

卫星边缘计算智能化技术研究进展*

张其阳¹, 邢若琳¹, 李元哲², 周 傲¹, 徐梦炜¹, 王尚广¹

¹(网络与交换全国重点实验室(北京邮电大学), 北京 100876)

²(清华大学 智能产业研究院, 北京 100084)

通信作者: 王尚广, E-mail: sgwang@bupt.edu.cn



摘 要:近年来,随着太空任务的日益复杂化,太空数据呈现爆炸式增长.然而,受限于星地链路带宽限制和稀缺的频谱资源,传统弯管架构在星地数据传输中遭遇瓶颈.此外,星上数据必须等待卫星经过地面站上空才能下传,而大规模建设地面站不仅成本高昂,还面临地缘政治风险和经济收益的不确定性.卫星边缘计算作为一种有效的解决方案,通过在卫星边缘引入移动边缘计算技术,能够显著提升用户体验,同时有效减少网络冗余流量.在轨处理星上原始数据不仅缩短了数据获取时间,还减少了对地面站的依赖.此外,卫星边缘计算结合人工智能技术,为应对当前挑战提供了高效且充满潜力的解决方案.综述卫星边缘计算智能化技术的研究现状:首先探讨其在多个典型场景下的需求与应用;随后分析该领域的关键挑战和研究进展;最后归纳若干开放性研究课题,并提出可借鉴的新思路.期望通过讨论,为推动卫星边缘计算智能化技术创新与实际应用提供有价值的参考.

关键词: 卫星边缘计算; 智能化技术; 边缘智能; 星地协同

中图法分类号: TP393

中文引用格式: 张其阳, 邢若琳, 李元哲, 周傲, 徐梦炜, 王尚广. 卫星边缘计算智能化技术研究进展. 软件学报, 2026, 37(1): 326–343. <http://www.jos.org.cn/1000-9825/7410.htm>

英文引用格式: Zhang QY, Xing RL, Li YZ, Zhou A, Xu MW, Wang SG. Research Progress on Intelligent Technologies for Satellite Edge Computing. Ruan Jian Xue Bao/Journal of Software, 2026, 37(1): 326–343 (in Chinese). <http://www.jos.org.cn/1000-9825/7410.htm>

Research Progress on Intelligent Technologies for Satellite Edge Computing

ZHANG Qi-Yang¹, XING Ruo-Lin¹, LI Yuan-Zhe², ZHOU Ao¹, XU Meng-Wei¹, WANG Shang-Guang¹

¹(State Key Laboratory of Networking and Switching Technology (Beijing University of Posts and Telecommunications), Beijing 100876, China)

²(Institute for AI Industry Research, Tsinghua University, Beijing 100084, China)

Abstract: In recent years, the increasing complexity of space missions has led to an exponential growth in space-generated data. However, limited satellites-to-ground bandwidth and scarce frequency resources pose significant challenges to traditional bent-pipe architecture, which faces severe transmission bottlenecks. In addition, onboard data must wait for satellites to pass over ground stations before transmission. The large-scale construction of ground stations is not only cost-prohibitive but also carries geopolitical and economic risks. Satellite edge computing has emerged as a promising solution to these bottlenecks by integrating mobile edge computing technology into satellite edges. This approach significantly enhances user experience and reduces redundant network traffic. By enabling onboard data processing, satellite edge computing shortens data acquisition times and reduces reliance on extensive ground station infrastructure. Furthermore, the integration of artificial intelligence (AI) and edge computing technologies offers an efficient and forward-looking path to address existing challenges. This study reviews the latest progress in intelligent satellite edge computing. First, the demands and applications of satellite edge computing in various typical scenarios are discussed. Next, key challenges and recent research advancements in this field are analyzed.

* 基金项目: 国家自然科学基金(62032003, 62372061, 62425203, U21B2016); 中央高校基本科研业务费专项资金(2024ZCJH11)
收稿时间: 2024-07-28; 修改时间: 2024-12-05, 2024-12-31; 采用时间: 2025-02-11; jos 在线出版时间: 2025-06-04
CNKI 网络首发时间: 2025-06-05

Finally, several open research topics are highlighted, and new ideas are proposed to guide future studies. This discussion aims to provide valuable insights to promote technological innovation and the practical implementation of satellite edge computing.

Key words: satellite edge computing; intelligent technology; edge intelligence; satellite-ground collaboration

1 引言

随着互联网技术的快速发展,物联网、移动通信和边缘计算等新兴技术相继涌现,深刻改变了人类的工作、生活与出行方式.然而,由于高昂的网络部署成本以及地理条件的限制,原始森林、海洋等偏远区域仍难以覆盖地面网络,且难以享受互联网带来的便利.此外,地面基础设施在应对地震等极端环境中易出现瘫痪,无法保障网络的稳定性.作为地面网络的重要补充,低轨卫星星座凭借通信延迟低、覆盖范围广、部署成本相对较低、环境适应性强等优势,正处于快速发展阶段.大规模星座计划(如 Starlink、OneWeb)通过部署大量小型卫星实现全球覆盖,在地球观测、天文研究、环境监测和气象预测等领域发挥了重要作用.例如,在国防演习活动中,低轨卫星星座平台可搭载通信、导航、探测及干扰等多种有效载荷,支持全天候、全天时、实时监控与连续侦察,从而提升全域作战指挥能力和跨域协同能力,助力实现海陆空天等多维作战单元的一体化^[1].

传统卫星多采用弯管结构,其特点是卫星节点仅具备信号透明转发功能,缺乏在轨数据处理能力.这种架构严重依赖地面站,且由于地球自转和星地连接间歇性,进一步加剧了传输延迟,难以满足实时业务需求.此外,弯管结构存在安全隐患,在星地传输过程中数据易遭窃听、篡改或丢失,地面处理中心也可能因网络攻击或自然灾害而导致通信中断.为克服弯管架构的不足,提高卫星数据处理和分析的效率与可靠性,卫星边缘计算逐渐成为研究热点.在卫星边缘上部署计算服务器,可实现数据预处理和特征提取,仅将有效信息传输至地面终端,显著减少数据传输量和网络带宽的占用.同时,随着商用现货 (commercial-off-the-shelf, COTS) 计算硬件的广泛应用和快速发展,边缘计算能力从地面数据中心有效迁移至卫星边缘,计算能力显著提升,从而实现了数据处理的高效迁移.近年来,卫星边缘计算的智能化技术^[2]作为一种融合人工智能与边缘计算的新兴范式,得到了学术界和产业界的高度关注.

通过将计算、存储和数据处理能力嵌入卫星节点,卫星边缘计算显著提升了在轨数据处理的智能化水平和实时响应能力.同时,与传统后地面计算相比,卫星边缘计算在并行度、实时性和运行环境的稳定性方面具备显著优势,广泛应用于通信、遥感、导航及自主控制等领域.尤其在低轨卫星星座平台中,通过搭载智能化有效载荷,利用星间链路实现数据智能分发与高效互联,构建多星协同的智能网格化体系,能够满足全球范围内的高实时性数据服务的需求.表 1 总结并展示了单星计算、多星协同计算与传统后地面计算这 3 种计算方式的特点比较.

表 1 卫星边缘计算方法特点比较

计算方式	与数据产生距离	计算能力	实时性	带宽节省	分布式计算能力	星地协同计算能力	计算卫星星座规模
单星计算	近	弱	低	是	否	否	可扩展性差
多星协同计算	近	较强	高	是	是	是	可扩展分布式计算
后地面计算	远	强	低	否	否	否	无

1.1 卫星边缘计算需求场景

随着数据处理需求的不断增长和应用场景的多样化,卫星边缘计算的智能化应用逐渐成为现代卫星技术的关键支撑,涵盖了增强传统遥感、拓展边缘计算以及科学实验与太空云等多个方向.

(1) 增强传统遥感

遥感卫星捕获数据的速率远超网络传输速率,形成了“数据积累速度与传输能力不匹配”的突出问题.例如,从 1972 年发射的 Landsat 1 到 2021 年发射的 Landsat 9,遥感技术的进步使得数据收集量大幅增长,仅 2022 年新增的遥感数据量就超过 6 000 PB^[2],这对数据传输和实时处理提出了严峻挑战.此外,如表 2 所示,当前国内外遥感卫星搭载光学或合成孔径雷达 (synthetic aperture radar, SAR) 等技术发展迅速,推动了卫星在多领域的应用.

针对传统遥感卫星数据量庞大和实时性需求的挑战, 星上数据处理技术提供了高效解决方案^[3]. 首先, 通过智能识别与筛选机制, 卫星在数据源头实时剔除冗余或不相关数据, 仅传输有价值或紧急信息至地面, 减少地面计算压力. 其次, 利用任务优先级驱动的数据压缩算法, 卫星可针对不同需求实时压缩数据. 例如, 在森林火灾或洪水等自然灾害发生时, 优先传输关键图像和传感器数据, 从而提升任务实时性. 随着机器学习和 AI 算法的不断优化, 未来星上数据处理将进一步实现智能化, 为地面提供精准、高效的数据支持.

表 2 当前国内外代表性遥感卫星相关属性数据比较

卫星名称	载荷类型	分辨率 (m)	发射时间 (年)	所属国家	卫星用途
WorldView-4	光学	0.31 (全色)、1.24 (多光谱)	2016	美国	成像与环境监测
SPOT-7	光学	1.5 (光学)、6 (多光谱)	2014	法国	地球观测
Landsat 9	光学	15 (全色)、30 (多光谱)	2021	美国	地球观测
Capella-3	SAR	0.5	2021	美国	地球观测
ALOS-2	SAR	1 (全色)、3 (多光谱)	2014	日本	地质灾害监测、森林监测
高分03星	光学	亚米级 (全色)、小于4 (多光谱)	2022	中国	海洋防灾减灾、环境监测
高分多模卫星	光学	0.5 (全色)、2 (多光谱)	2022	中国	调查监测等高精度应用需求
宏图一号	SAR	0.5 (全色)、2 (多光谱)	2023	中国	城市地质灾害监测

(2) 扩展边缘计算

卫星边缘计算作为一种特殊的边缘计算形式^[4], 核心目标是在卫星网络的边缘节点进行数据处理, 优化数据传输效率, 减少带宽占用^[5]. 卫星边缘计算适用于地理位置分散和网络环境复杂的场景. 例如, 在内容分发网络 (content distribution network, CDN) 中, 通过将数据缓存到靠近用户的卫星节点, 卫星边缘计算可以显著减少时延, 提升用户体验^[6]. 此外, 卫星边缘计算在多个垂直行业中展现了广阔应用前景. 例如, 在精细农业领域, 卫星边缘计算通过连接的土壤湿度传感器、天气监控系统和水资源 IoT 设备, 实现了农业生产的智能监控与管理. 在环境监测领域, 卫星边缘计算为生态保护提供了重要支撑. 总体来看, 卫星边缘计算作为未来空间网络与地面应用融合的重要技术支撑, 持续的技术创新和应用实践有望突破现实瓶颈.

(3) 科学实验与太空云

近年来, 太空计算技术的进展显著提升了数据处理能力, 推动了“太空云”这一概念的提出. 以阿联酋的“希望号”火星探测器和慧与星载计算机-2 (Spaceborne Computer-2) 为例, 成功验证了在轨数据处理的可行性与高效性. 此外, “太空云”概念是在卫星网络上构建的分布式计算与存储环境, 通过卫星节点协同处理数据, 突破地面网络带宽与延迟的限制. “太空云”的另一关键优势是依托近地轨道太阳能资源. 与地面光伏系统相比, 空间太阳能电站可全天候不断收集太阳能, 极大提高了能源利用效率. 总体而言, 太空计算和太空云不仅在数据处理和实时应用方面具有巨大潜力, 还在可持续能源利用方面展示了前瞻性.

1.2 卫星边缘计算智能化应用

卫星边缘计算能够在卫星端提供泛在的存储和计算能力, 尤其在遥感领域的智能化应用中表现突出, 极大推动了卫星技术的发展. 以下总结了卫星边缘计算的典型智能化应用场景, 包括但不限于以下内容.

(1) 目标监视与跟踪

传统的目标监视与跟踪任务中, 卫星获取的大量数据需要回传至地面中心进行集中处理, 这种方式不仅导致数据传输延迟, 还增加了数据泄露的风险. 引入卫星边缘计算后, 卫星能够在轨完成数据的初步筛选、分类和智能分析, 例如识别、跟踪地面目标及预测运动趋势, 大幅减少回传数据量. 通过缩短数据分析与决策的时间, 卫星边缘计算平台可迅速生成高价值的情报信息, 进而为具体任务提供实时决策支持. 此外, 在星上直接完成关键数据处理, 有效规避了数据回传过程中被拦截或篡改的风险.

(2) 灾害预警与应急响应

卫星边缘计算在自然灾害监测和资源管理领域具有广泛应用. 卫星能够实时监测地球的气候变化、自然灾害及资源分布, 并为环境保护和资源管理提供及时、准确的数据支持. 例如, 在森林火灾、地质灾害等突发事件中,

卫星能够迅速回传受灾区域的影像数据, 帮助救援人员掌握灾情并制定救援方案. 此外, 卫星边缘计算通过对数据的智能分析与处理, 为资源开发和管理提供决策支持, 为社会经济的可持续发展提供重要的信息保障.

(3) 海洋监测与保护

卫星边缘计算技术可集成多种遥感传感器, 实现对全球海洋环境的连续、高精度监测. 其能够提供海表温度、盐度、海流分布及海底地形等高精度数据, 为科学家们深入理解海洋生态系统的复杂性和动态变化提供有力支持. 在海洋监测领域, 卫星边缘计算技术能够实时监测海洋污染情况, 为环保部门提供应急响应支持, 并对渔业资源进行动态评估, 助力精准制定渔业管理决策. 在海洋保护领域, 卫星边缘计算通过长期监测海洋生态系统的变化, 为建立海洋保护区、保护生物多样性及评估人类活动对海洋的影响提供了科学依据.

本文第 1 节概述了卫星边缘计算的背景、需求场景及其典型应用. 第 2 节重点分析卫星边缘计算面临的关键挑战. 在此基础上, 第 3 节介绍相关智能化技术的研究进展. 第 4 节展望卫星边缘计算未来研究方向. 最后总结全文.

2 卫星边缘计算关键挑战

卫星边缘计算虽然正处于快速发展阶段, 但在实现广泛应用并充分释放潜力的过程中, 仍面临一系列复杂且多维度的挑战. 从硬件配置的局限性, 到动态复杂网络环境中的性能表现, 再到极端太空条件下的稳健性, 以及与地面和其他空中网络的高效融合, 各方面均对其实用性和稳定性提出了严峻考验. 本文从资源受限、网络协同、健壮可靠和架构融合这 4 个维度, 对卫星边缘计算面临的主要挑战进行了系统剖析, 并探讨了可能的解决路径.

2.1 资源受限

火箭载荷的计价标准主要基于重量和体积, 因此卫星在发射前必须严格控制这两方面的参数. 发射后, 卫星平台的热量控制与能源管理则成为主要的限制因素. 尤其是对于基于 COTS 计算设备的星载系统, 这些设备主要为地面边缘计算场景设计, 在极端太空环境中对温度控制和能耗管理提出了更高要求. 例如, 根据“北邮一号”卫星的实测数据^[7], 星载树莓派在高负载状态下温度迅速升高至 80 °C 以上, 远高于地面环境下的平均温度 (约 60 °C); 此外, 太阳能电池在单个轨道周期内提供的电能极为有限, 以 12U (20 cm×20 cm×30 cm) 主流卫星平台为例, 其电能需求通常在 15–40 Wh 之间. 这些因素严重限制了计算任务的长时间、高功耗持续运行.

总体而言, 资源受限主要体现在以下方面: 首先, 星上物理条件要求硬件必须具备低功耗、小尺寸的特性, 但这类设备的性能远不及地面数据中心的大规模计算设备, 成为计算效率的瓶颈. 其次, 卫星主要依赖太阳能供电, 但太阳能电池板产生的电力受限于日照条件和轨道因素制约, 难以实现持续、稳定的能源供应. 此外, 遥感卫星每天采集数据量远超其能够下传的数据量, 导致大量数据无法及时传输与处理. 因此, 在重量和体积的严格约束下, 构建温度控制、能源利用高效且计算效率优越的卫星边缘计算架构, 已成为亟待解决的重要问题.

2.2 网络协同

低轨卫星以约 7.8 km/s 的高速绕地运行, 其与地面站的单次过顶通信窗口时间仅 7–10 min. 这种高速移动性给网络协同带来了多层面的挑战, 主要包括: 在物理通信层方面, 在无线链路不稳定的条件下, 必须确保通信的鲁棒性. 例如, 针对星链的网络测量显示, 恶劣天气条件下无线链路的吞吐量显著下降, 延迟和丢包率增加^[8]; 在网络路由层方面, 低轨卫星网络的高度动态拓扑要求路由算法具备高速收敛和自适应能力, 以应对星座的频繁变化, 维持网络稳定与高效^[9]; 在网络应用层方面, 短暂的通信窗口时间要求卫星具有较高的自治能力. 在网络断连或弱网环境下, 卫星必须能够自主完成部分任务, 以保障应用服务的稳定性^[10].

此外, 卫星的下行链路限制是制约数据传输的关键瓶颈. 目前, 主流卫星平台的下行链路容量尚不足以满足其不断增长的数据传输需求^[11,12]. 例如, 单颗卫星捕获的遥感图像数据量通常远超下行链路的传输能力. 尽管卫星使用高频段 (如 Ka 频段) 数据传输^[13], 其下行速率通常不超过 100 Mb/s. 在典型的 6 min 星地通信时间内, 即便以最大带宽速率传输, 最多也只能下传约 4.39 GB 数据量^[14]. 进一步看, 受限于星地通信时间的短暂性、地球自转等因素, 下行链路成为影响卫星数据传输的核心瓶颈, 导致大量卫星观测数据无法及时传输到地面. 因此, 星地网络协同是提升网络性能和任务执行效率的解决方案.

2.3 健壮可靠

太空环境中, 卫星电子器件易受到单粒子效应的影响, 主要包括单粒子翻转、单粒子锁定、单粒子烧毁等现象. 这些效应是由高能粒子与电子器件相互作用引发, 其发生概率受轨道高度、太阳活动及电子器件设计等因素影响. 在太阳活动周期的峰值阶段, 这种风险尤为突出. 在航天器运行于恶劣的环境时, 宇宙射线或辐射带的高能带电粒子(如重离子或质子)具有极强的穿透能力. 例如, 单个高能粒子可能导致存储器或微处理器出现故障, 即发生单粒子事件. 在此类事件中, 单个高能粒子可能引发存储器或微电子器件状态改变, 导致数据位翻转^[15]. 虽然单粒子翻转不会直接损坏器件, 但若发生在重要电路(如控制计算机)中, 可能导致错误指令或程序异常运行, 严重时甚至危及航天器安全. 此外, 单粒子效应不仅会在低轨道明显, 在高轨道上更为显著. 受地磁场的影响, 南大西洋异常区与两极区域的高能粒子辐射更为集中, 进一步增加了单粒子事件的风险.

针对单粒子效应的主流应对策略主要分为两类. 第 1 类是采用辐射硬化器件以降低单粒子事件的发生概率. 然而, 这类器件的计算性能较低且成本极高. 例如, 典型的辐射硬化器件 RAD750 主频仅为 200 MHz, 造价却高达 20 万美元, ARM SoC (system on chip, 片上系统) 的主频可达 1.4 GHz, 造价仅约 50 美元^[16]. 第 2 类是在软件层面. 通过三模冗余、差错检测和硬件纠错等容错机制提高系统可靠性^[17]. 例如, 慧与星载计算机算力可达 TeraFLOP 级(每秒一万亿次浮点运算), 软件加固使其计算误差仅为 0.03%, 在一定程度验证了此类措施的有效性. 尽管上述措施在一定程度上提升了系统的可靠性, 但在太空辐射对卫星边缘计算的影响仍需进一步的研究和应对.

2.4 架构融合

卫星领域的标准化推进长期面临较大困难, 主要原因在于传统卫星应用已难以适应技术发展的新趋势. 早期卫星通信主要面向对可靠性要求极高的小范围客户群体, 采用专用的通信制式和设备, 且以高轨卫星为主. 例如, 广播电视采用高度定制化的电视广播协议进行信号传输^[18], 导致遥感、通信、导航等不同类型卫星之间缺乏统一标准, 形成孤立发展局面, 导致了应用部署不灵活、星座资源利用不充分等一系列问题. 此外, 由于星座间的通信规范和网络协议未能统一, 许多现有应用无法直接在卫星网络中运行, 这不仅阻碍了网络的互联互通, 也给空天地一体化网络融合发展带来巨大的挑战. 针对上述问题, 标准化组织(如 3GPP)提出了非地面网络的解决方案^[19], 该方案旨在通过统一的网络协议和架构设计, 实现星地网络的无缝融合. 然而, 这项工作仍处于早期阶段, 技术和工程层面面临诸多挑战. 从长远来看, 卫星与地面基础设施的架构融合不仅是技术发展的必然趋势, 也是推动未来卫星领域向标准化、协同化、智能化方向迈进的重要动力.

3 相关研究进展

卫星边缘计算智能化技术涵盖了多个关键领域的发展和探索, 是探索并实现智能化卫星的重要支撑. 首先, 云原生卫星平台的兴起标志着卫星边缘计算向高效、灵活和可扩展方向的迈进. 与此同时, 专为卫星设计的智能化操作系统不断演进, 以满足日益复杂的任务需求. 智能化在轨处理则大幅提升了卫星自主性和智能化水平, 使其能够高效完成任务并应对复杂外部环境. 最后, 卫星任务调度与规划的优化是确保各类卫星任务高效完成的核心环节.

3.1 云原生卫星平台

传统卫星载荷依赖专用硬件进行数据传输和处理, 这种方式使卫星载荷的管理和扩展面临诸多挑战. 为此, 许多国家和组织积极研发并应用云原生卫星平台. 例如, 美国国家航空航天局(NASA)和欧洲空间局(ESA)联合推进的项目, 通过云计算技术显著提升了卫星数据处理与传输的效率. NASA 的多任务算法和分析平台将相关数据、算法和计算能力集成于一个统一的云环境中, 从而有效解决数据共享和处理的难题. 随着边缘计算^[20]的快速发展, 虚拟化技术广泛应用在资源受限的设备上. 因此, 基于虚拟化技术的云原生卫星能够整合计算、通信、存储等功能, 在卫星上实现智能推理和多任务处理. 例如, 全球首颗云原生卫星“宝酝号”卫星利用卫星和地面站的计算资源, 成功实现了利用 AI 推理暴雨前后图像变化的能力, 从而达到提前预警灾害的效果. 针对卫星图像的高云覆盖率的问题, 系统会丢弃低质量图像以降低传输开销, 而高质量的卫星图像数据则传回地面进行高精度模型计算, 从而显著提升了数据处理效率和响应速度. 此外, 云原生卫星能够通过云原生应用实现按需更新. 例如, “北邮一号”卫星

通过将服务程序及其相关依赖打包到容器中统一部署, 依据实际需求动态调整资源, 按需启动或关闭服务, 显著提高了星载智能化服务的资源管理和调度方面的能力. 相比之下, 非云原生平台通常提供定制化服务, 因长时间占用有限的星载资源, 导致服务利用率下降. 表 3 总结了云原生卫星和传统卫星在轨部署、扩展性和成本费用等方面的差异.

表 3 传统卫星与云原生卫星的特点比较

特点	传统卫星	云原生卫星
部署	对模拟环境进行现场调试 通信协议适应星地链路	打包和随时可用的容器镜像 未感知IP通信
扩展性	不可更改的自定义任务 终端设备固定, 不能被动态管理	更换镜像, 更新应用 可插拔的设备管理框架
成本费用	由测控机构管理的遥测和遥控 数据传输站管理的下行数据	云统一管理卫星边缘云 负载均衡, 高效利用星座计算资源

图 1 展示了首颗集成卫星和地面计算资源共同处理星载任务的智能化云原生卫星架构. 该架构包括 3 大核心组成部分: 卫星平台、基础功能和泛在服务. 完整的卫星平台由电力系统、通信系统、有效载荷系统以及指挥和数据处理系统组成. 各子系统协同工作以完成复杂的太空任务. 其中, 通信子系统建立无线链路负责数据和命令的传输; 有效载荷系统管理任务数据生成与回传; 指挥和数据处理系统负责遥测和遥控, 确保航天器和地面操作员间的稳定通信.

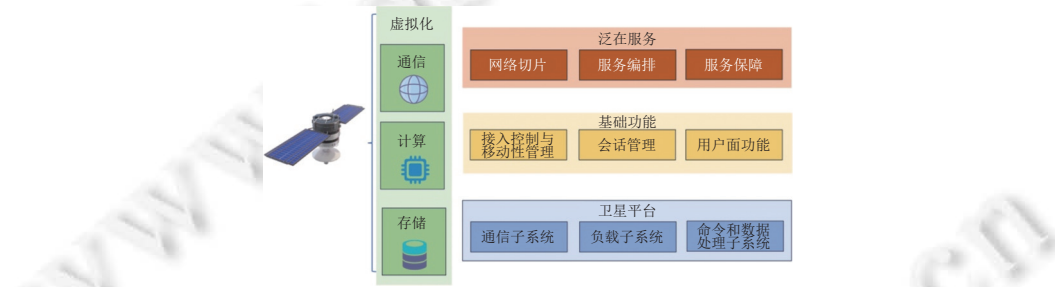


图 1 云原生卫星功能结构示意图

随着星载处理能力的提升, 卫星系统的智能化水平不断提高, 特别是在信号处理和分组处理方面. 信号处理增强了卫星波束的高效切换, 使卫星能够灵活分配无线资源, 实现精细化的策略管理与服务质量控制; 分组处理则支持全星座组网, 为卫星系统提供了更强的互联互通能力. 此外, 基于 SDN 和 NFV 技术的云原生卫星展现了灵活且可编程的资源管理方式, 通过相同的底层硬件提供多种服务, 具备一定程度的通用性, 展示出智能化的网络架构与服务功能. 以卫星核心网为例, 基本网元功能 (即 AMF、SMF、UPF) 满足用户注册和会话管理程序的基本需求, 而身份验证和数据持久化等功能则进一步完善了网络功能. 云原生卫星不仅提供基础的网络连接功能, 还实现了更高级的泛在服务特性, 为最终用户和多样化业务场景提供了更加全面且灵活的支持.

3.2 智能化操作系统

随着低轨卫星任务复杂性的不断提高, 传统操作系统难以同时满足实时性和通用性的需求. 例如, POSIX 系统的实时内核在兼容性方面存在问题, 而采用同时运行实时内核与通用内核的虚拟机方案又会引入大量核间通信开销. 为了解决这些问题, SpaceX 等公司采用了 COTS 器件和 Preempt-Linux 作为主要的硬件与软件方案, 然而, 这种方案仅能实现软实时性能, 实时性仍然不足. 此外, 国内外在新一代卫星操作系统的研发上也取得了一定进展. 例如, Sputnix 的软件定义卫星操作系统和 Starlink 基于传统 Linux 的操作系统. 提升卫星性能、增强容错能力、降低开发难度及优化资源调度等方面, 这些探索取得了积极成效, 但仍存在一些不足之处: 以 Sputnix 为代表的软件定义卫星在操作系统设计中较好地平衡了性能、容错、开发难度和调度资源能力, 但在系统稳定性和实时性方

面仍显不足. 此外, 其容错机制在效果与开销的权衡尚无完美解决方案, 同时普通容器的任务隔离方式存在一定的安全隐患; 而 Starlink 搭载的操作系统则注重生态兼容性、性能提升、实时性和稳定性. 然而, 该系统依赖抗辐射 CPU, 导致任务开发和部署相对滞后, 整体灵活性仍需进一步提高.

针对上述问题, 北京邮电大学研究团队开发了基于 Rust 语言的双内核操作系统. 该系统集成了通用的 Linux 内核与硬实时保证内核, 通过双内核架构的设计显著提升了系统的稳定性, 并实现了高吞吐量、低时延的核间通信能力. 在资源受限的平台上, RROS 系统能够同时满足高实时性和高通用性的任务处理需求, 展现出智能化的资源调度与优化性能. RROS 系统已于 2023 年 12 月搭载“天仪 33 号”卫星成功进入预定轨道并正常运行, 成为全球首个在卫星场景中正式应用的 Rust 编写的双内核操作系统. 目前, 该系统已实现了实时调度子系统、实时内存子系统、实时文件子系统以及实时网络子系统等模块, 其高通用性和高实时性表现如下: 通用内核支持运行 TensorFlow、PyTorch、Spark、MySQL 等任务, 运行开销控制在 5% 以内; 实时内核的响应时间与传统实时操作系统内核 (MicroC/OS、FreeRTOS) 相当. 此外, RROS 系统具备高度自适应能力, 可支持 x86/x86_64/ARMv7/v8 等商用硬件平台, 并成功运行于树莓派和 ARM SoC 集群服务器上. 该系统还提出了星载 ARM SoC 阵列方案及算网融合的探索, 通过提升卫星数据处理效率, 减轻计算负担, 进一步增强了卫星网络与计算的智能化水平.

3.3 智能化在轨处理

随着边缘设备计算能力的不断提升, 深度学习模型的推理过程可以在数据产生源附近完成, 从而实现低时延的实时服务响应. 然而, 受限于设备计算和存储资源, 模型推理在实时性和计算开销之间面临权衡挑战. 当前的研究主要集中在两个方向: 一是设计更高精度的深度学习算法^[21,22], 二是通过模型压缩技术优化训练后的模型以实现高效推理^[23,24]. 其中, 模型压缩技术在提升推理速度方面尤为突出, 其核心目标是通过精简模型参数和计算量, 在保证模型精度的同时降低计算和存储需求. 这项技术显著减少了资源占用, 为深度学习模型在受限设备端高效部署提供了可行性支持. 此外, 随着深度学习的广泛应用, 特别是基于神经网络的矩阵计算模型的普及, 高性能的星载计算平台正逐渐成为支持深度学习应用的关键硬件基础.

(1) 深度学习算法优化

深度学习模型设计: 卫星传感器捕获数据需要经过一系列预处理步骤, 包括数据解析、辐射校正、系统几何校正以及图像分片等. 在这些数据的基础上, 深度学习模型被设计用于进一步的分析和处理. 智能算法 (如表 4 所示) 被广泛应用于目标检测、语义分割和地物变化检测. 基于深度学习的轻量化模型通常小于 100 MB, 其中目标检测和语义分割算法已较为成熟, 并涌现出许多高效的深度学习模型. 相比之下, 遥感地物变化检测模型仍是当前研究的热点与难点, 该类模型因参数庞大、结构复杂, 对输入数据的精度要求较高, 因此难以在轨完成地理精度的校正. 因此, 星上地物变化检测仍需依赖地面处理支持. 同时, 对于高精度检测和识别任务, 地面站通过精确的辐射校正与几何校正, 提取高精度的地物语义和目标信息.

表 4 卫星边缘计算的智能算法与任务需求

算法/任务需求		卫星及其星座	地面站
预处理算法	图像分片、辐射校正、粗正射校正、系统几何校正	图像分片、精准几何校正、精准正射校正、影像相对配准	
目标检测模型	轻量级 YOLO-tiny 系列、MobileNet-SSD、MobileNet、NanoNet 等模型	YOLO 系列、Faster R-CNN、SSD 等模型	
语义分割模型	轻量级 DeepLab-v3、LU-Net、LRUNet、MUNet 等模型	DeepLab-v3、UNet、HRNetv3、PSPNet 等模型	
变化检测模型	—	ESNet、SECDNet、STANet、DTCDNet 等模型	
后处理算法	图像拼接算法、栅格矢量转化算法	图像拼接算法、栅格矢量转化算法	
任务需求	时效性要求高、精度要求低	精度要求高	

在卫星图像处理中, 目标识别算法通常分为两类: 一类是基于区域选择的两段式算法, 另一类是基于回归分析的单段式算法. 传统的目标识别流程采用两段式算法, 首先生成目标候选区域, 然后对这些区域进行分类与位置边框调整. 典型的两段式算法包括 R-CNN、Fast R-CNN 和 Faster R-CNN^[25]. R-CNN 模型是由 Facebook^[26]于 2014

年提出, 通过卷积神经网络提取图像中的候选区域特征, 但运行速度较慢且效率低下. Fast R-CNN 作为 R-CNN 的改进版本, 采用 VGG16 替代 AlexNet^[27], 将特征提取、目标分类和位置回归整合为一体, 显著提高了训练和测试速度, 同时提升了检测精度. 然而, Fast R-CNN 在提取候选区域时仍依赖于 CPU 上的选择性搜索算法, 限制了其运行效率. 为解决上述问题, Faster R-CNN 提出了利用区域建议网络代替选择性搜索算法, 并将候选区域选择、目标分类和定位统一集成到一个框架中, 从而加快了目标检测速度, 同时保持较高的准确性. 目前, 许多研究致力于改进深度学习检测框架并将其应用于卫星图像领域. 例如, Lin 等人^[28]结合级联 AdaBoost 算法与 R-CNN 模型, 提出了一种从粗粒度到细粒度的目标识别算法, 在减少计算量的同时显著提高了高分遥感影像中的目标检测精度. Fratini 等人^[29]基于 Faster R-CNN 提出了特征融合与软判决结合的目标检测算法, 有效提升了对图像中小目标的检测精度. Verfaillie 等人^[30]针对遥感影像中目标方向的不确定性, 将旋转区域网络引入 Faster R-CNN 框架, 通过在全连接层前增加卷积层以减少特征图维度, 从而提升分类器性能, 实现了更优的目标检测效果.

相较于复杂的两段式目标识别方法, 单段式目标识别算法在计算资源有限或对实时性要求较高的场景中展现出显著优势. 在卫星图像处理场景中, 由于目标检测系统对时效性的要求极高, 模型需要具备低计算复杂度和高运行速度. 单段式算法将目标识别问题建模为回归问题, 直接从输入图像中输出目标类别和位置, 无需生成候选区域, 从而显著提高了检测速度. YOLO 系列算法是单段式方法的典型代表, 通过摒弃候选区域生成步骤, 同时预测目标的位置和类别, 极大地提升了检测速度. 自 2015 年推出首个版本以来, YOLO 算法得到了持续优化: YOLOv1 首次将整幅图像作为网络输入, 在输出层对目标类别与位置进行回归. YOLOv2 则在多数据集训练检测器, 在提升检测速度与准确率等方面取得显著改进. YOLOv3 引入二值交叉熵函数和 Darknet-53 基础网络, 在速度和精度之间实现了良好的平衡. 目前, YOLO 系列已发展至第 9 版, 并广泛应用于卫星图像目标检测. 例如, van Etten 等人^[31]针对高分辨率图像和密集小目标检测优化了 YOLO 算法. Verfaillie 等人^[32]结合 YOLO 算法和图像分割技术, 在保证准确率与实时性的同时实现了卫星目标识别. Bensana 等人^[33]通过改进 YOLOv3 网络结构, 有效应对了小目标检测的挑战, 在地震受灾区域遥感图像中成功检测到损毁建筑.

为了降低深度学习模型的推理开销, 设计轻量级模型是一种有效途径. 例如, SqueezeNet^[34]通过创新的 fire module 模块, 将参数量压缩至 AlexNet 的 2%, 但精度保持相当. 该模块使用 1×1 的卷积压缩特征图通道数 (squeeze 层), 并结合 1×1 和 3×3 卷积生成输出特征 (expand 层), 在大幅减少模型参数的同时保持性能. ShuffleNet^[35]通过分组卷积减少参数量和计算量, 并通过 shuffle 操作重新排列通道, 从而实现分组卷积中的信息融合. 这些技术极大地提高了模型的效率, 为资源受限场景下的卫星图像处理提供了新的技术解决方案.

基于压缩技术的深度学习模型推理方法: 性能优越的深度学习模型通常具有庞大的规模, 这不仅带来了存储需求的显著增加, 而且使其很难部署在资源受限的设备上. 此外, 大规模模型的参数量和计算量的增加导致了推理过程中时间和能耗开销. 在卫星边缘环境中, 研究轻量化且高效的星载遥感数据压缩编码技术具有重要意义. 在数据编码阶段, 需显著减少计算需求, 以适应卫星边缘计算环境的资源限制; 即使在较高的数据丢失率条件下, 这些技术也能够准确重建高质量的遥感图像, 为卫星任务的实时处理提供可靠支持. 当前主流的星载遥感数据压缩编码方法总结如表 5 所示. 其中, 经典的模型压缩技术包括模型量化、参数剪枝和共享及知识蒸馏技术等.

表 5 目前主流的星载数据压缩编码方法比较

数据压缩编码方案	核心思想	压缩比	压缩速度	数据完整性	计算复杂度
有损压缩	丢失原始大型数据细节	高	高	中等	中等
无损压缩	无损压缩原始数据	中等	中等	高	高
近乎无损压缩	最小数据丢失可控	中等	中等	中等	中等
矢量化压缩	输入数据被转换成不同的域	高	中等	高	高
转换编码	输入数据被分类成一组有限的值	高	中等	高	高

(a) 模型量化技术. 模型量化是一种通过减少模型参数数值表示位数来降低模型复杂性和内存占用的有效方法. 传统的量化方式包括 16-bit、8-bit 和 2-bit 表示. 研究表明, 低比特量化在显著减少参数量和计算量的同时, 几乎不影响模型精度, 从而展现出优异的性能. 当前, 8-bit 量化是工业界最常用的方案, 它能够实现参数体积的 4 倍

压缩,同时降低内存占用并提升运行速度.2018年,谷歌提出了一种新的量化方案^[35],将神经网络的权重和激活值压缩为 8-bit 整数,并将部分参数(如偏置向量)保留为 32-bit.这种方法针对仅支持整数运算的硬件进行了优化,大幅提升了量化推理的效率.此外,研究者还提出了乘积量化和矢量量化技术^[36],进一步探索其在深度学习模型压缩中的应用.矢量量化通过将多个标量数据组合成矢量,并对整个矢量进行量化以实现数据压缩,同时尽量保持信息完整性.乘积量化则将输入空间划分为多个子空间,对每个子空间的权值矩阵进行聚类,生成包含多个码字的码书,通过少量码字近似表示相同维度的向量,从而显著缩减模型体积.实验表明,将乘积量化应用于卷积层和全连接层,不仅能够大幅减少模型体积,还能显著提升计算效率.例如,Asiyabi 等人^[37]提出了一种用于 SAR 卫星图像的随机量化噪声方法,通过自适应均匀量化格式在实现高效数据压缩的同时,保留了 SAR 原始数据的统计特性,为多种应用程序提供了高保真的数据基准.

(b) 参数剪枝和共享技术.剪枝技术通过移除模型中的冗余参数或网络连接,降低模型的复杂度和存储需求,根据剪枝粒度的不同,可以分为结构化剪枝和非结构化剪枝.结构化剪枝的操作粒度较大,通常删除整个卷积核或多个通道等参数组合,显著减小模型规模并提高硬件加速效果.然而,这种方法通常会引起一定的准确率降低.非结构化剪枝的操作粒度更小,可以灵活地移除任意冗余参数,不受特定结构的限制,可在保持模型精度的同时实现更高的稀疏化.然而,由于剪枝后模型结构不规则,硬件加速的实现较为困难.Han 等人^[38]提出了一种基于阈值选择的非结构化剪枝策略,通过设定阈值剔除权值较小的参数(如突触和神经元),在降低计算量的同时保持模型稀疏化效果.Lin 等人^[39]提出了一种基于剪枝的神经网络模式解释方法,通过修剪神经网络中的冗余参数以突出关键权重部分,从而提升了模型的可解释性,为卫星图像分析提供了优化新思路.

(c) 知识蒸馏技术.知识蒸馏通过“学生-教师”范式将复杂模型提炼为更紧凑的小模型,从而减少计算资源需求.其基本思想是通过 Softmax 函数学习教师模型的类别分布,并将大规模教师模型的知识传递给学生模型.这种方法显著减少了训练资源消耗,并在各种图像分类任务中取得了显著的性能提升.然而,知识蒸馏技术仅适用于具有 Softmax 损失函数的分类任务,限制了其应用范围.这些方法主要通过对模型的参数和结构来实现压缩,从而减少计算量,提升模型推理效率与性能.模型压缩本质是模型参数量、计算复杂度和推理性能之间的权衡.因此,在进行模型压缩时,需综合考虑其在特定应用场景下的实际需求,以避免性能损失和精度降低的过度影响.例如,Zhang 等人^[40]提出了一种两阶段检测网络,在精馏学习阶段引入自适应注意力机制,以捕捉教师网络与学生网络之间的差异特征,从而提升了卫星视频中目标跟踪的准确性,同时保持了较高的计算效率.Zhou 等人^[41]设计了一个多层语义知识对齐蒸馏框架,使学生模型能够学习复杂模型中提取的语义信息.该方法应用于遥感图像的密集预测,并通过知识蒸馏自适应地选择有用的语义信息,在目标分割任务中实现了出色的性能.

高性能星载计算平台发展现状:相比传统数据处理器,越来越多的硬件平台开始支持深度学习模型的快速部署与高效执行.这些平台不仅包括传统的 CPU 和 GPU,还涵盖了专为深度学习算法设计的硬件,如高性能 FPGA 等可编程逻辑加速器.同时,多个卫星项目的研究团队已成功完成高性能计算单元的在轨验证.例如,中国首颗商业 SAR 卫星“海丝一号”(代号 HISEA-1)搭载了英伟达 Jetson TX1 智能处理器,用于卫星在轨目标识别^[42].该卫星数据处理系统结合 GPU 执行复杂目标检测算法,CPU 完成图像分割等较低复杂度的任务,通过合理分配计算负荷,实现了高效目标识别.此外,由中国科学院主导的空间新技术试验卫星(代号 SATech-01)搭载异构多核智能处理单元,成功完成了深度学习算法的在轨验证^[43],该单元融合了多种智能处理单元架构,包括英伟达 Jetson TX2i、华为 Atlas 200 等,为在轨功能验证提供了丰富的计算资源.随着卫星逐渐小型化并降低成本,基于高性能 COTS 计算平台的应用预计将在各种更多低成本、短周期的空间任务中得到广泛推广.

(2) 卫星智能推理

当前深度学习模型对计算资源的高需求,给实际应用带来了巨大的挑战.现有研究主要聚焦于设计模型切分策略,通过优化能耗、通信或计算等代价函数,寻找满足用户或系统需求的分配方案.主流深度学习模型通常具有良好的结构特性,可采用纵向切分、横向切分及混合切分等方法,将切分后的子模型按照依赖关系部署至云端和边缘端.在纵向切分方面,DeepThings^[44]采用可伸缩的融合块分区切分 CNN,既最小化内存占用,又暴露并行性,同时通过创新的工作调度流程减少整体执行延迟.在横向切分方面,Neurosurgeon^[45]首次提出分层切分策略,有效

减轻了计算负担,并优化了端到端推理性能和能耗效率.在混合切分方面,SPINN^[46]则利用基于早退机制的分布式渐进推理系统,在高动态环境下实现了鲁棒的 CNN 推理.该系统采用早期退出策略和 CNN 分离的协同优化方法,有效降低了对设备与服务器通信的硬性要求.此外,Edgent^[47]提出了一个终端设备和边缘服务器协同的深度学习模型优化框架,通过离线训练分支网络 and 在线模型分割点搜索,有效平衡了模型准确率与推理需求.DeepSlicing^[48]则通过灵活的细粒度计划与高效调度方案,实现了模型与数据的协作切分,大幅缩短了计算和推理时间.然而,当前边缘云协同计算的研究在地面应用场景取得了显著进展,但针对星地协同网络边缘计算的研究却相对有限.卫星单节点资源受限,同时星地网络的通信时间窗口有限且带宽波动较大.因此,如何充分利用地面和卫星的特点,构建高效、灵活的深度学习模型推理框架,仍是一个亟待解决的关键性课题.

随着卫星通信技术和卫星边缘计算载荷算力的快速发展,将移动计算服务器部署在卫星节点上已成为研究热点.表 6 总结了近年来卫星边缘计算系统方法的比较,在轨边缘计算架构 (orbital edge computing, OEC)^[49]的提出突破弯管结构的限制.通过在卫星节点配置边缘计算能力,传感数据能够在轨处理,从而重构计算流程并优化任务分配.Kodan^[50]系统于 2023 年提出了一种面向单颗卫星的方案.该系统利用深度学习在轨部署实现了无效卫星图像的早期丢弃,显著减少了数据传输量.同时,Kodan 系统能够根据数据样本动态选择最优模型,以在计算资源受限的情况下最大化有价值数据的下行传输.与 Kodan 不同,Serval^[51]提出了针对固定神经网络模型的星地计算任务分离方法.通过利用卫星轨道的可预测性,Serval 能够有效降低高优先级请求的端到端时延.此外,北京邮电大学与天仪研究院联合发射了全球首个云原生卫星^[52],并基于此提出了一个面向地球观测任务的星地协同推理系统,如图 2 所示.该系统通过实时推理在轨图像数据,并在云层覆盖超过 50% 的情况下丢弃低质量的图像数据,从而有效缓解了星地传输压力.案例研究结果表明,该系统将推理精度提升了 50%,数据传输量减少了 90%.然而,该系统在下传数据过程中仍采用粗粒度处理,未充分考虑卫星能源消耗和星地通信的带宽等实际挑战,当前仅完成了在轨验证.在处理大型图像方面,现有方法多依赖缩放策略^[53]或将图像分块后分别使用 DNN 模型推理^[54].尽管这些研究能够提升图像推理性能,但可靠性不足且未考虑计算开销.在太空环境中,受限的能耗资源是计算的主要瓶颈.为此,TargetFuse^[55]提出了一种综合能耗资源和计算资源限制的端到端星地协同推理框架,旨在进一步提升模型推理精度和效率.

表 6 卫星边缘计算系统方法比较

已有工作	计算场景	核心理念
OEC ^[49]	星地	借助地面站位置和轨道特点,降低数据传输量
Kodan ^[50]	单星	在轨过滤无效数据,降低数据传输量
Serval ^[51]	星地	数据优先级排序,降低数据传输时延
Tiansuan ^[52]	星地	借助地面站模型,提升在轨模型的推理精度
TargetFuse ^[55]	星地	考虑卫星能耗和计算的限制,优化星地上传效率,提升推理精度

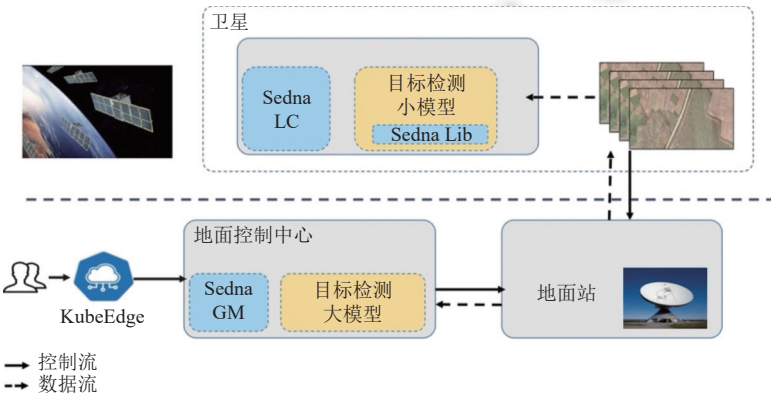


图 2 基于天算星座的首个星地协同推理系统示意图

(3) 卫星联邦学习

集中式机器学习方法在卫星边缘计算中的应用虽然受到广泛关注, 但存在诸多局限性. 首先, 集中式方法依赖参数服务器收集训练样本, 导致大量数据传输并引发严重的通信延迟问题. 其次, 传统集中式方法难以有效保障设备的数据隐私安全. 为了解决这些问题, 研究人员提出了分布式学习算法, 通过本地训练实现设备间的协同建模. 在此背景下, 基于星地联邦学习系统应运而生. 该系统将联邦学习这一协作式机器学习范式引入低轨卫星星座与地面站的协同网络. 如图 3 所示, 其具体设计如下: 将卫星作为端侧节点, 地面站作为服务器; 在参数服务器架构下, 卫星与地面站协同完成机器学习模型的训练. 基于星地的联邦学习系统为低轨卫星边缘计算提供了一种高效、隐私保护的解决方案, 该系统不仅支持大规模星座协同, 还能够有效降低通信成本并保障卫星数据的隐私安全.

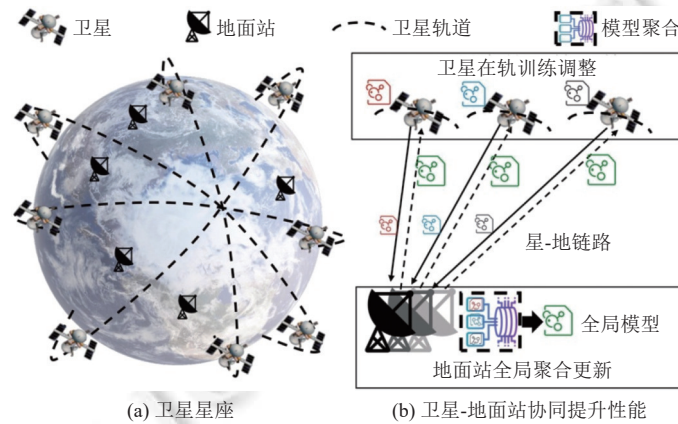


图 3 星地联邦学习示意图

此前研究主要致力于缓解卫星异构性带来的挑战, 尤其是同步或异步联邦学习算法中可能存在的滞后与过时问题. 例如, So 等人^[56]通过设置自适应聚合缓存维护聚合参数的新鲜度, 从而提升了系统精度. 针对地面与卫星之间带宽有限导致的协同训练低下问题, 已有研究提出了自适应感知带宽变化的方法, 动态选择对模型有益的参数进行传输, 以提升模型收敛速度和协同训练效率^[57-61]. 为降低通信成本, 现有神经量化方法主要聚焦于减少冗余梯度信息的传输, 例如使用 int-8 或 int-4 替代默认的 float-32 精度. Yang 等人^[58]提出了精度优先的联邦量化训练算法, 有效支持地星动态连接场景下的在轨卫星训练. 此外, 针对单个卫星在资源受限、样本不足导致的学习效率低下, 以及低轨卫星网络的时空约束带来的协同训练困难问题, Zhang 等人^[61]提出了一种分布式主导的联邦学习体系结构. 该体系通过星间链路评估动态星座的连通性和负载, 选出主导卫星和追随卫星, 实现深度学习模型全局和局部参数的连续传输与共享, 从而在无需依赖地面站的条件下完成模型高效训练. 为进一步优化资源分配, 该体系引入基于深度强化学习的自适应卫星资源分配方法, 通过分布式近端策略优化动态调控卫星计算能力和发射功率, 不仅加速自适应传输, 还有效降低了系统能耗.

3.4 任务智能调度与规划

卫星任务调度与规划问题是当前研究的热点, 旨在优化网络和系统资源的有效利用, 提升任务执行的整体性能. 传统卫星任务通常由地面指挥中心根据卫星在轨运行状态事先制定计划, 并通过星地链路将任务分配至卫星节点. 卫星到达目标区域时, 依据接收的任务信息自主执行. 然而, 随着卫星种类与数量的激增, 传统的卫星任务分配方法难以满足复杂卫星网络环境的需求^[62,63]. 在卫星边缘计算网络架构下, 计算资源分布于用户终端、卫星边缘计算节点、地面站数据中心等多层平台. 为了满足用户对服务质量 (QoE) 的多样化需求 (如时延、隐私、能耗等), 亟需设计动态灵活的多层任务调度机制. 当单颗卫星计算节点资源受限时, 可通过邻近卫星的协作分担任务, 从而提升整体计算性能和资源利用率. 图 4 展示了卫星任务调度的主要框架, 主要包括分层协调任务调度和多星协调任务调度.

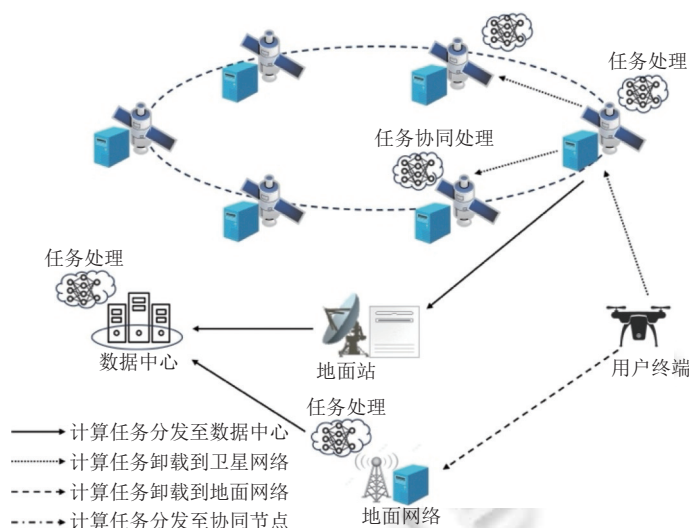


图4 卫星任务调度与规划示意图

(1) 分层协同任务调度

针对不同计算需求(如对时延、吞吐量、能耗和带宽等),需综合考虑用户终端、卫星边缘计算节点与地面站数据中心的资源特性,优化多层协同任务调度.该过程通常包含以下4个阶段:(a)终端用户与连接的卫星边缘计算节点交互,获取网络状态信息(如链路状态信息、数据队列信息)并提交任务请求;(b)卫星边缘计算节点根据其剩余计算资源和队列情况反馈处理能力;(c)终端用户依据自身计算需求(如低时延)评估不同任务分配调度方案的性能,选取最优方案;(d)其他计算节点(卫星边缘计算节点或地面数据中心)处理任务并返回结果,最终由终端用户汇总完成任务.

目前的卫星在轨任务规划模型包括约束规划模型、图模型、背包模型、整数线性规划模型和状态-动作模型等^[64]。由于卫星调度属于 NP-hard 问题, 小规模问题可通过精确策略方法求解。例如, Cao 等人^[65]针对低轨卫星场景, 提出了一种将物联网设备任务卸载至基站或低轨卫星的方案, 从而最大化节能。Zhu 等人^[66]构建了一个 3 层计算框架 (物联网设备-基站-中心云), 通过卫星将任务卸载至云端处理以降低总延迟。庞中华等人^[67]基于高低轨卫星协同管控模式, 建立任务分配模型并设计了求解算法。Li 等人^[68]通过启发式算法优化地观测收益与卫星能耗间的平衡。Zhang 等人^[69]设计了卫星对等卸载方案, 提高服务质量和资源利用率。Chen 等人^[70]提出了多层混合并行计算架构, 集成本地、边缘和云计算, 提升任务处理性能。然而, 随着问题规模扩大, 传统静态或预定义优化方法难以适应复杂动态场景。因此, 研究者逐渐引入智能优化方法, 例如禁忌搜索算法与深度强化学习等方法^[71, 72]。Cheng 等人^[73]提出了一种天空地一体化网络, 旨在最小化总延迟。Qin 等人^[74]设计了一个 3 层卸载架构 (物联网设备-卫星-中心云), 通过联合优化传输功率分配和卫星边缘计算资源, 最大限度地降低了延迟和能耗。这些方法虽显著提升了调度效率, 但也因动作空间复杂度的增加而面临新的挑战。

(2) 多星协同任务调度

单颗卫星的资源通常有限,不同卫星间的资源特性也存在差异.通过多星协同调度可高效利用资源,优化任务处理性能.在低轨卫星网络场景中,当单卫星节点资源不足时,多星协同调度能够实现任务的分布式分配与全局优化.其调度过程一般包含以下3个阶段.(a)本地卫星边缘节点接收任务请求,并与相邻卫星节点交互本地信息(如任务队列长度、执行时间和能耗成本等).(b)本地节点根据资源情况判断是否满足任务需求:若资源足够则本地处理;否则请求相邻卫星节点协同计算.相邻卫星节点根据计算能力与资源状态判断是否参与协同计算,并反馈状态信息.若整个卫星网络无法满足需求,则任务被卸载到地面数据中心进行远程处理.(c)相邻卫星边缘节点完成任务后将结果反馈到本地节点,最终完成任务.

研究者提出了多种多星协同调度策略. 例如, Kim 等人^[75]设计了基于遗传算法的星座级任务分配模型, 目标是最小化系统响应时间. Cui 等人^[76]针对分布式卫星任务调度分配问题, 提出了基于任务优先级的动态调度模型, 通过混合遗传搜索算法优化任务等待时间与收益. Cesta 等人^[77]提出了面向目标的任务执行及状态更新算法, 解决分布式多智能体的任务调度问题. Grey 等人^[78]设计了支持复杂的卫星任务分配与协商的智能化任务分配架构. 王海蛟^[79]基于马尔可夫链与多智能体强化学习, 提出了在线调度模型, 在低通信成本的在线调度模型, 实现了多星任务协同. 此外, 得益于高中低卫星系统的互补特性, 星座级协同也为卫星任务分配与调度提供了新思路. 低轨卫星在业务速率和时延等方法具有优势. 而高中低轨系统则在覆盖、速率等方面优势互补, 可进一步提升系统的多样化服务能力.

此外, 各国在卫星在轨任务规划系统方面取得了显著成果. 例如, NASA 开发的任务规划调度系统已应用于“深空一号”探测器上. 此外, NASA 构建了一套可持续的任务重新规划系统, 采用迭代修复技术, 实现了任务的实时更新与任务调整, 已成功应用于在对地观测 (earth observing-1, EO-1) 卫星以及高光谱红外成像 (hyperspectral infrared imager, HypsIRI)^[80]卫星. 在“毅力号”探测器的 Mars 2020 探测计划中, NASA 实现了任务调度和分配的可视化与解释分析^[81].

4 未来研究展望

卫星边缘计算作为应对卫星数据爆炸式增长并提升响应速度的重要技术手段, 正在推动卫星服务能力的变革. 然而, 当前领域仍存在诸多技术瓶颈与开放性研究问题. 如表 7 所示, 未来的主要研究方向与挑战主要包括智能处理算法、计算可靠性、多源数据在轨融合、卫星边缘云安全防护等, 并且分析了各个研究方向的现状和未来的目标.

表 7 未来研究方向的当前现状和目标

研究方向	当前现状	短期目标	中长期目标
智能处理算法	目标场景与模型训练场景存在较大差异, 算法适应性较弱; 小目标检测精度较低, 易产生误报	提升在轨智能处理算法的鲁棒性和精度; 开发去噪与图像增强技术; 优化小目标检测算法, 提高实时处理能力	研发具备强泛化能力的星载深度学习模型; 开发遥感大模型并优化训练策略, 提升复杂场景下的适应性
计算可靠性	COTS 硬件在太空环境中受能源、散热和辐射影响, 计算能力易受限; 计算效率受热管理影响	通过硬件设计与软件容错机制结合, 提升 COTS 器件抗辐射能力; 建立多维环境影响模型	设计自适应动态管理机制, 实现计算资源的智能调度与优化, 提高长期运行可靠性
多源数据在轨融合	现有方法难以充分挖掘多种传感器数据的互补信息; 数据融合计算成本高, 星载环境下难以部署	研发基于多模态的多源数据融合算法, 提升信息提取精度; 优化数据回传策略, 减少带宽占用	设计轻量级、低功耗的在轨融合计算方案; 推进智能化决策支持
卫星安全防护	传统云安全机制难以适用于卫星环境; 通信链路易受干扰, 数据窃取与攻击风险高	采用多层次安全防护措施, 以提升系统安全性	研究基于 SDN/NFV 的动态安全防护平台, 实现微隔离技术与自动化安全策略, 提高系统安全性

4.1 智能处理算法

当前智能处理算法在目标场景与模型训练场景存在较大差异, 适应性较弱, 尤其在受传感器噪声、大气干扰以及低质量遥感图像时, 处理能力受到限制. 此外, 现有检测识别算法在小目标检测识别方面仍有较大的提升空间. 例如, 在遥感图像中小目标检测 (如车辆或建筑) 中, 检测精度较低, 且在相似目标场景下产生误报. 为此, 未来研究应重点关注提升在轨智能处理算法的鲁棒性和精度. 具体而言, 需要开发具备强泛化能力和鲁棒性的机器学习和深度学习算法, 以适应复杂的星载环境. 在此基础上, 可引入去噪与图像增强技术, 以优化低质量图像的处理效果. 此外, 针对小目标检测问题, 应研发多尺度特征提取与增强模型, 以提升识别精度和实时处理能力. 同时, 在计算资源有限的星载环境中, 需优化算法的计算效率, 例如, 开发适用于星载环境的遥感大模型, 并结合高效的训练与微调策略, 实现更快速、更精确的智能数据处理.

4.2 计算可靠性

高性能 COTS 硬件的引入显著提升了卫星平台的计算能力, 但同时也受限于复杂的太空环境. 特别是在能源

供应和散热管理方面, 受卫星体积和重量的限制, 能源收集和存储能力有限, 而真空环境无法采用传统对流散热方式, 导致散热效率低下。COTS 器件在过热时会自动降频以减少热量产生, 从而维持安全温度范围, 但这也导致了计算效率显著降低。此外, 空间和内部环境受太空环境、卫星在轨位置变化及系统运行状态的动态耦合作用, 使其整体运行特性复杂多变。仅依赖单一抗干扰环境措施 (如元器件抗辐射能力) 或基于特定时间段的静态评估, 难以真实反映其在轨运行环境。因此, 未来研究应重点提升 COTS 器件的抗辐射能力, 并结合硬件设计与软件容错机制, 提高系统整体的可靠性; 同时, 需要构建涵盖空间环境与内部因素的多维度环境影响模型, 并设计自适应动态管理机制, 以应对复杂环境条件的变化。

4.3 多源数据在轨融合

多源数据融合是卫星边缘计算的重要研究方向之一, 旨在利用不同传感器 (如可见光、高光谱、SAR 等) 之间的互补信息, 提高目标与地物的精细化信息提取能力。在遥感领域, 不同类型传感器在分辨率、光谱信息和时间覆盖等方面各具特色。通过多源信息融合技术, 可显著提升数据利用效率, 同时减少数据回传量, 从而实现更高效的信息提取与决策支持。未来研究应重点开发基于多模态深度学习的多源数据融合算法, 深入挖掘不同传感器数据之间的互补信息, 以实现更精准的目标识别与智能决策支持。同时, 为适应星载环境的计算和能耗限制, 需要设计轻量级、低功耗的在轨融合计算方案。这些方案不仅能够提升遥感数据的处理效率, 还可广泛应用于灾害监测、资源管理等场景, 进一步提高卫星边缘计算的实际应用价值。

4.4 卫星边缘云安全防护

卫星边缘计算的实时性与复杂性, 以及星载资源的受限, 使得传统地面云中心的安全机制难以直接适用于卫星环境。卫星数据在传输过程中可能面临数据窃取、DDoS 攻击等安全威胁。特别是在跨越大气层的通信过程中, 链路极易受到干扰; 此外, 受限的星载计算和存储能力也对卫星数据的安全存储、备份以及处理效率提出了更高要求。为应对这些挑战, 卫星边缘云需要采用多层次安全防护措施, 包括接入认证技术、恶意攻击防范技术以及链路传输加密技术, 以提升系统整体安全性。与此同时, 随着网络功能的虚拟化技术的发展, 安全管理的复杂性进一步增加。因此, 未来研究应聚焦于基于 SDN/NFV 的动态安全防护平台, 结合微隔离技术与自动化安全策略, 以加强卫星载荷的安全防护能力, 以确保整体系统的可靠性与数据安全性。

5 总 结

近年来, 卫星边缘计算因其在拓展传统卫星功能、提高计算能力以及优化通信性能方面的巨大潜力, 已成为学术界和产业界的研究热点。然而, 其发展仍面临诸多挑战, 包括资源受限、网络协同复杂性、健壮性与可靠性不足以及架构融合等技术难题。本文系统梳理了卫星边缘计算智能化技术的研究现状与最新进展, 重点讨论了以下关键领域: 云原生卫星平台、智能化操作系统、智能化在轨处理以及任务智能调度与规划。这些领域的研究不仅为现阶段卫星边缘计算提供了理论与技术支持, 同时也暴露出许多亟待解决的关键问题。从根本上看, 卫星边缘计算的核心目标在于推动卫星及其地面基础设施从传统硬件驱动向服务驱动全面转型。这种转变不仅要求部署低成本、高效能的计算设备, 还需要充分利用以太阳光能为主的清洁能源, 从而实现可持续的全天候服务模式。展望未来, 卫星边缘计算将在资源优化、智能处理等方面释放更大的潜能, 为构建更加智能化、可持续的太空计算生态奠定坚实基础。

References

- [1] World Meteorological Organization. Weather-related disasters increase over past 50 years, causing more damage but fewer deaths. 2021. <https://wmo.int/media/news/weather-related-disasters-increase-over-past-50-years-causing-more-damage-fewer-deaths>
- [2] Wang SG, Li Q. Satellite computing: Vision and challenges. IEEE Internet of Things Journal, 2023, 10(24): 22514–22529. [doi: 10.1109/JIOT.2023.3303346]
- [3] Earth Observation Center. 60 petabytes for the German satellite data archive D-SDA. 2018. <https://www.dlr.de/en/eoc/latest/news/2018/60-petabytes-for-the-german-satellite-data-archive-d-sda>

- [4] Lucia B, Denby B, Manchester Z, Desai H, Ruppel E, Colin A. Computational nanosatellite constellations: Opportunities and challenges. *GetMobile: Mobile Computing and Communications*, 2021, 25(1): 16–23. [doi: [10.1145/3471440.3471446](https://doi.org/10.1145/3471440.3471446)]
- [5] Xu MW, Fu Z, Ma X, Zhang L, Li YN, Qian F, Wang SG, Li K, Yang JY, Liu XZ. From cloud to edge: A first look at public edge platforms. In: *Proc. of the 21st ACM Internet Measurement Conf.* ACM, 2021. 37–53. [doi: [10.1145/3487552.3487815](https://doi.org/10.1145/3487552.3487815)]
- [6] Bhattacharjee D, Kassing S, Licciardello M, Singla A. In-orbit computing: An outlandish thought experiment? In: *Proc. of the 19th ACM Workshop on Hot Topics in Networks*. ACM, 2020. 197–204. [doi: [10.1145/3422604.3425937](https://doi.org/10.1145/3422604.3425937)]
- [7] Xing RL, Xu MW, Zhou A, Li Q, Zhang YR, Qian F, Wang SG. Deciphering the enigma of satellite computing with COTS devices: Measurement and analysis. In: *Proc. of the 30th Annual Int'l Conf. on Mobile Computing and Networking*. Washington: ACM, 2024. 420–435. [doi: [10.1145/3636534.3649371](https://doi.org/10.1145/3636534.3649371)]
- [8] Kassem MM, Raman A, Perino D, Sastry N. A browser-side view of Starlink connectivity. In: *Proc. of the 22nd ACM Internet Measurement Conf.* Nice: ACM, 2022. 151–158. [doi: [10.1145/3517745.3561457](https://doi.org/10.1145/3517745.3561457)]
- [9] Michel F, Trevisan M, Giordano D, Bonaventure O. A first look at Starlink performance. In: *Proc. of the 22nd ACM Internet Measurement Conf.* Nice: ACM, 2022. 130–136. [doi: [10.1145/3517745.3561416](https://doi.org/10.1145/3517745.3561416)]
- [10] Ma S, Chou Y C, Zhao HY, Chen L, Ma XQ, Liu JC. Network characteristics of LEO satellite constellations: A Starlink-based measurement from end users. In: *Proc. of the 2023 IEEE Annual Joint Conf.: INFOCOM, IEEE Computer and Communications Societies*. New York: IEEE, 2023. 1–10. [doi: [10.1109/INFOCOM53939.2023.10228912](https://doi.org/10.1109/INFOCOM53939.2023.10228912)]
- [11] Ye B, Mo LH, Liu T, Sun YM, Liu J. Influence of orbital parameters on SEU rate of low-energy proton in nano-SRAM device. *Symmetry*, 2020, 12(12): 2030. [doi: [10.3390/sym12122030](https://doi.org/10.3390/sym12122030)]
- [12] Wang SG, Li Q, Xu MW, Ma X, Zhou A, Sun QB. Tiansuan constellation: An open research platform. In: *Proc. of the 2021 IEEE Int'l Conf. on Edge Computing*. Chicago: IEEE, 2021. 94–101. [doi: [10.1109/EDGE53862.2021.00022](https://doi.org/10.1109/EDGE53862.2021.00022)]
- [13] Panagopoulos AD, Arapoglou PDM, Cottis PG. Satellite communications at Ku, Ka, and V bands: Propagation impairments and mitigation techniques. *IEEE Communications Surveys & Tutorials*, 2004, 6(3): 2–14. [doi: [10.1109/COMST.2004.5342290](https://doi.org/10.1109/COMST.2004.5342290)]
- [14] Jiujiu. Given its downlink rate exceeding 100 Mbps, is Starlink a viable alternative to 5G? 2022 (in Chinese). <https://www.51cto.com/article/704582.html>
- [15] Terrasanta G, Ziarko MW, Bergamasco N, Poot M, Poliak J. Simulating optical single event transients on silicon photonic waveguides for satellite communication. *IEEE Trans. on Nuclear Science*, 2024, 71(2): 176–183. [doi: [10.1109/TNS.2024.3353489](https://doi.org/10.1109/TNS.2024.3353489)]
- [16] Landauer DC, Lovelley TM. Performance evaluation of the radiation-tolerant NVIDIA Tegra K1 system-on-chip. In: *Proc. of the 2023 IEEE Space Computing Conf.* Pasadena: IEEE, 2023. 24–33. [doi: [10.1109/SCC57168.2023.00014](https://doi.org/10.1109/SCC57168.2023.00014)]
- [17] Li ZQ, Jing XY, Zhu XK, Zhang HY. Heterogeneous defect prediction through multiple kernel learning and ensemble learning. In: *Proc. of the 2017 IEEE Int'l Conf. on Software Maintenance and Evolution*. Shanghai: IEEE, 2017. 91–102. [doi: [10.1109/ICSME.2017.19](https://doi.org/10.1109/ICSME.2017.19)]
- [18] El-Hajjar M, Hanzo L. A survey of digital television broadcast transmission techniques. *IEEE Communications Surveys & Tutorials*, 2013, 15(4): 1924–1949. [doi: [10.1109/SURV.2013.030713.00220](https://doi.org/10.1109/SURV.2013.030713.00220)]
- [19] Wang F, Jiang DD, Wang ZH, Chen JG, Quek TQS. Seamless handover in LEO based non-terrestrial networks: Service continuity and optimization. *IEEE Trans. on Communications*, 2023, 71(2): 1008–1023. [doi: [10.1109/TCOMM.2022.3229014](https://doi.org/10.1109/TCOMM.2022.3229014)]
- [20] Zhang QY, Che XY, Chen YJ, Ma X, Xu MW, Dustdar S, Liu XZ, Wang SG. A comprehensive deep learning library benchmark and optimal library selection. *IEEE Trans. on Mobile Computing*, 2024, 23(5): 5069–5082. [doi: [10.1109/TMC.2023.3301973](https://doi.org/10.1109/TMC.2023.3301973)]
- [21] Huang CY, Ye P, Chen T, He T, Yue XY, Ouyang WL. EMR-merging: Tuning-free high-performance model merging. In: *Proc. of the 38th Annual Conf. on Neural Int'l Processing Systems*. Vancouver: NeurIPS, 2024. 122741–122769.
- [22] Albashish D, Al-Sayyed R, Abdullah A, Ryalat MH, Almansour NA. Deep CNN model based on VGG16 for breast cancer classification. In: *Proc. of the 2021 Int'l Conf. on Information Technology*. Amman: IEEE, 2021. 805–810. [doi: [10.1109/ICIT52682.2021.9491631](https://doi.org/10.1109/ICIT52682.2021.9491631)]
- [23] Wu JX, Leng C, Wang YH, Hu QH, Cheng J. Quantized convolutional neural networks for mobile devices. In: *Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016. 4820–4828. [doi: [10.1109/CVPR.2016.521](https://doi.org/10.1109/CVPR.2016.521)]
- [24] Jacob B, Kligys S, Chen B, Zhu ML, Tang M, Howard A, Adam H, Kalenichenko D. Quantization and training of neural networks for efficient integer-arithmetic-only inference. In: *Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 2704–2713. [doi: [10.1109/CVPR.2018.00286](https://doi.org/10.1109/CVPR.2018.00286)]
- [25] Bharati P, Pramanik A. Deep learning techniques—R-CNN to mask R-CNN: A survey. In: Das AK, Nayak J, Naik B, eds. *Computational Intelligence in Pattern Recognition*. Singapore: Springer, 2020. 657–668. [doi: [10.1007/978-981-13-9042-5_56](https://doi.org/10.1007/978-981-13-9042-5_56)]
- [26] Girshick R, Jeff D, Trevor D, Jitendra M. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proc. of the 2014 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*. 2014. 580–587.
- [27] Yu W, Yang KY, Bai YL, Xiao TJ, Yao HX, Rui Y. Visualizing and comparing AlexNet and VGG using deconvolutional layers. In:

- Proc. of the 33rd Int'l Conf. on Machine Learning. New York: JMLR, 2016.
- [28] Lin M, Chen C, Lai C. Object detection algorithm based AdaBoost residual correction Fast R-CNN on network. In: Proc. of the 3rd Int'l Conf. on Deep Learning Technologies. 2019. 42–46.
 - [29] Fratini S, Policella N, Silva R, Guerreiro J. On-board autonomy operations for OPS-SAT experiment. *Applied Intelligence*, 2022, 52(6): 6970–6987. [doi: [10.1007/s10489-020-02158-5](https://doi.org/10.1007/s10489-020-02158-5)]
 - [30] Verfaillie G, Charneau MC. A generic modular architecture for the control of an autonomous spacecraft. In: Proc. of the 5th Int'l Workshop on Planning and Scheduling for Space. Baltimore: Space Telescope Science Institute, 2006.
 - [31] Van Etten A. You only look twice: Rapid multi-scale object detection in satellite imagery. arXiv:1805.09512, 2018.
 - [32] Verfaillie G, Lemaitre M. Tutorial on planning activities for earth watching and observation satellites and constellations: From off-line ground planning to on-line on-board planning. 2006. <https://icaps06.icapsconference.org/preprints/i06-tu5-allpapers.pdf>
 - [33] Bensana E, Verfaillie G, Agnese JC, Bataille N, Blumstein D. Exact and inexact methods for the daily management of an earth observation satellite. In: Proc. of the 1996 Int'l Symp. on Space Mission Operations and Ground Data Systems. 1996. 507–514.
 - [34] Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50× fewer parameters and <0.5 MB model size. In: Proc. of the 5th Int'l Conf. on Learning Representations. Toulon: ICLR, 2017. 1–13.
 - [35] Zhang XY, Zhou XY, Lin MX, Sun J. ShuffleNet: An extremely efficient convolutional neural network for mobile devices. In: Proc. of the 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 6848–6856. [doi: [10.1109/CVPR.2018.00716](https://doi.org/10.1109/CVPR.2018.00716)]
 - [36] Yue Q, Xu XL, Wang YX, Tao YK, Luo XLY. Routing-guided learned product quantization for graph-based approximate nearest neighbor search. In: Proc. of the 40th IEEE Int'l Conf. on Data Engineering. Utrecht: IEEE, 2024. 4870–4883. [doi: [10.1109/ICDE60146.2024.00370](https://doi.org/10.1109/ICDE60146.2024.00370)]
 - [37] Asiyabi RM, Anghel A, Focsa A, Datcu M, Martone M, Rizzoli P, Imbembo E. Adaptation of decoded Sentinel-1 SAR raw data for the assessment of novel data compression methods. In: Proc. of the 2024 IEEE Int'l Geoscience and Remote Sensing Symp. Athens: IEEE, 2024. 2541–2545. [doi: [10.1109/IGARSS53475.2024.10642269](https://doi.org/10.1109/IGARSS53475.2024.10642269)]
 - [38] Han S, Mao HZ, Dally WJ. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. arXiv:1510.00149, 2015.
 - [39] Lin HP, Yin JJ, Yang J, Xu F. Interpreting neural network pattern with pruning for PolSAR target recognition. *IEEE Trans. on Geoscience and Remote Sensing*, 2024, 62: 5227114. [doi: [10.1109/TGRS.2024.3493415](https://doi.org/10.1109/TGRS.2024.3493415)]
 - [40] Zhang WH, Deng WJ, Cui Z, Liu J, Jiao LC. Object knowledge distillation for joint detection and tracking in satellite videos. *IEEE Trans. on Geoscience and Remote Sensing*, 2024, 62: 4701213. [doi: [10.1109/TGRS.2024.3355933](https://doi.org/10.1109/TGRS.2024.3355933)]
 - [41] Zhou WJ, Li YZ, Huan J, Liu YY, Jiang QP. MSTNet-KD: Multilevel transfer networks using knowledge distillation for the dense prediction of remote-sensing images. *IEEE Trans. on Geoscience and Remote Sensing*, 2024, 62: 4504612. [doi: [10.1109/TGRS.2024.3384669](https://doi.org/10.1109/TGRS.2024.3384669)]
 - [42] Krebs GD. Haisi 1. 2025. https://space.skyrocket.de/doc_sdat/haisi-1.htm
 - [43] Krebs GD. SATech 01 (Kongjian Xinjishu Shiyan 01). 2025. https://space.skyrocket.de/doc_sdat/satech-01.htm
 - [44] Zhao ZR, Barijough KM, Gerstlauer A. DeepThings: Distributed adaptive deep learning inference on resource-constrained IoT edge clusters. *IEEE Trans. on Computer-aided Design of Integrated Circuits and Systems*, 2018, 37(11): 2348–2359. [doi: [10.1109/TCAD.2018.2858384](https://doi.org/10.1109/TCAD.2018.2858384)]
 - [45] Kang YP, Hauswald J, Gao C, Rovinski A, Mudge T, Mars J, Tang LJ. Neurosurgeon: Collaborative intelligence between the cloud and mobile edge. *ACM Sigarch Computer Architecture News*, 2017, 45(1): 615–629. [doi: [10.1145/3093337.3037698](https://doi.org/10.1145/3093337.3037698)]
 - [46] Laskaridis S, Venieris S I, Almeida M, Leontiadis I, Lane ND. SPINN: Synergistic progressive inference of neural networks over device and cloud. In: Proc. of the 26th Annual Int'l Conf. on Mobile Computing and Networking. London: ACM, 2020. 37. [doi: [10.1145/3372224.3419194](https://doi.org/10.1145/3372224.3419194)]
 - [47] Li E, Zeng LK, Zhou Z, Chen X. Edge AI: On-demand accelerating deep neural network inference via edge computing. *IEEE Trans. on Wireless Communications*, 2020, 19(1): 447–457. [doi: [10.1109/TWC.2019.2946140](https://doi.org/10.1109/TWC.2019.2946140)]
 - [48] Zhang S, Zhang S, Qian ZZ, Wu J, Jin YB, Lu SL. DeepSlicing: Collaborative and adaptive CNN inference with low latency. *IEEE Trans. on Parallel and Distributed Systems*, 2021, 32(9): 2175–2187. [doi: [10.1109/TPDS.2021.3058532](https://doi.org/10.1109/TPDS.2021.3058532)]
 - [49] Denby B, Lucia B. Orbital edge computing: Nanosatellite constellations as a new class of computer system. In: Proc. of the 25th Int'l Conf. on Architectural Support for Programming Languages and Operating Systems. Lausanne: ACM, 2020. 939–954. [doi: [10.1145/3373376.3378473](https://doi.org/10.1145/3373376.3378473)]
 - [50] Denby B, Chintalapudi K, Chandra R, Lucia B, Noghabi S. Kodan: Addressing the computational bottleneck in space. In: Proc. of the

- 28th ACM Int'l Conf. on Architectural Support for Programming Languages and Operating Systems, Vol. 3. Vancouver: ACM, 2023. 392–403. [doi: [10.1145/3582016.3582043](https://doi.org/10.1145/3582016.3582043)]
- [51] Tao B, Chabra O, Janveja I, Gupta I, Vasisht D. Known knowns and unknowns: Near-realtime earth observation via query bifurcation in serval. In: Proc. of the 21st USENIX Symp. on Networked Systems Design and Implementation. Santa: USENIX, 2024. 809–824.
 - [52] Wang SG, Zhang QY, Xing RL, Qi F, Xu MW. The first verification test of space-ground collaborative intelligence via cloud-native satellites. *China Communications*, 2024, 21(4): 208–217. [doi: [10.23919/JCC.fa.2022-0422.202404](https://doi.org/10.23919/JCC.fa.2022-0422.202404)]
 - [53] Ji JH, Zhong BJ, Wu QH, Ma KK. A channel-wise multi-scale network for single image super-resolution. *IEEE Signal Processing Letters*, 2024, 31: 805–809. [doi: [10.1109/LSP.2024.3372781](https://doi.org/10.1109/LSP.2024.3372781)]
 - [54] Wang XD, Li SF, Kallidromitis K, Kato Y, Kozuka K, Darrell T. Hierarchical open-vocabulary universal image segmentation. In: Proc. of the 37th Int'l Conf. on Neural Information Processing Systems. New Orleans: Curran Associates Inc., 2023. 936. [doi: [10.5555/3666122.3667058](https://doi.org/10.5555/3666122.3667058)]
 - [55] Zhang QY, Yuan X, Xing RL, Zhang YR, Zheng ZM, Ma X, Xu MW, Dustdar S, Wang SG. Resource-efficient in-orbit detection of earth objects. In: Proc. of the 2024 IEEE Conf. on Computer Communications. Vancouver: IEEE, 2024. 551–560. [doi: [10.1109/INFOCOM52122.2024](https://doi.org/10.1109/INFOCOM52122.2024)]
 - [56] So J, Hsieh K, Arzani B, Noghabi S, Avestimehr S, Chandra R. FedSpace: An efficient federated learning framework at satellites and ground stations. *arXiv:2202.01267*, 2022.
 - [57] Nguyen J, Malik K, Zhan HY, Yousefpour A, Rabbat M, Malek M, Huba D. Federated learning with buffered asynchronous aggregation. In: Proc. of the 25th Int'l Conf. on Artificial Intelligence and Statistics. Valencia: PMLR, 2022. 3581–3607.
 - [58] Yang C, Yuan JL, Wu YZ, Sun QB, Zhou A, Wang SG, Xu MW. Communication-efficient satellite-ground federated learning through progressive weight quantization. *IEEE Trans. on Mobile Computing*, 2024, 23(9): 8999–9011. [doi: [10.1109/TMC.2024.3358804](https://doi.org/10.1109/TMC.2024.3358804)]
 - [59] Razmi N, Matthiesen B, Dekorsy A, Popovski P. Ground-assisted federated learning in LEO satellite constellations. *IEEE Wireless Communications Letters*, 2022, 11(4): 717–721. [doi: [10.1109/LWC.2022.3141120](https://doi.org/10.1109/LWC.2022.3141120)]
 - [60] Tang FX, Wen C, Chen XH, Kato N. Federated learning for intelligent transmission with space-air-ground integrated network toward 6G. *IEEE Network*, 2023, 37(2): 198–204. [doi: [10.1109/MNET.104.2100615](https://doi.org/10.1109/MNET.104.2100615)]
 - [61] Zhang HY, Zhao HB, Liu RK, Gao XQ, Xu SZ. Leader federated learning optimization using deep reinforcement learning for distributed satellite edge intelligence. *IEEE Trans. on Services Computing*, 2024, 17(5): 2544–2557. [doi: [10.1109/TSC.2024.3376256](https://doi.org/10.1109/TSC.2024.3376256)]
 - [62] Xiang S, Chen YG, Li GL, Xing LN. Review on satellite autonomous and collaborative task scheduling planning. *Acta Automatica Sinica*, 2019, 45(2): 252–264 (in Chinese with English abstract). [doi: [10.16383/j.aas.c180068](https://doi.org/10.16383/j.aas.c180068)]
 - [63] Zhao P, Chen ZM. An adapted genetic algorithm applied to satellite autonomous task scheduling. *Chinese Space Science and Technology*, 2016, 36(6): 47–54 (in Chinese with English abstract). [doi: [10.16708/j.cnki.1000-758X.2016.0064](https://doi.org/10.16708/j.cnki.1000-758X.2016.0064)]
 - [64] Damiani S, Verfaillie G, Charneau MC. An earth watching satellite constellation: How to manage a team of watching agents with limited communications. In: Proc. of the 4th Int'l Joint Conf. on Autonomous Agents and Multiagent Systems. New York: ACM, 2005. 455–462. [doi: [10.1145/1082473.1082543](https://doi.org/10.1145/1082473.1082543)]
 - [65] Cao XL, Yang B, Shen YL, Yuen C, Zhang Y, Han Z, Poor HV, Hanzo L. Edge-assisted multi-layer offloading optimization of LEO satellite-terrestrial integrated networks. *IEEE Journal on Selected Areas in Communications*, 2023, 41(2): 381–398. [doi: [10.1109/JSAC.2022.3227032](https://doi.org/10.1109/JSAC.2022.3227032)]
 - [66] Zhu XM, Jiang CX. Delay optimization for cooperative multi-tier computing in integrated satellite-terrestrial networks. *IEEE Journal on Selected Areas in Communications*, 2023, 41(2): 366–380. [doi: [10.1109/JSAC.2022.3227083](https://doi.org/10.1109/JSAC.2022.3227083)]
 - [67] Pang ZH. Research on collaborative mission planning method for high and low earth observation satellites [MS. Thesis]. Harbin: Harbin Institute of Technology, 2013 (in Chinese).
 - [68] Li JT, Zhang S, Liu XL, He RJ. Multi-objective evolutionary optimization for geostationary orbit satellite mission planning. *Journal of Systems Engineering and Electronics*, 2017, 28(5): 934–945. [doi: [10.21629/JSEE.2017.05.11](https://doi.org/10.21629/JSEE.2017.05.11)]
 - [69] Zhang XY, Liu J, Zhang R, Huang YD, Tong JC, Xin N, Liu L, Xiong ZH. Energy-efficient computation peer offloading in satellite edge computing networks. *IEEE Trans. on Mobile Computing*, 2024, 23(4): 3077–3091. [doi: [10.1109/TMC.2023.3269801](https://doi.org/10.1109/TMC.2023.3269801)]
 - [70] Chen Q, Meng WX, Quek TQS, Chen SY. Multi-tier hybrid offloading for computation-aware IoT applications in civil aircraft-augmented SAGIN. *IEEE Journal on Selected Areas in Communications*, 2023, 41(2): 399–417. [doi: [10.1109/JSAC.2022.3227031](https://doi.org/10.1109/JSAC.2022.3227031)]
 - [71] Bensana E, Verfaillie G, Agnese J C, Bataille N. Exact and inexact methods for daily management of earth observation satellite. In: Proc. of the 1996 Conf. on Space Mission Operations and Ground Data Systems. 1996. 394–507.
 - [72] Sarkheyli A, Bagheri A, Ghorbani-Vaghei B, Askari-Moghadam R. Using an effective tabu search in interactive resources scheduling problem for LEO satellites missions. *Aerospace Science and Technology*, 2013, 29(1): 287–295. [doi: [10.1016/j.ast.2013.04.001](https://doi.org/10.1016/j.ast.2013.04.001)]

- [73] Cheng XW, Lyu F, Quan W, Zhou CH, He HL, Shi WS, Shen XM. Space/aerial-assisted computing offloading for IoT applications: A learning-based approach. *IEEE Journal on Selected Areas in Communications*, 2019, 37(5): 1117–1129. [doi: [10.1109/JSAC.2019.2906789](https://doi.org/10.1109/JSAC.2019.2906789)]
- [74] Qin ZY, Yao HP, Mai TL, Wu D, Zhang N, Guo S. Multi-agent reinforcement learning aided computation offloading in aerial computing for the Internet-of-Things. *IEEE Trans. on Services Computing*, 2023, 16(3): 1976–1986. [doi: [10.1109/TSC.2022.3190562](https://doi.org/10.1109/TSC.2022.3190562)]
- [75] Kim H, Chang YK. Mission scheduling optimization of SAR satellite constellation for minimizing system response time. *Aerospace Science and Technology*, 2015, 40: 17–32. [doi: [10.1016/j.ast.2014.10.006](https://doi.org/10.1016/j.ast.2014.10.006)]
- [76] Cui JT, Zhang X. Application of a multi-satellite dynamic mission scheduling model based on mission priority in emergency response. *Sensors*, 2019, 19(6): 1430. [doi: [10.3390/s19061430](https://doi.org/10.3390/s19061430)]
- [77] Cesta A, Ocon J, Rasconi R, Sanchez Montero AM. Simulating on-board autonomy in a multi-agent system with planning and scheduling. 2010. https://www.researchgate.net/profile/Jorge-Ocon/publication/266222948_Simulating_On-Board_Autonomy_in_a_Multi-Agent_System_with_Planning_and_Scheduling/links/55487a080cf26a7bf4dac59b/Simulating-On-Board-Autonomy-in-a-Multi-Agent-System-with-Planning-and-Scheduling.pdf
- [78] Grey S. Distributed agents for autonomous spacecraft [MS. Thesis]. Glasgow: University of Glasgow, 2013.
- [79] Wang HJ. Massive scheduling method under online situation for satellites based on reinforcement learning [Ph.D. Thesis]. Beijing: University of Chinese Academy of Sciences (National Space Science Center, CAS), 2018 (in Chinese).
- [80] Chien S, McLaren D, Rabideau G, Silverman D, Mandl D, Hengemihle J, Microtel L. Onboard processing for low-latency science for the HyspIRI mission. 2010. https://www.researchgate.net/publication/229022061_Onboard_Processing_for_Low-latency_Science_for_the_HyspIRI_Mission
- [81] Agrawal J, Yelamanchili A, Chien S. Using explainable scheduling for the Mars 2020 rover mission. arXiv:2011.08733, 2020.

附中文参考文献

- [14] 九九. 下行速率超过 100 Mbps, Starlink 真的能取代 5G 吗? 2022. <https://www.51cto.com/article/704582.html>
- [62] 向尚, 陈盈果, 李国梁, 邢立宁. 卫星自主与协同任务调度规划综述. *自动化学报*, 2019, 45(2): 252–264. [doi: [10.16383/j.aas.c180068](https://doi.org/10.16383/j.aas.c180068)]
- [63] 赵萍, 陈志明. 应用于卫星自主任务调度的改进遗传算法. *中国空间科学技术*, 2016, 36(6): 47–54. [doi: [10.16708/j.cnki.1000-758X.2016.0064](https://doi.org/10.16708/j.cnki.1000-758X.2016.0064)]
- [67] 庞中华. 高低轨对地观测卫星协同任务规划方法研究 [硕士学位论文]. 哈尔滨: 哈尔滨工业大学, 2013.
- [79] 王海蛟. 基于强化学习的卫星规模化在线调度方法研究 [博士学位论文]. 北京: 中国科学院大学 (中国科学院国家空间科学中心), 2018.

作者简介

张其阳, 博士生, 主要研究领域为卫星计算, 服务计算, 边缘智能.

邢若邈, 博士生, CCF 学生会员, 主要研究领域为卫星计算, 5G/6G 核心网.

李元哲, 博士, CCF 专业会员, 主要研究领域为 5G/6G 核心网, 边缘计算.

周傲, 博士, 副教授, CCF 专业会员, 主要研究领域为卫星计算, 边缘计算, 云计算.

徐梦炜, 博士, 副教授, CCF 专业会员, 主要研究领域为卫星操作系统, 边缘计算.

王尚广, 博士, 教授, CCF 杰出会员, 主要研究领域为卫星计算, 服务计算, 边缘计算.