

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(XXXX)XX-0001-15

论文引用格式: Wang Hanzheng, Li Wei, Zhang Jiaming, He Hongtao, Huang Xun. Hyperspectral aerial object tracking integrating spectral prompts and hybrid attention[J/OL]. Journal of Image and Graphics, XXXX:1-15. DOI: 10.11834/jig.260136. (空中目标跟踪, 王瀚正, 李伟, 张家铭, 何鸿涛, 黄荀. 融合光谱提示与混合注意力的高光谱[J/OL]. 中国图象图形学报, XXXX:1-15. DOI: 10.11834/jig.260136.) [DOI: 10.11834/jig.260136]

融合光谱提示与混合注意力的高光谱

空中目标跟踪^{1,2,3}, 王瀚正^{1,2}, 李伟^{1,2}, 张家铭³, 何鸿涛³, 黄荀³

1. 北京理工大学 信息与电子学院, 北京 100081; 2. 天基智能信息处理全国重点实验室, 北京 100081; 3. 北京空间机电研究所, 北京 100094

摘要: 目的 分割一切模型(segment anything model 2, SAM 2)在目标跟踪领域已取得显著进展,其具备强泛化的分割先验以及长时序建模能力,能够目标跟踪提供空间信息支持,缓解背景混淆与遮挡后重现易丢失等跟踪关键难题。然而,SAM 2无法利用高光谱数据中的光谱判别信息,当目标与背景在空间纹理上高度相似等情况时,仅依赖空间语义先验生成的掩码可能出现漂移。因此,针对高光谱目标跟踪中目标与背景分离不充分等问题,提出一种面向高光谱目标跟踪的光谱提示与混合注意力方法。**方法** 以分割一切模型SAM 2为基础构建跟踪框架:在空间维度,利用语义掩码进行像素级目标-背景分离;在时序维度,通过记忆库与记忆注意力建立跨帧长时依赖,增强遮挡重现后的关联与恢复;在光谱维度,提出混合嵌入网络以对齐高光谱数据分布,并引入混合注意力与光谱嵌入模块,将目标光谱信息编码为提示向量、学习波段间深层语义关系。此外,为了促进相关领域研究,构建了首个高光谱空中目标跟踪数据集AHOT。**结果** 在3个数据集上与6种最新方法对比,所提方法在HOTC-2024-RedNIR数据集上DP_20相比第二名提升4.08%,在AHOT数据集上DP_20相比第二名提升15.57%,对比实验结果证明所提算法改善了目标跟踪的性能。此外,本文方法的推理速度达到15.4 FPS,在实现较高精度的同时具有近实时的工程效率。**结论** 通过掩码级目标分离与长时记忆建模,并结合光谱嵌入与注意力融合策略,所提方法显著提升了复杂场景下高光谱目标跟踪的精度与鲁棒性。构建数据集将在科学数据银行公开(31253.11. sciencedb. j00240.00255)。

关键词: 高光谱目标跟踪;分割一切模型;注意力机制;光谱提示;特征提取

Hyperspectral aerial object tracking integrating spectral prompts and hybrid attention

Wang Hanzheng^{1,2,3}, Li Wei^{1,2}, Zhang Jiaming^{1,2}, He Hongtao³, Huang Xun³

1. School of Information and Electronics, Beijing Institute of Technology, Beijing 100081; 2. National Key Laboratory for Space-based Intelligent Information Processing, Beijing 100081; 3. Beijing Institute of Space Mechanics and Electricity, Beijing 100094

Abstract: **Objective** Segment Anything Model 2 (SAM 2) has demonstrated strong potential for object tracking, benefiting from its highly generalizable segmentation priors and long-term temporal modeling capability. These properties provide reliable spatial support for tracking and help mitigate key challenges such as background confusion and target loss after occlusion and reappearance. However, SAM 2 is not able to explicitly exploit the discriminative spectral information inherent in hyperspectral data. As a result, when the target and background exhibit highly similar spatial textures, mask prediction driven mainly by spatial semantic priors may drift. To address the insufficient target-background separation in hyper-

收稿日期: 2026-03-17; 修回日期: 2026-07-09

基金项目: 国家自然科学基金(项目编号: 625B2025)

Supported by: Project supported by the National Natural Science Foundation of China (Grant No. 625B2025)

© 中国图象图形学报版权所有

spectral object tracking, we propose a spectral-prompt and hybrid-attention method tailored for hyperspectral tracking, which enhances the model's ability to leverage spectral cues and improves robustness in visually ambiguous scenes.

Method The proposed HOT framework is built upon Segment Anything Model 2 (SAM 2), leveraging its high-precision mask generation to accurately delineate objects at the pixel level and thus achieve clean object-background separation in complex scenes. However, since SAM 2 is not designed to explicitly perceive hyperspectral characteristics, it struggles to capture discriminative differences across high-dimensional spectral bands. To address this limitation, we propose a hybrid embedding network that efficiently adapts the model to hyperspectral data distributions, and further introduce a hybrid attention mechanism together with a spectral embedding module. These components transform object spectral signatures into learnable prompt vectors and learn deep inter-band semantic representations, respectively, facilitating a collaborative fusion of spatial, temporal, and spectral information for robust hyperspectral object tracking. Furthermore, to foster research and provide a standardized benchmark for this emerging task, we construct AHOT, the first hyperspectral UAV object tracking dataset. AHOT is collected from real UAV flights and is specifically designed to reflect the unique challenges of low-altitude aerial tracking with hyperspectral imagery, where strong viewpoint changes, platform motion, and complex backgrounds are common. **Result** Comprehensive experiments against six recent trackers on three benchmarks further validate the overall effectiveness and generalization ability of our approach. We report both AUC and DP₂₀ to reflect complementary aspects of tracking quality, where AUC measures overall success across overlap thresholds and DP₂₀ evaluates localization accuracy under a 20-pixel criterion. Across different spectral configurations and scene complexities, our method exhibits consistently strong performance and stable ranking. In particular, it achieves the top DP₂₀ on HOTC-2024-RedNIR of 0.8416, outperforming the second-best competitor by 4.08%, while also attaining the highest AUC of 0.6034, indicating improvements not only in point-wise precision but also in overall tracking robustness. On the more challenging AHOT benchmark, the advantage becomes more pronounced: our method reaches DP₂₀ of 0.8779, exceeding the runner-up by 15.57%, and obtains the best AUC of 0.3885, suggesting enhanced resilience to severe appearance changes, background clutter, and long-term interruptions such as occlusion and object re-emergence. Moreover, on HOT-2024-NIR, our method remains highly competitive with DP₂₀ of 0.9740 and a better AUC of 0.7648, demonstrating reliable performance in near-infrared scenarios. Overall, these cross-dataset results consistently show that the proposed tracker delivers accurate localization and robust temporal consistency, making it well suited for practical deployment under diverse spectral modalities and challenging real-world conditions. In addition, the inference speed of the method proposed in this paper reaches 15.4 FPS, achieving near real-time engineering efficiency while maintaining high accuracy. **Conclusion** This paper presents a hyperspectral object tracking model that effectively integrates the complementary strengths of spectral representation learning and segmentation-based foundation models. The incorporation of spectral embedding enhances the model's ability to perceive and exploit high-dimensional inter-band differences, improving discrimination against visually similar but materially different distractors. Coupled with long-term temporal association modeling, the tracker demonstrates stronger robustness to severe appearance changes, deformation, and long-duration occlusions, and is able to re-localize targets more reliably after disappearance and re-emergence. Overall, the proposed method delivers more accurate and stable tracking performance across diverse hyperspectral scenarios, highlighting the effectiveness of jointly modeling spatial, temporal, and spectral cues for HOT.

Key words: Hyperspectral object tracking; segmentation-based model; attention mechanism; spectral prompt; feature extraction

论文引用格式:[DOI: 10.11834/jig.260136]

0 引言

高光谱成像是一种融合成像和光谱分析的感知技术,旨在获取连续窄带范围内每一个场景像素的

光谱信息(孙卫东等,2025)。与传统的彩色图像(RGB)相比,高光谱图像(hyperspectral image, HSI)同时提供目标的二维空间信息与三维诊断性光谱特征,反映不同材质对光的反射和吸收特性,进而揭示物体的本质属性,具有区分目标的能力(刘恩勤等,2025)。因此,高光谱成像在伪装目标探测(Lv等,

© 中国图象图形学报版权所有

2023)、森林防火、农作物生长评估(张影等, 2019)、城市交通监控(Altat 等, 2025)等领域得到广泛应用, 具有重要的科研价值与工程应用前景。

高光谱目标跟踪(hyperspectral object tracking, HOT)任务对高光谱图像序列中的给定目标进行实时定位, 是计算机视觉与遥感领域的重要研究方向(胡世宇等, 2024)。相较于基于RGB图像的目标跟踪, HOT旨在通过辨析目标的材料属性, 区分颜色相近但材质不同的目标(He 等, 2024), 进而对目标连续定位, 在复杂背景或遮挡情况下具备更高的鲁棒性。

传统的HOT研究主要依赖相关滤波方法, 提取目标特征并通过训练得到目标滤波器模板, 计算目标滤波器模板与当前待检测样本的相关响应图, 将响应值最大的位置当作当前帧的目标位置。例如, Qian 等人(2018)首先在目标区域创建一组归一化的三维立方体, 用于提取目标周围的光谱信息。随后将其作为固定的卷积滤波器, 结合核相关滤波框架(Henriques 等, 2015)实时更新滤波器模板, 从而区分目标与周围的环境。但是核相关滤波算法在滤波器训练过程中未考虑图像的光谱维度信息, 直接应用其处理HOT任务可能限制跟踪性能的提升。针对这一问题, Xiong 等人(2020)提出一种新的基于背景感知相关滤波器的相关滤波算法。通过设计空间-光谱多维梯度直方图增强光谱特征表达能力, 将其与丰度图特征联合使用, 嵌入到相关滤波器中联合编码潜在的材质分布, 实现对目标位置的准确预测。Hou 等人(2022)在相关滤波框架中设计一个空间-光谱加权正则化项, 通过抑制与中心像素不同的光谱特征来引导滤波器模板关注目标。使用改进的方向梯度直方图特征提取器和自适应尺度搜索策略, 进一步提高跟踪器的特征区分能力和尺度适应能力。虽然基于相关滤波的跟踪方法能够进行感兴趣目标的初步定位, 但其准确率仍有较大的提升空间。现有改进的相关滤波方法对通道信息与背景动态变化建模不足, 难以保持在背景快速变化的场景下的跟踪性能。相比之下, 深度学习模型能够自动学习多层次的空间-光谱联合特征, 在计算机视觉领域取得突破性进展, 研究者开始将其引入HOT任务中。一类方法尝试将现有RGB领域中的目标跟踪框架扩展到高光谱数据。Li 等人(2023)提出一种集成学习框架, 运用自监督训练得到高光谱波段重组

模型, 进而生成HSI的波段权重, 根据权重挑选生成具有不同光谱组合的三通道伪彩色图像, 输入至RGB预训练跟踪模型中进行微调推理。最后将骨架网络不同阶段的预测位置集成为最终结果, 确保目标定位的鲁棒性。Chen 等人(2025)通过引入频率感知滤波模块与自适应滤波生成器, 对波段重组得到的伪彩色图像的光谱和空间特征一致性进行细化, 实现高效的跨伪彩模态交互。Su 等人(2025)将混合专家模型引入HOT任务, 利用光谱分组和专家路由策略, 对多个伪彩色图像进行稀疏采样, 高效聚合不同的专家知识进行目标位置预测。虽然该方法部分缓解样本不足的压力, 但波段重组作为一种不增加额外信息的图像变换手段, 其产出的伪彩色图像存在冗余, 可能对模型泛化能力的提升产生影响。另一类方法则设计专门的高光谱特征提取网络, 采用3D卷积(Wang 等人, 2025b)、光谱注意力机制(Li 等人, 2020)等模块, 充分利用光谱维度的上下文信息以提高特征表达能力。Liu 等人(2022)设计一种空间-光谱注意力模块以实现特征融合, 该模块使伪彩色RGB跟踪网络能够指导光谱网络进行训练, 从而显著降低模型优化的难度。Wang 等人(2025a)提出一种用于空谱特征联合提取的骨架网络, 与伪彩色RGB骨架共同组成孪生网络提取视觉与光谱的联合判别信息, 并构建基于光谱角感知的预测头部网络, 提高对感兴趣目标光谱定位的准确率。Jiang 等人(2024)等人提出基于长短期记忆网络的通道自适应模块。利用3D卷积分离通道特征, 构建前向与后向通路执行双向扫描及状态转移, 增强后的特征用来进行目标位置的预测。Wang 等人(2024)在Transformer结构中引入内容和位置嵌入模块, 动态学习位置信息与内容信息, 并通过动态可靠性细化模块解决目标位置准确感知难题。

然而, 上述方法在处理复杂场景时仍面临瓶颈。首先, 传统HOT方法采用边界框表示目标位置, 易包含背景杂波, 目标光谱特征易被背景像素“污染”。这种混合像元现象将破坏目标独特的光谱指纹, 在对外观相似但材质不同的干扰物时, 模型预测结果易发生漂移。其次, 现有方法在时序建模上缺乏长时记忆能力。仅关注相邻帧之间的关联, 缺乏对目标在整个视频序列中位置变化的全局理解。当目标遭遇遮挡或消失后重现时, 模型往往因缺乏历史上下文信息的支撑而无法重新捕获目标, 导致跟踪

失败。相比之下,分割一切模型(segment anything model 2, SAM 2)(Ravi 等,2024)为解决上述问题提供新的视角。SAM 2能够生成高精度的分割掩码,可以将目标从复杂背景中精确分离,进而剔除背景光谱噪声的干扰。此外,SAM 2引入记忆库和记忆注意力模块,能够存储并检索目标在过去帧中的特征交互信息。在目标发生剧烈形变、旋转甚至被长时间遮挡后,仍能利用历史记忆准确地找回并分割目标,提升跟踪的鲁棒性。然而SAM 2缺乏对高维光谱信息的感知能力,无法捕捉光谱维度的差异。

基于此,为了充分挖掘SAM 2在视频分割与时序记忆上的潜力,同时弥补其在光谱感知上的缺失,本文提出一种基于SAM 2的HOT方法。冻结SAM 2的编码器以保留其强大的时序关联与抗遮挡能力,仅通过混合嵌入网络使其快速适应高光谱数据分布。同时,创新性地引入光谱提示机制与光谱嵌入模块,前者将目标的光谱特征转化为提示向量,赋予模型基于材质物理属性的引导能力;后者通过提取波段间深层语义,将高维光谱信息转化为模型可理解的特征嵌入。该方法实现空间、时序与光谱特征的深度融合,有效解决传统方法中背景污染严重及长时跟踪易丢失的难题,显著提升跟踪的鲁棒性与精度。

本文主要贡献如下:1)设计基于SAM的高光谱目标跟踪模型,将光谱信息与SAM的时序关联能力融合,通过混合嵌入网络提取光谱特征,丰富融合特征中的光谱可判别细节以增强特征表示。2)设计混合注意力模块,实现通道与空间的互补增强、抑制背景干扰;进一步通过提示编码器将目标光谱先验编码为可学习的提示嵌入,对跟踪过程进行材质属性引导,实现空间—时序—光谱信息的深度协同,提升目标重捕获能力。3)构建首个高光谱空中目标跟踪数据集,涵盖多种挑战性场景支持对长时记忆、重捕获与光谱判别能力的可量化评估,为高光谱空中目标跟踪方法的训练、对比与标准化评测提供统一基准。4)在3个高光谱目标跟踪数据集上的实验结果表明,提出方法能有效利用视觉和光谱信息,提高目标跟踪精度。

1 相关工作

1.1 注意力机制

注意力机制最初源于人类视觉系统的信息处理方式,人类会自动聚焦于复杂视觉场景的关键区域,过滤冗余信息,提高感知效率。受此启发,注意力机制被引入神经网络中,用于引导模型动态地分配计算资源,增强对关键信息的建模能力。近年来,注意力机制已成为深度学习领域的重要研究方向,空间注意力机制和通道注意力机制作为最早提出并广泛应用的两种形式,具有结构简单、计算高效、效果显著等优点,被广泛应用于图像分类、目标检测、图像分割等视觉任务中。

空间注意力机制以特征图的空间位置为加权依据,通过计算各个位置之间的相关性,对空间维度上的特征进行加权整合,生成空间注意力掩码,用以反映图像中不同区域的重要性分布,从而实现对各空间位置的权重调控与信息增强。Yang 等人(2025)提出一种空间注意力模块,对输入特征图进行通道维度的压缩,通过卷积层学习空间注意力权重,并将其作用于原始特征图中,增强目标区域的响应。通道注意力机制将每个通道视为语义特征的响应器,通过对特征分布的自适应学习,模型能够动态调整对输入特征的关注意度,增强对关键通道的响应能力,进而提升对关键信息的感知效果与表达能力。Hu 等人(2018)提出“压缩与激活”模型,对每个通道的特征图进行全局平均池化,提取全局上下文信息,随后由全连接层组成的非线性结构学习各通道的重要性权重。这些权重被用于重新标定原始特征图的通道响应,关注有用特征,抑制无关通道信息。

1.2 分割一切模型

随着计算机视觉领域向大模型和通用模型方向发展,Meta AI 于 2023 年提出的 SAM(Kirillov 等,2023)成为图像分割领域的重要突破,是首个面向“分割一切”任务的大规模视觉基础模型,具备强大的泛化能力,能够在无需重新训练的前提下,通过简单提示(如点、框、掩码)实现对任意图像中目标的高质量分割。其结构由图像编码器、提示编码器和掩码解码器三部分组成,其中图像编码器采用视觉Transformer架构以提取全局特征,提示编码器将输入转化为语义向量,掩码解码器则融合图像特征与

提示信息,生成目标掩码。SAM 2将模型任务扩展到视频分割(Cheng等,2023),通过引入记忆机制来跟踪动态视频序列中多帧内的对象。Yang等人(2024)提出一种增强SAM 2模型SAMURAI,引入时序运动线索与运动感知记忆选择机制,能够准确预测目标运动并优化掩码选择。Yin等人(2025)提出一种无需训练的视频分割方法,结合上下文感知与遮挡鲁棒的记忆机制,提升了在复杂场景下的分割准确性与稳定性。

1.3 研究思路

受以上工作启发,针对现有高光谱目标跟踪方法缺乏时序建模能力的问题,引入SAM结构作为基础框架,并结合改进的注意力模块,增强模型处理高光谱序列数据的能力。通过融合时序上下文信息与空间-光谱表征,构建空间-光谱混合注意嵌入模块,弥补SAM在高光谱维度上的建模缺陷。该模块融合空间注意力与光谱注意力机制,能够自适应挖掘图像中具有判别力的空间区域与关键波段信息,从而实现多维特征的联合增强表达。

2 本文方法

2.1 整体结构

所提基于SAM 2的高光谱目标跟踪网络架构如图1所示。整体框架以SAM 2为主干网络,结合提出的混合嵌入网络与提示编码器,实现从原始图像与光谱提示到精细目标掩码的高效映射。其中SAM 2由图像编码器、掩码解码器以及记忆注意力模块组成。在每帧图像处理过程中,从候选记忆槽中选择具有代表性的历史帧(首帧及临近预测帧),通过记忆注意力模块与当前帧的图像特征 F_t 进行融合,引导模型关注目标区域。记忆编码器将当前帧的目标掩码编码为记忆特征 F'_t ,动态更新候选记忆槽,实现多帧之间的时序信息传递。同时,SAM 2支持稀疏(点、框)与稠密(掩膜)提示的联合输入。提示信息首先通过提示编码器进行嵌入,其中稀疏提示通过位置编码与可学习向量进行表示,密集提示则通过卷积模块与图像特征融合。最终经过双向Transformer构成的掩码解码器对融合特征进行解码,并生成多个候选掩膜与置信度,选择得分最高者作为最终预测掩码 M 。此外,考虑到高光谱目标跟踪任务中缺乏可见光图像,难以直接提取传统视觉

特征,采用固定波段选择策略(Sun等,2023),从原始高光谱图像中选取3个代表性波段构建3通道伪彩色图像 I_v ,适配SAM 2输入格式。进而利用其预训练的通用视觉知识以增强特征表达能力,对模型进行任务相关的微调优化,实现高光谱图像与视觉模型之间的有效迁移与融合。

尽管SAM 2在自然图像跟踪中表现出色,但其原始设计主要针对可见光图像,未考虑高光谱图像中的波段关系与光谱特征分布,容易忽略关键波段间的光谱差异,导致目标边界模糊或类别混淆。此外,SAM 2的提示机制主要基于几何位置信息,难以适配光谱信息提取,限制其在高光谱目标跟踪中的应用。针对上述问题,在SAM 2的基础上引入混合嵌入网络,实现从原始图像与光谱提示到精细目标掩码的高效映射,保留SAM 2在提示驱动分割中的强泛化能力,同时弥补其在高维光谱建模方面的不足。混合嵌入网络通过卷积层对光谱图像进行初步压缩,提取局部关联特征;随后引入混合注意力网络,结合通道注意力与空间注意力机制,动态感知不同波段之间的语义联系与空间依赖;通过非线性激活函数和卷积层增强特征表达能力。最终将原始高维光谱信息映射为3通道的光谱特征图,具备

光谱判别性与空间一致性,输入至后续的光谱编码器进行特征提取和提示生成。混合编码器用于引入光谱提示,实现光谱-空间联合建模。在原有稀疏提示嵌入的基础上,增加针对光谱特征响应的编码结构,通过一组轻量卷积与对光谱提示进行建模,并与空间提示嵌入进行融合,引导模型响应光谱提示,提高现有跟踪框架的精度与泛化能力。

2.2 混合嵌入网络

为充分挖掘光谱判别特征,引导主干网络学习,设计混合嵌入网络(hybrid embedding network, HEN)。该网络整体结构如图2所示,由卷积模块、混合注意力模块、激活层与光谱编码器组成,最终产生一个3通道的光谱特征图,作为引导信息输入至光谱编码器形成光谱嵌入。具体而言,混合嵌入网络的输入为尺寸为 $H \times W \times C$ 的高光谱图像张量 I_s ,其中 H 和 W 分别为图像的高和宽, C 为光谱通道数。采用 1×1 卷积层对光谱通道进行压缩和整合,初步提取局部光谱联合特征,输出张量 I'_s 维度为 $H \times W \times C_s$, C_s 经实验验证设置为16;随后送入混合注意力模块,该模块融合通道注意力与空间注意力机制,

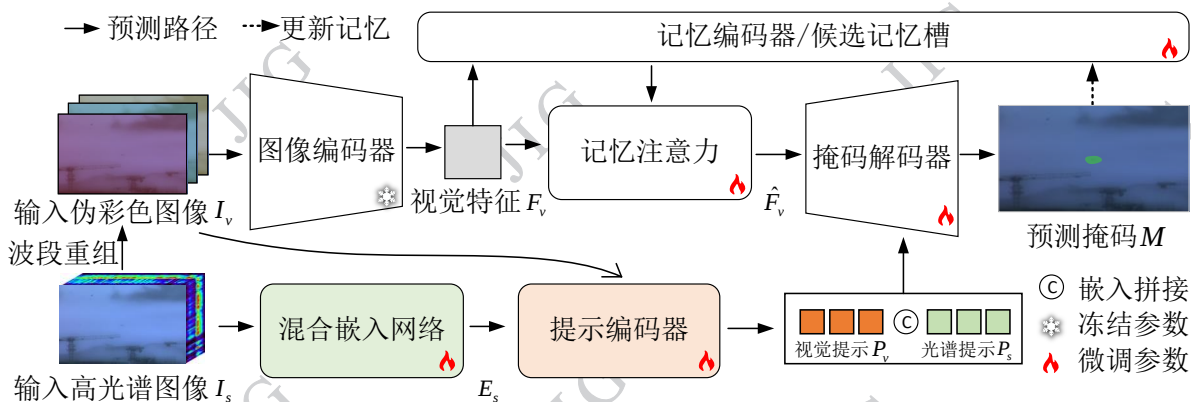


图1 网络整体结构

Fig. 1 Overall network structure

保持输出特征维度不变,增强模型对关键波段区域和空间位置的关注。随后经过非线性激活函数提升特征表达能力,最后通过 1×1 的卷积模块将通道数压缩为3,输出光谱特征图 I_s'' , 维度为 $H \times W \times 3$ 。

为进一步提升光谱提示的语义层次与结构感知能力,将上述光谱特征图输入至具有轻量化金字塔结构的Transformer网络(pyramid vision transformer, PVT)(Wang等,2021)中进行特征提取。该网络具有更小的计算开销,同时保持较强建模能力。通过多尺度特征编码机制,逐层聚合局部与全局信息,构建层次化语义表征的提示特征。选取最后一阶段的输出特征作为最终的光谱嵌入 E_s ($((H/32) \times (W/32) \times 512$, 其中512为通道数)。该设计实现了提示空间的扩展,不仅支持空间几何提示,引入光谱维度的语义引导,使模型在处理高光谱图像时能够根据光谱结构进行更精准的区域关注。最终,光谱嵌入输入至提示编码器生成光谱提示,与框级提示的联合输入至SAM掩膜解码器,引导其在空间-光谱双重语义下进行分割预测,提升模型对材质类别、波段差异显著区域的响应能力。

2.3 混合注意力网络

为挖掘光谱嵌入编码中的空间-光谱联合特征,提出混合注意力网络(hybrid attention network, HAN),在空间维度与通道维度上进行显式建模与交叉增强。不同于传统注意力机制仅考虑单一维度,该模块通过空间注意力与通道注意力的协同实现多维特征信息的深度融合与互补强化,提升模型对目标区域和关键波段的响应能力。

首先介绍空间注意力模块(spatial attention block, SAB),具体结构如图3所示。引入可学习归一化缩放因子 γ 的权重评估机制,结合组归一化(group normalization, GN)(Wu等,2018)的缩放参数以及通道间的归一化操作,评估每个通道空间特征的重要性。对于给定输入特征 $X_{in}^s \in \mathbf{R}^{H_f \times W_f \times C_f}$, 其中 H_f 和 W_f 分别为特征的高和宽, C_f 为通道数。其计算过程如下所述:

$$X_{out}^s = GN(X_{in}^s) = \gamma \frac{X_{in}^s - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta \quad (1)$$

式中 μ 和 σ 分别表示每组特征的均值和标准差, ε 为稳定性参数, GN 表示组归一化函数, γ 和 β 是可学习的缩放与偏移参数,这里 γ 作为衡量空间通道变化程度的指标,其数值越大,表示该通道内的特征波动越显著,包含的信息越丰富。此外,为使每个通道在空间注意力分配中具有相对权重的解释能力,使用 $Softmax$ 函数对所有通道的 γ 值进行加权归一化,得到通道注意力权重 W_r 。

最后,将归一化后的通道权重 W_r 与归一化特征 X_{out}^s 相结合,经过 $Sigmoid$ 激活函数以增强非线性表达,并送入门控函数,如公式(3)所示:

$$W = Gate(Sig(W_r \cdot X_{out}^s)) \quad (2)$$

式中 W 代表空间注意力权重图, $Gate$ 表示门控函数, Sig 表示激活函数,因此模型能够根据每个通道空间特征的变化程度,自适应地增强或抑制不同区域的响应,有效关注图像中的显著区域。随后将特征图沿通道维度划分为多个子组,分别与对应的空间权重进行加权融合,使每个分组在空间上可分,实现更

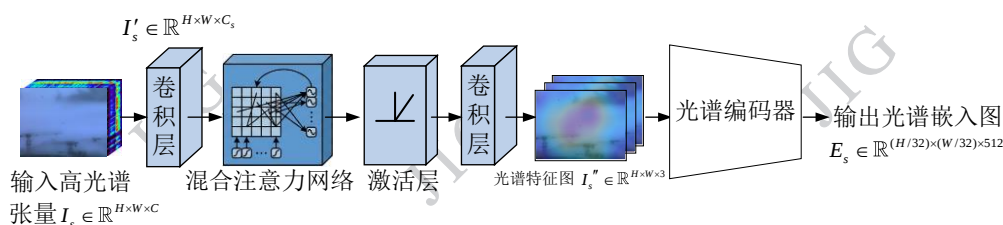


图2 混合嵌入网络结构

Fig. 2 Architecture of HEN

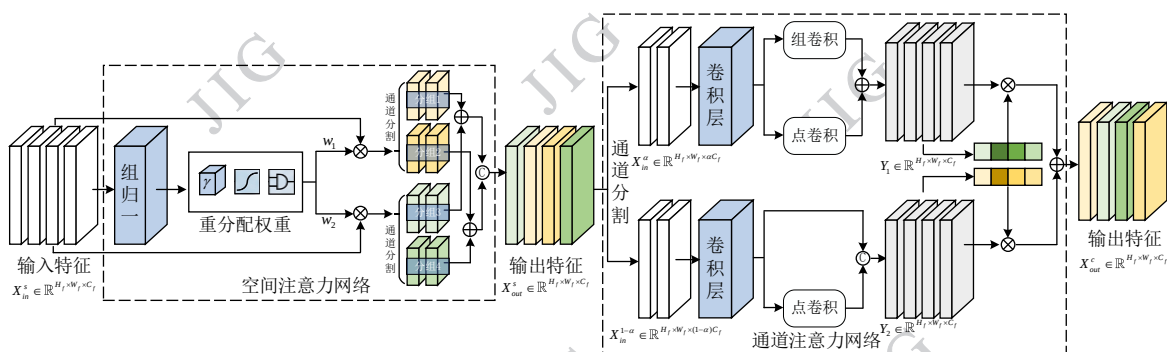


图3 混合注意力网络结构

Fig. 3 Architecture of HAN

细粒度的空间选择性关注。融合后的分组特征被拼接后传递至后续网络。

在高光谱图像中,不同光谱通道对特定目标的响应强度存在显著差异,这种波段间的异质性与互补性为特征选择与增强提供了关键依据。因此,为强化模型对判别性光谱信息的捕获能力,引入通道注意力模块(channel attention module, CAM),用于建模通道之间的语义依赖关系,并实现对重要光谱通道的显式增强,具体结构如图3所示。

给定输入特征图 $X_{in} \in \mathbb{R}^{H_j \times W_j \times C}$,沿通道维度进行两路划分,分别输入并行分支。其计算过程如下所述:

$$X_{up} = \text{Conv}(X_{in}^{\alpha}), X_{low} = \text{Conv}(X_{in}^{1-\alpha}) \quad (3)$$

式中 Conv 为卷积操作, $X_{in}^{\alpha}, X_{in}^{1-\alpha}$ 分别代表划分后通道数分别为 αC 和 $(1-\alpha)C$ 的输入特征, α 为超参数, $0 \leq \alpha \leq 1$ 。每一路特征通过共享的卷积层提取局部上下文信息,随后分别进行组卷积和点卷积操作,以捕捉不同层次的通道间依赖关系,可以表示为

$$Y_1 = \text{GC}(X_{up}) + \text{PC}(X_{up}) \quad (4)$$

$$Y_2 = \text{PC}(X_{low}) + X_{low} \quad (5)$$

式中 GC 代表组卷积, PC 代表点卷积。为自适应融合两分支输出 Y_1 和 Y_2 ,进行全局平均池化得到通道

向量 S_1 和 S_2 :

$$S_1 = \text{Pooling}(Y_1), S_2 = \text{Pooling}(Y_2) \quad (6)$$

式中 Pooling 代表全局平均池化。将两者拼接后通过全连接层和 Softmax 激活函数,生成权重系数 β_1 和 β_2 ,进行两个分支输出特征的加权融合:

$$X_{out}^c = \beta_1 Y_1 + \beta_2 Y_2 \quad (7)$$

式中 X_{out}^c 为输出特征。通道注意力模块采用双路结构,通过组卷积与点卷积,在扩展特征感受野的同时保持较高的参数效率,捕获不同层次的光谱上下文信息;模块内嵌的自适应通道加权机制,能够通过可学习的权重系数动态融合多路径特征,使模型聚焦判别性关键光谱通道,有效提升了高光谱特征的表征能力与任务适应性。

2.4 提出的数据集

为在复杂真实场景下进行性能评估,构建高光谱空中目标跟踪数据集 AHOT。数据采集采用短波红外波段的高光谱成像相机,以帧式推扫方式获取高光谱图像序列。每幅图像包含 5 个连续的光谱波段,覆盖 900 纳米至 1700 纳米的短波红外区间,空间分辨率统一为 640×512 像素。数据集共包含训练集视频 47 段,共 12294 张图像,测试集视频 26 段,共 6410 张图像。图 4 给出了数据集的典型图像与光谱

实例。拍摄过程中,相机的曝光时间、白平衡等关键参数均保持一致,确保了数据在光谱与辐射尺度上的一致性,为算法研究提供了可靠的实验基准。所有数据均在500米至2公里的远距离条件下获取,以模拟实际监测中空中目标的典型观测距离。拍摄背景涵盖天空、城市建筑等多种典型场景,增强了数据集的场景异质性与挑战性,能够有效检验算法在复杂背景干扰下对目标的感知与跟踪鲁棒性。

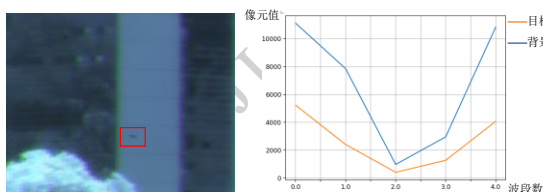


图4 AHOT数据集典型图像与光谱示例

Fig. 4 Typical image and spectral example from the AHOT dataset

3 实验与分析

3.1 数据集与评价指标

采用公开的高光谱目标跟踪数据集 HOTC-2024-NIR、HOTC-2024-RedNIR 和构建的 AHOT 数据集作为主要实验数据来源。其中 HOTC-2024-NIR、HOTC-2024-RedNIR 数据集的具体细节如下所示:

1) NIR: 包含 25 个光谱波段,覆盖约 700 - 1000 nm 的近红外区域,共有训练图像 12515 张,测试图像 5049 张;

2) RedNIR: 包含 15 个光谱波段,覆盖约 680 - 900 nm 的光谱区域。共有训练图像 15386 张,测试图像 8872 张;

该数据集包含 14 个属性,包括背景杂波(BC)、相机运动(CM)、变形(DEF)、快速运动(FM)、光照变化(IV)、低对比度(LC)、低亮度(LL)、相似颜色(OSC)、遮挡(OCC)、旋转(OPR)、视野外(OV)、反射(REF)、小目标(ST)和尺度变化(SV)。评估采用常见的 2 项评价指标:成功概率图曲线下面积(AUC)和阈值为 20 像素时的准确率。成功概率图描述的是超过不同交并比(IOUS)阈值的帧的百分比,形成阈值范围从 0 到 1 的成功概率曲线。准确率曲线基于预测边界框中心与实际边界框中心之间的像素偏移量,曲线由中心差异小于指定像素阈值(阈值范围

最高可达 50 像素)的帧的百分比生成。通常使用阈值为 20 的准确率曲线对跟踪器进行排名,记为 DP₂₀。

3.2 实验细节

提出的方法基于 PyTorch 库实现,使用硬件显卡为 NVIDIA RTX 4090 GPU。网络配置方面,使用 Hiera-L 作为图像编码器,使用 PVT 的轻量级版本作为光谱编码器。训练过程中,为了标准化输入,按照(Ravi 等, 2024)中的做法将输入图像的大小调整为 1024 × 1024。冻结图像编码器的参数,对其他模块进行微调,共运行 40 个 epoch。采用 Adam 优化器,学习动量设置为 0.1,学习率设置 5×10^{-6} ,使用余弦调度调整训练学习率。此外,对 HOTC-2024 和 AHOT 数据集进行像素级的掩码标注,确保分割模型的正常训练。

3.3 对比实验

3.3.1 定量评估

为了验证提出算法的有效性,在公开数据集 HOTC-2024 和构建数据集 AHOT 上进行对比实验。对比方法包括 RGB 领域中的先进跟踪方法 ARTrack (Wei 等, 2023)、ODTrack (Zheng 等, 2024)、Mixformer (Cui 等, 2024)、SAMURAI (Yang 等, 2024) 和 HOT 领域中的先进跟踪方法 SEE-Net (Li 等, 2023)、S3T-FFS (Wang 等, 2026)。结果如表 1 所示。

在 HOT-2024-NIR 数据集上,本文方法在 AUC 和 DP₂₀ 上分别取得 0.764 和 0.974, AUC 指标仅次于 S3T-FFS (0.790), 优于其他方法,如 SAMURAI (AUC: 0.756, DP₂₀: 0.955) 和 ODTrack (AUC: 0.748, DP₂₀: 0.927)。相较于 MixFormer 和 ARTrack,本文方法在两个指标上均有提升,证明在处理多波段近红外图像时具备良好的特征学习与目标感知能力。在 HOT-2024-RedNIR 数据集上,本文方法的 AUC 达到 0.603, DP₂₀ 达到 0.841, 优于现有对比方法。在构建的高光谱空中目标跟踪数据集 AHOT 上,多数对比方法出现性能下降。ODTrack 的 AUC 和 DP₂₀ 分别降至 0.237 与 0.477, MixFormer 下降至 0.327 与 0.563, 而相对鲁棒的 SAMURAI 和 ARTrack 降至 0.337 与 0.722 和 0.365 与 0.651。相比之下,本文方法在 AUC 和 DP₂₀ 上分别达到 0.388 和 0.877, 在对比方法中表现最优,证明本文方法能够更好地应对空中场景中存在的光照复杂、目标尺寸小、背景混杂等挑战。

与主流跟踪方法相比,本文方法挖掘高光谱图像中丰富的光谱维度信息,构建光谱嵌入网络,实现目标在不同波段下表征的一致性与互补性,使得模型在背景混淆等伪装场景中仍能保持稳定的判别能力。其次,为了解决直接迁移SAM到高光谱跟踪场景中存在的波段感知能力缺乏问题,设计混合嵌入网络对原始光谱图像进行压缩处理,引入混合注意力机制融合通道注意力与空间注意力,动态建模不同波段之间的空间依赖与语义联系,为现有SAM跟踪框架注入光谱提示信息,实现精准目标定位。

为进一步验证本文方法的推理效率与当前主流

高光谱目标跟踪方法在模型规模以及推理速度两个维度进行对比,评价指标包括模型参数量(M)以及模

型每秒处理的帧数(FPS)。实验结果如表2所示。MixFormer和SEE-Net分别以58.55MB和53.93MB的模型体积展现出较好轻量化程度,其中MixFormer也取得了最高的帧率(30.9FPS),显示出较好的实时处理能力。然而,这类方法通常在结构建模能力与跟踪精度上存在一定牺牲,难以兼顾复杂场景中的鲁棒性需求。

表 1 定量评估

Table 1 Quantitative evaluation

方法	出版	HOT-2024-NIR 数据集		HOT-2024-RedNIR 数据集		AHOT 数据集	
		AUC	DP_20	AUC	DP_20	AUC	DP_20
ODTrack (Zheng等,2024)	AAAI 2024	0.748	0.927	0.500	0.625	0.237	0.477
MixFormer (Cui等,2024)	TPAMI 2024	0.689	0.866	0.394	0.484	0.327	0.563
ARTrack (Wei等,2023)	CVPR 2023	0.640	0.784	0.423	0.540	0.365	0.651
SAMURAI (Yang等,2024)	Arxiv 2024	0.756	0.955	0.590	0.800	0.337	0.722
SEE-Net (Li等,2023)	TIP 2023	0.471	0.701	0.332	0.433	0.337	0.500
S3T-FFS (Wang等,2026)	TGRS 2026	0.790	0.973	0.550	0.721	0.347	0.608
本文方法	-	0.764	0.974	0.603	0.841	0.388	0.877

注:加粗字体为每列最优值,短线表示本文尚未正式出版。

相较之下,本文方法的模型体积为237.00MB,虽为所有方法中最大,但由于引入SAM以及光谱嵌入网络,在增强模型结构表达能力、提升跟踪性能方面起到关键作用。在推理速度方面,本文方法达到15.4FPS,略低于SAMURAI(24.0FPS)与ODTrack(23.6FPS)等方法,能够满足绝大多数实际空中监控及高光谱视频分析任务对准实时性的需求。

此外,与现有主流方法进行基于属性的对比实验,对比曲线图如表3所示。根据图中的属性成功率曲线,与当前主流的高光谱目标跟踪算法进行深入对比。本文方法在大多数属性场景中均取得领先表现。具体而言,在BC、CM、OCC、LC和LL等典型感知挑战下,本文方法均获得较高的成功率。在低光照条件下,取得0.809的成功率,显著优于S3T-FFS(0.695)和其他对比方法,表明在信息退化场景中仍能保持稳定跟踪。在复杂动态变化方面,如FM

和DEF中,分别取得0.730和0.735的成功率,展现出良好的时空建模能力。在SV方面,成功率为0.856,领先于其他方法,进一步表明本文提出的具备光谱提示网络的跟踪模型在处理空间变化时的优势。

针对一些具有高度挑战性的场景,如OSC、REF与OPR中,同样取得优异表现。在旋转类干扰中,本文方法准确率为0.772,与最佳方法SAMURAI几乎持平。在小目标跟踪任务中,本文方法取得0.403的准确率,表现优于大部分方法,体现出对低分辨率目标的良好感知能力。

总体来看,本文方法在14个属性场景中的综合表现较好,大多数情况下优于当前先进方法。在典型高动态高遮挡环境中,展现出更强的结构感知能力和复杂场景的鲁棒性,凸显本文方法在高光谱目标跟踪任务中的实际应用价值和推广前景。

为了直观对比不同高光谱目标跟踪方法在复杂场景中的表现,选取多个具有代表性的图像序列进行可视化分析,图5展示不同方法在连续帧中生成的目标预测框(每种方法以不同颜色标注)与真实边框的对比情况。对比方法包括 ODTrack、ARTrack、MixFormer、SEE-Net、S3T-FFS 和 SAMURAI,本文方法以红色框标出。从第一行的可视化结果可以看出,该场景中背景较为单一、目标运动平稳,多数方法在初始阶段均能准确跟踪。第二行场景为复杂背景干扰下的小目标序列,包含高密度结构纹理与伪目标区域,具有较强的干扰性。在第0015与0114帧中,多数对比方法(如 ARTrack、S3T-FFS、ODTrack)预测框发生了显著偏移,甚至出现框重叠或漂移问题,导致目标跟踪失败。本方法较其他方法有所改观但总体效果仍然较差,其原因可能是由于玻璃幕

墙反射淹没目标纹理,SAM原有的视觉提示信息失效,尽管设计了光谱提示的模块,但依旧破坏了时

表2 与其他方法的实时性及参数量对比
Table 2 Comparison of real-time performance and parameter size with other methods

方法	Size/M	FPS
ODTrack (Zheng 等, 2024)	92.12	23.6
MixFormer (Cui 等, 2024)	58.55	30.9
ARTrack (Wei 等, 2023)	173.10	20.1
SAMURAI (Yang 等, 2024)	224.00	24.0
SEE-Net (Li 等, 2023)	53.93	12.0
S3T-FFS (Wang 等, 2026)	97.27	22.6
本文方法	237.00	15.4

注:加粗字体为每列最优值

表3 与其他方法的基于属性的对比
Table 3 Comparison with other methods based on attributes

方法	BC	CM	DEF	FM	IV	LC	LL	OSC	OCC	OPR	OV	REF	ST	SV
ODTrack	0.525	0.437	0.761	0.337	0.444	0.643	0.688	0.527	0.498	0.728	0.057	0.042	0.305	0.670
MixFormer	0.446	0.289	0.523	0.450	0.335	0.092	0.590	0.473	0.440	0.591	0.058	0.044	0.205	0.433
ARTrack	0.368	0.316	0.269	0.336	0.470	0.610	0.519	0.443	0.400	0.549	0.083	0.253	0.336	0.594
SAMURAI	0.581	0.574	0.508	0.746	0.607	0.143	0.800	0.604	0.554	0.770	0.067	0.259	0.469	0.854
SEE-Net	0.361	0.155	0.297	0.279	0.289	0.078	0.540	0.364	0.319	0.561	0.007	0.039	0.100	0.492
S3T-FFS	0.557	0.633	0.578	0.373	0.586	0.613	0.609	0.557	0.568	0.587	0.065	0.325	0.427	0.716
本文方法	0.632	0.557	0.759	0.733	0.598	0.438	0.805	0.621	0.566	0.772	0.069	0.228	0.493	0.856

注:加粗字体为每列最优值

空谱的联合建模,导致跟踪失败。日后研究可以引入额外的位置预测模型,在纹理失效时依靠失效前的运动状态进行下一步预测,缓解跟踪框的漂移情况。第三行展示的是典型的密集树木干扰场景,目标被树叶、枝干等纹理较强的自然物体干扰,背景动态变化剧烈。多种方法(如 SEE-Net、S3T-FFS)在复杂树木区域中误识别或预测散乱,出现框体漂移或异常变大。而本文方法的跟踪框可框出正确目标,体现在自然干扰场景下对目标特征的识别能力和鲁棒性的优势。第四行为高目标复杂性与空间变化场景。第0118、0183和0288帧中,多个方法在处理远距离小目标或强结构干扰时均出现失误,预测框偏离甚至漂移至背景区域。而本文方法在这些典型困难帧中,保持预测精度与边界对齐,证明在

极端挑战条件下所具备的强大鲁棒性。综上所述,本文方法在跟踪框连续性和鲁棒性方面均存在优势。相比于现有主流高光谱目标跟踪方法常出现的预测框漂移、边界不稳定、遮挡误识等问题,本文方法借助光谱注意力与嵌入模块,能更有效地区分目标与复杂背景,提高目标聚焦与建模能力。

3.4 消融实验

使用构建的高光谱无人机目标跟踪数据集 AHOT 进行评估。从2项评价指标讨论提出模块在光谱特征嵌入、注意力嵌入的有效性,同时研究不同嵌入维度、模型尺寸以及通道与空间注意力对高光谱目标跟踪的影响。

注:加粗字体为每列最优值

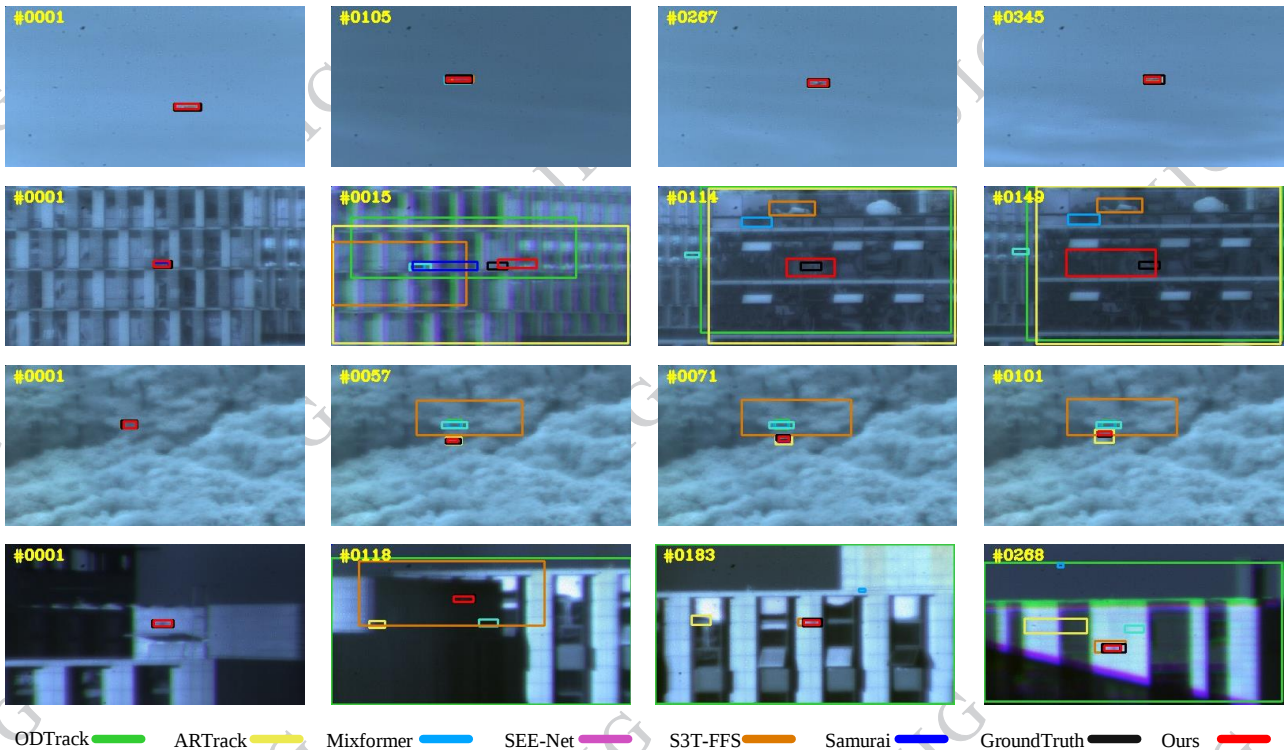


图5 跟踪可视化结果
Fig. 5 Tracking visualization results

3.4.1 提出各模块的有效性

为验证所提出各模块对整体模型性能的贡献,设计模块消融实验,实验结果如表4所示,分别比较“基线模型”、“基线模型+HEN”以及“基线模型+HEN+HAN”三种设置下的性能表现,评价指标采用 AUC 和 DP_20。从实验结果可见,基线模型的 AUC 和 DP_20 分别为 0.331 和 0.721。引入光谱嵌入模块后,AUC 提升至 0.357,DP_20 提升至 0.780,证明高光谱维度的信息显著增强模型对目标区域的区分能力,提升目标外观建模效果和匹配准确性。

进一步加入混合注意力机制后(包括空间注意力与通道注意力的耦合设计),AUC 达到 0.388,DP_20 提升至 0.878,体现融合注意力机制可增强模型对空间区域与光谱特征的联合建模能力,有效提高模型在复杂环境下的目标识别与定位鲁棒性。

3.4.2 混合注意力机制中特征嵌入维度的有效性

为进一步探究混合注意力机制中特征嵌入维度的设置对模型性能的影响,对嵌入维度进行参数消融实验,选取嵌入维度为 8、16、32 和 64 的四组配置进行评估。嵌入维度指的是图2混合嵌入网络第一个卷积层的输出通道维度 C_i 。实验结果如表5与表

表4 提出方法的消融实验

Table 4 Ablation studies of the proposed methods

方法	AUC	DP_20
基线	0.331	0.721
基线+HEN	0.357	0.780
基线+HEN+HAN	0.388	0.878

注:加粗字体为每列最优值

6所示。模型在嵌入维度为 16 时表现最优,AUC 在 AHOT 数据集上达到 0.388,在 HOT-2024-RedNIR 上达到 0.603,DP_20 在 AHOT 数据集上达到 0.878,在 HOT-2024-RedNIR 数据集上达到 0.841,显著优于其他维度设定。这表明在适中维度下,融合的空间与通道注意力更易实现有效的信息表达与判别建模,提升了跟踪精度和鲁棒性。

在 AHOT 数据集上,当嵌入维度较小(如 8 维)时,特征表达能力受限,难以充分捕捉复杂光谱空间信息,仅实现 AUC 0.361 与 DP_20 0.785 的次优表现。当嵌入维度逐步增大至 32 和 64 时,模型性能呈现下降趋势,在 64 维设定下,AUC 降至 0.327,DP_20 下降至 0.720。在 HOT-2024-RedNIR 数据集上,随着嵌入维度从 8 提升到 16,AUC 由 0.585 提升至

0.603, DP_20 也由 0.795 显著提升至 0.841, 说明过低维度会限制特征表征能力; 当维度继续增大到 32 时 AUC 进一步小幅提升至 0.606, 但 DP_20 下降至 0.829, 反映更高维度可能引入冗余或过拟合, 使定位稳定性受影响; 继续增大到 64 后, AUC 下降至 0.602, DP_20 为 0.833。过高的嵌入维度反而可能引入冗余特征和注意力噪声, 造成训练不稳定或对关键目标区域关注度分散。此外, 高维特征还可能带来额外的计算开销与过拟合风险, 影响模型的泛化能力。

表 5 HAN 中的特征嵌入维度实验 (AHOT 数据集)
Table 5 Ablation studies on the feature embedding dimension in HAN (AHOT dataset)

混合注意力特征嵌入维度	AUC	DP_20
8	0.361	0.785
16	0.388	0.878
32	0.359	0.782
64	0.327	0.720

注: 加粗字体为每列最优值

表 6 HAN 中的特征嵌入维度实验 (HOT-2024-RedNIR 数据集)
Table 6 Ablation studies on the feature embedding dimension in HAN (HOT-2024-RedNIR dataset)

混合注意力特征嵌入维度	AUC	DP_20
8	0.585	0.795
16	0.603	0.841
32	0.606	0.829
64	0.602	0.833

注: 加粗字体为每列最优值

3.4.3 不同规模 SAM 的有效性

为探究不同规模的 SAM 在高光谱目标跟踪任务中的性能差异, 选取四种不同尺寸的预训练模型版本 (SAM2-T、SAM2-S、SAM2-B 和 SAM2-L) 进行对比评估, 结果如表 7 所示。SAM2-L 模型表现最优, 在 AUC 和 DP_20 指标分别为 0.388 和 0.878 的最高成绩, 证明更大的模型容量可以刻画目标与背景之间的细节差异, 提升模型对复杂场景的适应能力和定位精度。相较之下, SAM2-S 与 SAM2-B 的性能存在局限性, SAM2-T 的性能甚至优于较大模型, 这可能与模型过拟合或者模型性能依赖训练集质量

有关。

综上所述, 大尺寸模型 (如 SAM2-L) 具备更强的特征表达与区域建模能力, 在复杂场景中优势明显, 而轻量模型 (如 SAM2-T) 在保证较高性能的同时具有更好部署效率。中等尺寸模型存在一定波动, 因此可以结合实际应用需求针对性选择模型结构, 在性能与效率之间取得最佳权衡。

表 7 SAM 模型尺寸实验
Table 7 Ablation studies on model size in SAM

模型尺寸	AUC	DP_20
SAM2-T	0.388	0.823
SAM2-S	0.386	0.786
SAM2-B	0.368	0.804
SAM2-L	0.388	0.878

注: 加粗字体为每列最优值

3.4.4 混合注意力机制的有效性

为探究混合注意力机制中不同注意力模块对模型性能的贡献, 在保持其它结构不变的前提下, 分别使用仅含空间注意力、仅含通道注意力以及去除注意力机制 (w/o 混合注意力) 的设置进行消融实验, 并与完整模型进行对比, 实验结果如表 8 所示。移除注意力机制后 (w/o 混合注意力), 模型性能出现明显下降, AUC 为 0.357, DP_20 为 0.780, 证明原有模型在目标特征建模方面存在一定缺陷, 难以充分抑制背景干扰与提取关键区域。当引入空间注意力模块后, 模型性能得到显著提升, AUC 提高至 0.370, DP_20 上升至 0.805, 表明空间注意力有效引导模型关注目标区域的结构与位置分布, 在复杂场景中有助于增强定位精度。相比之下, 仅引入通道注意力模块的效果有限, AUC 为 0.359, DP_20 为 0.781, 提升幅度较小。表明仅使用通道间特征重加权无法充分捕获局部空间变化对目标形态和边界的影响, 在高光谱目标跟踪任务中作用相对有限。最终, 融合空间注意力与通道注意力组成混合注意力机制后, 模型性能达到最优, AUC 提高至 0.388, DP_20 提升至 0.878, 超越单一形式注意力的表现。证明空间与通道注意力之间的互补性可以捕捉目标的判别信息, 提升模型对目标的定位精度与鲁棒性。

3.4.5 不同超参数 α 取值的影响

为了探究通道注意力模块中超参数 α (用于控

© 中国图象图形学报版权所有

表8 不同注意力机制的实验

Table 8 Experiments with different attention mechanisms

方法	AUC	DP_20
w/o 混合注意力	0.357	0.780
SAB	0.370	0.805
CAM	0.374	0.817
本文方法	0.388	0.878

注:加粗字体为每列最优值

制通道划分比例)对跟踪性能的影响,在保持其余设置不变的条件下进行不同取值的消融实验。实验结果如表9所示。当 α 取值为0.50时取得最高的AUC(0.374),相比 α 为0.25与0.75分别提升0.4%与0.6%;在DP 20指标上三者差异较小,分别为0.818、0.817与0.819。综合来看, α 对整体准确率(AUC)影响较大,其中 α 取值为0.50能在特征变换分支与信息保留分支之间取得更合适的平衡,从而带来更优的综合表现。

表9 不同超参数 α 取值影响的实验Table 9 Experiment on the influence of different values of hyperparameter α

取值	AUC	DP_20
0.25	0.370	0.818
0.50	0.374	0.817
0.75	0.368	0.819

注:加粗字体为每列最优值

4 结 论

本文面向复杂场景下高光谱目标跟踪易受背景杂波干扰、边界框引入混合像元导致光谱特征污染,以及缺乏长时记忆而在遮挡重现后易丢失目标等关键难题,提出了基于SAM 2的高光谱目标跟踪框架。该方法利用SAM 2的高精度掩码实现目标与背景的像素级分离,从源头抑制背景光谱噪声;同时借助记忆库与记忆注意力建立跨帧长时关联,提升遮挡、形变及消失重现场景下的重捕获能力。为弥补SAM 2对高维光谱差异感知不足,本文设计混合嵌入网络快速适配高光谱数据分布,并引入混合注意力与光谱嵌入模块,将目标光谱信息编码为提示向量并学

习波段间深层语义表征,实现空间—时序—光谱特征的协同融合。实验结果表明,所提方法在三个数据集上相较六种最新方法取得更优表现。此外,构建首个高光谱空中目标跟踪数据集 AHOT,为后续研究与公平评测提供了重要基准。综上,本文方法为高光谱目标跟踪提供了有效的新思路。

参考文献(References)

- Altaf M A and Kim M Y. 2025. Multiple object detection and tracking in autonomous vehicles: A survey on enhanced affinity computation and its multimodal applications. *ICT Express*, 11 (4): 809-818 [DOI:10.1016/j.ict.2025.06.005]
- Chen Y, Yuan Q, Tang Y, Wang X, Xiao Y, He J, et al. 2025. ProFiT: A prompt-guided frequency-aware filtering and template-enhanced interaction framework for hyperspectral video tracking. *ISPRS Journal of Photogrammetry and Remote Sensing*, 226: 164-186 [DOI:10.1016/j.isprsjprs.2025.05.008]
- Cheng H K, Oh S W, Price B, Schwing A, and Lee J Y. 2023. Tracking anything with decoupled video segmentation // *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Paris, France: IEEE: 1316-1326 [DOI: 10.1109/ICCV51070.2023.00127]
- Cui Y T, Jiang C, Wang L M and Wu G S. 2024. MixFormer: End-to-End Tracking With Iterative Mixed Attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46 (6): 4129-4146 [DOI:10.1109/TPAMI.2024.3349519]
- He W, Wu Z L, Yokoya N, and Yuan X. 2024. An interpretable and flexible fusion prior to boost hyperspectral imaging reconstruction. *Information Fusion*. 111: 102528 [DOI: 10.1016/j.inffus. 2024. 102528]
- 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.]
- Hou Z, Li W, Zhou J, and Tao R. 2022. Spatial - spectral weighted and regularized tensor sparse correlation filter for object tracking in hyperspectral videos. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1-12 [DOI:10.1109/TGRS.2022.3215431]
- Hu J, Shen L and Sun G. 2018. Squeeze-and-excitation networks // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE: 7132-7141 [DOI: 10.1109/CVPR.2018.00745]
- Hu S Y, Zhao X, and Huang K Q. 2024. Visual intelligence evaluation techniques for single object tracking: a survey. *Journal of Image and Graphics*, 29 (08): 2269-2302 (胡世宇, 赵鑫, 黄凯奇. 2024. 单目标跟踪中的视觉智能评估技术综述. *中国图象图形学* 2024. 单目标跟踪中的视觉智能评估技术综述. 中国图象图形学 © 中国图象图形学报版权所有)

- 学报, 29(08): 2269-2302 [DOI:10.11834/jig.230498]
- Jiang X, Wang X Y, Sun C, Zhu Z L, and Zhong Y F. 2024. A channel adaptive dual Siamese network for hyperspectral object tracking. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1-12 [DOI:10.1109/TGRS.2024.3378165]
- Kirillov A, Mintun E, Ravi N, Mao H Z, Rolland C, and Gustafson L. 2023. Segment anything//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Paris, France: IEEE: 4015-4026 [DOI:10.1109/ICCV51070.2023.00371]
- 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, Wang J, Lu J F, Qian Y T. 2020. BAE-Net: A band attention aware ensemble network for hyperspectral object tracking//*2020 IEEE International Conference on Image Processing*. Abu Dhabi, United Arab Emirates: IEEE: 2106-2110 [DOI:10.1109/ICIP40778.2020.9191105]
- Liu E Q, Huang W, Xu Y, Yang M, Gao B, and Mo D R. 2025. Hyperspectral Camouflaged Identification Driven by Spatial-Spectral Characteristics. *Spectroscopy and Spectral Analysis*, 45 (10): 2942-2949 (刘恩勤, 黄维, 徐勇, 杨漫, 高兵, 莫定儒. 2025. 空谱特征驱动的高光谱遥感迷彩伪装识别. *光谱学与光谱分析*, 45(10): 2942-2949) [DOI:10.3964/j.issn.1000-0593(2025)10-2942-08]
- Liu Z, Wang X, Zhong Y, Shu M, and Sun C. 2022. SiamHYPER: Learning a hyperspectral object tracker from an RGB-based tracker. *IEEE Transactions on Image Processing*, 31: 7116-7129 [DOI:10.1109/TIP.2022.3216995]
- Lv Y Q, Zhang J, Dai Y C, Li A X, Barnes N, and Fan D P. 2023. Toward deeper understanding of camouflaged object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 33 (7): 3462-3476 [DOI:10.1109/TCSVT.2023.3234578]
- Qian K, Zhou J, Xiong F C, Zhou H X, and Du J. 2018. Object tracking in hyperspectral videos with convolutional features and kernelized correlation filter//*Proceedings of the International Conference of Smart multimedia*. Toulon, France: Springer: 11010: 308-319 [DOI:10.1007/978-3-030-04375-9_26]
- Ravi N, Gabeur V, Hu Y T, Hu R H, Ryali C, Ma T Y, et al. 2024. Sam 2: Segment anything in images and videos[EB/OL].[2024-10-28].
<https://arxiv.org/pdf/2408.00714>
- Sun C, Wang X Y, Liu Z Q, Wan Y T, Zhang L P, and Zhong Y F. 2023. Siamohot: A lightweight dual siamese network for onboard hyperspectral object tracking via joint spatial-spectral knowledge distillation. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1-12 [DOI:10.1109/TGRS.2023.3307052]
- Sun W, Tan Y, Li J, Hou S W, Li X B, and Shao Y Z. 2025. Hotmoe: Exploring sparse mixture-of-experts for hyperspectral object tracking. *IEEE Transactions on Multimedia*, 27: 4072-4083 [DOI: 10.1109/TMM.2025.3535339]
- Sun W D and Han X L. 2025. Computational imaging of high-spatial-resolution hyperspectral remote sensed images: From fusion to spectral super-resolution. *National Remote Sensing Bulletin*, 29 (06): 1636-1648 (孙卫东, 韩晓琳. 2025 高分高光谱遥感图像计算成像: 从融合到光谱超分. *遥感学报*, 29(06): 1636-1648) [DOI:10.11834/jrs.20254209]
- Wang H Z, Li W, Xia X G, Du Q, and Tian J. 2025. Ssf-net: Spatial-spectral fusion network with spectral angle awareness for hyperspectral object tracking. *IEEE Transactions on Image Processing*, 34: 3518-3532 [DOI:10.1109/TIP.2025.3572812]
- Wang H Z, Li W, Xia X G, Cui B L, Shi Z C, and Lin H Y. 2026. Bring KANs to Tracker: a Nonlinear Fusion Strategy over Frequency Decoupling for Temporal Hyperspectral Object Tracking. *IEEE Transactions on Geoscience and Remote Sensing*, 64: 1-15 [DOI:10.1109/TGRS.2026.3658509]
- Wang H Z, Li W, Xia X G, and Du Qian. 2025. BihoT: A large-scale dataset and benchmark for hyperspectral camouflaged object tracking. *IEEE Transactions on Neural Networks and Learning Systems*, 36(9): 16392-16406. [DOI:10.1109/TNNLS.2025.3564059]
- Wang W H, Xie E Z, Li X, Fan D P, Song K T, Liang D, et al. 2021. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Montreal, QC, Canada: IEEE: 568-578 [DOI:10.1109/ICCV48922.2021.00061]
- Wang Y, Mei S, Ma M Y, Liu Y H, and Su Y R. 2025. HTACPE: A hybrid transformer with adaptive content and position embedding for sample learning efficiency of hyperspectral tracker. *IEEE Transactions on Multimedia*, 27: 2384-2398 [DOI: 10.1109/TMM.2024.3521819]
- Wei X, Bai Y F, Zheng Y C, Shi D H, Gong Y H. 2023. Autoregressive visual tracking//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, TN, USA: IEEE: 9697-9706 [DOI:10.1109/CVPR52729.2023.00935]
- Wu Y X and He K M. 2018. Group normalization//*Proceedings of the European Conference on Computer Vision*. Munich, Germany: Springer: 3-19 [DOI:10.1007/978-3-030-01261-8_1]
- Xiong F C, Zhou J, Qian Y T. 2020. Material based object tracking in hyperspectral videos. *IEEE Transactions on Image Processing*, 29: 3719-3733 [DOI:10.1109/TIP.2020.2965302]
- Yang Y, Wang X Z, Liu X, Huang S Y, Liu Z Y, Wang S Z. 2025. Dual-branch U-Net with hybrid attention for hyperspectral pansharpening. *Journal of Image and Graphics*, 30 (04): 989-1002. (杨勇, 王晓争, 刘轩, 黄淑英, 刘紫阳, 王书昭. 面向高光谱全色锐化的混合注意力双分支U型网络. 2025 中国图象图形学报, 30 (04): 989-1002) [DOI: 10.11834/jig.240410]
- Yang C Y, Huang H W, Chai W H, Jiang Z Y, and Hwang J N. 2024. Samurai: Adapting segment anything model for zero-shot visual

tracking with motion-aware memory.[EB/OL].[2024-11-30].
<https://arxiv.org/pdf/2411.11922>

Yin M, Wang F, Ye X, Meng Y D, and Fu Z Y. 2025. Memory-augmented sam2 for training-free surgical video segmentation//International Conference on Medical Image Computing and Computer-Assisted Intervention. Daejeon, South Korea: Springer: 328-337 [DOI:10.1007/978-3-032-05127-1_32]

Zhang Y, Zhao X J, and Wang D. 2019. Research advances on crop identification using hyperspectral remote sensing. China Agricultural Informatics. 31(5):1-12 (张影, 赵小娟, 王迪. 基于高光谱遥感的农作物分类研究进展. 中国农业信息. 31(5):1-12) [DOI:10.12105/j.issn.1672-0423.20190501]

Zheng Y Z, Zhong B N, Liang Q H, Mo Z Y, Zhang S P, and Li X X. 2024. Odtrack: Online dense temporal token learning for visual tracking[C]//Proceedings of the AAAI conference on artificial intel-

ligence. Vancouver, Canada: AAAI: 38(7): 7588-7596 [DOI:10.1609/aaai.v38i7.28591]

作者简介

王瀚正,男,博士研究生,研究方向为高光谱视频处理。E-mail:hzwangc@bit.edu.cn

李伟,男,教授,研究方向为高光谱图像处理。E-mail:liw@bit.edu.cn

张家铭,男,博士,研究方向为遥感器总体设计。E-mail:394463508@qq.com

何鸿涛,男,博士,研究方向为遥感器总体设计。E-mail:hehongtao@aliyun.com

黄荀,男,博士,研究方向为遥感器总体设计。E-mail:hxun15@outlook.com