**一、研究目标**

针对卫星网络应用中多模态影像数据的传输、分析与应用需求，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链路，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，突破多源异构数据融合、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，构建基于多模态语义理解的卫星智算网络智能调度平台，开展海洋、林草、动保等场景的验证，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。

**二、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。

**三、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、仿真等方式增强图像数据；对文本进行同义词替换、句式改写，提升模型泛化能力；对长文本分段处理，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**

将卫星影像数据归一化处理，适配视觉编码器；文本数据结合卫星词典 token 化，构建多模态输入序列。采用 LoRA 技术冻结 80% 预训练参数，微调低秩矩阵，设计多任务损失函数，强化 “影像特征 - 场景语义 - 决策关联”，精准理解 “森林火情”“动物活动” 等语义链路。

**（三）多模态语义跨域融合与深度推理技术**

构建跨模态注意力网络，动态调整时空 - 语义权重，解决模态异构鸿沟；集成行业知识图谱，实现 “影像采集 - 知识关联 - 智能决策” 的深度语义理解，提升特征融合精度。

**（四）开放域多模态理解增强**

混入 10%-20% 开放世界样本，训练模型区分已知 / 未知类别，适配新目标识别；搭建开放知识引擎，解决模型知识固化问题，支撑长期卫星场景应用。

**四、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**

* **边云协同网络**：含感知设备（光学 / 红外 / 遥感）、边缘智算（瑞芯微 RK3588，6TOPS 算力）、中星 26 号卫星（上行 4-6Mbps，下行 20-40Mbps）、卫通智算云（大模型部署环境）等。
* **感知能力**：1080P 及以上图像 / 视频采集，支持数据压缩与目标监测。

**（3）运行流程**

* **数据采集→边缘预处理→数据传输→云端处理→行业应用→应用反馈**，实现 “数据 - 模型 - 应用” 闭环进化。

**（二）难点问题**

**（1）多模态数据融合与对齐**

解决时空基准统一、语义鸿沟、数据缺失 / 噪声问题，开发时空配准模型、对比学习编码器、动态融合框架。

**（2）多模态语义理解应用领域泛化**

（此处原文未完整，需补充具体技术问题与攻关方向）

**五、扩展支撑**

**（一）与 L1 级航天行业大模型对接**

* **标准化数据接口**：Restful API 推送多模态语义元数据（JSON 格式）。
* **模型能力调用**：预留接口，设想引接后工作模式（如任务分发、语义增强分析）。

**（二）与开放场景多模态数据对接**

预留与 XXX 仿真平台接口，设想引接后工作模式（如数据融合验证、场景推演）。

**六、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。

七、计划

围绕 “数据 - 模型 - 系统” 核心链路，以 50 天为周期分三阶段攻坚：  
**第一阶段（前 15 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 20 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 15 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** | **计划描述** | **责任人** | **完成时间** | **关键交付物 / 验证指标** | **备注** |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | 数据采集与增强 | 1. 整合森林、海洋、工矿场景原始数据（光学 / 红外 / 文本），目标≥4.5 万有效样本； 2. 用旋转、裁剪增强图像（≥3 种变换），文本同义词替换（≥20% 改写率）。 | 赵宇琛 | 第 1-8 天 | 多模态数据集（样本量≥4.5 万） | 协调卫星数据接口权限 |
| 2 | 数据标注规范 | 1. 用 X-AnyLabeling 工具制定标注规则（图像 - 文本对齐、场景分类标签）； 2. 完成首批 1.8 万样本标注，验证覆盖度≥90%。 | 杨舒茜 | 第 6-13 天 | 标注规则文档、标注数据集（1.8 万样本） | 与模型团队同步标签需求 |
| 3 | 模型基础适配 | 1. 基于 Qwen2.5-VL 7B，冻结 80% 参数，配置 LoRA 微调环境（Pytorch+GPU）； 2. 测试多模态输入流程，验证数据加载成功率≥95%。 | 王运韬 | 第 11-15 天 | 模型适配环境、基础推理流程 | 依赖 GPU 算力资源 |
| 4 | 场景化模型微调 | 1. 针对森林火情场景，用增强数据微调模型（学习率 1e-5，批次 16）； 2. 验证火情识别精度≥85%，模态对齐误差≤10 像素。 | 王运韬 | 第 16-28 天 | 微调后模型（火情精度≥85%） | 监控训练 Loss 曲线 |
| 5 | 系统框架搭建 | 1. 设计 “数据解析 - 语义理解 - 任务分发” 原型系统架构； 2. 开发边缘端（RK3588）与云端基础模块，实现数据传输闭环。 | 王运韬 | 第 26-32 天 | 系统架构文档、端云基础模块 | 联调前单元测试 |
| 7 | 场景化验证 | 1. 在森林火情、动物监测场景测试全流程（数据采集→语义输出→决策）； 2. 输出验证报告，优化模型精度至≥90%。 | 王运韬 | 第 41-47 天 | 场景验证报告（精度≥90%） | 需真实场景测试数据 |
| 8 | 成果收尾优化 | 1. 整理模型参数、系统代码、测试报告； 2. 修复关键问题，完成最终版本迭代。 | 杨舒茜 | 第 48-50 天 | 完整技术文档、可部署系统包 | 同步准备展示材料 |

**多模态语义理解在卫星智算网络中的研究计划**

**一、研究目标**

在当下卫星网络应用场景中，多模态影像数据的高效运用已成为推动行业发展的关键驱动力。但该领域面临着多模态影像数据传输效率低、分析手段有限以及应用场景拓展困难等诸多挑战。为突破这些瓶颈，本研究将全力聚焦于语义理解、场景分类、任务分发以及专业模型深度分析等核心技术链路。

通过开展多源异构数据融合对齐研究，消除不同来源、不同类型数据间的差异，实现数据的无缝对接；推进多模态语义联合表征与场景知识识别研究，挖掘数据深层语义关联，精准识别各类场景知识。以此为基础，全力突破多源异构数据融合、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关键技术。

进而构建基于多模态语义理解的卫星智算网络智能调度平台，该平台将具备强大的数据处理与调度能力。同时，针对海洋、林草、动保等典型场景展开深度验证，融合多模态语义大模型的强大数据处理能力与领域专家知识的专业经验，形成创新的双驱智能体系。最终使该体系具备高效的多模态数据价值提取能力，能够从海量复杂数据中精准提炼关键信息，实现精准的语义检索等智能信息服务，为卫星网络应用开拓全新的发展局面。

**二、研究内容**

**（一）多模态卫星通信场景设计**

智算网络的发展对多模态语义理解与任务分发提出了更高要求。本部分将深入剖析海洋应用、森林防火、动物保护等典型卫星应用场景下多模态数据的特性，这些场景的数据具有来源广泛、格式多样、时空特性复杂等特点。基于此，定义清晰、准确的场景分类规则，比如依据海洋应用中的船只航行轨迹、水体颜色变化等特征进行分类；同时制定科学合理的数据价值评估指标，像以数据对场景监测的关键程度、数据的准确性和时效性等作为评估标准。

精心设计端到端处理流程，从应用数据的采集传输开始，通过高效的传输协议保障数据快速、稳定地到达处理端；接着进行多模态语义理解，运用先进算法挖掘数据语义；再依据分类规则实现场景分类分发，将数据精准导向相应的专业模型进行深入分析；最后生成有价值的信息服务并反馈应用。通过这样的全周期服务流程，大幅提升系统服务时效性，减少数据处理延迟，优化算力分配合理性，避免算力浪费，提升整体系统性能。

**（二）多模态语义理解数据集制备**

多模态语义理解数据集的质量直接影响模型的性能。本研究将基于开源数据集广泛收集原始数据，涵盖光学图像、文本描述、视频片段等多类型数据，构建丰富的数据基础。运用数据增强技术，对图像进行旋转、裁剪、色彩变换，为模型训练提供更多样化的数据，提升模型泛化能力；对文本进行同义词替换、句式改写，丰富文本表达形式。

利用开源标注工具 X - AnyLabeling 完成数据标注工作，精准标注图像中的目标物体、文本对应的图像内容等，支撑图像 - 文本对齐任务。设计合理的数据分层存储架构，按照数据的重要程度、使用频率、场景类别等进行分层，提高数据访问效率。并通过覆盖度、一致性等多维度指标对数据集进行严格质量评估，如检查标注是否覆盖所有关键数据、不同标注员标注结果是否一致等，形成高质量、标准化的数据集，为多模态语义理解任务提供坚实的数据支撑。

**（三）多模态语义理解技术研究**

采用当下先进的 Pytorch + Transformers 技术路线，以 Qwen2.5 - VL 7B 多模态大模型为基石，构建高度适配卫星应用场景的语义理解体系。选择山火防控与动物监测场景作为切入点，这两个场景对卫星数据的实时性、准确性要求极高。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，逐步优化模型性能。

在场景化微调阶段，利用特定场景数据对模型进行微调，强化模型对该场景特征的识别能力；能力验证阶段，通过一系列指标评估模型在该场景下的语义理解、目标识别等能力；复杂任务扩展阶段，增加任务难度和复杂度，如在山火防控中加入火势蔓延预测等任务，突破模态对齐效率瓶颈，实现不同模态数据快速、精准对齐；加深语义推理深度，让模型能够理解复杂语义关系；平衡算力资源，在保障模型性能的同时，合理分配计算资源，避免算力浪费。

**（四）多模态语义理解原型系统搭建**

基于前面多模态语义理解技术研究成果以及标准化数据底座，构建卫星应用的 “数据智能解析 - 场景化语义理解 - 智能化任务分发” 原型系统。该系统能够对多源卫星数据进行智能解析，将原始数据转化为可处理的格式；通过场景化语义理解模块，准确理解数据在不同场景下的语义；依据智能化任务分发机制，将处理任务合理分配到不同的计算资源上。

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，为海洋应用中的海洋资源监测、山火防控中的火灾预警、动物监测中的动物种群动态跟踪、工矿监管中的安全生产监测等场景提供智能化支撑，推动卫星应用在各行业的深入发展。

**三、主要技术途径**

**（一）多模态卫星应用数据制备**

全面整合森林烟火、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，确保有效样本量≥10 万，保证数据的丰富性和全面性。

运用多种方式增强图像数据，通过旋转改变图像角度，模拟不同拍摄视角；裁剪突出关键区域，提升模型对局部特征的识别能力；进行色彩变换，丰富图像色彩特征；添加噪声，增强模型对真实环境中噪声数据的适应性；叠加目标，增加数据的复杂程度；利用仿真技术生成模拟数据，补充真实数据的不足。对于文本数据，采用同义词替换丰富词汇表达，改写句式调整文本结构，提升模型对不同文本表达方式的理解能力。针对长文本，合理分段处理，保留上下文语义，避免信息截断，确保模型能够准确理解长文本含义。

**（二）Qwen2.5 - VL 7B 大模型适配微调**

对卫星影像数据进行归一化处理，统一数据格式和范围，适配视觉编码器，使其能够更好地处理卫星影像特征。将文本数据结合卫星词典进行 token 化，将文本转化为模型可理解的 token 序列，构建多模态输入序列。采用 LoRA 技术冻结 80% 预训练参数，减少训练参数量，降低计算成本，仅对低秩矩阵进行微调。设计多任务损失函数，综合考虑影像特征与场景语义的匹配度、模型预测结果与真实标签的差异等因素，强化 “影像特征 - 场景语义 - 决策关联”，使模型能够精准理解 “森林火情”“动物活动” 等复杂语义链路，提升模型在卫星应用场景下的性能。

**（三）多模态语义跨域融合与深度推理技术**

构建跨模态注意力网络，该网络能够动态调整时空 - 语义权重，根据不同模态数据的特点和任务需求，自动分配注意力资源，有效解决模态异构鸿沟，实现不同模态数据的深度融合。集成行业知识图谱，将行业内的专业知识结构化，与多模态数据相结合，实现 “影像采集 - 知识关联 - 智能决策” 的深度语义理解，提升特征融合精度，使模型能够利用行业知识进行更准确的推理和决策。

**（四）开放域多模态理解增强**

在训练数据中混入 10%-20% 开放世界样本，这些样本包含未知类别和新的场景，训练模型区分已知 / 未知类别，使其具备对新目标的识别能力，能够快速适应开放域中的新情况。搭建开放知识引擎，持续更新和补充模型的知识储备，解决模型知识固化问题，让模型能够跟上卫星应用场景不断变化的需求，支撑长期卫星场景应用，保持模型的长期有效性和适应性。

**四、场景想定**

**（一）基本场景**

**1. 场景描述**

以 Qwen2.5 - VL 7B 为核心构建多模态语义体系，围绕森林火情、野生动物监测、海洋航运、工矿安全四大场景展开。在这些场景中，充分融合光学、红外、遥感等多模态数据，实现从数据采集到最终决策的全流程语义理解。通过准确理解数据语义，为生态保护提供野生动物数量变化、栖息地状况等信息；为海事监管提供船舶航行轨迹、海洋气象等数据支持，形成完整的业务闭环，推动相关行业的智能化发展。

**2. 场景要素**

* **边云协同网络**：由多种设备和平台构成，感知设备包含光学相机用于获取高分辨率图像、红外设备探测温度异常、遥感设备提供大面积监测数据；边缘智算采用瑞芯微 RK3588，具备 6TOPS 算力，可在数据采集端进行初步处理，减轻传输压力；中星 26 号卫星负责数据的远距离传输，上行速率 4 - 6Mbps，下行速率 20 - 40Mbps，保障数据可靠传输；卫通智算云作为大模型部署环境，提供强大的计算资源支撑模型运行。
* **感知能力**：具备 1080P 及以上图像 / 视频采集能力，能够清晰捕捉目标物体细节。支持数据压缩，在不损失关键信息的前提下减少数据量，便于传输；具备目标监测能力，能够实时检测出感兴趣目标，如森林中的火源、野生动物的活动踪迹等。

**3. 运行流程**

遵循 “数据采集→边缘预处理→数据传输→云端处理→行业应用→应用反馈” 的流程。数据采集设备获取多模态数据后，在边缘端进行数据清洗、压缩、初步特征提取等预处理工作；通过卫星链路将数据传输至云端；云端利用多模态语义理解模型进行深度处理和分析；处理结果应用于相应行业场景，如森林部门根据火情分析结果制定灭火方案；行业应用过程中的反馈信息再次进入系统，用于优化数据采集、模型训练等环节，实现 “数据 - 模型 - 应用” 闭环进化，不断提升系统性能。

**（二）难点问题**

**1. 多模态数据融合与对齐**

多模态数据存在时空基准不统一问题，不同设备采集数据的时间和空间坐标系可能不同，导致数据难以直接融合。语义鸿沟也是一大挑战，不同模态数据的语义表达方式差异大，如文本描述与图像特征之间的语义关联难以建立。数据缺失 / 噪声问题也较为常见，部分数据可能因设备故障、传输干扰等原因丢失或包含错误信息。

为解决这些问题，将开发时空配准模型，通过算法统一不同模态数据的时空基准；构建对比学习编码器，挖掘不同模态数据间的相似性和差异性，缩小语义鸿沟；设计动态融合框架，根据数据特点和任务需求，灵活调整融合策略，有效处理数据缺失和噪声问题，实现多模态数据的高质量融合与对齐。

**2. 多模态语义理解应用领域泛化**

随着应用场景的不断拓展，模型面临在新领域和复杂场景下泛化能力不足的问题。一方面，新领域数据与训练数据分布存在差异，导致模型难以准确识别和理解新场景中的语义。例如在一些特殊地理环境或新兴应用场景中，数据特征与传统训练数据大不相同，模型预测精度大幅下降。另一方面，数据稀缺问题在某些小众领域尤为突出，样本量过少使得模型无法学习到足够的特征和规律，难以构建有效的语义理解模型。

针对这些问题，将采用迁移学习技术，利用在相似领域已训练好的模型参数，快速适应新领域数据，降低对大规模新数据的依赖。同时，结合生成对抗网络（GAN）技术，生成稀缺领域的虚拟数据，扩充数据集规模，提升模型泛化能力。此外，设计自适应学习算法，使模型能够在实际应用中根据新数据不断调整优化，持续提升在不同领域的语义理解能力 。

**五、扩展支撑**

**（一）与 L1 级航天行业大模型对接**

* **标准化数据接口**：采用 Restful API 技术搭建标准化数据接口，以 JSON 格式推送多模态语义元数据。这种格式具有良好的可读性和通用性，能够方便地与各类系统进行交互。通过该接口，可将本研究生成的多模态语义数据准确、高效地传输至 L1 级航天行业大模型，实现数据共享与融合。
* **模型能力调用**：预留模型能力调用接口，设想引接后，当遇到复杂任务或需要更广泛知识支持时，可调用 L1 级航天行业大模型的强大能力。如在处理全球性航天监测任务时，借助其更全面的知识体系和更强的计算能力进行语义增强分析，提升任务处理的准确性和效率；在任务分发方面，依据 L1 级大模型的分析结果，更合理地分配任务，优化资源利用。

**（二）与开放场景多模态数据对接**

预留与 XXX 仿真平台接口，实现与开放场景多模态数据的对接。引接后，可利用仿真平台生成的多样化数据对模型进行训练和验证，拓展模型的适用范围。例如在模拟极端天气下的海洋监测场景、复杂地形中的林草监测场景等，通过数据融合验证，检验模型在不同复杂情况下的性能，发现模型存在的问题并加以改进。同时，基于对接的数据进行场景推演，提前预测不同场景下可能出现的情况，为实际应用提供决策参考，提升系统应对复杂多变场景的能力。

**六、预期成果与展示汇报**

**（一）预期成果**

* **模型成果**：成功构建适配卫星场景的多模态语义理解大模型，在关键场景的语义理解任务中，精度≥90%，能够准确识别和理解卫星影像中的各类目标和语义信息。
* **数据成果**：形成标准化多模态数据集，样本量≥10 万，标注覆盖率≥95%，为多模态语义理解研究和相关应用提供高质量的数据基础。
* **原型系统**：开发出用于多模态语义理解大模型应用演示的原型系统，能够直观展示模型在卫星数据处理、场景语义理解、任务分发等方面的功能和效果。
* **其他**：产出原型系统技术文档，详细记录系统架构、技术实现细节等；完成部署文档，指导系统在不同环境下的部署；编写应用手册，方便用户了解和使用系统，推动研究成果的实际应用和推广。

**（二）展示汇报**

结合卫星智算云平台，展示系统在实际卫星数据处理中的运行情况，包括数据传输速度、处理效率等指标。利用 X - AnyLabeling 工具，现场演示数据标注过程以及标注数据对模型训练的支持作用。重点汇报 “跨模态时空对齐” 技术如何实现不同模态数据在时空维度上的精准匹配；“开放域识别” 技术怎样使模型具备处理未知类别和新场景的能力。通过实际案例和数据对比，展示模型和系统在森林火情防控、动物监测等场景中的卓越性能，如在森林火情防控中，模型能够提前 X 小时准确预警，在动物监测中，对动物种类识别准确率达到 X% 等，让观众直观感受研究成果的实际价值和创新性 。