ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 41,No. 1,Jan. 2026,pp. 28－52
DOI:10. 16337/j. 1004-9037. 2026. 01. 003

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

无人机多模态超宽谱认知仪研究

施云鹤，张小飞，吴启晖

（南京航空航天大学电子信息工程学院，南京 211106）

摘　要：本文设计了一种无人机（Unmanned aerial vehicle，UAV）多模态超宽谱认知仪，通过深度融合可见光、红外、合成孔径雷达（Synthetic aperture radar，SAR）及无线频谱等多模态传感器构建智能遥感系统，旨在攻克传统无人机遥感的根本性瓶颈：续航时间短严重制约探测范围、有效载荷不足限制多模态感知能力、机载算力薄弱导致实时处理延迟、通信容量有限阻碍高保真态势评估。本文设计方案针对续航挑战，采用活塞发动机与锂电池协同的混合能源构型，结合垂直起降（Vertical take-off and landing，VTOL）飞翼布局，显著提升航时效能；为应对载荷限制，开发复眼多目相机实现大视场高分辨率成像，集成W波段轻小型SAR突破亚毫米级振动补偿技术，支撑空-时-频多维度协同感知；为化解实时处理困境，基于时空配准框架与轻量化深度学习模型，构建数据层-特征层-语义层多层次融合机制，将低可观测目标检测精度提升至90％以上；针对通信瓶颈，创新生成式编码技术结合知识图谱驱动的态势重建，通过无参考质量评估模型量化语义保真度，实现超400倍压缩下的高保真三维态势生成。该仪器在国防侦察领域成功实现复杂电磁环境中隐蔽目标实时追踪，在应急救援中完成洪涝灾害监测与三维重建等关键任务，验证了多模态超宽谱认知在复杂场景的实用价值。

关键词：无人机；混合能源；多模态载荷集成；多模态融合；质量评估

中图分类号：V279；V243；TP79　　**文献标志码：**A

引用格式：施云鹤，张小飞，吴启晖．无人机多模态超宽谱认知仪研究[J]．数据采集与处理，2026，41(1)：28-52.
SHI Yunhe，ZHANG Xiaofei，WU Qihui. Research on UAV multi-modal ultra-wide spectrum cognitive instrument[J]. Journal of Data Acquisition and Processing，2026，41(1)：28-52.

引　　言

《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》指出需打造全球覆盖、高效运行的通信、导航、遥感空间基础设施体系[1]。国家自然资源部《关于加快测绘地理信息事业转型升级更好支撑高质量发展的意见》指出，强化无人机（Unmanned aerial vehicle，UAV）遥感布局，推进应急测绘保障能力“平急结合”，对于国家的生态环境监测、洪涝灾害监测和应急救援等具有重要意义[2]。国家自然科学基金委“十四五”优先发展领域“精准探测与信息融合处理”明确要围绕复杂环境和复杂目标信息获取与处理难题，探索多源融合探测成像、多维度稀疏信号处理、智能遥感信息处理与目标识别等新机理、新方法，为国家应急响应系统建设及应用拓展提供技术支撑[3]。

无人机遥感（Unmanned aerial vehicle remote sensing，UAVRS）[4-5]，即利用先进的无人驾驶飞行器、遥感传感器、遥测遥控、通信、全球定位系统（Global positioning system，GPS）差分定位等技术，实现自动化、智能化、专用化快速获取国土资源、自然环境、地震灾区、重要目标等空间遥感信息，完成遥感数

基金项目：国家自然科学基金（62427801）。
收稿日期：2025-11-12；**修订日期：**2026-01-15

据处理、建模和应用分析[6-8]。相比卫星遥感,无人机遥感具有分辨率高[9]、机动性强、不受云覆盖限制等优势,已成为世界各国争相研究的热点[10],将成为未来的主要航空遥感技术之一[11-20]。随着无人机遥感与国防安全、国土资源、农业、林业、水利、环境保护、城市规划及灾害防治等领域的深度融合,亟需向多模态超宽谱遥感探测方向发展。传统的无人机遥感面临着四大挑战:一是无人机续航时间受限,探测范围小;二是无人机有效载荷受限,探测模态少;三是无人机机载算力受限,实时处理难;四是无人机通信容量受限,实时评估难。因此,传统无人机遥感仪器亟需从单模态向多模态、从图像生成向态势生成、从短航时向长航时、从地面计算向空地协同计算跃升,设计制造出新型科研仪器-无人机多模态超宽谱认知仪,以推动认知科学、人工智能理论与无人机遥感的深度融合,探索目标与环境的多模态超宽谱特性,揭示面向目标与环境的多模态认知、多模态决策与多模态学习等作用机理,实现对目标与环境的多特性分析认知、多语义态势生成与实时自主航迹规划,为无人机遥感探测的迭代升级提供科学指导。

重要目标及其环境的特性表征、感知融合、计算成像是无人机遥感探测中的关键问题。本文通过无人机多模态超宽谱认知仪的研制,针对重要目标及其环境要素的多模态超宽谱特性关联表征、多模态多层次融合认知机理、多模态语义态势生成模型3个重要科学问题展开研究。低可观测目标探测问题是一个探索性强、难度大而又具有强烈背景需求的研究领域,强杂波背景、目标复杂运动、低辐射、天气复杂多变等会使得探测信号微弱,时频域难以区分,部分目标会呈现低可观测特性。多模态认知是研究机器怎样模拟人类“联觉”,分析不同感知通道的联系,以实现对目标及环境进行时空频以及语义知识等维度上的高效感知和综合理解。态势生成是指通过数据采集与分析获取实时或准实时的目标环境状态趋势信息,利用图形化、形象化、逼真化等可视化技术,将状态趋势信息转换成人眼可观测的态势图,以利于人机协同决策。其拟解决的关键技术问题是:轻小型高精度多模态探测载荷设计集成技术、轻量化高精度多模态融合协同认知技术、大压缩比高保真多模态混合态势生成技术。

综上所述,面向国家空间基础设施体系建设、应急测绘保障与精准探测等重大需求,开展无人机多模态超宽谱认知仪研制,不仅是推动无人机遥感技术从单一感知向智能认知跨越的必然趋势,也是解决当前探测范围有限、载荷模态单一、实时处理困难等瓶颈问题的关键途径。本文通过多模态超宽谱特性的关联表征、多层次融合认知机理探索以及语义态势生成模型构建,为重要目标及其环境的精准感知、智能解译与高效决策提供坚实的理论支撑与先进的技术途径。

1 研究现状

1.1 光学与红外载荷技术研究现状

多光谱载荷主要包含光学相机与红外相机。光学相机具有较高分辨率,可进行较清晰的成像。最早用于无人机遥感的传感器是可见光相机,可见光相机的视角扩展已成为无人机遥感特别关注的问题。2019年,Fujita等[21]提出了一种从光场相机阵列中提取、处理多视角图像的方法。2022年,Qi等[22]总结了多孔径光学成像系统的数学模型,以便于多目相机的设计与评估。红外热像仪是能够在恶劣天气条件或低照度下工作的设备,适合于火灾预警、目标跟踪等。2018年,文献[23]提出了一种基于热红外图像的空域监测系统用于监控目标空域的小型飞行物,通过开发的热红外图像增强算法可实时输出增强后视频序列,在低能见度情况下可对小型飞机、直升机和无人机等进行检测跟踪。2020年,Zhang等[24]将卷积神经网络模型用于VIF算法中;2024年,Wang等[25]结合传统计算机视觉中的算子方法与深度学习,通过梯度聚合残差稠密块和并行注意机制来驱动融合过程,可生成高质量的融合图像。

1.2 合成孔径雷达小型化与实时处理研究现状

合成孔径雷达(Synthetic aperture radar,SAR)是利用电磁波进行成像感知的传感器,不受云、雨、雾、雪等天气影响,能够全天时、全天候地获取地表的高分辨图像。2013年,Aguasca等[26]提出了一种小型无人机机载SAR平台(ARBRES),该平台具有重量轻、独立可操作的特点,工作在X或C频段,采用的是阶跃线性调频信号。2018年,美国喷气实验室研制了无人机上实时SAR处理系统,可以实现25 MB/s原始L波段的实时处理,并将数据转换为连续条带图像,且功耗不超过8 W[27]。2022年,Yocky等[28]探讨了使用小型无人机SAR为SPE三期项目提供快速、廉价的合成孔径雷达干涉(Interferometric synthetic aperture radar,InSAR)数据的可行性,并通过无人机实验采集的数据证明了InSAR高度测量精度;文献[29]为克服单一平台的局限性以及保证集群无人机合成孔径雷达系统能够执行先进的SAR任务,为集群无人机SAR系统提出了一种混合分辨率增强(Hybrid resolution enhancement,HRE)框架。

1.3 无线频谱监测与定位技术研究现状

机载无线频谱监测接收机开始主要应用于军事领域,主要为中型、大型无人机系统提供无线电信号搜索截获、测量分析和定位等能力,如美国诺斯罗普·格鲁曼公司研制的"战术雷达电子战系统"是一种无人机机载无源定位系统,能在3 s内对地面雷达定位,圆概率误差为50 m[30]。美国的"捕食者"、法国的"玛尔特""狐狸"、英国的"不死鸟""守望者"等,都可装载机载侦察设备,快速有效地获取战场信息。中国无人机系统如"翼龙"无人机、"W-50"暗剑遥控无人机等都可携带各种侦察、电子对抗设备等,以执行监视、侦查等任务。近年来,无人机无线频谱监测平台开始在民用领域得到快速发展,相比陆基监测平台具有受杂波影响小、多径少、移动性强等优势,同时能够克服卫星平台定位粒度粗、定位不够精准的局限,可有效提升辐射源定位的效率和成功率。根据观测量如强度、时间、方向等的不同,目标定位方法可分为波达方向(Direction of arrival, DOA)、到达时间(Time of arrival, TOA)、到达时间差(Time difference of arrival, TDOA)以及基于接收信号强度(Received signal strength-based localization, RSS)的定位。文献[31]提出了一种不需要任何矩阵运算的迭代方法,能够实现利用多传感器的DOA估计结果对三维空间中的目标进行定位。与以上传统定位不同,直接定位(Direct position determination,DPD)技术[32]通过利用信号数据域信息直接完成目标位置估计,避免两步定位方法中两步分离造成的信息损失,具有更高的定位精度。

1.4 多模态融合技术发展现状

无人机遥感是近年来快速发展的技术之一,当前主要集中可见光、红外、SAR雷达及无线频谱监测等遥感传感器单模态或跨模态设计,尚缺少多模态传感器无人机载集成设计研究,亟需从单模态传感器向多模态传感器设计跃升。

1.4.1 SAR与可见光图像融合的研究

SAR和可见光图像有着天然的互补关系,两者融合可以充分发挥各自的优势,既可以保留图像的光谱信息,又可以不受时间和天气的影响,更全面、更直观地反映地物属性,为后续图像分析以及信息提取提供支持。现如今,如何高效地将SAR和光学图像进行精准融合,从而获得一幅更全面的融合图像成为融合领域中的研究热点[33]。传统方法主要有基于多尺度变换、子空间和稀疏表示等方法,其中基于多尺度变换的常用方法有基于小波变换(Wavelet transform,WT)、金字塔变换(Pyramid)、非下采样剪切波变换(Non-subsampled shearlet transform,NSST)、非下采样轮廓波变换(Non-subsampled contourlet transform,NSCT)、主成分分析(Principal component analysis,PCA)、强度-色调-饱和度(Intensity-hue-saturation,IHS)以及非负矩阵分解(Non-negative matrix factorization,NMF)等。Zhang

等[34]将拉普拉斯金字塔与稀疏表示进行组合。Yehia等[35]将小波变换和IHS变换组合,能够将SAR数据注入可见光图像。Fu等[36]为了解决融合图像中存在严重的光谱和空间失真的问题,提出一种耦合增益注入法和引导滤波器的可见光和SAR图像融合方法,增强了图像的边缘细节。近年来,深度学习在计算机视觉领域表现突出,并应用于图像融合领域。2020年,Li等[37]首次采用双线性池化层的融合网络对光学和SAR图像提取的特征进行融合。Ye等[38]提出一种结构与优化分离的双流融合网络(Structure- and optimization-separated two-stream fusion network,SOSTF)新型融合框架,该算法能够有效地融合SAR和可见光图像的互补特征。美国NASA兰利研究中心[39]构建了一套光学-雷达融合的飞行器探测与跟踪系统与算法,用于实现空中碰撞回避的安全操作。

1.4.2 多光谱融合技术进展

传统可见光图像是指包含红-绿-蓝三通道的光学图像,具有高分辨率、低成本的优势。然而,可见光图像在暗光或黑夜、环境隐蔽性强的探测场景中感知性能会大打折扣。因此,产生了多光谱图像融合处理技术,借助可见光与红外等图像多谱段融合,完成高精度、智能化的环境与目标感知。2021年,中国科学院长春光学精密机械与物理研究所[40]提出了一种基于红外与可见图像融合对无人机等低空低速小目标的远程探测算法,该算法融合后的图像细节和目标都有增强,在单传感器视觉效果较差时仍有较好的检测精度。2022年,文献[41]针对红外-可见光不同成像机理和异构探测器空间分辨率不匹配导致的低质量融合成像问题,提出一种基于深度学习的红外-可见跨模态颜色融合网络,将融合任务表述为保持红外-可见光图像的结构和强度比问题。文献[42]提出了一种基于稀疏先验联合显著性检测和潜在低秩表示的四阶微分方程的红外-可见图像融合方法,该显著性算法具有较高的检测精度和良好的背景抑制能力,其信息熵和信噪比都有显著增加。当前无人机遥感融合算法研究主要集中在光学图像与雷达,光学图像与红外等双模态间融合处理,尚缺可见光、红外、雷达和无线频谱等多模态的融合处理研究。为提高对低可观测目标及其环境的检测、识别、跟踪和行为推理性能,亟需从跨模态融合向多模态认知跃升,开展多模态多层级融合认知处理前沿研究。

1.4.3 生成模型的研究现状

图像视频压缩分为传统方法和基于深度学习的方法。传统方法基于数学变换、量化和编码的原理减少不重要信息。基于深度学习的方法主要基于神经网络的建模和优化能力,重建高质量的图像视频[43],在解码端生成模型重建图像。对于视频任务,深度学习方法则通过引入视频生成模型或逐帧压缩的方式来实现[44]。文献[45]使用跨模态语言模型从图片中生成文本描述,最后重建出图像。文献[46]在语义保真上进行跨模态图像压缩,在编码端提取文本描述并压缩传输,解码端使图像生成模型ControlNet生成图片。文献[47]提出三级跨模态图像重建框架,以文本描述、结构草图与隐空间特征三者为传输内容,先从文本中解码出语义保真的图片,结合结构草图生成结构一致的图片和隐空间特征生成信号保真的图片并输出。文献[48]提出的视频生成框架CustomVideo支持以文本和语义分割图为输入,基于目标注意力控制策略生成质量较高的视频。Song等[49]在2018年较早提出了用于跨模态遥感图像生成的端对端卷积神经网络模型。Ma等[50]较早地提出了一种基于生成对抗网络(Generative adversarial network,GAN)的红外与可见光图像生成方法。Xiao等[51]通过对跨模态图像特征潜在空间的探索,提出了基于GAN的图像融合生成算法。2021年,Mildenhall等[52]提出神经辐射场(Neural radiance field,NeRF)的图像渲染生成方法。同年,Ho等[53]提出了去噪扩散概率模型(Denoising diffusion probabilistic models,DDPM)。2021年,Nichol等[54]提出了基于DDPM的文本到图像跨模态生成模型GLIDE(Guided language to image diffusion for generation and editing)。2022年,Ramesh等[55]提出的DALL-E2将多模态生成过程拆分为从文本到CLIP图像隐向量和从CLIP图像隐向量到图像这两个过程,使用两个DDPM分别训练,取得了比GLIDE更加精细而高质量的生成结果。2023年,Blattmann

等[56]提出了基于扩散模型的典型框架AYL。最早的无参考图像/视频质量评估算法(Blind image/video quality assessment, BIQA/BVQA)是基于对少量图像或帧级特征的测量,设计用于分析和量化单个失真类型。这些已经基本被通用BIQA/BVQA模型所取代。文献[57]实现了一个类似的集成模型,该模型使用组合用户生成内容(User generated content, UGC)超集上的监督特征选择程序融合了从几个性能最好的BVQA模型中选择的特征。Kang等[58]提出了一个利用卷积神经网络准确预测图像感知质量的通用BIQA框架。PaQ-2-PiQ[59]是一系列深度质量预测模型,包括P2P基线模型、P2P RoI模型和P2P反馈模型,其中预训练的ResNet-18作为骨干,RoIPool层聚合图片和补丁质量信息。将Transformers应用于VQA问题的第1项工作是StarVQA[60],它由预处理、L编码块和矢量化回归损失模块组成。You等[61]提出了一种感知分层网络来预测帧质量评级和特征向量的分布,以及一种长短期卷积变换器(Long short-term convolutional transformer,LSCT)架构来融合剪辑特征并预测整体视频质量。

无人机多模态超宽谱认知仪技术对比如表1所示。当前无人机遥感态势生成研究主要集中在图像视频压缩、生成与质量评估,多模态、语义化研究还比较少。受限于无人机遥感通信容量的限制,亟需从图像视频重建生成向多模态态势生成跃升,开展可见光、红外、激光雷达、SAR雷达、无线频谱等多模态数据压缩、态势生成、质量评估一体化前沿研究与相关软件开发。

表1 无人机多模态超宽谱认知仪技术现状对比分析

Table 1 Comparative analysis of the state of the art in UAV multi-modal ultra-wide spectrum cognitive sensors

技术方向	代表性进展	技术瓶颈	本文突破方向
光学/红外载荷	•复眼相机阵列(2022) •红外-可见光CNN融合(2020)	•夜间性能差 •异构分辨率配准误差大	•复眼多目相机+红外联合标定 •端到端参数生成网络优化图像
SAR小型化	•轻量化X/C波段SAR(2013) •实时L波段处理(2018)	•W波段功放不成熟 •振动补偿精度不足(>1 mm)	•W波段固态功率合成 •亚毫米级振动补偿技术
频谱监测	•DPD直接定位(2019) •多通道测向系统(2023)	•同频信号分离难 •动态目标定位延迟高	•超分辨率同频多信号测向 •移动侦测数据动态融合定位
多模态融合	•SAR-光学稀疏融合(2022) •红外-可见光GAN生成(2021)	• 仅限于双模态 • 实时性差(>500 ms)	•四模态时空配准 •轻量化多层次融合架构
态势生成	•神经辐射场(NeRF,2021) •扩散模型生成(DALL-E2, 2022)	•未适配无人机通信约束 •缺乏语义保真评估	•生成式编码(400倍压缩) •知识图谱驱动重建+无参考质量评估

2 无人机多模态超宽谱认知仪系统架构

认知仪研制内容主要包括无人机多模态超宽谱认知仪平台、低功耗轻重量小体积多模态载荷模块、多模态多层次融合认知机载计算模块以及多模态态势生成地面站4个方面,如图1所示。

无人机多模态超宽谱认知仪软件系统由机载软件子系统和地面站软件子系统组成。机载软件包含各模态探测载荷、控制器、通信设备、机载计算中心等不同的操作系统(国产华为Atlas 200I DK计算模块上的openEuler系统、GIGABYTE QBPi-1115G4MS计算模块上搭载的Windows系统、英伟达Jetson NX上搭载的Ubuntu系统)、分布式异步消息中间件Mqtt、数据库中间件JDBC、感知数据处理与存储、多模态融合处理与存储、多模态数据压缩处理与存储、航迹控制处理等。地面站软件子系统由地面服务端软件及地面显控软件组成。地面服务端软件包含地面大数据量的多模态融合认知计算模块、组网通信协调模块、地面多模态数据存储管理模块,地面显控软件包含地面多模态传感器一体化指控调度模块、无人机系统任务规划与调度模块、系统状态管理模块和综合态势呈现模块。

图1 无人机多模态超宽谱认知仪

Fig.1 UAV-borne multi-modal ultra-wide spectrum cognitive instrument

无人机多模态超宽谱认知软件系统基于B/S架构设计,用服务端部署的方式通过划分功能来满足用户的需求,如图2所示。软件架构主要分为4层。

数据层:负责数据汇聚、存储和对上层的调用接口,包括多模态感知数据、机载数据缓存、轻量级数据中台、清洗后数据和融合后数据等,将各个传感器的数据进行整合和处理,提高数据质量和一致性,以便于进行更全面、准确的分析计算。在数据汇集的过程中,对多模态感知数据进行预处理,以提高数据的质量和可用性。多模态数据以基础关系型数据库或数据文件的形式分类分级存储,由数据层提供访问接口供多模态融合算法提取计算。数据层中间件包括数据库中间件JDBC用于接收机载多模态融合处理数据存储到地面数据库、地面端异步消息中间件用于整套软件系统的消息数据异步传输。

算法层:负责软件系统算法的调用接口管理、运行资源管理分配,包括机载多模态认知算法、多模态融合增强认知算法、无人机自主规划算法、智能通信频点决策算法、机载认知网络切换算法、地面多模态认知算法、地面多模态融合增强算法、伪装目标探测识别算法、多目标稳定跟踪算法、同频目标定位跟踪算法、多模态数据清洗算法以及综合态势认知生成算法等,提供对上层的算法接入和调度接口。具备扩展能力,能够接入符合接口要求的多级多类认知计算算法。

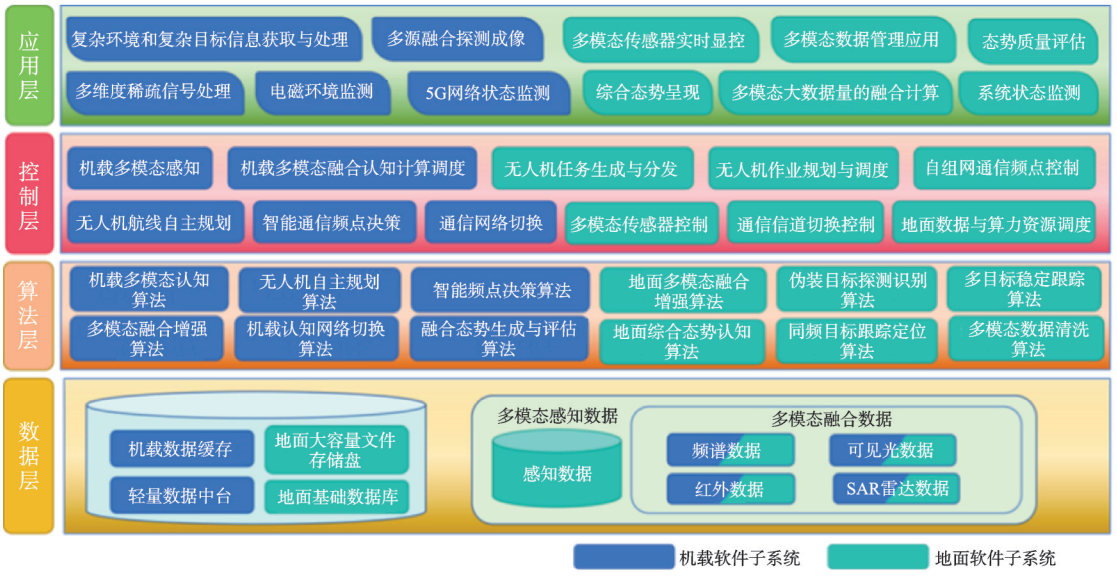

图2 无人机多模态超宽谱认知软件系统

Fig.2 Multi-modal ultra-wide spectrum cognitive software system for UAVs

控制层:包括对多模态传感器的指令控制、无人机作业任务的控制调度、对系统空地通信信道切换的控制、多模态融合认知算法的计算调度等,根据任务需求和环境条件,动态调度各种认知算法,并根据计算资源的可用性合理分配计算任务,确保计算任务的高效执行。软件控制层中间件包括分布式任务中间件,用于各模态探测子系统的协同任务调度,以及任务注册中心,用于管理各模态探测子系统的任务管理以及对应用层提供任务调度接口服务。

应用层:负责系统多模态传感器实时显控、通信网络管理、用户的指控信令生成、大数据量的多模态融合认知计算、多模态数据管理应用、电磁环境认知、5G网络状态监测、综合态势呈现、态势质量评估、复杂环境和复杂目标信息获取与处理、多维度稀疏信号处理以及多源融合探测成像等一系列应用。将各种传感器数据和分析结果以直观的方式实时展示出来,帮助全面了解无人机系统的运行状态和周围环境,通过将多种感知数据进行融合增强,提高对目标的检测、识别和跟踪能力。

3 关键技术研究进展

3.1 长航时混合能源无人机平台技术

围绕应急救援、巡检任务、军事侦察等的迫切需求,本文研究可在复杂气象和地形条件下起降、长时间续航飞行的垂直起降无人机,突破长航时混合能源无人机总体构型、高效布局设计、新能源动力匹配与驱动技术等关键技术研究,形成长航时无人机总体技术。

3.1.1 垂直起降无人机总体构型设计

垂直起降(Vertical take-off and landing,VTOL)技术的应用显著提升了长航时无人机在复杂环境下的作业能力,使其摆脱了对传统起降场地的依赖。然而,这种技术改进也带来了新的设计挑战,特别是在采用混合能源系统的情况下,系统复杂度显著增加。VTOL模块不仅影响无人机的巡航性能,还对飞行速度、续航时间等关键指标提出了更高要求,这使得垂直起降固定翼无人机的总体设计面临严峻挑战。混合能源无人机的设计过程存在明显的能量-重量耦合效应,这要求从能量平衡、重量平衡和功重匹配等多个维度进行综合考量。与传统能源无人机相比,混合动力系统的重量会随飞行条件动态

变化,因此需要建立精确的系统模型。首先需要构建包含燃油推进系统和锂电池系统在内的混合能源重量模型,并基于能量守恒原理建立系统内部的功率传递模型。通过对不同飞行状态下动力需求的差异性分析,可以建立完整的飞行剖面功率模型,实现混电系统内部功率的精确分配。

在重量/能量耦合分析的基础上,研究建立了包含结构、动力、燃油、载荷和机载设备等在内的全系统重量预测模型和功率预测模型,通过各分系统重量模型的协同优化,在满足功率平衡和重量平衡约束条件的前提下,最终实现了混合能源无人机总体参数的优化设计。结合总体参数和飞行性能需求对超长航时无人机的外形参数,如展弦比、翼展、机身容积等进行分析,并结合燃油需求和混合能源动力需求进行初步总体布置,对比多种布局方案,完成混合能源无人机总体构型方案。此无人机平台拟采用大展弦比机翼、翼身融合的飞翼布局形式,并添加翼尖小翼。机身头部为任务载荷舱,机身中部为燃油及动力装置舱,机身尾部螺旋桨为主要推进装置,其基本构型为翼展8 m,展弦比22。原型机设计图见图3。

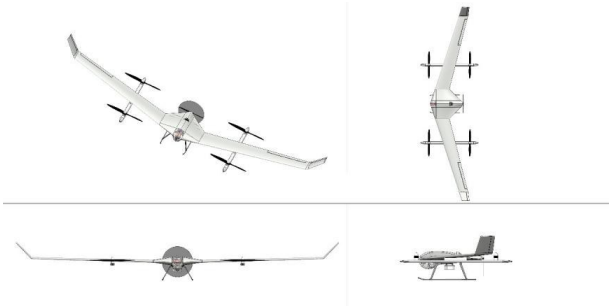

图3 无人机平台外形图

Fig.3 UAV platform profile

3.1.2 混合能源动力系统优化

混合能源动力系统的关键在于如何优化能量管理和动力匹配,首先保证动力系统匹配优化,可巡航速度要大于失速速度,保证飞机巡航时的安全性。其次,还需要确定功率要求,对于往复活塞式航空发动机,起飞阶段的功率是恒定的,在忽略滚动阻力的情况下,采用平均推力和阻力进行计算,则起飞距离 d_{tk} 可以表示为[62]

$$d_{\mathrm{tk}}=\frac{W_g V_{\mathrm{LO}}^2}{2g(T_{\mathrm{avg}}-D_{\mathrm{avg}})} \tag{1}$$

式中:W_g 为飞机重量;g 为重力加速度;爬升速度 V_{LO} 应不小于1.1倍的失速速度;平均推力 T_{avg} 和阻力 D_{avg} 应该等于0.7倍 V_{LO} 情况下的推力和阻力。对于混合动力无人机的参数匹配中,待优化匹配的参数有:发动机的最大功率 $P_{\mathrm{ICE,max}}$,电机最大功率 $P_{\mathrm{EM,max}}$,所需燃油质量 m_{fuel},飞行时间 t_{f},功率分配系数 σ_{p}。优化目标函数为

$$\min J_1 = m_{\mathrm{fuel}} \tag{2}$$

$$\min J_2 = -t_{\mathrm{f}} \tag{3}$$

根据电池荷电状态(State of charge,SOC)与电池关系[63]即可得到系统预测模型为

$$\dot{\mathrm{SOC}}=-\frac{I}{Q_{\mathrm{b}}}=\frac{U_{\mathrm{b}}-\sqrt{U_{\mathrm{b}}^2-4R_{\mathrm{b}}P_{\mathrm{b}}}}{2R_{\mathrm{b}}} \tag{4}$$

式中:Q_{b} 为电池的额定容量,电池开路电压 U_{b} 和电池内阻 R_{b} 为电池SOC的函数,动力电池输出功率 P_{b} 可视为发动机-发电机组输出功率与系统需求功率之差,也即

$$P_{\mathrm{b}}=\omega_{\mathrm{e}}\cdot T_{\mathrm{gen}}-P_{\mathrm{need}} \tag{5}$$

在系统状态转移方程中,系统状态量 $\boldsymbol{x}$ 为电池SOC,系统控制量 $\boldsymbol{u}$ 为发动机转速 ω_{e} 和发电机转矩 T_{gen},系统干扰量 $\boldsymbol{v}$ 为需求功率 P_{need},系统输出量 $\boldsymbol{y}$ 为电池SOC和发动机燃油消耗率 m_{f}。非线性预测模型的离散形式可以描述为

$$\boldsymbol{x}(k+1)=f(\boldsymbol{x}(k),\boldsymbol{u}(k),\boldsymbol{v}(k))\quad \boldsymbol{y}(k+1)=g(\boldsymbol{x}(k),\boldsymbol{u}(k),\boldsymbol{v}(k)) \tag{6}$$

式中:$\boldsymbol{x}=[\mathrm{SOC}]$, $\boldsymbol{u}=[\omega_{\mathrm{e}}\ T_{\mathrm{gen}}]^{\mathrm{T}}$, $\boldsymbol{v}=[P_{\mathrm{need}}]^{\mathrm{T}}$, $\boldsymbol{y}=[\mathrm{SOC}\ m_{\mathrm{f}}]^{\mathrm{T}}$,等效燃油消耗因子 S_{e} 计算公式为[64]

$$S_{\mathrm{e}}=2-2\cdot(1-0.7z)\cdot z^3 \tag{7}$$

$$z=\frac{\mathrm{SOC}-\dfrac{\mathrm{SOC_H}+\mathrm{SOC_L}}{2}}{\mathrm{SOC_H}-\mathrm{SOC_L}} \tag{8}$$

式中：$\mathrm{SOC_H}$为电池SOC最大限值，$\mathrm{SOC_L}$为电池SOC的最小限值。

综合发动机油耗和SOC变动量限制两个因素，预测时域之内的优化目标成本函数可以表述为

$$J=\int_{t_0}^{t_h}(m_f+\omega_s(\mathrm{SOC_t}-\mathrm{SOC_r})^2)\mathrm{d}t+\omega_h(\mathrm{SOC_{th}}-\mathrm{SOC_r}) \tag{9}$$

式中：t_0为预测时域起始时刻，t_h为预测时域终止时刻，$\mathrm{SOC_{th}}$为终止时刻的电池SOC值，$\mathrm{SOC_r}$为电池参考SOC基准，$\omega_s(\mathrm{SOC_t}-\mathrm{SOC_r})^2$表示对电池SOC波动量的惩罚，$\omega_s$为电池SOC波动量的权重，$\omega_s(\mathrm{SOC_t}-\mathrm{SOC_r})$表示对电池预测时域终止时刻电池SOC变动量的惩罚，ω_h为其惩罚权重。

基于以上模型完成仿真与设计，从而实现整机飞行过程中动态能量管理，如图4所示。以150 kg级混合能源动力无人机为研究对象，形成无人机混合能源和分布式推进新型高效混合动力系统总体技术方案。其中：混合动力系统最大起飞质量不小于150 kg，载荷能力不低于35 kg，速度100 km/h，续航时间不低于10 h。

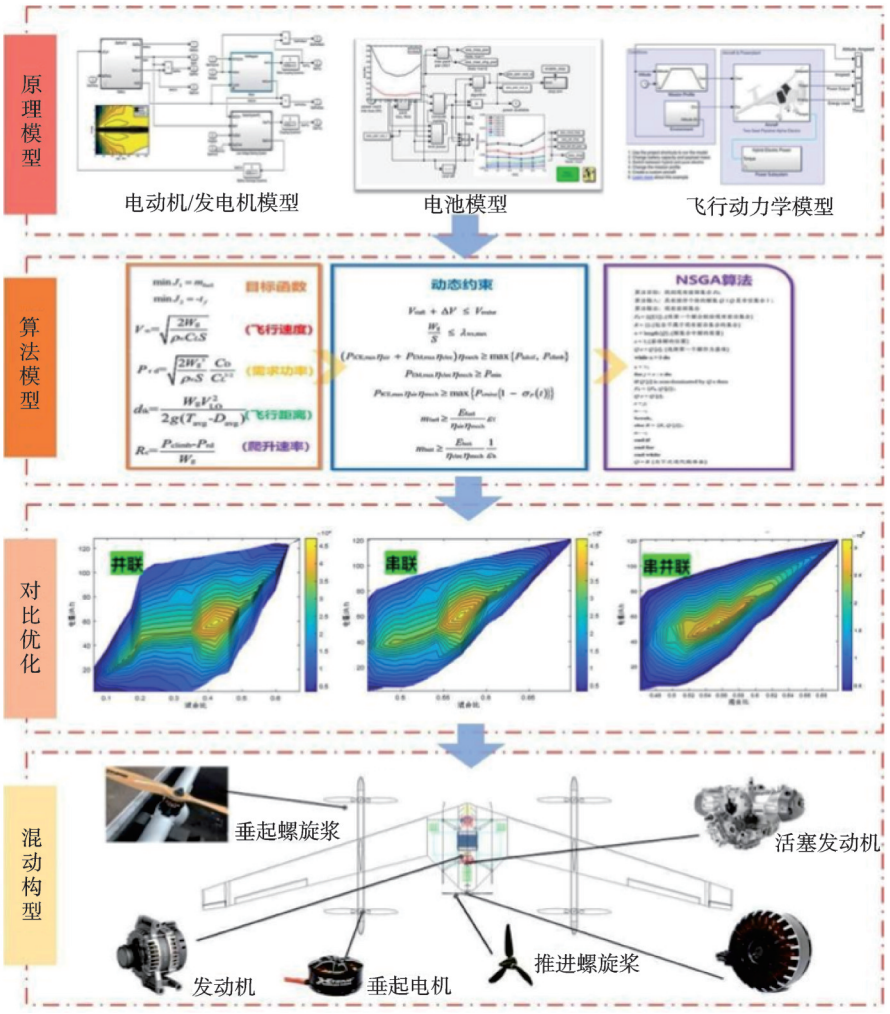

图4 活塞发动机发电和锂电池的混合动力系统构型

Fig.4 Configuration of a hybrid power system with a piston engine-generator and lithium batteries

通过表2对比能够看出,与国内外的垂直起降无人机平台相比,本文方案无人机遥感平台通过采用大展弦比机翼+旋翼的高效布局形式,以及混合动力系统优化匹配设计,可以有效提升平台的载荷能力,实现多种载荷的选装,同时提升航时达到10 h以上。

表2 国内外垂直起降无人机平台性能与本文方案对比

Table 2 Performance comparison of domestic/international VTOL UAV platforms and the proposed solution

型号	CW-100	HC-141	白鹤	TU-150	本文方案
起飞质量/kg	110	160	162	140	≤150
动力形式	油动	油动	油动	混合动力	混合动力
飞行速度/(km•h^{-1})	100	108	120	120	100
最大载荷/kg	≤25	30	30	25	≥35
续航时间/h	8	10	5	8	≥10

3.2 多模态探测载荷技术

多模态探测载荷技术是提升无人机遥感性能的核心之一。不同类型的传感器能够从不同维度获取目标和环境的信息,通过综合利用这些信息,可以得到更加准确和全面的遥感数据。为提高无人机的探测能力,需设计轻量化的多模态传感器,如可见光相机、红外热像仪和SAR雷达等。这些载荷的集成设计不仅要考虑重量、体积等因素,还要在电磁兼容性、互不干扰等方面进行优化,确保各类传感器能够协同工作,实现多模态数据的高效采集和处理。

3.2.1 复眼多目相机技术

目标识别跟踪任务通常需要高分辨率图像作为输入,大场景、大视野下的精准小目标检测需要2 000万以上的像素,而传统相机的像素数量有限,无法满足对小目标准确识别和跟踪的需求。因此使用大视场高分辨率复眼多目相机的关键技术,为低可观测目标的检测和追踪提供了有效解决方案。

复眼多目相机技术通过配置多个相机,能够获得更广阔的视场和更高的图像分辨率。在无人机遥感应用中,复眼相机常用于地面物体的检测、跟踪和识别等任务。其优势在于通过对不同角度拍摄的图像进行融合,能够克服单一视角下的盲区,提供更为清晰和立体的图像。基于多个可见光传感器的相机系统可以充分发挥复眼光学相机的大场景、高分辨率图像优势,结合红外传感器在夜视方面的成熟优势,能够实现低成本、高分辨率、高准确性、全天候的“全场景”成像与感知。复眼计算成像系统为了获取尽可能大的融合分辨率和视场角,一方面各子镜头的分视场之间需要尽可能减少重叠,另一方面增加子镜头数目以获取更多的视野,最终实现复眼相机的大视场覆盖,主要包含可见光拼接融合、复眼光学相机校准与标定、可扩展视频拼接专用处理器3方面的技术。

不同传感器在数据分辨率、帧率、成像内容上差异巨大,需要充分发挥多传感器的物理器件优势,实现在异质分辨率、视场角、空间尺度上对复眼相机视场内物体的联合标定,最终实现在超大视野范围下,结合图像的高分辨率特性和多源传感器在距离探测、运动捕捉和多光谱成像的特性。在可见光与红外融合方面,通过融合红外和可见光传感器数据,获得对时间、空间、角度的特性高维调制,利用联合标定设计的思路,实现多模态传感器对场景同一目标的多维度描述,突破了传统单一传感器的物理限制,构建了合适的内参和外参模型,把所有子镜头相位点及传感器物理特性都合理利用起来,建立统一的联合参数给予多源复眼光学相机建模。同时利用端对端的参数生成网络生成图像处理核参数,避免了不同场景下非最优解以及重新训练的问题,可以使得各子镜头根据自己所拍摄的场景强化图像高频和低频分量,使图像中物体轮廓清晰、细节明显,从而保证所拍摄图像的一致性和优化性,如图5所示。

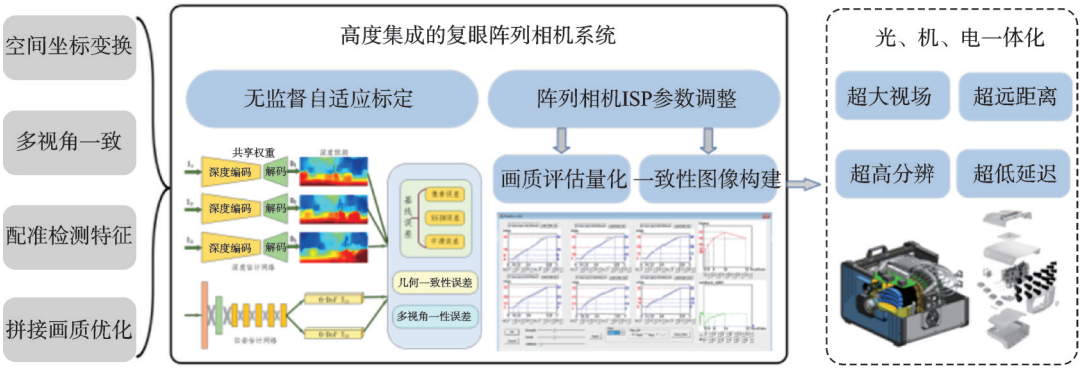

图 5 复眼光学相机校准和图像优化

Fig.5 Camera array calibration and image optimization

3.2.2 W 波段轻小型 SAR 雷达

SAR 雷达是一种通过主动发射电磁波来探测目标的技术，具备全天候、全天时的优势。W 波段雷达作为一种高频率的 SAR 雷达，其具有更高的分辨率和更强的穿透能力，特别适用于地表和植被下的目标探测。为了使其适用于无人机平台，需对雷达的重量和功耗进行优化，确保其能够在有限的空间内发挥最大效能。当前技术条件下，W 波段合成孔径雷达技术还面临多方面挑战。在总体设计方面，如何在小型化高集成度要求下降低重量、减小功耗、优化散热是一大难题；在器件方面，现阶段 W 波段固态功放等关键元器件不成熟，输出功率等核心指标不满足探测要求；在天线设计方面，平面天线合成馈电网络路径长损耗大，导致天线效率大幅降低。因此，需针对以上挑战，开展 W 波段雷达一体化集成技术、W 波段固态功率合成技术、W 波段高增益平面天线技术研究。W 波段无人机 SAR 原理样机主要由雷达吊舱和后端设备等组成，如图 6 所示。

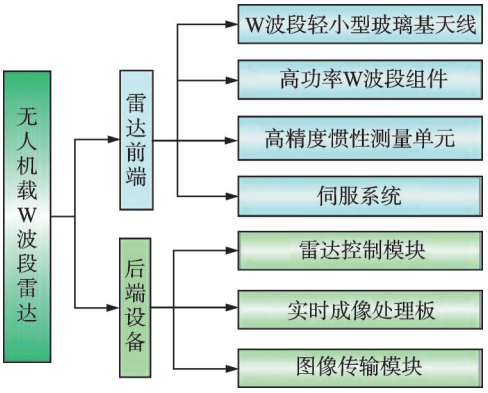

图 6 雷达系统设备组成图

Fig.6 Radar system block diagram

3.2.3 移动侦测数据融合定位

无线频谱种类多、传播复杂、变化快，给无线频谱测向带来极大挑战。同时，机载平台有限载荷和功耗给机载参数估计、辐射源定位等算法时效性提出较高要求。项目通过研制机载多通道无线频谱侦测模块，突破超分辨率同频多信号快速测向、移动侦测数据动态融合辐射源定位等技术难题，实现复杂无线频谱的快速准确测向。其中，信号处理单元实物图见图 7，天线阵列概念图见图 8。同时，发挥无人机平台的灵活性，结合动态优化的路径，通过多点观测获得最优观测性能。在目标位置和传播环境未知时，研究基于费雪信息矩阵、具身智能、强化学习等模型、数据驱动路径动态优化方法，通过机载动态路径规划实现对目标位置的精确认知。移动侦测数据融合定位技术能够结合多种传感器的数据，如红外、雷达、无线频谱等，进行综合分析，从而提高目标的定位精度。在实际应用中，利用此技术可以实现对高速、移动目标的实时监测和定位，对于执行救援任务、敌情侦察等具有重要意义。

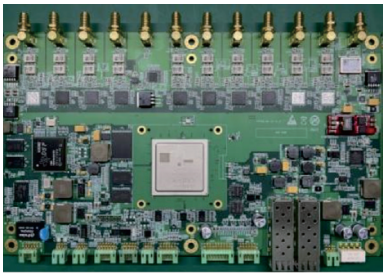

图 7 信号处理单元

Fig.7 Signal processing unit

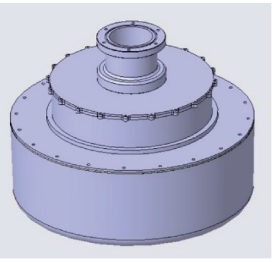

图 8 天线阵列

Fig.8 Antenna array

3.3 多模态融合认知技术

多模态融合认知技术是无人机多模态超宽谱认知仪的核心技术之一,其目的是将不同传感器的数据进行有效融合,提高目标识别和环境感知能力。

针对无人机计算和存储资源受限问题,设计适配多模态超宽谱数据融合处理的机载计算中心单元以及多模态载荷计算单元的算力分配技术,提高协同计算和资源分配效率。针对低可观测目标、时敏目标、辐射目标等多种类型目标,在可见光、红外、SAR雷达、无线频谱等多模态观测下呈现的异构异质性、互补性与冗余性等特点,以及受干扰与天气等条件影响,研究开发新算法提升多模态融合成像精度与目标识别准确率。针对SAR图像目标微弱等情况,实现了非下采样轮廓波变换的SAR、红外、多光谱、频谱图像融合算法:首先对预处理的SAR、红外、多光谱、无线频谱图像进行多尺度多方向分解,得到低频和带通方向图像;然后,由于低频和带通方向图像具有不同的信息特征,选择不同融合算法对其进行融合,得到融合后的低频和带通方向图像;最后对融合后的低频和带通方向图像进行非下采样轮廓波逆变换得到融合图像。

3.3.1 多模态时空配准算法

多模态时空配准是指将不同时间、不同传感器采集的图像或数据对齐,以便进行后续的分析和处理。由于SAR图像[65]、红外图像、多光谱图像以及无线频谱图像来自不同的传感器设备且时间-空间呈现不一致性,需要对多模态图像进行交叉映射,即多模态图像配准。本研究提出基于特征的配准方法,所提取的特征应该是高度清晰、稳定且可在图像之间重复的,即图像中的点、轮廓、边缘或区域特征;由于多模态图像强度和纹理的差异较大,所提取特征通常具有低重复性,这将明显降低匹配性能。因此,本研究拟构建基于局部不变特征的多模态图像配准模型,提取能克服非线性强度差异的特征描述符定向梯度通道特征,实现多模态图像配准,其中局部不变特征[66]主要指的是从可见光、红外、SAR 等异构模态图像中提取的、能够抵抗成像机理与条件差异的稳健特征点及其描述符。由于不同传感器的成像原理不同,导致同一场景在不同模态图像中呈现出截然不同的辐射特性、几何形态与噪声结构,使得直接进行像素匹配或使用传统特征极易失效。因此,构建基于局部不变特征的多模态图像配准模型的核心在于提取如“定向梯度通道特征”这类描述符,该特征对非线性强度变化保持稳定,不依赖于跨模态间差异巨大的绝对灰度值,同时对尺度、旋转等几何微变具有一定鲁棒性,并立足于局部邻域信息以应对局部形变。通过检测并匹配这些跨越模态仍能保持可重复性的关键点,算法能够建立不同模态图像间可靠的特征对应关系,从而解算出将它们对齐到统一时空坐标系所需的变换参数,为后续的多层次融合认知与态势生成奠定基础。多模态时空对齐原理如图9所示。

图10为本文方案在角度误差为0.5°时的辐射源位置仿真估计散点图。由图10可以看出,本方案可实现多干扰源被动定位,且对多个干扰源同时定位时无需额外的步骤去除模糊解。

以上方案进行了初步融合实验,辐射源目标为空中无人机,释放图传信号。实验中,给定相机的观

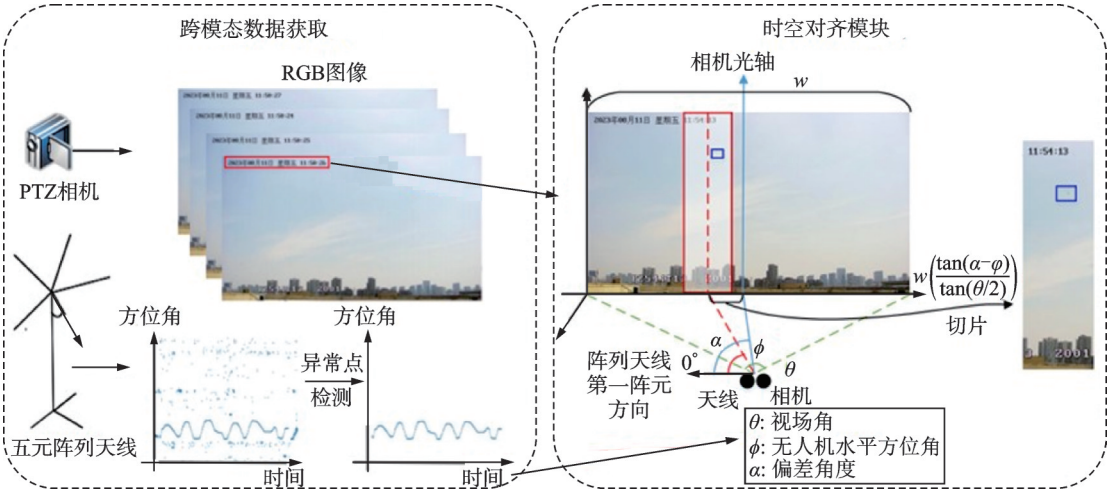

图 9 多模态时空对齐

Fig.9 Multi-modal spatio-temporal alignment

察位置，能够在视野范围内观察到真实场景角度大小。随着焦距倍数增长，监控摄像头空间分辨率不断提升，同时视场角逐渐减小。当监控 4 倍焦距时，空间分辨率与多光谱相机相当，此时球机水平视场角为 15.6°，监控 1 倍焦距时，水平视场角度为 57.6°。图 11(a)为输入图像；图 11(b)的鸟类和直升机区域显示出更高的显著度，这表明其他空中物体严重干扰了无人机小目标定位的性能；图 11(c)中无人机区域的特征显著性较高，而背景的显著性较低，这证明本文融合方法能更好地捕捉无人机目标在这些特定区域的特征。利用天线作为补充信息，将焦点限制在特定区域，可以过滤掉无关的小目标干扰，从而提高无人机定位的精确度和效率。

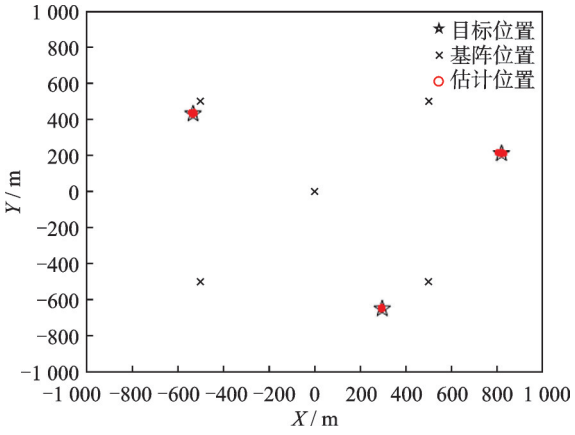

图 10 基于多测向信息聚类定位散点图

Fig.10 Localization scatter plot based on multi-DOA information clustering

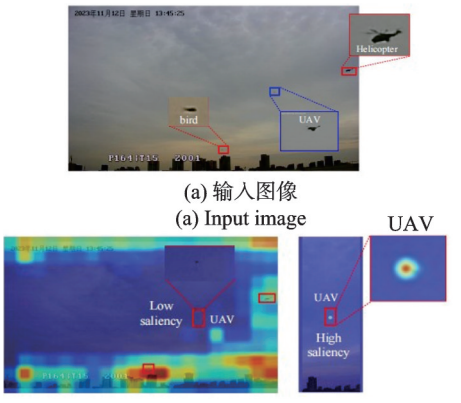

图 11 多模态融合图

Fig.11 Multi-modal fusion diagram

3.3.2 多层次融合架构

多层次融合架构是指在数据层、特征层和语义层等多个层次上进行融合，以实现对多模态数据的深度分析。数据层融合主要关注传感器数据的匹配与对齐，特征层融合则提取数据中的关键特征进行融合，语义层融合则侧重于对目标和环境进行高层次的语义理解。这一架构能够有效提升多模态数据融合

的精度和稳定性。数据层融合需要将多模态数据映射到统一的信息空间,但由于杂波、干扰、噪声等破坏了数据的真实性,给后续的特征级融合和决策级融合引入误差;特征层融合先对多模态数据分别提取特征再进行融合,越过原始异构数据对齐,对传感器对准没有像素级要求,但存在信息损失与误差传播;语义层融合把多模态数据的语义信息进行融合处理,存在语义不一致的融合难题。多模态融合旨在将多模态信息整合成稳定的多模态表征,以获得全面特征,提升模型鲁棒性,并且保证模型在某些模态缺失时仍能有效工作。按融合层次,多模态融合可分为数据、特征、语义、混合等类别,分别对原始数据、抽象特征、决策语义进行融合或混叠融合。面向多模态多层次融合,无人机多模态超宽频认知仪的机载各模态计算单元与机载多模态计算中心的高效协同亦是仪器研制中必须解决的关键科学难题。

3.3.3 轻量化深度学习模型设计

随着深度学习在遥感领域的应用,如何在无人机平台的有限计算资源下实现高效的深度学习模型成为了一个关键问题。为此,研究者们提出了轻量化深度学习模型设计方法,通过剪枝、量化等技术,减少模型的计算量和存储需求,从而保证实时性和高效性。

由于可见光、红外、SAR和无线频谱监测是不同传感器或技术,它们各自具有独特的优势和劣势。可见光图像具有丰富的色彩和纹理信息,在日常场景和自然光照条件下表现良好。然而,可见光图像受光照和环境因素的影响较大,限制了其在夜间和恶劣天气条件下的应用。相比之下,红外图像对热源和低照度环境中的目标具有较强的检测能力,能够在低光、夜间以及复杂天气条件下发挥更好的性能。SAR图像具有穿透能力,可以穿透障碍物进行观测,但其分辨率较低且数据处理复杂。无线频谱则通过无线信号的传播和反射进行非接触式检测,具有高穿透能力,但受到信号干扰的影响。无线频谱作为非成像感知模态,通过接收和分析目标的电磁辐射信号实现探测、定位与识别,虽在数据形态上与可见光、红外、SAR等成像模态差异显著,但其具备全天候全天时工作、对电磁辐射目标准确响应,以及对植被和地形穿透性较好等独特优势,使其在低可观测目标探测、复杂电磁环境感知中具有不可替代的互补价值。融合的可行性首先基于信息互补性:光学影像提供高分辨空间纹理但受限于光照天气,SAR可实现穿透成像但对弱散射目标敏感,而无线频谱专长于探测无视觉特征的电磁辐射目标,如通信设备与干扰源,各模态在探测机制与环境适应性上形成天然互补;其次,无人机平台通过“同机同时同视场”的多模态载荷一体化集成设计,为实现时空同步的数据采集与后续对齐提供了物理基础;此外,融合可在多层次展开,不仅包括将无线频谱定位结果与光学图像进行坐标对齐的数据层融合,更涉及特征层与决策层的深度融合,例如提取信号的调制类型、到达方向等特征并与图像CNN特征联合输入多模态检测网络,或通过知识图谱关联目标的电磁属性与视觉外观进行综合推理。具体技术路径上,可以通过轻小型高精度无线频谱监测接收机实现多通道超分辨率测向,系统框图如图12所示,为融合提供高精度输入。

对于该多通道机载接收机,假设监测通道数为32,模拟中频信号采样率为102.4 MHz,而音频解调所需的最大采样率为3.2 MHz,则数字下变频器多级抽取模块第1级抽取可使用级联积分梳状滤波器组进行32倍抽取,将其采样率降低为3.2 MHz。系统时钟为102.4 MHz,设2倍抽取有限脉冲响应半带滤波器组包括12个2倍抽取有限脉冲响应半带滤波器,则可完成32个通道复用。天线阵列方面采用均匀圆阵,且阵元均为各向同性阵元,主要用于对无线频谱信号的接收,由测向天线阵和测向开关组成。算法方面,相比常规波束形成(Conventional beamforming, CBF)方法,多信号分类(Multiple signal classification, MUSIC)方法可获得超分辨率空间方向信息。图13为对3个同频信号的测向结果,可以发现MUSIC方法能进行有效测向,且测向精度优于2°,可实现对复杂混叠多信号的超分辨率测向,结合基于可见光、红外、SAR、无线频谱的多模态融合感知方法预计可以达到最低检测精度不低于90%、漏检率不超过10%。因此尽管无线频谱数据具有异构性,但通过时空配准、特征融合与决策协同,可有效融入

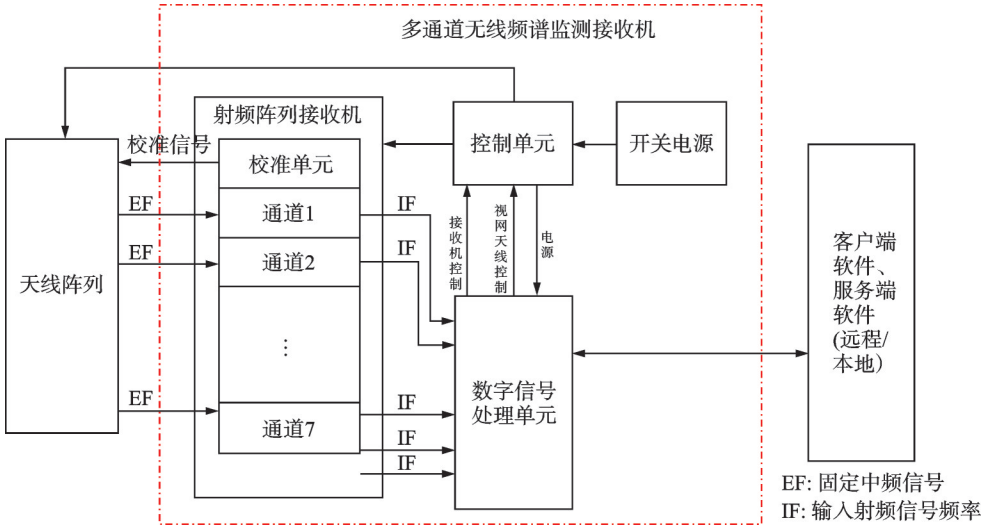

图12 机载多通道无线频谱监测接收机系统框图

Fig.12 System block diagram of the airborne multi-channel wireless spectrum monitoring receiver

多模态认知框架。文献[67]给出了一个成熟的无线频谱多模态感知方案,证明了无线频谱多模态融合的可行性。因此,联合可见光、红外图像、SAR雷达和无线频谱,通过这4种模态信息融合,可以提高目标检测的鲁棒性,使系统更适用于各种环境和光照条件。基于以上问题,可以设计一种可见光、红外、SAR和无线频谱的深度学习融合网络,其主要可由特征提取骨干网络、融合单元和目标检测头3部分组成。(1)特征骨干网络:由多个层级的池化层和卷积层所组成,从而逐渐减小输入图像的空间维度,并增加通道的数量。为了更好地提取主模态和辅模态的特征,分别采用2个独立的特征提取骨干网络,以获取两种模态中的有效信息。(2)融合单元:将主模态和辅模态的特征进行融合,从而更好地发挥其优势。其主要有特征加权融合和特征最大两种不同的形式,加权求和具有灵活性和综合性的优势,但需要考虑权重设定的敏感性。最大值简单明确,直观性强,但可能没有办法充分利用所有的数据。为了更好地融合其信息,本文将两个混合特征进行拼接。(3)目标检测头:可以将语义特征征与目标的类别和位置进行关联,从而实现对目标的检测和定位。利用目标检测头,即可以对之前得到的混合特征进行处理,从而得到最后的目标检测结果。目标检测网络设计图见图14。

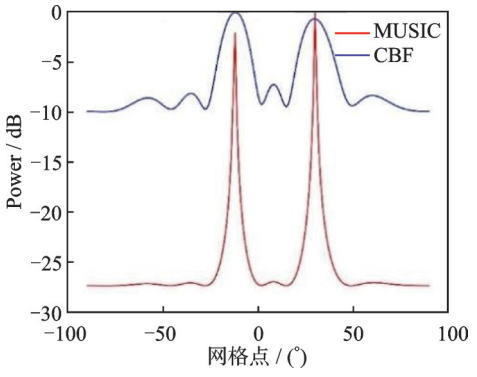

图13 超分辨率空间谱估计

Fig.13 Super-resolution spatial spectrum estimation

基于上述的可见光+红外+SAR+无线频谱融合框架,可对当前的通用目标检测网络进行改造以实现可见光与红外数据的数据融合与目标检测。随着目标检测的发展,它可以整合最先进的技术,从而实现轮次迭代。在YOLOv3达到峰值性能后,YOLO扩展板开始关注无锚点检测器、高级标签分配策略和端到端检测器,这是近年来的主要进展[68]。升级后,在COCO数据集上,与YOLOv3相比,它表现出了显著的性能,而单次多框检测器结合了几个不同分辨率的特征图,从而通过多尺度训练提高了模型的性能。该检测网络对于不同大小的物体有很好的检测效果,且只涉及单个网络,使得模型易于

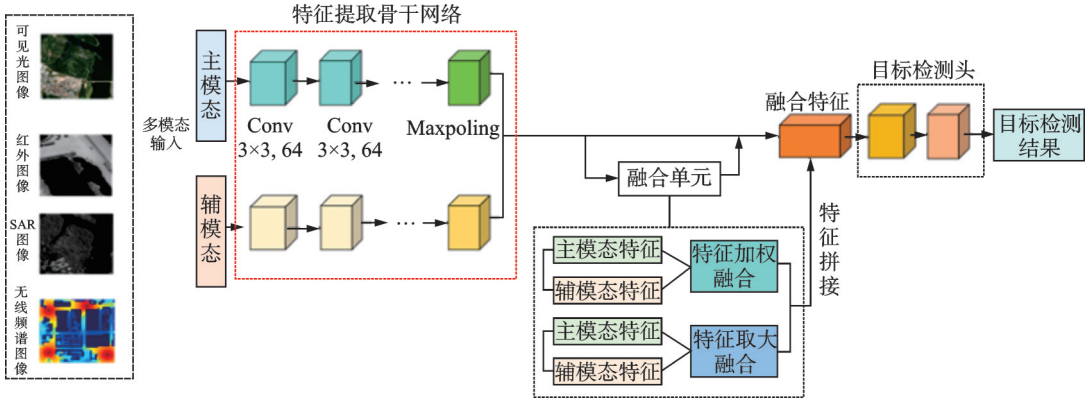

图14 基于可见光、红外、SAR、无线频谱的目标检测网络

Fig.14 Target detection network based on visible light, infrared, SAR, and wireless spectrum

训练。

目标检测算法通常使用平均精度均值(mean Average precision, mAP)来进行性能准确性的评价。mAP表示在数据集中检测到的所有目标的平均精度。计算mAP,首先需要计算精确率(Precision)和召回率(Recall),计算公式如下[69]

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}} \tag{10}$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}} \tag{11}$$

式中:TP(True positive)代表的是被模型正确检测出的正样本数量,FP(False positive)代表的是被模型错误地检测为目标的负样本数量,FN(False negative)代表的是真实存在但未被模型检测出的目标数量。在目标检测网络中,一般认为当预测框和真实框的交并比(Intersection over union,IoU)>50%时,该预测框为正样本。针对目标检测问题,考虑无人机载SAR全天时全天候工作的优势,以SAR模态为主,在特征层和语义层融合SAR图像以及可见光、红外图像,实现可见光、红外辅助融合的SAR动目标检测功能;针对目标识别问题,利用可见光模态高分辨成像的特点,以可见光模态为主,在特征层、语义层构建可见光、红外融合模型,实现对图像目标的精准识别;针对辐射目标定位问题,利用目标的辐射特性,以无线频谱感知模态为主,设计无线频谱与可见光在语义层融合的定位算法,实现对辐射目标定位性能的提升;针对目标跟踪问题,利用可见光成像效率高以及红外模态夜间成像优势,在数据层、特征层、语义层融合可见光与红外图像,实现对图像目标的有效跟踪。传统的单模态检测方法在理想条件下精度通常在80%左右[70]。近年来,面向无人机的轻量化检测模型将单模态精度提升至接近90%的水平[71]。而通过引入多模态融合,特别是借鉴基于Transformer的先进融合架构[72],可在复杂场景下实现更鲁棒、更精准的感知。传统基于可见光、红外、SAR、无线频谱的单模态方法检测精度不低于80%、漏检率不超过20%,通过多模态融合实现多场景、多目标、全天候的检测性能,最低检测精度不低于90%、漏检率不超过10%[73]。

与同类型轻量化深度学习方法相比,本文提出的研究框架核心优势在于其并非单一的模型压缩或加速,而是一个面向无人机平台资源严格受限条件(载荷、算力、续航、通信)进行全方位软硬件协同设计与多层次优化的“系统级”轻量化框架,构建了覆盖“近传感器计算-单模态计算-多模态融合计算”三层异构计算架构,并实现了机载多模态数据的异步协同计算与动态资源调度,这使得可见光、红外、SAR与无线频谱等处理负载迥异的模态能在统一管理下高效协同,避免了简单堆砌单模态轻量化模型

带来的资源冲突与效率瓶颈。该框架还将轻量化从算法层延伸至感知-传输-计算闭环,创新性地提出了基于知识图谱与生成式模型的语义级态势压缩与重建方法,在通信带宽受限条件下,实现了比传统编码方法高压缩比的同时保持高层语义保真,解决了“轻量传输”与“丰富信息”之间的矛盾。

因此,本框架的优势在于其系统性、协同性与任务适应性,能够在确保无人机长航时、大范围作业的物理约束下,实现多模态融合认知从数据到决策的全流程高效运行,而不仅仅是单个深度学习模型的轻量化。以SAR图像为例,与国内外的无人机载单模态SAR图像动目标检测与跟踪性能相比,本算法中基于多模态融合的SAR图像动目标检测与跟踪模块,可以实现更低的目标最小可检测速度和同时跟踪更多目标的功能,有效提升了无人机载SAR图像的目标检测与跟踪性能,性能对比如表3所示。

表 3 SAR 图像动目标检测与跟踪性能指标先进性对比

Table 3 Performance comparison of moving target detection and tracking in SAR imagery

指标名称	现有SAR雷达单模态图像目标检测与跟踪性能	本项目中多模态融合SAR图像目标检测与跟踪性能
运动目标最小可检测速度/(m•s^{-1})	≤7	≤5
运动目标同时跟踪数	≥6	≥8

3.4 多模态态势生成技术

多模态态势生成技术旨在通过对多模态遥感数据的处理和融合,实时生成目标和环境的态势图,帮助决策人员进行实时决策。在无人机平台上实现超低码率且高质量的多模态态势生成模型具有深远的意义。首先,它极大地提高了无线传输效率,使得在带宽受限的情况下也能传输8 000像素高分辨率图像,这对于扩大无人机的作业范围和提高任务执行的灵活性至关重要。其次,通过保证压缩重建后图像的语义保真度,确保了信息的准确性,这对于无人机执行侦察、监视等高精度任务非常关键,能够确保目标检测与追踪的信息准确无误。可以利用多模态态势生成模型范式,结合高效的多模态视频图像编解码方法,建立起具有端到端运行多模态态势生成算法能力的边缘计算平台,以低功耗、低延时、高精度的性能构建多模态态势生成终端。算法上,多模态图像成像系统可以分成编码端和解码端两部分,算法框架如图15所示。其中,语义提取[74]旨在从文本中捕捉语言符号背后的深层语义表示,其核心是通过计算模型将离散符号映射为连续向量空间中的稠密向量。该过程通常依赖于词嵌入技术,将词汇转化为低维分布式向量,并借助注意力机制动态加权不同上下文成分的贡献。近年来,预训练语言模型(Pre-trained language models, PLMs)如基于变换器的双向编码器表示(Bidirectional encoder representations from Transformer, BERT)通过大规模语料库的掩码语言建模(Masked language modeling, MLM)任务,学习融合上下文信息的深层语义特征。这些模型通过多层变换器(Transformer)编码器捕获词汇间复杂依赖关系,从而生成能够支撑下游自然语言处理任务的通用语义表示。态势生成终端具有通信带宽紧缺、供电紧俏、物理空间狭小、可编程度低和面向人类视觉的特点。因此要求态势生成及编解码算法复杂度不可过高,具有场景和气候普适性,重建图像质量以人眼观感为首要倾向,同时尽可能压缩态势生成所需的信息源,以达到节省带宽的目的。

3.4.1 大压缩比编码算法

在无人机遥感中,数据传输的带宽有限,尤其是在执行大范围监测任务时,大量的遥感数据会导致通信瓶颈。大压缩比编码算法能够在保持数据质量的前提下,实现数据的高效压缩,从而减少数据传输所需的带宽,提高通信效率。因此,可以使用“压缩特征-信息增强视频”的生成式算法,采用生成对抗式训练方式,从而实现具有强大生成能力的轻量级成像网络。

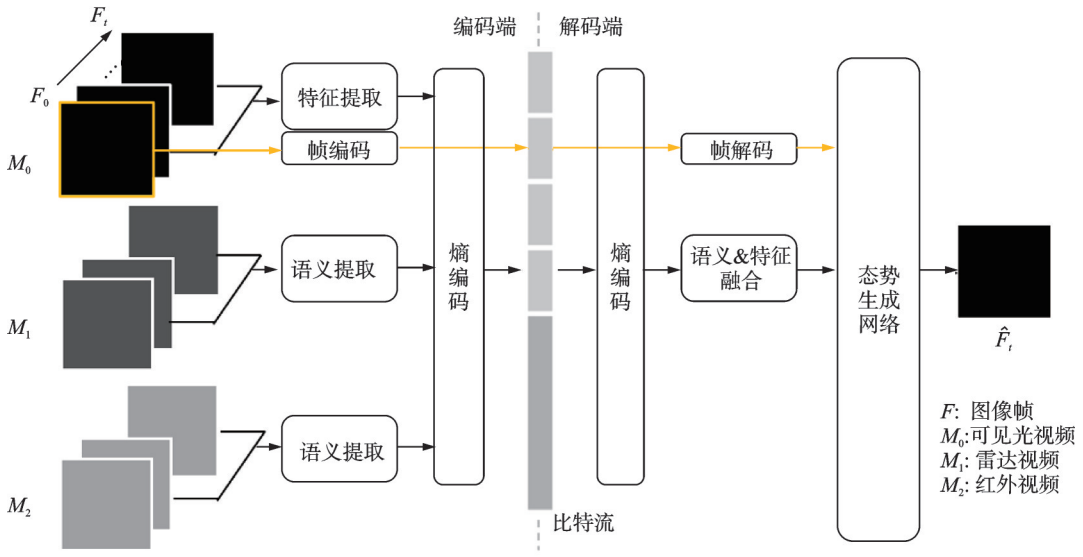

图15 多模态态势生成算法示意图

Fig.15 Schematic diagram of the multi-modal situation awareness generation algorithm

算法上,多模态图像成像系统可以分成编码端和解码端两部分。该算法以多模态视频M_m为输入,其中M_0为可见光视频,其他为雷达视频、红外视频等其他语义视频。该算法于编码端实现语义信息和特征图像的提取和压缩,将多模态视频转入隐式的特征空间,在高度保留语义信息的同时压缩比特流大小,以满足无人机受限的通信带宽。对可见光视频非关键帧进行特征提取时,使用类U-Net网络;对关键帧进行图像编码时,使用传统的帧内编码或图像编码算法;对其他模态视频进行语义提取时,使用混合式提取方法。算法于解码端实现由压缩信息至完整多维图像的生成过程,在获得高保真视频画面的同时集合多模态语义信息。首先,二进制比特流经过对应的熵解码和图像解码过程,得到关键帧图像和非关键帧特征及语义信息;其次,特征和语义信息经过融合后得到增强后的隐式空间信息;最终,该信息经过强大的生成式网络,生成出最终的高纬度、多语义、高保真视频。整个编解码算法以端到端、生成-对抗的训练方式进行基于大数据的学习,收敛后固定权重,该算法即可离线使用。

解码端是该态势生成算法的重点,其生成能力决定着该态势生成终端的成像质量及成本。解码端可以分成稀疏运动估计、密集运动估计和生成网络3部分。稀疏运动估计用于生成初步光流图,以表征当前帧与关键帧之间的大致运动关系,该步骤不涉及深度学习部分,仅以传统计算方式完成;密集运动估计由传统计算和深度学习网络结合而成:前模块生成的帧在通道维度堆叠,经U-Net处理后,分两路经两个卷积层处理,一个用于生成多个表征变形发生处的蒙版图,另一个用于生成表征遮挡位置的遮挡示意图;末端的生成网络首先对关键帧进行带池化的卷积处理,使其分辨率降至与运动图相同,以便在密集运动和遮挡示意图的指引下进行变形,最终经由多层卷积后生成出高质量的原始分辨率图像。

多模态态势生成模型基于大量数据训练形成的网络先验知识,构成了高知识水准、信息结构完备的知识图谱。端侧定期的微调训练可以深化多模态态势生成模型对复杂情景处理的能力和智能化水平,具有高扩展度。知识图谱利用图论原理,通过实体(节点)和实体间关系(边)的结构化表示,构建了一个庞大且复杂的知识网络,为多模态态势生成模型提供了一个广泛的知识基础。这一结构化的知识表示方式使得模型能够基于已有的知识进行响应生成,并通过学习知识图谱中新增的实体和关系信

息,实现知识库的动态扩展和更新,进而保持模型输出的时效性和准确性。更重要的是,知识图谱的深层语义关系和逻辑推理能力,使得模型在处理包含文本、图像、声音等多模态数据时,能够进行更准确的语义分析和逻辑推理,显著提高了决策和预测的准确性。此外,知识图谱的跨领域知识整合能力,通过实体识别和关系映射技术,极大地扩展了模型的适应性和应用范围,使其能够灵活应对各种领域的特定需求。这种知识图谱原理与离线推理模型的结合,不仅促进了模型在知识理解、逻辑推理、跨领域适应等方面的显著进步,而且为构建更为高效、智能的多模态态势生成系统奠定了坚实的基础。此多模态态势生成模型[75]可将视频频谱等多模态信息压缩 400 倍以上,与国内外其他相关技术相比,压缩能力有2~4倍的提升,具有世界技术水平。

3.4.2 质量评估指标体系

态势生成的质量评估是确保生成结果符合实际应用需求的关键。针对多模态态势生成的特点,研究者提出了一系列质量评估指标,如生成的态势图的分辨率、准确性、时效性等,确保在不同环境和任务条件下生成的态势图具有较高的质量和可靠性。设多模态态势生成的图像为I,经过专用于检测人眼感兴趣区域(Region of interest,ROI)的预训练神经网络得到ROI区域掩膜$\mathrm{Mask}_{\mathrm{ROI}}$,将$\mathrm{Mask}_{\mathrm{ROI}}$划分为不交叠、尺寸为$Q\times Q$的图像块,根据图像块中ROI像素点占比进一步将图像块分类为ROI块和非ROI块。设图片中的ROI块数量为N_{r},非ROI块数量为N_{nr}。对ROI块进行语义维度和信号维度两方面的质量评估,对非ROI块仅进行信号维度方面的质量评估。整幅图像的质量分数将由所有图像块的质量分数加权平均计算得到。语义维度的质量评估由对比语言-图像预训练模型(Contrastive language-image pre-training,CLIP)衡量。由于缺乏可参考的高质量图像,语义维度的质量评估需要首先建立语义信息丰富、完整、清晰的先验特征集,以供待测质量的图像块提取的特征进行对比和误差衡量。建立语义先验特征集的过程如下:

(1)从高级机器视觉任务的深度学习数据集中整理出语义信息丰富、完整、清晰的图片数据集,设图片数量为M;

(2)使用CLIP的图像编码器对(1)中搜集的图片进行语义特征提取,设每张图片提取到的语义特征向量维度为P。

(3)使用多元高斯模型(Multivariate Gaussian model,MGM)对$M\times P$的特征矩阵进行参数拟合,得到均值矩阵$\boldsymbol{v}_1$(维度为$1\times P$)和协方差$\boldsymbol{\Sigma}_1$(维度为$P\times P$)矩阵,这将作为语义保真对比的高质量先验特征集的量化表达。

对于所有ROI图像块,使用CLIP图像编码器进行特征提取,得到$N_{\mathrm{r}}\times P$的待测特征矩阵,使用MVG拟合得到均值矩阵$\boldsymbol{v}_2$和协方差矩阵$\boldsymbol{\Sigma}_2$。使用Mahalanobi距离计算[76]待测语义特征和高质量先验语义特征的误差分数,作为待测图像ROI区域的语义保真分数,Mahalanobis距离计算过程如式(12)所示,S_{semantic}数值越小,表明语义保真质量越高。

$$S_{\mathrm{semantic}}(\boldsymbol{v}_1,\boldsymbol{v}_2,\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)=\sqrt{(\boldsymbol{v}_1-\boldsymbol{v}_2)^{\mathrm{T}}\left(\frac{\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2}{2}\right)^{-1}(\boldsymbol{v}_1-\boldsymbol{v}_2)} \tag{12}$$

信号维度的质量评估由自然图像质量评估指标(Natural image quality evaluator,NIQE)衡量[77]。自然图像质量评估指标是一个在图像视频压缩、重建、增强等领域应用广泛的无参考质量评估方法,侧重衡量图像信号自然度上的质量水平。NIQE以块级为单位提取自然场景统计特性相关的特征,同从高质量自然图像集中建立的先验特征集表征进行误差计算,得到自然度分数。将所有ROI块和非ROI块组成的待测图片进行NIQE分数计算,得到图像信号自然度的分数,如式(13)所示,S_{signal}数值越小,表示图像信号自然度越高。

$$S_{\mathrm{signal}}=\mathrm{NIQE}(I) \tag{13}$$

计算多模态态势生成图像的综合质量评估指标之前,需要将语义质量分数和信号质量分数进行逆转换,即

$$\begin{cases} S'_{\text{signal}} = \dfrac{1}{1+S_{\text{signal}}} \\ S'_{\text{semantic}} = \dfrac{1}{1+S_{\text{semantic}}} \end{cases} \tag{14}$$

多模态态势生成图像的综合质量评估指标由两方面质量分数加权平均得到,计算如下

$$S_{\text{overall}} = \frac{S'_{\text{signal}} \times W_{\text{signal}} + S'_{\text{semantic}} \times W_{\text{semantic}}}{W_{\text{signal}} + W_{\text{semantic}}} \tag{15}$$

式中 W 表示不同方面的权重系数。

在无人机遥感领域,图像质量评估的核心矛盾在于传统基于自然场景统计的无参考方法(如NIQE)关注低层信号保真度,却与高层语义信息保真度这一核心任务需求脱节。为此,本文提出将语义保真度与信号自然度分离评估。其理论依据在于,所采用的CLIP模型将图像映射至高维语义嵌入空间,使得特征距离直接反映语义内容差异;进而通过多元高斯模型对该语义特征分布进行建模,以马氏距离量化语义偏离程度。该方法实现了从评估"统计自然性"到评估"信息正确性"的范式转变,能够敏锐捕捉生成过程中关键的语义失真,为面向任务的高保真无人机遥感系统提供了更直接的质量度量依据。

与国内外现有的图像质量评估技术相比,此多模态态势生成质量评估方法能够实际应用在无人机探测的无参考场景,针对在无人机巡航监测中对目标区域的画质要求,对多模态态势生成的图像进行ROI区域监测,并对语义和信号两方面进行综合评估,具有领先水平。

4 结束语

本文系统性地提出了无人机多模态超宽谱认知仪的创新设计方案和传统无人机遥感续航的一些核心问题。方案通过设计活塞发动机与锂电池混合能源构型及垂直起降飞翼布局,显著提升了航时效能;开发复眼多目相机与轻小型W波段SAR雷达,突破亚毫米级振动补偿技术,实现空-时-频多维协同感知;构建时空配准框架下的多层次融合机制(数据层-特征层-语义层),结合轻量化深度学习模型,将低可观测目标检测精度提升至90%以上,优于近期基于轻量化YOLO的检测方案及先进的Transformer融合模型;通过无参考质量评估模型保障语义保真度,达成400倍压缩下的高保真三维态势生成,在压缩比与重建质量的权衡上达到了前沿水平[45,78]。

本研究为构建智能无人机遥感系统提出了一些设计和改进方案。然而,作为一个前瞻性的综合设计方案,其在迈向工程化与大规模应用的道路上仍面临若干内在挑战,这亦是未来研究需要着力推进的方向。首先,多模态高性能载荷的集成在提升感知能力的同时,也带来了系统层级的复杂性,其电磁兼容性、热管理与长期野外作业的可靠性需通过更严苛的工程样机试验进行验证与优化。其次,尽管所采用的轻量化深度学习模型在测试场景中取得了高精度,但其对未知极端环境与强对抗干扰的泛化能力仍有局限,且模型决策过程缺乏可解释性,未来需融入因果推理等机制以增强鲁棒性与可信度。进一步,系统的能量供给与异构计算资源之间尚未形成全局动态协同,开发跨平台、载荷与计算模块的自适应资源调度策略,是平衡任务效能、续航时间与数据质量的关键。最终,当前设计主要服务于单平台智能提升,而未来实战应用必然依赖于多机集群协作。因此,如何将本仪器的感知认知能力扩展至分布式群体,实现多机间的知识共享与协同决策,是构建体系化、自主化遥感网络所必须攻克的前沿课题。正视这些局限并围绕其开展持续研究,将推动该技术从卓越的原理验证走向广泛的实际应用。

参考文献：

[1] 中华人民共和国国务院. 中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要[R]. 北京: 中华人民共和国国务院, 2021.

[2] 国家自然资源部. 关于加快测绘地理信息事业转型升级更好支撑高质量发展的意见[R]. 北京: 中华人民共和国自然资源部, 2023.

[3] 国家自然科学基金委员会. 国家自然科学基金“十四五”发展规划[R]. 北京: 国家自然科学基金委员会, 2022.

[4] WATTS A C, AMBROSIA V G, HINKLEY E A. Unmanned aircraft systems in remote sensing and scientific research: Classification and considerations of use[J]. Remote Sensing, 2012, 4(6): 1671-1692.

[5] (美) GLEASON T J. 吴汉平, 等译. 无人机系统导论[M]. 北京: 电子工业出版社, 2003.

[6] JEREMIAH G. Unmanned aerial systems[R]. USA:Congressional Research Service, 2012.

[7] 李德仁, 龚健雅, 邵振峰. 从数字地球到智慧地球[J]. 武汉大学学报(信息科学版), 2010, 35(2): 127-132, 253-254.
LI Deren, GONG Jianya, SHAO Zhenfeng. From digital earth to smart earth[J]. Geomatics and Information Science of Wuhan University, 2010, 35(2): 127-132, 253-254.

[8] 李德仁. 摄影测量与遥感学的发展展望[J]. 武汉大学学报(信息科学版), 2008, 33(12): 1211-1215.
LI Deren. Development prospect of photogrammetry and remote sensing[J]. Geomatics and Information Science of Wuhan University, 2008, 33(12): 1211-1215.

[9] 臧克, 孙永华, 李京, 等. 微型无人机遥感系统在汶川地震中的应用[J]. 自然灾害学报, 2010, 19(3): 162-166.
ZANG Ke, SUN Yonghua, LI Jing, et al. Application of miniature unmanned aerial vehicle remote sensing system to Wenchuan earthquake[J]. Journal of Natural Disasters, 2010, 19(3): 162-166.

[10] 李德仁, 眭海刚, 单杰. 论地理国情监测的技术支撑[J]. 武汉大学学报(信息科学版), 2012, 37(5): 505-512, 502.
LI Deren, SUI Haigang, SHAN Jie. Discussion on key technologies of geographic national conditions monitoring[J]. Geomatics and Information Science of Wuhan University, 2012, 37(5): 505-512, 502.

[11] 晏磊, 廖小罕, 周成虎, 等. 中国无人机遥感技术突破与产业发展综述[J]. 地球信息科学学报, 2019, 21(4): 475-495.
YAN Lei, LIAO Xiaohan, ZHOU Chenghu, et al. The impact of UAV remote sensing technology on the industrial development of China: A review[J]. Journal of Geo-Information Science, 2019, 21(4): 475-495.

[12] BERNI J A J, ZARCO-TEJADA P J, SUAREZ L, et al. Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle[J]. IEEE Transactions on Geoscience and Remote Sensing, 2009, 47(3): 722-738.

[13] XIANG H, TIAN L. Development of a low-cost agricultural remote sensing system based on an autonomous unmanned aerial vehicle (UAV)[J]. Biosystems Engineering, 2011, 108(2): 174-190.

[14] KHAN A, SCHAEFER D, TAO L, et al. Low power greenhouse gas sensors for unmanned aerial vehicles[J]. Remote Sensing, 2012, 4(5): 1355-1368.

[15] KELCEY J, LUCIEER A. Sensor correction of a 6-band multispectral imaging sensor for UAV remote sensing[J]. Remote Sensing, 2012, 4(5): 1462-1493.

[16] WATTS A C, PERRY J H, SMITH S E, et al. Small unmanned aircraft systems for low-altitude aerial surveys[J]. The Journal of Wildlife Management, 2010, 74(7): 1614-1619.

[17] 马瑞升, 孙涵, 林宗桂, 等. 微型无人机遥感影像的纠偏与定位[J]. 南京气象学院学报, 2005, 28(5): 632-639.
MA Ruisheng, SUN Han, LIN Zonggui, et al. Geometric correction and registration of optical remote sensing image from miniature unmanned aerial vehicle[J]. Journal of Nanjing Institute of Meteorology, 2005, 28(5): 632-639.

[18] 桂德竹. 基于组合宽角相机低空影像的城市建筑物三维模型构建研究[D]. 徐州: 中国矿业大学, 2010.
GUI Dezhu. Research on 3D model construction of urban buildings based on low-altitude images of combined wide-angle camera [D]. Xuzhou: China University of Mining and Technology, 2010.

[19] 程远航. 无人机航空遥感图像动态拼接技术的研究[D]. 沈阳: 东北大学, 2009.
CHENG Yuanhang. Research on the technology of mosaic of aviation remote sensing image from UAV[D]. Shenyang: Northeastern University, 2009.

[20] LI M, LI D, FAN D. A study on automatic UAV image mosaic method for paroxysmal disaster[J]. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2012(B6): 123-128.

[21] FUJITA S, TAKAHASHI K, FUJII T. High-quality multi-view image extraction from a light field camera considering its

physical pixel arrangement[J]. IEICE Transactions on Information and Systems, 2019(4): 702-714.
[22] QI Q, FU R, SHAO Z, et al. Multi-aperture optical imaging systems and their mathematical light field acquisition models[J]. Frontiers of Information Technology & Electronic Engineering, 2022, 23(6): 823-844.
[23] SOSNOWSKI T, BIESZCZAD G, MADURA H, et al. Thermovision system for flying objects detection[C]//Proceedings of 2018 Baltic URSI Symposium (URSI). Poznan, Poland: IEEE, 2018: 141-144.
[24] ZHANG Y, LIU Y, SUN P, et al. IFCNN: A general image fusion framework based on convolutional neural network[J]. Information Fusion, 2020, 54: 99-118.
[25] WANG Y, PU J, MIAO D, et al. SCGRFuse: An infrared and visible image fusion network based on spatial/channel attention mechanism and gradient aggregation residual dense blocks[J]. Engineering Applications of Artificial Intelligence, 2024, 132: 107898.
[26] AGUASCA A, ACEVO-HERRERA R, BROQUETAS A, et al. ARBRES: Light-weight CW/FM SAR sensors for small UAVs[J]. Sensors, 2013, 13(3): 3204-3216.
[27] HAWKINS B P, TUNG W. UAVSAR real-time embedded GPU processor[C]//Proceedings of IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium. Yokohama, Japan: IEEE, 2019: 545-547.
[28] YOCKY D, WEST R. Unmanned aerial vehicle synthetic aperture RADAR for surface change monitoring[R]. Albuquerque, NM, USA: Sandia National Lab, 2022.
[29] REN H, SUN Z, YANG J, et al. Swarm UAV SAR for 3-D imaging: System analysis and sensing matrix design[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5238316.
[30] 高松, 滕克难, 段哲. 美军核心电子战支援装备及其发展趋势分析[J]. 飞航导弹, 2019(11): 12-17.
GAO Song, TENG Kenan, DUAN Zhe. Analysis of US army's core electronic warfare support equipment and its development trend[J]. Aerodynamic Missile Journal, 2019(11): 12-17.
[31] ZOU Y, WU L, FAN J, et al. A convergent iteration method for 3-D AOA localization[J]. IEEE Transactions on Vehicular Technology, 2023, 72(6): 8267-8271.
[32] WU G Z, ZHANG M, GUO F C. High-resolution direct position determination based on eigenspace using a single moving ULA[J]. Signal, Image and Video Processing, 2019, 13(5): 887-894.
[33] 何仲涛. SAR光学遥感图像融合技术研究[D]. 成都: 电子科技大学, 2022.
HE Zhongtao. Research on fusion technology of sar optical remote sensing image[D]. Chengdu: University of Electronic Science and Technology of China, 2022.
[34] ZHANG H, SHEN H, YUAN Q, et al. Multispectral and SAR image fusion based on Laplacian pyramid and sparse representation[J]. Remote Sensing, 2022, 14(4): 870.
[35] YEHIA A, SAFY M, AMEIN A S. Fusion of high-resolution SAR and optical imageries based on a wavelet transform and IHS integrated algorithm[J]. International Journal of Engineering Research in Africa, 2021, 52: 62-72.
[36] FU Y, YANG S, YAN H, et al. Optical and SAR image fusion method with coupling gain injection and guided filtering[J]. Journal of Applied Remote Sensing, 2022, 16(4): 046505.
[37] LI X, LEI L, SUN Y, et al. Multimodal bilinear fusion network with second-order attention-based channel selection for land cover classification[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 1011-1026.
[38] YE Y, LIU W, ZHOU L, et al. An unsupervised SAR and optical image fusion network based on structure-texture decomposition[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 4028305.
[39] DOLPH C, MINWALLA C, LOMBAERTS T, et al. Ground to air testing of a fused optical-radar aircraft detection and tracking system[C]//Proceedings of AIAA SCITECH 2022 Forum. San Diego, CA, USA: AIAA, 2022: AIAA2022-0498.
[40] SUN H, LIU Q, WANG J, et al. Fusion of infrared and visible images for remote detection of low-altitude slow-speed small targets[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 2971-2983.
[41] WANG B, ZOU Y, ZHANG L, et al. Multimodal super-resolution reconstruction of infrared and visible images via deep learning[J]. Optics and Lasers in Engineering, 2022, 156: 107078.
[42] LI Y, LIU G, BAVIRISETTI D P, et al. Infrared-visible image fusion method based on sparse and prior joint saliency detection and LatLRR-FPDE[J]. Digital Signal Processing, 2023, 134: 103910.
[43] LIU D, LI Y, LIN J, et al. Deep learning-based video coding: A review and a case study[J]. ACM Computing Surveys, 2021, 53(1): 1-35.
[44] YU S, SOHN K, KIM S, et al. Video probabilistic diffusion models in projected latent space[C]//Proceedings of 2023 IEEE/

CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, BC, Canada: IEEE, 2023: 18456-18466.

[45] CAREIL M, MUCKLEY M J, VERBEEK J, et al. Towards image compression with perfect realism at ultra-low bitrates[EB/OL]. (2023-10-10). https://arxiv.org/abs/2310.10325.

[46] LEI E, USLU Y B, HASSANI H, et al. Text + sketch: Image compression at ultra low rates[EB/OL]. (2023-07-08). https://arxiv.org/abs/2307.01944.

[47] ZHANG P, WANG S, WANG M, et al. Rethinking semantic image compression: Scalable representation with cross-modality transfer[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(8): 4441-4445.

[48] WANG Z, LI A, ZHU L, et al. CustomVideo: Customizing text-to-video generation with multiple subjects[EB/OL]. (2024-01-01). https://arxiv.org/abs/2401.09962.

[49] SONG H, LIU Q, WANG G, et al. Spatiotemporal satellite image fusion using deep convolutional neural networks[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018, 11(3): 821-829.

[50] MA J, YU W, LIANG P, et al. FusionGAN: A generative adversarial network for infrared and visible image fusion[J]. Information Fusion, 2019, 48: 11-26.

[51] XIAO Erliang, LIN Huaxi, JIAN Xianzhong. Medical image fusion algorithm adopting generative adversarial network to explore latent space[J]. Information and Control, 2021, 50(5): 538-549.

[52] MILDENHALL B, SRINIVASAN P P, TANCIK M, et al. NeRF: Representing scenes as neural radiance fields for view synthesis[C]//Proceedings of Computer Vision—ECCV 2020. Cham: Springer, 2020: 405-421.

[53] HO J, HO J, JAIN A, et al. Denoising diffusion probabilistic models[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, BC, Canada: ACM, 2020: 6840-6851.

[54] NICHOL A, DHARIWAL P, RAMESH A, et al. GLIDE: Towards photorealistic image generation and editing with text-guided diffusion models[EB/OL]. (2021-12-12). https://arxiv.org/abs/2112.10741.

[55] RAMESH A, DHARIWAL P, NICHOL A, et al. Hierarchical text-conditional image generation with CLIP latents[EB/OL]. (2022-04-01). https://arxiv.org/abs/2204.06125.

[56] BLATTMANN A, ROMBACH R, LING H, et al. Align your latents: High-resolution video synthesis with latent diffusion models[C]//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, BC, Canada: IEEE, 2023: 22563-22575.

[57] TU Z, WANG Y, BIRKBECK N, et al. UGC-VQA: Benchmarking blind video quality assessment for user generated content [J]. IEEE Transactions on Image Processing, 2021, 30: 4449-4464.

[58] KANG L, YE P, LI Y, et al. Convolutional neural networks for No-reference image quality assessment[C]//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014: 1733-1740.

[59] YING Z, NIU H, GUPTA P, et al. From patches to pictures (PaQ-2-PiQ): Mapping the perceptual space of picture quality [C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 3572-3582.

[60] XING F, WANG Y G, WANG H, et al. StarVQA: Space-time attention for video quality assessment[C]//Proceedings of 2022 IEEE International Conference on Image Processing (ICIP). Bordeaux, France: IEEE, 2022: 2326-2330.

[61] YOU J, YOU J. Long short-term convolutional transformer for no-reference video quality assessment[C]//Proceedings of the 29th ACM International Conference on Multimedia. [S.l.]: ACM, 2021: 2112-2120.

[62] ANDERSON J D. Aircraft performance and design[M]. Boston: WCB/McGraw-Hill, 1999: 216-218.

[63] JIAO S, ZHANG G, ZHOU M, et al. A comprehensive review of research hotspots on battery management systems for UAVs[J]. IEEE Access, 2023, 11: 84636-84650.

[64] RIZZONI G, GUZZELLA L, BAUMANN B M. Unified modeling of hybrid electric vehicle drivetrains[J]. IEEE/ASME Transactions on Mechatronics, 1999, 4(3): 246-257.

[65] REN B, LIU B, WANG Q, et al. DCIFNet: Cross-modal fusion with correction and interaction for optical-SAR land cover classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 5643218.

[66] PHAN D, OH C M, KIM S H, et al. Object recognition by combining binary local invariant features and color histogram[C]//Proceedings of 2013 2nd IAPR Asian Conference on Pattern Recognition. Naha, Japan: IEEE, 2014: 466-470.

[67] CHENG X, HUANG Z, BAI L, et al. M^3SC: A generic dataset for mixed multi-modal (MMM) sensing and communication integration[J]. China Communications, 2023, 20(11): 13-29.

[68] ZHANG W, ZHENG L, YAO G, et al. Research on the application of intelligent extraction of wind turbines in satellite remote sensing images based on YOLO-world[C]//Proceedings of IGARSS 2025—2025 IEEE International Geoscience and Remote

Sensing Symposium. Brisbane, Australia: IEEE, 2025: 7663-7666.
[69] VAN RIJSBERGEN C J. Information retrieval[M]. 2nd ed. London: Butterworths, 1979: 123-125.
[70] YANG X, YANG J, YAN J, et al. SCRDet: Towards more robust detection for small, cluttered and rotated objects[C]// Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea: IEEE, 2020: 8231-8240.
[71] WANG C, ZHANG M, LIU H. Lightweight YOLOv5: An efficient UAV-based target detection model for remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-15.
[72] LIU R, LING J, ZHANG H. SoftFormer: SAR-optical fusion transformer for urban land use and land cover classification[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2024, 218: 277-293.
[73] DI Z, CHEN X, WU Q, et al. Learned compression framework with pyramidal features and quality enhancement for SAR images[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 4505605.
[74] ZHAO Y, GUAN M, LUO J, et al. Research on IC-LOGO extraction algorithm based on deep semantic segmentation network[C]//Proceedings of 2024 6th International Conference on Artificial Intelligence and Computer Applications (ICAICA). Dalian, China: IEEE, 2025: 386-389.
[75] WAN R, ZHENG Q, FAN Y. M3-CVC: Controllable video compression with multimodal generative models[C]// Proceedings of ICASSP 2025—2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Hyderabad, India: IEEE, 2025: 1-5.
[76] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision[C]// Proceedings of International Conference on Machine Learning. San Diego, CA, USA: JMLR, 2021.
[77] MITTAL A, SOUNDARARAJAN R, BOVIK A C. Making a "completely blind" image quality analyzer[J]. IEEE Signal Processing Letters, 2013, 20(3): 209-212.
[78] YANG L, WANG S, ZHANG H. Semantic-aware video compression for UAV surveillance using neural radiance fields and adaptive quantization[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(2): 876-890.

作者简介:

施云鹤(1998-),男,博士研究生,研究方向:阵列信号处理与阵列设计,E-mail:bx2404005@nuaa.edu.cn。

张小飞(1977-),**通信作者**,男,教授,博士生导师,研究方向:移动通信、阵列信号处理、通信信号处理,E-mail:zhangxiaofei@nuaa.edu.cn。

吴启晖(1970-),男,教授,博士生导师,研究方向:认知无线网络的算法与优化、软件定义无线电以及无线通信系统,E-mail:wuqihui2014@sina.com。

(编辑:张黄群)

Research on UAV Multi-modal Ultra-Wide Spectrum Cognitive Instrument

SHI Yunhe, ZHANG Xiaofei*, WU Qihui

(College of Electronic and Information Engineering, Nanjing University of Aeronautics & Astronautics, Nanjing, 211106, China)

Abstract: The unmanned aerial vehicle (UAV) multi-modal ultra-wide spectrum cognitive instrument constructs an intelligent remote sensing system by deeply integrating visible light, infrared, synthetic aperture radar (SAR), and wireless spectrum sensors. It aims to overcome fundamental bottlenecks in traditional UAV remote sensing: Limited endurance severely constraining detection range, insufficient payload capacity restricting multi-modal perception, weak onboard computing capability causing real-time processing delays, and finite communication capacity hindering high-fidelity situational assessment. To address endurance challenges, the design employs a hybrid energy configuration combining piston engines and lithium batteries with a vertical take-off and landing (VTOL) flying-wing layout, significantly enhancing operational longevity. For payload limitations, it develops a compound-eye multi-camera array for wide-field high-resolution imaging and integrates a W-band miniaturized SAR radar with submillimeter-level vibration compensation technology, enabling air-time-frequency multi-dimensional collaborative perception. To resolve real-time processing constraints, a spatiotemporal registration framework and lightweight deep learning model establish a multi-level fusion mechanism (data-feature-semantic layers), elevating detection accuracy for low-observable targets beyond 90%. Targeting communication bottlenecks, innovative generative coding combined with knowledge-graph-driven situational reconstruction achieves high-fidelity 3D situational generation under 400-fold compression, quantified via a no-reference quality assessment model for semantic fidelity. Validated in defense reconnaissance for real-time tracking of concealed targets in complex electromagnetic environments and in emergency response for flood monitoring and 3D reconstruction, the instrument demonstrates practical value in complex scenarios. Future research should deepen cross-modal semantic understanding optimization and dynamic cooperative control of UAV swarms to advance intelligent remote sensing toward real-time, autonomous cognitive evolution.

Highlights:

1. A design scheme for a UAV platform integrating hybrid energy sources (piston engine + lithium battery) with a vertical take-off and landing flying-wing configuration is proposed, and a comparative analysis is conducted on the current research status of multi-modal ultra-wideband cognitive sensor technology for UAVs.
2. A multi-level "data-feature-semantic" fusion cognitive mechanism is established, which combines lightweight deep learning models with four-modal (visible light, infrared, SAR, and wireless spectrum) spatiotemporal registration. This is expected to significantly enhance the detection accuracy of low-observable targets in complex scenarios.
3. A high-compression-ratio situational generation technology based on generative coding is proposed, which is anticipated to achieve high-fidelity 3D situational reconstruction and ensure semantic fidelity through a no-reference quality assessment model.

Key words: unmanned aerial vehicle; hybrid energy; multi-modal payload integration; multi-modal fusion; quality assessment

Foundation item: National Natural Science Foundation of China (No. 62427801).

Received: 2025-11-12; **Revised**: 2026-01-15

***Corresponding author**, **E-mail:** zhangxiaofei@nuaa.edu.cn.