面向卫星智算网络应用的多模态语义理解大模型研训工作策划

中国卫通集团股份有限公司

2025年6月

目录

**[一、 示范项目 1](#_Toc29283)**

**[二、 场景想定 1](#_Toc20158)**

[（一）基本场景 1](#_Toc20701)

[（二）难点问题 4](#_Toc12315)

[（1）多模态数据融合与对齐 4](#_Toc8187)

[（2）多模态语义理解应用领域泛化 4](#_Toc11327)

**[三、 研究目标 4](#_Toc9655)**

**[四、 研究内容 5](#_Toc2225)**

[（一）多模态卫星通信场景设计 5](#_Toc13070)

[（二）多模态语义理解数据集制备 5](#_Toc13766)

[（三）多模态语义理解技术研究 5](#_Toc29986)

[（四）多模态语义理解原型系统搭建 6](#_Toc14768)

**[五、 主要技术途径 6](#_Toc6288)**

[（一）多模态卫星应用数据制备 6](#_Toc6407)

[（二）Qwen2.5-VL 7B大模型适配微调 6](#_Toc6226)

[（三）多模态语义融合与深度推理技术 7](#_Toc20623)

[（四）开放域多模态理解增强 7](#_Toc7403)

[（五）模型训练全流程设计 7](#_Toc21071)

**[六、 扩展支撑 9](#_Toc13738)**

[（一）与L1航天行业大模型的对接 9](#_Toc18481)

[（二）与L2中队多模态大模型的对接 9](#_Toc17734)

[（三）与中国卫通云端其他各智能应用的对接 10](#_Toc12477)

**[七、 预期成果与展示汇报 10](#_Toc1061)**

[（一）预期成果 10](#_Toc20153)

[（二）展示汇报 10](#_Toc19975)

**[八、 工作思路与工作计划 10](#_Toc27310)**

[（一）工作思路 10](#_Toc24052)

[（二）工作计划 11](#_Toc7583)

1. 示范项目

**项目名称：边云协同网络智算服务系统应用**

面向天空地海全域智能服务需求，构建以边云协同为技术基座、以感通算用一体化为应用模式的卫星智能服务体系，突破传统卫星“海量回传-地面处理-滞后响应”的效能瓶颈。聚焦海洋权益、林草防护、应急救灾、低空巡检等典型卫星应用场景，打造“星地实时感知-边缘智能处理-云端服务融合-应用精准灵活”的全链条服务范式，推动卫星应用由透明转发向智算服务跃升，为数字中国建设提供全天候、强时效、高可靠、自主可控的空天智能基础设施。

1. 场景想定

（一）基本场景

（1）场景描述

构建以Qwen2.5-VL 7B为核心的多模态语义体系，围绕森林火情、野生动物监测、海洋航运、工矿安全四大场景，融合多模态数据，实现全流程语义理解，支撑生态保护、海事监管等业务闭环。

（2）场景要素

边云协同智算网络包含感知设备、边缘智算系统、卫星通信终端、高轨通信卫星、卫星地面主站、卫通智算云服务系统、用户数据中心以及地面各环节有线网络连接等。

感知设备包含光学摄像头、红外相机、天基遥感等设备设施，具备1080P及以上光学图像与视频采集能力。

高轨通信卫星为中星26号高通量卫星，支持上行4-6Mbps、下行20-40Mbps通信速率，通过卫星通信终端与地面主站实现整个卫星通信链路的构建。

边缘智算系统、卫通智算云服务系统为智能化建设点，边缘智算系统基于瑞芯微RK3588，搭载6 TOPS算力，支持数据压缩、目标监测等智能化边缘应用。卫通智算云服务系统为大模型部署环境，对各类场景、各类应用回传的数据进行价值挖掘。

（3）运行流程

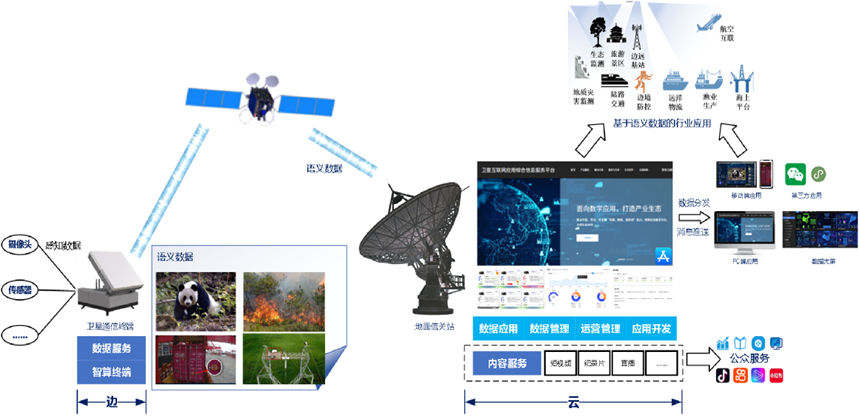
边云协同网络智算系统的运行流程为“数据采集→边缘处理→数据传输→云端处理→行业应用→应用反馈”6个环节。

感知采集：感知设备（光学摄像头、红外相机、天基遥感等）从不同角度进行信息采集，获取光学图像/视频数据，标记场景、时间、位置等元信息。

边缘预处理：基于边缘智能算力，对采集到的影像数据围绕“人事环物”等要素进行分析处理，并进行数据压缩，减少无效数据传输。

云端深度处理：云端智算服务系统基于接收到的影像数据进行粗粒度语义提取，快速提取场景分类、核心物体、事件类型等语义，再分发给各场景大模型进行细粒度分析。

应用反馈与闭环优化：云端分析结果以API、SaaS服务、系统模组等方式推送至用户数据中心或业务管理系统，支撑用户的场景化应用，并基于应用端反馈的信息处理效果，反向驱动模型迭代更新，实现“数据 - 模型 - 应用”的闭环进化。



|  |  |  |
| --- | --- | --- |
| **输入** | **粗粒度分析**  **任务分发** | **细粒度分析**  **处理** |
| IMG_256 | 基于多模态语义大模型的粗粒度内容理解，生成影像语义描述、场景分类和后继分析方法 | 分发给动物分析大模型及小模型集、用户业务系统 |
| IMG_256 | 分发给林火防控大模型及小模型集、用户业务系统 |
| IMG_256 | 分发给海洋生产大模型及小模型集、用户业务系统 |
| IMG_256 | 分发给海洋航运大模型及小模型集、用户业务系统 |

（二）难点问题

（1）多模态数据融合与对齐

卫星多模态数据（光学/红外/遥感影像、文本描述等）存在时空分辨率差异，需解决跨模态时空基准统一问题。图像与文本模态的语义表达差异显著，产生语义鸿沟。同时，关键影像数据缺失、噪声扰动等问题加剧融合难度。为此，开发多模态时空特征编码算法，构建时空配准模型；利用对比学习训练多模态编码器，缩小语义鸿沟；设计多模态动态融合框架，实现“数据-模型-任务”适配。

（2）多模态语义理解应用领域泛化

卫星应用场景存在显著领域偏移，模型在新领域语义理解精度可能会骤降；小众场景数据稀缺，传统训练方式泛化能力不足；卫星场景动态演化也会导致模型语义理解滞后。为此，构建卫星场景元学习框架，通过小样本快速适配新领域，实现跨领域语义迁移，开发领域适配评估指标。开发在线学习机制，实时更新模型参数，结合注意力机制捕捉时序特征，降低动态场景误判率。

1. 研究目标

针对卫星网络应用中多模态影像数据的多元化应用需求，聚焦语义理解、场景分类、任务分发与专业模型深度分析等关键技术环节，开展多源异构数据融合对齐、多模态语义联合表征与场景知识识别等研究，突破多模态语义理解与知识提取、领域知识引导的任务分发与资源优化等技术，构建基于语义理解的卫星智算网络智能调度平台，面向海洋、林草等典型应用场景进行验证，形成“多模态语义大模型+领域专家知识”的双驱智能体系，具备多模态数据价值提取、语义检索等智能信息服务能力。

1. 研究内容

（一）多模态卫星通信场景设计

基于智算网络多模态语义理解与任务分发应用需求，分析海洋应用、森林防火、动物保护等典型卫星应用场景的多模态数据特性，定义场景分类规则和数据价值评估指标。设计端到端处理流程，实现从应用数据传输、多模态语义理解、场景分类分发、专业模型分析到信息服务生成的全周期服务，提升系统服务时效性与算力分配合理性。

（二）多模态语义理解数据集制备

聚焦多模态语义理解数据集的系统性制备，基于开源数据集完成原始数据收集，结合数据增强技术提升数据多样性，基于开源标注工具X-AnyLabeling完成数据标注工作，支持图像-文本对齐任务，设计数据分层存储架构，并通过覆盖度、一致性等多维度指标进行质量评估，形成支撑多模态语义理解任务的标准化数据集。

（三）多模态语义理解技术研究

通过Pytorch + Transformers技术路线，基于Qwen2.5-VL 7B多模态大模型为基础，构建适配卫星应用场景的语义理解体系。从山火防控与动物监测场景入手，通过“场景化微调-能力验证-复杂任务扩展”的递进式研究路径，突破模态对齐效率、语义推理深度与算力资源平衡等关键问题。

（四）多模态语义理解原型系统搭建

基于多模态语义理解技术研究成果与标准化数据底座，构建卫星应用“数据智能解析-场景化语义理解-智能化任务分发”原型系统，实现从多源卫星数据输入到行业应用价值输出的全链路贯通，为海洋应用、山火防控、动物监测、工矿监管等场景提供智能化支撑。

1. 主要技术途径

（一）多模态卫星应用数据制备

整合森林烟火、动物活动、海洋航行、工矿生产等应用场景的光学图像、光学视频、红外图像、光学遥感等原始数据，构建跨场景多模态数据集，有效样本量≥10万。

通过旋转、裁剪、色彩变换、噪声添加、目标叠加、仿真等方式进行图像类数据的增强。对文本进行同义词替换、句式改写，提升模型泛化能力。对长文本进行分段处理，结合滑动窗口技术保留上下文语义，避免信息截断。

（二）Qwen2.5-VL 7B大模型适配微调

将卫星应用多模态影像数据按分辨率、频段特性等要素进行归一化处理，适配模型视觉编码器，文本数据结合卫星应用词典进行token化，构建多模态输入序列。采用LoRA技术，冻结模型80%以上预训练参数，对注意力层、输出层的低秩矩阵进行微调，结合算力消耗实现对模型语义理解通用性平衡。结合卫星应用场景，设计多任务损失函数，强化“卫星影像特征-场景语义-决策关联”，使模型精准理解“森林火点分布-火情等级-扑救策略”、“动物活动轨迹-种群状态-栖息地状态”等语义链路。

（三）多模态语义融合与深度推理技术

构建卫星应用场景跨模态注意力网络，基于影像的时空特性与文本的语义逻辑，动态调整注意力权重，解决多模态数据的“模态异构鸿沟”问题，提升特征融合精度。集成行业知识图谱，实现卫星应用“影像数据采集-领域知识关联-智能决策生成”的深度语义理解。

（四）开放域多模态理解增强

在场景数据集中混入10%-20%开放世界样本，提升模型学习开放场景的语义逻辑。引入开放世界识别头，训练模型区分“已知类别”和“未知类别”，适配卫星应用场景中频繁出现的新目标识别需求。同时，搭建开放知识引擎，解决静态模型“知识固化”问题，支撑长期卫星场景应用。

（五）模型训练全流程设计

（1）分阶段训练策略

场景微调阶段：针对森林火情、动物活动等核心场景，采用“端到端 + 任务分解”训练模式：

端到端训练：直接输入影像 - 文本对，学习“影像特征→语义决策”的直接映射；

任务分解训练：先通过目标检测子模块定位火点区域，再与语义分类模块联合优化，提升细粒度理解能力。

泛化增强阶段：混入15%开放世界样本（如极地影像、新型工业设施），通过“已知/未知类别”分类头训练模型泛化能力，新领域适配精度≥80%。

（2）优化器与学习率调度

优化算法：采用AdamW优化器（=0.9，=0.999），结合梯度裁剪防止参数震荡，确保训练稳定性。

学习率策略：

预热阶段：线性升温至 5e-5，避免初始训练梯度爆炸；

主体阶段：采用余弦衰减策略降至1e-6，配合早停机制，验证集损失连续3 轮无下降则终止训练，防止过拟合。

（3）分布式训练架构

数据并行策略：基于4 × A100 GPU进行数据并行训练，通过NCCL通信框架保证批量同步效率≥90%，单轮训练时间缩短至非分布式的 1/4。

算力动态调度：对接卫通云端算力管理平台，根据训练阶段自动调整资源配置，提升资源利用率。

（4）闭环迭代优化流程

应用反馈收集：通过行业用户（如林草局、海事部门）采集真实场景中的误判样本，建立难例样本库。

增量训练机制：采用EM更新模型参数，对难例样本进行重标注后纳入训练集，通过增量学习方式保留历史知识并适应新场景，后续将每季度完成一次模型迭代。

模型压缩部署：通过知识蒸馏技术将7B大模型压缩至1.3B参数，结合量化技术（INT8）进一步降低边缘端算力需求，确保在6 TOPS边缘设备上高效运行。

1. 扩展支撑

（一）与L1航天行业大模型的对接

**标准化数据接口**：基于Restful API构建多模态语义元数据推送机制，采用JSON 格式封装卫星影像特征向量（如视觉语义编码、时空坐标）及文本标注信息，实现与 L1 大模型的标准化数据交互。

**模型能力协同调用**：预留模型并行推理接口，当遇到复杂场景时，触发L1大模型的全球尺度全场景行业语义推理能力，与本模型的子场景分析能力形成互补，提升跨场景语义理解精度。

（二）与L2中队多模态大模型的对接

**数据共享与协同标注：**与L2中队建立多模态数据共享平台，互通海洋监测、森林防火、无人机、遥感等场景的原始数据，联合开展跨模态数据标注，构建更全面的行业数据集。

**模型共建与能力协同：**L2中队聚焦天基/空基影像判读特殊性，构建航天领域多模态模型。卫通多模态模型聚焦天空地海一体化场景，采集数据涵盖地表固定与移动影像。可预留模型并行推理接口，当遇到天基、空基应用场景时，可直接调用L2中队多模态大模型能力。

（三）与中国卫通云端其他各智能应用的对接

API集成与业务联动：通过微服务架构集成卫通云端的船舶追踪、气象预警等智能应用，将多模态语义理解结果实时推送至对应业务系统，形成“语义理解-业务决策” 闭环。

算力资源弹性调度：对接卫通云端的算力管理平台，根据多模态模型训练及推理需求，动态申请/释放GPU资源。

1. 预期成果与展示汇报

（一）预期成果

模型成果：适配卫星场景的多模态语义理解大模型（精度≥90%）。

数据成果：标准化多模态数据集（样本量≥10万，标注覆盖率≥95%）。

原型系统：用于多模态语义理解大模型应用演示。

其他：原型系统技术文档、部署文档与应用手册。

（二）展示汇报

对模型训练成果进行原型系统展示，结合典型应用场景进行展示，重点汇报智能模型解决卫星应用数据泛化性上的能力。

1. 工作思路与工作计划

（一）工作思路

（1）技术分层推进

**数据层**：三家单位联合构建多模态数据治理中台，统一采集标准、标注规范，确保训练数据一致性。

**模型层**：以Qwen2.5-VL 7B为基础，采用“主干模型共享 + 场景适配器独立训练”模式，各单位针对海洋、林草等场景开发专用适配器，实现“统一底座 + 差异化能力”。

**应用层**：搭建原型系统联合调试，通过“边云协同测试床”模拟卫星链路，验证模型推理延迟与精度。

（2）人才协同机制

大模型训练组（6人）负责数据预处理、模型微调，每周输出模型评估报告。

算法验证组（4人）复现 OLMO、LLaVa 等模型积累经验，每两周开展技术分享会，同步核心知识至训练任务。

（二）工作计划

围绕“数据-模型-系统”核心链路，以6月11日起，60天为周期，分三阶段攻坚：

第一阶段（前10天）：完成多模态数据采集、预处理与标注，构建训练基础；

第二阶段（中间44天）：聚焦模型适配、微调与能力验证，突破核心技术；

第三阶段（后6天）：搭建原型系统并联调，实现场景化验证与迭代闭环。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 序号 | 工作阶段 | 工作计划 | 计划描述 | 责任人 | 完成时间 | 备注 |
| 1 | 数据采集与预处理 | 多模态数据采集与增强 | 整合森林火情、海洋监测等场景的光学 / 红外影像、文本数据，完成≥10 万有效样本采集；通过旋转、裁剪等方式增强图像数据，对文本进行同义词替换（改写率≥25%）。 | 王运韬 | 2025.06.01  -  2025.06.10 |  |
| 2 | 工作策划 | 完成工作策划编制与研讨 | 针对分队任务、实训任务目标，基于导师与技术委员会建议，完成工作策划修改。 | 王运韬 | 2025.06.11  -  2025.06.17 |  |
| 3 | 数据集标注与验证 | 标准化数据集制备 | 使用 X-AnyLabeling完成10万样本标注，构建图像 - 文本对齐标注体系；设计数据分层存储架构，通过覆盖度、一致性指标评估数据质量。 | 赵宇琛 | 2025.06.16  -  2025.06.26 | 标注规则需与模型团队同步，避免标签偏差 |
| 4 | 模型基础适配 | Qwen2.5-VL 7B 适配改造 | 卫星影像数据归一化处理，适配视觉编码器；文本结合卫星词典token化，构建多模态输入序列；测试数据加载成功率≥95%。 | 杨舒茜 | 2025.06.27  -  2025.07.05 | 需 GPU 集群支持（≥4 块 A100），确保环境稳定性 |
| 5 | 场景化模型微调 | 多模态语义理解模型训练 | 采用LoRA技术冻结80%参数，针对森林火情场景微调；设计多任务损失函数，强化“影像特征 - 场景语义”关联，验证火情识别精度≥85%。 | 王运韬 | 2025.07.06  -  2025.07.21 | 监控训练 Loss 曲线，避免过拟合；需行业专家参与损失函数设计 |
| 6 | 跨域融合与泛化增强 | 多模态语义泛化能力优化 | 构建跨模态注意力网络，集成行业知识图谱，解决模态异构问题；混入15%开放世界样本，训练模型区分已知/未知类别，新领域适配精度≥80%。 | 王运韬 | 2025.07.22  -  2025.07.31 | 需生成稀缺场景数据（如极地影像），扩充训练集 |
| 7 | 原型系统搭建 | 边云协同原型系统开发 | 设计“数据解析 - 语义理解 - 任务分发” 架构；完成云端演示系统搭建，实现数据传输闭环。 | 王运韬 | 2025.08.01  -  2025.08.05 | 联调前完成单元测试，确保模块接口兼容性 |
| 8 | 模型与系统联调 | 多模态语义理解全链路验证 | 打通“数据采集 - 边缘处理 - 云端推理” 流程；对接L2中队模型，验证天基 - 空基协同能力。 | 王运韬 | 2025.08.06  -  2025.08.10 |  |
| 9 | 场景化验证与优化 | 行业场景落地验证 | 在森林火情、海洋航运场景测试全流程；收集应用反馈，优化模型精度至≥90%，输出场景验证报告 | 杨舒茜  赵宇琛 | 2025.08.10  -  2025.08.12 | 需真实场景数据支撑，协调行业用户参与测试 |
| 10 | 成果整理与验收 | 技术成果汇总与展示 | 整理模型参数、系统代码、测试报告；准备演示材料，展示多模态语义理解能力（如跨模态时空对齐精度≤50 米） | 王运韬 | 2025.08.12  -  2025.08.15 |  |