究方向为计算机视觉与深度学习。E-mail: tangjin@ahu.edu.cn